Raman Microspectroscopy for the Discrimination of Thyroid and Lung Cancer Subtypes for Application in Clinical Cytopathology

Declan O’Dea
Technological University Dublin

Follow this and additional works at: https://arrow.tudublin.ie/sciendoc

Part of the Medicine and Health Sciences Commons

Recommended Citation

This work is licensed under a Creative Commons Attribution-Noncommercial-Share Alike 4.0 License
Raman microspectroscopy for the discrimination of thyroid and lung cancer subtypes for application in clinical cytopathology

Declan O’Dea BSc. (Hons)

Thesis submitted in partial fulfilment for the degree of Doctor of Philosophy (PhD)

Supervised by:

Alison Malkin

Prof. Fiona M Lyng

School of Biological and Health Sciences, Technological University Dublin

And

Centre for Radiation and Environmental Science FOCAS Research Institute

Technological University Dublin

February 2020
Abstract

The branch of cytology known as cytopathology, studies and diagnoses diseases at a cellular level, and is a useful method for detecting cancer. The procedures used to attain cytological samples for diagnostic purposes, such as aspiration and exfoliative methods are safe, accurate and cost-effective. Histochemical and immunohistochemical (IHC) techniques are commonly applied to cytological samples to aid cancer diagnosis, however multiple limitations occur using these methods for the diagnosis of thyroid cancer (TC) and non-small cell lung cancer (NSCLC). Fine needle aspiration cytology (FNAC) is the prominent diagnostic method used for the initial investigation of thyroid nodules but is limited by the inability to accurately diagnose malignancy in follicular-patterned lesion. As a result, more than 20% of cases under investigation for TC are classified as cytologically “indeterminate”, requiring surgical resection for accurate diagnosis. In the case of NSCLC, with the advent of targeted therapies, it is imperative to accurately differentiate (NSCLC) subtypes in order to ensure efficacy of treatment for patients. Immunohistochemistry and molecular techniques are increasingly part of the diagnostic work-up of NSCLC patients however due to the limitation of small sample size, overlapping morphological features and molecular characterisation, differential diagnosis of NSCLC still proves challenging. Raman spectroscopy has shown promising results for the detection of a variety of cancers, however to date there has been no evaluation of Raman spectroscopy on cytology bronchoscopy samples or thyroid FNAC samples, which may eliminate the limitations of current methods. This thesis explores the use of Raman spectroscopy as an alternative or adjunct tool for the diagnosis of TC and NSCLC using cytological specimens.
Declaration

I certify that this thesis which I now submit for examination for the award of Doctor of Philosophy (PhD), is entirely my own work and has not been taken from the work of others, save and to the extent that such work has been cited and acknowledged within the text of my work.

This thesis was prepared according to the regulations for graduate study by research of the Technological University Dublin and has not been submitted in whole or in part for another award in any other third level institution.

The work reported on in this thesis conforms to the principles and requirements of the TU Dublin's guidelines for ethics in research.

TU Dublin has permission to keep, lend or copy this thesis in whole or in part, on condition that any such use of the material of the thesis be duly acknowledged.

Signature __________________________ Date ____________

(Declan O’Dea)
Acknowledgements

I would like to express my sincere appreciation and gratitude for my supervisors, Ms. Alison Malkin and Prof. Fiona Lyng, for their support and advice throughout. I thank them for their guidance, for all the time and effort they put into helping me with this project, and for helping me grow as a scientist.

I would also to thank everybody in the RESC for their help along the way, in particular I’d like to thank Damien Traynor, who taught me how to take a Raman spectrum and whose help has been invaluable throughout. Thank you to Daniel Cullen, Adrian Maguire, and Sebastian Nagorski who were willing to give up their time to help me with the data mining aspect of the project.

This work would not have been possible without our collaborators, Dr. Finbar O’Connell, Dr. Siobhán Nicholson, and Dr. Aoife Maguire from St. James’s Hospital Dublin, Prof. Renzo Vanna from the Nanomedicine and Molecular Imaging Lab at Istituti Clinici Scientifici Maugeri Spa – Società Benefit IRCCS, and Prof. Massimo Bongiovanni from University Hospital of Lausanne CHUV. My sincerest thanks to them all.

An additional thank you to Hugh Byrne, the students, and post docs at FOCAS, for making the last four years an enjoyable experience.

Thank you to my friends for their incredible support, and for helping me de-stress when I needed it. I am forever grateful to my Mam, Dad, and family for all the encouragement they have given me over the last four years, without it this would not have been possible.

And finally, I would like to say thank you to my wife Sabrina Moris. After stressful days and late evenings, she always listened to me vent and made me laugh when I got home. I’m incredibly lucky to have had her support.
### Abbreviations list

<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>AC</td>
<td>Adenocarcinoma</td>
</tr>
<tr>
<td>ALK</td>
<td>Anaplastic lymphoma kinase</td>
</tr>
<tr>
<td>ATCC</td>
<td>American Type Culture Collection</td>
</tr>
<tr>
<td>ATR</td>
<td>Attenuated total reflection</td>
</tr>
<tr>
<td>BAL</td>
<td>Bronchioalveolar lavage</td>
</tr>
<tr>
<td>BRF2</td>
<td>B-related factor 2</td>
</tr>
<tr>
<td>BW</td>
<td>Bronchial wash</td>
</tr>
<tr>
<td>CARS</td>
<td>Coherent anti-stokes Raman scattering</td>
</tr>
<tr>
<td>CCD</td>
<td>Charge coupled device</td>
</tr>
<tr>
<td>CT scan</td>
<td>Computerized axial tomography</td>
</tr>
<tr>
<td>CTLA-4</td>
<td>Cytotoxic T-lymphocyte-associated protein 4</td>
</tr>
<tr>
<td>DNA</td>
<td>Deoxyribonucleic acid</td>
</tr>
<tr>
<td>DTT</td>
<td>Dithiothreitol</td>
</tr>
<tr>
<td>EGFR</td>
<td>Epidermal growth factor receptor</td>
</tr>
<tr>
<td>EMT</td>
<td>Epithelial mesenchymal transition</td>
</tr>
<tr>
<td>FBS</td>
<td>Fetal bovine serum</td>
</tr>
<tr>
<td>FDA</td>
<td>Food and drug administration</td>
</tr>
<tr>
<td>FGFR</td>
<td>Fibroblast growth factor receptor</td>
</tr>
<tr>
<td>FISH</td>
<td>Fluorescent in situ hybridisation</td>
</tr>
<tr>
<td>FNA</td>
<td>Fine needle aspiration</td>
</tr>
<tr>
<td>FNAC</td>
<td>Fine needle aspiration cytology</td>
</tr>
<tr>
<td>FTC</td>
<td>Follicular thyroid carcinoma</td>
</tr>
<tr>
<td>FTIR</td>
<td>Fourier transform infrared</td>
</tr>
<tr>
<td>HER2</td>
<td>Human epidermal growth factor receptor 2</td>
</tr>
<tr>
<td>HPV</td>
<td>Human papillomavirus</td>
</tr>
<tr>
<td>IHC</td>
<td>Immunohistochemistry</td>
</tr>
<tr>
<td>Abbreviation</td>
<td>Description</td>
</tr>
<tr>
<td>--------------</td>
<td>-------------</td>
</tr>
<tr>
<td>IMS</td>
<td>Industrial methylated spirits</td>
</tr>
<tr>
<td>LDA</td>
<td>Linear discriminant analysis</td>
</tr>
<tr>
<td>LV</td>
<td>Latent variable</td>
</tr>
<tr>
<td>MET</td>
<td>Mesenchymal Epithelial Transition</td>
</tr>
<tr>
<td>MRI</td>
<td>Magnetic resonance imaging</td>
</tr>
<tr>
<td>MTC</td>
<td>Medullary thyroid carcinoma</td>
</tr>
<tr>
<td>NNLS</td>
<td>Non-negative least squares analysis</td>
</tr>
<tr>
<td>NMCS</td>
<td>No malignant cells seen</td>
</tr>
<tr>
<td>NSCLC</td>
<td>Non-small cell lung cancer</td>
</tr>
<tr>
<td>NSCLC NOS</td>
<td>Non-small cell lung cancer not otherwise specified</td>
</tr>
<tr>
<td>Pap</td>
<td>Papanicolaou</td>
</tr>
<tr>
<td>PBS</td>
<td>Phosphate buffered saline</td>
</tr>
<tr>
<td>PCA</td>
<td>Principal component analysis</td>
</tr>
<tr>
<td>PCR</td>
<td>Polymerase chain reaction</td>
</tr>
<tr>
<td>PDGF</td>
<td>Platelet derived growth factors</td>
</tr>
<tr>
<td>PD-L1</td>
<td>Programmed death-ligand 1</td>
</tr>
<tr>
<td>PLSDA</td>
<td>Partial least squares discriminant analysis</td>
</tr>
<tr>
<td>PTC</td>
<td>Papillary thyroid carcinoma</td>
</tr>
<tr>
<td>RPMI</td>
<td>Roswell Park Memorial Institute medium</td>
</tr>
<tr>
<td>SCC</td>
<td>Squamous cell carcinoma</td>
</tr>
<tr>
<td>SCLC</td>
<td>Small cell lung cancer</td>
</tr>
<tr>
<td>TC</td>
<td>Thyroid cancer</td>
</tr>
<tr>
<td>TKI</td>
<td>Tyrosine kinase inhibitor</td>
</tr>
<tr>
<td>TTF-1</td>
<td>Thyroid transcription factor 1</td>
</tr>
<tr>
<td>US</td>
<td>Ultrasound</td>
</tr>
<tr>
<td>VEGFR-2</td>
<td>Vascular endothelial growth factor receptor 2</td>
</tr>
</tbody>
</table>
Contents

Chapter 1: Cancer cytopathology and Raman spectroscopy ......................................................... 16

1.1 Thesis outline ......................................................................................................................... 16

1.2 Cytopathology ....................................................................................................................... 17

1.2.1 Introduction to cytopathology .......................................................................................... 17

1.2.2 Normal cellular morphology and cancer nomenclature .................................................. 18

1.2.3 The identification of neoplastic changes with cytopathology ......................................... 19

1.2.4 Benefits and limitations of current methods ....................................................................... 21

1.3 Raman spectroscopy ............................................................................................................. 22

1.3.1 Raman spectroscopy for cancer diagnosis ........................................................................ 22

1.3.2 Principle ............................................................................................................................ 23

1.3.3 Instrumentation .................................................................................................................. 26

1.3.4 Raman spectroscopic analysis of cytological samples ...................................................... 28

1.4 Summary ............................................................................................................................... 29

1.4.1 Thesis objectives ............................................................................................................... 29

References .................................................................................................................................. 30

Chapter 2: Recent advances in the vibrational spectroscopic diagnosis of non-small cell lung cancer .......................................................................................................................... 34

Abstract ....................................................................................................................................... 35

2.1 Lung cancer ........................................................................................................................... 36

2.1.1 Introduction to Lung Cancer ............................................................................................ 36

2.1.2 Current Treatment Options ............................................................................................ 38

2.1.3 Current Methods for Differential Diagnosis of NSCLC Subtypes ................................. 40

2.1.4 Bronchoscopy .................................................................................................................. 40

2.1.5 NSCLC Tumour Classification Utilising Cytology Specimens .......................................... 41

2.1.6 Immunohistochemistry .................................................................................................... 41

2.1.7 Fluorescent in situ Hybridisation ...................................................................................... 43

2.1.8 Current Research on Genomic Profiling ......................................................................... 43

2.1.9 Limitations of Current Diagnostic Methods .................................................................... 44

2.2 Vibrational Spectroscopy ..................................................................................................... 44

2.2.1 FTIR Spectroscopy for Lung Cancer Diagnosis ............................................................. 45

2.2.2 Raman Spectroscopy for Lung Cancer Diagnosis .......................................................... 48

2.2.3 Summary and Future Perspectives .................................................................................... 55

2.3 Conclusion ............................................................................................................................ 56
Chapter 3: Methodologies .......................................................... 66

3.1 Study subjects, sample collection and preparation ............................................. 66
   3.1.1 Sample processing for BAL and BW samples ........................................ 66
   3.1.2 Slide preparation .................................................................................. 66

3.2 Raman spectral acquisition and data pre-processing and analysis ......................... 67
   3.2.1 Raman Measurement ........................................................................... 67
   3.2.2 Recording position of targeted cells unstained slide ................................ 67
   3.2.3 Smoothing ............................................................................................ 68
   3.2.4 Baseline removal .................................................................................. 68
   3.2.5 Normalisation ....................................................................................... 68
   3.2.6 Glass removal by non-negative least squares ......................................... 68
   3.2.7 Outlier removal .................................................................................... 69
   3.2.8 Principal component analysis .................................................................. 69
   3.2.9 Linear discriminant analysis .................................................................... 70
   3.2.10 Partial least squares discriminant analysis ........................................... 70
   3.2.11 Cross validation .................................................................................. 71
   3.2.12 Calculation of classification models performance scores ....................... 71

3.3 Pap staining ................................................................................................. 72

Chapter 4: Raman spectroscopic analysis of NSCLC using bronchoscopy cytology samples .... 75

4.1 Introduction ................................................................................................. 75

4.2 Methods ........................................................................................................ 77
   4.2.1 Study subjects and sample collection ..................................................... 77
   4.2.2 Sample processing and slide preparation ............................................... 77
   4.2.3 Raman Measurement ............................................................................ 77
   4.2.4 Data pre-processing and analysis ............................................................ 78
   4.2.5 Pap staining .......................................................................................... 78

4.3 Results .......................................................................................................... 79
   4.3.1 Mean spectral analysis .......................................................................... 79
   4.3.2 Spectral comparison of the lymphocyte, bronchial, and squamous nuclei from negative cases ........................................................................ 84
   4.3.3 Mean spectral analysis of cell nuclei from negative cases, SCC cases, and AC cases 86
   4.3.4 Cytoplasmic spectra analysis .................................................................. 96
6.3.2 Comparison of the mean Raman spectra from bronchial tissues................................. 210
6.3.3 Principal component analysis ......................................................................................... 214
6.3.4 PLSDA classification of lung cancer subtypes .............................................................. 220
6.4 Discussion ....................................................................................................................... 223
References .......................................................................................................................... 232
Chapter 7: Raman spectroscopy for the preoperative diagnosis of thyroid cancer and its
subtypes: an in vitro proof-of-concept study ................................................................. 235
7.1 Introduction ..................................................................................................................... 237
7.2 Materials and Methods ................................................................................................. 239
  7.2.1 Cell cultures ................................................................................................................ 239
  7.2.2 Raman Measurement ................................................................................................. 240
  7.2.3 Data pre-processing and analysis .............................................................................. 241
7.3 Results ........................................................................................................................... 241
  7.3.1 Mean spectral analysis ............................................................................................... 241
  7.3.2 Linear Discriminant Analysis .................................................................................... 246
7.4 Discussion ....................................................................................................................... 253
7.5 Conclusion ..................................................................................................................... 257
References .......................................................................................................................... 257
Chapter 8: Development of a blood removal protocol suitable for thyroid FNAC samples .... 261
8.1 Introduction ..................................................................................................................... 261
8.2 Section A .......................................................................................................................... 263
  8.2.1 Materials and Methods .............................................................................................. 263
  8.2.2 Results ....................................................................................................................... 267
  8.2.3 Discussion .................................................................................................................. 279
8.3 Section B - Use of Raman spectroscopy for the discrimination between normal and
malignant FNAC samples prepared from resected thyroid specimens ............................ 282
  8.3.1 Methods ..................................................................................................................... 283
  8.3.2 Results and discussion ............................................................................................. 285
8.4 Section C – Optimising the blood removal method for thyroid FNAC samples ............. 293
  8.4.1 Methods ..................................................................................................................... 293
  8.4.2 Results and discussion ............................................................................................. 295
8.5 Summary and Conclusion ............................................................................................. 304
References .......................................................................................................................... 305
Chapter 9: Conclusion ......................................................................................................... 307
9.1 Summary of thesis components ................................................................. 307
9.2 Clinical Relevance................................................................................... 309
9.3 Future Work............................................................................................ 311
List of Publications ..................................................................................... 313
Modules completed: .................................................................................... 314
  • Discipline specific.................................................................................. 314
  • Employability skills................................................................................. 314
Conference presentations: ........................................................................ 314
### Table of figures.

<table>
<thead>
<tr>
<th>Figure</th>
<th>Page</th>
<th>Caption</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.1</td>
<td>18</td>
<td>ThinPrep vial for liquid based cytology collection.</td>
</tr>
<tr>
<td>1.2</td>
<td>20</td>
<td><strong>A-D</strong> are stained with the Pap stain to provide contrast between cell types and cell components. Slides were prepared from BALs using a ThinPrep 2000. <strong>A</strong> The flat, polygonal cells marked by red arrows are indicative of morphologically normal squamous cells. <strong>B</strong> The blue arrow indicates morphologically normal bronchial cells. The ciliated epithelium is columnar with cilia bound on the flat surface by a terminal plate. <strong>C</strong> Depiction of malignant squamous cells with abnormal nuclei and increased keratin in the cytoplasm, evident by the uptake of the orange OG6 stain. <strong>D</strong> Malignant bronchial cells forming an acinar structure with pale chromatin and rounded edges.</td>
</tr>
<tr>
<td>1.3</td>
<td>24</td>
<td>Energy level diagram for Rayleigh, Stokes and anti-stokes scattering.</td>
</tr>
<tr>
<td>1.4</td>
<td>28</td>
<td>Schematic of the instrumentation of the HORIBA-Jobin-Yvon XploRA Raman spectrometer.</td>
</tr>
<tr>
<td>2.1</td>
<td>45</td>
<td>Flow diagram of the process involved in applying vibrational spectroscopy to bronchoscopy obtained specimens for cancer detection. After obtaining the cytology/biopsy sample the biological material is stored in a preservative. The tissue section or cellular material is then fixed to a substrate such as a glass slide and left unstained for spectroscopic analysis. Using a laser or an infrared source, spectra are obtained by collecting the Raman scattered light or transmitted light. The spectral data are then pre-processed to minimise noise and spectral contaminations. Diagnostic features are then extracted from the data with multivariate analytical techniques.</td>
</tr>
<tr>
<td>2.2</td>
<td>49</td>
<td>Depiction of a Raman spectrum showing the different spectral regions and the corresponding biomolecules associated with each region. The spectrum which represents a bronchial epithelial cell nucleus was baseline corrected and the non-negatively constrained least squares method was used to remove spectral contaminants from the substrate.</td>
</tr>
<tr>
<td>4.1</td>
<td>80</td>
<td>Depiction of stained and unstained bronchial cells (a), squamous cells (b), and lymphocytes (c). The Pap stain was applied to provide contrast between cell types and components. The corresponding mean Raman spectra (± standard deviation) taken from the nuclei of each cell type are shown. (d) The mean Raman spectrum of bronchial epithelial cell nuclei taken from negative cases (blue, n=70) (n=the number of spectra used), (e) the mean Raman spectrum of squamous cell nuclei taken from negative cases (black, n=94), and (f) the mean Raman spectrum of lymphocyte nuclei taken from negative cases (red, n=80).</td>
</tr>
<tr>
<td>4.2</td>
<td>81</td>
<td>Keratinizing tumour cells with cytoplasmic projections taken from patients with squamous cell carcinoma. The cells are stained with the Pap stain which was applied subsequent to Raman analysis. A-C are images taken at x400 magnification, D is taken at x200 magnification.</td>
</tr>
<tr>
<td>4.3</td>
<td>82</td>
<td>Tumourous glandular cells taken from patients with adenocarcinoma carcinoma. The cells are stained with the Pap stain which was applied subsequent to Raman analysis. A is taken at x200 magnification, B-D are taken at x400 magnification.</td>
</tr>
<tr>
<td>4.4</td>
<td>86</td>
<td>The difference spectrum between the mean Raman spectrum of squamous cell nuclei taken from negative cases and the mean Raman spectrum of bronchial epithelial cell nuclei taken from negative cases is depicted in blue. The difference spectrum between the mean Raman spectrum of squamous nuclei taken from negative cases and the mean Raman spectrum of lymphocyte nuclei taken from negative cases is</td>
</tr>
</tbody>
</table>
shown in red. The difference spectrum between the mean Raman spectrum of lymphocyte nuclei taken from negative cases and the mean Raman spectrum of bronchial nuclei taken from negative cases is depicted in green. Shading indicates regions of the spectrum that were significantly different ($P < .01$).

### 4.5 88
Mean Raman spectrum of negative bronchial nuclei ($n=70$). Mean Raman spectrum of a bronchial cell nucleus from SCC cases ($n=150$). Mean Raman spectrum of a bronchial nucleus from AC cases ($n=63$). Shading denotes the standard deviation.

### 4.6 89
Difference spectrum between the mean Raman spectrum of bronchial epithelial cell nuclei taken from negative cases and the mean Raman spectrum of bronchial epithelial cell nuclei taken from SCC cases (blue, $n=150$), and the mean Raman spectrum of bronchial epithelial cell nuclei taken from AC cases (red, $n=63$) the difference spectrum between the bronchial epithelial cell nuclei taken from SCC and AC cases (green). Shading indicates regions of the spectrum that were significantly different ($P < .01$).

### 4.7 92
Mean Raman spectrum of a negative squamous cell nucleus ($n=94$). Mean Raman spectrum of a squamous nucleus from SCC cases ($n=189$). Mean Raman spectrum of a squamous nucleus from AC cases ($n=67$). Shading denotes the standard deviation.

### 4.8 93
Difference spectrum between the mean Raman spectrum of squamous cell nuclei taken from negative cases and the mean Raman spectrum of squamous cell nuclei taken from SCC cases (blue, $n=189$). Difference spectrum between the mean Raman spectrum of squamous cell nuclei taken from negative cases and the mean Raman spectrum of squamous cell nuclei taken from AC cases (red, $n=67$). Difference spectrum between the mean Raman spectrum of squamous cell nuclei taken from SCC cases and the mean Raman spectrum of squamous cell nuclei taken from AC cases (green). Shading indicates regions of the spectrum that were significantly different ($P < .01$).

### 4.9 95
Mean Raman spectrum of a lymphocyte nucleus form negative cases ($n=80$). Mean Raman spectrum of a lymphocyte nucleus from SCC cases ($n=114$). Mean Raman spectrum of a lymphocyte nucleus from AC cases ($n=58$). Shading denotes the standard deviation.

### 4.10 96
Difference spectrum between the negative lymphocyte nuclear spectrum and the SCC lymphocyte nuclear spectrum (blue, $n=114$), (b) the AC lymphocyte nuclear spectrum (red, $n=58$). Difference spectrum between the mean Raman spectrum of lymphocyte nuclei taken from SCC cases and the mean Raman spectrum of squamous cell nuclei taken from AC cases (green). Shading indicates regions of the spectrum that were significantly different ($P < .01$).

### 4.11 97
Mean Raman spectra of cytoplasm from negative bronchial cells (black, $n=33$), and negative squamous cells (blue, $n=98$). Shading denotes the standard deviation.

### 4.12 98
The difference spectrum between the mean Raman spectrum from the cytoplasm of squamous cells from negative cases and the mean Raman spectrum from the cytoplasm of bronchial cells from negative cases is depicted in blue. Shading indicates regions of the spectrum that were significantly different ($P < .01$).
4.13  99-100  Mean Raman spectrum of a negative bronchial cell cytoplasm (n= 33). Mean Raman spectrum of a bronchial cell cytoplasm from SCC cases (n=68). Mean Raman spectrum of a bronchial cell cytoplasm from AC cases (n=46). Shading denotes the standard deviation.

4.14  100  Difference spectrum between the mean Raman spectrum of negative bronchial epithelial cell cytoplasm (n=33) and (a) and the mean Raman spectrum of SCC bronchial epithelial cell cytoplasm (blue, n=68), (b) the mean Raman spectrum of AC bronchial epithelial cell cytoplasm (red, n=46). Difference spectrum between the mean Raman spectrum of bronchial cell cytoplasm taken from SCC cases and the mean Raman spectrum of bronchial cell cytoplasm taken from AC cases (green). Shading indicates regions of the spectrum that were significantly different (P < .01).

4.15  102  Mean Raman spectrum of a negative squamous cell cytoplasm (n= 98). Mean Raman spectrum of a squamous cytoplasm from SCC cases (n=91). Mean Raman spectrum of a squamous cytoplasm from AC cases (n=43). Shading denotes the standard deviation.

4.16  103  Difference spectrum between the mean Raman spectrum of negative squamous cell cytoplasm (n=98) and (a) and the mean Raman spectrum of SCC squamous cell cytoplasm (blue, n=91), (b) the mean Raman spectrum of AC squamous cell cytoplasm (red, n=43). Difference spectrum between the mean Raman spectrum of squamous cell cytoplasm taken from SCC cases and the mean Raman spectrum of squamous cell cytoplasm taken from AC cases (green). Shading indicates regions of the spectrum that were significantly different (P < .01).

4.17  105  A PCA scatter plot of bronchial, squamous and lymphocyte nuclei from negative cases shows that the bronchial and squamous spectra form two distinct clusters along the first principal component. The lymphocyte spectra are dispersed evenly along PC1 and PC2 and cannot be discriminated from the bronchial or squamous cells. B The loadings of PC1 show peaks assigned to glycogen/cholesterol/protein/phospholipids (420-690 cm\(^{-1}\)), nucleic acids (762-788 cm\(^{-1}\)), glycogen/cholesterol/saccharides/protein/nucleic acids (805-1011 cm\(^{-1}\)), glycogen/peaks related to collagen (1024-32 cm\(^{-1}\)), amide III/nucleic acid (1245-60 cm\(^{-1}\)), lipids/protein/amide I (1436-45, 1620-60 cm\(^{-1}\)) on the positive side, and peaks assigned to lipid/protein/nucleic acids/phosphates (1129-1200, 1351 cm\(^{-1}\)), lipid/protein/COO/nucleic acids/cytosine (1450-1546, 1564-82, 1611 cm\(^{-1}\)), and amide I disordered structure/β sheets/lipids (1680-1719 cm\(^{-1}\)) on the negative side.

4.18  106  Three-dimensional scatter plot of the first three principal components showing no discrimination of spectra taken from the bronchial nuclei of negative, SCC and AC cases.
| 4.19 | 107-108 | A Scatter plot showing spectra taken from the bronchial nuclei of SCC and AC cases. The AC spectra form a tight cluster in the positive side of PC4, whereas the SCC spectra are widely distributed along PC4. The loadings of PC4 indicate variation in methionine (715 cm\(^{-1}\)), nucleic acids (779 cm\(^{-1}\)), saccharides (840, 860 cm\(^{-1}\)), phenylalanine (1004 cm\(^{-1}\)), amide III/nucleic acids (1220 cm\(^{-1}\)), and amide I (1660 cm\(^{-1}\)) associated with the positive side, and peaks assigned to protein (1100 cm\(^{-1}\)), glycogen (1145 cm\(^{-1}\)), nucleic acids/protein (1340 cm\(^{-1}\)), lipid (1379 cm\(^{-1}\)), aspartic/glutamic acid (1400 cm\(^{-1}\)), protein/collagen (~930, 1445 cm\(^{-1}\)), and COO\(^{-}\) (1585 cm\(^{-1}\)) associated with the negative side. |
| 4.20 | 108 | PCA scatter plot of the first two principal components showing no discrimination of spectra taken from the squamous nuclei of negative, SCC and AC cases. PCA did not discriminate between the negative, SCC and AC cases using a combination of PC that explained the first 95% of the variance in the dataset. |
| 4.21 | 109 | PCA scatter plot of the first two principal components showing no discrimination of spectra taken from the squamous nuclei of SCC and AC cases. PCA did not discriminate between the negative, SCC and AC cases using a combination of PCs that explained the first 95% of the variance in the dataset. |
| 4.22 | 110 | A Scatter plot showing the distribution of spectra taken from lymphocyte nuclei of negative cases, SCC, and AC cases, along the first and second PCs. PCA did not discriminate between the negative, SCC and AC cases using a combination of PCs that explained the first 95% of the variance in the dataset. |
| 4.23 | 111 | PCA scatter plot showing spectra from the nuclei of lymphocytes, bronchial, and squamous cells of negative cases, SCC cases and AC cases, grouped according to a negative or malignant diagnosis. The negative cases (blue circles) cannot be discriminated from the malignant cases (blue circles) using PCA, which is depicted by the distribution of spectra along the first and second PCs. |
| 4.24 | 112 | PCA scatter plot showing spectra from the nuclei of lymphocytes, bronchial and squamous cells of SCC, and AC cases grouped according to the lung cancer subtype. PCA did not discriminate between the SCC and AC groups using a combination of PCs that explained the first 95% of the variance in the dataset. |
A PCA scatter plot of spectra taken from the cytoplasm of bronchial and squamous cells of negative cases shows that the bronchial and squamous spectra form two distinct clusters along the first and second principal components. B The loadings of PC1 show peaks assigned to cholesterol/lipids/methionine (409-695 cm\(^{-1}\)), protein/proline/hydroxyproline/saccharides/glycogen (853-57, 896-970 cm\(^{-1}\)), saccharides/lipids/proteins (1106-1125 cm\(^{-1}\)), lipids/proteins/nucleic acids (1296-1333 cm\(^{-1}\)), and lipids (1699-1751 cm\(^{-1}\)) on the positive side and the pyrimidine ring breathing mode at 766 cm\(^{-1}\), phenylalanine (1006 cm\(^{-1}\)), tyrosine/nucleic acids (1159-91 cm\(^{-1}\)), amide III/nucleic acids (1217-1268 cm\(^{-1}\)), the protein/lipid/saccharide peaks at 1356-1415 cm\(^{-1}\), the lipid/protein peak at 1445 cm\(^{-1}\), disaccharides/amide II/nucleic acids/COO\(^-\) (1460-1595 cm\(^{-1}\)), and amide I of disordered structure/β sheets (1674-1690 cm\(^{-1}\)) on the negative side. The loadings of PC2 show peaks assigned to proline/valine/saccharides (~930 cm\(^{-1}\)) and amide I (1656 cm\(^{-1}\)) on the negative side.

A PCA scatter plot showing the discriminatory performance of a PCA model developed using spectra taken from the bronchial cytoplasm of negative, SCC and AC cases. A three dimensional scatter plot demonstrates the discrimination of the three clusters of spectra. PC1 differentiates the SCC cases and the AC cases. A combination of PC2 and PC3 separates the negative cases from the SCC and AC cases. PC2 is mainly responsible for discriminating negative and AC spectra. PC3 discriminates the negative and SCC clusters. B The loadings of PC1 show positive peaks associated with amide III/nucleic acids (1250 cm\(^{-1}\)), proteins (~1330 cm\(^{-1}\)), lipids (1445 cm\(^{-1}\)), and negative peaks associated with amide I and II (~1500-1650 cm\(^{-1}\)). The loadings of PC2 show peaks assigned to proteins/carbohydrates/nucleic acids (735-890 cm\(^{-1}\)), phenylalanine (1000 cm\(^{-1}\)), amide III/nucleic acids (1200-1280 cm\(^{-1}\)), and amide I of disordered structure/β sheets (1690 cm\(^{-1}\)) on the positive side and nucleotides/phosphatidylinositol/glycerol (570-620 cm\(^{-1}\)), carbohydrates/proteins/lipids (1100-1145 cm\(^{-1}\)), collagen (~1330 cm\(^{-1}\)), protein (~1410 cm\(^{-1}\)), and amide I (1650 cm\(^{-1}\)) on the negative side. The loadings of PC3 indicate positive peaks associated with increased saccharides (850 cm\(^{-1}\)), proline/valine (930 cm\(^{-1}\)), and protein (1410 cm\(^{-1}\)), and decreased phosphatidylinositol (~570-80 cm\(^{-1}\)), and amide I (1650 cm\(^{-1}\)).
A PCA scatter plot of spectra from the cytoplasm of squamous cells showing variance between the negative cluster and the SCC and AC clusters along PC1, with some overlap. The majority of spectra from negative cases are distributed in the negative side of PC1, whereas the majority of spectra SCC and AC cases are distributed along the positive side of PC1. The SCC and AC spectra form overlapping clusters and cannot be discriminated with PCA. B The loadings of PC1 show peaks assigned to protein/collagen (1169, 1223, 1240-47, 1265-69, 1278, 1280, 1283, 1336, 1344, 1488 cm\(^{-1}\)), phenylalanine (995-1004 cm\(^{-1}\)), tyrosine (1165-1171 cm\(^{-1}\)), amide III/ lipids (1219-92, ~1337 cm\(^{-1}\)), amide II (1472-89 cm\(^{-1}\)), and amide I of disordered structure/\(\beta\) sheets (1674-1688 cm\(^{-1}\)) on the positive side, and cholesterol (418, 424-432, 608, 614, 702 cm\(^{-1}\)), saccharides (451-502 cm\(^{-1}\)), phosphatidylinositol/proteins (508-527 cm\(^{-1}\)), proteins/nucleic acids/carbohydrates/lipids (568-726 cm\(^{-1}\)), proline/valine (935, ~950 cm\(^{-1}\)), amide I (1594-1657 cm\(^{-1}\)), and lipid (~1750 cm\(^{-1}\)) on the negative side.

A PCA scatter plot showing spectra from the cytoplasm of bronchial and squamous cells of negative cases, SCC cases and AC cases, grouped according to a negative or malignant diagnosis. The negative and malignant cases cannot be discriminated. The malignant cases (red circles) form a cluster along PC1, whereas the negative cases (blue circles) are evenly distributed along PC1. B The loadings of PC1 revealed a greater variation in phosphatidylinositol (575 cm\(^{-1}\)), glucose (908 cm\(^{-1}\)), proline/valine/saccharides (~935 cm\(^{-1}\)), saccharide (1110 cm\(^{-1}\)) and amide I (1650 cm\(^{-1}\)) in the negative cases. The loadings of PC1 are similar to the loadings discriminating spectra from the cytoplasm of bronchial and squamous cells from negative cases, indicating that the variation explained by PC1 is likely the difference between the bronchial and squamous cells.

A PCA scatter plot showing spectra from the cytoplasm of bronchial and squamous cells of SCC, and AC cases grouped according to the lung cancer subtype. The SCC and AC cases cannot be discriminated with PCA, which is depicted with a plot of the first two PCs. The AC cases (red circles) form a two distinct clusters along PC1. B The loadings of PC1 are similar to the loadings discriminating spectra from the cytoplasm of bronchial and squamous cells from negative cases, indicating that the two distinct clusters are likely a separation of the bronchial and squamous cells from AC cases.

LV score scatter plot of Raman spectra from bronchial nuclei of negative cases (blue circles), squamous nuclei of negative cases (green squares), and lymphocyte nuclei of negative cases (solid red circles). The three-way classification model used the first 3 latent variables and yielded sensitivities and specificities of 69-94% for discrimination based on cell type. B and C Latent variables loadings of the developed PLS-DA model for the dataset obtained from the bronchial, squamous, and lymphocyte nuclei of negative cases.

A Scatter plot of the latent variables (LVs) obtained from the Raman spectral dataset of bronchial nuclei from negative cases (blue circles), SCC cases (green squares), and AC cases (solid red circles). B and C
Latent variables loadings of the developed PLS-DA model for the dataset obtained from bronchial nuclei.

<table>
<thead>
<tr>
<th>Page</th>
<th>Line</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>4.32</td>
<td>128</td>
<td><strong>A</strong> Scatter plot of the first two LVs obtained from the Raman spectral dataset of squamous nuclei from negative cases (blue circles), SCC cases (green squares), and AC cases (solid red circles). <strong>B</strong> Scatter plot of the first and third LVs obtained from the Raman spectral dataset of squamous nuclei from negative cases (blue circles), SCC cases (green squares), and AC cases (solid red circles).</td>
</tr>
<tr>
<td>4.33</td>
<td>129</td>
<td>Latent variables from PLSDA of Raman spectra from squamous nuclei of negative cases, SCC cases, and AC cases. <strong>A-C</strong> The Latent variables loadings of the developed PLS-DA model for the dataset obtained from squamous nuclei. <strong>A</strong> depicts LV1, <strong>B</strong> shows the loadings of LV2, and <strong>C</strong> illustrates the loadings of LV3.</td>
</tr>
<tr>
<td>4.34</td>
<td>131</td>
<td><strong>A</strong> Scatter plot of the LVs obtained from the Raman spectral dataset of lymphocyte nuclei from negative cases (blue circles), SCC cases (green squares), and AC cases (solid red circles). <strong>B</strong> and <strong>C</strong> Latent variables loadings of the developed PLS-DA model for the dataset obtained from lymphocyte nuclei.</td>
</tr>
<tr>
<td>4.35</td>
<td>134-135</td>
<td><strong>A</strong> Scatter plot of the first and third LVs obtained from the Raman spectral dataset of combined bronchial, squamous, and lymphocyte nuclei. Representative spectra from each cell type from the negative, SCC, and AC cases were grouped according to a negative or malignant diagnosis. The negative cases (red circles) are discriminated from the malignant cases (blue circles) along the first and third LVs. <strong>B</strong> and <strong>C</strong> Latent variables loadings of the developed PLS-DA model for the dataset obtained from bronchial, squamous, and lymphocyte nuclei.</td>
</tr>
<tr>
<td>4.36</td>
<td>137</td>
<td><strong>A</strong> Scatter plot showing spectra from the bronchial, squamous, and lymphocyte nuclei of SCC and AC cases grouped according to the lung cancer subtype. The AC cases (red circles) are discriminated from the SCC cases (blue circles) along the first and second LVs. <strong>B</strong> and <strong>C</strong> Latent variables loadings of the developed PLS-DA model for the dataset obtained from bronchial, squamous, and lymphocyte nuclei.</td>
</tr>
<tr>
<td>4.37</td>
<td>139</td>
<td><strong>A</strong> Scatter plot of the LVs obtained from the Raman spectra taken from the cytoplasm of bronchial cells from negative cases (blue circles), SCC cases (green squares), and AC cases (solid red circles). <strong>B</strong> and <strong>C</strong> Latent variables loadings of the developed PLS-DA model for the dataset obtained from the cytoplasm of bronchial cells.</td>
</tr>
<tr>
<td>4.38</td>
<td>141</td>
<td><strong>A</strong> Scatter plot of the LVs obtained from the Raman spectra taken from the cytoplasm of squamous cells from negative cases (blue circles), SCC cases (green squares), and AC cases (solid red circles). <strong>B</strong> and <strong>C</strong> Latent variables loadings of the developed PLS-DA model for the dataset obtained from the cytoplasm of squamous cells.</td>
</tr>
<tr>
<td>4.39</td>
<td>144</td>
<td><strong>A</strong> Scatter plot showing spectra from the cytoplasm of bronchial and squamous cells of negative cases, SCC cases and AC cases, grouped according to a negative or malignant diagnosis. The negative cases (green circles) are discriminated from the malignant cases (blue circles) along the first and second LVs. <strong>B</strong> and <strong>C</strong> Latent variables loadings of the developed PLS-DA model for the dataset obtained from the cytoplasm of bronchial and squamous cells.</td>
</tr>
<tr>
<td>4.40</td>
<td>146</td>
<td><strong>A</strong> Scatter plot showing spectra from the cytoplasm of bronchial and squamous cells of SCC, and AC cases grouped according to the lung cancer subtype. The AC cases (red circles) are discriminated from the SCC cases (blue circles) along the first LV. <strong>B</strong> The Latent variable</td>
</tr>
</tbody>
</table>
loadings of the developed PLS-DA model for the dataset obtained from the cytoplasm of bronchial and squamous cells.

5.1 174 The mean Raman spectra (± standard deviation) taken from the nuclei of bronchial epithelial cells from negative cases (black, n=70), TP-SCC cases (blue, n=93), and FN-SCC cases (red, n=48).

5.2 175 Difference spectrum between the mean Raman spectrum of bronchial epithelial cell nuclei taken from negative cases and (a) the mean Raman spectrum of bronchial epithelial cell nuclei from TP-SCC cases (blue, n=93), (b) the mean Raman spectrum of bronchial epithelial cell nuclei from FN-SCC cases (red, n=48). (c) Difference spectrum between the mean Raman spectrum of bronchial epithelial cell nuclei from TP-SCC cases and the mean Raman spectrum of bronchial epithelial cell nuclei from FN-SCC cases (green). Shading indicates regions of the spectrum that were significantly different (P < 0.01).

5.3 176 The mean Raman spectra (± standard deviation) taken from the nuclei of squamous cells from negative cases (black, n=94), TP-SCC cases (blue, n=137), and FN-SCC cases (red, n=35).

5.4 177 Difference spectrum between the mean Raman spectrum of squamous cell nuclei taken from negative cases and (a) the mean Raman spectrum of squamous cell nuclei from TP-SCC cases (blue, n=137), (b) the mean Raman spectrum of squamous cell nuclei from FN-SCC cases (red, n=35). (c) Difference spectrum between the mean Raman spectrum of squamous cell nuclei from TP-SCC cases and the mean Raman spectrum of squamous cell nuclei from FN-SCC cases (green). Shading indicates regions of the spectrum that were significantly different (P < 0.01).

5.5 181 The mean Raman spectra (± standard deviation) taken from the nuclei of bronchial epithelial cells from negative cases (black, n=70), TP-AC cases (blue, n=25), and FN-AC cases (red, n=21).

5.6 182 Difference spectrum between the mean Raman spectrum of bronchial epithelial cell nuclei taken from negative cases and (a) the mean Raman spectrum of bronchial epithelial cell nuclei from TP-AC cases (blue, n=25), (b) the mean Raman spectrum of bronchial epithelial cell nuclei from FN-AC cases (red, n=21). (c) Difference spectrum between the mean Raman spectrum of bronchial epithelial cell nuclei from TP-AC cases and the mean Raman spectrum of bronchial epithelial cell nuclei from FN-AC cases (green). Shading indicates regions of the spectrum that were significantly different (P < 0.01).

5.7 183 The mean Raman spectra (± standard deviation) taken from the nuclei of squamous cells from negative cases (black, n=94), TP-AC cases (blue, n=25), and FN-AC cases (red, n=20).

5.8 184 Difference spectrum between the mean Raman spectrum of squamous cell nuclei taken from negative cases and (a) the mean Raman spectrum of squamous cell nuclei from TP-AC cases (blue, n=25), (b) the mean Raman spectrum of squamous cell nuclei from FN-AC cases (red, n=20). (c) Difference spectrum between the mean Raman spectrum of squamous cell nuclei from TP-AC cases and the mean Raman spectrum of squamous cell nuclei from FN-AC cases (green). Shading indicates regions of the spectrum that were significantly different (P < 0.01).

5.9 186 A PCA scatter plot of bronchial cells from from negative cases, TP-SCC cases, and FN-SCC cases. The spectra from TN cases and TP-SCC cases form two adjacent, overlapping clusters along the first PC. The spectra taken from bronchial cells of FN-SCC cases were distributed evenly.
along PC1 and PC2 and could not be discriminated using PCA. B On the positive side of the first principal component peaks at 650, 850, 925, 1004, 1210-1280 cm\(^{-1}\), ~1350, 1450, and 1680 cm\(^{-1}\) can be distinguished. While peaks at and 780 cm\(^{-1}\) and 1500-1645 cm\(^{-1}\) can be distinguished on the negative side.

5.10 187  PCA scatter plot of spectra taken from squamous cells from TN cases, TP-SCC cases, and FN-SCC cases shows no separation with even distribution along the first two principal components.

5.11 188  A PCA scatter plot of bronchial cells from TN cases, TP-AC cases, and FN-AC cases. The spectra from TP-AC cases and FN-AC cases form two distinct clusters along the first principal component. The spectra from negative cases are dispersed along the first two PCs and cannot be discriminated. B. On the positive side of the first principal component, peaks at 650, 850, 1210-1280, 1350, 1420-50, 1680 cm\(^{-1}\) can be distinguished. While peaks at 780, 1105, 1500-1650 cm\(^{-1}\) can be distinguished on the negative side.

5.12 189  A PCA scatter plot of squamous cells from TN cases, TP-AC cases, and FN-AC cases. The spectra from TP-AC cases and FN-AC cases form two distinct clusters along the first principal component. The spectra from TN cases are dispersed along the first two PCs and cannot be discriminated. B. On the positive side of the first principal component, peaks at 650, 850, 1004 cm\(^{-1}\), 1210-1270, ~1360, 1420-50, and ~1680 cm\(^{-1}\) can be distinguished. While peaks at 780 cm\(^{-1}\) and 1480-1655 cm\(^{-1}\) can be distinguished on the negative side.

6.1 203  Cases of SCC stained with Haematoxylin and Eosin. A-F show nests of malignant squamous epithelial cells with large eosinophilic cytoplasm and intercellular bridges. A-D are at x200 magnification, E and F are at x400 magnification.

6.2 204  Cases of AC stained with Haematoxylin and Eosin. A-F show adenocarcinomas with acinar formation, mucinous vacuoles and signet ring cells. A-D are at x200 magnification, E and F are at x400 magnification.

6.3 205  Normal tissue stained with Haematoxylin and Eosin. A shows normal bronchial epithelium with mucinous glands at x200 magnification. B shows normal mucinous glands at x400 magnification. C shows normal bronchial epithelium at x400 magnification. D shows normal bronchial epithelium at x200 magnification. E and F show normal bronchial epithelium at x400 magnification.

6.4 206  Case of NSCLC-NOS stained with Haematoxylin and Eosin. A and B show abnormal cellular and nuclear morphology however there are no defining characteristics for AC or SCC. A is at x200 magnification, B is at x400 magnification.

6.5 206  Case of Atypical carcinoid stained with Haematoxylin and Eosin. A and B Sheets of tumour cells that have uniform round nuclei. The tumour cells form nests which are sharply distinct from the stroma. A is at x200 magnification, B is at x400 magnification.

6.6 208  Mean Raman spectrum of the grouped negative tissue (black, n= 112). Mean Raman spectrum of the grouped malignant tissue (red, n=119). Shading denotes the standard deviation. The spectral data has been subject to NNLS for removal of wax peaks, however strong wax peaks remain in the spectra at 1060 and 1294 cm\(^{-1}\).

6.7 209  A PCA scatter plot of processed negative and malignant tissue shows a separation on the fifth principal component with even distribution along the first principal component. B Wax spectral peaks influence
the fifth principal component, along which the spectra from negative and malignant tissue are separated. Shading denotes the wax related peaks influencing the fifth principal component.

6.8 212 Mean spectrum taken from negative tissue (n= 112 spectra). Wax related peaks in the ranges of 877 to 895, 1050 to 1072, 1115 to 1183, 1280 to 1311, and 1400 to 1500 cm\(^{-1}\) were removed for analysis. Shading denotes the standard deviation.

6.9 213 Difference spectrum between the mean Raman spectrum of negative tissue and the mean Raman spectrum taken from SCC tissue (blue, n=141), AC tissue (red, n=70), NSCLC-NOS (green, n=25), atypical carcinoid (magenta, n=25), and grouped spectra from malignant tissue (light blue, n=119). Shading indicates regions of the spectrum that were significantly different (P < 0.01). Variables were excluded from spectral analysis in the ranges of 877 to 895, 1050 to 1072, 1115 to 1183, 1280 to 1311, and 1400 to 1500 cm\(^{-1}\), as the spectra indicated wax contamination.

6.10 214 Difference spectrum between the mean Raman spectrum of SCC tissue (n=141) and the mean Raman spectrum taken from AC tissue (n=70). Variables were excluded from spectral analysis in the ranges of 877 to 895, 1050 to 1072, 1115 to 1183, 1280 to 1311, and 1400 to 1500 cm\(^{-1}\), as the spectra indicated wax contamination.

6.11 215 A PCA scatter plot of processed negative and malignant tissue shows a separation on the fourth principal component with comparable distribution along the first principal component. B Fourth PC after wax contribution was removed. On the positive side of the fourth principal component, phenylalanine related peak 570 cm\(^{-1}\), 1004 cm\(^{-1}\) and 1585 cm\(^{-1}\) can be distinguished. While peaks at 715, 850, 925, 1233, and 1645 cm\(^{-1}\) can be distinguished on the negative side. Variables were excluded from PCA analysis in the ranges of 877 to 895, 1050 to 1072, 1115 to 1183, 1280 to 1311, and 1400 to 1500 cm\(^{-1}\), as the spectra indicated wax contamination.

6.12 216 PCA scatter plot of SCC and AC tissue shows no separation with even distribution along the first two principal components. Variables were excluded from PCA analysis in the ranges of 877 to 895, 1050 to 1072, 1115 to 1183, 1280 to 1311, and 1400 to 1500 cm\(^{-1}\), as the spectra indicated wax contamination.

6.13 217-218 A. PCA scatter plot of processed SCC, AC and NSCLC-NOS tissue shows that the SCC and NSCLC-NOS spectra form two distinct clusters along the second principal component. The AC spectra are dispersed along the second PC and cannot be discriminated from SCC or NSCLC. B. On the positive side of the second principal component, peaks at 715, 780, 1185, 1200, ~1360, and ~1575 cm\(^{-1}\) can be distinguished. While some peaks at 560, 1004, 860, 930, 1030, 1075-1114, 1230, and 1650 cm\(^{-1}\) can be distinguished on the negative side. Variables were excluded from PCA analysis in the ranges of 877 to 895, 1050 to 1072, 1115 to 1183, 1280 to 1311, and 1400 to 1500 cm\(^{-1}\), as the spectra indicated wax contamination.

6.14 219 A Scatter plot of the principal component scores of representative spectra from negative, SCC, AC, NSCLC-NOS, and atypical carcinoid tissue. A combination of PC1 and PC3 separates the SCC and AC clusters from the negative, NSCLC-NOS and atypical carcinoid spectra, with some overlap. The negative, NSCLC-NOS, and atypical carcinoid tissue shows similar distribution along the first and third principal components. B Loadings from PC1 and PC3. Variables were
excluded from PCA analysis in the ranges of 877 to 895, 1050 to 1072, 1115 to 1183, 1280 to 1311, and 1400 to 1500 cm\(^{-1}\), as the spectra indicated wax contamination.

6.15 221 LV score scatter plot of Raman spectra from negative tissue (blue circles), and grouped spectra from malignant tissue (red circles). Variables were excluded from PCA analysis in the ranges of 877 to 895, 1050 to 1072, 1115 to 1183, 1280 to 1311, and 1400 to 1500 cm\(^{-1}\), as the spectra indicated wax contamination.

6.16 222 LV score scatter plot of Raman spectra from negative tissue (blue circles), SCC tissue (green circles), and AC tissue (red circles). Variables were excluded from PCA analysis in the ranges of 877 to 895, 1050 to 1072, 1115 to 1183, 1280 to 1311, and 1400 to 1500 cm\(^{-1}\), as the spectra indicated wax contamination.

7.1 242 Mean Raman spectrum of the benign cell line (NThy-ori 3-1) with the standard deviation shown as shading, and the different spectral regions with corresponding biomolecules.

7.2 242-243 Difference spectrum between (a) Nthy-ori 3-1 (benign) and FTC cell line XTC1; (b) Nthy-ori 3-1 and PTC cell line K1; (c) Nthy-ori 3-1 and PTC cell line TPC1. Shading indicates regions of the spectrum that were significantly different (p<0.05).

7.3 243 Difference spectrum between (a) Nthy-ori 3-1 (benign) and 8505C (UTC); and (b) Nthy-ori 3-1 and C643 (UTC). Shading indicates regions of the spectrum that were significantly different (p<0.05).

7.4 244 Difference spectrum between (a) Nthy-ori 3-1 (benign) and CRL1803TT (MTC); (b) Nthy-ori 3-1 and MZCRC1 (MTC). Shading indicates regions of the spectrum that were significantly different (p<0.05).

7.5 252 Scatter plot of the linear discriminant scores of the benign spectra and representative spectra from each TC subtype using PC-LDA.

8.1 264 Blood scale index used to classify patient samples as having no blood contamination (Group 0), low contamination (Group 1), medium contamination (Group 2), and heavy contamination (Group 3).

8.2 267 Group 2 samples which received 5 cytolyt washes and 3 minutes H\(_2\)O\(_2\) on the slide. (a) Prepared using a gynae filter for ThinPrep®, and (b) Prepared using a non-gynae filter. The red arrows highlight areas with an accumulation of blood debris. (a) and (b) are at x400 magnification.

8.3 268 (a) Group 0 cells representing a negative control for the presence of blood (sample 10). (b) Group 1 which received blood removal treatment (sample 2). The comparable presentation of the slides indicates that microscopic traces of blood have been successfully removed from sample 2. (a) and (b) are at x200 magnification.

8.4 269 A positive control for the presence of blood (sample 11). (b) Group 2 which received blood removal treatment (sample 6). The red arrows highlight areas with an accumulation of blood debris. (a) and (b) are at x200 magnification.

8.5 270 (a) A positive control for the presence of blood (sample 11). (b) Group 3 which received blood removal treatment (sample 8). The red arrows highlight areas with an accumulation of blood debris. (a) and (b) are at x200 magnification.

8.6 271 (a) Group 0 cells representing a negative control for the presence of blood (sample 10). (b) Group 1 which received blood removal treatment (sample 3). The comparable presentation of the slides indicates that microscopic traces of blood have been successfully removed from sample 3. (a) and (b) are at x400 magnification.
(a) A positive control for the presence of blood (sample 11). (b) Group 2 which received blood removal treatment (sample 5). The red arrows highlight areas with an accumulation of blood debris. (a) and (b) are at x400 magnification.

(a) A positive control for the presence of blood (sample 11). (b) Group 3 which received blood removal treatment (sample 7). The red arrows highlight areas with an accumulation of blood debris. (a) and (b) are at x400 magnification.

Mean spectrum of a HeLa cell line. The cells were not contaminated with blood and are used as a reference spectrum or negative control.

Spectrum of blood representing the positive control.

Raman spectrum taken from a group I sample indicating burning. Cells were exposed to H$_2$O$_2$ on the ThinPrep slide. There are no distinct biological features in the spectrum.

Difference spectra of the HeLa cells in group 1. (a) corresponds to sample 1 in Table 8.2 (b) the difference spectra for sample 2, (c) the difference spectra for sample 3, this sample was subjected to 4 CytoLyt washes and 3 minutes exposure to hydrogen peroxide.

Difference spectra of the HeLa cells in group 2. (a) corresponds to sample 4 in Table 8.2 (b) the difference spectra for sample 5, (c) the difference spectra for sample 6, this sample was subjected to 5 CytoLyt washes and 5 minutes exposure to hydrogen peroxide.

Difference spectra of the HeLa cells in group 3 (a) corresponds to sample 7 in Table 8.2 (b) the difference spectra for sample 8, (c) the difference spectra for sample 9.

Depiction of how the samples were combined to yield sufficient cellular material for Raman spectroscopic analysis.

(a) Mean spectra of the benign thyroid cells (N1) and the thyroid cancer cells (T1), (b) the difference spectrum between benign thyroid cells and thyroid cancer cells with the statistically significant Raman peaks highlighted in grey. Shading indicates regions of the spectrum that were significantly different (p<0.05).

(a) Mean spectra of the benign thyroid cells N2 and the thyroid cancer cells (T2), (b) the difference spectrum between benign thyroid and thyroid cancer. Shading indicates regions of the spectrum that were significantly different (p<0.05).

PCA scores plot of benign thyroid cells (orange) and tumour cells (blue) attained from FNAC samples. (a) Scatter plot of T1 (blue) and N1 (orange) using PC2 and PC3. (b) Scatter plot of T2 (blue) and N2 (orange) using PC2 and PC5.

Loadings plot of PC2 discriminating the benign and cancer cells in Figure 8.18 (a), Peaks with higher intensity highlight the prominent differences between benign and cancerous thyroid cells.

Loadings plot of PC5 discriminating the benign and cancer cells in Figure 8.18 (b).

Steel slide ThinPrep® made with sample 1 post treatment. No cellular material can be seen.

Depiction of cellular material and debris on the steel slide after the sample received treatment. The cell marked with a red box and shown at a greater magnification was targeted for spectroscopic analysis. Acquired with the laser power set to 50% the raw spectra indicate the presence of biological peaks with a high background signal.
Depiction of cellular material and debris on the steel slide after the sample received treatment. The cell marked with a red box and shown at a greater magnification was targeted for spectroscopic analysis. Acquired with the laser power set to 50% the raw spectra indicate the presence of biological peaks with a high background signal.

Depiction of cellular material and debris on the steel slide after the sample received treatment. The cell marked with a red box and shown at a greater magnification was targeted for spectroscopic analysis. Acquired with the laser power set to 100% the raw spectra indicate a strong biological signal with a high background signal.

Depiction of cellular material and debris on the steel slide after the sample received treatment. The cell marked with a red box and shown at a greater magnification was targeted for spectroscopic analysis. Acquired with the laser power set to 100% the raw spectra indicate a strong biological signal with a high background signal.

Depiction of cellular material and debris on the steel slide after the sample received treatment. The cell marked with a red box and shown at a greater magnification was targeted for spectroscopic analysis. Acquired with the laser power set to 100% the raw spectra indicate a strong biological signal with a high background signal.

The mean spectrum taken from the nuclei of cells in sample 5. The spectrum has been processed with Savitsky-Golay smoothing, rubberband baseline subtraction, and vector normalisation, to maximise the intensity of cellular signals and minimise noise. Strong cellular peaks in the processed spectrum are not swamped by blood peaks.

List of tables

<table>
<thead>
<tr>
<th>Table</th>
<th>Page</th>
<th>Title</th>
</tr>
</thead>
<tbody>
<tr>
<td>2.1</td>
<td>37</td>
<td>Progression stage, estimated 5 year survival rate, and appropriate treatments for each stage of NSCLC</td>
</tr>
<tr>
<td>4.1</td>
<td>83</td>
<td>Tentative peak assignments for the prominent significant Raman bands identified by spectral variation analysis</td>
</tr>
<tr>
<td>4.2</td>
<td>146</td>
<td>Performance of PLSDA in discriminating bronchial cell nuclei of negative cases, SCC cases, and AC cases, in separate two-way classifications</td>
</tr>
<tr>
<td>4.3</td>
<td>146</td>
<td>Performance of PLSDA in discriminating squamous cell nuclei of negative cases, SCC cases, and AC cases, in separate two-way classifications</td>
</tr>
<tr>
<td>4.4</td>
<td>146</td>
<td>Performance of PLSDA in discriminating lymphocyte nuclei of negative cases, SCC cases, and AC cases, in separate two-way classifications</td>
</tr>
<tr>
<td>4.5</td>
<td>146</td>
<td>Performance of PLSDA in discriminating the bronchial cell nuclei of negative cases, SCC cases and AC cases</td>
</tr>
<tr>
<td>4.6</td>
<td>148</td>
<td>Performance of PLSDA in discriminating the squamous cell nuclei of negative cases, SCC cases and AC cases</td>
</tr>
<tr>
<td>Section</td>
<td>Pages</td>
<td>Description</td>
</tr>
<tr>
<td>---------</td>
<td>-------</td>
<td>-------------</td>
</tr>
<tr>
<td>4.7</td>
<td>148</td>
<td>Performance of PLSDA in discriminating the lymphocyte nuclei of negative cases, SCC cases and AC cases</td>
</tr>
<tr>
<td>4.8</td>
<td>148</td>
<td>Performance of PLSDA in discriminating negative cases from malignant cases, using spectra from the nuclei of bronchial cells, squamous cells, and lymphocytes</td>
</tr>
<tr>
<td>4.9</td>
<td>149</td>
<td>Performance of PLSDA in discriminating negative cases from malignant cases, and SCC cases from AC cases using combined spectra from the nuclei of bronchial cells, squamous cells, and lymphocytes</td>
</tr>
<tr>
<td>4.10</td>
<td>149</td>
<td>Performance of PLSDA in discriminating the cytoplasm of bronchial cells from negative cases, SCC cases and AC cases</td>
</tr>
<tr>
<td>4.11</td>
<td>149</td>
<td>Performance of PLSDA in discriminating the cytoplasm of squamous cells from negative cases, SCC cases and AC cases</td>
</tr>
<tr>
<td>4.12</td>
<td>149-150</td>
<td>Performance of PLSDA in discriminating negative cases from malignant cases, and SCC cases from AC cases using spectra from the cytoplasm of bronchial cells and squamous cells</td>
</tr>
<tr>
<td>4.13</td>
<td>150</td>
<td>Performance of PLSDA in discriminating negative cases from malignant cases, and SCC cases from AC cases using combined spectra from the cytoplasm of bronchial and squamous cells</td>
</tr>
<tr>
<td>5.1</td>
<td>169-170</td>
<td>Patient samples with the corresponding cytology and biopsy reports, as characterised by a pathologist</td>
</tr>
<tr>
<td>5.2</td>
<td>173-174</td>
<td>Tentative peak assignments for the prominent significant Raman bands identified by spectral variation analysis</td>
</tr>
<tr>
<td>5.3</td>
<td>181</td>
<td>Performance of PLSDA in discriminating bronchial nuclei of negative cases from bronchial nuclei of SCC cases with SCC on cytology and SCC with NMCS on cytology. The first 3 LVs were used in each case</td>
</tr>
<tr>
<td>5.4</td>
<td>191-192</td>
<td>Performance of PLSDA in discriminating squamous nuclei of negative cases from squamous nuclei of SCC cases with SCC on cytology and SCC with NMCS on cytology. The first 2 LVs were used in each case</td>
</tr>
<tr>
<td>5.5</td>
<td>192</td>
<td>Performance of PLSDA in discriminating bronchial nuclei of negative cases from bronchial nuclei of AC cases with AC on cytology and AC with NMCS on cytology. The first 4 LVs were used in each case</td>
</tr>
<tr>
<td>5.6</td>
<td>192</td>
<td>Performance of PLSDA in discriminating squamous nuclei of negative cases from squamous nuclei of AC cases with AC on cytology and AC with NMCS on cytology. The first 2 LVs were used in each case</td>
</tr>
<tr>
<td>5.7</td>
<td>192</td>
<td>Cytology reports with their corresponding biopsy reports and ranges of sensitivity for malignancy using Raman spectroscopy</td>
</tr>
<tr>
<td>6.1</td>
<td>223</td>
<td>Performance of PLSDA in discriminating Raman spectra from negative and malignant tissue</td>
</tr>
<tr>
<td>6.2</td>
<td>223</td>
<td>Performance of PLSDA in discriminating the Raman spectra from negative, SCC, and AC tissue</td>
</tr>
<tr>
<td>6.3</td>
<td>223</td>
<td>Performance of PLSDA in discriminating Raman spectra taken from negative tissue, NSCLC tissue, and atypical carcinoid tissue</td>
</tr>
<tr>
<td>7.1</td>
<td>244-245</td>
<td>Tentative peak assignments for the significant Raman bands identified by spectral variation analysis</td>
</tr>
<tr>
<td>7.2</td>
<td>246-247</td>
<td>Performance of the PC-LDA classification model in discriminating the benign thyroid follicular cell line Nthy-ori 3-1 from each TC cell line</td>
</tr>
<tr>
<td>7.3</td>
<td>247-248</td>
<td>Performance of the PC-LDA classification model in discriminating the benign thyroid follicular cell line Nthy-ori 3-1 from TC cell lines grouped by TC subtype</td>
</tr>
<tr>
<td>7.4</td>
<td>248</td>
<td>Performance of the PC-LDA classification model in discriminating PTC from UTC cell lines</td>
</tr>
<tr>
<td>7.5</td>
<td>249</td>
<td>Performance of the PC-LDA classification model in discriminating MTC from PTC cell lines</td>
</tr>
<tr>
<td>Section</td>
<td>Pages</td>
<td>Description</td>
</tr>
<tr>
<td>---------</td>
<td>-------</td>
<td>-------------</td>
</tr>
<tr>
<td>7.6</td>
<td>249-250</td>
<td>Performance of PC-LDA classification of FTC lines from PTC, MTC and UTC cell lines</td>
</tr>
<tr>
<td>7.7</td>
<td>250-251</td>
<td>Performance of the PC-LDA classification model for discriminating the MTC from UTC cell lines</td>
</tr>
<tr>
<td>7.8</td>
<td>251-252</td>
<td>Performance of the PC-LDA classification model for discriminating the benign and TC subtypes.</td>
</tr>
<tr>
<td>7.9</td>
<td>253</td>
<td>Performance of the PC-LDA classification model for discriminating the benign cell line from all TC lines.</td>
</tr>
<tr>
<td>8.1</td>
<td>265</td>
<td>Stepwise protocol for blood removal with H2O2 treatment on the slide</td>
</tr>
<tr>
<td>8.2</td>
<td>266</td>
<td>Stepwise protocol for blood removal with H₂O₂ treatment in the tube</td>
</tr>
<tr>
<td>8.3</td>
<td>274</td>
<td>Tentative peak assignments for the significant Raman bands identified by spectral variation analysis</td>
</tr>
<tr>
<td>8.4</td>
<td>275</td>
<td>Peak assignments for the significant Raman bands of the spectrum of blood.</td>
</tr>
<tr>
<td>8.5</td>
<td>288-289</td>
<td>Tentative peak assignments for the significant Raman bands identified by spectral variation analysis</td>
</tr>
<tr>
<td>8.6</td>
<td>292</td>
<td>The sensitivity, specificity and classification accuracy of a PC-LDA model for discriminating the benign spectra from the thyroid cancer cells.</td>
</tr>
<tr>
<td>8.7</td>
<td>293-294</td>
<td>Overview of the number of washes and method of H₂O₂ treatment used for each sample.</td>
</tr>
</tbody>
</table>
Chapter 1: Cancer cytopathology and Raman spectroscopy

1.1 Thesis outline

This thesis explores the use of Raman spectroscopy as an alternative or adjunct tool for the diagnosis of cancer using cytological specimens, in particular non-small cell lung cancer (NSCLC) and thyroid cancer (TC). The ability of Raman spectroscopy to detect cancer is based on the production of a cellular biochemical fingerprint in the form of a spectrum. The first chapter will describe the background of cytopathology and the principle of Raman spectroscopy. As the primary objective of the PhD is based on investigating the use of Raman spectroscopy for the discrimination of NSCLC subtypes, Chapter 2 reviews the recent advances made for the diagnosis of NSCLC using vibrational spectroscopy. The spectral processing and analytical methods used throughout the thesis are detailed in Chapter 3. The work presented in Chapters 4 and 5 provide a body of evidence that demonstrates the suitability of using Raman spectroscopy on bronchoscopy cytology samples for the detection and differentiation of NSCLC subtypes. In Chapter 6, Raman spectroscopy was applied to bronchial tissue sections with the aim to identify the diagnostically relevant peaks, if any, discriminating the negative tissue and malignant tissues, and to comparatively analyse the diagnostic effectiveness of applying Raman spectroscopy to cytology specimens and tissue specimens.

As the secondary objective of the PhD is based on investigating the use of Raman spectroscopy for the discrimination of thyroid cancer subtypes, Chapter 7 demonstrates the ability of Raman spectroscopy to discriminate between thyroid cancer subtypes using cell line models. The first section of Chapter 8 describes a blood removal protocol, developed using cell lines, to prepare thyroid FNAC (fine needle aspiration cytology) samples for Raman analysis. The blood removal method is then applied thyroid FNAC samples taken from surgically removed thyroid specimens and the method is optimised
further, allowing the identification of molecular differences between benign and malignant thyroid samples. In the final section of Chapter 8, the application of Raman spectroscopy to residual thyroid FNAC samples is achieved with further optimisation of the blood removal protocol.

1.2 Cytopathology

1.2.1 Introduction to cytopathology

Cytology, the study of cells, and pathology, which is the study of disease, have been combined since the 19th century to form a branch of cytology known as cytopathology, which studies and diagnoses diseases at a cellular level. Cytological analysis allows a more complete analysis of cell components in comparison to tissue sections, as cytology allows whole cells to be visualised on a slide, whereas tissue sections are often cut and viewed in profiles. Another advantage cytology has over histology is that the procedures used to attain cytological samples for diagnostic purposes are safe and complications rarely occur.

Aspiration cytology, such as fine needle aspiration (FNA) utilises a thin needle to aspirate cells from the area of interest. It is frequently used to investigate readily accessible palpable lesions in cases of thyroid cancer, breast cancer, and malignancy in the lymph nodes. Lesions identified by CT-scan can also be sampled by ultrasound guided FNA. Exfoliative cytology, which is instrumental in cancer detection and screening, involves the collection of exfoliated cells from the surface of the suspicious lesion. Exfoliative specimens are obtained from brushings, washings, urine, sputum, scrapings, and smears, and are routinely used in the diagnostic workup of lung cancer, oral cancer, bladder cancer, skin cancer and cervical cancer. After the clinician obtains the cytology sample, the biological material is stored in a vial with preservative solution, as shown in
Figure 1.1. For cytological analysis, aspirated or exfoliated cells are fixed to a slide and identified microscopically.

![ThinPrep vial for liquid based cytology collection.](image)

1.2.2 Normal cellular morphology and cancer nomenclature

To use cytopathology for the identification of disease, one must first have an understanding of the regular function and growth patterns of the normal cells. Using cytopathology for cancer diagnosis, cells are firstly categorised as either of epithelial or nonepithelial origin, based on patient data and the cell morphology. There are four prominent types of epithelia including squamous epithelium, glandular epithelium, ciliated epithelium, and mesothelium, each have different morphologies\textsuperscript{11}. On cytology, non-cancerous squamous cells are commonly flat, polygonal, vary in size, have distinguished cytoplasm, and they tend to retain their shape during preparation due to high keratin content. Normal glandular epithelium generally presents as cuboidal or
columnar cells on cytology. Their cytoplasm may be transparent due to the accumulation of secretory products, and the nuclei are typically polarised, and are clear or have moderate granularity with small nucleoli. The cytological presentation of non-cancerous ciliated epithelium is often columnar with cilia bound on one flat surface by a terminal plate. These cells are regularly seen in respiratory tract. Lining the body’s cavities, mesothelial cells are cuboidal on cytology with transparent or lightly granular nuclei, which may have nucleoli, and a dense perinuclear space. Cancers derived from epithelial origin are classified as carcinomas and can be further classified based on the cell type from which the cancer originated\textsuperscript{11}. Squamous cell carcinomas are derived from squamous epithelium, glandular epithelium give rise to adenocarcinomas, and carcinomas originating from the mesothelia are called mesotheliomas. Conversely, cancers derived from non-epithelial origins include sarcomas (malignancies of muscle, connective tissue and bone), leukemias (cancers of blood cells), and other cancers of specialised tissues such as lymphomas, which are cancers of the lymphatic system\textsuperscript{11}.

1.2.3 The identification of neoplastic changes with cytopathology

Microscopically observable changes in cell morphology have been recognised as indicators for cancer since the 19\textsuperscript{th} century \textsuperscript{1}. Stains are routinely applied to allow the observation of cells. The field of cytopathology advanced in the 1930s and 1940s with the development of the Papanicolaou (Pap) stain, which provides a visible contrast between cell components and different cell types, allowing the observation of an array of cytoplasmic and nuclear features \textsuperscript{1,12}. Cell chromatin pattern is revealed by the Haematoxylin stain, and cytoplasmic contents are highlighted with acidic stains. Keratin is stained a bright orange and metabolically active cells are stained as green, enabling the differentiation and accurate identification of cell types\textsuperscript{12}.
Knowledge of normal cell configurations permits abnormalities and pathological processes such as carcinogenesis to be identified\textsuperscript{11,13}. Cell abnormalities indicative of cancer includes changes in the cell size or shape, cell nuclear size, nuclear configuration, and changes in the number and size of nucleoli, in comparison to normal cells of the same origin. Specific nuclear aberrations may be used as prognostic indicators, and can be characteristic of certain tumour types. For example, nuclear polylobulation is a typical characteristic of many adenocarcinomas, and diffuse heterochromatin is a prevalent trait in small cell lung cancers\textsuperscript{11,14}. Recognising the abnormal cytological patterns associated with cancer is imperative for making an accurate diagnosis on stained cytology samples. Figure 1.2 depicts morphologically normal and abnormal squamous cells and bronchial cells stained with a Pap stain.

Figure 1.2 A-D are stained with the Pap stain to provide contrast between cell types and cell components. Slides were prepared from BALs using a ThinPrep 2000. A The flat,
polygonal cells marked by red arrows are indicative of morphologically normal squamous cells. Image was taken at x200 magnification. B The blue arrow indicates morphologically normal bronchial cells. The ciliated epithelium is columnar with cilia bound on the flat surface by a terminal plate. Image was taken at x400 magnification. C Depiction of malignant squamous cells with abnormal nuclei and increased keratin in the cytoplasm, evident by the uptake of the orange OG6 stain. Image was taken at x1000 magnification. D Malignant bronchial cells forming an acinar structure with pale chromatin and rounded edges. Image was taken at x400 magnification.

As well as the Pap stain, Immunohistochemistry (IHC) is another diagnostic method commonly applied to cytological samples to aid cancer diagnosis. IHC is concerned with the identification and distribution of cellular components through pinpointing specific antigens. The application of IHC can be used to identify specific cell types by labelling tissue specific biomarkers. The method is based on antigen-antibody interactions, where the location of antibody binding is visualised by direct or secondary antibody labelling methods.

1.2.4 Benefits and limitations of current methods

Cytopathology for cancer diagnosis is a cost effective, accurate and safe diagnostic tool, commonly used for breast, thyroid, lung, and pancreatic cancers. Overall diagnostic sensitivities for breast, thyroid, pancreatic and lung cancers range from 77 – 100%, and diagnostic specificities range from 75.4% - 100%, using cytology. Despite the benefits of cytology for cancer diagnosis, a substantial rate of false negatives is reported, with indeterminate and false negatives on cytology of 1.2-10.6% of breast cancer cases, 20% of pancreatic cancer cases, 16% of squamous cervical cancers, 52% of glandular cervical cancers, and over 20% of thyroid cancer cases being reported. With current diagnostic techniques, well differentiated tumours exhibiting minimal neoplastic changes, and benign features indistinguishable from malignant features lead to diagnostic challenges for cytopathologists. As heterogeneity is a feature of many tumours and morphological interpretation can be difficult, the use of adjunct diagnostic tools such as
IHC are used to improve the accuracy of diagnosis and reduce the rate of indeterminate and false diagnoses. Despite the effectiveness of IHC, the use of labels inhibits subsequent testing of the cellular material.

To overcome current limitations and maximise the cell sample, there is a need for a new label-free diagnostic tool for cytology that can detect neoplastic changes not evident by cellular morphology, and can objectively distinguish benign and malignant cells. As many enhanced diagnostic techniques such as nucleic acid amplification techniques and biomarkers are destructive of the diagnostic sample, they do not overcome limitations facing cytopathology. The application of deep learning technology in medical image classification is a new field for cancer detection, and automated image interpretation is very promising, yet deep learning for medical image processing is still in its infancy and most likely will not aid diagnostic cytology in the near future. Vibrational spectroscopic techniques, including Fourier transform infrared (FTIR) spectroscopy and Raman spectroscopy have been shown to be highly sensitive tools for detecting malignancy, however Raman spectroscopy is better suited to cytology due to having a smaller laser point size that can target single cells and cellular components. Raman spectroscopy provides comprehensive detail about the intracellular biochemistry in a non-destructive manner and without the use of labels. As Raman spectroscopy is being investigated here as an alternative or adjunct diagnostic method for cytology, it will be discussed in the next section in more detail.

1.3 Raman spectroscopy

1.3.1 Raman spectroscopy for cancer diagnosis

Using Raman spectroscopy on cytological material could present as a new minimally invasive diagnostic tool for cancer, as this spectroscopic technique has been shown to be highly sensitive to neoplastic changes. Raman spectroscopy is a label-free diagnostic
technique which requires minimal sample preparation and allows the rapid acquisition of detailed biochemical information. As the energy of the detected scattered radiation corresponds to vibrational energy modes of specific molecules, a detailed spectral fingerprint of a cell can be obtained, allowing for an objective assessment of the cellular components. Statistical analysis of the spectra is needed to expose details about the biochemical composition of a sample, and any minute molecular and structural changes induced by a cancerous state can be detected.

1.3.2 Principle
Raman spectroscopy is based on a scattering phenomenon. Scattering results from the interaction between the photons and materials, such as molecules. As shown schematically in figure 1.3, most photons colliding with molecules do not change their energy after the collision. These collisions are called elastic collisions and the result of these collisions is elastic scattering called Rayleigh scattering. Rayleigh scattering is a two-photon process as a result of which there may be a change in direction of light but no net change in frequency. On the other hand, inelastic collisions, after which the incident photons change their energy, lead to inelastic scattering during which exchange of vibrational energy occurs. The process is defined as Raman scattering. Raman scattering is also a two-photon process by which a change in the energy of the colliding photon occurs from $h\nu_0$ to $h\nu_0 \pm h\nu_s$ where $\nu_0$ is the incident photon frequency, $\nu_s$ is the frequency of the characteristic vibration of the scattering molecule and $h$ is Planck’s constant.
For a molecular vibration to be Raman active, the net molecular polarisability must be changed. The polarisability ($\alpha$) represents the ability of an applied electric field, $E$, to induce a dipole moment, $\mu_0$, in an atom or molecule. This process is represented mathematically by the equations below:\textsuperscript{31–34}:

\[ \mu = \alpha E \] \hspace{1cm} \text{Equation 1.1}

where $\alpha$ is the polarisability and $E$ is the strength of the electric field of the applied electromagnetic radiation. This external electric field varies with time $t$ and oscillates at frequency $\nu$ with amplitude $E_0$. This can be given as:

\[ E = E_0 \cos 2\pi \nu t \] \hspace{1cm} \text{Equation 1.2}
This external oscillating electric field will induce an oscillating dipole moment in the molecule, the frequency of which will be the same as that of the electric field. This is given as:

\[ \mu = \alpha E_0 \cos 2\pi vt \]  \hspace{1cm} \text{Equation 1.3}

For a molecule vibrating with a frequency \( \nu_0 \), the nuclear displacement \( Q \) is written as

\[ Q = Q_0 \cos 2\pi \nu_0 t \]  \hspace{1cm} \text{Equation 1.4}

where \( Q_0 \) is a vibrational amplitude of nuclear displacement.

The polarisability of a molecule bond thus varies with nuclear displacement. \( \alpha \) is a linear function of \( Q \) if the amplitude of vibration is small, the polarisability of the molecule can be approximated using the Taylor series expansion:

\[ \alpha = \alpha_0 + \left( \frac{\delta \alpha}{\delta Q} \right) Q + \ldots \]  \hspace{1cm} \text{Equation 1.5}

where \( \alpha_0 \) is the equilibrium polarisability, \( \left( \frac{\delta \alpha}{\delta Q} \right)_0 \) is polarisability derivative which describes the change in polarisability.

Combining equations 1.4 and 1.5 yields:

\[ \alpha = \alpha_0 + \left( \frac{\delta \alpha}{\delta Q} \right)_0 Q_0 \cos 2\pi \nu_0 t \]  \hspace{1cm} \text{Equation 1.6}

and substituting this value for \( \alpha \) into equation 1.3 gives:

\[ \mu = \alpha_0 E_0 \cos 2\pi vt + \left( \frac{\delta \alpha}{\delta Q} \right)_0 Q_0 E_0 (\cos 2\pi \nu_0 t) (\cos 2\pi vt) \]  \hspace{1cm} \text{Equation 1.7}

Finally, a standard trigonometric identity can be applied:

\[ \cos A \cos B = \frac{1}{2} \left( \cos (A - B) + \cos (A + B) \right) \]  \hspace{1cm} \text{Equation 1.8}
This allows equation 1.7 to be written as:

\[
\mu = \alpha_0 E_0 \cos 2\pi vt + \left( \frac{\delta \alpha}{\delta \Omega} \right)_0 \frac{Q_0 E_0}{2} \left[ \cos 2(v - v_0) t + \cos 2\pi (v + v_0) t \right]
\]  \hspace{1cm} \text{Equation 1.9}

The oscillating polarisation reradiates light at the same frequency and thus the scattered light has components of the frequencies represented in Equation 1.9. The first term of equation 1.9 represents Rayleigh scattering (the frequency of the scattered photon is equal to the frequency of the incident photon, \(v\)), while the second term corresponds to the Raman scattering of frequency (Stokes scattering) and (anti-Stokes scattering). At ambient temperature, most of the molecules are in their vibrational ground state. According to the Boltzmann distribution, a much smaller number of molecules are in a vibrationally excited state. Therefore, the chances for a Raman interaction which transfer vibrational energy to a molecule and leave a quantum of lower energy are much higher than an interaction which leaves a quantum of higher energy. Raman scattering processes are referred to as Stokes, resulting in quanta of lower energy, or anti-Stokes resulting in quanta of higher energy \(^{32}\). At room temperature, as most of the molecules are generally in the ground vibrational state, Stokes Raman scattering most commonly occurs and is usually measured. A plot of intensity of scattered light versus energy difference between the incident light and the Raman scattered light is called Raman shift, is a Raman spectrum. Raman shift is usually expressed in wavenumbers (cm\(^{-1}\)).

1.3.3 Instrumentation

The source of monochromatic laser light can be different wavelengths, depending on the intent of the experiment. The intensity of the laser light, which is adjustable, is determined by a neutral density filter or by pin holes. In order to ensure only the laser light stimulates the sample, an interference filter acts as a clean-up tool. Joining the Raman spectrometer with a microscope makes it possible to analyse biological samples as the objective lenses
deliver the laser while collecting the back scattered light. To inhibit interference from Rayleigh scattering, the notch filter reflects scattered light of the same wavelength as the incident laser and allows inelastically scattered light to pass through. A grating then disperses the light. The spectral resolution is determined by multiple factors including: the groove intensity of the grating (typically 300-1800 grooves/mm), the wavelength of the excitation source, and the distance between the grating and the detector. A charge coupled device (CCD) detector is used to collect the light as it allows a multichannel operation, so that the entire Raman spectrum can be detected in a single spectrum. A schematic of the HORIBA-Jobin-Yvon XploRA Raman spectrometer is shown in figure 1.4.

The material the sample is fixed to, the substrate, may contribute background noise to the spectrum. Glass is the most applicable substrate for the clinical setting as it is cost effective, and the signals derived from glass can be removed computationally to derive the cellular signal\textsuperscript{35}. A 785 nm laser source is capable of revealing comprehensive detail about the intracellular biochemistry of a sample, however glass cannot be used a substrate with this laser line as the fluorescence is strong at 785 nm, swamping the biological peaks in the Raman spectrum. As the 785 nm laser requires more expensive substrates, this laser line may not be suitable for clinical application, and may be better utilised in research. Another difficulty in translating Raman spectroscopy to routine clinical use is the insensitivity of the technique, as only 1 in 1x10\textsuperscript{8} photons are Raman scattered\textsuperscript{36}. Higher wavelength laser lines such as 1064 nm, produce weaker Raman signals, and may not be optimal for use in clinics. In order to be effective in a clinical setting, the method of Raman spectroscopy utilised must therefore be cheap and maximise sensitivity. In this project, a 532 nm laser was utilised to maximise the Raman signal, and glass was used as a substrate to support the tissue.
1.3.4 Raman spectroscopic analysis of cytological samples

Raman spectroscopy has a multitude of attributes making it suitable for this study. As cytological specimens are less invasive than tissue biopsies, an accurate diagnostic technique which is applicable to cytology samples would reduce the need for invasive biopsy procedures. With a laser resolution of 1 µm, Raman spectroscopy is capable of analysing single cells and cellular components and is an ideal technique for investigating cytology samples. Further, this spectroscopic technique has demonstrated the ability to detect malignancies in a multitude of studies. In particular, the diagnostic capabilities of applying Raman spectroscopy to cytology samples has previously been demonstrated with bladder cancer, oral cancer, and cervical cancer.
1.4 Summary
Cytological methods are minimally invasive and are routinely used for the diagnosis of many cancers. However, many tumours are heterogeneous and morphological interpretation can be difficult with current staining and IHC techniques. Raman spectroscopy is a label free vibrational spectroscopic technique based on the inelastic scattering of light, and has been shown in many studies to be a promising tool for cancer diagnosis. As Raman spectra contain detailed biochemical information, statistical analysis can be applied to the spectral dataset to detect molecular alterations associated with cancer, allowing an objective analysis of cytological specimens. If Raman spectroscopy shows the ability to differentiate tumour types based on their biochemical fingerprint, then the potential to improve the diagnostic accuracy of cytology and preserve residual material for subsequent molecular analysis for therapeutics is of clinical value.

1.4.1 Thesis objectives
The broad objectives of this thesis are to investigate the feasibility of applying Raman spectroscopy to cytological specimens to detect non-small cell lung cancer (NSCLC) and thyroid cancer. A more detailed description of the aims are;

1) To explore the feasibility of applying Raman spectroscopy to bronchoalveolar lavage (BAL) and bronchial wash (BW) samples to provide a marker free diagnosis of NSCLC subtypes, and to identify the biochemical changes responsible for the classification.

2) To investigate the potential of Raman spectroscopy to identify malignant signatures in NSCLC cases with clinical cytological analyses reported as diagnostic or no malignant cells seen (NMCS).
3) To analyse Raman spectra taken from bronchial tissues and compare the diagnostic effectiveness of applying Raman spectroscopy to bronchoscopy attained cytological samples and tissue samples for NSCLC diagnosis.

4) To investigate the Raman profiles of thyroid cancer cell lines, and to build a classifier based on the Raman spectra to discriminate between the benign and malignant cell lines.

5) To optimise a blood removal protocol suitable to prepare thyroid FNAC samples for Raman analysis.

6) To discriminate between benign and malignant cells from thyroid FNAC samples.

References


Chapter 2: Recent advances in the vibrational spectroscopic diagnosis of non-small cell lung cancer


Declan O’Dea¹, ², Fiona M. Lyng², ³, Siobhan Nicholson⁴, Finbar O’Connell⁴, Aoife Maguire⁴, Alison Malkin¹

¹ School of Biological and Health Sciences, Technological University Dublin, Dublin, Ireland
² Radiation and Environmental Science Centre, FOCAS Research Institute, Technological University Dublin, Dublin, Ireland
³ School of Physics & Clinical & Optometric Sciences, Technological University Dublin, Dublin, Ireland
⁴ St James's Hospital, James's Street, Dublin 8.
Abstract

Lung cancer is the leading cause of cancer deaths worldwide accounting for 1.69 million deaths in 2015. Studies have indicated a 5 year survival rate of 8% to 15% in western countries, although a survival rate as low as 1% has been demonstrated for late stage diagnosis. With the advent of targeted therapies, it is imperative to accurately differentiate non-small cell lung cancer (NSCLC) subtypes in order to ensure efficacy of treatment for patients. Immunohistochemistry and molecular techniques for the diagnosis of NSCLC are increasingly part of the diagnostic algorithm and clinical work-up of lung cancer patients, however due to the limitation of small sample size, overlapping morphological features and molecular characterisation, differential diagnosis of NSCLC still proves challenging. Vibrational spectroscopy has shown promising results for the detection of a variety of cancers and a limited number of studies have focused on lung cancer. Yet to date there has been no published evaluation of vibrational spectroscopy on cytology bronchoscopy samples which may eliminate the necessity for an invasive biopsy procedure. Following an introduction to the epidemiology and etiological factors associated with NSCLC, currently used diagnostic methods and their limitations are presented. A thorough review of Raman and FTIR spectroscopic methods in lung cancer diagnosis is then presented. On review of the literature, vibrational spectroscopy offers an alternative or adjunct diagnostic method to be applied in bronchoscopy cytology samples.
2.1 Lung cancer

2.1.1 Introduction to Lung Cancer

With an estimated 2.1 million new cases and 1.76 million deaths in 2018, lung cancer persists as the most common cancer worldwide, and represents the highest cancer incidence and mortality rates in both developed and less developed countries \(^1,2\). Studies have indicated a 5 year survival rate of 8% to 15% in developed countries, although a survival rate as low as 1% has been reported for late stage diagnosis \(^3\text{-}5\). Global lung cancer incidence trends indicate that although western regions such as Northern America and Europe have the highest rates of incidence, the rate of incidence is declining in these regions, while increasing in less developed regions \(^6,7\). These trends can be directly correlated with smoking prevalence, as cigarette smoking is responsible for approximately 85% of lung cancer cases \(^8\). Other risk factors for non-smokers include passive cigarette smoking and air pollution \(^9\).

Lung cancer is divided into two main categories, small cell lung cancer (SCLC) accounting for an estimated 13%-14% of cases, and non-small cell lung cancer (NSCLC), which is the most common form of lung cancer accounting for approximately 85% of cases \(^10\text{-}13\). Furthermore, NSCLC includes different histological subtypes including squamous cell carcinoma (SCC), adenocarcinoma (AC) and large cell carcinoma. AC is more prevalent in non-smokers and accounts for the largest portion of all lung cancer cases at 44%, SCC is the second most common subtype, accounting for approximately 26% of lung cancer cases \(^10\). An epidemiological study\(^14\) examining sex differences in lung cancer incidence and survival reported that women are more commonly diagnosed at an earlier stage of lung cancer, and at an earlier age than men. The study also indicated that regardless of age and stage of the disease at diagnosis, women have an increased survival in comparison to men, and this is most evident with the AC subtype. Irrespective
of sex, patient prognosis is dependent on the stage of the disease at diagnosis. Table 2.1 illustrates the approximate 5 year survival rates for different stages of NSCLC.

Table 2.1. Progression stage, estimated 5 year survival rate, and appropriate treatments for each stage of NSCLC.

<table>
<thead>
<tr>
<th>Stage of NSCLC</th>
<th>Progression stage</th>
<th>Approximate 5-year survival rate (%)</th>
<th>Treatment</th>
</tr>
</thead>
<tbody>
<tr>
<td>Stage IA</td>
<td>Localised (&lt;3cm)</td>
<td>75</td>
<td>Lobectomy/radiotherapy</td>
</tr>
<tr>
<td>Stage IB</td>
<td>Localised (3-5cm)</td>
<td>55</td>
<td>Lobectomy/radiotherapy</td>
</tr>
<tr>
<td>Stage IIA</td>
<td>Localised (5-7cm) OR Spread to nearby lymph nodes (&lt;5cm)</td>
<td>50</td>
<td>Lobectomy/adjoint chemotherapy/radiotherapy</td>
</tr>
<tr>
<td>Stage IIB</td>
<td>Spread to nearby lymph nodes/diaphragm/mediastinal pleura/parietal pericardium/bronchus (5-7 cm) OR no spreading and &gt;7 cm</td>
<td>40</td>
<td>Lobectomy/adjoint chemotherapy/radiotherapy</td>
</tr>
<tr>
<td>Stage IIIA</td>
<td>Spread to local lymph nodes/heart/trachea/oesophagus/phrenic nerve/ present in more than one lobe (&gt;7 cm)</td>
<td>10-35</td>
<td>Radiotherapy/ chemotherapy/immunotherapy/ tyrosine kinase inhibitors</td>
</tr>
<tr>
<td>Stage IIIB</td>
<td>Spread to lymph nodes of opposite lung OR spread to lymph nodes in the mediastinum and one or more of the areas mentioned in stage IIIA or a main blood vessel</td>
<td>&lt;5</td>
<td>Radiotherapy/ chemotherapy/immunotherapy/ tyrosine kinase inhibitors</td>
</tr>
<tr>
<td>Stage IV</td>
<td>Metastasised to another body part, other lung, caused malignant pleural effusion/pericardial effusion</td>
<td>&lt;5</td>
<td>Combination of cytotoxic chemotherapies/ palliative care/ tyrosine kinase inhibitors/radiotherapy</td>
</tr>
</tbody>
</table>
2.1.2 Current Treatment Options

Therapeutic strategies and patient management are dependent on the stage, subtype, and mutation status of the malignancy. Patients diagnosed with early stage (stage I and stage II) lung cancer may avail of potentially curative treatment through surgical resection. Although patients with early stage lung cancer have a 5 year survival rate of over 70%, and chemotherapy post-surgery has been shown to benefit some stage II patients, the risk of metastatic recurrence after surgery remains as high as 70%. However, over half of patients are diagnosed at advanced stage and are not eligible for surgical resection. Treatments available for advanced stage disease include platinum based chemotherapies, radiotherapy, immunotherapies and molecular targeted therapies, including tyrosine kinase inhibitors (TKIs). In recent years combining chemotherapy with the appropriate TKIs has been shown to increase progression free survival of patients with advanced stage AC, although there is currently no cure.

Understanding the underlying molecular pathways involved with cancer cell progression drives the development of novel therapeutics, and identifying and targeting molecules that drive neoplastic proliferation has become a major tool for combatting lung cancer. Specific therapies that inhibit carcinogenic pathways are now available for NSCLC patients with genetic aberrations such as epidermal growth factor receptor (EGFR) mutations, BRAF mutations and anaplastic lymphoma kinase (ALK) fusion translocations. The EGFR gene encodes a receptor tyrosine kinase and mutations of this gene are commonly observed in lung AC. EGFR mutations are reported in approximately 50% of Asian AC patients, and 20% of Western patients with AC. Patients that are EGFR mutation positive have exhibited sensitivity to TKIs, such as Gefitinib, which has been implemented in treatment regimens since 2003. Another molecular targeted therapy, Crizotinib, is recommended as first-line treatment for patients...
with ALK gene rearrangements, which are present in approximately 3-6% of AC patients, particularly in non-smokers 19,32–34. The BRAF gene encodes a protein in the serine threonine kinase family, and mutations are present in 1-3% of NSCLC patients 35,36. Two therapies that target proteins in the serine threonine kinase family, dabrafenib and trametinib, have been approved for patients with the BRAF mutation, and have been shown to slow tumour growth 37,38.

As EGFR, ALK and BRAF mutations are targetable aberrations observed in AC, and as TKIs are associated with increased progression free survival and a lower toxicity than chemotherapeutic agents, molecular testing for these mutations on bronchoscopy cytology samples has now become a major part of the diagnostic work up of NSCLC, and is considered imperative for formulating treatment strategies 26,39–41.

Chemotherapy remains the cornerstone treatment strategy for most patients with SCC, as targeting the potentially actionable mutations currently shows little therapeutic benefit for patients with SCC. Although, the recent approval of the monoclonal antibody, Necitumumab, as first-line therapy in metastatic SCC has been shown to improve overall survival of patients with EGFR mutations 42.

The use of immunotherapeutics for patients with SCC is increasing. With the recent approval of three immune checkpoint inhibitors, nivolumab, atezolizumab and pembrolizumab, for SCC patients with PD-L1 positive tumours, patients receiving immunotherapy demonstrate prolonged overall survival, increased objective response rate and progression free survival, compared with systemic chemotherapy 21–25,43–45.

Testing for PD-L1, which is an immune molecule capable of inhibiting an anticancer immune response, is now part of the routine diagnostic algorithm of SCC 46–49, and expression is measured by the tumour proportion score. Patients with a proportion score greater than 50% may avail of first-line treatment with immunotherapy.
The increasing preference for personalised therapeutics for the treatment of NSCLC, is not only due to the negative side effects imparted on patients by non-selective chemotherapy, but also the increased progression free survival exhibited by patients in receipt of targeted therapies compared to chemotherapy \(^{50,51}\). Research for novel therapeutic markers is ongoing and includes studies investigating targetable molecules that drive carcinogenic progression such as FGFR, VEGFR-2, DDR2, P13 Kinase and PDGF, as well as novel immunotherapies\(^{49,52}\). As more targets and subsequent therapies are discovered, retaining adequate sample during the diagnostic work up of NSCLC for molecular testing will be imperative for future treatment planning.

2.1.3 Current Methods for Differential Diagnosis of NSCLC Subtypes

With the introduction of targeted therapies and immunotherapies, distinction between NSCLC and small cell lung cancer (SCLC) although paramount, is no longer adequate and further sub-classification of NSCLC is required. Although the most common symptoms associated with lung cancer are cough, dyspnea, chest pain and haemoptysis, lung cancer often presents with no symptoms until it has reached advanced stages, resulting in the majority of patients presenting with distant metastasis of the disease at diagnosis\(^{16,53}\). The following paragraphs detail the current methods used to acquire lung samples for analysis and the diagnostic tests used.

2.1.4 Bronchoscopy

Bronchoscopy is employed as an investigatory technique subsequent to the detection of an abnormality within the lung using CT scanning or X-rays\(^ {54}\). Flexible bronchoscopy was first developed during the 1960s and remains one of the principal diagnostic methods in pulmonology. In the case of suspected lung cancers, the main focus of bronchoscopy is to attain a sample of the suspicious lesion for diagnostic analysis\(^ {55}\). At present, morphological evaluation of small specimens is the principal method used to
differentially diagnose lung cancers. Accordingly, flexible bronchoscopy has evolved with the addition of adjunct applications including navigational bronchoscopy, endobronchial ultrasound, and bronchial thermoplasty, making it an effective method for obtaining minimally invasive cytology samples.

2.1.5 NSCLC Tumour Classification Utilising Cytology Specimens

Small biopsy and cytology samples are the primary methods used for NSCLC diagnosis as these inflict less discomfort on the patients and still allow accurate identification of malignancy. Abrasive and exfoliative methods such as bronchial washing, brushing and bronchoalveolar lavage (BAL) are frequently utilised to attain samples from bronchoscopy for cytological analysis. According to international multidisciplinary classification systems approximately 70% of lung cancer histological subtypes are diagnosed by assessing the predominant cytological pattern of small biopsies and cytological specimens, making cytological analysis a commonly implemented diagnostic method for NSCLC.

When a tumour does not show standard morphologic criteria for SCC or AC, immunohistochemical markers are used for further classification. Current guidelines suggest a vigilant use of immunohistochemistry (IHC) for diagnosis as limited cell sample remains for molecular testing for therapeutics, such as Fluorescent in situ hybridisation (FISH) and molecular sequencing.

2.1.6 Immunohistochemistry

IHC is concerned with the identification and distribution of cellular components through pinpointing specific antigens. The application of IHC can be used to identify specific cell types by labelling tissue specific biomarkers. The method is based on antigen-antibody interactions, where the location of antibody binding is visualised by direct or secondary antibody labelling methods. This is an effective technique used to
discriminate between the different subtypes of NSCLC, and double IHC staining has been shown to be effective in the detection of AC 65.

A systematic review by Desai et al. 29 analysed data published between 2000 and 2012 and identified IHC as the most commonly used molecular testing technique for identifying biomarkers for the differential diagnosis of SCC and AC. The most frequently used biomarkers for this distinction are thyroid transcription factor 1 (TTF-1), EGFR, and the tumour suppressor gene p63 29,49,50. TTF-1 modulates genes responsible for pulmonary morphogenesis and differentiation, and has exhibited elevated immunoreactivity in ACs 66,67. The genomic profile of AC has also displayed an association with EGFR mutations 68, therefore the high prevalence of EGFR mutations in AC, alongside the recent development of molecular targeted therapies, makes it an effective marker for diagnosis as well as treatment strategy 69. In contrast, p63 has been noted to display significantly higher positivity in SCCs than in ACs, making it a useful diagnostic indicator for SCC 66.

While IHC may be required to aid diagnosis in some cases, the technique is becoming increasingly applied in the realm of therapy prediction. It is not only imperative for the detection of treatment sensitive tumours, but also for the identification of those patients who may suffer ill effects where specific agents may be contraindicated. Despite the efficacy of IHC for the differential diagnosis of NSCLC subtypes, this diagnostic technique has multiple limiting factors. Although dual staining methods are now applicable with IHC, an additional slide is still required, significantly reducing the scant cellular material available for subsequent molecular testing. IHC also relies on the use of large molecules such as antibodies to label molecular structures, which interfere with further investigation of intrinsic molecular properties 70,71.
2.1.7 Fluorescent in situ Hybridisation

FISH is an effective molecular testing technique for identifying fusion genes and has been approved by the FDA as the gold standard method for the detection of ALK rearrangements in NSCLC.\(^{72,73}\) As IHC does not directly detect the ALK fusion gene, not all fusion positive tumours are positive using IHC.\(^{74,75}\) FISH however has the ability to directly detect ALK fusions and also has an overall sensitivity and specificity higher than IHC for the identification of ALK gene aberrations in NSCLC.\(^{74,76}\) Despite being a powerful technique needed for ALK testing and identifying patients that may respond to molecular targeted therapy, FISH requires the limited residual sample left over from diagnosis.

2.1.8 Current Research on Genomic Profiling

In the era of personalised therapies and as new molecular markers are discovered, molecular testing is becoming part of the routine diagnostic workup of NSCLC. Current research is investigating potential molecular markers which may act as therapeutic targets or predictors of response to therapy.\(^{77–79}\)

Genomic and molecular profiles of AC and SCC are being developed using techniques applicable on archival cancer samples such as mass spectrometry and next generation sequencing. These techniques have been used to investigate the genomic profile of AC and have identified some high frequency mutations associated with the AC subtype, such as EGFR, MET, FGFR and HER2.\(^{39,77–79}\)

Genomic hybridisation has been used in a number of studies to explore the gene-phenotype relationship of NSCLC subtypes, and research using this technique has identified entire chromosomal regions and individual genes which are applicable for the discrimination of AC and SCC. Using residual bronchial brush samples, Su et al.\(^{80}\) identified 19 genes that are applicable for the discrimination of SCC and AC, many of the
genes associated with epidermis development and cell adhesion and may present as potential therapeutic markers. Lockwood et al.\textsuperscript{81} identified an entire chromosomal region (8p12) which may be valuable for profiling NSCLC subtypes. A gene within this region, BRF2, was identified as a SCC specific genetic amplification, and indicates that this gene is a promising marker for SCC. As BRF2 encodes a protein involved in transcription initiation and facilitates in transcribing proteins required for cell growth, it may present as a novel target for future therapies.

With the expanding demand for genomic profiling there is an increased demand on limited cell sample. The optimisation of diagnostic processes is therefore paramount to retain cellular material for subsequent molecular analysis such as genomic profiling.

2.1.9 Limitations of Current Diagnostic Methods

Although IHC is a highly accurate technique for discriminating SCC and AC, this diagnostic technique significantly reduces the amount of sample available for subsequent molecular testing, and as previously discussed, molecular profiling of the tumour is paramount in order to ensure efficacy of treatment. The demand for limited samples is ever increasing, this drives the need for an alternative diagnostic method which can be applied to small tissue and cytology samples without compromising subsequent applications \textsuperscript{70,71}. In response to the need for new diagnostic methods, the investigation of alternative methods such as vibrational spectroscopy has been growing in recent years. The following section will discuss vibrational spectroscopy methods, such as Fourier Transform Infrared (FTIR) spectroscopy and Raman spectroscopy, as alternative techniques for the diagnosis of lung cancer.

2.2 Vibrational Spectroscopy

Vibrational spectroscopy uses electromagnetic radiation to probe vibrations within molecules, and allows analysis of such molecular vibrations \textsuperscript{82}. Molecular vibrational
energies are routinely investigated using Fourier transform infrared (FTIR) spectroscopy and Raman spectroscopy. Figure 2.1 demonstrates the process involved in applying vibrational spectroscopy to biological specimens for cancer detection.

Figure 2.1. Flow diagram of the process involved in applying vibrational spectroscopy to bronchoscopy obtained specimens for cancer detection. After obtaining the cytology/biopsy sample the biological material is stored in a preservative. The tissue section or cellular material is then fixed to a substrate such as a glass slide and left unstained for spectroscopic analysis. Using a laser or an infrared source, spectra are obtained by collecting the Raman scattered light or transmitted light. The spectral data are then pre-processed to minimise noise and spectral contaminations. Diagnostic features are then extracted from the data with multivariate analytical techniques.

2.2.1 FTIR Spectroscopy for Lung Cancer Diagnosis

Fourier transform infrared (FTIR) spectroscopy is a label free vibrational spectroscopic technique, which necessitates minimal sample pre-treatment \(^{82-84}\). When applied to
biological tissue, the FTIR spectrum reveals comprehensive detail about the molecular content of the tissue. Intracellular changes induced by cancer states may be detected as variations from the spectral profile of normal tissue. The use of FTIR spectroscopy for lung cancer diagnosis has been investigated for over 20 years, with previous reports examining the applicability of FTIR spectroscopy for lung cancer investigation using in vitro models and ex vivo samples. The in vitro studies included exploring the ability of FTIR spectroscopy to discriminate normal and cancer cells, and assessing lung cancer cells response to therapy. Lee et al. differentiated normal and lung cancer cell lines and reported differences in nucleic acid and phospholipid bands in the malignant lines. Sulé-Suso et al. demonstrated the ability of FTIR spectroscopy to detect therapeutic response in lung cancer cell lines after the addition of a chemotherapeutic agent. The FTIR spectra indicated an increase in phosphate groups in the cells after the addition of the chemotherapeutic agent. The spectral changes detected were correlated with cell survival, showing that FTIR spectroscopy may potentially identify treatment sensitive cells. The study also demonstrates the suitability of FTIR spectroscopy for single cell analysis.

Although cell line studies provide proof of concept, stronger evidence to support the feasibility of using FTIR spectroscopy for lung cancer diagnosis are provided by applying the technique to surgically resected samples. A pilot study conducted by Sun et al. used surgically resected tissue from 60 patients and applied a variation of FTIR spectroscopy, attenuated total reflection (ATR) FTIR spectroscopy to collect spectra from malignant and non-malignant tissues. The study showed the ability to detect the presence of lung malignancy with a sensitivity, specificity and accuracy of 96.7%. In addition to allowing the classification of neoplastic changes with high diagnostic accuracy, analysis of FTIR spectra reveals the molecular differences between malignant and non-malignant lung
tissue. Yano et al.\textsuperscript{85,90} described the increase in glycogen content of NSCLC tissue sections compared to non-cancerous tissue sections. Akalin et al.\textsuperscript{91} used a machine based learning approach to show that spectral profiles can distinguish between normal and cancerous tissues with high accuracy. They also reported the ability of IR spectroscopy to accurately classify SCLC, as well as the NSCLC subtypes AC and SCC. The authors indicated that this method may potentially be used to identify clinically significant subtypes of AC, although training and test data sets of adequate size were not available for their study. Further classification of AC into prognostically different groups was achieved by Grosserueschkamp et al.\textsuperscript{92}. Using representative spectra for each subtype a supervised classifier was constructed. The first and second level random forests separated between normal tissue, diseased tissue and lung cancer subtypes. The third level random forest enabled further classification of AC into prognostically relevant subtypes, the favourable non-mucinous lipidic subtype, less favourable papillary and acinar subtypes, and the poor solid and micropapillary subtypes. A discrimination accuracy of 97\% for cancer subtypes and 95\% for AC subtypes was reported\textsuperscript{92}.

As cytological samples are the least invasive samples to attain for the diagnosis of NSCLC, several studies have assessed the ability of FTIR spectroscopy to characterise lung cancer using pulmonary cytological samples. Ghosal et al.\textsuperscript{93,94} have demonstrated the feasibility of applying FTIR spectroscopy to sputum samples to detect lung cancer. Using two wavenumbers (1031.7, 1409.7 cm\textsuperscript{-1}) to develop a predictive model, lung cancer cells were identified with a sensitivity and specificity over 91\%. Lewis et al.\textsuperscript{95} generated FTIR spectra from sputum cell pellets, and identified an increase in glycogen in the lung cancer cells. This finding is in accordance with multiple studies and signifies that elevated glycogen content may be a useful diagnostic biomarker for lung cancer using FTIR.
spectroscopy. These studies demonstrate that applying this powerful diagnostic technique to cytological specimens for the detection of lung cancer in a clinical setting is a realistic prospect. In addition, FTIR spectroscopic analysis of sputum samples may be a cost effective and fast tool for lung cancer screening due to the non-invasive procedure of obtaining sputum samples and high throughput FTIR technology. In order for the application of FTIR spectroscopy to progress from a research setting to preclinical or clinical trials, an interdisciplinary approach is required to validate the performance of spectral biomarkers with a sufficient number of patients. The complexity and current availability of these validated models is a significant limitation for the application FTIR for lung cancer diagnosis.

2.2.2 Raman Spectroscopy for Lung Cancer Diagnosis

Raman spectroscopy has been discussed as a promising tool for cancer diagnostics for over 20 years. Multiple reviews have illustrated how this spectroscopic technique has demonstrated favourable results for the identification of cancerous tissue and has the potential to improve cancer diagnostics in a variety of cancers including brain, breast, skin, lung and gastrointestinal cancer. As with FTIR spectroscopy, Raman spectroscopy is a form of non-invasive, label free vibrational spectroscopy, however Raman spectroscopy relies on the detection of inelastically scattered light. The Raman spectrum is a plot of the intensity of the scattered light versus the change in energy given in wavenumbers (cm$^{-1}$), and represents a detailed biochemical fingerprint of the cellular components, as shown in figure 2.2. As Raman spectroscopy provides comprehensive details about the composition of tissues, biochemical changes in cancerous tissue may be detected in the Raman spectra. Combining Raman spectroscopy with multivariate analysis can enable a highly accurate classification of tissue types and malignancy. In recent years, many studies have shown the efficacy of Raman
spectroscopy for detecting lung cancer \textit{in vitro}, \textit{ex vivo} and intraoperatively. Work by Jess et al.\textsuperscript{103} showed the potential of Raman spectroscopy for grading of lung neoplasia in cellular samples. The cell lines utilised in the study were representative of three groups, normal cells (primary normal bronchial epithelial cells HBEpCs), cells with extended lifespan (HBEpCs retrovirally transduced with either human papillomavirus (HPV) type 16 E7 or CDK4), and immortalised or malignantly transformed cells (BEP2D and AsbTB2A). BEP2D is a human bronchial epithelial cell line expressing HPV18, and AsbTB2A is a transformed cell line derived from BEP2D following exposure to asbestos. A multivariate classification model based on linear discriminant analysis classified the cell types with accuracies ranging from 58\% - 77\%. Normal cells were discriminated from all abnormal cell types with a sensitivity of 91\% and specificity of 75\%. The Raman spectra indicated that normal cells, cells with extended lifespan, and immortalised and transformed cells, could be characterised by their DNA, protein, amide and lipid content, suggesting that Raman spectroscopy may identify stage of lung cancer development\textsuperscript{103}. 

![Graph of Raman spectra](image)

<table>
<thead>
<tr>
<th>Protein/Carbohydrates</th>
<th>Tyrosine/Phenylalanine</th>
<th>Nucleic acids</th>
<th>Protein/Carbohydrates</th>
<th>Phenylation</th>
<th>Amide III/Lipids</th>
<th>NAA/Protein</th>
<th>Protein/Lipids</th>
<th>Amide I/Lipids</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intensity (Normalised)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>400</td>
<td></td>
<td></td>
<td>800</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1200</td>
<td></td>
<td></td>
<td>1600</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Figure 2.2. Depiction of a Raman spectrum showing the different spectral regions and the corresponding biomolecules associated with each region. The spectrum which represents a bronchial epithelial cell nucleus was baseline corrected and the non-negatively constrained least squares method was used to remove spectral contaminants from the substrate.

The efficacy of Raman spectroscopy for the in vitro discrimination of normal and cancerous cells was also investigated by Oshima et al.\textsuperscript{104}. Five lung cell lines from different histological origins were analysed (MRC-5 from normal, RERF-LC-MS from AC, EBC-1 from SCC, Lu-65 from large cell undifferentiated, and RERF-LC-MA from small cell carcinoma). The study showed that step-wise linear discrimination analysis of the Raman spectra allowed the label-free detection of each histological cell type with an accuracy of 100%, validating the effectiveness of Raman spectroscopy for cancer cell diagnosis.

In addition to the assessment of in vitro models, the applicability of using Raman spectroscopy for the assessment of surgically resected samples has been explored. Using a 1064nm laser to eliminate background fluorescence, Kaminaka et al.\textsuperscript{105} investigated the molecular differences between normal and malignant lung tissue. Spectra with a high signal to noise ratio were obtained and distinct Raman bands at 1448 and 1666 cm\textsuperscript{-1} differentiated the normal and cancer spectral profiles. This demonstrated that an increase in collagen content in lung cancer tissue is detectable with Raman spectroscopy.

In 2012 Pavićević et al.\textsuperscript{106} applied Raman spectroscopy to tissue samples from an array of different tumour types, which included lung SCC, and AC. Using PCA and a neural network algorithm, the SCC and AC subtypes were correctly classified with accuracies of 83.3\% and 91.7\% respectively. Normal and tumour tissue spectra were also discriminated with an overall sensitivity of 95\%. Huang et al.\textsuperscript{107} applied Raman spectroscopy to biopsy samples to differentiate normal, AC, and SCC tissue and discovered significant differences between the spectra of normal and cancerous tissue.
The Raman spectra indicated that the malignant tissue was characterised by an increase of nucleic acids, tryptophan and phenylalanine, in comparison to the normal tissue, while the AC and SCC subtypes were discriminated by a different ratio of peak intensities at 1455 cm\(^{-1}\) and 1655 cm\(^{-1}\), corresponding to a variation in lipid to protein ratio.

Gao et al.\(^{108,109}\) applied another form of Raman spectroscopy, Coherent anti-Stokes Raman Scattering (CARS), to frozen lung tissue samples, differentiating normal and benign tumours, cancerous and non-cancerous tissues, and subsets of lung cancers. The authors developed a highly accurate classification system which delineated malignant tissue from normal tissue with classification accuracy over 91%, and successfully differentiated small cell and NSCLC. An investigation by Magee et al.\(^{110}\) using surgically resected samples achieved a sensitivity of 84% and a specificity of 61% for the discrimination of normal and cancer tissue, reporting that tumour tissue could be characterised by an increase in DNA and a decreased level of porphyrin in comparison to the normal tissue. In addition to showing the effectiveness of Raman spectroscopy to detect cancerous tissue, this study was the first to demonstrate a prognostic ability of Raman spectroscopy to predict cancer recurrence in patients with a sensitivity and specificity of 73% and 74% respectively\(^{110}\).

The CARS variant of Raman spectroscopy utilises multiple pulsed laser beams that interact with the sample through a wave mixing process and produces a signal that is much higher than the spontaneous Raman signal. With no interference from fluorescence, it has been shown to be a suitable method for the differential diagnosis of NSCLC, while preserving tissue specimens for follow up diagnostic tests\(^{111-113}\). Xu et al.\(^{112}\) applied CARS as part of a multimodal imaging technique to unfixed, unstained lung tissue to effectively differentiate normal lung tissue, cancer tissue, and desmoplastic tissue. In another study by Gao et al.\(^{113}\), a label free diagnostic system was developed by applying
a 3D imaging strategy in conjunction with CARS to differentiate NSCLC subtypes. Human AC and SCC cell lines were grown in mouse models, and the tumours were excised after 2 weeks of growth. By acquiring individual 2D images of the same field of view from different imaging depths, the authors created a 3D data structure of the tissue. The system enabled a more accurate analysis of whole cells and nuclear size, overcoming the limitations of 2D sections. The NSCLC subtypes were differentiated with an accuracy of 97%, proving the ability of vibrational spectroscopy to improve the current diagnostic algorithm with minimal sample consumption and without hindering additional molecular testing.

Surface enhanced Raman spectroscopy (SERS) has also been applied to biological samples including saliva and cell derived exosomes to characterise biomolecular differences between normal lung and lung cancer. Zhang et al. examined the prospects of implementing SERS in lung cancer diagnostics by applying SERS technology to lung tissue sections and developing a PC-LDA classifier to analyse the data. The study identified characteristic changes in the cancer spectra associated with increased DNA, carotenoids, lipids and proteins. After performing PC-LDA, a ROC curve was constructed to evaluate the performance of the developed classifier, and a diagnostic sensitivity of 95.7% and specificity of 95.7% was achieved for the discrimination of normal and malignant lung tissue.

The use of Raman spectroscopy to improve the in vivo diagnosis of lung cancer has also been investigated and presents as an exciting clinical application for the real-time diagnosis of NSCLC, which may prevent the need for surgical biopsy procedure. In 2008 Short et al. designed an endoscopic Raman probe and acquired the first in vivo lung Raman spectra. With 1-2 second integration times and filters to reduce noise and fluorescence, quality Raman spectra were acquired in the high frequency spectral range.
from normal, dysplastic, and tumour sites. Analysis of the spectral profiles showed large variations between the SCC and normal tissue, although due to the low number of patients involved in the study the authors could not rule out interpatient variability as a potential cause. The following year a pilot study by Magee et al.\textsuperscript{118} then provided proof of concept for the suitability of applying Raman spectroscopy for the \textit{in vivo} diagnosis of lung cancer. Using a biomedical filtered fibre optic probe and a method to reduce background fluorescence, clear Raman spectra were acquired from \textit{ex vivo} normal and malignant lung tissue. Spectral variation analysis identified increased peak intensities for amide I (1655 cm\textsuperscript{-1}), amide III (1260 cm\textsuperscript{-1}), and phenylalanine (1002 cm\textsuperscript{-1}) in the tumour spectra. The normal and tumour spectra were also classified with 100\% accuracy using PCA leave one out cross validation. In 2016, a single centre clinical investigation with 80 patients explored the adjunct application of Raman spectroscopy with bronchoscopic methods for the \textit{in vivo} detection of lung cancer \textsuperscript{119}. Real time point laser Raman spectroscopy (1 second per spectrum) was performed on 280 sites from non-malignant and malignant lung lesions. Multivariate analysis of the spectra differentiated the normal, benign and malignant tissue with a sensitivity of 90\% and specificity of 65\%.

Raman technology also has the potential to help clinicians decide which lesions are suitable for biopsy, as the adjunct use of \textit{in vivo} Raman spectroscopy with bronchoscopy methods has been shown to detect preneoplastic lesions with sensitivity and specificity above 90\% \textsuperscript{120}. These studies show that Raman spectroscopy is a highly sensitive method for the \textit{in vivo} diagnosis of lung cancer, and now with the ability to detect tumours in the peripheral lung with novel miniature Raman probe technology \textsuperscript{121}, and the capacity to produce clear spectra from weak Raman signals, further multicentre clinical trials are warranted.
In addition to *in vivo* and *ex vivo* studies, others have explored the feasibility of applying Raman spectroscopy combined with data mining methods to body fluid samples for lung cancer diagnosis. These are promising, cost effective methods and would allow simple and reliable screening to detect early stage malignancy. Analysing the serum of patients with NSCLC, Wang et al.\textsuperscript{122} were able to obtain distinct Raman spectral profiles for stage I, stage II, and advanced stage NSCLC. Comparative analysis of the mean spectral profiles for each stage of NSCLC revealed multiple peaks involved in the carcinogenic progression of NSCLC. Cancer progression corresponded with a decrease in all of the analysed peak intensities. The authors noted that protein and phospholipid content was significantly reduced in the serum samples of NSCLC patients, evident by a decreased peak intensity at 1658 cm\(^{-1}\). With PCA and discriminant analysis, spectra taken from the sera of healthy people and patients with stage I-IV NSCLC were differentiated with an overall accuracy of 92%. Other studies have shown that Raman spectroscopy also has the potential to characterise lung cancer and detect malignancy with high diagnostic sensitivities and specificities with non-invasive samples such as saliva and urine\textsuperscript{115,123,124}. Applying Raman spectroscopy to these minimally invasive specimen types may be an effective tool for NSCLC screening and staging without the need for bronchoscopy. The translation of these technologies into a clinical setting now requires successful randomised clinical trials before they are accepted.

As discussed in this review, applying Raman spectroscopy to tissue sections has shown the ability to discriminate between non-malignant and malignant lung tissue, and between NSCLC subtypes. Raman technology has also been successfully applied to single cells, effectively identifying malignancy. However, based on a comprehensive review of the literature it appears no studies have investigated the use of Raman spectroscopy on bronchoscopy cytology samples, which would align with the current sampling techniques.
for lung cancer. In light of the research to date, applying this technique to cytology samples may present as a minimally invasive method for identifying and discriminating NSCLC subtypes and treatment sensitive tumours. This method may be used to provide an accurate diagnosis and retain sufficient sample for ancillary molecular tests, maximising the use of limited samples.

2.2.3 Summary and Future Perspectives

Small tissue and cytology samples are often the only specimens available for the diagnosis and molecular analysis of NSCLC. Current diagnostic methods for NSCLC consume much of the available sample, and with an increasing demand for the sample for ancillary molecular tests, an alternative accurate diagnostic method is desirable. As vibrational spectroscopic techniques are label free, highly sensitive and allow the rapid detection of intracellular biochemical information, they address the limitations of current diagnostic methods and are a potential option for the differential diagnosis of NSCLC. As vibrational spectroscopy is non-destructive, the unstained slides can be used for subsequent analysis.

Numerous studies have demonstrated the ability of vibrational spectroscopic techniques to accurately detect and discriminate lung cancer subtypes using tissue and cell samples. As the process of obtaining cell specimens is minimally invasive, future research may investigate the viability of using vibrational spectroscopic methods on bronchoscopy cytology samples. A previous study has already shown the ability of FTIR spectroscopy to characterise lung cancer cells using cytological samples, and although Raman spectroscopy has been investigated on lung carcinoma samples \textit{in vitro}, \textit{ex vivo}, and \textit{in vivo}, to our knowledge it has never been used on bronchoscopy attained cytology samples\textsuperscript{95,104,107,109,120}.

As probe based Raman spectroscopy is developing and is a promising technique for the \textit{in vivo} diagnosis of cancer, development in the application of Raman spectroscopy for
the in vivo diagnosis of NSCLC may present with many promising advantages for clinical setting 119,125,126. It can be used at a lower cost than other imaging techniques including magnetic resonance imaging (MRI) and ultrasound, in addition to giving real time information on biochemical composition in high resolution with no stains or labels 99.

2.3 Conclusion
As the identification of NSCLC subtype is pertinent for the development of treatment regimens, an enhanced label-free, rapid, and even automated diagnostic technique would be largely beneficial. Integrating vibrational spectroscopic methods into the diagnostic algorithm for lung cancer offers the ability to provide an accurate diagnosis and discover new molecular markers for the discrimination of NSCLC subtypes. Vibrational spectroscopic methods may be useful for the non-destructive analysis of small unstained bronchoscopy attained samples, without compromising subsequent ancillary molecular tests. These techniques have the potential to reduce the need for surgical biopsies, and may be economically advantageous by saving medical resources.

References


49. Derman BA, Mileham KF, Bonomi PD, Batus M, Fidler MJ. Treatment of advanced squamous cell carcinoma of the lung: a review. Transl Lung Cancer Res. 2015;4(5):524-
532. doi:10.3978/j.issn.2218-6751.2015.06.07


62


Chapter 3: Methodologies

3.1 Study subjects, sample collection and preparation

This study had approval from the Research Ethics Committee, Dublin Institute of Technology (now Technological University Dublin) as well as SJH/AMNCH Research Ethics Committee. BALs and BWs were collected from patients during bronchoscopy for suspected lung cancer. Informed consent to use residual cell sample was obtained from each patient at a previous clinical appointment. Lung cancer was subsequently confirmed through final clinical diagnosis and final histology was recorded.

3.1.1 Sample processing for BAL and BW samples

Cellular material was isolated by breaking down the mucus with a solution consisting of 10µM Dithiothreitol (DTT), a mucolytic agent, in Cytolyt. Samples then underwent centrifugation at 3000 rpm for 10 min and supernatant was poured off leaving a cell pellet. The pellet was then resuspended in PreservCyt and transferred to a ThinPrep® vial.

3.1.2 Slide preparation

The sample vial was placed into a ThinPrep® 2000 Processor which first thoroughly mixed the sample to disperse the cells by rotating the filter. A gentle vacuum is then created in the ThinPrep® filter to collect a monolayer of cells on the filter membrane. The cells collected on the filter membrane are then gently pressed against the ThinPrep® glass slide and natural attraction and positive air pressure helps the cells to adhere to the slide in a 20 mm circle. The slide was then deposited into a fixative solution of 95% ethanol before being removed from the alcohol and allowed to air dry. ThinPrep® glass slides were used as the substrate for Raman spectroscopic analysis throughout this study.
3.2 Raman spectral acquisition and data pre-processing and analysis

3.2.1 Raman Measurement

Raman spectra were acquired using a HORIBA Jobin Yvon XploRA™ system (Villeneuve d’Ascq, France) which consists of an Olympus microscope BX41 supplied with a x100 objective (MPlanN, Olympus, NA = 0.9). A 532 nm diode laser source was utilised during the study and was set at 100% power, which gave 8 mW at the objective. The confocal hole was set at 100 µm and coupled to a slit of aperture 100 µm. The system was calibrated to the 520 cm⁻¹ spectral line of Silicon and the 1200 lines per mm grating was used. Using Silicon, the laser line was calibrated to yield maximum signal intensity. Backscattered light was detected using an air-cooled CCD detector (Andor, 1024 x 256 pixels). The software implemented to manage the spectrometer was Labspec V6.0. The spectral range was 400-1800 cm⁻¹ and each spectrum corresponds to the average of two accumulations of 30 seconds. Spectra were also recorded from 300 vacant locations on a ThinPrep® glass slide with identical exposure time in order to obtain representative background glass spectra.

3.2.2 Recording position of targeted cells unstained slide

Each ThinPrep® glass slide has the word ‘ThinPrep®’ printed on the upper side of the glass. Before recording spectra from the slide, the centre of the dot in the ‘i’ in the printed word ‘ThinPrep’ was set as the 0-0 XY co-ordinate. Having a reference point on the slide allowed the XY co-ordinates of target cells to be recorded and stored in a separate ‘co-ordinate file’ that was correlated to the spectra. At a later date, targeted cells could be identified by setting the 0-0 point as previously explained, and opening the co-ordinate file using the Labspec 6 software.
To reduce interference from random and irrelevant variations within the spectral data, pre-processing of the Raman spectra is needed before chemometric analysis.

3.2.3 Smoothing
The Savitzky-Golay smoothing technique reduces spectral noise and can enhance the biochemical signal. The Savitzky-Golay method uses a least squares technique to fit a window of wavenumbers to a polynomial function, retaining features such as band width and height\(^1\). Fifth order filtering with a 15-point window was used throughout.

3.2.4 Baseline removal
The rubberband baseline correction method finds a polygonal line with edges at the spectral minima or ‘troughs’. Another computational method of background subtraction is the subtraction of a polynomial of certain order\(^2\). The rubberband method was applied throughout this thesis, aside from Chapter 7. A fifth order polynomial baseline correction was applied to the cell line data in Chapter 7 as the rubberband method did not adequately correct the baseline.

3.2.5 Normalisation
The spectra were vector normalised to reduce intensity variations caused by excitation power, this allows for reliable analysis of minute changes in the spectra. Vector normalisation is a scaling method whereby each variable is divided by a scaling factor which is different for each variable, returning a vector of unit length 1. A vector normalised spectrum is normalised to the square root of the sum of the squared value of all variables for the spectrum\(^2,3\).

3.2.6 Glass removal by non-negative least squares
As described by Ibrahim et al.\(^4\), non-negative least squares analysis (NNLS) is a technique which uses a function to calculate the weighted contribution of a set of input spectra in a
sample spectrum. Similar to the classical least squares fitting technique, it assumes that a sample spectrum is the sum of the base components contributing to the spectrum. As the input spectra take on only non-negative values, the NNLS function incorporates non-negative constraints on the input spectra, meaning only positive values can contribute to the approximation of the sample spectrum. Thus, spectra from cellular components, such as nucleic acids, proteins, lipids, and the independently recorded glass spectra were used as input spectra to fit the sample spectra and remove the glass contribution.

3.2.7 Outlier removal
Grubbs’ test was used to identify and remove outliers from the data. The outlier is detected by calculating the difference between the test value and the mean, and then dividing the value of the difference by the standard deviation of the dataset. If the ratio is greater than +/-2, the object is considered an outlier and is removed. The Grubbs test detects one outlier in the dataset at a time. If an outlier is detected, it is removed and the test is reiterated.

3.2.8 Principal component analysis
The identification of cancer biomarkers using Raman spectroscopy requires the translation of the Raman signal into a diagnosis with data mining techniques. Alterations between spectra are detected and evaluated using machine learning software and correlated to histopathological results attained from gold standard diagnostic methods. This data can then be used to design a classification system to discriminate between normal and cancerous tissues. Additionally, the wavenumbers of important spectral bands may also be correlated to specific biomolecules and may identify pathways involved with disease progression, proliferation and more.

There are numerous statistical techniques for the analysis of biological Raman spectra. Principal component analysis (PCA) is an unsupervised data reduction technique that is
extensively used in relation to cancer research. It is used to identify variances within the data which may classify individual spectra into groups\textsuperscript{6-8}.

PCA reduces the dimensionality of the data while retaining most of the variance within the dataset. This is achieved through several mathematical steps, initially the mean is subtracted from the dataset to produce a new data set, the mean centred matrix. Then the covariance matrix of the mean centred matrix is computed by calculating the linear relationships between individual spectra. Principal components (PC) and scores, or eigenvectors and eigenvalues, are then calculated for the covariance matrix, with each PC explaining independent sources of variance and having a corresponding score. The PC with the highest score explains most of the variance in the dataset, and is called the first PC. The PCs are organised in order of the proportion of variance they explain, from the highest to lowest. Further analysis of the data is then done using the calculated scores of each variable in relation to each PC\textsuperscript{6-8}.

3.2.9 Linear discriminant analysis

Linear discriminant analysis (LDA) is another technique that is routinely amalgamated with PCA to develop a predictive model based on the Raman spectra. When implemented with the PCA model, the LDA classification system increases classification accuracy by using the PC scores. LDA is a member of a class of methods which use linear decision boundaries for classification. The LDA method identifies linear functions that maximise the separation between different classes, and minimises the within-class covariance matrix, allowing classification of data into different classes\textsuperscript{6,9}.

3.2.10 Partial least squares discriminant analysis

Partial least squares discriminant analysis (PLS-DA) mean centres the Raman spectral data to identify the diagnostically significant variations between the groups. PLS-DA reduces dimensionality of the data by considering the correlated relationships between
the spectral dataset and the class membership. The PLS components, or latent variables (LVs), are rotated to maximise the covariance between spectral variation and class membership. This supervised approach finds the LVs in the multivariate space that maximise the separation of groups and enables discrimination between known classifications of samples (negative, cancer etc.) based on the diagnostically relevant spectral variations. The required number of latent variables was utilised in each model to set the cross-validation classification error average at 5%. The analysis was performed using the PLS toolbox (Eigenvector Research, Wenatchee, WA) in the Matlab (Mathworks Inc., Natick, MA) scripting environment, and cross validated using leave one out cross validation.

3.2.11 Cross validation
Cross validation is implemented to evaluate the performance of a classification model in accurately determining pathological status. The method used in this study was venetian blinds cross validation. In venetian blinds validation, each validation set is developed using every n\textsuperscript{th} spectrum in the data, starting at spectrum 1 through n. For example, in a 10-fold venetian blinds validation, each validation set or fold i is built taking spectra from the dataset of a 10-multiple position until the end of the dataset (samples \(i, i + 10, i + 2(10), i + 3(10), \ldots\)). Once the folds are developed, 10 models are trained with 10–1 folds, and tested with the remaining fold, rotating the testing fold until all of them have been used. The overall performance of the model is determined as the average performance of the 10 models.

3.2.12 Calculation of classification models performance scores
The ability of a classifier to detect pathology can be measured by evaluating the sensitivity, specificity, and diagnostic accuracy of the model. The sensitivity of a technique is determined as the proportion of cases with a disease that are accurately
detected. The specificity of a model is the ability of the model to accurately detect those without disease, and diagnostic accuracy describes the proportion of correctly classified cases among all the classified cases.

![Actual condition](chart)

<table>
<thead>
<tr>
<th>Test result</th>
<th>Disease</th>
<th>Healthy</th>
</tr>
</thead>
<tbody>
<tr>
<td>Positive</td>
<td>True positive</td>
<td>False positive</td>
</tr>
<tr>
<td>Negative</td>
<td>False negative</td>
<td>True negative</td>
</tr>
</tbody>
</table>

Sensitivity = \( \frac{\text{True positive}}{\text{True positive} + \text{False negative}} \times 100 \)

Specificity = \( \frac{\text{True negative}}{\text{True negative} + \text{False positive}} \times 100 \)

Diagnostic accuracy = \( \frac{\text{true positive} + \text{true negative}}{(\text{true positive} + \text{true negative} + \text{false positive} + \text{false negative})} \times 100 \)

3.3 Pap staining

This was conducted after collecting Raman spectra from the target cells. Slides were rehydrated in normal saline (0.9% NaCl) for 60 seconds before being washed in water. Harris haematoxylin was added for 5 minutes and then washed off with water. The Haematoxylin was then differentiated with acid alcohol for 2 seconds and checked. Complete removal of the haematoxylin from the cell cytoplasm was checked promptly with a light microscope, ensuring the slide did not dry. The slide was then submerged in water to blue the bound stain. Following quick washes in 70% and 95% ethanol, O.G 6
solution was applied for 3 minutes. The slides were then washed rapidly in 95% alcohol and stained with E.A 50 solution for 4 minutes. After washing rapidly in 95% ethanol, the slides were transferred through a series of two washes of absolute alcohol and then into histoclear before coverslipping. The stained slides were then reviewed microscopically. The recorded cells were mapped by their XY coordinates and accurately categorised into databases according to their cell type.

References


Chapter 4: Raman spectroscopic analysis of NSCLC using bronchoscopy cytology samples

4.1 Introduction

Lung cancer persists as the most common cancer worldwide\(^1\). According to the GLOBOCAN 2018 report \(^2\), lung cancer represents the highest cancer incidence and mortality rates in both developed and less developed countries, with approximately 2.1 million new cases and 1.76 million deaths in 2018. Studies have indicated a 5 year survival rate of 8\% to 15\% in developed countries, although a survival rate as low as 1\% has been reported for late stage diagnosis \(^3\)–\(^5\). Incidence trends can be directly correlated with smoking prevalence, as cigarette smoking is responsible for 85\% of lung cancer cases \(^6\)–\(^8\).

There are two prominent types of lung cancer, small cell lung cancer (SCLC) accounting for an estimated 13\%–14\% of cases, and non-small cell lung cancer (NSCLC), which is the most common form of lung cancer accounting for approximately 85\% of cases \(^9\)–\(^12\). NSCLC constitutes two prominent histological phenotypes including squamous cell carcinoma (SCC), adenocarcinoma (AC) and large cell carcinoma. AC accounts for the largest portion of all lung cancer cases at 44\%, followed by SCC at 26\% \(^9\). Historically, both SCC and AC were treated clinically and therapeutically the same, however with the recent availability of differential treatment options for AC and SCC patients, accurate subtyping of NSCLC is becoming increasingly important in the therapeutic algorithm \(^13\),\(^14\),\(^15\)–\(^22\),\(^23\)–\(^26\).

In approximately 80\% of cases patients are diagnosed through the analysis of cytological material or a small tumour biopsy. For decades, bronchoscopic techniques such as bronchial brushing, bronchial washing (BW) and bronchoalveolar lavage (BAL) have been used to obtain cytological samples of suspicious pulmonary lesions. Studies
investigating the efficacy of these techniques for diagnosing malignancy have indicated sensitivities for malignancy ranging from approximately 30% - 71% \(^{27-30}\). Immunohistochemistry (IHC) can be applied to improve the differential diagnosis of NSCLC to approximately 85%\(^{31}\), and additional molecular testing techniques may be utilised to aid the detection of targetable mutations. These include fluorescent in situ hybridisation (FISH), and sequencing techniques \(^{32-33}\). However, these methods cannot always be applied and for many patients the small diagnostic sample is often fully utilised to classify the NSCLC subtype, and ancillary molecular tests cannot be performed. For these patients a repeat biopsy is needed to complete the clinical work-up of the patient. As more molecular targets and subsequent personalised therapies are discovered, retaining adequate sample during the diagnostic work up of NSCLC for molecular testing is becoming increasingly important. Implementing a label free diagnostic technique in the diagnostic algorithm of NSCLC would allow subsequent use of the limited sample for molecular profiling. In this context Raman spectroscopy is a promising emerging tool.

Raman spectroscopy is a label free vibrational spectroscopic technique based on the inelastic scattering of light \(^{34,35}\). It has shown favourable results for detecting malignancy in multiple cancer types including brain, breast, skin, gastrointestinal, oesophageal and lung cancer \(^{36-39}\). The ability of Raman spectroscopy to detect lung malignancy has been investigated using biopsy samples, and in vivo, with reported diagnostic sensitivities up to 91% \(^{40-45}\). Despite the efficacy of Raman spectroscopy for the detection of lung cancer, to our knowledge it has never been applied to bronchoscopy attained bronchial washes, or BAL cytology samples.

Applying Raman spectroscopy for cancer prediction requires data mining techniques to translate the Raman signal into a diagnosis. A supervised classifier is developed using a database of spectra from patient samples with known histopathological results attained
from gold standard diagnostic methods. The spectra are evaluated to identify relationships between the spectra and their known histopathological result. This data can then be used to design a classification model to predict the diagnosis of new patient samples. Additionally, the wavenumbers of important spectral bands may also be correlated to specific biomolecules and may give insight into the biochemical composition of the cells.

The aim of this study was to investigate the feasibility of applying Raman spectroscopy to BAL or BW samples to provide a marker free diagnosis of NSCLC subtypes, and to identify the biochemical changes responsible for the classification. For this study, samples were grouped according to their final gold standard histopathological diagnosis as characterised by a pathologist.

4.2 Methods

4.2.1 Study subjects and sample collection
As described in chapter 3 section 3.1.

This study utilised BALs and BWs collected from 24 patients (9 female, 15 males). Lung cancer was confirmed through final clinical diagnosis and final histology was recorded: 12 squamous cell carcinoma, 5 adenocarcinoma, and 7 negative (no cancer diagnosed).

4.2.2 Sample processing and slide preparation
As described in chapter 3 section 3.1.1-2.

4.2.3 Raman Measurement
As described in chapter 3 section 3.2.1-2.

Spectra were taken from the nuclei of morphologically normal looking bronchial epithelial cells, squamous cells, and lymphocytes. For each bronchial and squamous cell, spectra were also recorded from the cell cytoplasm. On SCC cases, spectra were also taken from the nuclei of morphologically abnormal squamous cells, although spectra
could not be acquired from the cytoplasm due to photodegradation using a 532 nm laser\textsuperscript{46}. On AC cases morphologically abnormal cells (Figure 4.3) were not targeted as they could not be identified on the unstained slides. For each sample, we aimed to take spectra from fifteen of each cell type. For each cell, one measurement was recorded from the nucleus, with the system set us as described in chapter 3 section 3.2.1-2. As the cell cytoplasm is thin, and therefore there was more interference from glass, it was difficult to acquire quality spectra. Thus for each sample we aimed to take spectra from the cytoplasm of 5 to 15 cells of each cell type.

4.2.4 Data pre-processing and analysis
All spectral data analysis was conducted using MATLAB software. Data were processed as described in chapter 3 section 3.2.3-.7. The data was analysed using Principal Component Analysis (PCA), as described in chapter 3 section 3.2.8. Partial least squares discriminant analysis (PLS-DA) with venetian blind cross-validation was also employed to the mean centered Raman spectral data to identify the diagnostically significant variations between the groups, as described in chapter 3 section 3.2.10-12.

4.2.5 Pap staining
Following Raman spectroscopy, slides were stained using the Papanicolaou staining method as described in chapter 3 section 3.3. Using the XY co-ordinates recorded as described in chapter 3 section 3.2.2, the spectra were correlated with their cell type and categorised into separate databases according to their cell type.
4.3 Results

4.3.1 Mean spectral analysis

Figure 4.1 depicts three cell types targeted in this study and the mean nuclear spectra for each cell type taken from negative cases. Applying the Pap stain to slides analysed with Raman spectroscopy enabled the targeted cell types to be identified as bronchial epithelial cells (Figure 4.1 (a)), squamous cells (Figure 4.1(b)), and lymphocytes (Figure 4.1(c)). The morphologically abnormal or malignant squamous cells (Figure 4.2) were also targeted. These cells were observed in SCC cases where malignancy was evident on cytology. The standard deviations in Figure 4.1 d-f show that there was minimal variation between the spectra from individual cells. The significant spectral differences between the nuclei of lymphocytes, bronchial and squamous cells from negative cases are shown by the difference spectra in Figure 4.4. The features in the nuclear spectra taken from bronchial cells, squamous cells, and lymphocytes of AC and SCC cases that differ from the cells of negative cases are demonstrated by the difference spectra in Figures 4.6, 4.8 and 4.10. Significance was determined using Student’s t-test and values of p<0.01 were considered to be statistically significant. Table 4.1. shows the tentative peak assignments for the prominent significant Raman bands identified by spectral variation.
Figure 4.1. Depiction of stained and unstained bronchial cells (a), squamous cells (b), and lymphocytes (c). The Pap stain was applied to provide contrast between cell types and components. The corresponding mean Raman spectra (± standard deviation) taken from the nuclei of each cell type are shown. (d) The mean Raman spectrum of bronchial epithelial cell nuclei taken from negative cases (blue, $n=70$) ($n$=the number of spectra used), (e) the mean Raman spectrum of squamous cell nuclei taken from negative cases (black, $n=94$), and (f) the mean Raman spectrum of lymphocyte nuclei taken from negative cases (red, $n=80$). Images were taken at x1000 magnification.
Figure 4.2. Representative keratinizing tumour cells with cytoplasmic projections taken from patients with squamous cell carcinoma. The cells are stained with the Pap stain which was applied subsequent to Raman analysis. Images were taken at x1000 magnification.
Figure 4.3. Representative tumourous glandular cells taken from patients with adenocarcinoma carcinoma. The cells are stained with the Pap stain which was applied subsequent to Raman analysis. Image A was taken at x400 magnification, B-D were taken at x1000 magnification.
Table 4.1. Tentative peak assignments for the prominent significant Raman bands identified by spectral variation analysis.47,48.

<table>
<thead>
<tr>
<th>Wavenumber (cm$^{-1}$)</th>
<th>Raman peak assignments</th>
</tr>
</thead>
<tbody>
<tr>
<td>~420/30/40</td>
<td>cholesterol</td>
</tr>
<tr>
<td>509</td>
<td>protein</td>
</tr>
<tr>
<td>580</td>
<td>phosphatidylinositol</td>
</tr>
<tr>
<td>486-526</td>
<td>Polysaccharides, proteins</td>
</tr>
<tr>
<td>590</td>
<td>Protein/glycerol</td>
</tr>
<tr>
<td>640</td>
<td>Tyrosine/methionine</td>
</tr>
<tr>
<td>657/~700/10</td>
<td>Methionine</td>
</tr>
<tr>
<td>735</td>
<td>Thiocyanate</td>
</tr>
<tr>
<td>~780</td>
<td>Nucleic acids</td>
</tr>
<tr>
<td>830-920</td>
<td>Polysaccharides/proline/valine/collagen</td>
</tr>
<tr>
<td>~930</td>
<td>Proline/valine/saccharides</td>
</tr>
<tr>
<td>1006</td>
<td>phenylalanine</td>
</tr>
<tr>
<td>1024-32</td>
<td>glycogen/peaks related to collagen</td>
</tr>
<tr>
<td>1106-1125</td>
<td>saccharides/lipids/proteins</td>
</tr>
<tr>
<td>1175-1200</td>
<td>Nucleic acids/phosphates</td>
</tr>
<tr>
<td>1210-1280</td>
<td>Amide III/nucleic acids</td>
</tr>
<tr>
<td>~1340-50</td>
<td>Nucleic acids/protein/lipids</td>
</tr>
<tr>
<td>~1380</td>
<td>lipid</td>
</tr>
<tr>
<td>1420-50</td>
<td>Proteins/lipids</td>
</tr>
<tr>
<td>~1480</td>
<td>Amide II</td>
</tr>
<tr>
<td>~1500-1600</td>
<td>Amide II/ nucleic acids</td>
</tr>
<tr>
<td>~1580</td>
<td>COO-</td>
</tr>
<tr>
<td>1600-1666</td>
<td>Amide I</td>
</tr>
<tr>
<td>~1680</td>
<td>Amide I disordered structure/β sheets/lipids</td>
</tr>
<tr>
<td>~1700-1740</td>
<td>lipids</td>
</tr>
</tbody>
</table>
4.3.2 Spectral comparison of the lymphocyte, bronchial, and squamous nuclei from negative cases

Based on many significant differences the Raman spectra from lymphocytes, squamous cells and bronchial cells could be differentiated from each other. Spectral differences between the squamous cell nuclei of negative cases (Figure 4.4) and the bronchial cell nuclei of negative cases corresponded to peaks relative to proteins and lipids (421-421-500, 530-690, 762-84, 1024-32, 1129-200, 1205-06, 1345-61, 1379-1459, 1470-1546, 1611, 1620-60, 1680-1727 cm\(^{-1}\)), saccharides (421-500, 920-68, 1024-32, 1129-200 cm\(^{-1}\)) and nucleic acids (762-84, 1129-200, 1315-20, 1345-61, 1450-59, 1564-82 cm\(^{-1}\)). The squamous cells showed an increased peak intensity in peaks assigned to glycogen / cholesterol / protein / phospholipids (421-690 cm\(^{-1}\)), glycogen / cholesterol / saccharides / protein/nucleic acids (762-1011 cm\(^{-1}\)), glycogen and peaks related to collagen (1024-32 cm\(^{-1}\)), lipids / protein / amide I (1436-44, 1620-60 cm\(^{-1}\)) in comparison to the bronchial cells. The bronchial cells showed an increased peak intensity in peaks assigned to lipid / protein / nucleic acids / phosphates (1129-1200, 1205, 1230, 1315-20, 1345-61, 1379-1430 cm\(^{-1}\)), and lipid / protein / COO\(^{-}\) / nucleic acids / cytosine (1450-1546, 1564-82, 1611 cm\(^{-1}\)) in comparison to the squamous cells.

The squamous cells from negative cases displayed an elevated peak intensity at 460-95, 519-615, 633-90, 710-963, 991-1017, 1040-62, 1070-96, 1221-40, 1245-60, and 1607-1665 cm\(^{-1}\) compared to the lymphocytes from negative cases. This corresponds to more glycogen / cholesterol / protein / phospholipids / nucleic acids / phenylalanine (460-1017 cm\(^{-1}\)), saccharides / protein / lipid (1040-62, 1070-96 cm\(^{-1}\)),amide III / nucleic acid (1221-40, 1245-60 cm\(^{-1}\)) lipid / protein (1435-45 cm\(^{-1}\)), and amide I (1607-1665 cm\(^{-1}\)) content in the squamous cells (Figure 4.4). The lymphocytes from negative cases were associated with elevated peak intensities assigned to lipids / saccharides / nucleic acids (1107-16, 1130-51, 1172-96, 1338-63, 1373-80 cm\(^{-1}\)), amide II / lipid / protein (1450-1547 cm\(^{-1}\)),
COO⁻ (1578-96 cm⁻¹), cytosine (1605 cm⁻¹), and amide I disordered structure/β sheets/lipids (1680-1719 cm⁻¹) in comparison to the squamous cells from negative cases. The difference spectrum between the lymphocytes from negative cases and bronchial cells from negative cases shows significant differences across the entire spectrum (Figure 4.4). The lymphocytes from negative cases displayed an increased peak intensity in peaks assigned to proteins and lipids (421-683), amide II / protein / lipid (1450-1562 cm⁻¹), COO⁻ (1581-97 cm⁻¹), amide I (1607-57 cm⁻¹) and amide I disordered structure / β sheets / lipids (1680-1719 cm⁻¹) in comparison to the bronchial cells from negative cases. The bronchial cells showed increased peak intensities in peaks assigned to glycogen / cholesterol / saccharides / protein / nucleic acids (752-1017 cm⁻¹), glycogen / protein / lipids / saccharides (1022-1099 cm⁻¹), nucleic acids / amide III / proteins (1146-1325 cm⁻¹), lipid / protein (1433-47 cm⁻¹), nucleic acids (1576 cm⁻¹), and the amide I peak from 1662-70 cm⁻¹ in comparison to the lymphocytes from negative cases.
Figure 4.4. The difference spectrum between the mean Raman spectrum of squamous cell nuclei taken from negative cases and the mean Raman spectrum of bronchial epithelial cell nuclei taken from negative cases is depicted in blue. The difference spectrum between the mean Raman spectrum of squamous nuclei taken from negative cases and the mean Raman spectrum of lymphocyte nuclei taken from negative cases is shown in red. The difference spectrum between the mean Raman spectrum of lymphocyte nuclei taken from negative cases and the mean Raman spectrum of bronchial nuclei taken from negative cases is depicted in green. Shading indicates regions of the spectrum that were significantly different (P < 0.01).

4.3.3 Mean spectral analysis of cell nuclei from negative cases, SCC cases, and AC cases

4.3.3.1 Bronchial cell analysis

The mean Raman spectra taken from the bronchial cell nuclei of negative, SCC, and AC cases are shown in Figure 4.5. Spectral differences between the bronchial cell nuclei of
negative cases and the bronchial nuclei from AC and SCC cases corresponded to peaks relative to lipids (433, 569-626, 1282-1297, 1288-1310, 1468, 1470-1477 cm⁻¹) proteins (502-513, 525, 569-626, 640, 650, 819-826, 834-899, 920, 915-928, 972-1000, 1008, 1214-1258, 1429, 1468, 1419-1448, 1597-1663 cm⁻¹), and nucleic acids (676-688, 764-788, 1231-1258, 1323-1326, 1416, 1429, 1576 cm⁻¹), as shown in Figure 4.6. The difference spectra of the bronchial cells in AC and SCC cases had similar profiles, with peaks at 509, 599-626, 788, 848, 920, 988, 1023, 1223, 1240-1248, 1288-1297, 1429, 1467, 1597-1661 cm⁻¹ differentiating them from the bronchial cells of negative cases. The bronchial cells from AC and SCC patients showed a decreased peak intensity at 509, 599-626, 1470, 1597-1661 cm⁻¹, and an increased peak intensity at 788, 848, 920, 988, 1223, 1230, 1240-1248, 1288-1297, and 1429 cm⁻¹, relative to the bronchial cells of negative cases. This corresponds to a lower level of cholesterol, phosphatidylinositol, lipids, and amide I, and an increase in collagen related peaks, nucleic acids, polysaccharides, proline / glycogen, amide III and lipids in the bronchial cells of NSCLC patients.

Bronchial cells from SCC patients displayed an elevated peak intensity at 509, 1100, 1340, 1379, 1401, 1410, ~1445 cm⁻¹, 1480 and 1655 cm⁻¹ compared to the bronchial cells from AC patients, corresponding to increased lipid, protein, nucleic acid and amide I content (Figure 4.6). An increase in methionine (~700 cm⁻¹), tryptophan / nucleic acids/ phosphatidylinositol (760-776 cm⁻¹), saccharides (~840, 860-900 cm⁻¹), phenylalanine (1006 cm⁻¹), lipids / saccharides / proteins / nucleic acids (1130-1206 cm⁻¹) amide III / nucleic acids (1215, 1280 cm⁻¹), amide II (1555 cm⁻¹), amide I peaks at 1611-1628 cm⁻¹, and amide I of disordered structure / β sheets (1690 cm⁻¹) was associated with spectra from AC patients.
Figure 4.5. Mean Raman spectra of negative bronchial cell nuclei (n= 70), bronchial cell nuclei (n= 150) from SCC cases, and bronchial cell nuclei (n= 63) from AC cases. Shading denotes the standard deviation.
Figure 4.6. Difference spectrum between the mean Raman spectrum of bronchial epithelial cell nuclei taken from negative cases and the mean Raman spectrum of bronchial epithelial cell nuclei (blue, $n=150$) taken from SCC cases, and the mean Raman spectrum of bronchial epithelial cell nuclei (red, $n=63$) taken from AC cases the difference spectrum between the bronchial epithelial cell nuclei taken from SCC and AC cases (green). Shading indicates regions of the spectrum that were significantly different ($P < 0.01$).

4.3.3.2 Squamous cell analysis

The mean Raman spectra taken from the squamous cell nuclei of negative, SCC, and AC cases are shown in Figure 4.7. The differences between the spectra taken from the squamous cell nuclei of negative cases and the spectra taken from the squamous cell nuclei of AC and SCC cases corresponded to peaks relative to proteins and lipids (408, 504-518, 584-88, 596-620, ~645, ~652, ~700, ~840, ~880, 989-1006, 1150-1200, 1205,
1219-43, ~1330, 1435-46, 1456-1477, 1520-1570, and 1611-1703 cm\(^{-1}\)) carbohydrates (~495 467-73, 486-98, 551, and 836-840 cm\(^{-1}\), and nucleic acids (~692, ~780, 1219-43, ~1320, 1330, 1346-58, 1513-16, and 1520-72 cm\(^{-1}\)), as shown in Figure 4.8. The difference spectra of the squamous cells taken from AC and SCC cases display peaks at 495, 504, 597-607, 626, 1200-206, 1219-29, 1318-1323, 1456-62, 1526, 1534-38, 1567-72, and >1600 cm\(^{-1}\) differentiating them from the squamous cells of negative cases (Figure 4.8). The squamous cells from the AC and SCC cases showed decreased peak intensities at 495, 504, 597-608, 626, 1318-1323, 1456-1462, and >1600 cm\(^{-1}\), and increased peak intensities at 1200-1206, 1219-1229, 1526, 1534-1538, and 1567-1572 cm\(^{-1}\), relative to the squamous cells from negative cases. This corresponds to a lower level of proteins (504, 597-607, 1318-23, 1456, >1600 cm\(^{-1}\)), glycerol/methionine (626 cm\(^{-1}\)), and guanine (1318-23 cm\(^{-1}\)), and an increase in protein (1200-206 cm\(^{-1}\)), glycogen (1022/23 cm\(^{-1}\)), amide III / nucleic acids (1220-29 cm\(^{-1}\)), amide II (1526 cm\(^{-1}\)) and tryptophan / nucleic acids (1534-38, 1567-72 cm\(^{-1}\)) in the squamous cells of AC and SCC cases relative to the squamous cells of negative cases. Squamous cells from SCC cases also have an alteration in nucleotide conformation (600 cm\(^{-1}\)), and display elevated nucleic acids (~780, 1513-1516 cm\(^{-1}\)), protein / collagen related peaks (878-84, 1435-46 cm\(^{-1}\)), and a lower level of cholesterol / lipids (408, 429-36, 440-45 cm\(^{-1}\)), polysaccharides (467-73, 486-95 cm\(^{-1}\)), methionine (652-56 cm\(^{-1}\)), guanine (1346-58 cm\(^{-1}\)), and phenylalanine (989-1006 cm\(^{-1}\)) compared to the squamous cells of negative cases. Squamous cells from SCC cases displayed an elevated peak intensity at 586, 627-47, 1370, and 1387-91 cm\(^{-1}\), compared to the squamous cells from AC cases, corresponding to increased lipid and methionine content (Figure 4.8). Elevated cholesterol (408, ~430cm\(^{-1}\)), phospholipids (717 cm\(^{-1}\)), amide II (1548-59 cm\(^{-1}\)), phenylalanine (1002 cm\(^{-1}\)), tyrosine (632 cm\(^{-1}\)), nucleic acids (762-774, 1175-1199 cm\(^{-1}\)), polysaccharides (~840,
860-900 cm\(^{-1}\), amide III (1280 cm\(^{-1}\)), and amide I (1625-1628 cm\(^{-1}\)) was associated with the spectra from AC patients in comparison to the SCC patients.

The spectra taken from the nuclei of squamous cells with cytoplasmic projections (Figure 4.8) indicated many biochemical changes in the malignant cells. A reduced level of carbohydrates, cholesterol, phosphatidylinositol, and methionine is evident in the malignant cells by the lower peak intensities at ~408, ~430, 457-98, 567-70, 575-83, 589-92, 595-641, 655-660, 681-707, and 711-24 cm\(^{-1}\) (Figure 4.8). The malignant cells also show an increase in tryptophan (760 cm\(^{-1}\)), DNA (780 cm\(^{-1}\)), proline / tyrosine (847-56, 864-86 cm\(^{-1}\)), and amide III / lipids / nucleic acids / saccharides (1220-54, 1278-85, 1350-1415, 1420-25, 1435-46, 1473-78 cm\(^{-1}\)). The nuclei of both morphologically normal and abnormal squamous cells in SCC cases are similar. In comparison to the squamous cells of negative cases, both the squamous cells with cytoplasmic projections and morphologically normal squamous cells from SCC cases display decreased peak intensities at 408, 430-36, 441-45, 457-60, 467-73, 486-95, 505-18, 597-620, 626-41, 655-56, 692, 700-07, 1318-21, 1611, and 1620-31 cm\(^{-1}\), and increased peak intensities at 780, 878-84, 1022, 1048, 1220-43, and 1435-46 cm\(^{-1}\). This corresponds to an increase in DNA (780 cm\(^{-1}\)), the protein / lipid band at ~1445 cm\(^{-1}\), glycogen (1023, 1048 cm\(^{-1}\)) and amide III (1220-54, 1278-85 cm\(^{-1}\)), and a decrease in carbohydrates, phosphatidylinositol, methionine and amide I.
Figure 4.7. Mean Raman spectra of negative squamous cell nuclei \((n=94)\), squamous cell nuclei \((n=189)\) from SCC cases, and squamous cell nuclei \((n=67)\) from AC cases. Shading denotes the standard deviation.
Figure 4.8. Difference spectrum between the mean Raman spectrum of squamous cell nuclei taken from negative cases and the mean Raman spectrum of squamous cell nuclei (blue, $n=189$) taken from SCC cases. Difference spectrum between the mean Raman spectrum of squamous cell nuclei taken from negative cases and the mean Raman spectrum of the nuclei (black, $n=60$) of cells with cytoplasmic projections taken from SCC cases. Difference spectrum between the mean Raman spectrum of squamous cell nuclei taken from negative cases and the mean Raman spectrum of squamous cell nuclei (red, $n=67$) taken from AC cases. Difference spectrum between the mean Raman spectrum of squamous cell nuclei taken from SCC cases and the mean Raman spectrum of squamous cell nuclei taken from AC cases (green). Shading indicates regions of the spectrum that were significantly different ($P < 0.01$).
4.3.3.3 Lymphocytes

The mean Raman spectra taken from the lymphocyte nuclei of negative, SCC, and AC cases are shown in Figure 4.9. Spectral differences between the lymphocytes of negative cases and the lymphocyte nuclei from AC and SCC cases corresponded to peaks relative to proteins (524, ~630, 650, ~704-714, 719-45, 847-55, 1005, 1205, 1219-54, ~1298, 1346-60, 1387-1426, 1440-53, 1458, 1491-1534, 1584-93, 1603-1674, 1611-55, 1667-72 cm\(^{-1}\)), lipids (415-54, 502-35, 574-636, ~1298, 1304-09, ~1368, 1387-1426, 1440-53, 1458, 1702-1707, ~1747--54 cm\(^{-1}\)), carbohydrates (460-79, 522-67, 847-55, cm\(^{-1}\)) and nucleic acids (673-76, 719-45, 789-821, 828-30, 1060, 1219-54, 1304, 1346-60, 1422-29, 1458, 1476-80 cm\(^{-1}\)), as shown in Figure 4.10. The difference spectra of lymphocytes from AC and SCC cases had decreased peaks at 419-425, 470-473, 487-489, 519-22, 590, 623-628, 1611-1655, ~1750 cm\(^{-1}\), and increased peaks at 1367-69, 1387-94 and 1422-26 cm\(^{-1}\) differentiating them from the lymphocytes of negative cases (Figure 4.10). This corresponds to a lower level of cholesterol (419-25 cm\(^{-1}\)), polysaccharides (470-73, 487-89), phosphatidylinositol (519-22), glycerol (590), protein (1611-55), and lipid (~1750 cm\(^{-1}\)) in the lymphocytes of AC and SCC cases relative to the lymphocytes of negative cases. This also indicates an increase in phospholipids (1367), lipids (1387-94), and deoxyribose (1422-1426) in the lymphocytes of SCC and AC patients.

Lymphocytes from SCC cases displayed elevated cytosine and tyrosine (1204, 1210, 1154-1178, 1605-1608 cm\(^{-1}\)), saccharides (847-55 cm\(^{-1}\)), phenylalanine (1004 cm\(^{-1}\)), amide III (1219-54 cm\(^{-1}\)), protein and lipid peaks at 1034 cm\(^{-1}\), 1052 cm\(^{-1}\), and 1400-1445 cm\(^{-1}\), and amide I (1655 cm\(^{-1}\)) compared to the lymphocytes from AC cases (Figure 4.10). Conversely, increased cholesterol and polysaccharides (430, 462-65, 524-87, 608 cm\(^{-1}\)), phosphatidylinositol (520, 594, 766-74 cm\(^{-1}\)), methionine (631-35, 651-55 cm\(^{-1}\)), thiocyanate (735 cm\(^{-1}\)), nucleic acids (729-51, 782-822, 1183-93, 1341-57 cm\(^{-1}\)), and
protein and lipid peaks at 1106, 1307-13, 1341-57, 1378-9, 1472-1539, 1576-94, 1702, and 1747-53 cm$^{-1}$, were associated with the spectra from the lymphocytes from AC cases in comparison to the lymphocytes from SCC cases.

Figure 4.9. Mean Raman spectra of lymphocyte nuclei from negative cases ($n=80$), lymphocyte nuclei ($n=114$) from SCC cases, and lymphocyte nuclei ($n=58$) from AC cases. Shading denotes the standard deviation.
Figure 4.10. Difference spectrum between the negative lymphocyte nuclear spectrum and the SCC lymphocyte nuclear spectrum (blue, \( n=114 \)), (b) the AC lymphocyte nuclear spectrum (red, \( n=58 \)). Difference spectrum between the mean Raman spectrum of lymphocyte nuclei taken from SCC cases and the mean Raman spectrum of squamous cell nuclei taken from AC cases (green). Shading indicates regions of the spectrum that were significantly different (\( P < 0.01 \)).

### 4.3.4 Cytoplasmic spectra analysis

Figure 4.11 shows the mean spectra taken from the cytoplasm of bronchial and squamous cells from negative cases. As the lymphocyte cytoplasm is characteristically small, spectra cannot be acquired from the cytoplasm of lymphocytes without spectral...
contamination from the nucleus as the Raman laser has a spot size of approximately 1 µm. For this reason lymphocytes were not included in this section.

![Figure 4.11. Mean Raman spectra of cytoplasm from negative bronchial cells (black, n=33), and negative squamous cells (blue, n=98). Shading denotes the standard deviation.](image)

4.3.5 Spectral comparison of the cytoplasmic components of bronchial and squamous cells from negative cases

The difference spectrum between the spectra taken from the cytoplasm of squamous cells from negative cases and the spectra taken from the cytoplasm of bronchial cells from negative cases displays significant peaks at 409-32, 449-537,542-47, 556-608, 620-695, 732-43, 764-68, 853-87, 896-970, 985-1054, 1068-85, 1106-1114, 1121-25, 1159-91, 1217-1268, 1296-1333, 1356-1415, 1434-43, 1462-1523, 1531-36, 1542-95, 1636-66, 1673-82, 1699-1721, 1727-32, and 1736-51 cm⁻¹ differentiating the two cell types (Figure 4.12). The squamous cells showed increased peak intensities assigned to cholesterol / lipids / methionine (409-695 cm⁻¹), protein / proline / hydroxyproline / saccharides / glycogen (853-57, 896-970 cm⁻¹), saccharides / lipids / proteins (1106-1125 cm⁻¹), lipids
proteins / nucleic acids (1296-1333 cm\(^{-1}\)), amide I (1656 cm\(^{-1}\)), and lipids (1699-1751 cm\(^{-1}\)), relative to the bronchial cells. In comparison to the bronchial cells the squamous cells displayed decreased peak intensities assigned to the pyrimidine ring breathing mode at 766 cm\(^{-1}\), phenylalanine (1006 cm\(^{-1}\)), tyrosine / nucleic acids (1159-91 cm\(^{-1}\)), amide III / nucleic acids (1217-1268 cm\(^{-1}\)), the protein / lipid / saccharide peaks at 1356-1415 cm\(^{-1}\), the lipid / protein peak at 1445 cm\(^{-1}\), disaccharides / amide II / nucleic acids / COO\(^{-}\) (1460-1595 cm\(^{-1}\)), and amide I of disordered structure / β sheets (1674-1690 cm\(^{-1}\)).

Figure 4.12. The difference spectrum between the mean Raman spectrum from the cytoplasm of squamous cells from negative cases and the mean Raman spectrum from the cytoplasm of bronchial cells from negative cases is depicted in blue. Shading indicates regions of the spectrum that were significantly different (P < 0.01).

4.3.6 Mean cytoplasm spectral analysis of cells from negative cases, SCC cases, and AC cases

4.3.6.1 Bronchial cell analysis

The mean Raman spectra taken from the cytoplasm of bronchial cells from negative, SCC, and AC cases are shown in Figure 4.13. Spectral differences between the spectra from the cytoplasm of bronchial cells from negative cases and spectra from the cytoplasm of bronchial cells from AC and SCC cases corresponded to peaks relative to lipids (433, 576-633, 1285-1289, ~1425, ~1445, 1465-1468 cm\(^{-1}\)) proteins (658-65, 918-40, 1003-
1005, 1159-1163, 1223, 1249-65, 1285-89, 1330-54, 1360, 1402-1412, 1423-27, 1445-52, 1465-68, 1492-95, 1567-75, 1590-1666 cm\(^{-1}\), methionine (638-43, 696-731 cm\(^{-1}\)) carbohydrates (453-67, 488-529, ~568, ~846 cm\(^{-1}\)), and nucleic acids (~672, 676-90, 794-99, 1238-43 cm\(^{-1}\)), as shown in Figure 4.14.

The difference spectra of the bronchial cells from AC and SCC cases had similar profiles, with peaks at 509, 599-626, 788, 848, 922, 988, 1230, 1240-1248, 1288-1297, 1429, 1470, 1597-1661 cm\(^{-1}\) differentiating them from the bronchial cells of negative cases.

The bronchial cells from AC and SCC patients had decreased peak intensities at ~440, 462-65, 488, ~492, 576-611, 658-65, 1106-1142 cm\(^{-1}\), and 1621-1661 cm\(^{-1}\), and increased peak intensities at 846-867, 929-932, ~1005, 1048, 1228-30, 1238-43, 1250-54, 1259-62, 1285-89 cm\(^{-1}\), relative to the bronchial cells from negatives cases. This corresponds to a lower level of proteins, cholesterol, saccharides, phosphatidylinositol, and amide I, and an increase in collagen related peaks, glycogen, glucose, proline, tyrosine, valine, phenylalanine, nucleic acids, and amide III in the bronchial cells of NSCLC patients.
Figure 4.13. Mean Raman spectra of cytoplasm from negative bronchial cells (n=33), bronchial cells (n=68) from SCC cases, and bronchial cells (n=46) from AC cases. Shading denotes the standard deviation.

Figure 4.14. Difference spectrum between the mean Raman spectrum of negative bronchial epithelial cell cytoplasm (n=33) and (a) and the mean Raman spectrum of SCC bronchial epithelial cell cytoplasm (blue, n=68), (b) the mean Raman spectrum of AC bronchial epithelial cell cytoplasm (red, n=46). Difference spectrum between the mean Raman spectrum of bronchial cell cytoplasm taken from SCC cases and the mean Raman spectrum of bronchial cell cytoplasm taken from AC cases (green). Shading indicates regions of the spectrum that were significantly different (P < 0.01).

4.3.6.2 Squamous cell analysis

The mean Raman spectra taken from the cytoplasm of squamous cells from negative, SCC, and AC cases are shown in Figure 4.15. As shown in Figure 4.16, the differences between the spectra taken from the cytoplasm of squamous cells from negative cases and
The spectra taken from the cytoplasm of squamous cells from AC and SCC cases corresponded to peaks relative to proteins, carbohydrates, nucleic acids (403 - 1005 cm\(^{-1}\)), amide I, II, III, and lipids (1162-1171, 1181-1192, 1214-1292, 1337, 1344-1414, 1466-1491, 1595-1665, 1674-1688, 1757-1771 cm\(^{-1}\)). The difference spectra of the squamous cells taken from AC and SCC cases display significant peaks at 418, 424-432, 451-502, 508-527, 568-726, 737-744, 748-751, 935, 950-954, 995, 1004, 1165-1171, 1181-1192, 1214-1292, 1336, 1344-1377, 1472-1489, 1594-1657, 1674-1688, 1770 cm\(^{-1}\) differentiating them from the squamous cells of negative cases. Both AC and SCC cases showed decreased peak intensities at ~418, 424-432, 451-502, 508-527, 568-726, 737-744, 748-751, 935, 950-54, 1594-1657, 1750 cm\(^{-1}\), and increased peak intensities at 995, 1004, 1165-1171, 1181-1192, 1214-1292, 1336-1337, 1472-1489, 1674-1688 cm\(^{-1}\), relative to the negative cases (Figure 4.16).

This corresponds to a lower level of cholesterol (418, 424-432, 608, 614, 702 cm\(^{-1}\)), saccharides (451-502 cm\(^{-1}\)), phosphatidylinositol / proteins (508-527 cm\(^{-1}\)), proteins / nucleic acids / carbohydrates / lipids (568-726 cm\(^{-1}\)), saccharides / proline / valine(935, ~950 cm\(^{-1}\)), amide I (1594-1657 cm\(^{-1}\)), and lipid (~1750 cm\(^{-1}\)), and an increase in collagen related protein peaks (1169, 1223, 1240-47, 1265-69, 1278, 1280, 1283, 1336, 1344, 1488 cm\(^{-1}\)), phenylalanine (995-1004 cm\(^{-1}\)), glycogen (1048 cm\(^{-1}\)), tyrosine (1165-1171 cm\(^{-1}\)), amide III / lipids (1219-92, ~1337 cm\(^{-1}\)), amide II (1472-89 cm\(^{-1}\)), and amide I of disordered structure / β sheets (1674-1688 cm\(^{-1}\)) in the cytoplasm of squamous cells from AC and SCC cases relative to the negative cases. The squamous cells from SCC cases also have reduced peak intensities associated with DNA at (772-791 cm\(^{-1}\)), protein (799-860 cm\(^{-1}\)), tryptophan (~880 cm\(^{-1}\)), and carbohydrates/proteins (896-975 cm\(^{-1}\)) compared to the squamous cells of negative cases.
In comparison to the cytoplasm of squamous cells from AC cases, squamous cells from SCC cases are associated with higher peak intensities at 628, 650, 1410, 1445, 1558, 1590, 1611, and 1622 cm\(^{-1}\), and lower peaks at 834, 860, 910, 942, 964, and 1660 cm\(^{-1}\). This corresponds to higher peaks intensities assigned to tyrosine / methionine (628, 650 cm\(^{-1}\)), lipids / proteins (1410, 1445 cm\(^{-1}\)), amide II (1558 cm\(^{-1}\)), COO\(^-\) (1590 cm\(^{-1}\)), and the tryptophan / cytosine / amide I peaks at 1611 and 1622 cm\(^{-1}\) associated with the SCC cases, and higher peak intensities assigned to saccharides (834, 860, 910 cm\(^{-1}\)), proline / valine / saccharides (942 cm\(^{-1}\)), protein (964 cm\(^{-1}\)), and amide I (1660 cm\(^{-1}\)) associated with the AC cases.

Figure 4.15. Mean Raman spectra of cytoplasm from negative squamous cells \((n=98)\), squamous cells \((n=91)\) from SCC cases, and squamous cells \((n=43)\) from AC cases. Shading denotes the standard deviation.
Figure 4.16. Difference spectrum between the mean Raman spectrum of negative squamous cell cytoplasm ($n=98$) and (a) the mean Raman spectrum of SCC squamous cell cytoplasm (blue, $n=91$), (b) the mean Raman spectrum of AC squamous cell cytoplasm (red, $n=43$). Difference spectrum between the mean Raman spectrum of squamous cell cytoplasm taken from SCC cases and the mean Raman spectrum of squamous cell cytoplasm taken from AC cases (green). Shading indicates regions of the spectrum that were significantly different ($P < 0.01$).
4.3.7 Principal component analysis

PCA was then applied to the dataset to investigate if the different data sets could be discriminated in an unsupervised approach. Figures 4.17-29 demonstrate the PCA scatter plots developed using the spectral data. The PCs which showed the best discriminatory performance were used in the scatter plots. If PCA did not effectively discriminate the negative, SCC and AC subtypes, the first two PCs were depicted in a scatter plot to show the overlapping clusters. Figure 4.17 depicts a PCA scatter plot of bronchial, squamous and lymphocyte nuclei from negative cases. The spectra from bronchial and squamous nuclei form two distinct clusters which are separated along the first principal component. The lymphocyte spectra are dispersed evenly along PC1 and PC2 and cannot be discriminated from the bronchial or squamous cells. The loadings of PC1 show peaks assigned to glycogen / cholesterol / protein / phospholipids (420-690 cm\(^{-1}\)), nucleic acids (762-788 cm\(^{-1}\)), glycogen / cholesterol / saccharides / protein / nucleic acids (805-1011 cm\(^{-1}\)), peaks related to collagen (1024-32 cm\(^{-1}\)), amide III / nucleic acid (1245-60 cm\(^{-1}\)), lipids / protein / amide I (1436-45, 1620-60 cm\(^{-1}\)) associated with the squamous cells, and peaks assigned to lipid / protein / nucleic acids / phosphates (1129-1200, 1351 cm\(^{-1}\)), lipid / protein / COO\(^{-}\) / nucleic acids / cytosine (1450-1546, 1564-82, 1611 cm\(^{-1}\)), and amide I disordered structure/β sheets/lipids (1680-1719 cm\(^{-1}\)) associated with the bronchial cells. The loadings of PC1 are in accordance with the significant peaks identified in the difference spectrum between the mean bronchial nuclei from negative cases and the mean squamous nuclei from negative cases. Although it is not shown, PCA was applied to the bronchial and lymphocyte data but the cell types were not successfully discriminated. Similarly, PCA of the data from squamous cells and lymphocytes showed no discriminatory effect.
Figure 4.17. A PCA scatter plot of bronchial, squamous and lymphocyte nuclei from negative cases shows that the bronchial and squamous spectra form two distinct clusters along the first principal component. The lymphocyte spectra are dispersed evenly along PC1 and PC2 and cannot be discriminated from the bronchial or squamous cells. B The loadings of PC1 show peaks assigned to glycogen / cholesterol / protein / phospholipids (420-690 cm\(^{-1}\)), nucleic acids (762-788 cm\(^{-1}\)), glycogen / cholesterol / saccharides / protein / nucleic acids (805-1011 cm\(^{-1}\)), glycogen / peaks related to collagen (1024-32 cm\(^{-1}\)), amide III / nucleic acid (1245-60 cm\(^{-1}\)), lipids / protein / amide I (1436-45, 1620-60 cm\(^{-1}\)) on the positive side, and peaks assigned to lipid / protein / nucleic acids / phosphates (1129-1200, 1351 cm\(^{-1}\)), lipid / protein / COO / nucleic acids / cytosine (1450-1546, 1564-82, 1611 cm\(^{-1}\)), and amide I disordered structure / \(\beta\) sheets / lipids (1680-1719 cm\(^{-1}\)) on the negative side.
Figure 4.18 shows the scatter plot of the PC scores along the first three principal components, demonstrating overlapping of the spectra taken from the bronchial nuclei of negative, SCC and AC cases. Although it is not shown, separate PCA of negative cases versus AC cases, and negative cases versus SCC cases, yielded no discriminatory effect.

![Image](image_url)

Figure 4.18. Three-dimensional scatter plot of the first three principal components showing no discrimination of spectra taken from the bronchial nuclei of negative, SCC and AC cases.

In Figure 4.19, a scatter plot shows the distribution of AC and SCC spectra after PCA. Although the AC and SCC spectra form overlapping clusters, the AC spectra form a tight cluster in the positive side of PC4, whereas the SCC spectra are widely distributed along PC4. The loadings of PC4 indicate variation in methionine (715 cm\(^{-1}\)), nucleic acids (779 cm\(^{-1}\)), saccharides (840, 860 cm\(^{-1}\)), phenylalanine (1004 cm\(^{-1}\)), amide III / nucleic acids (1220 cm\(^{-1}\)), and amide I (1660 cm\(^{-1}\)) associated with the positive side, and peaks assigned to protein (1100 cm\(^{-1}\)), glycogen (1145 cm\(^{-1}\)), nucleic acids / protein (1340 cm\(^{-1}\)), lipid
(1379 cm\(^{-1}\)), aspartic / glutamic acid (1400 cm\(^{-1}\)), protein / collagen (~930, 1445 cm\(^{-1}\)), and COO\(^-\) (1585 cm\(^{-1}\)) associated with the negative side. This indicates a greater variation in glycogen (1145 cm\(^{-1}\)), nucleic acids / protein (1340 cm\(^{-1}\)), lipid (1379 cm\(^{-1}\)), aspartic / glutamic acid (1400 cm\(^{-1}\)), protein / collagen (~930, 1445 cm\(^{-1}\)), and COO\(^-\) (1585 cm\(^{-1}\)) content in the bronchial cell nuclei of SCC cases in comparison to the bronchial cell nuclei of AC cases.

![Figure 4.19](image.png)

Figure 4.19. A Scatter plot showing spectra taken from the bronchial nuclei of SCC and AC cases. The AC spectra form a tight cluster in the positive side of PC4, whereas the
SCC spectra are widely distributed along PC4. The loadings of PC4 indicate variation in methionine (715 cm\(^{-1}\)), nucleic acids (779 cm\(^{-1}\)), saccharides (840, 860 cm\(^{-1}\)), phenylalanine (1004 cm\(^{-1}\)), amide III/nucleic acids (1220 cm\(^{-1}\)), and amide I (1660 cm\(^{-1}\)) associated with the positive side, and peaks assigned to protein (1100 cm\(^{-1}\)), glycogen (1145 cm\(^{-1}\)), nucleic acids / protein (1340 cm\(^{-1}\)), lipid (1379 cm\(^{-1}\)), aspartic / glutamic acid (1400 cm\(^{-1}\)), protein / collagen (~930, 1445 cm\(^{-1}\)), and COO\(^{-}\) (1585 cm\(^{-1}\)) associated with the negative side.

Figures 4.20 and 4.21 shows the score plots for the analysis of spectra taken from squamous nuclei. The negative, SCC and AC cases were not distinguishable using the model depicted in Figure 4.20. Similarly, the AC and SCC spectra overlap in Figure 4.21 and cannot be differentiated. Although it is not shown, separate PCA of negative cases versus AC cases, and negative cases versus SCC cases, yielded no discriminatory effect.

![PCA scatter plot](image)

**Figure 4.20.** PCA scatter plot of the first two principal components showing no discrimination of spectra taken from the squamous nuclei of negative, SCC and AC cases. PCA did not discriminate between the negative, SCC and AC cases using a combination of PC that explained the first 95% of the variance in the dataset.
Figure 4.21. PCA scatter plot of the first two principal components showing no discrimination of spectra taken from the squamous nuclei of SCC and AC cases. PCA did not discriminate between the negative, SCC and AC cases using a combination of PCs that explained the first 95% of the variance in the dataset.

In Figure 4.22 the distribution of PC1 and PC2 was plotted to visualize the variances between the lymphocytes from negative cases, SCC, and AC cases. The scatter plot using PC1 and PC2 was used to demonstrate that the negative, SCC and AC cases could not be discriminated. Although it is not shown, separate PCA of negative cases versus AC cases, and negative cases versus SCC cases, yielded no discriminatory effect.
Figure 4.22. A Scatter plot showing the distribution of spectra taken from lymphocyte nuclei of negative cases, SCC, and AC cases, along the first and second PCs. PCA did not discriminate between the negative, SCC and AC cases using a combination of PCs that explained the first 95% of the variance in the dataset.

Representative spectra from lymphocyte nuclei, bronchial nuclei, and squamous nuclei of negative cases were selected at random and combined into one negative group. Similarly, representative spectra from the nuclei of each cell type from each lung cancer subtype were selected at random and combined into one malignant group. Figure 4.23 shows that the negative cases could not be discriminated from the malignant cases using PCA. The first two PCs were used to depict the overlapping distribution of negative and malignant spectra.
Figure 4.23. PCA scatter plot showing spectra from the nuclei of lymphocytes, bronchial, and squamous cells of negative cases, SCC cases and AC cases, grouped according to a negative or malignant diagnosis. The negative cases (blue circles) cannot be discriminated from the malignant cases (blue circles) using PCA, which is depicted by the distribution of spectra along the first and second PCs.

Spectra taken from the nuclei of lymphocytes, bronchial and squamous cells from SCC cases were combined into one SCC group. Similarly, spectra taken from the nuclei of lymphocytes, bronchial and squamous cells from AC cases were combined into one AC group. Figure 4.24 shows the PCA scatter plot of the combined SCC group and the combined AC group. The SCC and AC cases cannot be discriminated with PCA, which is depicted with a plot of the first two PCs.
Figure 4.24. PCA scatter plot showing spectra from the nuclei of lymphocytes, bronchial and squamous cells of SCC, and AC cases grouped according to the lung cancer subtype. PCA did not discriminate between the SCC and AC groups using a combination of PCs that explained the first 95% of the variance in the dataset.

Figure 4.25 demonstrates the performance of PCA in discriminating spectra taken from the cytoplasm of bronchial and squamous cells of negative cases. The scatter plot shows that the spectra from bronchial and squamous cells form two distinct clusters that are separated along the first and second principal components. The loadings of PC1 show peaks assigned to cholesterol / lipids / methionine (409-695 cm\(^{-1}\)), protein / proline / hydroxyproline / saccharides / glycogen (853-57, 896-970 cm\(^{-1}\)), saccharides / lipids / proteins (1106-1125 cm\(^{-1}\)), amide I (1656 cm\(^{-1}\)), and lipids (1699-1751 cm\(^{-1}\)) associated with the squamous cells, and peaks assigned to phenylalanine (1006 cm\(^{-1}\)), tyrosine / nucleic acids (1159-91 cm\(^{-1}\)), amide III / nucleic acids (1217-1268 cm\(^{-1}\)), the protein / lipid / saccharide peaks at 1356-1415 cm\(^{-1}\), the lipid / protein peak at 1445 cm\(^{-1}\), disaccharides / amide II / nucleic acids / COO\(^-\) (1460-1595 cm\(^{-1}\)), and amide I of disordered structure / β sheets (1674-1690 cm\(^{-1}\)) associated with the bronchial cells. The
loadings of PC2 show peaks assigned to proline / valine / saccharides (~930 cm\(^{-1}\)) and amide I (1656 cm\(^{-1}\)) associated with the squamous cells.

Figure 4.25. A PCA scatter plot of spectra taken from the cytoplasm of bronchial and squamous cells of negative cases shows that the bronchial and squamous spectra form two distinct clusters along the first and second principal components. B The loadings of PC1 show peaks assigned to cholesterol / lipids / methionine (409-695 cm\(^{-1}\)), protein / proline / hydroxyproline / saccharides / glycogen (853-57, 896-970 cm\(^{-1}\)), saccharides / lipids / proteins (1106-1125 cm\(^{-1}\)), lipids / proteins / nucleic acids (1296-1333 cm\(^{-1}\)), amide I (1656 cm\(^{-1}\)), and lipids (1699-1751 cm\(^{-1}\)) on the positive side and the pyrimidine ring breathing mode at 766 cm\(^{-1}\), phenylalanine (1006 cm\(^{-1}\)), tyrosine / nucleic acids (1159-91 cm\(^{-1}\)), amide III / nucleic acids (1217-1268 cm\(^{-1}\)), the protein / lipid / saccharide
peaks at 1356-1415 cm⁻¹, the lipid / protein peak at 1445 cm⁻¹, disaccharides / amide II / nucleic acids / COO⁻ (1460-1595 cm⁻¹), and amide I of disordered structure / β sheets (1674-1690 cm⁻¹) on the negative side. The loadings of PC2 show peaks assigned to proline / valine / saccharides (~930 cm⁻¹) and amide I (1656 cm⁻¹) on the negative side.

Figure 4.26 shows the PCA scatter plot of spectra taken from the bronchial cytoplasm of negative, SCC and AC cases. The three dimensional scatter plot demonstrates the discrimination of the three clusters of spectra. The first PC separates the SCC cases from the AC cases. The loadings of PC1 show positive peaks associated with amide III / nucleic acids (1250 cm⁻¹), proteins (~1330 cm⁻¹), lipids (1445 cm⁻¹), and amide I (1656 cm⁻¹) associated with the SCC cases, and negative peaks assigned to COO⁻ / amide II / tyrptophan / amide I (~1500-1620 cm⁻¹) associated with the AC cases. A combination of PC2 and PC3 separates the negative cases from the SCC and AC cases. The second PC is mainly responsible for discriminating negative and AC spectra. The loadings of PC2 show peaks assigned to proteins / carbohydrates / nucleic acids (735-890 cm⁻¹), phenylalanine (1000 cm⁻¹), amide III / nucleic acids (1200-1280 cm⁻¹), and amide I of disordered structure / β sheets (1690 cm⁻¹) associated with the AC cases, and peaks assigned to nucleotides / phosphatidylinositol / glycerol (570-620 cm⁻¹), carbohydrates / proteins / lipids (1100-1145 cm⁻¹), phopsholipids / protein / DNA (~1330 cm⁻¹), protein / lipid (~1410 cm⁻¹), and amide I (1650 cm⁻¹) with the negative cases. PC3 discriminates the negative and SCC clusters and the loadings indicate peaks assigned to with increased saccharides (850 cm⁻¹), proline / valine (930 cm⁻¹), and protein / lipid (1410 cm⁻¹) associated with the SCC cases and peaks assigned to phosphatidylinositol (~570-80 cm⁻¹), and amide I (1650 cm⁻¹) associated with the negative cases. The loadings of PC1, PC2 and PC3 are in accordance with the significant peaks identified in the difference spectra between the cytoplasm of bronchial cells from negative, SCC and AC cases.
Figure 4.26. A PCA scatter plot of spectra taken from the bronchial cytoplasm of negative, SCC and AC cases. A three dimensional scatter plot demonstrates the discrimination of the three clusters of spectra. PC1 differentiates the SCC cases and the AC cases. A combination of PC2 and PC3 separates the negative cases from the SCC and AC cases. PC2 is mainly responsible for discriminating negative and AC spectra. PC3 discriminates the negative and SCC clusters. B The loadings of PC1 show positive peaks associated
with amide III/nucleic acids (1250 cm$^{-1}$), proteins (~1330 cm$^{-1}$), lipids (1445 cm$^{-1}$), and negative peaks associated with amide I and II (~1500-1650 cm$^{-1}$). The loadings of PC2 show peaks assigned to proteins/carbohydrates/nucleic acids (735-890 cm$^{-1}$), phenylalanine (1000 cm$^{-1}$), amide III/nucleic acids (1200-1280 cm$^{-1}$), and amide I of disordered structure/β sheets (1690 cm$^{-1}$) on the positive side and nucleotides/phosphatidylinositol/glycerol (570-620 cm$^{-1}$), carbohydrates/proteins/lipids (1100-1145 cm$^{-1}$), collagen (~1330 cm$^{-1}$), protein (~1410 cm$^{-1}$), and amide I (1650 cm$^{-1}$) on the negative side. The loadings of PC3 indicate positive peaks associated with increased saccharides (850 cm$^{-1}$), proline/valine (930 cm$^{-1}$), and protein (1410 cm$^{-1}$), and decreased phosphatidylinositol (~570-80 cm$^{-1}$), and amide I (1650 cm$^{-1}$).

PCA of spectra from the cytoplasm of squamous cells shows variance between the negative cluster and the NSCLC clusters along PC1, with some overlap (Figure 4.27). The majority of spectra from negative cases are distributed in the negative side of PC1, whereas the majority of spectra SCC and AC cases are distributed along the positive side of PC1. The SCC and AC spectra were not discriminated using PCA. The loadings of PC1 show peaks assigned to protein/collagen (1169, 1223, 1240-47, 1265-69, 1278, 1280, 1283, 1336, 1344, 1488 cm$^{-1}$) phenylalanine (995-1004 cm$^{-1}$), glycogen (1048 cm$^{-1}$), tyrosine (1165-1171 cm$^{-1}$), amide III/lipids (1219-92, ~1337 cm$^{-1}$), amide II (1472-89 cm$^{-1}$), and amide I of disordered structure/β sheets (1674-1688 cm$^{-1}$) associated with the SCC and AC cases, and peaks assigned to cholesterol (418, 424-432, 608, 614, 702 cm$^{-1}$), saccharides (451-502 cm$^{-1}$), phosphatidylinositol/proteins (508-527 cm$^{-1}$), proteins/nucleic acids/carbohydrates/lipids (568-726 cm$^{-1}$), proline/valine (935, ~950 cm$^{-1}$), amide I (1594-1657 cm$^{-1}$), and lipid (~1750 cm$^{-1}$) associated with the negative cases. The loadings of PC1 are in accordance with the significant peaks identified in the difference spectra between the cytoplasm of squamous cells from negative, SCC and AC cases.
Figure 4.27. A PCA scatter plot of spectra from the cytoplasm of squamous cells showing variance between the negative cluster and the SCC and AC clusters along PC1, with some overlap. The majority of spectra from negative cases are distributed in the negative side of PC1, whereas the majority of spectra SCC and AC cases are distributed along the positive side of PC1. The SCC and AC spectra form overlapping clusters and cannot be discriminated with PCA. B The loadings of PC1 show peaks assigned to protein/collagen (1169, 1223, 1240-47, 1265-69, 1278, 1280, 1283, 1336, 1344, 1488 cm\(^{-1}\)) phenylalanine (995-1004 cm\(^{-1}\)), glycogen (1048 cm\(^{-1}\)), tyrosine (1165-1171 cm\(^{-1}\)), amide III/ lipids (1219-92, ~1337 cm\(^{-1}\)), amide II (1472-89 cm\(^{-1}\)), and amide I of disordered structure/β sheets (1674-1688 cm\(^{-1}\)) on the positive side, and cholesterol (418, 424-432, 608, 614, 702 cm\(^{-1}\)), saccharides (451-502 cm\(^{-1}\)), phosphatidylinositol/proteins (508-527 cm\(^{-1}\)).
proteins/nucleic acids/carbohydrates/lipids (568-726 cm\(^{-1}\)), proline/valine(935, ~950 cm\(^{-1}\)), amide I (1594-1657 cm\(^{-1}\)), and lipid (~1750 cm\(^{-1}\)) on the negative side.

Representative spectra from the cytoplasm of bronchial and squamous cells of negative cases were selected at random and combined into one negative group. Similarly, representative spectra from the cytoplasm of bronchial and squamous cells from SCC and AC cases were selected at random and combined into one malignant group. Figure 4.28 depicts a PCA scatter plot of the grouped negative cases and grouped malignant cases. The negative and malignant cases cannot be discriminated, however the malignant cases form a cluster along PC1, whereas the negative cases are evenly distributed along PC1. The loadings of PC1 reveal a greater variation in phosphatidylinositol (575 cm\(^{-1}\)), glucose (908 cm\(^{-1}\)), proline / valine / saccharides (~935 cm\(^{-1}\)), saccharide (1110 cm\(^{-1}\)) and amide I (1650 cm\(^{-1}\)) in the cytoplasm of cells from negative cases. The loadings of PC1 are similar to the loadings discriminating spectra from the cytoplasm of bronchial and squamous cells from negative cases, indicating that the variation explained by PC1 is likely differences between the bronchial and squamous cells.
Figure 4.28. A PCA scatter plot showing spectra from the cytoplasm of bronchial and squamous cells of negative cases, SCC cases and AC cases, grouped according to a negative or malignant diagnosis. The negative and malignant cases cannot be discriminated. The malignant cases (red circles) form a cluster along PC1, whereas the negative cases (blue circles) are evenly distributed along PC1. B The loadings of PC1 revealed a greater variation in phosphatidylinositol (575 cm\(^{-1}\)), glucose (908 cm\(^{-1}\)), proline/valine/saccharides (~935 cm\(^{-1}\)), saccharide (1110 cm\(^{-1}\)) and amide I (1650 cm\(^{-1}\)) in the negative cases. The loadings of PC1 are similar to the loadings discriminating spectra from the cytoplasm of bronchial and squamous cells from negative cases, indicating that the variation explained by PC1 is likely the difference between the bronchial and squamous cells.

Spectra taken from the cytoplasm of bronchial and squamous cells from SCC cases were combined into one SCC group. Similarly, spectra taken from the cytoplasm of bronchial
and squamous cells from AC cases were combined into one AC group. Figure 4.29 depicts the PCA scatter plot of the combined SCC group and the combined AC group. The SCC and AC cases cannot be discriminated with PCA, which is depicted with a plot of the first two PCs. The AC cases (red circles) form a two distinct clusters along PC1. The loadings of PC1 are similar to the loadings discriminating spectra from the cytoplasm of bronchial and squamous cells from negative cases, indicating that the two distinct clusters are likely a separation of the bronchial and squamous cells from AC cases.
Figure 4.29. A PCA scatter plot showing spectra from the cytoplasm of bronchial and squamous cells of SCC, and AC cases grouped according to the lung cancer subtype. The SCC and AC cases cannot be discriminated with PCA, which is depicted with a plot of the first two PCs. The AC cases (red circles) form two distinct clusters along PC1. B The loadings of PC1 are similar to the loadings discriminating spectra from the cytoplasm of bronchial and squamous cells from negative cases, indicating that the two distinct clusters are likely a separation of the bronchial and squamous cells from AC cases.

4.3.8 PLSDA classification of NSCLC subtypes

In order to evaluate the efficacy of Raman spectroscopy to accurately classify the NSCLC subtypes, PLSDA classification models were developed. PLSDA employs the same fundamental principle of PCA, but the PLS components, or latent variables (LVs), are rotated to maximise the covariance between spectral variation and class membership. This obtains separation between classes based on the diagnostically relevant spectral variations. To produce an unbiased classifier, the number of spectra used for each class was matched. As the diagnostically significant peaks are identified in the first few LVs, the required number of latent variables in each model was used to set the cross-validation classification error average at 5%. Utilising more LVs to develop the models increases the likelihood that classification is based on noise. As Figure 4.30 illustrates, cell types can be discriminated with PLSDA. The LV score scatter plot of Raman spectra taken from the bronchial cell nuclei of negative cases, squamous cell nuclei of negative cases, and lymphocyte nuclei of negative cases, shows separation of the three cell types using the first 2 latent variables. The spectra from lymphocyte, bronchial and squamous cell nuclei form three distinct clusters which are separated with the first two latent variables. The first LV discriminates the cluster of squamous cell nuclei from the bronchial cell and lymphocyte clusters. The loadings of LV1 show peaks assigned to glycogen / cholesterol / protein / phospholipids (421-690 cm\(^{-1}\)), glycogen / cholesterol / saccharides / protein / nucleic acids (762-1011 cm\(^{-1}\)), glycogen / peaks related to collagen (1024-32 cm\(^{-1}\)), amide
III / nucleic acid (1221-40, 1245-60 cm\(^{-1}\)), lipids / protein / amide I (1436-44, 1620-60 cm\(^{-1}\)) associated with the squamous cells, and peaks assigned to lipid / protein / nucleic acids / phosphates (1129-1200, 1315-20, 1351 cm\(^{-1}\)), and lipid / protein / COO\(^{-}\) / nucleic acids / cytosine (1450-1546, 1564-82, 1611 cm\(^{-1}\)), and amide I disordered structure / β sheets / lipids (1680-1719 cm\(^{-1}\)) associated with the bronchial and lymphocyte nuclei. The lymphocyte nuclei from negative cases are differentiated from the bronchial cell nuclei of negative cases along the second LV. The loadings of LV2 show peaks assigned to proteins and lipids (421-683), amide II / protein / lipid (1450-1562 cm\(^{-1}\)), COO\(^{-}\) (1581-97 cm\(^{-1}\)), amide I (1607-57 cm\(^{-1}\)) and amide I disordered structure / β sheets / lipids (1680-1719 cm\(^{-1}\)) associated with the lymphocyte nuclei from negative cases, and peaks assigned to glycogen / protein / lipids / saccharides (1022-1099 cm\(^{-1}\)), nucleic acids / amide III / proteins (1146-1325 cm\(^{-1}\)), lipid / protein (1433-47 cm\(^{-1}\)), nucleic acids (1576 cm\(^{-1}\)), and the amide I peak from 1662-70 cm\(^{-1}\) associated with the bronchial cell nuclei from negative cases. The loadings of LV1 and LV2 are in accordance with the significant peaks identified in the difference spectra between the lymphocytes, bronchial and squamous cells from negative cases.
Figure 4.30. LV score scatter plot of Raman spectra from bronchial cell nuclei of negative cases (blue circles), squamous cell nuclei of negative cases (green squares), and lymphocyte nuclei of negative cases (solid red circles). The three-way classification model used the first 3 latent variables and yielded sensitivities and specificities of 69-94% for discrimination based on cell type. B and C Latent variables loadings of the
developed PLS-DA model for the dataset obtained from the bronchial, squamous, and lymphocyte nuclei of negative cases.

Cell type can therefore inaccurately influence the performance of a classification model used to classify the NSCLC subtypes. To ensure classification was based on cancer subtype and not discrepancies in cell type, spectra taken from bronchial cells were analysed together, spectra from squamous cells were analysed together, and spectra from lymphocytes were analysed together.

Table 4.2 shows the results of three separate PLSDA models in discriminating spectra from bronchial cell nuclei of negative cases from the bronchial cell nuclei of SCC or AC cases, and the bronchial nuclei of SCC cases from the bronchial nuclei of AC cases. The performance of each two-way classification is shown using the sensitivity and specificity of each model. The first 3 latent variables were used in each case to develop the PLSDA classifier. PLSDA classification of bronchial cells from negative cases vs. bronchial cells from SCC or AC cases yielded sensitivities of ≥91%, and specificities of ≥89%. The bronchial cells from SCC cases were discriminated from the bronchial cells of AC cases with a sensitivity and specificity of 90% and 86% respectively.

Table 4.3 shows the results of the two-way PLSDA classification models used for the discrimination of spectra from the squamous cell nuclei of negative cases from the squamous cell nuclei of SCC or AC cases, and spectra from the squamous cell nuclei of SCC cases from the squamous cell nuclei of AC cases. Each model was developed using 4 latent variables. Each two-way classification model yielded a sensitivity above 73% and a specificity above 78%.

Table 4.4 shows the results for the PLSDA classification of spectra taken from the lymphocyte nuclei of negative cases versus spectra taken from the lymphocyte nuclei of SCC or AC cases, and spectra taken from the lymphocyte nuclei of SCC cases from the
lymphocyte nuclei of AC cases. Each classification is the result of separate two-way models using 3 latent variables each. The models yielded sensitivities above of 82% and specificities above 82% for the discrimination of the SCC and AC subtypes.

Table 4.5 demonstrates the performance of a three-way classification model in discriminating the bronchial cell nuclei of negative cases, the bronchial cell nuclei of SCC, and the bronchial cell nuclei of AC cases. The model yielded diagnostic sensitivities from 81-90% and specificities from 79-91% using the first 5 latent variables. Figure 4.31 is a LV score scatter plot of Raman spectra from bronchial cell nuclei of negative cases, SCC cases, and AC cases, showing separation of the negative cases and NSCLC subtypes. The negative cases are separated from the SCC and AC subtypes along LV1. The loadings of LV1 show peaks assigned to nucleic acids (788 cm\(^{-1}\)), proteins / saccharides (848 cm\(^{-1}\)) amide III / nucleic acids (1223-1297 cm\(^{-1}\)), proteins / lipids / nucleic acids (1350 cm\(^{-1}\)) and amide I of disordered structure / \(\beta\) sheets (1686 cm\(^{-1}\)) associated with the SCC and AC cases, and peaks assigned to protein (509 cm\(^{-1}\)), cholesterol / lipid / methionine (599-626 cm\(^{-1}\)), lipids / amide II (1470 cm\(^{-1}\)), and amide I (1597-1661 cm\(^{-1}\)) associated with the negative cases. The AC and SCC subtypes form two distinct clusters and can be discriminated along the second latent variable. The loadings of LV2 indicate variation in methionine (~700 cm\(^{-1}\)), tryptophan / nucleic acids / phosphatidylinositol (760-776 cm\(^{-1}\)), saccharides (~840, 860 cm\(^{-1}\)), saccharides / proline / valine (930 cm\(^{-1}\)), phenylalanine (1006 cm\(^{-1}\)), lipids / saccharides / proteins / nucleic acids (1130-1206 cm\(^{-1}\)) amide III / nucleic acids (1215, 1280 cm\(^{-1}\)), protein / lipid (1445 cm\(^{-1}\)), amide II (1555 cm\(^{-1}\)), and the amide I peaks at 1611-1628 cm\(^{-1}\) associated with the AC cases, and peaks assigned to protein (509 cm\(^{-1}\)), nucleic acids (788 cm\(^{-1}\)), the protein peak at 1100 cm\(^{-1}\), nucleic acids / protein (1340 cm\(^{-1}\)), lipid (1379 cm\(^{-1}\)), amide II / nucleic acids (1480 cm\(^{-1}\)), and COO\(^{-}\) (1585 cm\(^{-1}\)) associated with the SCC cases.
Figure 4.31. **A** Scatter plot of the latent variables (LVs) obtained from the Raman spectral dataset of bronchial cell nuclei from negative cases (blue circles), SCC cases (green squares), and AC cases (solid red circles). **B** and **C** Latent variables loadings of the developed PLS-DA model for the dataset obtained from bronchial nuclei.
Table 4.6 summarises the results for the three-way classification of the negative, SCC and AC subtypes, using spectra taken from squamous cell nuclei. The first 6 latent variables were used to develop the three-way PLSDA model. Squamous cells from negative cases were discriminated from the squamous cells of SCC and AC cases with a sensitivity and specificity above 91%. Squamous cells from AC cases were identified with a sensitivity and specificity of 63 and 79% respectively. The SCC subtype was classified with a sensitivity of 60%, and a specificity of 78%. Figure 4.32 depicts the scatter plots of the spectra taken from the squamous cell nuclei of negative cases and spectra from the squamous cell nuclei of each NSCLC subtype, along the first three latent variables. The negative spectra can be discriminated from the SCC and AC spectra using a combination of the first two latent variables. The loadings of the LV1 (figure 4.33) and indicate peaks assigned to phenylalanine (1006 cm\(^{-1}\)), nucleic acids / proteins (1318-1323, 1346-58 cm\(^{-1}\)), protein / lipid (1456-1462 cm\(^{-1}\)), and amide I (1660 cm\(^{-1}\)) associated with the negative cases, and peaks assigned to nucleic acids (780, 1513-1516 cm\(^{-1}\)), amide II (1526 cm\(^{-1}\)) and tryptophan / nucleic acids (1534-38, 1567-72 cm\(^{-1}\)) associated with the SCC and AC cases. The second LV responsible for discriminating the negative cases is LV2 (Figure 4.33). The loadings of LV2 show peaks assigned to polysaccharides (495 cm\(^{-1}\)), protein (504, 597-608), , and amide I (1656 cm\(^{-1}\)) associated with the negative cases, and peaks assigned to amide III / nucleic acids (1200-1206, 1219-1229) and amide I of disordered structure / β sheets 1686 (cm\(^{-1}\)) associated with the SCC and AC cases. LV3 is responsible for discriminating the SCC and AC clusters. The loadings of LV3 show peaks assigned to DNA (779 cm\(^{-1}\)), amide III / nucleic acids (1230 cm\(^{-1}\)), and amide II / COO\(^{-}\) / amide I (1480-1656 cm\(^{-1}\)) associated with the AC cases, and peaks assigned to protein /
carbohydrates (~930 cm\(^{-1}\)), proteins / lipids / saccharides / nucleic acids (1370, 1387-1485 cm\(^{-1}\)) associated with the SCC cases (figure 4.33).

Figure 4.32. A Scatter plot of the first two LVs obtained from the Raman spectral dataset of squamous cell nuclei from negative cases (blue circles), SCC cases (green squares), and AC cases (solid red circles). B Scatter plot of the first and third LVs obtained from the Raman spectral dataset of squamous nuclei from negative cases (blue circles), SCC cases (green squares), and AC cases (solid red circles).
Figure 4.33. Latent variables from PLSDA of Raman spectra from squamous cell nuclei of negative cases, SCC cases, and AC cases. A-C The Latent variables loadings of the developed PLS-DA model for the dataset obtained from squamous cell nuclei. A depicts LV1, B shows the loadings of LV2, and C illustrates the loadings of LV3.

Table 4.7 shows the performance of a three-way PLSDA model for the classification of negative, SCC, and AC using spectra from lymphocyte nuclei. Using 5 latent variables,
spectra in the negative group were differentiated from the SCC and AC subtypes with a sensitivity and specificity above 78%. The classification model identified the spectra from the SCC subtype with a sensitivity and specificity of 80% and 78% respectively, and spectra from the AC subtype were detected with a sensitivity of 81% and specificity of 88%. Figure 4.34 is a LV score scatter plot of Raman spectra from lymphocyte nuclei of negative cases, SCC cases, and AC cases. The SCC spectra and the AC spectra form two separate clusters along the first latent variable, which can be discriminated from the negative cases along LV2, with some overlap. The loadings of LV1 indicate variation in thiocyante (735 cm\(^{-1}\)), nucleotide conformation (794 cm\(^{-1}\)), protein (1106 cm\(^{-1}\)), protein / nucleic acids (1341 cm\(^{-1}\)), lipid / amide II (1465-1493 cm\(^{-1}\)), COO\(^{-}\) (1585 cm\(^{-1}\)), and lipid / aspartic / glutamic acid (1702 cm\(^{-1}\)) associated with the AC cases, and peaks assigned to phenylalanine (1006 cm\(^{-1}\)), cytosine / tyrosine (1161-1174 cm\(^{-1}\)), amide III / nucleic acids (1280 cm\(^{-1}\)), lipid / protein / aspartic / glutamic acid (1400-1445 cm\(^{-1}\)), amide II (1560 cm\(^{-1}\)), amide I (1606-67 cm\(^{-1}\)), and amide I disordered structure / β sheets (1690 cm\(^{-1}\)) associated with the SCC cases. The loadings of LV1 are in accordance with the significant peaks identified in the difference spectrum between the lymphocytes of SCC cases and the lymphocyte nuclei of AC cases. The loadings of LV2 show peaks assigned to amide III / nucleic acids (~1240 cm\(^{-1}\)), proteins and lipids (1387-1453 cm\(^{-1}\)) associated with the SCC and AC cases, and a peak associated with amide I (1650 cm\(^{-1}\)) associated with the negative cases. The loadings of LV2 are a combination of the two difference spectra between the lymphocyte nuclei of negative cases and the lymphocyte nuclei of SCC and AC cases.
Figure 4.34. A Scatter plot of the LVs obtained from the Raman spectral dataset of lymphocyte nuclei from negative cases (blue circles), SCC cases (green squares), and AC cases (solid red circles). B and C Latent variables loadings of the developed PLS-DA model for the dataset obtained from lymphocyte nuclei.
Table 4.8 shows the results of a PLSDA model in discriminating spectra taken from bronchial cell nuclei of negative cases from the spectra taken from bronchial cell nuclei of SCC and AC cases combined into one malignant group. In order to produce an unbiased classifier, the total number of spectra from malignant cases used to develop the classifier was identical to the number of spectra from negative cases being used. The first 3 latent variables were used to develop the PLSDA classifier. PLSDA classification of bronchial cells from negative cases vs. bronchial cells from malignant cases yielded a sensitivity of 82%, and a specificity of ≥88%. Table 4.8 also shows the results of the PLSDA model used to differentiate between spectra taken from the squamous cell nuclei of negative cases, and the squamous cell nuclei of AC and SCC cases combined into one malignant group. The total number of spectra from malignant cases used to develop the classifier was identical to the number of spectra from negative cases being used. Using 3 LVs the two-way classifier yielded a sensitivity of 88% and a specificity of 84% for the classification of squamous cell nuclei from negative and malignant cases. Table 4.8 also shows the results of a PLSDA model in discriminating spectra taken from the lymphocyte nuclei of negative cases, and spectra taken from the lymphocyte nuclei of SCC and AC cases combined into one malignant group. The total number of spectra from malignant cases used to develop the classifier was identical to the number of spectra from negative cases being used. The negative and malignant lymphocyte nuclei were classified with a sensitivity of 81%, and a specificity of 77% using 3 LVs.

Then the ability of PLSDA to correctly classify the NSCLC subtypes using a dataset of combined cell types was evaluated. Representative spectra from lymphocyte nuclei, bronchial cell nuclei, and squamous cell nuclei of negative cases were selected at random and combined into one negative group. Similarly, representative spectra from the nuclei
of each cell type from each lung cancer subtype were selected at random and combined into one malignant group. In order to produce an unbiased classifier, the total number of spectra from malignant cases used to develop the classifier was identical to the number of spectra from negative cases being used. In addition, the number of spectra used from each cell type was also matched in the negative and malignant groups. That is, the number of bronchial spectra used to develop the negative group matched the number of bronchial spectra used to develop the malignant group, and so on. Table 4.9 shows results achieved for discriminating spectra from the negative and malignant groups. The two-way classifier yielded a sensitivity of 89% and a specificity of 79% using 3 LVs. Figure 4.35 shows the discriminatory performance of PLSDA in differentiating the grouped negative and grouped malignant spectra. The negative cases were discriminated from the malignant cases with some overlap using the first and third LVs. The spectra from negative cases are predominantly distributed in the positive side of LV1, whereas spectra from the malignant cases are predominantly distributed in the negative side of LV1. The loadings of LV1 show peaks assigned to tyrosine / methionine (628, 658 cm$^{-1}$), phenylalanine (1006 cm$^{-1}$), lipid (~1128 cm$^{-1}$), the lipid and protein peak at 1445 cm$^{-1}$, and amide I (1650 cm$^{-1}$) associated with the negative group, and peaks assigned to DNA (780 cm$^{-1}$), the saccharide / protein / lipid / nucleic acids peaks at 1020-1105 cm$^{-1}$, lipid (1379 cm$^{-1}$), amide II (1489 cm$^{-1}$), and COO$^{-}$ 1575 associated with the malignant group. The spectra from negative cases are predominantly distributed in the negative side of LV3, whereas spectra from the malignant cases are predominantly distributed in the positive side of LV3. The loadings of LV3 show peaks assigned to proteins (~509 cm$^{-1}$), nucleic acids / saccharides / protein / lipids (729-1013 cm$^{-1}$), protein / lipids (1430 cm$^{-1}$), and amide I (~1650 cm$^{-1}$) associated with the negative group, and peaks assigned to carbohydrates / proteins / lipids / nucleic acids (1110-1198 cm$^{-1}$), nucleic acids / protein
(-1340 cm\(^{-1}\)), lipid (1379 cm\(^{-1}\)), \text{COO}^- / \text{amide II} / \text{tryptophan} / \text{amide I} (-1475-1610 \text{ cm}^{-1}), \text{amide I of disordered structure} / \beta \text{ sheets} (1690 \text{ cm}^{-1}) \text{ associated with the malignant group.}

Figure 4.35. A Scatter plot of the first and third LVs obtained from the Raman spectral dataset of combined bronchial, squamous, and lymphocyte nuclei. Representative spectra from each cell type from the negative, SCC, and AC cases were grouped according to a
negative or malignant diagnosis. The negative cases (red circles) are discriminated from the malignant cases (blue circles) along the first and third LVs. B and C Latent variables loadings of the developed PLS-DA model for the dataset obtained from bronchial, squamous, and lymphocyte nuclei.

Spectra taken from the nuclei of lymphocytes, bronchial and squamous cells from SCC cases were combined into one SCC group. Similarly, spectra taken from the nuclei of lymphocytes, bronchial and squamous cells from AC cases were combined into one AC group. In order to produce an unbiased classifier, the total number of spectra from SCC cases used to develop the classifier was identical to the number of spectra from AC cases being used. As before, the number of spectra used from each cell type was also matched in the SCC and AC groups. Table 4.9 shows results achieved for discriminating spectra from the SCC and AC groups. The two-way classifier yielded a sensitivity of 84% and a specificity of 69% using 3 LVs. Figure 4.36 shows the PLSDA scatter plot of the combined SCC group and the combined AC group. The SCC cases were discriminated from the AC cases with some overlap using the first and second LVs. The spectra from SCC cases are predominantly distributed in the positive side of LV1, whereas spectra from the AC cases are predominantly distributed in the negative side of LV1. The loadings of LV1 indicate variation in protein (509 cm\(^{-1}\)), thiocyanate (735 cm\(^{-1}\)), nucleic acids (790 cm\(^{-1}\)), protein (1106 cm\(^{-1}\)), lipid (1380 cm\(^{-1}\)), amide II (1493 cm\(^{-1}\)), and COO\(^{-}\) (1585 cm\(^{-1}\)) associated with the AC cases, and peaks assigned to saccharides (840 cm\(^{-1}\)), proline / valine / saccharides (930 cm\(^{-1}\)), phenylalanine (1006 cm\(^{-1}\)), cytosine / tyrosine (1161-1174 cm\(^{-1}\)), amide III / nucleic acids (1224 cm\(^{-1}\)), lipid / protein / aspartic / glutamic acid (1400-1445 cm\(^{-1}\)), and amide I (1666 cm\(^{-1}\)) associated with the SCC cases. The spectra from SCC cases are predominantly distributed in the negative side of LV2, whereas spectra from the AC cases are predominantly distributed in the positive side of LV1. The loadings of LV2 indicate variation in protein (509 cm\(^{-1}\)), tyrosine / methionine
(631, 646 cm\(^{-1}\)), nucleic acids (790 cm\(^{-1}\)), protein (1106 cm\(^{-1}\)), lipid (1380 cm\(^{-1}\)), amide II (1493 cm\(^{-1}\)), and COO\(^{-}\) (1585 cm\(^{-1}\)) associated with the SCC cases, and peaks assigned to saccharides (840, 860 cm\(^{-1}\)), proline / valine / saccharides (930 cm\(^{-1}\)), phenylalanine (1006 cm\(^{-1}\)), cytosine / tyrosine (1161-1174 cm\(^{-1}\)), lipid / protein (1445 cm\(^{-1}\)), and amide I (>1600 cm\(^{-1}\)) associated with the AC cases.
Figure 4.36. A Scatter plot showing spectra from the bronchial, squamous, and lymphocyte nuclei of SCC and AC cases grouped according to the lung cancer subtype. The AC cases (red circles) are discriminated from the SCC cases (blue circles) along the first and second LVs. B and C Latent variables loadings of the developed PLS-DA model for the dataset obtained from bronchial, squamous, and lymphocyte nuclei.
To evaluate the ability of PLSDA to discriminate negative cases from NSCLC subtypes using spectra taken from the cytoplasm of cells, spectra taken from the cytoplasm of bronchial cells were used to develop a supervised classifier. Another classifier was developed with spectra taken from the cytoplasm of squamous cells. As spectra from the cytoplasm of bronchial and squamous cells of negative cases were effectively discriminated using PCA, the two cell types were not comparatively analysed using PLSDA. Table 4.10 shows that the 3 way classification model discriminated the bronchial cytoplasmic spectra from negative cases, SCC cases, and AC cases, with sensitivities of 68-97%, and specificities of 85-95% using the first 4 latent variables. Figure 4.37 depicts the discriminatory performance of the PLSDA model developed using spectra taken from the cytoplasm of bronchial cells from negative, SCC and AC cases. The first LV separates the SCC cases from the AC cases. The loadings of LV1 show negative peaks assigned to proteins (~1330 cm\(^{-1}\)), saccharides/proteins (~930 cm\(^{-1}\)), lipids (1445 cm\(^{-1}\)) and amide I (1656 cm\(^{-1}\)) associated with the SCC cases, and positive peaks assigned to cholesterol/lipids/nucleic acids (705-820 cm\(^{-1}\)), COO\(^{-}\)/amide II/tyrptophan/amide I (~1500-1620 cm\(^{-1}\)) and the nucleic acids and phosphates peak at 1200 cm\(^{-1}\) associated with the AC cases. LV2 separates the negative cases from the SCC and AC cases. The loadings of LV2 show peaks assigned to proteins/carbohydrates/nucleic acids (735-890 cm\(^{-1}\)), phenylalanine (1000 cm\(^{-1}\)), amide III/nucleic acids (1200-1280 cm\(^{-1}\)), saccharides/proline/valine (930 cm\(^{-1}\)), and protein (1410 cm\(^{-1}\)) associated with the SCC and AC cases, and peaks assigned to nucleotides/phosphatidylinositol/glycerol (570-620 cm\(^{-1}\)), carbohydrates/proteins/lipids (1100-1145 cm\(^{-1}\)), phospholipids/protein/DNA (~1330 cm\(^{-1}\)), and amide I (1650 cm\(^{-1}\)) associated with the negative cases.
Figure 4.37. **A** Scatter plot of the LVs obtained from the Raman spectra taken from the cytoplasm of bronchial cells from negative cases (blue circles), SCC cases (green squares), and AC cases (solid red circles). **B** and **C** Latent variables loadings of the developed PLS-DA model for the dataset obtained from the cytoplasm of bronchial cells.
Table 4.11 shows the results for the discriminant analysis of spectra taken from the cytoplasm of squamous cells. The three-way PLSDA classification model, which was developed using the first 4 latent variables, differentiated spectra from the negative cases from the NSCLC cases with a sensitivity of 88% and a specificity of 94%. Spectra taken from the cytoplasm of squamous cells from SCC cases were classified with a sensitivity and specificity above 76%, and the spectra from AC cases were identified by the model with a sensitivity and specificity of over 78%. Figure 4.38 shows the scatter plot of the latent variable scores of the spectra taken from the cytoplasm of squamous cells from negative cases, SCC cases, and AC cases. The spectra from SCC and AC cases are discriminated from the spectra of negative cases using the first latent variable. The loadings of LV1 show peaks assigned to cholesterol (418, 424-432, 608, 614, 702 cm$^{-1}$), saccharides (451-502 cm$^{-1}$), phosphatidylinositol / proteins (508-527 cm$^{-1}$), proteins / nucleic acids / carbohydrates / lipids (568-726 cm$^{-1}$), saccharides / proline / valine (935, ~950 cm$^{-1}$), amide I (1594-1657 cm$^{-1}$), and lipid (~1750 cm$^{-1}$) associated with the negative cases, and peaks assigned to phenylalanine (995-1004 cm$^{-1}$), glycogen (1048 cm$^{-1}$), tyrosine (1165-1171 cm$^{-1}$), amide III / lipids (1219-92, ~1337 cm$^{-1}$), amide II (1472-89 cm$^{-1}$), and amide I of disordered structure / β sheets (1674-1688 cm$^{-1}$) associated with the cytoplasm of squamous cells from AC and SCC cases. The second latent variable differentiates the SCC and AC clusters. The loadings of LV2 show peaks assigned to tyrosine / methionine (628, 650 cm$^{-1}$), nucleic acids / lipids / proteins (1371-1446 cm$^{-1}$), amide II (1558 cm$^{-1}$), COO$^-$ (1590 cm$^{-1}$), and the tryptophan / cytosine / amide I peaks at 1611 and 1622 cm$^{-1}$ associated with the SCC cases, and peaks assigned to saccharides (834, 860, 910 cm$^{-1}$), proline / valine / saccharides (942 cm$^{-1}$), protein (964 cm$^{-1}$), and amide I (1660 cm$^{-1}$) associated with the AC cases.
Figure 4.38. A Scatter plot of the LVs obtained from the Raman spectra taken from the cytoplasm of squamous cells from negative cases (blue circles), SCC cases (green squares), and AC cases (solid red circles). B and C Latent variables loadings of the developed PLS-DA model for the dataset obtained from the cytoplasm of squamous cells.
Table 4.12 shows the results of a PLSDA model in discriminating spectra taken from the cytoplasm of bronchial cells from negative cases, and spectra taken from the cytoplasm of bronchial cells from SCC and AC cases combined into one malignant group. In order to produce an unbiased classifier, the total number of spectra in the combined malignant group was identical to the number of spectra from negative cases being used. The first 2 latent variables were used to develop the PLSDA classifier. PLSDA classification of bronchial cells from negative cases vs. bronchial cells from malignant cases yielded a sensitivity of 94%, and a specificity of 94%. Table 4.12 also shows the results of the PLSDA model used to differentiate between spectra taken from the cytoplasm of squamous cells from negative cases, and spectra taken from the cytoplasm of squamous cells of AC and SCC cases combined into one malignant group. The total number of spectra in the malignant group matched the number of spectra in the negative group. Using 3 LVs the two-way classifier yielded a sensitivity of 96% and a specificity of 98% for the classification of squamous nuclei from negative and malignant cases. In a two-way classifier using 4 LVs, spectra from the cytoplasm of bronchial cells from SCC cases were discriminated from spectra from the cytoplasm of bronchial cells from AC cases with a sensitivity and specificity of 89% and 90% respectively (Table 4.12). In addition, Table 4.12 shows the results for the discriminant analysis of spectra from the cytoplasm of squamous cells from SCC cases and spectra from the cytoplasm of squamous cells from AC cases. The SCC and AC cases were distinguished with a sensitivity of 81% and a specificity of 74%.

Representative spectra from the cytoplasm of bronchial and squamous cells of negative cases were selected at random and combined into one negative group. Similarly, representative spectra from the cytoplasm of bronchial and squamous cells from SCC and
AC cases were selected at random and combined into one malignant group. In order to produce an unbiased classifier, the total number of spectra from malignant cases used to develop the classifier was identical to the number of spectra from negative cases being used. As before, the number of spectra used from each cell type was also matched in the negative and malignant groups. Table 4.13 shows results achieved for discriminating spectra from the negative and malignant groups. The two-way classifier yielded a sensitivity of 87% and a specificity of 90% using 4 LVs. Figure 4.39 depicts a PLSDA scatter plot of the grouped negative cases and grouped malignant cases. The negative and malignant groups are discriminated by the first two LVs. The spectra from negative cases are predominantly distributed in the positive side of LV1, whereas spectra from the malignant cases are predominantly distributed in the negative side of LV1. The loadings of LV1 reveal peaks assigned to phosphatidylinositol (575 cm\(^{-1}\)), proline / valine / saccharides (~935 cm\(^{-1}\)), protein / lipid / saccharides (1101-40 cm\(^{-1}\)) and amide I (1650 cm\(^{-1}\)) associated with the cytoplasm of cells from negative cases, and peaks assigned to phenylalanine (1006 cm\(^{-1}\)), amide III / nucleic acids (1220-1260 cm\(^{-1}\)), lipid (1375 cm\(^{-1}\)), amide II (1489 cm\(^{-1}\)), amide I of disordered structure / β sheets (1690 cm\(^{-1}\)) associated with the malignant group. The spectra from negative cases are predominantly distributed in the positive side of LV2, whereas spectra from the malignant cases are predominantly distributed in the negative side of LV2. The loadings of LV2 show peaks assigned to phenylalanine (1006 cm\(^{-1}\)), lipid (1375 cm\(^{-1}\)), protein / aspartic / glutamic acid (1400 cm\(^{-1}\)), lipid / protein / nucleic acids (1430 cm\(^{-1}\)), amide II (1555 cm\(^{-1}\)), amide I / cytosine / tryptophan (1611-26 cm\(^{-1}\)), associated with the negative group, and peaks assigned to glucose (908 cm\(^{-1}\)), proline / valine / saccharides (~935 cm\(^{-1}\)), saccharide (1110 cm\(^{-1}\)), 1280-1376 cm\(^{-1}\)) and amide I (1650 cm\(^{-1}\)) associated with the malignant group.
Figure 4.39. A Scatter plot showing spectra from the cytoplasm of bronchial and squamous cells of negative cases, SCC cases and AC cases, grouped according to a negative or malignant diagnosis. The negative cases (green circles) are discriminated from the malignant cases (blue circles) along the first and second LVs. B and C Latent variables loadings of the developed PLS-DA model for the dataset obtained from the cytoplasm of bronchial and squamous cells.
Spectra taken from the cytoplasm bronchial and squamous cells from SCC cases were combined into one SCC group. Similarly, spectra taken from the cytoplasm of bronchial and squamous cells from AC cases were combined into one AC group. In order to produce an unbiased classifier, the total number of spectra from SCC cases used to develop the classifier was identical to the number of spectra from AC cases being used. As before, the number of spectra used from each cell type was also matched in the SCC and AC groups. Table 4.13 shows results achieved for discriminating spectra from the SCC and AC groups. The two-way classifier yielded a sensitivity of 85% and a specificity of 71% using 2 LVs. Figure 4.40 shows the PLSDA scatter plot of the combined SCC group and the combined AC group along the first two latent variables. The negative and malignant groups are discriminated by the first two LVs. The spectra from AC cases are predominantly distributed in the positive side of LV1, whereas spectra from the SCC cases are predominantly distributed in the negative side of LV1. The loadings of LV1 reveal peaks assigned to protein (508 cm$^{-1}$) nucleic acids (790 cm$^{-1}$), phenylalanine (1006 cm$^{-1}$), amide III / nucleic acids (1201-1245 cm$^{-1}$), and amide II / COO$^{-}$ / nucleic acids / amide I (1495-1656 cm$^{-1}$) associated with the cytoplasm of cells from AC group, and peaks assigned to saccharides / proline / valine (930 cm$^{-1}$), lipids (1131 cm$^{-1}$), nucleic acids (1320 cm$^{-1}$), aspartic / glutamic acid (~1405 cm$^{-1}$) and lipids / proteins (1445 cm$^{-1}$) associated with the SCC group.
Figure 4.40. A Scatter plot showing spectra from the cytoplasm of bronchial and squamous cells of SCC, and AC cases grouped according to the lung cancer subtype. The AC cases (red circles) are discriminated from the SCC cases (blue circles) along the first LV. B The Latent variable loadings of the developed PLS-DA model for the dataset obtained from the cytoplasm of bronchial and squamous cells.
Table 4.2. Performance of PLSDA in discriminating bronchial cell nuclei of negative cases, SCC cases, and AC cases, in separate two-way classifications.

<table>
<thead>
<tr>
<th>Comparison</th>
<th>Sensitivity (%)</th>
<th>Specificity (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Negative v SCC</td>
<td>91</td>
<td>89</td>
</tr>
<tr>
<td>Negative v AC</td>
<td>95</td>
<td>90</td>
</tr>
<tr>
<td>SCC v AC</td>
<td>90</td>
<td>86</td>
</tr>
</tbody>
</table>

Table 4.3. Performance of PLSDA in discriminating squamous cell nuclei of negative cases, SCC cases, and AC cases, in separate two-way classifications.

<table>
<thead>
<tr>
<th>Comparison</th>
<th>Sensitivity (%)</th>
<th>Specificity (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Negative v SCC</td>
<td>86</td>
<td>84</td>
</tr>
<tr>
<td>Negative v AC</td>
<td>87</td>
<td>93</td>
</tr>
<tr>
<td>SCC v AC</td>
<td>73</td>
<td>78</td>
</tr>
</tbody>
</table>

Table 4.4. Performance of PLSDA in discriminating lymphocyte nuclei of negative cases, SCC cases, and AC cases, in separate two-way classifications.

<table>
<thead>
<tr>
<th>Comparison</th>
<th>Sensitivity (%)</th>
<th>Specificity (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Negative v SCC</td>
<td>89</td>
<td>82</td>
</tr>
<tr>
<td>Negative v AC</td>
<td>89</td>
<td>97</td>
</tr>
<tr>
<td>SCC v AC</td>
<td>82</td>
<td>85</td>
</tr>
</tbody>
</table>

Table 4.5. Performance of PLSDA in discriminating the bronchial cell nuclei of negative cases, SCC cases and AC cases.

<table>
<thead>
<tr>
<th>Histological diagnosis</th>
<th>Sensitivity (%)</th>
<th>Specificity (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Negative</td>
<td>90</td>
<td>91</td>
</tr>
<tr>
<td>SCC</td>
<td>82</td>
<td>84</td>
</tr>
<tr>
<td>AC</td>
<td>81</td>
<td>79</td>
</tr>
</tbody>
</table>
Table 4.6. Performance of PLSDA in discriminating the squamous cell nuclei of negative cases, SCC cases and AC cases.

<table>
<thead>
<tr>
<th>Histological diagnosis</th>
<th>Sensitivity (%)</th>
<th>Specificity (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Negative</td>
<td>91</td>
<td>91</td>
</tr>
<tr>
<td>SCC</td>
<td>60</td>
<td>78</td>
</tr>
<tr>
<td>AC</td>
<td>63</td>
<td>79</td>
</tr>
</tbody>
</table>

Table 4.7. Performance of PLSDA in discriminating the lymphocyte nuclei of negative cases, SCC cases and AC cases.

<table>
<thead>
<tr>
<th>Histological diagnosis</th>
<th>Sensitivity (%)</th>
<th>Specificity (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Negative</td>
<td>78</td>
<td>85</td>
</tr>
<tr>
<td>SCC</td>
<td>80</td>
<td>78</td>
</tr>
<tr>
<td>AC</td>
<td>81</td>
<td>88</td>
</tr>
</tbody>
</table>

Table 4.8. Performance of PLSDA in discriminating negative cases from malignant cases, using spectra from the nuclei of bronchial cells, squamous cells, and lymphocytes.

<table>
<thead>
<tr>
<th>Comparison</th>
<th>Bronchial classification</th>
<th>Squamous classification</th>
<th>Lymphocyte classification</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Sensitivity (%)</td>
<td>Specificity (%)</td>
<td>Sensitivity (%)</td>
</tr>
<tr>
<td>Negative vs Malignant</td>
<td>82</td>
<td>88</td>
<td>88</td>
</tr>
</tbody>
</table>
Table 4.9. Performance of PLSDA in discriminating negative cases from malignant cases, and SCC cases from AC cases using combined spectra from the nuclei of bronchial cells, squamous cells, and lymphocytes.

<table>
<thead>
<tr>
<th>Comparison</th>
<th>Sensitivity (%)</th>
<th>Specificity (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Negative vs Malignant</td>
<td>89</td>
<td>79</td>
</tr>
<tr>
<td>SCC vs AC</td>
<td>84</td>
<td>69</td>
</tr>
</tbody>
</table>

Table 4.10. Performance of PLSDA in discriminating the cytoplasm of bronchial cells from negative cases, SCC cases and AC cases.

<table>
<thead>
<tr>
<th>Histological diagnosis</th>
<th>Sensitivity (%)</th>
<th>Specificity (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Negative</td>
<td>97</td>
<td>95</td>
</tr>
<tr>
<td>SCC</td>
<td>68</td>
<td>88</td>
</tr>
<tr>
<td>AC</td>
<td>80</td>
<td>85</td>
</tr>
</tbody>
</table>

Table 4.11. Performance of PLSDA in discriminating the cytoplasm of squamous cells from negative cases, SCC cases and AC cases.

<table>
<thead>
<tr>
<th>Histological diagnosis</th>
<th>Sensitivity (%)</th>
<th>Specificity (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Negative</td>
<td>88</td>
<td>94</td>
</tr>
<tr>
<td>SCC</td>
<td>76</td>
<td>79</td>
</tr>
<tr>
<td>AC</td>
<td>81</td>
<td>78</td>
</tr>
</tbody>
</table>

Table 4.12. Performance of PLSDA in discriminating negative cases from malignant cases, and SCC from AC cases using spectra from the cytoplasm of bronchial cells and squamous cells.

<table>
<thead>
<tr>
<th>Comparison</th>
<th>Bronchial classification</th>
<th>Squamous classification</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Sensitivity (%)</td>
<td>Specificity (%)</td>
</tr>
<tr>
<td>Negative vs Malignant</td>
<td>94</td>
<td>94</td>
</tr>
</tbody>
</table>
Table 4.13. Performance of PLSDA in discriminating negative cases from malignant cases, and SCC cases from AC cases using combined spectra from the cytoplasm of bronchial and squamous cells.

<table>
<thead>
<tr>
<th>Comparison</th>
<th>Sensitivity (%)</th>
<th>Specificity (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Negative vs Malignant</td>
<td>87</td>
<td>90</td>
</tr>
<tr>
<td>SCC vs AC</td>
<td>85</td>
<td>71</td>
</tr>
</tbody>
</table>

4.4 Discussion

To our knowledge, this is the first study to apply Raman spectroscopy to bronchoscopy cytology samples to obtain a biomolecular characterisation of NSCLC subtypes and provide a marker free diagnosis of NSCLC. Bronchoscopic methods for ascertaining cellular material, such as BAL and BW, collect a mixture of different cell types. In this study, to eradicate variation based on cell type, spectral data from the distinct cell types were processed and analysed separately. This enabled the identification of spectral differences across the negative, SCC and AC cases. In addition, to conduct a comprehensive analysis of the dataset, spectral data from different cell types were grouped by a negative or malignant diagnosis, and by SCC or AC subtype. The different groupings were comparatively analysed to determine the diagnostic efficacy of combining the Raman data from different cell types.

It was possible to identify bronchial and squamous cells on the unstained slides due to their unique morphologies. As lymphocytes are more difficult to identify on unstained slides, the locations of targeted lymphocytes were mapped during Raman spectroscopic analysis. Through the application of the Papanicolaou stain and analysis of the nuclear
patterns and cell morphology, the cell lineage was determined to be immune cell, more specifically a lymphocyte. The locations of targeted abnormal squamous cells on the unstained slides were also marked during spectral acquisition. The Pap stain enabled visualisation of the neoplastic hyperchromatic cells with abundant cytoplasm and cytoplasmic projections, and the cells were determined as malignant squamous cells. Further, cytoplasmic spectra could not be acquired from the morphologically malignant squamous cells due to photodegradation using the 532 nm laser line. Indeed, further studies may investigate alternative laser lines to lessen irradiation induced damage. In addition, the characteristically small cytoplasm of lymphocytes, and the laser spot size of approximately 1 µm, prohibited the inclusion of representative spectra from the cytoplasm of lymphocytes without spectral contaminations from the nucleus. On AC cases, morphologically abnormal cells were not targeted as they could not be identified on the unstained slides. The primary objective was to comparatively analyse the spectra from negative cases and NSCLC cases, and evaluate the differences, if any, between them. Spectral differences were consistently observed between the negative cases and NSCLC cases with the most common signals occurring at ~430, ~780, 930, 1006, ~1106, 1175-200, ~1220-50, 1340, 1379, ~1450, 1575, ~1660, and 1686 cm⁻¹, representing significant differences in the molecular composition of cholesterol, nucleic acids, saccharides, lipids, protein structures and amides, across the negative and NSCLC cases. The spectral profiles taken from the morphologically normal and abnormal cells of NSCLC cases are commonly characterised in this study by an increase in collagen related peaks, glycogen, nucleic acids and amide III, and a decrease in amide I, showing that Raman spectroscopy can be used to detect malignancy in morphologically normal cells of SCC and AC cases. The peaks differentiating negative cases and NSCLC cases are in
accordance with several previous studies investigating lung tissue samples and cell lines with spectroscopic methods \textsuperscript{50–53}. It has been previously reported that elevated glycogen is a feature in lung cancers. A number of studies generated FTIR spectra from lung cancer cells and tissue and found a characteristic increase in glycogen associated peaks in comparison to the negative tissue \textsuperscript{50–52}. The increased peak intensities at 1022 cm\(^{-1}\) and 1048 cm\(^{-1}\) observed in the spectra taken from the cells of NSCLC cases may indicate that increased glycogen levels are detectable in Raman spectra, and could potentially be a diagnostic biomarker for NSCLC. Of particular interest, peaks relating to an elevated level of collagen were evident in both the nuclei and cytoplasm of the squamous and bronchial cells. An increase in collagen related peaks in the Raman spectra of lung cancer tissue was also reported by Kaminaka et al.\textsuperscript{53} in a study investigating the molecular differences between normal lung tissue and SCC using a 1064nm laser. Intracellular collagen is commonly found in collagen producing cells such as mesenchymal cells, yet it is not a common protein found in epithelial cells. Epithelial mesenchymal transition (EMT) is a process that results in the downregulation of epithelial genes and characteristics, and the upregulation of mesenchymal genes and characteristics \textsuperscript{54}. While this process is integral in development, pathological reactivation is associated with fibrosis and cancer progression. The EMT process induces the transition of epithelial cells to collagen producing cells and promotes cell invasion \textsuperscript{54,55}. Studies have indicated that EMT may be activated in lung cancer, which may explain the collagen related peaks observed in this study \textsuperscript{56,57}. Extracellular collagen is routinely observed in lung tissue as a response to lung damage, and in many lung cancer cases the patients also present with interstitial lung diseases, where pulmonary fibrosis is a common feature. Therefore, it needs to be established if utilising collagen peaks as spectral biomarkers for NSCLC could lead to potential false positive diagnosis in the presence of fibrosis. In addition, as
the assignments of the main Raman bands in this chapter were based on the biological peak assignments confirmed by other studies, further research is required to confirm that the collagen related peaks observed in this study are indeed related to intracellular collagen. The increased nucleic acid peak intensities in the NSCLC spectra is a feature reported in previous studies that investigated lung cancer with vibrational spectroscopic methods\textsuperscript{50,52}. Aneuploid DNA content is a feature in up to 85% of NSCLC tumours, which could explain the elevated peak intensity observed in malignant cases in this study\textsuperscript{58}.

The tumour related biochemical changes observed in the morphologically normal bronchial and squamous cells of NSCLC cases may be explained by the concept of field cancerization\textsuperscript{59}. First described by Slaughter et al\textsuperscript{60} in 1953, field cancerization suggests early genetic events in carcinogenesis may trigger expansion of the pre neoplastic cells in the tumour area or field. Further aberrations causes some of these cells to transform into the malignant phenotype, and are detectable with histopathological methods, however the remaining genetically altered cells are morphologically normal and are not identified\textsuperscript{61–63}. This concept has been described in a variety of cancers including lung\textsuperscript{64}, and could explain the high rate of recurrence for lung cancer.

Principal component analysis, an unsupervised classification method, was then applied to reduce the dimensionality of the data and analyse variance within the dataset. Applied to spectra acquired from nuclei, PCA was found to be an ineffective tool for discriminating the negative, SCC and AC cases. The PC scores plots indicate poor separation of the different groupings when analysing data from the distinct cell types separately, and when analysing combined data from different cell types. Similarly, PCA of the cytoplasmic data from the bronchial and squamous cells combined yielded a poor discriminatory performance, as the loadings spectra of the PCs responsible for the variation between the negative and malignant groups, and the SCC and AC groups, are similar to the PC
loadings spectrum discriminating the cytoplasm of bronchial cells from negative cases and the cytoplasm of squamous cells from negative cases. This indicates that the variation explained by PCA is likely the difference between the bronchial and squamous cells, and not the difference between different groups. However, applying PCA to the cytoplasmic data achieved good differentiation of the negative and NSCLC subtypes when analysing data from the bronchial and squamous cells separately. Consistent with the significant peaks identified in the difference spectra, both the squamous and bronchial cells from NSCLC cases were discriminated from their respective cell types from negative cases based on an increase in collagen, glycogen, phenylalanine, tyrosine, nucleic acids, and amide III, and a decrease in cholesterol, saccharides, phosphatidylinositol, and amide I. This suggests that when analysing the cytoplasm of bronchial and squamous cells with Raman spectroscopy, these may be effective diagnostic biomarkers for the identification of NSCLC. Further, applying PCA to spectra taken from the cytoplasm of bronchial cells, the SCC and AC cases were discriminated based on a variation in amide III / nucleic acid (1250 cm\(^{-1}\)), protein (~1330 cm\(^{-1}\)), lipid (1445 cm\(^{-1}\)), COO\(^{-}\) / amide II / tyrptophan / amide I (~1500-1620 cm\(^{-1}\)), and amide I (1656 cm\(^{-1}\)) content. These diagnostic peaks are in accordance with the significant peaks identified in the difference spectrum between the cytoplasm of bronchial cells from SCC and AC cases, and may be useful diagnostic markers for the differential diagnosis of SCC and AC. Based on the results of this study, applying PCA to cytoplasmic spectra is a more effective method for discriminating negative and NSCLC cases than applying PCA to nuclear spectra. The cytoplasm is a rich broth of complex biomolecules and most chemical reactions occur within the cell cytoplasm, this may account for the greater variation observed in the cytoplasm of cells from NSCLC cases using PCA.
As the visual differences between the spectra from cytology samples from negative, SCC and AC cases are discrete, partial least squares discriminant analysis (PLSDA), a supervised classification technique, was applied to the data to maximise the diagnostically relevant spectral variations. Analysing the bronchial nuclei, the SCC and AC subtypes were detected and discriminated with sensitivities >90% and specificities >86%, and examining the squamous nuclei, the SCC and AC subtypes were detected and discriminated with sensitivities >73% and specificities >78%. The model used to analyse the lymphocyte spectra achieved sensitivities and specificities above 82% for the identification of and discrimination between the SCC and AC subtypes.

PLSDA was then applied to investigate the ability to discriminate the bronchial nuclei from negative cases and the bronchial nuclei from the SCC and AC subtypes, all within the same model. The model yielded a sensitivity of 90% and specificity of 91% for identification of negative cases. The SCC and AC subtypes were classified with sensitivities of 82 and 81% respectively although multiple SCC and AC spectra were misclassified. The LVs distinguishing the negative cases from the NSCLC cases, and discriminating the SCC and AC cases matched the significant peaks identified in the difference spectra between the bronchial nuclei of negative, SCC, and AC cases. The LVs revealed diagnostic information about nucleic acids (788, 1223-1297 cm\(^{-1}\)), cholesterol / lipids / methionine (599-626, 700, 1379 cm\(^{-1}\)), proteins / saccharides (848 cm\(^{-1}\)), amide III (1223-1297 cm\(^{-1}\)), amide II (1470 cm\(^{-1}\)), COO\(^{-}\) (1585 cm\(^{-1}\)) and amide I (1597-1661, 1686 cm\(^{-1}\)). A PLSDA model was then developed to discriminate the squamous nuclei from negative cases and the squamous nuclei from SCC and AC cases, within in the same classifier. The negative cases were classified with a good diagnostic sensitivity of 91% and specificity of 91%. The SCC and AC cases were discriminated with sensitivities of 60-63% and specificities of 78-79%, as observed visually in the scatter plots of the first
three LVs, there was some overlap of the SCC and AC clusters which explains the lower performance scores of the model. A combination of the diagnostic peaks identified in the first two LVs match the significant differences revealed in the difference spectra. The negative cases were again differentiated from the NSCLC cases based on saccharide, protein, amide III, amide I, nucleic acid, and lipid content, whereas the diagnostically relevant peaks discriminating the SCC and AC cases were assigned to saccharides, nucleic acids, amide II, COO-, amide I, and proteins.

The diagnostic relevance of lymphocytes in NSCLC was demonstrated as the performance of a three-way PLSDA classifier showed sensitivities of 78-81% and specificities of 78-88% for the classification of negative, SCC and AC cases. The dissimilarities in the spectral profiles observed in this study between the lymphocyte nuclei of the two NSCLC subtypes may be explained by differing immune responses to tumour subtype. As part of the adaptive immune system, both B and T lymphocytes are involved in the immune anti-cancer response. Although B and T lymphocytes are morphologically indistinguishable, they have a variety of different functions in the immune response to cancer. When activated, B cells have the ability to differentiate into plasma cells and produce highly specific antibodies. Activated T cells comprise of helper T cells (CD4+), and cytotoxic T cells (CD8+), depending on the presence of a surface CD4 or CD8 T cell receptor. Helper T cells secrete cytokines which induce an inflammatory response to attract other immune cells. Cytotoxic T cells infiltrate the tumour and are capable of killing cancer cells. The differences in the spectra taken from lymphocytes of NSCLC cases and lymphocytes of negative cases may be spectral signatures for these immune biochemical responses to the presence of tumour. Although data is limited in this study, with further investigation the characteristic Raman peaks and
unique spectral profiles of lymphocytes may prove valuable for delineating NSCLC and elucidating the role of the immune system in lung cancer.

Notably, similar diagnostic information was provided in the LVs discriminating the lymphocyte nuclei of SCC cases from the lymphocyte nuclei of AC cases, and the bronchial nuclei of SCC cases from the bronchial nuclei of AC cases. The spectral profile of the LVs may represent a similar malignancy dependent change in the biochemical composition of bronchial cells and lymphocytes. However, it is also possible that abnormal glandular cells were incorrectly classified as lymphocytes upon microscopic examination of the stained slides. This may explain the overlap in spectral features between the bronchial nuclei and lymphocyte nuclei observed in this study. To clarify that the Raman peaks are indicative of changes detected in lymphocytes, future studies may consider identification of the cell type through immunohistochemical staining.

Analysing the cytoplasmic data with PLSDA revealed the diagnostic value of the cell cytoplasm for NSCLC. Objective comparative analysis of bronchial cytoplasm spectra was achieved by the application of a three-way PLSDA algorithm to the data. The model achieved sensitivities of 68-97% and specificities of 85-95% for the distinction of negative, SCC and AC subtypes. The LVs revealed diagnostic peaks in accordance with the PC loadings discriminating the cytoplasm of bronchial cells from negative, SCC and AC cases. Regardless of the statistical method used to classify the NSCLC subtypes, diagnostic peaks relating to nucleotides / phosphatidylinositol / glycerol (570-620 cm\(^{-1}\)), amide III / nucleic acids (1200-1280 cm\(^{-1}\)), saccharides / proline / valine (930 cm\(^{-1}\)), proteins (~1330 cm\(^{-1}\)), saccharides / proteins (~930 cm\(^{-1}\)), lipids (1445 cm\(^{-1}\)), and amide I (1656 cm\(^{-1}\)) are found to discriminate the cytoplasm of bronchial cells from negative, SCC and AC cases.
In addition, a PLSDA model applied to the spectral data acquired from the cytoplasm of squamous cells exhibited the ability to identify negative cases with a sensitivity and specificity of 88% and 94% respectively. The SCC and AC subtypes were discriminated with sensitivities and specificities ranging from 76-81%. In accordance with the PC loadings and difference spectra between the cytoplasm of squamous cells from negative, SCC and AC cases, the LVs extracted information about cholesterol, glycogen, saccharides, phosphatidylinositol, saccharides / proline / valine, phenylalanine, amide III / nucleic acids / lipids, amide II and amide I differentiating the negative cases and NSCLC subtypes.

Combining the datasets from different cell nuclei, the PLSDA algorithm classified the negative, malignant, SCC and AC groups with sensitivities and specificities of 69-89%. The LVs discriminating the combined groups retained many peaks found throughout the difference spectra and PC loadings associated with nucleic acids, proteins, amides, lipids, and saccharides. Similarly, combining the datasets from the cytoplasm of different cell types, the PLSDA algorithm classified the negative, malignant, SCC and AC groups with sensitivities and specificities of 71-90%. Despite the good performance of the model for differentiating the negative and malignant groups, the information revealed by the LV indicates that the spectral variations between the cell types are greater than the variations between negative and malignant spectral signatures. Combining spectra from different cell types for PLSDA does not affect the accuracy of disease diagnosis, however analysing the different cell types separately is a better approach to study the specific spectral alterations associated with each NSCLC subtype. The results achieved using Raman spectroscopy and PLSDA with venetian blinds cross-validation show that all cell types and cell components targeted in this study were found to be diagnostically significant. In addition, the PLSDA performance scores achieved in this study are similar
to the findings in other studies that explored the use of Raman spectroscopy to identify lung cancer using cells or tissue specimens. As previously mentioned, the discrimination of SCC and AC, along with molecular profiling of the tumour is now imperative to ensure efficacy of treatment. As the current diagnostic algorithm for lung cancer greatly reduces the amount of sample for molecular tests, the demand for diagnostic specimen is rising and a new marker free diagnostic technique would be largely beneficial to the healthcare system. The results of this study show that Raman spectroscopy may be utilized to improve the diagnostic algorithm of NSCLC. In our study good sensitivities (60-90%) and specificities (69-90%) were achieved for the differentiation of NSCLC subtypes using Raman spectroscopy on BAL and BW specimens prepared as ThinPrep® cytology slides.

Although the results identify promising diagnostic spectral biomarkers for NSCLC, this study is limited by several factors. The primary limitation of the study is the low number of patients. Biomarker discovery customarily requires a minimum of 10-100 patient samples. In addition, to validate the use of a spectral biomarker for NSCLC diagnosis, the performance characteristics must be well established with the backing of a wealth of evidence. To evaluate the performance measures of the diagnostic spectral biomarkers for clinical application, such as the sensitivity, specificity, reproducibility, and diagnostic accuracy must be determined with a larger cohort of approximately 200-1000 patients. Another limitation of the study is the low number of AC cases. In order to develop unbiased PLSDA classifiers, the number of AC, SCC, and negative cases had to be matched in each model. As a result, the number of SCC and negative cases were reduced to 5 when developing a PLSDA model with AC cases. As residual cytology samples were used for this study, the amount of cellular material left in the sample vials for analysis meant we were limited to a maximum of 15 spectra per cell type, per sample. For this
study this was adequate to develop well working PLSDA models as Beleites et al\textsuperscript{74} illustrate 75-100 spectra per class is required for a good classifier, however while the minimum required spectra were met for most classifications in this study, more patient samples are needed to be able to perform leave one patient out cross-validation. Further studies may also evaluate the use of alternative laser lines to reduce photo damage of morphologically abnormal squamous cells.

4.4.1 Summary

In summary Raman signals were acquired from 24 BAL and BW samples, representing 7 negative, 12 SCC, and 5 AC cases. High quality Raman spectra were obtained from the nuclei of bronchial cells, squamous cells, morphologically abnormal squamous cells, and lymphocytes, and the cytoplasm of bronchial and squamous cells. Mean spectral analysis revealed distinctive spectral variations in each cell component and each cell type across the negative and NSCLC cases. The consistent spectral differences between the negative, SCC and AC cases, associated with collagen, glycogen, and amide I and III, could prove to be diagnostically significant, providing insight into the carcinogenic process of NSCLC. These results also indicate that Raman spectroscopy can detect malignant signatures in morphologically normal cells. Of particular interest, the spectra from lymphocytes also showed differences in malignant and benign cases. Future studies may explore the use of Raman spectroscopy to further understand the role and diagnostic relevance of lymphocytes in NSCLC.

The PLSDA algorithm was successfully applied to the data and yielded diagnostic sensitivities and specificities of 60-97\% for the discrimination of negative, SCC, and AC cases using the spectra acquired from the nuclei and cytoplasms of morphologically normal bronchial and squamous cells. This indicates that both cell types and both cell components are diagnostically significant. Applying PLSDA to spectra acquired from
lymphocyte nuclei also yielded high performance scores for the differentiation of negative, SCC and AC cases. The spectral differences influencing the classification may be related to differences in immune response to SCC and AC.

The combination of Raman spectroscopy and PLSDA was an effective tool for the classification of NSCLC, using bronchoscopy attained BAL and BW specimens. Although data is limited, these results demonstrate the ability of Raman spectroscopy to effectively detect malignancy on ThinPrep® cytology samples. Applying this technique to cytology samples may present as a minimally invasive method for identifying and discriminating NSCLC subtypes while retaining the sample for ancillary molecular tests.

In some NSCLC cases, the diagnosis reported on cytology differs from the histological diagnosis. Therefore, to build on the findings of this chapter, the next chapter will investigate the accuracy of Raman spectroscopy on cytology in comparison to the known histological result.

References


164


Chapter 5: Examination of Raman spectra from NSCLC cases with cytological analyses reported as diagnostic or no malignant cells seen

5.1 Introduction

Lung cancer diagnostics are predominantly based on cytology or small biopsy specimens. As morphological analysis remains the primary method for lung cancer diagnosis, the overlapping cytomorphologic features of lung cancers, low cellularity of diagnostic specimens, and inter-observer variability are some of the reported limitations for current cytological analysis\(^1\).

To evaluate the diagnostic accuracy of cytological procedures for lung cancer, Sakr et al\(^2\) conducted a retrospective chart review of patients that underwent fibreoptic bronchoscopy for suspicion of lung cancer, and had both cytological and histological methods performed. The authors examined the diagnostic agreement between cytological and histological results and found that only 30% of cases classified as non-malignant on cytology were indeed false negatives. In addition, false positives were confirmed in only 12% of cases. Further, 60% of negative cases on cytology were deemed true negatives, and 52% of positive cases on cytology were considered true positives.

As a label free vibrational spectroscopic technique with the ability to provide comprehensive detail about intracellular biochemistry, Raman spectroscopy could potentially be an alternative or adjunct tool for the objective measurement of cytological samples. Raman spectroscopy can detect malignancy associated changes in morphologically normal cells, and has been shown to be a highly sensitive and specific tool for detecting lung malignancies\(^3\)\(^{-10}\).

There is a clinical need for a technology that improves the diagnostic accuracy of cytological analysis for lung cancer. The purpose of this study was to build on the findings of chapter 4 by investigating the feasibility of utilising Raman spectroscopy to detect
malignant signatures in cells from NSCLC cases with NMCS on cytology using bronchoscopy attained cytology samples, and assess the accuracy of Raman spectroscopy on cytology in comparison to the known histological result.

5.2 Methods

5.2.1 Study subjects and sample collection

As described in chapter 3 section 3.1.

This study utilised BALs and BWs collected from 19 patients (8 females, 11 males). Lung cancer was confirmed though final clinical diagnosis and final histology was recorded as: 9 squamous cell carcinoma, 4 adenocarcinoma, and 6 negative (no malignancy seen).

The samples were grouped as depicted in Table 5.1. Spectra taken from samples 1-6 represent true negatives, spectra from samples 7-11 represent SCC on cytology and SCC on histology, samples 12-15 represent NMCS on cytology and SCC on histology, samples 16 and 17 are AC on cytology and AC on histology, and samples 18 and 19 are NMCS on cytology and AC on histology. For this chapter, where cytology and histology correlate, these cases are termed true positive (TP-SCC / TP-AC) and true negative (TN) as appropriate. In cases where cytology was reported as negative and histology was positive, these are termed false negative (FN-SCC / FN-AC).

Table 5.1. Patient samples with the corresponding cytology and biopsy reports, as characterised by a pathologist.

<table>
<thead>
<tr>
<th>Sample number</th>
<th>Gender</th>
<th>Cytology report</th>
<th>Biopsy report</th>
<th>Termed</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>M</td>
<td>NMCS</td>
<td>Negative</td>
<td>TN</td>
</tr>
<tr>
<td>2</td>
<td>F</td>
<td>NMCS</td>
<td>Negative</td>
<td>TN</td>
</tr>
<tr>
<td>3</td>
<td>M</td>
<td>NMCS</td>
<td>Negative</td>
<td>TN</td>
</tr>
<tr>
<td>4</td>
<td>M</td>
<td>NMCS</td>
<td>Negative</td>
<td>TN</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>---</td>
<td>---</td>
<td>---</td>
<td>---</td>
<td>---</td>
</tr>
<tr>
<td>5</td>
<td>M</td>
<td>NMCS</td>
<td>Negative</td>
<td>TN</td>
</tr>
<tr>
<td>6</td>
<td>M</td>
<td>NMCS</td>
<td>Negative</td>
<td>TN</td>
</tr>
<tr>
<td>7</td>
<td>F</td>
<td>SCC</td>
<td>SCC</td>
<td>TP-SCC</td>
</tr>
<tr>
<td>8</td>
<td>F</td>
<td>SCC</td>
<td>SCC</td>
<td>TP-SCC</td>
</tr>
<tr>
<td>9</td>
<td>M</td>
<td>SCC</td>
<td>SCC</td>
<td>TP-SCC</td>
</tr>
<tr>
<td>10</td>
<td>F</td>
<td>SCC</td>
<td>SCC</td>
<td>TP-SCC</td>
</tr>
<tr>
<td>11</td>
<td>M</td>
<td>SCC</td>
<td>SCC</td>
<td>TP-SCC</td>
</tr>
<tr>
<td>12</td>
<td>M</td>
<td>NMCS</td>
<td>SCC</td>
<td>FN-SCC</td>
</tr>
<tr>
<td>13</td>
<td>F</td>
<td>NMCS</td>
<td>SCC</td>
<td>FN-SCC</td>
</tr>
<tr>
<td>14</td>
<td>M</td>
<td>NMCS</td>
<td>SCC</td>
<td>FN-SCC</td>
</tr>
<tr>
<td>15</td>
<td>M</td>
<td>NMCS</td>
<td>SCC</td>
<td>FN-SCC</td>
</tr>
<tr>
<td>16</td>
<td>F</td>
<td>AC</td>
<td>AC</td>
<td>TP-AC</td>
</tr>
<tr>
<td>17</td>
<td>F</td>
<td>AC</td>
<td>AC</td>
<td>TP-AC</td>
</tr>
<tr>
<td>18</td>
<td>F</td>
<td>NMCS</td>
<td>AC</td>
<td>FN-AC</td>
</tr>
<tr>
<td>19</td>
<td>M</td>
<td>NMCS</td>
<td>AC</td>
<td>FN-AC</td>
</tr>
</tbody>
</table>

*M, male; F, female.*

5.2.2 Sample processing and slide preparation

As described in chapter 3 section 3.1.1-2.

5.2.3 Raman Measurement

As described in chapter 3 section 3.2.1-2

For this study, spectra were taken from the nuclei of bronchial epithelial cells and squamous cells.

5.2.4 Data pre-processing and analysis

All spectral data analysis was conducted using MATLAB software. Data were process as described in chapter 3 section 3.2.3-.7.
The data was analysed using Principal Component Analysis (PCA), as described in chapter 3 section 3.2.8. Partial least squares discriminant analysis (PLS-DA) with venetian blind cross-validation was also employed to the mean centered Raman spectral data to identify the diagnostically significant variations between the groups, as described in chapter 3 section 3.2.10-12.

5.2.5 Pap staining
Following Raman spectroscopy, slides were stained using the Papanicolaou staining method as described in chapter 3 section 3.3. Using the XY co-ordinates recorded as described in chapter 3 section 3.2.2, the spectra were correlated with their cell type and categorised into separate databases according to their cell type.

5.3 Results
5.3.1 Comparative spectral analysis

5.3.1.1 Spectral analysis of bronchial and squamous nuclei from SCC cases with SCC on cytology, and SCC cases with NMCS on cytology

Figures 5.1 and 5.3 depict the mean bronchial and squamous cell spectra from TN cases, TP-SCC cases, and FN-SCC cases. The standard deviation shows that there was minimal variation between the spectra from individual cells and this was the case for all targeted cells. The features in the bronchial and squamous cell spectra from TP-SCC cases and FN-SCC cases that differ from the cells from TN cases are demonstrated by the difference spectra in Figures 5.2 and 5.4. The tentative peak assignments for the significant Raman bands identified by spectral variation analysis in this study are shown in Table 5.2. As depicted in Figure 5.2, the bronchial cells from all SCC cases show elevated peak intensities at ~780, 1300-1310, 1395 and 1421 cm\(^{-1}\), and decreased peak intensities at 440, ~840, 1622 and 1662 cm\(^{-1}\) in comparison to the bronchial cells from TN cases. This indicates that the bronchial cells from SCC cases had increased nucleic acids, lipids, and
decreased cholesterol, proteins and polysaccharides in comparison to the bronchial cells from TN cases.

In comparison to the mean negative bronchial cell spectrum in Figure 5.1, the bronchial cells from TP-SCC cases have more significant differences than the bronchial cells from FN-SCC cases. In addition to the aforementioned spectral differences, bronchial cells from TP-SCC cases have elevated peaks at 820, 924, 935, 1221-1270, 1290, 1329-1351, and 1674 cm\(^{-1}\), and decreased peak intensities at 470-491, 506-533, 556, 570, 577-636, 652, 690-727, 1501-1575, and 1584-1621 cm\(^{-1}\), in comparison to the bronchial cells from TN cases.

The difference spectra in Figure 5.4 show that the squamous cells from TP-SCC cases and FN-SCC cases have a different biochemistry than the squamous cells from TN cases. The squamous cells from TP-SCC cases and FN-SCC cases show an increase in nucleic acids (788, 1515, 1575 cm\(^{-1}\)), collagen (1203 cm\(^{-1}\)), and amide II (1535 cm\(^{-1}\)), and a decrease in cholesterol (433 cm\(^{-1}\)), protein / amide I (615-620, 1640 cm\(^{-1}\)), and methionine (~700 cm\(^{-1}\)) compared to the squamous cells from TN cases. Squamous cell nuclei from TP-SCC cases also had an increase in collagen (978 cm\(^{-1}\)), phenylalanine (989-1007 cm\(^{-1}\)), and amide III / nucleic acids (1215-1245 cm\(^{-1}\)), and a decrease in cholesterol (408, 420 cm\(^{-1}\)), polysaccharides and protein (470, 486-526 cm\(^{-1}\)), phosphatidylinositol (580 cm\(^{-1}\)), protein / glycerol (590-640 cm\(^{-1}\)), methionine (657 cm\(^{-1}\)), protein / nucleic acids (1320 cm\(^{-1}\)), and amide I (1621-1701 cm\(^{-1}\)), in comparison to squamous cells from TN cases.

From TP-SCC cases (Figures 5.2 and 5.4), the bronchial and squamous cells showed consistent peaks at 435, 470, 506-533, 595-621, 627-636, 700, 789, 1221-1245, and 1621-1662 cm\(^{-1}\) differentiating them from their respective cell types in TN cases.
On the FN-SCC cases (Figures 5.2 and 5.4), the bronchial and squamous cells had elevated peaks at 788, 1535, and 1575 cm\(^{-1}\), and lower peak intensities at 433, 837, 975, and 1640 cm\(^{-1}\), in comparison to their respective cell types in TN cases. Decreased peak intensities at 435 and \(~1640\) cm\(^{-1}\), and an increased peak intensity at \(~780-789\) cm\(^{-1}\) was consistent among the bronchial and squamous cells from all SCC cases.

Comparative analysis of the bronchial and squamous cells from TP-SCC cases, and FN-SCC cases revealed differences in their spectral profiles (Figures 5.2 and 5.4). The bronchial and squamous cells from TP-SCC cases indicated a higher level of proline / hydroxproline / glycogen (911-18 cm\(^{-1}\)), pheylalanine (\(~1004\) cm\(^{-1}\)) and collagen (1030-40 cm\(^{-1}\)) than FN-SCC cases. The TP-SCC cases also exhibited a decrease in tyrosine / methionine (638-40 cm\(^{-1}\)) and amide I (1651-93 cm\(^{-1}\)) content when compared to the FN-SCC cases.

Table 5.2. Tentative peak assignments for the prominent significant Raman bands identified by spectral variation analysis\(^{11,12}\).

<table>
<thead>
<tr>
<th>Wavenumber (cm(^{-1}))</th>
<th>Raman peak assignments</th>
</tr>
</thead>
<tbody>
<tr>
<td>(~420/30/40)</td>
<td>cholesterol</td>
</tr>
<tr>
<td>580</td>
<td>phosphatidylinositol</td>
</tr>
<tr>
<td>486-526</td>
<td>Polysaccharides, proteins</td>
</tr>
<tr>
<td>590</td>
<td>Protein/glycerol</td>
</tr>
<tr>
<td>640</td>
<td>Tyrosine/methionine</td>
</tr>
<tr>
<td>657/(~700/10)</td>
<td>methionine</td>
</tr>
<tr>
<td>(~780)</td>
<td>Nucleic acids</td>
</tr>
<tr>
<td>830-920</td>
<td>Polysaccharides/proline/valine/collagen</td>
</tr>
<tr>
<td>(~930)</td>
<td>Proline/valine/saccharides</td>
</tr>
<tr>
<td>1175-1200</td>
<td>Nucleic acids/phophates</td>
</tr>
<tr>
<td>1210-1280</td>
<td>Amide III/nucleic acids</td>
</tr>
<tr>
<td>(~1350)</td>
<td>Tyrptophan/guanine/saccharides</td>
</tr>
<tr>
<td>1420-50</td>
<td>Proteins/lipids</td>
</tr>
<tr>
<td>Wavenumber</td>
<td>Description</td>
</tr>
<tr>
<td>-------------</td>
<td>--------------------------------------------------</td>
</tr>
<tr>
<td>~1500-1600</td>
<td>Amide II/COO-/nucleic acids</td>
</tr>
<tr>
<td>1600-1660</td>
<td>Amide I</td>
</tr>
<tr>
<td>~1680</td>
<td>Cholesterol/amide I Amide I disordered structure/β sheets/lipids</td>
</tr>
<tr>
<td>~1740</td>
<td>lipids</td>
</tr>
</tbody>
</table>

Figure 5.1. The mean Raman spectra (± standard deviation) taken from the nuclei of bronchial epithelial cells from negative cases (black, $n=70$), TP-SCC cases (blue, $n=93$), and FN-SCC cases (red, $n=48$).
Figure 5.2. Difference spectrum between the mean Raman spectrum of bronchial epithelial cell nuclei taken from negative cases and (a) the mean Raman spectrum of bronchial epithelial cell nuclei from TP-SCC cases (blue, n=93), (b) the mean Raman spectrum of bronchial epithelial cell nuclei from FN-SCC cases (red, n=48). (c) Difference spectrum between the mean Raman spectrum of bronchial epithelial cell nuclei from TP-SCC cases and the mean Raman spectrum of bronchial epithelial cell nuclei from FN-SCC cases (green). Shading indicates regions of the spectrum that were significantly different (P < 0.01)
Figure 5.3. The mean Raman spectra (± standard deviation) taken from the nuclei of squamous cells from negative cases (black, n=94), TP-SCC cases (blue, n=137), and FN-SCC cases (red, n=35).
Figure 5.4. Difference spectrum between the mean Raman spectrum of squamous cell nuclei taken from negative cases and (a) the mean Raman spectrum of squamous cell nuclei from TP-SCC cases (blue, $n=137$), (b) the mean Raman spectrum of squamous cell nuclei from FN-SCC cases (red, $n=35$). (c) Difference spectrum between the mean Raman spectrum of squamous cell nuclei from TP-SCC cases and the mean Raman spectrum of squamous cell nuclei from FN-SCC cases (green). Shading indicates regions of the spectrum that were significantly different ($P < 0.01$).
5.3.1.2 Spectral analysis of bronchial and squamous nuclei from AC cases with AC on cytology and AC cases with NMCS on cytology

Figures 5.5 and 5.7 depict the mean bronchial and squamous cell spectra from TN cases, TP-AC cases, and FN-AC cases. The standard deviation shows that there was minimal variation between the spectra from individual cells and this was the case for all targeted cells. The features in the bronchial and squamous cell spectra from TP-AC cases and FN-AC cases that differ from the cells from TN cases are demonstrated by the difference spectra in Figures 5.6 and 5.8.

As depicted in Figure 5.6, the bronchial cells from the AC cases showed elevated peak intensities at ~775, 1179-1182, 1188-1202, and 1291-94 cm\(^{-1}\), and decreased peak intensities at 1652-61, and 1677-83 cm\(^{-1}\) in comparison to the bronchial cells from TN cases. This indicates that the bronchial cells from AC cases had increased nucleic acids (775, 1179-82 cm\(^{-1}\)), phosphates (1188-202), and cytosine / methylene (1291-94 cm\(^{-1}\)) and decreased amide I (1652-61, 1677-82 cm\(^{-1}\)) in comparison to the bronchial cells from TN cases. In addition to the aforementioned spectral differences evident in the bronchial cells from AC cases, bronchial cells from TP-AC cases have elevated peaks at 830-921, 936-947, 978-1001, 1012-1054, ~1215, 1220-1268, ~1276, ~1283, and 1684-1696 cm\(^{-1}\), and decreased peak intensities at 425-432, 503-514, 524-529, 567,628, 701-704, 1402-1405, 1542-1552, 1558-1576, and 1583-1650 cm\(^{-1}\), in comparison to the bronchial cells from TN cases. These significant peaks show that in AC cases where malignancy is evident on cytology, the bronchial cells have a higher level of collagen related peaks (830-91, 936-47, 1012-54, 1283 cm\(^{-1}\)), glycogen (1012-54 cm\(^{-1}\)), pheylalanine (978-1001 cm\(^{-1}\)), proline/hydroxyproline / saccharides (830-91 cm\(^{-1}\)), amide III / nucleic acids (1215, 1220-68, 1276 cm\(^{-1}\)) and non-hydrogen bonded amide I (1684-96 cm\(^{-1}\)) than the bronchial cells of TN cases. The profile of the difference spectrum also indicates a decrease in cholesterol (425-32, 701-04 cm\(^{-1}\)), amide II (1542-52 cm\(^{-1}\)), phosphatidylinositol /
phosphatidylserine / protein (503-29 cm⁻¹), glycerol (630 cm⁻¹), COO⁻ (1558-76 cm⁻¹) and amino acid (567, 1402-05 cm⁻¹) content in the TP-AC cases.

The bronchial cells from FN-AC cases exhibited an increase in peaks associated with thiocyanate (446 cm⁻¹), phosphatidylserine (~524 cm⁻¹), cholesterol (539-54 cm⁻¹), nucleic acids (695, 720, 1421-25, 1429-38 cm⁻¹), methionine (708 cm⁻¹), phosphate ions (804-07 cm⁻¹), collagen (817-821 cm⁻¹), triglycerides (1068-71 cm⁻¹), saccharides (1112-1116 cm⁻¹), glycogen / collagen (1151-1167 cm⁻¹), aspartic / glutamic acid (1404-1416 cm⁻¹), lipid / protein (1429-1438 cm⁻¹), and amide II 1516-49 cm⁻¹ in comparison to the bronchial cells from TN cases. Bronchial cells from FN-AC cases also showed decreased peak intensities at 1004, 1220-25, 1320-33, and 1466-73 cm⁻¹, in comparison to the bronchial cells from TN cases, corresponding to a lower level of phenylalanine, guanine, amide III, amide II, phospholipids, and lipids.

The difference spectra in Figure 5.8 show that the squamous cells from AC cases have an increase in collagen / nucleic acids (1169-84cm⁻¹) and phosphates (1194-97cm⁻¹), and a decrease in fatty acids (1130-35 cm⁻¹) compared to the squamous cells from TN cases. The squamous cells from TP-AC cases also had an increase in nucleic acids (775, 966-70, 1217-49 cm⁻¹), hydroxyproline / tryptohpan (878 cm⁻¹), lipid / phosphorylated protein (966-70 cm⁻¹), amide III (1217-1249 ,1257-60 cm⁻¹), and lipids (1715-18cm⁻¹) in comparison to the squamous cells from TN cases. A decrease in phosphatidylinositol / phosphatidylserine (501-525 cm⁻¹), phenylalanine (621 cm⁻¹), lipids (1438-51 cm⁻¹), lipids / disaccharides (1459-62 cm⁻¹), and protein (1603-1642 cm⁻¹) was also evident in the squamous cells from TP-AC cases in comparison to the squamous cells from TN cases. The squamous cells from FN-AC cases exhibited increased peak intensities at 425-28, 439-44, 474-77, 551-54, 710-715, 1104-08, 1118-1120, 1139-1164, 1424-29, 1490-95, 1501-1623, and 1738-49 cm⁻¹, and decreased peak intensities at 494-98, 648, 835-53,
859-71, 902, 952, 970, 1004, 1211-17, 1233-36, 1249-60, 1320, and 1651-1702 cm⁻¹, in comparison to the squamous cells from TN cases. This corresponds to a higher level of cholesterol and thiocyanate (425-44 cm⁻¹), polysaccharides (474-77 cm⁻¹), glycogen (551-54, 1139-64 cm⁻¹), methioine (710-15 cm⁻¹), carotene (1139-64 cm⁻¹), amide II / nucleic acids (1490-95, ~1501-1600 cm⁻¹) and lipids (1738-49 cm⁻¹) and a lower level of tyrosine (646, 859-71 cm⁻¹), saccharides (835-53, 902 cm⁻¹), phenylalanine (1004 cm⁻¹), amide III / nucleic acids (1211-1260 cm⁻¹), guanine (1320 cm⁻¹) and amide I (1651-1702 cm⁻¹) in the FN-AC cases.

From cases where AC was evident on cytology (Figures 5.6 and 5.8), the bronchial and squamous cells showed consistent peaks at 504, 513, 525, 597-604, 624-26, 770-76, 878, 1016, 1179-82, 1194-98, 1220-49, 1257-60, and 1603-42 cm⁻¹ differentiating them from their respective cell types in TN cases. This indicated a decrease in the level of phosphatidylinositol, phosphatidylserine, carbohydrates (1016 cm⁻¹), and amide I, and an increase in nucleic acids, collagen, phosphates, and amide III in the bronchial and squamous cells of TP-AC cases.

The bronchial and squamous cells from FN-AC cases (Figures 5.6 and 5.8), had increased thiocyanate, methionine, glycogen, collagen, phospholipids, deoxyribose, and amide II, and a decreased phenylalanine, amide III (β sheet structure), guanine, and amide I in comparison to their respective cell types in TN cases, exhibited by the elevated peaks at 446, 710, 1151-1164, 1175-97, ~1425, and 1516-1549 cm⁻¹, and lower peak intensities at 1004, 1224, 1319- 1333, and 1651-83 cm⁻¹.

The variations in spectral profiles of bronchial and squamous cells from TP-AC cases and FN-AC cases indicate different biochemical profiles of cells from AC cases with different cytology reports (Figures 5.6 and 5.8). Cells from the TP-AC cases have an increase in proline/hydroxyproline / glycogen / collagen (829-73, 877-79, 882-86 cm⁻¹),
phenylalanine (~1004 cm\(^{-1}\)), amide III / nucleic acids / phospholipids (1216-62, 1268-72 cm\(^{-1}\)) and a region of the amide I peak (1662-1696 cm\(^{-1}\)) in comparison to the cells in FN-AC cases. The bronchial and squamous cells from TP-AC cases also show a decrease in Raman peaks associated with phosphatidylinositol (514-17 cm\(^{-1}\)), phosphatidylserine (524-30 cm\(^{-1}\)), cholesterol (609-23 cm\(^{-1}\)), lipids (1393-98 cm\(^{-1}\)), proteins / lipids (1434-38 cm\(^{-1}\)), and proteins / amide I / II / nucleic acids (1500-1642 cm\(^{-1}\)), in comparison to the FN-AC cases.

Figure 5.5. The mean Raman spectra (± standard deviation) taken from the nuclei of bronchial epithelial cells from negative cases (black, \(n=70\)), TP-AC cases (blue, \(n=25\)), and FN-AC cases (red, \(n=21\)).
Figure 5.6. Difference spectrum between the mean Raman spectrum of bronchial epithelial cell nuclei taken from negative cases and (a) the mean Raman spectrum of bronchial epithelial cell nuclei from TP-AC cases (blue, n=25), (b) the mean Raman spectrum of bronchial epithelial cell nuclei from FN-AC cases (red, n=21). (c) Difference spectrum between the mean Raman spectrum of bronchial epithelial cell nuclei from TP-AC cases and the mean Raman spectrum of bronchial epithelial cell nuclei from FN-AC cases (green). Shading indicates regions of the spectrum that were significantly different (P < 0.01)
Figure 5.7. The mean Raman spectra (± standard deviation) taken from the nuclei of squamous cells from negative cases (black, $n=94$), TP-AC cases (blue, $n=25$), and FN-AC cases (red, $n=20$).
Figure 5.8. Difference spectrum between the mean Raman spectrum of squamous cell nuclei taken from negative cases and (a) the mean Raman spectrum of squamous cell nuclei from TP-AC cases (blue, $n=25$), (b) the mean Raman spectrum of squamous cell nuclei from FN-AC cases (red, $n=20$). (c) Difference spectrum between the mean Raman spectrum of squamous cell nuclei from TP-AC cases and the mean Raman spectrum of squamous cell nuclei from FN-AC cases (green). Shading indicates regions of the spectrum that were significantly different ($P < 0.01$).
5.3.2 Principal component analysis

5.3.2.1 PCA of bronchial and squamous nuclei from negative cases, NSCLC cases with diagnostic cytology and NSCLC cases with NMCS on cytology

Principal component analysis was then applied to reduce the data and extract diagnostic information. Figure 5.9 shows the scatter plot of the PC scores along the first two principal components, demonstrating separation of the Raman spectra from bronchial cells from TN cases and bronchial cells from TP-SCC cases, with some overlap. The spectra from TN cases and TP-SCC cases form two adjacent, overlapping clusters along the first PC. The loadings of PC1 indicate variation in tyrosine (650 cm\(^{-1}\)), polysaccharides / proline / valine (850 cm\(^{-1}\)), saccharides / proline / valine (925 cm\(^{-1}\)), phenylalanine (1004 cm\(^{-1}\)), amide III / nucleic acids (1210-1280 cm\(^{-1}\)), tryptophan / guanine / cytosine (~1350 cm\(^{-1}\)), nucleic acids / protein / lipids (1450 cm\(^{-1}\)), and cholesterol / amide I (β sheet non-hydrogen bonded) (1680 cm\(^{-1}\)) on the positive side, and DNA (780 cm\(^{-1}\)) and amide II / I / tryptophan / tyrosine (1500-1645 cm\(^{-1}\)) on the negative side. Spectra taken from bronchial cells of FN-SCC cases were distributed evenly along PC1 and PC2 and were not discriminated using PCA.
Figure 5.9. A PCA scatter plot of bronchial cells from negative cases, TP-SCC cases, and FN-SCC cases. The spectra from TN cases and TP-SCC cases form two adjacent, overlapping clusters along the first PC. The spectra taken from bronchial cells of FN-SCC cases were distributed evenly along PC1 and PC2 and could not be discriminated using PCA. B On the positive side of the first principal component peaks at 650, 850, 925, 1004, 1210-1280 cm$^{-1}$, ~1350, 1450, and 1680 cm$^{-1}$ can be distinguished, while peaks at and 780 cm$^{-1}$ and 1500-1645 cm$^{-1}$ can be distinguished on the negative side.

Figure 5.10 shows the scatter plot along the first two PCs of spectra taken from squamous cells of TN cases, squamous cells of TP-SCC cases, and FN-SCC cases. The spectra were not distinguishable using a combination of the first five PCs, which explained 95% of the variance in the dataset.
Figure 5.10. PCA scatter plot of spectra taken from squamous cells from TN cases, TP-SCC cases, and FN-SCC cases shows no separation with even distribution along the first two principal components.

In Figure 5.11 the distribution of PC1 and PC2 was plotted to visualise the variances between spectra taken from bronchial cells of TN cases, TP-AC cases, and FN-AC cases. The TP-AC and FN-AC spectra are separated along PC1. Due to the even distribution of the spectra taken from TN cases along PC1 and PC2, neither the TP-AC nor the FN-AC clusters can be distinguished from TN cases. On the positive side of the first principal component, which is responsible for discriminating the TP-AC and FN-AC clusters, peaks at 650, 850, 1210-1280, 1350, 1420-1500, 1680 cm⁻¹ can be distinguished. While peaks at 780, saccharides (1105 cm⁻¹), and 1500-1650 cm⁻¹ can be distinguished on the negative side.
Figure 5.11. A PCA scatter plot of bronchial cells from TN cases, TP-AC cases, and FN-AC cases. The spectra from TP-AC cases and FN-AC cases form two distinct clusters along the first principal component. The spectra from negative cases are dispersed along the first two PCs and cannot be discriminated. B. On the positive side of the first principal component, peaks at 650, 850, 1210-1280, 1350, 1420-50, 1680 cm\(^{-1}\) can be distinguished, while peaks at 780, 1105, 1500-1650 cm\(^{-1}\) can be distinguished on the negative side.

Figure 5.12 demonstrates the PCA scatter plot of spectra taken from the squamous cells of TN cases, TP-AC cases, and FN-AC cases. The spectra from TP-AC cases and FN-AC cases form two distinct clusters along PC1. Due to the even distribution of the spectra
taken from TN cases along PC1 and PC2, the negative cases cannot be discriminated. The loadings of PC1 indicate variation in 650 cm\(^{-1}\), 850, phenylalanine 1004 cm\(^{-1}\), 1210-1270 cm\(^{-1}\), \(~\)1360, 1420-50, and \(~\)1680 cm\(^{-1}\) on the positive side, and 780 cm\(^{-1}\) and 1480-1655 cm\(^{-1}\) on the negative side.

Figure 5.12. A PCA scatter plot of squamous cells from TN cases, TP-AC cases, and FN-AC cases. The spectra from TP-AC cases and FN-AC cases form two distinct clusters along the first principal component. The spectra from TN cases are dispersed along the first two PCs and cannot be discriminated. B. On the positive side of the first principal component, peaks at 650, 850, 1004 cm\(^{-1}\), 1210-1270, \(~\)1360, 1420-50, and \(~\)1680 cm\(^{-1}\) can be distinguished, while peaks at 780 cm\(^{-1}\) and 1480-1655 cm\(^{-1}\) can be distinguished on the negative side.
5.3.3 PLSDA classification

The PLSDA classification model was applied to the data to examine whether Raman spectroscopy could detect malignancy in the NSCLC cases with NMCS on cytology, and cases where cancer was evident by cytological analysis. The number of spectra in each class was matched. The required number of latent variables in each model was used to set the cross validation classification error average at 5%. To ensure discrimination was based on the detection of malignant signatures in the spectra and not differences in cell type, the bronchial and squamous cell data were analysed separately.

Table 5.3 shows the results of two separate PLSDA models in differentiating spectra taken from the bronchial nuclei of TN cases, and the bronchial nuclei of TP-SCC cases or FN-SCC cases. The first 3 latent variables were used to develop each model. The spectra taken from bronchial nuclei of TP-SCC cases were discriminated from the bronchial nuclei of TN cases with a sensitivity and specificity above 91%. The spectra taken from bronchial nuclei of FN-SCC cases were discriminated from the bronchial nuclei of TN cases with a sensitivity of 73% and a specificity of 86%.

Table 5.4 shows the performance scores of two PLSDA classifiers in discriminating spectra taken from the squamous nuclei of TN cases, and spectra taken from the squamous nuclei of TP-SCC cases or FN-SCC cases. Each two-way model used two latent variables each. The squamous nuclei of TP-SCC cases were distinguished from the squamous nuclei of TN cases with a sensitivity and specificity of 87% and 78% respectively. The squamous nuclei of FN-SCC cases were identified with a sensitivity of 84% and a specificity of 95%.

Table 5.5 shows the results of the two-way PLSDA classification models used for the discrimination of spectra taken from the bronchial nuclei of TN cases, and the bronchial nuclei of TP-AC, or FN-AC cases. Each model was developed using four latent variables.
Each two-way classification model yielded sensitivities above 86% and specificities above 81%.

Table 5.6 shows the results of two separate PLSDA models, each using the first 2 latent variables, in differentiating spectra taken from the squamous nuclei of TN cases, and the squamous nuclei of TP-AC cases or FN-AC cases. The spectra taken from squamous nuclei of TP-AC cases were discriminated from the squamous nuclei of TN cases with a sensitivity and specificity of 73% and 92% respectively. The spectra taken from squamous nuclei of FN-AC cases were discriminated from the squamous nuclei of TN cases with a sensitivity and specificity above 91%.

Based on a compilation of the sensitivity ranges achieved in Tables 5.3-5.6, Table 5.7 shows the accuracy of Raman spectroscopy on cytology in comparison to the known histological result.

Table 5.3. Performance of PLSDA in discriminating bronchial nuclei of TN cases from bronchial nuclei of TP-SCC cases and FN-SCC cases. The first 3 LVs were used in each case.

<table>
<thead>
<tr>
<th>SCC cases distinguished from TN cases</th>
<th>Sensitivity</th>
<th>Specificity</th>
</tr>
</thead>
<tbody>
<tr>
<td>TP-SCC cases</td>
<td>91</td>
<td>92</td>
</tr>
<tr>
<td>FN-SCC cases</td>
<td>73</td>
<td>86</td>
</tr>
</tbody>
</table>

Table 5.4. Performance of PLSDA in discriminating squamous nuclei of TN cases from squamous nuclei of TP-SCC cases and FN-SCC cases. The first 2 LVs were used in each case.

<table>
<thead>
<tr>
<th>SCC cases distinguished from TN cases</th>
<th>Sensitivity</th>
<th>Specificity</th>
</tr>
</thead>
<tbody>
<tr>
<td>TP-SCC cases</td>
<td>87</td>
<td>78</td>
</tr>
</tbody>
</table>
Table 5.5. Performance of PLSDA in discriminating bronchial nuclei of TN cases from bronchial nuclei of TP-AC cases and FN-AC cases. The first 4 LVs were used in each case.

<table>
<thead>
<tr>
<th>AC cases distinguished from TN cases</th>
<th>Sensitivity</th>
<th>Specificity</th>
</tr>
</thead>
<tbody>
<tr>
<td>TP-AC cases</td>
<td>86</td>
<td>92</td>
</tr>
<tr>
<td>FN-AC cases</td>
<td>91</td>
<td>81</td>
</tr>
</tbody>
</table>

Table 5.6. Performance of PLSDA in discriminating squamous nuclei of TN cases from squamous nuclei of TP-AC cases and FN-AC cases. The first 2 LVs were used in each case.

<table>
<thead>
<tr>
<th>AC cases distinguished from TN cases</th>
<th>Sensitivity</th>
<th>Specificity</th>
</tr>
</thead>
<tbody>
<tr>
<td>TP-AC cases</td>
<td>73</td>
<td>92</td>
</tr>
<tr>
<td>FN-AC cases</td>
<td>91</td>
<td>100</td>
</tr>
</tbody>
</table>

Table 5.7. Cytology reports with their corresponding biopsy reports and ranges of sensitivity for malignancy using Raman spectroscopy.

<table>
<thead>
<tr>
<th>Cytology report</th>
<th>Biopsy report</th>
<th>Sensitivity for malignancy</th>
</tr>
</thead>
<tbody>
<tr>
<td>SCC</td>
<td>SCC</td>
<td>87-91%</td>
</tr>
<tr>
<td>NMCS</td>
<td>SCC</td>
<td>73-84%</td>
</tr>
<tr>
<td>AC</td>
<td>AC</td>
<td>73-86%</td>
</tr>
<tr>
<td>NMCS</td>
<td>AC</td>
<td>91%</td>
</tr>
</tbody>
</table>
5.4 Discussion

Though limited material on cytology samples and morphological challenges add to the difficulty of diagnostic cytology, it is still one of the first line diagnostic specimen types due to being less invasive and having fewer risks than histology sampling. Objective technology such as Raman spectroscopy may identify underlying disease in cytology samples and improve diagnostic accuracy on cytology. The companion histological reports allowed us to group the cytology spectra according to their final diagnosis and determine the ability of Raman spectroscopy to accurately identify underlying malignancy. The results showed that using PLSDA, the Raman spectra of bronchial and squamous cells could be used to effectively discriminate between TN and NSCLC histology confirmed cases with negative cytology with sensitivities from 73-91%, and specificities of 81-100%. Comparable performance scores were achieved for discriminating the TN cases from the NSCLC histology confirmed cases with TP cytology. These findings show that both the squamous and bronchial cell types are diagnostically significant in the SCC and AC subtypes, and that the algorithm can classify spectral changes in the cells from malignant cases with FN cytology. Field cancerization is a concept which may explain the good discrimination of spectra from TN cases and NSCLC cases with FN cytology. Field cancerization is when the genetically altered cells in a tumour area or field remain morphologically normal and are not identified\textsuperscript{13–17}. However, these malignant associated changes may be detectable in the Raman spectra. The performance of PLSDA in classifying the NSCLC cases with FN cytology is in accordance with the sensitivities and specificities achieved for discriminating the NSCLC subtypes in chapter 4. This suggests that phenotypic features of malignancy are not required to identify NSCLC with high diagnostic accuracy using Raman spectroscopy on bronchoscopy cytology samples. The ability of biospectroscopy to detect underlying
malignancy was also demonstrated in a study by Gajjar et al\textsuperscript{18}. Infrared (IR) spectroscopy was used to determine the effectiveness of cytological methods in identifying cervical malignancies using the companion histology results as gold standard for comparison. Analysis of the spectra revealed underlying pre-invasive cervical disease in cases with negative cytology. Our study as well as the study by Gajjar et al\textsuperscript{18} show that objective spectroscopic techniques have the ability to improve the diagnostic efficiency of cytological methods and may overcome the current limitations of using cytological analysis for lung cancer diagnosis.

Comprehensive analysis of the spectra shows that malignant signatures can be detected in morphologically normal cells in NSCLC cases. Further, there is a difference in the biochemical composition of cells from NSCLC cases with different cytological reports. The prominent differences, identified using mean spectral analysis and PCA, between the TN cases and NSCLC cases are evident in spectral regions associated with saccharides, lipids, collagen, nucleic acids, amide I / III, glycogen, protein and thiocyanate. After examination of the spectral data, a reduced phenylalanine peak at 1004 cm\textsuperscript{-1}, and decreased amide I content is a recurring feature in NSCLC cases with FN cytology. These characteristic peaks could prove to be effective for detecting NSCLC in cases limited by the current diagnostic algorithm. With SCC and AC cases, PCA and direct comparison of the spectra taken from cases with TP cytology and FN cytology indicated that cases with malignancy evident on cytology were associated with an increase in phenylalanine, collagen, and glycogen, and a decrease in amide I, in comparison to the cases with FN cytology. The spectra indicate that as cytomorphology progresses to the malignant phenotype, the glycogen and collagen content in the cells is elevated, and amide I is reduced, suggesting that alterations either in the quantity or content of these biomolecules is a feature of NSCLC in bronchoscopy cytology samples. The significant peaks
identified in this chapter are similar to the diagnostic peaks discriminating the negative, SCC and AC cases in chapter 4. This suggests that the differentiation of the negative cases and NSCLC subtypes in chapter 4 was based on a combination of peaks from cases with diagnostic cytology and cases with FN cytology.

Importantly, although significant diagnostic Raman peaks were identified in NSCLC cases with FN cytology, the biological assignments of the main Raman bands were based on the findings of other studies\textsuperscript{11,12} and were not experimentally confirmed in this study. Presently, the cytomorphologic diagnosis of lung cancer leads to highly varied results. In particular, the identification and subtyping of NSCLC through cytological analysis generally yields poor accuracy\textsuperscript{1,2,19}. The reporting of false positives and false negatives based on morphology can have devastating effects on patients and can result in delayed treatment. This study demonstrates that Raman spectroscopy may be an effective tool for detecting lung cancer on morphologically normal cytology samples, as detailed analysis of the Raman spectra revealed malignant signatures in the cases with FN cytology. Adjunct application of this technique may reduce the rate of false negative diagnoses and improve the accuracy of cytology reports, minimising the need for invasive biopsy procedures. Future studies could also evaluate the ability to detect false positives with Raman spectroscopy on cytology, as false positives can lead to psychological stress for patients.

This study has some limitations, similar to chapter 4. The main limitations of the study are the low number of patients and the low amount of cellular material left in the sample vials for analysis. As a result, the size of the spectral dataset is small. As the dataset is limited in this study, venetian-blinds cross validation was used to validate PLSDA. Thus, the performance characteristics of the PLSDA classifier must be determined with a larger cohort of patients using leave one patient out cross validation. In addition, an inherent
problem when determining the accuracy of a new diagnostic technique is the limitations of the current gold standard, to which the new method is being compared. In this study, the classifications are based on histological reports, which are subject to inter-observer variability.

Despite the limitation of data size in this study, recurrent patterns were revealed in Raman spectra of cells from NSCLC cases with differing cytology reports, indicating consistency in the spectral data. With further investigation, the diagnostic peaks identified could prove to be useful spectral markers for the detection of NSCLC subtypes. In addition, Raman spectroscopy used with PLSDA classification distinguished between the TN cases and NSCLC cases with high sensitivities and specificities. This illustrates the potential of Raman spectroscopy to improve the diagnostic accuracy of cytological analysis for lung cancer, unimpeded by the limitations of current cytopathological methods.

As the suitability of using Raman spectroscopy on cytology for NSCLC diagnosis has been demonstrated, the next chapter aims to compare the diagnostic effectiveness of using Raman spectroscopy on bronchoscopy attained cytological and tissue samples.

References


Chapter 6: Comparison of the diagnostic effectiveness of applying Raman spectroscopy to cytological specimens and tissue sections for NSCLC diagnosis

6.1 Introduction

Histological analysis of tissue specimens is the current gold standard diagnostic method for lung cancer. Histopathology involves the microscopic analysis of cellular patterns and tissue architecture to make an accurate diagnosis, through the application of histochemical stains and immunohistochemical markers. For the classification of cancer type, histopathologists look for specific characteristic features to classify into a diagnosis, such as the size and shape of cells, size and shape of nuclei, the morphology and distribution of cells within in the tissue. The tissue architecture is preserved for histological evaluation, making it advantageous over cytopathology in the current context of lung cancer diagnosis.

Studies exploring the use of Raman spectroscopy for lung cancer diagnosis on histological sections have shown the ability of this technique to delineate between normal and malignant lung tissue with high diagnostic sensitivities and specificities\textsuperscript{1–9}. With conventional Raman spectroscopy, the 785 nm and 1064 nm laser lines have been used to reveal distinctive malignant features in the Raman spectra\textsuperscript{3,10}. These studies exhibit the potential of using Raman spectroscopy for the optical diagnosis of lung cancer, yet the ability to translate these methods to a clinical setting is a limiting factor.

The primary aim of this chapter was to characterize the Raman spectra taken from bronchial tissues and compare the diagnostic effectiveness of using Raman spectroscopy on bronchoscopy attained cytological and tissue samples. The diagnostically relevant information differentiating the tissue spectra was compared to the diagnostic peaks identified in chapter 4. To our knowledge, no studies have compared the biochemical information extracted from the spectra taken from these different diagnostic sample types.
6.2 Methods

6.2.1 Study subjects and sample preparation

As described in chapter 3 section 3.1, this study utilised residual tissue blocks from 11 patients who had undergone biopsy during bronchoscopy for suspected lung cancer. The final histopathological diagnoses, characterised by a pathologist using gold standard diagnostic methods, was recorded as: 6 squamous cell carcinoma, 3 adenocarcinoma, 1 NSCLC-not otherwise specified (NSCLC NOS), and 1 atypical carcinoid.

6.2.2 Sample processing and slide preparation

10 µm parallel sections were cut from the FFPP tissue blocks and mounted on glass slides for Raman spectroscopy. The tissue sections were chemically dewaxed by sequential immersion of the sections in two xylene baths, and two IMS baths (100%), for 10 minutes in each bath. Parallel tissue sections were cut at 5 µm and stained with Haematoxylin and Eosin to allow differentiation between different tissue structures.

6.2.3 Raman Measurement

Raman spectra were acquired using a HORIBA Jobin Yvon Labram HR800 UV system, which consists of a 50 mW diode laser with 532 nm wavelength. The laser was delivered to the sample through a x100 objective (NA = 0.9) and the confocal hole was set at 100 µm. The system was calibrated to the 520 cm\(^{-1}\) spectral line of silicon and the spectral resolution was defined by the grating which was ruled with 600 lines per mm grating. Backscattered light was detected using an air-cooled CCD detector. Regions of the section with morphological changes associated with disease (Figures 6.1, 6.2, 6.4 and 6.5), and regions associated with normal morphology (Figure 6.3) were targeted with the laser. These regions were identified and marked using parallel H & E stained sections shown in Figures 6.1-5. Spectra were also recorded from 300 vacant locations on a glass slide, and
25 locations on wax regions with no underlying tissue. A thirty second integration time averaged over two accumulations was used in order to obtain representative background glass and wax spectra.

6.2.4 Data pre-processing and analysis
All spectral data analysis was conducted using MATLAB software. Data were processed as described in chapter 3 section 3.2.3-7.

The data was analysed using Principal Component Analysis (PCA), as described in chapter 3 section 3.2.8. Partial least squares discriminant analysis (PLS-DA) with venetian blind cross-validation was also employed to the mean centered Raman spectral data to identify the diagnostically significant variations between the groups, as described in chapter 3 section 3.2.10-12. As described in chapter 3 section 3.2.6 11, spectra from cellular components, such as nucleic acids, proteins, lipids etc, and the independently recorded wax spectra were used to fit the sample spectra and remove the wax contribution.

6.2.5 Haematoxylin and Eosin
Sections were dewaxed and rehydrated to water then stained with Harris’s Haematoxylin for 5 minutes. After washing in water, the Haematoxylin was then differentiated with acid alcohol for 2 seconds and checked. Complete removal of the haematoxylin from the cell cytoplasm was checked promptly with a light microscope, ensuring the slides do not dry. The slides were then submerged in water to blue the bound stain. Eosin was then added to the slides for 3 minutes before removing the excess stain with water. After washing rapidly in 95% IMS, the slides were transferred through absolute alcohols to histoclear before coverslipping.

6.3 Results
Figures 6.1 to 6.5 show H&E stained tissue sections from each malignant type analysed. Nuclei are stained dark blue purple, and Eosin has stained cytoplasm and connective
tissue pink. SCC lesions with islands of large eosinophilic cells with intercellular bridges are depicted in Figure 6.1, while Figure 6.2 depicts adenocarcinomas showing acinar formation, mucinous vacuoles and signet ring cells. Normal bronchial epithelium with mucinous glands is shown in Figure 6.3 and NSCLC-NOS tissue is depicted in Figure 6.4 which shows abnormal cell and nuclear morphology. Atypical carcinoid tissue is shown in Figure 6.5 with sheets of tumour cells that have uniform round nuclei. The tumour cells form nests which are sharply distinct from the stroma\textsuperscript{12,13}. 
Figure 6.1. Cases of SCC stained with Haematoxylin and Eosin. A-F show nests of malignant squamous epithelial cells with large eosinophilic cytoplasm and intercellular bridges. A-D are at x200 magnification, E and F are at x400 magnification.
Figure 6.2. Cases of AC stained with Haematoxylin and Eosin. A-F show adenocarcinomas with acinar formation, mucinous vacuoles and signet ring cells. A-D are at x200 magnification, E and F are at x400 magnification.
Figure 6.3. Normal tissue stained with Haematoxylin and Eosin. A shows normal bronchial epithelium with mucinous glands at x200 magnification. B shows normal mucinous glands at x400 magnification. C shows normal bronchial epithelium at x400 magnification. D shows normal bronchial epithelium at x200 magnification. E and F show normal bronchial epithelium at x400 magnification.
Figure 6.4. Case of NSCLC-NOS stained with Haematoxylin and Eosin. A and B show abnormal cellular and nuclear morphology however there are no defining characteristics for AC or SCC. A is at x200 magnification, B is at x400 magnification.

Figure 6.5. Case of Atypical carcinoid stained with Haematoxylin and Eosin. A and B Sheets of tumour cells that have uniform round nuclei. The tumour cells form nests which are sharply distinct from the stroma. A is at x200 magnification, B is at x400 magnification.
6.3.1 Wax contamination after digital wax removal

As wax contributions in the spectra can cause inaccurate results, the effectiveness of digital wax removal across the entire dataset was evaluated. To obtain a representation of wax peaks remaining in all tissue spectra, the Raman spectra from morphologically normal regions of tissue were combined into a “Negative” group, and spectra taken from morphologically abnormal regions of tissue were combined into a “Malignant” group for comparative analysis. The mean Raman spectra of grouped negative tissues and grouped malignant tissues after chemical and digital dewaxing are depicted in Figure 6.6. The standard deviation is denoted by shading. Principal component analysis was implemented on the negative and malignant spectra to determine the discriminatory performance of PCA and the peak assignments responsible for discrimination. Figure 6.7A shows the scatter plot of the PC scores along the first and fifth principal components, demonstrating separation of the Raman spectra along PC5. In Figure 6.7B the loadings of PC5 indicate that wax peaks, denoted by grey shading, are influencing the separation of negative and malignant spectra. Retaining variables affected by Raman peaks of wax may inaccurately influence analysis. To ensure reliability of results, the Raman peaks of wax in the spectral regions of 877 to 895, 1050 to 1072, 1115 to 1183, 1280 to 1311, and 1400 to 1500 cm\(^{-1}\) were removed from the dataset for all further analysis.
Figure 6.6. Mean Raman spectrum of the grouped negative tissue (black, \( n = 112 \)). Mean Raman spectrum of the grouped malignant tissue (red, \( n = 119 \)). Shading denotes the standard deviation. The spectral data has been subject to NNLS for removal of wax peaks, however strong wax peaks remain in the spectra at 1060 and 1294 cm\(^{-1}\).
Figure 6.7. A PCA scatter plot of processed negative and malignant tissue shows a separation on the fifth principal component with even distribution along the first principal component. B Wax spectral peaks influence the fifth principal component, along which the spectra from negative and malignant tissue are separated. Shading denotes the wax related peaks influencing the fifth principal component.
6.3.2 Comparison of the mean Raman spectra from bronchial tissues

Good quality Raman spectra were obtained from the chemically and digitally dewaxed lung tissue sections. There was no intereference of fluorescence from the tissue, and the Raman signals from the substrate did not appear in the processed spectra. The Raman peaks of wax in the spectral regions of 877 to 895, 1050 to 1072, 1115 to 1183, 1280 to 1311, and 1400 to 1500 cm\(^{-1}\), appeared in the processed spectra (Figures 6.6 and 6.7). Accordingly, the corresponding variables were removed from the dataset for comparative analysis of the mean spectra. Figure 6.8 shows the mean spectrum taken from regions associated with normal morphology. Shading in Figure 6.8 denotes the standard deviation. The features in the spectra taken from regions with morphological changes associated with disease that differ from the morphologically normal tissue are demonstrated by the following difference spectra in Figure 6.9. The biological components for the significant Raman peaks were tentatively assigned according to the studies by Movasaghi et al\textsuperscript{14}, and Naumann et al\textsuperscript{15}.

Spectral differences between the negative tissue and grouped spectra from malignant tissue corresponded to peaks relative to cholesterol (436-39 cm\(^{-1}\)), DNA / glycogen (480 cm\(^{-1}\)), glucose / saccharides / phosphatidylinositol / cholesterol (560 cm\(^{-1}\)), phospholipids / methionine (715 cm\(^{-1}\)), saccharides (860 cm\(^{-1}\)), proline / valine (930 cm\(^{-1}\)), nucleic acids (1185-1200 cm\(^{-1}\)), amide III / nucleic acids (1230 cm\(^{-1}\)), lipids / CH rocking / C=O symmetric stretch (1394-99 cm\(^{-1}\)), amide II (1545-61 cm\(^{-1}\)), COO\(^{-}\) / DNA / cytosine / tyrosine / amide I (1575-1625 cm\(^{-1}\)), and amide I (1637 cm\(^{-1}\)). The difference spectra of the SCC and AC tissues had similar profiles, with significant peaks assigned to at glucose / saccharide / phosphatidylinositol / cholesterol (560 cm\(^{-1}\)), phospholipids / methionine (715 cm\(^{-1}\)), saccharides (~860 cm\(^{-1}\)), proline / valine (925 cm\(^{-1}\)), collagen (~1030), glycogen / collagen (1045-50 cm\(^{-1}\)), saccharides / lipids / collagen (1075-1114 cm\(^{-1}\)),
nucleic acids (1185-1201 cm$^{-1}$), amide III / nucleic acids (1214-35 cm$^{-1}$), adenine (1510 cm$^{-1}$), tyrosine / tryptophan(1616 cm$^{-1}$), and amide I (1657-77 cm$^{-1}$) differentiating them from the negative tissue (Figure 6.9). In comparison to AC tissue, SCC tissue had elevated levels of proteins (623-25 cm$^{-1}$), phospholipids (~720 cm$^{-1}$), nucleic acids (771-804 cm$^{-1}$), amide III / nucleic acids (1209-15, 1317-23 cm$^{-1}$), and aspartic / glutamic acid (1693-1709 cm$^{-1}$), and lower levels of cholesterol / thiocyanate (438-44 cm$^{-1}$), protein / saccharides / phosphates (843-51, 872-76 cm$^{-1}$) proline / glycogen (~918 cm$^{-1}$), collagen (1028-29 cm$^{-1}$), amide III / collagen (1272-79 cm$^{-1}$), lipids / CH rocking / C=O symmetric stretch (1384-99 cm$^{-1}$), amide II / tryptophan / tyrosine (1536-56 cm$^{-1}$), and amide I / tryptophan / tyrosine (1596-1651 cm$^{-1}$) (Figure 6.10).

For the NSCLC-NOS tissue, spectral differences were observed at 780 cm$^{-1}$, 860 cm$^{-1}$, 1019-32 cm$^{-1}$, 1040-50 cm$^{-1}$, 1075-1114 cm$^{-1}$, 1185-1200 cm$^{-1}$, 1270 cm$^{-1}$, 1342-50 cm$^{-1}$, and 1650 cm$^{-1}$. This corresponds to a variation in DNA, protein / phosphate, saccharides, glycogen, collagen, amide III / nucleic acids, and amide I in the NSCLC-NOS tissue in comparison to the negative tissue (Figure 6.9).

Compared to the negative tissue, the atypical carcinoid tissue showed spectral differences associated with glycogen (480 cm$^{-1}$), glucose / saccharides / phosphatidylinositol / cholesterol (560 cm$^{-1}$), tyrosine (640 cm$^{-1}$), cytosine (661 cm$^{-1}$), phospholipids / methionine (715 cm$^{-1}$), saccharides (860 cm$^{-1}$), proline / valine (~930 cm$^{-1}$), nucleic acids / amide III (1185-1296 cm$^{-1}$), collagen (1335 cm$^{-1}$), saccharides (~1370 cm$^{-1}$), COO$^-$ (1585 cm$^{-1}$), and amide I (>1600 cm$^{-1}$) (Figure 6.9).
Figure 6.8. Mean spectrum taken from negative tissue (n= 112 spectra). Wax related peaks in the ranges of 877 to 895, 1050 to 1072, 1115 to 1183, 1280 to 1311, and 1400 to 1500 cm$^{-1}$ were removed for analysis. Shading denotes the standard deviation.
Figure 6.9. Difference spectrum between the mean Raman spectrum of negative tissue and the mean Raman spectrum taken from SCC tissue (blue, n=141), AC tissue (red, n=70), NSCLC-NOS (green, n=25), atypical carcinoid (magenta, n=25), and grouped spectra from malignant tissue (light blue, n=119). Shading indicates regions of the spectrum that were significantly different (P < 0.01). Variables were excluded from spectral analysis in the ranges of 877 to 895, 1050 to 1072, 1115 to 1183, 1280 to 1311, and 1400 to 1500 cm\(^{-1}\), as the spectra indicated wax contamination.
Figure 6.10. Difference spectrum between the mean Raman spectrum of SCC tissue \((n=141)\) and the mean Raman spectrum taken from AC tissue \((n=70)\). Variables were excluded from spectral analysis in the ranges of 877 to 895, 1050 to 1072, 1115 to 1183, 1280 to 1311, and 1400 to 1500 cm\(^{-1}\), as the spectra indicated wax contamination.

6.3.3 Principal component analysis

Principal component analysis of the data was conducted on the processed spectra to extract diagnostic information from the spectra. The five spectral regions from 877 to 895, 1050 to 1072, 1115 to 1183, 1280 to 1311, and 1400 to 1500 cm\(^{-1}\) were affected by the Raman peaks of wax, therefore the corresponding variables were removed from the dataset for principal component analysis. To evaluate the ability of PCA to discriminate between the negative tissue, malignant tissue, and LC subtypes, spectra were grouped for analysis. Representative spectra from each patient were selected at random and combined by their subtype and the number of spectra were matched in each group. Figure 6.11 shows the scatter plot of the PC scores along the first and fourth principal components, demonstrating distinctive clustering of the Raman spectra from malignant and negative tissue. The malignant and negative spectra form two adjacent, overlapping clusters along the fourth PC. The loadings of PC4 indicate variation in phosphatidylinositol (570 cm\(^{-1}\)), phenylalanine (1004 cm\(^{-1}\)), protein (1106 cm\(^{-1}\)), and COO\(^-\) (1585 cm\(^{-1}\)) associated with
the malignant tissue, and peaks assigned to phospholipids / methionine (715 cm\(^{-1}\)), saccharides (850 cm\(^{-1}\)), proline / valine / saccharides (925 cm\(^{-1}\)), amide III / nucleic acids (1233 cm\(^{-1}\)), and amide I (1645 cm\(^{-1}\)) associated with the negative tissue.

Figure 6.11. A PCA scatter plot of processed negative and malignant tissue shows a separation on the fourth principal component with comparable distribution along the first principal component. B Fourth PC after wax contribution was removed. On the positive side of the fourth principal component, phenylalanine related peak 570 cm\(^{-1}\), 1004 cm\(^{-1}\) and 1585 cm\(^{-1}\) can be distinguished. While peaks at 715, 850, 925, 1233, and 1645 cm\(^{-1}\) can be distinguished on the negative side. Variables were excluded from PCA analysis in the ranges of 877 to 895, 1050 to 1072, 1115 to 1183, 1280 to 1311, and 1400 to 1500 cm\(^{-1}\), as the spectra indicated wax contamination.
Figure 6.12 shows the score plot of the first two PCs developed for the analysis of grouped SCC spectra versus grouped AC spectra. The SCC and AC spectra were not distinguishable using a combination of the first five PCs, which explained 95% of the variance in the dataset.

![Figure 6.12. PCA scatter plot of SCC and AC tissue shows no separation with even distribution along the first two principal components. Variables were excluded from PCA analysis in the ranges of 877 to 895, 1050 to 1072, 1115 to 1183, 1280 to 1311, and 1400 to 1500 cm\(^{-1}\), as the spectra indicated wax contamination.](image)

In Figure 6.13 the distribution of PC2 and PC3 was plotted to visualize the variances between SCC, AC, and NSCLC-NOS spectra. The scatter plot using PC2 and PC3 gave the best discrimination performance, although the clusters overlap. The SCC and NSCLC-NOS spectra are separated along PC2, yet neither the SCC nor NSCLC-NOS clusters can be distinguished from AC due to the even distribution of AC spectra along PC2 and PC3. The loadings of PC2, which is responsible for discriminating SCC and NSCLC-NOS, indicate peaks assigned to phospholipids/methionine (715 cm\(^{-1}\)), nucleic
acids (780 cm	extsuperscript{-1}), lipid (~1380 cm	extsuperscript{-1}), and COO\textsuperscript{-} / amide II / typtophan / amide I (1501-1600 cm	extsuperscript{-1}) associated with the NSCLC-NOS tissue, and peaks assigned to phosphatidylinositol / cholesterol (560 cm	extsuperscript{-1}), saccharides (850 cm	extsuperscript{-1}), saccharides / proline / valine (945 cm	extsuperscript{-1}), phenylalanine (1004 cm	extsuperscript{-1}), amide III / nucleic acids (1240 cm	extsuperscript{-1}), and amide I (1650 cm	extsuperscript{-1}) associated with SCC tissue.

Figure 6.13. A. PCA scatter plot of processed SCC, AC and NSCLC-NOS tissue shows that the SCC and NSCLC-NOS spectra form two distinct clusters along the second principal component. The AC spectra are dispersed along the second PC and cannot be discriminated from SCC or NSCLC. B. On the positive side of the second principal component, peaks at 715, 780, 1185, 1200, ~1360, and ~1575 cm\textsuperscript{-1} can be distinguished. While some peaks at 560, 1004, 860, 930, 1030, 1075-1114, 1230, and 1650 cm\textsuperscript{-1} can be distinguished on the negative side. Variables were excluded from PCA analysis in the
ranges of 877 to 895, 1050 to 1072, 1115 to 1183, 1280 to 1311, and 1400 to 1500 cm\(^{-1}\), as the spectra indicated wax contamination.

Figure 6.14 visually demonstrates the discriminatory performance of PCA using representative spectra from negative tissue and each lung malignancy. Representative spectra from each subtype were selected at random and combined into groups as previously described. The SCC spectra and the AC spectra form two close clusters using PC1 and PC3. With a combination of PC1 and PC3, the SCC and AC clusters can be discriminated from the negative, NSCLC-NOS, and atypical carcinoid spectra, with some overlap. The loadings from PC1 indicate peaks assigned to proteins / saccharides (660, 845, 860 cm\(^{-1}\)) and DNA (780 cm\(^{-1}\)) associated with the negative and NSCLC-NOS tissues. The loadings spectrum of PC3 indicates peaks assigned to nucleic acids / glycogen / saccharides / phosphatidylinositol / cholesterol (480, 560 cm\(^{-1}\)), amide II / COO\(^-\) (1580 cm\(^{-1}\)), and amide I (>1600 cm\(^{-1}\)) associated with the SCC and AC tissue, and peaks assigned to phospholipids / methionine (715 cm\(^{-1}\)), nucleic acids (780 cm\(^{-1}\)), amide III / nucleic acids (1185-1280 cm\(^{-1}\)), and saccharides (1370 cm\(^{-1}\)) associated with the negative and NSCLC-NOS tissue. The negative, NSCLC-NOS, and atypical carcinoid spectra are similarly distributed along both PC1 and PC3, and so can not be distinguished.
Figure 6.14. A Scatter plot of the principal component scores of representative spectra from negative, SCC, AC, NSCLC-NOS, and atypical carcinoid tissue. A combination of PC1 and PC3 separates the SCC and AC clusters from the negative, NSCLC-NOS and atypical carcinoid spectra, with some overlap. The negative, NSCLC-NOS, and atypical carcinoid tissue shows similar distribution along the first and third principal components.

B Loadings from PC1 and PC3. Variables were excluded from PCA analysis in the ranges of 877 to 895, 1050 to 1072, 1115 to 1183, 1280 to 1311, and 1400 to 1500 cm\(^{-1}\), as the spectra indicated wax contamination.
6.3.4 PLSDA classification of lung cancer subtypes

PLSDA was used for discriminant analysis and cross-validation. To construct each model the required number of latent variables were used to set the cross validation classification error average at 5% and the number of spectra used for each class was matched to prevent a biased classifier. The performance of each classification is shown using the sensitivity and specificity of each model.

The five spectral regions from 877 to 895, 1050 to 1072, 1115 to 1183, 1280 to 1311, and 1400 to 1500 cm\(^{-1}\) were affected by the Raman peaks of wax (Figures 6.6 and 6.7), thus the corresponding variables were removed from the dataset for partial least squares discriminant analysis.

Table 6.1 shows the performance of the PLSDA model in discriminating the grouped spectra from morphologically normal tissue and representative spectra from each lung cancer subtype combined into one malignant group. The first four latent variables were used to develop the two-way PLSDA model, which produced a sensitivity of 84% and a specificity of 89% for detecting malignancy. Figure 6.15 is a LV score scatter plot depicting Raman spectra from morphologically normal tissue, and morphologically malignant tissue, showing separation of the two groups. The negative spectra are separated from the malignant spectra along LV with some overlap.
Figure 6.15. LV score scatter plot of Raman spectra from negative tissue (blue circles), and grouped spectra from malignant tissue (red circles). Variables were excluded from PCA analysis in the ranges of 877 to 895, 1050 to 1072, 1115 to 1183, 1280 to 1311, and 1400 to 1500 cm\(^{-1}\), as the spectra indicated wax contamination.

Table 6.2 summarises the results for the three-way classification of spectra from negative, SCC and AC tissue. The first 5 latent variables were used to develop the PLSDA model. Negative tissue was discriminated from SCC and AC tissue with a sensitivity and specificity above 83%. Spectra from SCC tissue were identified with a sensitivity and specificity of 76% and 79% respectively, and the AC subtype was classified with a sensitivity of 90%, and a specificity of 89%. Figure 6.16 depicts the scatter plot of the latent variable scores of the spectra taken from negative, SCC, and AC tissue. Using the first two latent variables to plot the scatter plot, the negative, SCC, and AC spectra form three distinct clusters. The negative spectra can be discriminated from the SCC and AC spectra using a combination of the first 2 latent variables, and the SCC and AC clusters are separated along the second LV.
Table 6.3 shows the results of 2 separate PLSDA models in discriminating spectra from negative tissue from the spectra taken from tissue classified as NSCLC-NOS or atypical carcinoid tissue. The first 4 latent variables were used in each case to develop the PLSDA classifier. PLSDA classification of negative tissue and NSCLC-NOS or atypical carcinoid yielded sensitivities and specificities over 92%.

Figure 6.16. LV score scatter plot of Raman spectra from negative tissue (blue circles), SCC tissue (green circles), and AC tissue (red circles). Variables were excluded from PCA analysis in the ranges of 877 to 895, 1050 to 1072, 1115 to 1183, 1280 to 1311, and 1400 to 1500 cm\(^{-1}\), as the spectra indicated wax contamination.
Table 6.1. Performance of PLSDA in discriminating Raman spectra from negative and malignant tissue.

<table>
<thead>
<tr>
<th>Histological diagnosis</th>
<th>Sensitivity (%)</th>
<th>Specificity (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Negative</td>
<td>89</td>
<td>84</td>
</tr>
<tr>
<td>Malignant</td>
<td>84</td>
<td>89</td>
</tr>
</tbody>
</table>

Table 6.2. Performance of PLSDA in discriminating the Raman spectra from negative, SCC, and AC tissue.

<table>
<thead>
<tr>
<th>Histological diagnosis</th>
<th>Sensitivity (%)</th>
<th>Specificity (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Negative</td>
<td>83</td>
<td>85</td>
</tr>
<tr>
<td>SCC</td>
<td>76</td>
<td>79</td>
</tr>
<tr>
<td>AC</td>
<td>90</td>
<td>89</td>
</tr>
</tbody>
</table>

Table 6.3. Performance of PLSDA in discriminating Raman spectra taken from negative tissue, NSCLC tissue, and atypical carcinoid tissue.

<table>
<thead>
<tr>
<th>Tumour type distinguished from negative tissue</th>
<th>Sensitivity (%)</th>
<th>Specificity (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>NSCLC-NOS</td>
<td>92</td>
<td>96</td>
</tr>
<tr>
<td>Atypical carcinoid</td>
<td>100</td>
<td>92</td>
</tr>
</tbody>
</table>

6.4 Discussion

Diagnostic samples such as lung tissue biopsies are commonly preserved in paraffin wax. Chemical dewaxing is generally sufficient to remove wax from embedded sections and prepare the tissue for histochemical diagnostic techniques, however due to the sensitivity of Raman spectroscopy, further digital dewaxing is often required. In this study the non-negatively constrained least squares method was applied to the data to remove
contaminating wax peaks, although spectral contaminants were still observed in the data. This may be caused by the microcrystalline structures within paraffin wax, which are randomly oriented in relation to the polarizing laser, resulting in an inhomogeneous response depending on sampling location. A larger matrix of wax spectra, accounting for more of the inhomogeneity may therefore result in more effective wax removal. To ensure the results of this study are reliable, the Raman peaks associated with wax contamination were removed from the data for all further analysis.

The results of this exploratory study demonstrate that there are significant differences in the Raman peak intensities of negative tissue and lung cancer subtypes. Spectral differences were consistently observed between the negative and lung cancer subtypes with the strongest signals occurring at ~560, 715, ~780, ~860, ~925, ~1045, 1075-1114, 1185-1200, ~1230, 1278, ~1395, 1510, 1545, 1575-1700 cm\(^{-1}\), representing significant differences in the molecular composition of saccharides, phosphatidylinositol, cholesterol nucleic acids, phospholipids, protein structures and amides, across the negative and lung cancer tissues. These consistent spectral differences between the negative and malignant tissues are reflective of the malignancy-associated changes that occur in lung carcinogenesis and are mainly in accordance with the characteristic lung cancer peaks identified in previous studies using tissue samples. The narrow and broad Raman peaks differentiating the malignant tissues from the negative tissue represent differences in the vibrational modes of biomolecules in the malignant cells. As the broad Raman bands are a result of Raman scatter from overlapping vibrational modes of biomolecules, mapping these broad bands to a single vibrational mode is not reliable. For this reason, the full spectral region of broad bands is utilised for the biomolecular correlation of Raman peaks.
The difference spectra of SCC, AC, atypical carcinoid and grouped malignant tissue spectra are characterized by reduced signals for phospholipids/methionine (715 cm\(^{-1}\)) and amino acids/saccharides (860 and 930 cm\(^{-1}\)), and elevated broad bands for collagen and glycogen (~1030, 1045-50 cm\(^{-1}\)) saccharides/phosphatidylinositol/cholesterol (560 cm\(^{-1}\)) and amide I (>1600 cm\(^{-1}\)), indicating alterations in the content of these biomolecules relative to the negative tissue. This is a finding also reported by Huang et al\(^3\) in a study comparatively analysing spectra from negative and malignant lung tissue. The Raman peaks consistent with malignancy in this chapter have previously been reported in other tissue types including cervix, and skin\(^{20,21}\).

The profile of the difference spectrum between the negative tissue and the NSCLC-NOS tissue is inconsistent with the other difference spectra. In NSCLC-NOS tissue, the broad amide I band (>1650 cm\(^{-1}\)) was found to decrease relative to the negative tissue. In addition, the DNA signal at 780 cm\(^{-1}\) was increased in the NSCLC-NOS tissue. The significant spectral differences between the negative and NSCLC-NOS tissue are consistent with the significant spectral differences between cells from negative cases and cells from SCC and AC cases identified in chapter 4. The unique spectral profile observed for NSCLC-NOS tissue may be a result of the heterogeneity of NSCLC. This is reflected by the difficulty had by pathologists in classifying the tumour type. The limited spectral data obtained for NSCLC-NOS in this study is another possible reason for the unique spectral profile observed.

As the differential diagnosis of SCC and AC is now a part of the diagnostic algorithm of NSCLC, spectroscopic analysis may be utilised to identify diagnostically significant Raman peaks that aid discrimination. In this study, the difference spectrum between the negative tissue and the SCC tissue shows similar prominent peaks to the difference spectra between the negative tissue and the AC tissue. In comparison to the negative
tissue, the SCC and AC subtypes showed increased peak intensities at 560, 1075-1114, 1616 and ~1650 cm\(^{-1}\), and decreased peak intensities at 715, 860, ~930, 1214-1230, and 1510 cm\(^{-1}\). This correlates with an increase in the level of lipids, collagen, tryptophan/tyrosine, cholesterol, phosphatidylinositol and amide I, and a decrease in the amount of phospholipids, saccharides, methionine, proline and valine, amide III, and nucleic acids in the SCC and AC tissue. The biochemical profile of these subtypes shows similarities to the characteristic molecular alterations associated with SCC and AC identified in a study by Kaznowska et al\(^{16}\) using vibrational spectroscopy on lung tissue. The spectral profiles of these two lung cancer subtypes reveal a comparable biochemistry. In chapter 4, the bronchial cells from SCC and AC cases, and the normal and abnormal squamous cells from SCC and AC cases also displayed elevated peak intensity at ~1030-50 cm\(^{-1}\). The reduced signals for phospholipids/methionine (715 cm\(^{-1}\)) and saccharides (~930 cm\(^{-1}\)) observed in the difference spectra presented in this chapter were also seen in the cytoplasm of bronchial and squamous cells of SCC and AC cases presented in chapter 4. Significant peaks associated with amide III and I differentiate negative from the SCC and AC subtypes on both cytology and histology, however the relative peak intensities are reversed. In the spectra obtained from cytological samples, increased amide III/nucleic acids and decreased amide I is associated with the NSCLC subtypes, whereas in the spectra obtained from tissue specimens, decreased amide III/nucleic acids and increased amide I is associated with the NSCLC subtypes. These results indicate that regardless of sample type, variations in these specific biomolecules differentiate negative from the NSCLC subtypes, however the content of these biomolecules can vary depending on the sampling method. The cells targeted in chapter 4 are from the surface layer of epithelium, due to the exfoliative and abrasive nature of cytology samples. The sections of tissue targeted in this chapter represent cells from tumours that have grown
within the deeper layers of the tissue (as seen in the images), as malignant cells that have breached the basement membrane into the lamina propria can be sampled with tissue biopsy. The variation may also be a result of tissue formalin fixation, and processing through alcohols and xylene, all of which may alter the biochemical composition.

In a clinical setting, difficulty differentiating between SCC and AC is common due to overlapping morphological and molecular features. It is therefore not surprising that they have similar spectral profiles in this study. Despite their similarities, comparative analysis of the mean SCC and AC spectra revealed the biomolecular differences between the two lung cancer subtypes. An elevation in protein, phospholipid, nucleic acid, amide III, and glutamine content was associated with SCC on tissue sections, whereas AC was characterised by greater peak intensities associated with carbohydrates, cholesterol / thiocyanate, proline and valine, glycogen, collagen, amide II, and amide I. AC had an increased amide I and II in comparison to SCC both in cytological and tissue samples. As discussed in chapter 4, alterations in glycogen, collagen, and nucleic acid content are recurring features in lung malignancies. The results of this study is in accordance with other studies that have demonstrated the ability of Raman spectroscopy to detect variations in these biomolecules in the spectra of malignant lung tissues.1,17–19,22

PCA was then conducted to reduce the dimensionality of the data and summarise the variation between the negative and malignant groups. The fourth PC was the main contributor to the discrimination of malignant and negative spectra. The loadings plot of PC4 shares features with the difference spectrum between negative and malignant tissue, indicating that biochemical alterations in saccharides, phosphatidylinositol, cholesterol, phospholipids, methionine, proline, valine, phenylalanine, lipid, collagen, nucleic acids, amide III and amide I in the malignant tissue contribute to the discrimination of malignant and negative tissue. PCA of the data from SCC and AC tissues could not differentiate
between the two NSCLC subtypes. The inability of this unsupervised classification technique to discriminate between these two lung cancer subtypes shows similarities in their biochemistry, revealed by comparable spectral profiles. The overlapping spectral features are depicted visually using a scatter plot of PC1 and PC2.

In the PCA scatter plot of spectra from SCC, AC and NSCLC-NOS tissue, the SCC and NSCLC-NOS spectra formed two separate clusters along PC2, although the AC spectra could not be distinguished from SCC or NSCLC-NOS spectra. The peaks influencing discrimination of the SCC and NSCLC-NOS tissue were tentatively assigned to saccharides, phosphatidylinositol, cholesterol, phospholipids, methionine, proline, valine, phenylalanine, collagen, lipids, collagen, nucleic acids, amide III, amide II, COO⁻, and amide I. As the spectral data for NSCLC-NOS came from a single case in this study, interpatient variability is a possible cause of variation between the SCC and NSCLC-NOS spectra. However, the pronounced differences between the SCC and NSCLC-NOS spectra are consistent with the malignant peak assignments observed throughout this study, which suggests that the variation between the two diagnoses may be a result of malignant associated alterations. The NSCLC-NOS spectra could not be classified as either SCC or AC. Increasing the number of patients in the study and thus increasing the spectral data may help classify future NSCLC-NOS cases.

PCA of Raman spectra taken from negative, SCC, AC, NSCLC-NOS, and atypical carcinoid tissues showed that the SCC and AC spectra could be discriminated from the negative spectra based on proteins / saccharides (660, 845, 860, 1004, 1650 cm⁻¹), and DNA (780 cm⁻¹), nucleic acids / glycopigen / saccharides / phosphatidylinositol / cholesterol (480, 560, 780, 1185-1280, 1370 cm⁻¹), phospholipids / methionine (715 cm⁻¹), amide III (1185-1280 cm⁻¹), amide II / COO⁻ (1580 cm⁻¹), amide I (>1600 cm⁻¹). The PC loadings spectra display features that are in the difference spectra between the
negative tissue and the SCC and AC tissue. The recurring differences between the tissues analysed in this study are evident in spectral regions associated with nucleic acids, carbohydrates, phospholipids, lipids, amide, and proteins. These biomolecular variations between malignancy type are consistent with the findings of Kaznowska et al\textsuperscript{16}, which utilised FTIR spectroscopy in the investigation of lung cancer tissue. With further research these peaks assignments could prove to be effective for the characterization of lung cancer tissues.

Partial least squares discrimination analysis, a supervised classification method, was then applied to the dataset to differentiate between negative and malignant tissue, and lung cancer subtypes. The results of this study demonstrate that PLSDA on processed Raman spectra can be used to delineate between negative and malignant tissue, with sensitivity and specificity over 84\%. The diagnostic performance of Raman spectroscopy in this study is in accordance with other studies that applied Raman spectroscopy to lung tissue to detect malignancy\textsuperscript{2,3,6}. It is plausible that perfect sensitivity and specificity was not achieved because morphologically normal tissue adjacent to the tumour tissue is exposed to the same carcinogenic insults as the tumour. These normal cells may have undergone malignant associated changes which is reflected in the cellular biochemistry and detected in the Raman spectra\textsuperscript{23}. This may explain the misclassified malignant and negative spectra. Field cancerization is another plausible reason for the misclassified spectra, whereby genetically altered cells in the tumour area or field remain morphologically normal and are not identified\textsuperscript{24–28}. PLSDA was then applied to discriminate the spectra taken from negative tissue and the spectra from SCC and AC tissue, all within the same model. The model yielded a sensitivity of 83\% and specificity of 85\% for identification of negative tissue. AC was detected with a sensitivity of 90\%, and the SCC subtype was classified with a sensitivity and specificity of 76 and 79\% respectively. The algorithm
misclassified 13 SCC spectra as AC. The lower discrimination performance for the classification of SCC using PLSDA on lung tissue spectra is similar to the results yielded using PLSDA on spectra from cytology samples in chapter 4. The misclassification of SCC and AC spectra observed in this study complies with the reported difficulty in differentiating these subtypes in a clinical setting. The heterogeneity of these malignancies results in overlapping features in the cellular biochemistry, which are reflected in the spectra. The SCC subtype was classified with marginally lower accuracy than the other subtypes in each of the bronchoscopic diagnostic sample types analysed with Raman spectroscopy.

Objective comparative analysis of negative, NSCLC-NOS, and atypical carcinoid tissue spectra was achieved by the application of two-way PLSDA algorithms to the data. The models achieved sensitivities and specificities over 92% for the detection of NSCLC-NOS and atypical carcinoid tissue. Although, as these results are limited by the low number of patients, classification may be based on interpatient variability. The PLSDA performance scores achieved for the classification of lung malignancies using tissue sections in this chapter (sensitivities and specificities of 76-100%) are comparable to the diagnostic sensitivities and specificities yielded using cytological specimens in chapter 4 (sensitivities and specificities of 60-97%), however cytology samples yielded more varied results.

The difference spectra in this study show that there are many broad Raman bands differentiating the negative and malignant tissue types. The broad bands observed in the difference spectra in this chapter indicate substantial overall variation between the spectra from different tissue types. Comparatively, the difference spectra between cells from cytological samples in chapter 4 have more narrow peaks differentiating the lung cancer subtypes, suggesting less overall variation between the spectra taken from different
subtypes. This is not surprising as specific cells and components were being targeted and directly compared on the cytology samples. When applying Raman spectroscopy to tissue samples the cell types and components are not easily distinguished and therefore cannot be targeted. Hence, the spectra are influenced by a combination of nuclei and cell cytoplasm. Greater variation between individual spectra and tissue types is therefore inevitable and leads to broader bands of significant differences between tissue types. As narrow Raman peaks are assigned to specific vibrational modes of biomolecules more easily than broad peaks, using cytology samples for Raman spectroscopic analysis may be more reliable for identifying diagnostically significant peaks for lung cancer. In addition, cytology may be more appropriate for analysis of the entire spectrum, as large spectral ranges were removed from the tissue spectra due to wax interference.

In this study, good sensitivities and specificities were achieved for the identification of lung cancer using Raman spectroscopy on formalin fixed paraffin embedded tissue sections. Our findings compare well with those of other studies and suggest that Raman spectra may potentially be used to effectively identify characteristic biochemical changes in lung cancer subtypes. Comparing the findings of this chapter and chapter 4, lung malignancies were detected with comparable sensitivities and specificities using cytological samples and tissue sections. Using Raman spectroscopy on cytological samples could be a minimally invasive and effective technique for lung cancer diagnosis, as cytological sampling methods are less invasive than biopsies for obtaining material to investigate lung lesions, and are not limited by wax interference. However, the results show that Raman spectroscopy appears to be diagnostically effective on both tissue and cytology samples.
The secondary objective of the PhD is based on investigating the suitability of Raman spectroscopy for detecting malignant thyroid lesions on FNAC samples. The following chapters aim to prove this concept, and to identify and address complications arising from using Raman spectroscopy on clinical FNAC samples. The next chapter is a proof of concept study demonstrating the ability of Raman spectroscopy to discriminate between thyroid cancer subtypes using cell line models.

References


Chapter 7: Raman spectroscopy for the preoperative diagnosis of thyroid cancer and its subtypes: an in vitro proof-of-concept study


Declan O’Dea¹,²*, Massimo Bongiovanni³*, Gerasimos P. Sykiotis⁴, Panos G. Ziros⁴, Aidan D. Meade²,⁵, Fiona M. Lyng²,⁵# and Alison Malkin¹#

¹School of Biological Sciences, Dublin Institute of Technology
²DIT Centre for Radiation and Environment Science, Focas Research Institute, DIT
³Institut of Pathology, Lausanne University Hospital, Lausanne, Switzerland
⁴Service of Endocrinology, Diabetology and Metabolism, Lausanne University Hospital, Lausanne, Switzerland
⁵School of Physics & Clinical & Optometric Sciences, Dublin Institute of Technology

* Joint first authors
# Joint senior authors

Key words: Raman spectroscopy, thyroid nodule, thyroid neoplasms, cell lines
Abstract

Objective: In 2016 there were an estimated 56,870 new cases of thyroid cancer (TC) in the United States. Fine needle aspiration cytology (FNAC) is the most safe, accurate and cost-effective method for the initial investigation of thyroid nodules. FNAC is limited by the inability to accurately diagnose malignancy in follicular-patterned lesions, and as a result 20% to 30% of cases under investigation for TC are classified as cytologically indeterminate, illustrating a problem with current FNAC procedure. Raman spectroscopy has shown promising results for the detection of many cancers however to date there has been no report on the performance of Raman spectroscopy on thyroid cytological samples. The aim of this study was to examine whether Raman spectroscopy could be used to correctly classify cell lines representing benign thyroid cells and various subtypes of TC.

Methods: A benign thyroid cell line and seven TC cell lines were prepared as ThinPrep® cytology slides and analysed with Raman spectroscopy. Principal components analysis (PCA) and linear discriminant analysis (LDA) were implemented to develop effective diagnostic algorithms for classification of Raman spectra of different TC subtypes.

Results: The spectral differences separating benign and TC cell lines were assigned to differences in the composition of nucleic acids, lipids, carbohydrates and protein in the benign and cancer cells. Good sensitivities (74 - 85%), specificities (65 - 93%) and diagnostic accuracies (71 – 88%) were achieved for the identification of TC.

Conclusion: These findings suggest that Raman spectroscopy has potential for preoperative TC diagnosis on FNAC samples.
7.1 Introduction

In 2016 there were an estimated 56,870 new cases of thyroid cancer (TC) in the United States. Women account for approximately 75% of these cases, exhibiting a substantially higher rate of incidence of the disease compared to men, but similar mortality. TC incidence has been steadily increasing in recent decades, as reported in the TC epidemiology studies. Although improved diagnostic methods, notably ultrasound (US) examination, are the likely reason for the global increase in incidence (especially of smaller nodules), recent studies show that the incidence of aggressive forms associated with higher mortality is also increasing.

Of the wide array of malignancies that are characterised as TC, papillary thyroid carcinoma (PTC) is the most common, accounting for approximately 80% of cases. Other histological types include follicular thyroid carcinoma (FTC), Hürthle cell carcinoma, poorly differentiated carcinoma, medullary thyroid carcinoma (MTC), and undifferentiated or anaplastic carcinoma (UTC or ATC, respectively). PTC and FTC emanate from thyroid follicular epithelial cells and are commonly referred to as well differentiated thyroid carcinomas (DTC). MTCs are derived from neuroendocrine cells and represent approximately 5 to 10% of TC.

The most common manifestation of TC is as a thyroid nodule. Thyroid nodules are common in the general population, with a prevalence of 2-7% when detected by palpation and 50% when searched for using US examination. Despite thyroid nodules having a high incidence rate, approximately 60%-80% are classified as benign, with only 3.5%-10% presenting as malignancies. Fine needle aspiration cytology (FNAC) is the most accurate and cost-effective method for the initial management of thyroid nodules, and its accuracy is higher when it is conducted under US guidance. FNAC is limited by the inability to accurately diagnose malignancy in follicular-patterned lesions, and as a result
more than 20% of cases under investigation for TC are classified as cytologically “indeterminate” 9–12. The indeterminate category often requires surgical resection to definitively exclude malignancy; however, since approximately 70% of these lesions are finally shown to be benign, surgery is an unnecessary procedure for the vast majority of cases 13,14.

In an attempt to overcome limitations of cytology in the detection of cancer, optical spectroscopic techniques have been investigated as adjunct or alternative approaches 15. Optical spectroscopic methods use the interaction of light with matter to provide a detailed description of the molecular composition of biological tissue, and so these techniques can be employed to detect biochemical profiles associated with health or disease.

Raman spectroscopy is a form of vibrational spectroscopy based on inelastic scattering. Raman spectra are obtained by irradiating a sample with monochromatic laser light, which interacts with molecules within the sample and induces molecular vibrations. The change in energy between the incident and inelastically scattered light is called the Raman shift. The Raman spectrum is a plot of the intensity of the scattered light versus the change in energy, given in wavenumbers, cm\(^{-1}\) 16–18. The spectrum produced by the probed molecular vibrations represents a detailed biochemical fingerprint of the cellular components.

Raman spectroscopy has shown promising results for the detection of multiple cancers including cervical, gastrointestinal, breast, brain, and lung cancer 19–22. The application of Raman spectroscopy for TC diagnosis has also been investigated in studies utilizing cell lines 23–25, and tissue sections 26–32. Harris et al. 23 applied Raman spectroscopy to two thyroid cell lines, one cancer cell line (8505C, representing UTC) and one benign cell line (Nthy-ori 3-1), to analyse the cellular differences between benign and malignant cells.
Distinct differences in the nucleic acid content of the cancer cells were revealed, and, using neural network analysis a diagnostic sensitivity of 95% was achieved for discrimination of the cancer cell line. Building on Harris’ work, Lones et al. applied Raman spectroscopy to analyse one benign thyroid cell line and four TC cell lines representing PTC, FTC, MTC and UTC. Alterations in either the molecular conformation or concentration of DNA/RNA, amide I and aromatic amino acids were found to be indicative of cancer. The authors reported good discrimination of the benign cell line from MTC and UTC, and a lower discrimination between cell lines with the same cell of origin, namely benign versus DTC cell lines.

Recent studies have also investigated the feasibility of applying Raman spectroscopy to thyroid tissue samples to detect cancer. These studies used multivariate statistical analysis of Raman spectra to achieve high diagnostic sensitivities and specificities for discrimination between healthy thyroid tissue and TC, and follicular patterned thyroid.

Application of Raman spectroscopy to thyroid FNAC has not yet been reported. As a first step, the present proof-of-concept study builds on the aforementioned studies by investigating the application of Raman spectroscopy to a benign thyroid cell line and a large number of TC cell lines prepared as ThinPrep® cytology slides, representing FNAC cytological specimens.

7.2 Materials and Methods

7.2.1 Cell cultures

Cell lines used were representative of benign thyroid cells and four subtypes of TC: one benign follicular epithelial cell line (NThy-ori 3-1), two PTC cell lines (K1 and TPC1), one FTC cell line (XTC1), two UTC cell lines (8505C and C643) and two MTC lines (CRL-1803TT and MZCRC1). The Nthy-ori 3-1 cell line [catalog no. 90011609, lot no. 239
13B007, passage no. 16 (p16)] was purchased from Sigma and used at p18. The TT cell line was obtained from the American Type Culture Collection (ATCC CRL1803 TT, lot no. 58785858, no passage information) and was used at p3. All other cell lines (K1, used at p9; TPC1, used at p29; XTC1, used at p17; 8505C used at p21; C643, used at p31; MZCRC1, used at p14) were kindly provided by Prof. Paula Soares (Institute of Molecular Pathology and Immunology of the University of Porto, Portugal). All cell lines were maintained in standard culture conditions at 37°C in a humidified incubator under 5% carbon dioxide. RPMI-1640 with 10% fetal bovine serum (FBS) and 1% Pen-Strep was the medium for most cell lines, with the following exceptions: K1 cells were cultured in DMEM/F12 medium supplemented with 10% FBS; CRL1803TT cells were cultured in F-12K Nutrient Mixture Kaighn’s Modification with 10% FBS; MZCRC1 cells were cultured in DMEM with 15% FBS. All culture media were from Invitrogen. Slides for Raman analysis were prepared using a ThinPrep® processor and air dried without final staining.

7.2.2 Raman Measurement

As described in chapter 3 section 3.1.1. Thirty cells were recorded from each slide. For each cell, a single spectrum was recorded from the cell nucleus, corresponding to the average of two accumulations of 30 seconds. Spectra were also recorded from 30 vacant locations on a ThinPrep® glass slide with identical exposure time in order to obtain representative background glass spectra. The signal to noise ratio (SNR) was calculated by dividing the difference between the peak signal at 1665 cm\(^{-1}\) and the baseline intensity (smooth region after 400 cm\(^{-1}\)) by the standard deviation of the peak signal, as described by Desroches et al.\(^{33}\). The mean SNR of all of the spectra included was 102 dB, which was calculated from a range of 77 - 159 dB.
7.2.3 Data pre-processing and analysis

All spectral data analysis was conducted using R software. Data were processed as described in chapter 3 section 3.2.3-7. The data was analysed using Principal Component Analysis (PCA) and Linear Discriminant Analysis (LDA), as described in chapter 3 section 3.2.8 and section 3.2.9 respectively. The first PCs that explained approximately 95% of the variance within the data were used in the PC-LDA classification models.

7.3 Results

7.3.1 Mean spectral analysis

Figure 7.1 depicts the mean benign thyrocyte cell line (Nthy-ori3-1) spectrum and the biological components corresponding to each spectral region. The standard deviation shows that there was minimal variation between the spectra from individual cells and this was also the case for all cell lines. The features in the TC cell line spectra that differ from the benign cell line are demonstrated by the difference spectra in Figures 7.2-4. The significant Raman peaks and their corresponding biological components are shown in Table 7.1. Spectral differences between the benign thyroid cell line and the TC cell lines corresponded to peaks relative to lipids and proteins (524, 820, 1225, 1278, 1310, 1343, 1402, 1435, 1453, 1570, 1661, 1677 cm\(^{-1}\)), carbohydrates (477, 941, 1343 cm\(^{-1}\)), phosphates (1190 cm\(^{-1}\)) and nucleic acids (780, 1330 cm\(^{-1}\)), as shown in Figures 7.2-4. The difference spectra of the FTC and PTC cell lines had similar profiles, with significant peaks at 1225, 1435, 1456 and 1690 cm\(^{-1}\) differentiating them from the benign cell line (Figure 7.2). For the UTC cell lines, consistent spectral differences were observed at 780, 1343, 1190, 1280, 1450 cm\(^{-1}\) (Figure 7.3). Compared to the benign cell line, the MTC cell lines showed consistent spectral differences at 780, 1450, 1570, 1645, and 1673 cm\(^{-1}\) (Figure 7.4).
Figure 7.1. Mean Raman spectrum of the benign cell line (NThy-ori 3-1) with the standard deviation shown as shading, and the different spectral regions with corresponding biomolecules.
Figure 7.2. Difference spectrum between (a) Nthy-ori 3-1 (benign) and FTC cell line XTC1; (b) Nthy-ori 3-1 and PTC cell line K1; (c) Nthy-ori 3-1 and PTC cell line TPC1. Shading indicates regions of the spectrum that were significantly different (p<0.05).

Figure 7.3. Difference spectrum between (a) Nthy-ori 3-1 (benign) and 8505C (UTC); and (b) Nthy-ori 3-1 and C643 (UTC). Shading indicates regions of the spectrum that were significantly different (p<0.05).
Figure 7.4. Difference spectrum between (a) Nthy-ori 3-1 (benign) and CRL1803TT (MTC); (b) Nthy-ori 3-1 and MZCRC1 (MTC). Shading indicates regions of the spectrum that were significantly different (p<0.05).

Table 7.1. Tentative peak assignments for the significant Raman bands identified by spectral variation analysis 34,35.

<table>
<thead>
<tr>
<th>Wavenumber (cm⁻¹)</th>
<th>Raman peak assignments</th>
</tr>
</thead>
<tbody>
<tr>
<td>430</td>
<td>Cholesterol ester</td>
</tr>
<tr>
<td>454</td>
<td>Ring torsion of phenyl</td>
</tr>
<tr>
<td>477</td>
<td>Polysaccharides</td>
</tr>
<tr>
<td>509</td>
<td>S-S disulphide stretching band of collagen</td>
</tr>
<tr>
<td>524</td>
<td>S-S disulphide stretching in proteins Phosphatidylserine</td>
</tr>
<tr>
<td>614</td>
<td>Cholesterol ester</td>
</tr>
<tr>
<td>645</td>
<td>C-C twisting mode of phenylalanine</td>
</tr>
<tr>
<td>700</td>
<td>Amino acid methionine</td>
</tr>
<tr>
<td>729</td>
<td>Adenine</td>
</tr>
<tr>
<td>780</td>
<td>Uracil based ring breathing mode</td>
</tr>
<tr>
<td>Wave Number</td>
<td>Description</td>
</tr>
<tr>
<td>-------------</td>
<td>-------------</td>
</tr>
</tbody>
</table>
| 820         | Protein band  
Structural protein modes of tumours  
Collagen I |
| 855         | Proline, tyrosine |
| 880         | Tryptophan |
| 941         | Skeletal modes of polysaccharides |
| 970         | Phosphate monoester groups of phosphorylated proteins and nucleic acids |
| ~1000       | Phenylalanine |
| 1100-375    | Several bands of moderate intensity belonging to amide III and other proteins |
| 1310        | CH3/CH2 twisting or bending mode of lipid/collagen |
| 1230        | Antisymmetric phosphate stretching vibration |
| 1330        | Region associated with DNA & phospholipids Collagen |
| 1343        | CH3,CH2 wagging (collagen assignment)  
Glucose  
CH residual vibrations |
| 1392        | C-N stretching, in quinoid ring-benzoid ring-quinoid ring |
| 1402        | Bending modes of methyl groups  
Amino acids aspartic and glutamic acid |
| 1436        | CH$_2$ scissoring |
| 1450        | C-H deformation bands (CH functional groups in lipids, amino acids side chains of the proteins and carbohydrates) |
| 1453        | Protein bands  
Structural protein modes of tumors |
| 1499        | C=C stretching in benzenoid ring |
| 1570        | COO-  
Aspartate and glutamate |
| 1645        | Amide I (α helix) |
| 1650        | C=C Amide I  
Protein amide I absorption |
| 1667        | Protein band  
Carbonyl stretch C=O |
| 1677        | Amide I (proteins)  
C=O stretching (lipids)  
T, G, C ring breathing modes of DNA/RNA bases |
7.3.2 Linear Discriminant Analysis

In order to evaluate the efficacy of Raman spectroscopy to accurately classify the benign and TC cell lines, PC-LDA classification models were developed using the principal components from PCA of the benign thyroid cell line and each of the TC cell lines. Table 7.2 shows the results of seven separate PC-LDA models used to classify the benign thyroid cells and each of the TC cell lines. The performance of each two way classification is shown using the sensitivity, specificity and diagnostic accuracy of each model. Diagnostic accuracy is used to describe the proportion of correctly classified spectra (true positive + true negative) among all the classified spectra (true positive + true negative + false positive + false negative) \(^{36}\). The first 5 PCs in each case were used to develop the PC-LDA classifiers. PC-LDA classification of Nthy-ori 3-1 (benign) vs. PTC, FTC and MTC cell lines yielded sensitivities of \(\geq 90\%\), specificities of \(\geq 80\%\) and accuracies of \(\geq 89\%\). Lower performance was observed for the models discriminating the benign cell line from the UTC cell lines with sensitivity, specificity, and diagnostic accuracy of 77\%, 73\% and 75\% for 8505C, and 87\%, 77\%, 82\% for C643, respectively.

Table 7.2. Performance of the PC-LDA classification model in discriminating the benign thyroid follicular cell line Nthy-ori 3-1 from each TC cell line.

<table>
<thead>
<tr>
<th>TC cell line distinguished from Nthy-ori 3-1</th>
<th>Subtype of TC</th>
<th>Sensitivity (%)</th>
<th>Specificity (%)</th>
<th>Diagnostic accuracy (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>XTC1</td>
<td>FTC</td>
<td>100</td>
<td>80</td>
<td>90</td>
</tr>
<tr>
<td>K1</td>
<td>PTC</td>
<td>90</td>
<td>87</td>
<td>89</td>
</tr>
<tr>
<td>TPC1</td>
<td>PTC</td>
<td>97</td>
<td>87</td>
<td>92</td>
</tr>
<tr>
<td>8505C</td>
<td>UTC</td>
<td>77</td>
<td>73</td>
<td>75</td>
</tr>
<tr>
<td>C643</td>
<td>UTC</td>
<td>87</td>
<td>77</td>
<td>82</td>
</tr>
</tbody>
</table>
To evaluate the ability of PC-LDA to discriminate benign cells from TC subtypes, cell lines were grouped by their TC subtype. To produce an unbiased classifier, representative spectra from each cell line were selected at random and combined by TC subtype, matching the number of spectra used for the benign cell line. As XTC1 was the only cell line representing FTC, all spectra from this cell line were used, matching the number of spectra used for the benign cell line. Table 7.3 shows the results achieved for discriminating the benign cell line from the TC cell lines grouped by subtype. Separate two-way classification models using 5 PCs each were developed to discriminate the benign cells from each TC subtype. The MTC and PTC subtypes were discriminated from the benign cell line with diagnostic accuracies of at least 82%. The UTC subtype was classified at a lower accuracy of 71%. Due to similarities in their cellular biochemistry observed in this study, the well differentiated PTC and FTC subtypes were grouped together for analysis, yielding a diagnostic accuracy of 74%.

Table 7.3. Performance of the PC-LDA classification model in discriminating the benign thyroid follicular cell line Nthy-ori 3-1 from TC cell lines grouped by TC subtype.

<table>
<thead>
<tr>
<th>TC subtype distinguished from Nthy-ori 3-1</th>
<th>Sensitivity (%)</th>
<th>Specificity (%)</th>
<th>Diagnostic accuracy (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>PTC†</td>
<td>85</td>
<td>79</td>
<td>82</td>
</tr>
<tr>
<td>UTC‡</td>
<td>77</td>
<td>65</td>
<td>71</td>
</tr>
<tr>
<td>MTC§</td>
<td>84</td>
<td>93</td>
<td>89</td>
</tr>
<tr>
<td>DTC¶</td>
<td>74</td>
<td>74</td>
<td>74</td>
</tr>
</tbody>
</table>
PTC includes K1 and TPC1. UTC includes 8505C and C643. MTC includes CRL1803TT and MZCRC1. DTC includes PTC (K1 and TPC1) and FTC (XTC1).

Table 7.4 shows the results of the two-way PC-LDA classification models used for the discrimination of the PTC and UTC cell lines, each was developed using 4 PCs. Each two-way classification model yielded sensitivities, specificities and diagnostic accuracies above 92% for each cell line classification. Grouped together, the PTC and UTC subtypes were discriminated with a diagnostic accuracy of 81%.

Table 7.4. Performance of the PC-LDA classification model in discriminating PTC from UTC cell lines.

<table>
<thead>
<tr>
<th>Comparison</th>
<th>Sensitivity (%)</th>
<th>Specificity (%)</th>
<th>Diagnostic accuracy (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>K1 vs. 8505C</td>
<td>93</td>
<td>90</td>
<td>92</td>
</tr>
<tr>
<td>K1 vs. C643</td>
<td>93</td>
<td>97</td>
<td>95</td>
</tr>
<tr>
<td>TPC1 vs. 8505C</td>
<td>97</td>
<td>93</td>
<td>95</td>
</tr>
<tr>
<td>TPC1 vs. C643</td>
<td>93</td>
<td>93</td>
<td>93</td>
</tr>
<tr>
<td>PTC† vs. UTC‡</td>
<td>80</td>
<td>82</td>
<td>81</td>
</tr>
</tbody>
</table>

Table 7.5 shows the results for the PC-LDA classification of MTC cell lines versus the PTC cell lines. Each classification is the result of separate two-way models using 4 PCs each. The models yielded specificities and diagnostic accuracies ≥93% for each cell line classification, and a diagnostic accuracy of 86% for the two-way discrimination of the MTC and PTC subtypes.

†PTC includes K1 and TPC1. ‡UTC includes 8505C and C643.
Table 7.5. Performance of the PC-LDA classification model in discriminating MTC from PTC cell lines.

<table>
<thead>
<tr>
<th>Comparison</th>
<th>Sensitivity (%)</th>
<th>Specificity (%)</th>
<th>Diagnostic accuracy (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>CRL1803TT vs. K1</td>
<td>100</td>
<td>97</td>
<td>99</td>
</tr>
<tr>
<td>MZCRC1 vs. K1</td>
<td>100</td>
<td>93</td>
<td>97</td>
</tr>
<tr>
<td>CRL1803TT vs. TPC1</td>
<td>100</td>
<td>100</td>
<td>100</td>
</tr>
<tr>
<td>MZCRC1 vs. TPC1</td>
<td>100</td>
<td>100</td>
<td>100</td>
</tr>
<tr>
<td>MTC† vs. PTC‡</td>
<td>90</td>
<td>82</td>
<td>86</td>
</tr>
</tbody>
</table>

†MTC includes CRL1803TT and MZCRC1. ‡PTC includes K1 and TPC1.

Table 7.6 shows the performance of separate two-way PC-LDA models in discriminating FTC cell lines from PTC, MTC and UTC cell lines. Discrimination of the follicular XTC1 cell line from the PTC cell lines was achieved using 4 PCs in each PC-LDA model. Diagnostic accuracies of 87% and 85% were obtained for the K1 and TPC1 cell lines respectively. The FTC cell line was discriminated from the UTC cell lines with diagnostic accuracies above 94%, and from the MTC cell lines with 100% accuracy, each model using 4 PCs. When cell lines were grouped by subtype, FTC was distinguished from MTC with an accuracy of 99% (using 4 PCs), from UTC with a diagnostic accuracy of 89% (using 3 PCs), and from PTC with 79% accuracy (using 5 PCs).

Table 7.6. Performance of PC-LDA classification of FTC lines from PTC, MTC and UTC cell lines.

<table>
<thead>
<tr>
<th>Comparison</th>
<th>Sensitivity (%)</th>
<th>Specificity (%)</th>
<th>Diagnostic accuracy (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>XTC1 vs. K1</td>
<td>97</td>
<td>77</td>
<td>87</td>
</tr>
</tbody>
</table>
XTC1 vs. TPC1  89  80  85

XTC1 vs. 8505C  100  87  94

XTC1 vs. C643  97  97  97

XTC1 vs. CRL1803TT  100  100  100

XTC1 vs. MZCRC1  100  100  100

FTC† vs. PTC‡  89  69  79

FTC† vs. MTC§  100  98  99

FTC† vs. UTC¶  97  81  89

†FTC includes XTC1. ‡PTC includes K1 and TPC1. §MTC includes CRL1803TT and MZCRC1. ¶UTC includes 8505C and C643.

Table 7.7 summarises the results achieved for separate two-way classifications of the MTC and UTC cell lines. The MTC cell lines were discriminated from the UTC cell lines with diagnostic accuracies ≥97% using 3 PCs in each model, and when grouped together, the MTC cell lines were discriminated from the UTC cell lines with an accuracy of 84% using 4 PCs.

Table 7.7. Performance of the PC-LDA classification model for discriminating the MTC from UTC cell lines.

<table>
<thead>
<tr>
<th>Comparison</th>
<th>Sensitivity (%)</th>
<th>Specificity (%)</th>
<th>Diagnostic accuracy (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>CRG1803TT vs. 8505C</td>
<td>100</td>
<td>93</td>
<td>97</td>
</tr>
<tr>
<td>MZCRC1 vs. 8505C</td>
<td>100</td>
<td>97</td>
<td>99</td>
</tr>
<tr>
<td>CRG1803TT vs. C643</td>
<td>96</td>
<td>97</td>
<td>97</td>
</tr>
</tbody>
</table>
Table 7.8 demonstrates the performance of a five-way classification model developed using the benign cell spectra and spectra from all the TC subtypes. Representative spectra from each cell line were selected at random and combined into TC subtypes as previously described. The model yielded diagnostic sensitivities from 57-100%, specificities ≥91%, and an overall diagnostic accuracy of 78% using 5 PCs. Figure 7.5 depicts the scatter plot of the linear discriminant scores of the benign spectra and spectra from each TC subtype. The FTC spectra and the PTC spectra form two close clusters using the first two discriminant functions, which can be discriminated from the benign spectra with some overlap. The MTC spectra form a cluster along the first discriminant function which is distinct from the benign spectra. The UTC subtype clusters adjacent to the benign spectra along the first discriminant function, although using only the first two discriminant functions to plot the scatter plot of the linear discriminant scores, the UTC and benign spectra overlap.

Table 7.8. Performance of the PC-LDA classification model for discriminating the benign and TC subtypes.

<table>
<thead>
<tr>
<th>Comparison</th>
<th>Sensitivity (%)</th>
<th>Specificity (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Benign</td>
<td>57</td>
<td>98</td>
</tr>
<tr>
<td>FTC†</td>
<td>94</td>
<td>92</td>
</tr>
<tr>
<td>PTC‡</td>
<td>63</td>
<td>91</td>
</tr>
<tr>
<td>UTC¶</td>
<td>79</td>
<td>96</td>
</tr>
</tbody>
</table>

†MTC includes CRL1803TT and MZCRC1. ‡UTC includes 8505C and C643.
Table 7.9 shows the performance of the PC-LDA model in discriminating the benign cell line from all of the TC subtypes combined into one group. As before, in order to produce an unbiased classifier, representative spectra from each TC subtype were selected at random and combined into one TC group, giving a total number of TC spectra that was similar to the number of benign spectra. The first 5 PCs were used to develop the two-way PC-LDA model which produced a sensitivity of 74%, specificity of 87%, and overall accuracy of 81% for discriminating benign spectra from the TC spectra.
Table 7.9. Performance of the PC-LDA classification model for discriminating the benign cell line from all TC lines.

<table>
<thead>
<tr>
<th>Comparison</th>
<th>Sensitivity (%)</th>
<th>Specificity (%)</th>
<th>Diagnostic accuracy (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Benign vs TC†</td>
<td>74</td>
<td>87</td>
<td>81</td>
</tr>
</tbody>
</table>

†TC includes representative spectra from FTC, PTC, UTC, and MTC.

7.4 Discussion

Raman microspectroscopy was performed in this study to obtain a biomolecular characterisation of eight thyroid cell lines. Spectral differences were consistently observed between the benign and TC cell lines with the strongest signals occurring at ~470, ~780, 855, 941, ~1230, 1278, 1343, 1402, 1436, 1456, 1571, 1650, 1690 and 1677 cm\(^{-1}\), representing significant differences in the molecular composition of carbohydrates, nucleic acids, lipids, protein structures and amides, as mapped by Movasaghi et al.\(^{35}\). These peaks differentiating benign thyroid cells and TC cells are in accordance with several of the characteristic TC peaks identified in previous studies using tissue samples and cell lines\(^{23,24,31}\). The consistent spectral differences between the benign and TC cell lines observed in our study may be reflective of the malignant transformation that occurs due to the carcinogenic progression of TC\(^{35,37}\). These Raman peaks represent the vibrational modes of biomolecules that are altered in quantity or conformation in the malignant cells.

The difference spectrum between the benign cell line and the FTC cell line (XTC1), shows identical strong bands to the difference spectra between the benign cell line and the PTC cell lines. The similar profiles of the difference spectra reveal a comparable biochemistry in the FTC and PTC subtypes. In comparison to the benign cell line, the PTC and FTC cell lines showed an increased peak intensity at 1225 cm\(^{-1}\) and 1435 cm\(^{-1}\), and a decreased peak intensity at 1456 cm\(^{-1}\) and 1690 cm\(^{-1}\). This indicates an increase in
the level of amide III, proteins, and nucleic acids, and a decrease in the amount of lipids and amide I in the cancer cells\textsuperscript{35}. The spectral profile of these DTC cell lines show similarities to the characteristic FTC peaks identified in a study by Teixeira \textit{et al.} using thyroid tissue\textsuperscript{30}. As both PTC and FTC emanate from epithelial follicular cells and are often even analysed as one group in terms of prognosis, it is plausible that the similarities in their biochemistry observed may be explained by the common origin of PTC and FTC\textsuperscript{5,6}.

Regarding the UTC cell lines (8505C and C643), both exhibited elevated nucleic acids (~1343 cm\textsuperscript{-1}) and polysaccharides (477 cm\textsuperscript{-1}) in comparison to the benign cell line. A lower peak intensity for amide III was observed for both cell lines in comparison to the benign cell line, which is evident by the broad band in the difference spectrum between ~1190 cm\textsuperscript{-1} and 1280 cm\textsuperscript{-1}. The 8505C cell line showed an increased peak intensity at both 780 cm\textsuperscript{-1} and 830 cm\textsuperscript{-1}, corresponding to increased nucleic acids, which is consistent with the findings of Harris \textit{et al}\textsuperscript{23}. The 8505C cell line also showed a higher peak intensity at 1450 cm\textsuperscript{-1} in comparison to the benign cell line, which suggests an increase in the level of proteins and lipids in the cancer cells\textsuperscript{35}. UTC is a heterogeneous disease at a molecular level, making it difficult to find commonalities between cases, which may explain the dissimilarities in the spectral profiles observed in this study between the two UTC cell lines\textsuperscript{38}.

The difference spectra of the benign versus the MTC cell lines (MZCRC1 and CRL1803TT) showed consistent spectral bands at 780, 1450, 1570, 1645, and 1673 cm\textsuperscript{-1} that differentiate the MTC cell lines from the benign cell line. In comparison to the benign cell line, the MTC cell lines exhibited a lower peak intensity at 1450 cm\textsuperscript{-1}, indicating lower levels of lipids, carbohydrates and amino acids in the MTC cells. In comparison to the benign cell line, both MTC cell lines exhibited greater peak intensities at 780 cm\textsuperscript{-1},
1570 cm⁻¹ and 1670 cm⁻¹, indicating an increase in the level of nucleic acids (780 cm⁻¹), aspartate and glutamate (1570 cm⁻¹), and amide I (1670 cm⁻¹). As MTCs emanate from the neuroendocrine parafollicular C cells and not from the follicular epithelial cells like the other cell lines, the distinct overlapping spectral features that differentiate the MZCRC1 and CRL1803TT cell lines from the benign cell line may be explained by the distinct origin of MTC. The main focus of this investigation was to evaluate the potential use of Raman spectroscopy as an enhanced diagnostic technique for TC. Objective comparative analysis of the spectra was achieved using linear discriminant analysis (LDA) on already PCA-transformed data sets for maximum class separation. Application of the PC-LDA algorithm to the data achieved diagnostic accuracies up to 99% for the distinction of benign and TC cell lines. The well differentiated PTC and FTC cell lines were detected with sensitivities >90% and specificities >80%, although the model yielded lower performance scores for identifying the UTC cell lines (sensitivities of 77 and 87%, specificities of 73 and 77%, and diagnostic accuracies of 75 and 82% for the 8505C and C643 cell lines, respectively). The heterogeneity and wide variability of molecular profiles of UTCs may explain the poorer performance of the model for discriminating benign cells from individual UTC cell lines as well as the grouped UTC cell lines (diagnostic accuracy of 71%).

A PC-LDA model was applied to examine whether Raman spectroscopy could be used to correctly classify cell lines representing the various subtypes of TC. Diagnostic accuracies from 92%-97% were achieved for the discrimination of the UTC cell lines from FTC and PTC cell lines. The MTC cell lines were discriminated from the PTC, FTC and UTC cell lines with classification model performance scores >93%. The high accuracy observed for the classification of the MTC cell lines may be due to the different
origin of MTCs from other TCs, as discussed previously. FTC (XTC1) and PTC (K1) cell lines were discriminated with an accuracy of 87% and the FTC cell line (XTC1) and PTC cell line (TPC1) were discriminated with an accuracy of 85%. Cell lines from the same subtype of TC were then grouped together for analysis. Diagnostic accuracies >81% were achieved for discriminating UTC from the PTC, FTC and MTC subtypes. The model also retained a high level of performance for the discrimination of the MTC subtype from PTC and FTC, with respective diagnostic accuracies of 86 and 99%. The well differentiated PTC and FTC subtypes were then discriminated with a marginally lower diagnostic accuracy of 79%, which may be explained by the common origin of PTC and FTC.

PC-LDA was then applied to investigate the ability to discriminate the benign cells and all the TC subtypes within the same model. The model yielded a poor sensitivity (57%) for the benign cells as benign spectra were misclassified as PTC. Multiple PTC spectra were classified as FTC, resulting in a low diagnostic sensitivity of 63% for PTC. The low discrimination accuracy achieved in this study between cells derived from the same origin is similar to the findings of Lones et al. The MTC and FTC subtypes were discriminated with sensitivities of 100 and 94% respectively, as observed visually in the scatter plot of the linear discriminant scores. The model also yielded a sensitivity of 79% for the UTC subtype. Multiple UTC spectra were misclassified as PTC and benign, as observed in the overlap of the clusters of these subtypes in the scatter plot of the linear discriminant scores.

The benign thyroid cell line could be discriminated from all of the TC subtypes combined into one group with a diagnostic accuracy of 81%. This is slightly higher than the 78.3% reported by Medeiros Neto et al. 31, which was achieved applying PC-LDA to Raman spectra obtained from patient thyroid tissue.
The paramount information required from the thyroid FNAC procedure is the differentiation between benign and malignant thyroid nodules. FNAC is the current favoured diagnostic method for the initial investigation of thyroid nodules, with reported sensitivity and specificity ranges of 65%-98% and 78%-100%, respectively. The results of this study show that Raman spectroscopy may also be utilized to improve the accuracy of FNAC. In our study, good sensitivities (63%-100%), specificities (65%-100%) and diagnostic accuracies (71%-99%) were achieved for the identification of TC using Raman spectroscopy on TC cell lines prepared as ThinPrep® cytology slides.

7.5 Conclusion

In summary, Raman spectroscopy together with PC-LDA classification models was applied to identify the molecular differences between a benign thyroid epithelial cell line and seven TC cell lines. The spectral differences separating benign and TC cell lines were assigned to differences in the molecular composition of nucleic acids, lipids, carbohydrates and protein in the benign and cancer cells. PC-LDA classification models discriminated between the benign and cancer cell lines as well as between the different TC subtypes with high diagnostic accuracy. This chapter showed good potential of Raman spectroscopy for TC diagnosis using cell line models prepared as ThinPrep® cytology slides. The next chapter will aim to investigate clinical thyroid FNAC samples, but first a methodology needs to be developed for sample preparation, as blood contamination in FNAC samples interferes with the cellular Raman spectrum.

References


Chapter 8: Development of a blood removal protocol suitable for thyroid FNAC samples

8.1 Introduction

Based on the previous chapter, Raman spectroscopy could be implemented as an adjunct diagnostic technique to thyroid FNAC samples to improve the limitations of current methods and provide an objective diagnosis. Applying Raman spectroscopy to biological specimens ordinarily requires minimal sample pretreatment, however as the thyroid is a very vascular organ receiving a high blood supply from the superior and inferior thyroid arteries, thyroid FNAC samples often contain blood. Blood in the FNA sample will often cover and surround the diagnostic cell groups and interfere with microscopic interpretation. To reduce interference from blood, during the ThinPrep® procedure for thyroid FNA, the patient cellular material is transferred to CytoLyt, an alcohol based lysing agent for red blood cells. However, Raman spectroscopy is a highly sensitive method and can detect trace amounts of blood.

The main limitations with Raman spectroscopy in cellular analysis include the autofluorescence of biological material creating a background in the spectra, substrate interference, and residual blood can be a limiting factor in the case of clinical specimens. The presence of these contaminants reduces the reliability of the information attained from the biological Raman spectrum and it is therefore paramount to remove any unwanted spectral signatures prior to analysis. Applying algorithms for baseline correction and substrate spectrum removal to the data reduces spectral contamination from autofluorescence and substrates. Blood contamination however must be removed from the sample by chemical means prior to acquiring Raman spectra. Previous studies explored the utilisation of hydrogen peroxide to remove blood contamination from cytological samples in preparation for Raman spectroscopic analysis. Bonnier et al...
reportedly obtained reproducible Raman spectra with no blood features by applying a hydrogen peroxide wash step on the prepared ThinPrep® slide during cytological sample preparation. The authors noted that their blood removal method was applicable to cervical liquid based cytology samples with low levels of blood contamination. Traynor et al² improved the protocol and prepared excessively bloody cervical liquid based cytology samples for Raman spectroscopic analysis. The improved blood removal was achieved following a hydrogen peroxide treatment in the collection vial, as opposed to on the prepared ThinPrep® slide. The findings of Traynor et al² and Bonnier et al³ indicate that H₂O₂ may be useful to remove blood from thyroid FNAC samples in preparation for Raman spectroscopic analysis.

The aim of this study was to develop a blood removal method, suitable for application to thyroid FNAC specimens, which allows uncontaminated Raman spectra to be obtained from the cell samples. In section A, a HeLa cell line was contaminated with human blood to replicate the complications encountered when applying Raman spectroscopy to thyroid FNAC samples. Different combinations of washes and chemical treatments were applied to the HeLa cells. The optimum blood removal protocol was investigated by assessing the presence of blood associated peaks in the cellular Raman spectra. Section B is a preliminary study investigating the use of Raman spectroscopy for the discrimination between normal and malignant thyroid FNAC samples. The blood removal method developed in section A was applied, with minor adjustments, to ‘mock’ thyroid FNAC samples taken from thyroid biopsies and Raman spectra were acquired and analysed. Finally, in section C, clinical thyroid FNAC samples were analysed and the method developed in section A and adjusted in section B was found to be not effective for these samples and required further optimisation.
8.2 Section A

8.2.1 Materials and Methods

8.2.1.1 Cell culture

The immortal cervical cancer HeLa cell line was provided by the American Type Culture Collection (ATCC) and were cultured in Roswell Park Memorial Institute medium (RPMI 1640) supplemented with 10% fetal bovine serum (FBS), and L Glutamine. The culture was maintained at 37 degrees and 5% carbon dioxide in an air humidified atmosphere. In order to prevent the cells from being exposed to a deprivation of nutrients, the cells were split into new T-75 culture flasks once 80 percent confluency was reached.

8.2.1.2 Blood contamination

The experimental component of section A was conducted in conjunction with a final year student as part of their research project in DIT (TU Dublin). This study utilised human blood, which was donated by a healthy donor. The TU Dublin School of Biological and Health Sciences have ethical approval for staff members to donate blood for teaching and educational purposes.

After culturing, the cells were transferred to falcon tubes and contaminated with human blood in an effort to resemble the conditions of clinical thyroid FNAC samples. The blood scale index, which is used in cytology laboratories, in Figure 8.1 was used as a reference to develop different levels of blood contamination. Group 0, which has no blood, was used as a negative control. Samples with group 1-3 contamination were utilised to develop the blood removal protocol. Group 1 represented low level contamination, group 2 represented medium level contamination, and group 3 represented heavy contamination. A sample with group 2 contamination which did not receive blood removal treatment was utilised as a positive control.
8.2.1.3 ThinPrep filter

To evaluate whether gynae or non-gynae filters enhanced blood removal while preparing slides, two group II samples were treated with an identical blood removal process of 5 Cytolyt washes and 3 minutes exposure to H₂O₂ on the prepared slide. One slide was prepared using a gynae ThinPrep® filter, and the other using a non-gynae ThinPrep® filter.

8.2.1.4 Development of an optimal blood removal process

Three groups representing low (Group 1), medium (Group 2) and heavy (Group 3) blood contamination were utilised. To analyse whether H₂O₂ had a greater blood removal effect when applied at various stages of sample preparation, the experiments shown in Tables 8.1 and 8.2 were designed. Table 8.1 demonstrates the combination of washes and H₂O₂ treatments which were applied to each sample to remove contaminating blood. After each wash in CytoLyt, the sample was centrifuged, the CytoLyt was decanted and the cell
pellet vortexed to mix. After the appropriate number of CytoLyt washes, the cell pellet was transferred directly into PreservCyt for step 2. H₂O₂ was applied to the cells in step 3 after they were fixed to the ThinPrep® slide. Sample 10 and 11 represent the negative and positive controls respectively.

Table 8.1. Stepwise protocol for blood removal with H₂O₂ treatment on the slide.

<table>
<thead>
<tr>
<th>Blood contamination</th>
<th>Step 1: No. of CytoLyt washes</th>
<th>Step 2: ThinPrep filter</th>
<th>Step 3: Time treated with H₂O₂ (minutes)</th>
</tr>
</thead>
<tbody>
<tr>
<td>(Group 1) Sample 1</td>
<td>3</td>
<td>Non-gyn</td>
<td>3</td>
</tr>
<tr>
<td>Sample 2</td>
<td>3</td>
<td>Non-gyn</td>
<td>5</td>
</tr>
<tr>
<td>Sample 3</td>
<td>4</td>
<td>Non-gyn</td>
<td>3</td>
</tr>
<tr>
<td>(Group 2) Sample 4</td>
<td>4</td>
<td>Non-gyn</td>
<td>5</td>
</tr>
<tr>
<td>Sample 5</td>
<td>5</td>
<td>Non-gyn</td>
<td>3</td>
</tr>
<tr>
<td>Sample 6</td>
<td>5</td>
<td>Non-gyn</td>
<td>5</td>
</tr>
<tr>
<td>(Group 3) Sample 7</td>
<td>5</td>
<td>Non-gyn</td>
<td>5</td>
</tr>
<tr>
<td>Sample 8</td>
<td>6</td>
<td>Non-gyn</td>
<td>3</td>
</tr>
<tr>
<td>Sample 9</td>
<td>6</td>
<td>Non-gyn</td>
<td>5</td>
</tr>
<tr>
<td>(Group 0) Sample 10</td>
<td>-</td>
<td>Non-gyn</td>
<td>3</td>
</tr>
<tr>
<td>(Group 2) Sample 11</td>
<td>-</td>
<td>Non-gyn</td>
<td>-</td>
</tr>
</tbody>
</table>

Table 8.2 demonstrates the combination of washes and H₂O₂ treatments that were applied to each sample to remove contaminating blood. After each wash in CytoLyt, the sample was centrifuged, the CytoLyt was decanted and the cell pellet vortexed to mix. After the appropriate number of CytoLyt washes, the CytoLyt was decanted and H₂O₂ was added into the tube directly to the cell pellet for the time indicated. The H₂O₂ was then diluted.
with PreservCyt and the cells were transferred directly to a PreservCyt vial for step 3. Sample 10 and 11 represent the negative and positive controls respectively.

Table 8.2. Stepwise protocol for blood removal with H\textsubscript{2}O\textsubscript{2} treatment in the tube.

<table>
<thead>
<tr>
<th>Blood contamination</th>
<th>Step 1: No. of CytoLyt washes</th>
<th>Step 2: Time treated with H\textsubscript{2}O\textsubscript{2}</th>
<th>Step 3: ThinPrep filter</th>
</tr>
</thead>
<tbody>
<tr>
<td>(Group 1) Sample 1</td>
<td>3</td>
<td>3</td>
<td>Non-gyn</td>
</tr>
<tr>
<td>Sample 2</td>
<td>3</td>
<td>5</td>
<td>Non-gyn</td>
</tr>
<tr>
<td>Sample 3</td>
<td>4</td>
<td>3</td>
<td>Non-gyn</td>
</tr>
<tr>
<td>(Group 2) Sample 4</td>
<td>4</td>
<td>5</td>
<td>Non-gyn</td>
</tr>
<tr>
<td>Sample 5</td>
<td>5</td>
<td>3</td>
<td>Non-gyn</td>
</tr>
<tr>
<td>Sample 6</td>
<td>5</td>
<td>5</td>
<td>Non-gyn</td>
</tr>
<tr>
<td>(Group 3) Sample 7</td>
<td>5</td>
<td>5</td>
<td>Non-gyn</td>
</tr>
<tr>
<td>Sample 8</td>
<td>6</td>
<td>3</td>
<td>Non-gyn</td>
</tr>
<tr>
<td>Sample 9</td>
<td>6</td>
<td>5</td>
<td>Non-gyn</td>
</tr>
<tr>
<td>(Group 0) Sample 10</td>
<td>-</td>
<td>3</td>
<td>Non-gyn</td>
</tr>
<tr>
<td>(Group 2) Sample 11</td>
<td>-</td>
<td>-</td>
<td>Non-gyn</td>
</tr>
</tbody>
</table>

8.2.1.5 Slide preparation of cytological specimens (ThinPrep method)

As described in chapter 3 section 3.1.2.

8.2.1.6 Raman Measurement

As described in chapter 3 section 3.2.1.

For each cell, a single spectrum was recorded from the cell nucleus, corresponding to the average of 2 accumulations of 30 seconds. A total of 15 spectra were acquired for each sample. Spectra were also recorded from 30 vacant locations on a ThinPrep® glass slide with identical exposure time in order to get representative background glass spectra.
8.2.1.7 Data mining and spectral analysis

Spectral data analysis was conducted using R software. Data were processed as described in chapter 3 section 3.2.3-7. The mean spectra of the cells for each treatment were then comparatively analysed.

8.2.2 Results

8.2.2.1 ThinPrep® filter influence on blood removal

Figure 8.2 (a) and (b) are both treated group 2 which received an identical blood removal protocol, differing only by the type of ThinPrep® filter (gynae or non-gynae) used to prepare the slides. The comparable quantities of residual blood observable on the slides indicate that gynae filters do not provide superior blood removal than non-gynae filters. For this reason, non-gynae filters were utilised for the remainder of the investigation.

Figure 8.2. Group 2 samples which received 5 cytolyt washes and 3 minutes H$_2$O$_2$ on the slide. (a) Prepared using a gynae filter for ThinPrep®, and (b) Prepared using a non-gynae filter. The red arrows highlight areas with an accumulation of blood debris. (a) and (b) are at x400 magnification.
8.2.2.2 Slide analysis - H₂O₂ treatment on slide

Figures 8.3 to 8.5 show the unstained HeLa cells after receiving the assigned blood removal treatments shown in Table 8.1. Residual blood is highlighted by red arrows to demonstrate the level of blood on the ThinPrep® slides. Figure 8.3 depicts a comparison of the negative control (sample 10) and the group 1 cells corresponding to sample 2. The comparable presentation of the 2 slides indicates that microscopic traces of blood have been successfully eliminated from sample 2. The combination of 3 CytoLyt washes and 5 minutes hydrogen peroxide in sample 2 was effective, producing a similar ThinPrep® to the negative control.

![Figure 8.3](image)

Figure 8.3. (a) Group 0 cells representing a negative control for the presence of blood (sample 10). (b) Group 1 which received blood removal treatment (sample 2). The comparable presentation of the slides indicates that microscopic traces of blood have been successfully removed from sample 2. (a) and (b) are at x200 magnification.

The group 2 and group 3 samples showed similar patterns of blood contamination on the slides, thus one sample from each group is depicted as a representative for that group. Accumulations of blood debris are visible on the group 2 and 3 slides, shown in Figures
8.4 and 8.5 respectively. Although the group 2 and 3 slides received blood removal treatment according to Table 8.1, they exhibit similarities to the positive control as blood is still microscopically visible, suggesting that the blood removal processes applied did not effectively remove the contaminating blood.

Figure 8.4. (a) A positive control for the presence of blood (sample 11). (b) Group 2 which received blood removal treatment (sample 6). The red arrows highlight areas with an accumulation of blood debris. (a) and (b) are at x200 magnification.
Figure 8.5. (a) A positive control for the presence of blood (sample 11). (b) Group 3 which received blood removal treatment (sample 8). The red arrows highlight areas with an accumulation of blood debris. (a) and (b) are at x200 magnification.

8.2.2.3 Slide analysis – H$_2$O$_2$ treatment in falcon tube

Figures 8.6 to 8.8 show the unstained HeLa cells after receiving the assigned blood removal treatments shown in Table 8.2. Residual blood is highlighted to demonstrate the level of blood on the ThinPrep® slides. Figure 8.6 depicts a comparison of the negative control (sample 10) and the group 1 cells corresponding to sample 3. The comparable presentation of the 2 slides indicates that microscopic traces of blood have been successfully eliminated from sample 3. The combination of 4 CytoLyt washes and 3 minutes hydrogen peroxide in sample 3 was sufficient to produce an identical ThinPrep® to the negative control.
Figure 8.6. (a) Group 0 cells representing a negative control for the presence of blood (sample 10). (b) Group 1 which received blood removal treatment (sample 3). The comparable presentation of the slides indicates that microscopic traces of blood have been successfully removed from sample 3. (a) and (b) are at x400 magnification.

As most group 2 and group 3 samples showed similar patterns of blood contamination on the slides, only one sample from each group is depicted. Residual blood can be observed coating the group 2 and group 3 cells, shown in Figures 8.7 and 8.8 respectively, and accumulations of blood debris on the slides are visible. Although the group 2 and 3 slides shown in Figures 8.7 and 8.8 received blood removal treatment, they exhibit similarities to the positive control as blood is still microscopically visible, suggesting that the blood removal processes applied were not successful in eradicating the contaminating blood.
Figure 8.7. (a) A positive control for the presence of blood (sample 11). (b) Group 2 which received blood removal treatment (sample 5). The red arrows highlight areas with an accumulation of blood debris. (a) and (b) are at x400 magnification.

Figure 4.8. (a) A positive control for the presence of blood (sample 11). (b) Group 3 which received blood removal treatment (sample 7). The red arrows highlight areas with an accumulation of blood debris. (a) and (b) are at x400 magnification.
8.2.2.4 Mean spectral analysis
Blood was successfully removed from group I samples and remained in the group 2 and 3 samples at a microscopic level. To evaluate whether blood interfered with the cellular Raman signals of the treated group 1-3 cells, Raman spectroscopic analysis was applied to the slides. Figure 8.9 depicts the mean HeLa cell spectrum with no blood contamination (group 0), utilized as a negative control. The positive control, which is a spectrum taken from blood, is depicted in Figure 8.10.

![Graph showing mean spectrum of a HeLa cell line.](image)

Figure 8.9. Mean spectrum of a HeLa cell line. The cells were not contaminated with blood and are used as a reference spectrum or negative control.
Figure 8.10. Spectrum of blood representing the positive control.

The significant Raman peaks and their corresponding biological components are shown in Table 8.3, and the peak assignments for the spectrum of blood are shown in Table 8.4.

Table 8.3. Tentative peak assignments for the significant Raman bands identified by spectral variation analysis.

<table>
<thead>
<tr>
<th>Wavenumber (cm$^{-1}$)</th>
<th>Raman peak assignments</th>
</tr>
</thead>
<tbody>
<tr>
<td>481</td>
<td>DNA</td>
</tr>
<tr>
<td>524</td>
<td>S-S disulphide stretching in proteins</td>
</tr>
<tr>
<td></td>
<td>Phosphatidylserine</td>
</tr>
<tr>
<td>678</td>
<td>Ring breathing modes in the DNA bases</td>
</tr>
<tr>
<td>720-780</td>
<td>DNA</td>
</tr>
<tr>
<td>1453</td>
<td>Guanine/Adenine</td>
</tr>
<tr>
<td></td>
<td>CH deformation (proteins, lipids, carbohydrates, DNA and RNA)</td>
</tr>
<tr>
<td>1575</td>
<td>Ring breathing modes of Guanine/Adenine</td>
</tr>
<tr>
<td>1667</td>
<td>Amide I (proteins)</td>
</tr>
</tbody>
</table>
Table 8.4. Peak assignments for the significant Raman bands of the spectrum of blood.

<table>
<thead>
<tr>
<th>Wavenumber (cm⁻¹)</th>
<th>Blood component</th>
</tr>
</thead>
<tbody>
<tr>
<td>746</td>
<td>Haemoglobin/ hemin/ proto porphyrin</td>
</tr>
<tr>
<td>995</td>
<td>Haemoglobin/ hemin</td>
</tr>
<tr>
<td>1126</td>
<td>Haemoglobin/ hemin/ proto porphyrin</td>
</tr>
<tr>
<td>1171</td>
<td>Haemoglobin/ hemin/ proto porphyrin</td>
</tr>
<tr>
<td>1220</td>
<td>Haemoglobin/ hemin/ proto porphyrin</td>
</tr>
<tr>
<td>1304</td>
<td>Haemoglobin/ hemin</td>
</tr>
<tr>
<td>1338</td>
<td>Haemoglobin/ hemin/ proto porphyrin</td>
</tr>
<tr>
<td>1357</td>
<td>Haemoglobin/ hemin/ proto porphyrin</td>
</tr>
<tr>
<td>1394</td>
<td>Haemoglobin/ hemin</td>
</tr>
<tr>
<td>1430</td>
<td>Haemoglobin/ hemin</td>
</tr>
<tr>
<td>1550</td>
<td>Haemoglobin/ hemin/ proto porphyrin</td>
</tr>
<tr>
<td>1585</td>
<td>Haemoglobin/ hemin/ proto porphyrin</td>
</tr>
</tbody>
</table>

8.2.2.5 Spectral variation analysis - H₂O₂ treatment on slide

The Raman spectra obtained from the group 1-3 samples exposed to H₂O₂ treatment on the slide (Table 8.1) indicated burning. As there are no distinct biological features in the spectra from these samples, spectral variation analysis was not conducted. Figure 8.11 depicts the typical spectral profile obtained for the group 1-3 samples exposed to H₂O₂ treatment on the slide.
Figure 8.11. Raman spectrum taken from a group 1 sample indicating burning. Cells were exposed to \( \text{H}_2\text{O}_2 \) on the ThinPrep slide. There are no distinct biological features in the spectrum.

8.2.2.6 Spectral variation analysis - \( \text{H}_2\text{O}_2 \) treatment in falcon tube

Raman spectra could be obtained from the group 1-3 samples exposed to \( \text{H}_2\text{O}_2 \) treatment in the tube (Table 8.2). The difference spectra depicted in Figures 8.12-8.14, are the features in the cell spectra from groups 1-3 which differ from the mean HeLa spectrum. The difference spectra between the mean HeLa cell spectrum and the group 1 cells in Figure 8.12 (a) and (b) have similar profiles to the spectrum of blood (Figure 8.10) with prominent peaks at \( \sim 750, \sim 1220, 1305, 1357, 1430, 1550, \) and \( 1585 \text{ cm}^{-1} \), indicating that the blood removal methods applied to samples 1 and 2 were not effective. Blood is not responsible for the difference spectra between the mean HeLa cell spectrum and sample 3, shown in Figure 8.12 (c). The blood removal method applied to sample 3 is an effective protocol to eradicate blood from cell samples with group 1 level contamination.
significant spectral differences in Figure 8.12 (c) correspond to nucleic acids (481 and 730 cm\(^{-1}\)), and demonstrate that blood is not contaminating the spectrum, as spectral variations are due to cellular biochemistry. Significant spectral differences between the mean HeLa cell spectrum and cells from group 2 and 3 were revealed at \(~750, \sim1220, 1305, 1357, 1430, 1550, 1585\) cm\(^{-1}\), as shown in Figures 8.13 and 8.14. These spectral peaks are identical to the prominent peaks in the positive control spectrum, indicating that the spectral contaminants convey the presence of blood. The profile of the difference spectra between the mean HeLa cell spectrum and sample 6 (Figure 8.13(c)) displays prominent peaks corresponding to cellular components such as nucleic acids (783, 1575, 1453 cm\(^{-1}\)) and proteins (1667 cm\(^{-1}\)), as well as peaks associated with blood contamination shown in Figure 8.10, suggesting that although spectral information could be obtained from the cellular components, blood contamination is still evident in the spectra. The blood removal processes applied in this study to tubes with group 2 and 3 contamination did not effectively remove blood from the samples.

![Figure 8.12](image)

Figure 8.12. Difference spectra of the HeLa cells in group 1. (a) corresponds to sample 1 in Table 8.2 (b) the difference spectra for sample 2, (c) the difference spectra for sample 3, this sample was subjected to 4 CytoLyt washes and 3 minutes exposure to hydrogen peroxide.
Figure 8.13. Difference spectra of the HeLa cells in group 2. (a) corresponds to sample 4 in Table 8.2 (b) the difference spectra for sample 5, (c) the difference spectra for sample 6, this sample was subjected to 5 CytoLyt washes and 5 minutes exposure to hydrogen peroxide.

Figure 8.14. Difference spectra of the HeLa cells in group 3 (a) corresponds to sample 7 in Table 4.2 (b) the difference spectra for sample 8, (c) the difference spectra for sample 9.

8.2.2.7 Blood removal protocol for group 1 samples

The difference spectra from the mean HeLa cell spectrum and sample 3 (Figure 8.12) indicates that blood has been successfully removed. The combination of CytoLyt washes and H₂O₂ treatment in the falcon tube applied to sample 3 (Table 8.2) effectively removed
blood from the group 1 sample. The working method applied to sample 3, described in
detail below, may be suitable for preparing group I thyroid FNAC samples for Raman
spectroscopic analysis.

**Working method for group 1 samples:**

1. Centrifuge sample at 1200 rpm for 5 minutes to pellet the cells.
2. Resuspend pellet in CytoLyt and vortex.
3. Centrifuge at 1200 rpm for 5 minutes.
4. Repeat steps 2 and 3 until blood is no longer visible (or for a total of 4 washes).
5. Remove CytoLyt from the cell pellet, vigorously vortex the pellet and pass through a
pipette to break up clumps of cells.
6. Add 1 mL of hydrogen peroxide (30%) to the pellet. Vortex and allow to stand for 3
minutes, inverting occasionally.
6. Dilute sample with 15 mL of PreservCyt.
7. Centrifuge at 1200 rpm for 5 minutes.
8. Transfer cells to PreservCyt vial for ThinPrep®.

**8.2.3 Discussion**

Blood contamination interferes with the Raman spectrum and inhibits reliable analysis.
To overcome this problem a blood removal protocol was developed to prepare thyroid
FNAC samples for single cell analysis with Raman spectroscopy. Both gynae and non-
gynae filters were utilised to investigate their ability to disaggregate and remove blood
from the samples. Inspection of the slides revealed that the type of ThinPrep® filter used
did not impact the amount of blood observed on the prepared slides. Accordingly, the
appropriate non-gynae filters were utilised for the study.

Subsequent to implementing the blood removal treatments, the prepared slides were
inspected microscopically to evaluate the effectiveness of each method at removing
blood. Blood remained microscopically visible on all group 2 and 3 slides, showing
similarities to the positive control. Subsequent to CytoLyt washing, trace amounts of blood remained in the cell pellets of the group 2 and 3 samples (samples 4-9). Although small amounts of blood were visible in the cell pellets, additional CytoLyt washes were not added in order to keep the working method applicable in a clinical setting. This may explain the aggregations of blood observed on the prepared slides. The group 1 samples treated with H₂O₂ in the tube, and the group 1 samples treated with H₂O₂ on the slides, both exhibited similarities to the negative control, indicating the complete removal of blood at a microscopic level. No blood was visible in the cell pellets of the group 1 samples during processing.

Good quality spectra could not be obtained from any of the samples which were exposed to H₂O₂ on the slide. Despite the microscopic removal of blood from the group 1 samples, the Raman spectra indicated burning. It is plausible that the H₂O₂ could not act upon all of the remaining blood coating the cells as fixing the cells prior to the application of H₂O₂ may conceal portions of the cell membranes. Uncontaminated Raman spectra could not be obtained from cells which were exposed to H₂O₂ after preparation of the ThinPrep® slide. This process was therefore not suitable for preparing thyroid FNAC samples for Raman spectroscopic in this study.

Applying the H₂O₂ to the cell pellet after the CytoLyt washes while they were still in the tube allowed the chemical to act on a greater surface area of the cells, increasing the probability of removing blood and enabling Raman spectra to be obtained. In the case of the group 1 samples which were treated with H₂O₂ in the tube, the difference spectra between the mean HeLa cell spectrum and samples 1 and 2 show significant peaks at ~750, ~1220, 1305, 1357, 1430, 1550, 1585 cm⁻¹, which are consistent with the spectral profile of blood. This indicates that the methods applied to samples 1 and 2 were not effective for removing blood from samples with low level contamination. However, the
The difference spectrum between the mean HeLa cell spectrum and sample 3 demonstrates the complete removal of blood from the sample. The significant spectral differences between the mean HeLa cell spectrum and sample 3 observed at 481 and 730 cm\(^{-1}\) correlates to nucleic acids and may be explained by variations in the cell cycle, as although many carcinogenic processes dysregulate cell cycle inhibition, the lack of available nutrients may impact the rate of proliferation, and therefore a cancer cell line grown in coincident cultures may have different proportions of proliferating cells with differing protein and DNA content \(^7,8\). The difference in proliferating status of cell lines is observable in the Raman spectrum, therefore the variance observed between the mean HeLa cell spectrum and the mean sample 3 spectrum may be attributed to variations in cell cycle progress \(^9\), explaining the significant spectral differences observed at 481 and 730 cm\(^{-1}\). This suggests that the combination of CytoLyt washes and H\(_2\)O\(_2\) treatment in the tube which was applied to sample 3 may be utilised as a new working method to prepare group 1 thyroid FNAC samples for Raman spectroscopy. This finding also indicates that the stage at which H\(_2\)O\(_2\) is applied to the sample during preparation is crucial, as good quality spectra with no blood features could only be obtained from cells receiving H\(_2\)O\(_2\) treatment in the tube.

The difference spectra for the group 2 and group 3 samples (samples 4-9) which received H\(_2\)O\(_2\) treatment in the tube indicate that the methods utilised in this study did not effectively remove blood contamination. The significant peaks in the difference spectra at ~750, ~1220, 1305, 1357, 1430, 1550, 1585 cm\(^{-1}\) are identical to the spectral profile of the positive control, signifying that residual blood is coating the cell membranes and swamping the cellular signals. The profile of the difference spectra between the mean HeLa cell spectrum and sample 6 displays prominent peaks corresponding to cellular components including nucleic acids (783, 1575, 1453 cm\(^{-1}\)) and proteins (1667 cm\(^{-1}\)), as
well as peaks associated with blood contamination. The presence of peaks corresponding
to cellular components in the Raman spectra are promising, signifying the partial removal
of contaminating blood, although further analysis of the spectra obtained from sample 6
would be unreliable as blood contamination is still evident in the spectra. Further
investigation could explore implementing additional washes or an increased H₂O₂
exposure time to effectively eradicate blood from group 2 samples.

Despite implementing an array of washes and chemical treatments, persistent blood
contamination inhibits reliable and reproducible data to be obtained from group 2 and
group 3 samples (samples 4-9). Blood was successfully removed from group 1 samples
with an optimised protocol of 4 CytoLyt washes and 3 minutes exposure to H₂O₂ in the
tube. This removes blood from the cell surface and allows spectral information to be
obtained from the intracellular biochemistry. The blood removal protocol developed in
this study is only applicable to group 1 samples and may be suitable for preparing group
1 thyroid FNAC samples for Raman spectroscopic analysis. Further investigation is
needed to develop a method for preparing group 2 and 3 samples for Raman spectroscopy.

8.3 Section B - Use of Raman spectroscopy for the discrimination between
normal and malignant FNAC samples prepared from resected thyroid
specimens

The primary aim of this study was to assess the efficacy of the blood removal method
developed in section A in preparing ‘mock’ thyroid FNAC samples for Raman
spectroscopic analysis. In addition, the study aims to examine the ability of Raman
spectroscopy to discriminate normal and malignant thyroid patient samples.
8.3.1 Methods

8.3.1.1 Study subjects and sample collection

This study had approval from the research ethics committee at Lausanne University Hospital, Switzerland. Thyroid resections were collected from patients under investigation for TC. FNA samples were subsequently prepared from the resected thyroid specimens. Thyroid cancer was confirmed through final clinical diagnosis and final histology was recorded. The study included samples from 6 patients with thyroid cancer. Patients one to six had different histological subtypes of TC, listed consecutively from patient one to six as follows; follicular adenoma, papillary carcinoma, papillary carcinoma, medullary carcinoma, follicular adenoma, oncocytic follicular carcinoma. Cytological specimens of both tumour and benign tissue were obtained from each patient. To yield sufficient cellular material for Raman spectroscopic analysis, patient samples were combined as depicted in Figure 8.15. Tumour samples from patients one to four were combined and the new combined sample is referred to in this study as T1. The tumour samples from patients five and six were also combined into one new sample referred to as T2. Similarly, benign samples from patients one to four were combined to form a new sample referred to in this study as N1, and the benign samples from patients five and six were combined into one new sample referred to as N2.
8.3.1.2 Blood removal from sample vials

The blood removal protocol developed in section A was applied to the FNA samples with two minor adjustments, as shown with the adjustments highlighted below. Firstly, to ensure clumps of cells were dispersed so that H₂O₂ could act on the entire cell surface, the samples required vigorous vortexing before exposing the cells to H₂O₂. Secondly, the ‘mock’ FNAC samples were exposed to H₂O₂ for 90 seconds to eradicate blood contamination.

*Working method for FNAC samples prepared from resected thyroid sections:*

1. Centrifuge sample at 1200 rpm for 5 minutes to pellet the cells.
2. Resuspend pellet in CytoLyt and vortex.
3. Centrifuge at 1200 rpm for 5 minutes.

4. Repeat steps 2 and 3 until blood is no longer visible (or for a total of 4 washes).

5. Remove CytoLyt from the cell pellet, vigorously vortex the pellet and pass through a pipette to break up clumps of cells. (First adjustment)

6. Add 1 mL of hydrogen peroxide (30%) to the pellet. Vortex and allow to stand for 90 seconds, inverting occasionally. (Second adjustment)

6. Dilute sample with 15 mL of PreservCyt.

7. Centrifuge at 1200 rpm for 5 minutes.

8. Transfer cells to PreservCyt vial for ThinPrep®.

8.3.1.3 Data pre-processing and analysis

All spectral data analysis was conducted using R software. Data were process as described in chapter 3 section 3.2.3-.7. The data was analysed using Principal Component Analysis (PCA) and Linear Discriminant Analysis (LDA), as described in chapter 3 section 3.2.8 and section 3.2.9 respectively. The first PCs that explained approximately 95% of the variance within the data were used in the PC-LDA classification models.

8.3.2 Results and discussion

8.3.2.1 Mean spectral analysis

The blood removal method completely eradicated blood contamination. Combining the sample vials to yield sufficient cellular material eliminated the possibility to classify individual cases, although classifiers can still be developed to discriminate benign and cancerous cells. The mean Raman spectra of benign and cancerous cells are displayed in Figures 8.16 and 8.17 with the difference spectra. Spectral analysis shows clear separation between the benign and malignant cells. Table 8.5 collates each Raman peak with their tentative corresponding biological component. The spectral profiles in Figure 8.16 indicate a higher level of lipids (1439 cm\(^{-1}\)), proteins/lipids (1445 cm\(^{-1}\)), proteins (1664 cm\(^{-1}\)),
cm⁻¹), and increased amino acids aspartic and glutamic acid (1400 cm⁻¹) in the tumour cells (T1) in comparison to the benign cells (N1). The difference spectrum in Figure 8.16 also demonstrates decreased peak intensities at 524-600, 785, 1484, and 1574 cm⁻¹ in the cancer cells, which is correlated to lower levels of nucleic acids and proteins.

![Figure 8.16](image)

Figure 8.16. (a) Mean spectra of the benign thyroid cells (N1) and the thyroid cancer cells (T1), (b) the difference spectrum between benign thyroid cells and thyroid cancer cells with the statistically significant Raman peaks highlighted in grey. Shading indicates regions of the spectrum that were significantly different (p<0.05).

The difference spectrum in Figure 8.17 demonstrates decreased peak intensity at 929 and 970 cm⁻¹ in the tumour cells, which is correlated with a lower level of phosphorylated proteins and nucleic acids. The mean cancer cell spectrum (T2) also indicates an increase in the level of tryptophan in comparison to the benign cells, evident by the peak at ~1360 cm⁻¹. The amide I band at 1664 cm⁻¹, which corresponds to stretching vibrations in
proteins, is decreased in the tumour cell spectrum. The decreased protein content evident in the tumour spectrum is most likely due to pathogenic mechanisms of TC which result in the downregulation of tumour suppressor proteins.  

Figure 8.17. (a) Mean spectra of the benign thyroid cells N2 and the thyroid cancer cells (T2), (b) the difference spectrum between benign thyroid and thyroid cancer. Shading indicates regions of the spectrum that were significantly different (p<0.05).

The spectral variations between normal and tumour cells may reflect the molecular alterations occurring due to carcinogenic progression. The increased lipid content in the tumour cells evident by the peak at 1439 cm\(^{-1}\) may indicate dysregulation of lipid metabolism, which is a common characteristic in malignancies including anaplastic TC. The reduced nucleic acid content in the tumour cell spectra is evident by the lower peak intensity at 785 cm\(^{-1}\) in comparison to the benign Raman spectra. Genomic deletions are a type of genomic instability which are a typical component in cancers, and have been shown to be a feature in some instances of TC. This may explain the
alteration in nucleic acid content of cancer cells observed in this study. The tumour cells are characterised by a lower level of nucleic acids and an increase of amino acids (aspartic and glutamic acid) in comparison to the benign cells. The decreased peak intensities at 929 and 970 cm\(^{-1}\) in the tumour cells indicate a lower level of proline / valine, phosphorylated proteins, and nucleic acids in the tumour cells. Protein phosphorylation regulates pathways involved in the cell cycle, the diminished rate of phosphorylated proteins in the tumour cells may indicate a mechanism for thyroid cancer progression 19.

Table 8.5. Tentative peak assignments for the significant Raman bands identified by spectral variation analysis 6

<table>
<thead>
<tr>
<th>Wavenumber (cm(^{-1}))</th>
<th>Raman peak assignments</th>
</tr>
</thead>
<tbody>
<tr>
<td>490</td>
<td>Glycogen</td>
</tr>
<tr>
<td>524-600</td>
<td>Disulphide stretching in proteins Phosphatidylserine Phosphatidylinositol Cholesterol</td>
</tr>
<tr>
<td>725</td>
<td>Adenine ring breathing mode of RNA/DNA</td>
</tr>
<tr>
<td>785</td>
<td>DNA</td>
</tr>
<tr>
<td>880-920</td>
<td>Tryptophan Saccharides</td>
</tr>
<tr>
<td>929</td>
<td>C-C stretching of amino acids proline &amp; valine/saccharides</td>
</tr>
<tr>
<td>970</td>
<td>Phosphate monoester groups of phosphorylated proteins/nucleic acids</td>
</tr>
<tr>
<td>1247</td>
<td>Amide III</td>
</tr>
<tr>
<td>1335</td>
<td>CH(_3)CH(_2) wagging, collagen (protein assignment) CH(_3)CH(_2) wagging, nucleic acid</td>
</tr>
<tr>
<td>~1360</td>
<td>Tryptophan</td>
</tr>
<tr>
<td>~1370</td>
<td>Saccharide band Ring breathing modes of RNA/DNA (T,A,G)</td>
</tr>
<tr>
<td>1400</td>
<td>NH in-plane deformation</td>
</tr>
</tbody>
</table>
Amino acids aspartic & glutamic acid

| 1439 | CH$_2$ deformation |
| 1445 | Bending modes of proteins and phospholipids |
| 1484 | Amide II (Coupling of CN stretching & in-plane bending of the N-H group) |
| 1574 | COO$^-$ |
| Nucleic acids |
| 1600-1670 | Amide I band of proteins due to C=O stretching |

8.3.1.2 Principal component analysis

The distribution of the PCs was visualised by plotting them in scatter plots representing the component space. The plots in Figure 8.18 illustrate the discrimination of benign thyroid cells from thyroid cancer cells. It is evident from the PC scores plot in Figure 8.18 (a) that PC2 is responsible for the discrimination of the benign (N1) and cancer spectra (T1). The discrimination of the N2 and T2 cells was achieved using PC5, shown in Figure 8.18 (b).
Figure 8.18. PCA scores plot of benign thyroid cells (orange) and tumour cells (blue) attained from FNAC samples. (a) Scatter plot of T1 (blue) and N1 (orange) using PC2 and PC3. (b) Scatter plot of T2 (blue) and N2 (orange) using PC2 and PC5.

As PC2 is responsible for the discrimination of the benign and cancer spectra in Figure 8.18 (a), the loading profile was attained and is displayed in Figure 8.19. The loading profile of PC2 shows that peak intensity is highest at 783, 1335, 1371, 1484, 1571, and 1611 cm$^{-1}$, indicating that these peaks have the highest weights for the PCA discrimination of benign and cancerous thyroid cells. These peaks are assigned to nucleic acids and proteins, shown in Table 8.5. The tumour cells have a decreased level of nucleic acids (783, 1371, 1574 cm$^{-1}$) and proteins / COO$^-$ (1335, 1371, 1484, 1571 cm$^{-1}$), and an increase in the level of lipids (1439 cm$^{-1}$) and amide I (1664 cm$^{-1}$) proteins/lipids (1445 cm$^{-1}$), and amino acids aspartic and glutamic acid (1400 cm$^{-1}$) in comparison to the benign cells.

Figure 8.19. Loadings plot of PC2 discriminating the benign and cancer cells in Figure 8.18 (a), Peaks with higher intensity highlight the prominent differences between benign and cancerous thyroid cells.
As the benign and thyroid cancer spectra in Figure 8.18 (b) were discriminated by PC5, the loading profile of the PC is displayed in Figure 8.20. The prominent peaks discriminating the benign and cancer cells correspond to a decrease in the amino acids proline / valine (929 cm\(^{-1}\)) and amide I (1670 cm\(^{-1}\)), and an increase of tryptophan (~1360 cm\(^{-1}\)) in the tumour cells in comparison to the benign cells.

![Loading profile of PC5](image)

Figure 8.20. Loadings plot of PC5 discriminating the benign and cancer cells in Figure 8.18 (b).

8.3.1.3 Linear discriminant analysis classification

Linear discriminant analysis (LDA) was used on the PCs from PCA to develop a classifier. The first PCs that explained approximately 95% of the variance within the data were used in the PC-LDA classification model. The benign and cancer groups were discriminated as shown in Table 8.6. The two-way classification models yielded sensitivities of 87%, specificities of 87% and 93%, and diagnostic accuracies of 87% and 90% for the detection of malignancy.
Table 8.6. The sensitivity, specificity and classification accuracy of a PC-LDA model for discriminating the benign spectra from the thyroid cancer cells.

<table>
<thead>
<tr>
<th>Benign thyroid vs Tumour sample</th>
<th>Sensitivity (%)</th>
<th>Specificity (%)</th>
<th>Diagnostic accuracy (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>N1 v T1</td>
<td>87</td>
<td>93</td>
<td>90</td>
</tr>
<tr>
<td>N2 v T2</td>
<td>87</td>
<td>87</td>
<td>87</td>
</tr>
</tbody>
</table>

The paramount information required from the thyroid FNAC procedure is the differentiation between benign and malignant thyroid nodules. It is an easy method to apply and is presently the favoured diagnostic method for the initial investigation of thyroid nodules, however sensitivity rates as low as 65% and specificity rates as low as 72% have been reported, representing a limitation with the method. In our study, high diagnostic accuracies were achieved (over 87%), as were high sensitivities (over 87%) and specificities (over 93%), using Raman spectroscopy to detect malignancy on thyroid FNAC samples prepared from resected thyroid specimens (‘mock’ FNAC samples). Significant spectral differences were observed between the benign and malignant cells corresponding to differences in the molecular composition of nucleic acids, proteins and lipids. However, the peaks are not consistent across T1 and T2, as each sample is comprised of a combination of patient samples with differing diagnoses. The variation in nucleic acid, protein and lipid content revealed in the malignant thyroid cells is in accordance with the biomolecules identified using cell lines in chapter 7.

The findings of this study suggest potential for using Raman spectroscopy to improve the preoperative diagnosis of TC, although the method needs to be applied on clinical FNA of thyroid to evaluate sensitivity, specificity and accuracy as a diagnostic tool. In the next section, the efficacy of the blood removal protocol developed in sections A and B will be evaluated using clinical FNAC samples.
8.4 Section C – Optimising the blood removal method for thyroid FNAC samples

To evaluate the clinical suitability of the blood removal method devised in sections A and B, the protocol was applied to, and further developed, using clinical thyroid FNAC samples.

8.4.1 Methods

8.4.1.1 Study subjects and sample collection

This section was completed in the Nanomedicine and Molecular Imaging Lab in the ICS Maugeri hospital in Pavia, Italy. This study had approval from the research ethics committee at ICS Maugeri, Pavia. Six thyroid FNAC samples taken from five patients. Four of the samples were taken from TIR3A lesions (low risk intermediate lesions), and two samples were taken from TIR2 lesions (non-malignant/benign)\(^23\).

8.4.1.2 Blood removal

Table 8.7 demonstrates the combination of CytoLyt washes and method of H\(_2\)O\(_2\) treatment that was applied to each sample to remove contaminating blood. After each wash in CytoLyt, the sample was centrifuged, the CytoLyt was decanted and the cell pellet vortexed to mix. This was repeated for the indicated number of CytoLyt washes. H\(_2\)O\(_2\) was then added directly to the cell pellet in samples 1, 2A, 3 and 4, before being diluted with PreservCyt. H\(_2\)O\(_2\) was added to the samples 2B and 5 after they were fixed to the slide.

<table>
<thead>
<tr>
<th>Sample Number</th>
<th>CytoLyt washes</th>
<th>H(_2)O(_2) treatment/Time (seconds)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sample 1</td>
<td>1</td>
<td>In vial / 90 seconds</td>
</tr>
<tr>
<td>Sample</td>
<td>Time</td>
<td>Storage</td>
</tr>
<tr>
<td>----------</td>
<td>----------</td>
<td>---------------</td>
</tr>
<tr>
<td>2A</td>
<td>2</td>
<td>In vial / 90 s</td>
</tr>
<tr>
<td>3</td>
<td>3</td>
<td>In vial / 90 s</td>
</tr>
<tr>
<td>4</td>
<td>4</td>
<td>In vial / 90 s</td>
</tr>
<tr>
<td>2B</td>
<td>1</td>
<td>On the slide/30 s</td>
</tr>
<tr>
<td>5</td>
<td>1</td>
<td>On the slide/30 s</td>
</tr>
</tbody>
</table>

8.4.1.3 Working method for clinical thyroid FNAC samples:

Adjustments were made to the blood removal method. The new adjusted working method for clinical thyroid FNAC samples, which is described below, was applied to samples 2B and 5.

1. Vortex sample for few seconds.
2. Split sample in two 15 mL Falcon tubes.
3. Centrifuge at 1000 rpm for 5 minutes to pellet the cells.
4. Decant supernatant and resuspend in 1 mL Cytolyt, combining samples.
5. Vortex the pellet and pipette sample to break up clumps of cells.
6. Centrifuge at 1000 rpm for 5 min.
7. Decant supernatant and resuspend in 1 mL PreservCyt.
8. Vortex the pellet and pass through a pipette to break up clumps of cells.
10. Treat slides with 35% H₂O₂ for 30 seconds
11. Wash with 70% ethanol for 2 minutes, followed by 100% ethanol for two minutes.
12. Allow the slide to air dry.
8.4.1.2 Slide preparation

As described in chapter 3 section 3.1.2.

The sample preparation laboratory in ICS Maugeri utilised mirrored stainless steel slides instead of ThinPrep® glass slides for Raman spectroscopic analysis. Mirrored stainless steel slides have a low background signal and increase the cellular Raman signal. This is attributed to a double-pass of the laser through the sample. Photons from the incident laser and forward scattered Raman signal are reflected from the mirrored surface and focused towards the collection optics. Stainless steel slides were also used because glass slides are not suitable to use with a 633 nm laser.

8.4.1.3 Raman measurement

As this section was completed in ICS Maugeri, a different Raman spectroscopic system was used for analysis. The Laboratory of Nanomedicine in ICS Maugeri is equipped with a Renishaw inVIA confocal Raman microscope system with a 633 nm laser line. The laser was delivered to the sample through a x100 objective (NA = 0.9) and the confocal hole was set at 100 µm. The system was calibrated to the 520 cm⁻¹ spectral line of silicon and the spectral resolution was defined by the grating which was ruled with 1200 lines per mm grating.

8.4.1.4 Data pre-processing

All spectral data analysis was conducted using Matlab software. Data were process as described in chapter 3 section 3.2.3.7.

8.4.2 Results and discussion

When applied to clinical thyroid FNAC samples, the blood removal protocol developed in section A and B resulted in no cell pellet. Despite repeated CytoLyt washes, the addition of H₂O₂ to the cell pellet sheared the cellular material. Figure 8.21 depicts the slide prepared from sample 1 after treatment, with no cellular material. Identical results
were observed for samples 2A, 3 and 4. This indicated that buried red blood cells remained in the cellular material, and the oxidation of haemoglobin by H$_2$O$_2$ may have produced oxygen bubbles, which may have disrupted the plasma membranes of the cellular material. Therefore, Raman spectra could not be acquired from the clinical thyroid FNAC samples using the method developed in section A and B. The artificial samples used in section A may not have accurately resembled true thyroid FNAC samples. As thyroid FNAC samples contain follicular conglomerates and are taken from a lesion with a direct blood supply, the red blood cells appear to be mixed with the cell clumps, and are difficult to lyse with haemolytic agents. Conversely, the samples in section A contained single cells mixed with blood, and so haemolytic agents were effective with low contamination levels.

The FNAC samples prepared from resected thyroid sections used in section B naturally contained less blood than the FNAC samples in this section as the excised tissue no longer had a blood supply. Typically, the mock FNAC samples in section B matched Group 0 level contamination in section A. As the clinical thyroid FNAC samples in section C were obtained directly from the thyroid of a patient undergoing investigation for TC, these samples contained blood and matched Group 1 level contamination in section A.
As the blood removal protocol developed in sections A and B was not suitable for preparing clinical thyroid FNAC samples for Raman spectroscopic analysis, adaptations were made to the method. The new method for clinical thyroid FNAC samples, applied to samples 2B and 5 reduces the number of CytoLyt washes to a single wash, as additional washes had no effect on blood removal. It was noted in section B that pipetting breaks up the naturally forming cell groups, increasing the haemolytic action of CytoLyt and producing a distribution of more single cells on the slide to facilitate Raman spectroscopy. Thus, vortexing and pipetting was carried out during the CytoLyt wash and before transferring the cells to a PreservCyt vial. In the new method for clinical thyroid FNAC samples, cells were fixed to the steel slide prior to adding H$_2$O$_2$ to retain enough cellular material spectroscopic analysis. Alcohol washes were then used to remove the H$_2$O$_2$ and dehydrate the sample. The addition of H$_2$O$_2$ and alcohol washes are similar to the method devised by Bonnier et al\textsuperscript{5}, however wash times with H$_2$O$_2$ and alcohols are reduced in this method.
Figures 8.22 to 8.26 show that the new method for clinical thyroid FNAC samples retains cellular material for analysis. The raw spectra shown have not been processed to maximise the intensity of cellular signals and minimise noise. The high background observed in the raw spectra can be removed digitally or may naturally decrease using a 532 nm laser.

Figure 8.22. Depiction of cellular material and debris on the steel slide after the sample received treatment. The cell marked with a red box and shown at a greater magnification was targeted for spectroscopic analysis. Acquired with the laser power set to 50% the raw spectra indicate the presence of biological peaks with a high background signal.
Sample 2B

Figure 8.23. Depiction of cellular material and debris on the steel slide after the sample received treatment. The cell marked with a red box and shown at a greater magnification was targeted for spectroscopic analysis. Acquired with the laser power set to 50% the raw spectra indicate the presence of biological peaks with a high background signal.
Figures 8.24-26 indicate that increasing the laser power to 100% effectively increases the signal to noise ratio to produce strong biological peaks. Good quality Raman spectra could thus be acquired from the cells of a clinical thyroid FNAC sample.

Sample 5

Figure 8.24. Depiction of cellular material and debris on the steel slide after the sample received treatment. The cell marked with a red box and shown at a greater magnification was targeted for spectroscopic analysis. Acquired with the laser power set to 100% the raw spectra indicate a strong biological signal with a high background signal.
Figure 8.25. Depiction of cellular material and debris on the steel slide after the sample received treatment. The cell marked with a red box and shown at a greater magnification was targeted for spectroscopic analysis. Acquired with the laser power set to 100% the raw spectra indicate a strong biological signal with a high background signal. Acquired with the laser power set to 100% the raw spectra indicate a strong biological signal with a high background signal.
Figure 8.26. Depiction of cellular material and debris on the steel slide after the sample received treatment. The cell marked with a red box and shown at a greater magnification was targeted for spectroscopic analysis. Acquired with the laser power set to 100% the raw spectra indicate a strong biological signal with a high background signal. Acquired with the laser power set to 100% the raw spectra indicate a strong biological signal with a high background signal.

Figure 8.27 shows mean spectrum taken from the nuclei of cells in sample 5. The spectrum has been processed to maximise the intensity of cellular signals and minimise noise. No blood peaks are visible in the processed spectrum.
Figure 8.27. The mean spectrum taken from the nuclei of cells in sample 5. The spectrum has been processed with Savitsky-Golay smoothing, rubberband baseline subtraction, and vector normalisation, to maximise the intensity of cellular signals and minimise noise. Strong cellular peaks in the processed spectrum are not swamped by blood peaks.

The method optimised in this study effectively prepared clinical thyroid FNAC samples for Raman spectroscopic analysis. Blood peaks do not swamp the cellular signal, and clear biological peaks are observed. Therefore, by utilising this method, further investigations may expand on the results observed in section B in this chapter, and explore the diagnostic capabilities of Raman spectroscopy for the discrimination of normal and malignant thyroid FNAC samples. However, the method needs to be evaluated on a larger sample cohort as low sample numbers limits this study. In addition, a limitation of the new method for clinical thyroid FNAC samples is that pipetting and vortexing breaks up the normal presentation the sample. In future studies, this may complicate the identification of cell types if subsequent morphological assessment of the slide is required.
8.5 Summary and Conclusion

When applying Raman spectroscopy to cytological samples, the presence of blood can swamp cellular Raman signals and inhibit a reliable analysis of spectra. Thyroid FNAC samples regularly contain blood due to the vascularity of the organ. In this study, HeLa cells were contaminated with human blood to replicate the complications encountered when applying Raman spectroscopy to thyroid FNAC samples. Different blood removal methods were analysed and although they were unsuccessful for moderate to high levels of blood (group 2 and 3), a successful method was developed for low levels (group 1). The spectral profile of the group 2 and 3 samples indicated burning as blood was not effectively removed from the cell surface. Subsequently the blood removal protocol developed in section A was applied to ‘mock’ FNAC samples prepared from thyroid biopsies in section B. With an additional vortex step and reduced exposure time to H$_2$O$_2$, good quality Raman spectra were acquired and significant spectral differences were identified between the benign and malignant cases. The benign and malignant cases were discriminated with sensitivities and specificities above 87%, although further investigation is required as there was insufficient cellular material to analyse each patient sample individually, this preliminary study indicates that applying Raman spectroscopy to thyroid FNAC samples is a useful tool for the discrimination of malignant and benign thyroid cells. Finally, in section C, when applied to clinical thyroid FNAC samples, the protocol developed in section A and B proved to be ineffective and adding H$_2$O$_2$ to the cell pellet damaged the cellular material. An alternative method including a single CytoLyt wash, pipetting and vortexing steps, and H$_2$O$_2$ treatment after cells were fixed to the slide, retained the cellular material and enabled good quality Raman spectra to be obtained. Although, the study was limited by low sample numbers and needs to be evaluated on a larger sample cohort, this method may be applied in the future to thyroid
FNAC samples with blood contamination in preparation for Raman spectroscopic analysis.

References


Chapter 9: Conclusion

9.1 Summary of thesis components

The concept underpinning the thesis is the use of Raman spectroscopy as a diagnostic tool for cancer using cytological samples, and so the thesis begins with Chapter 1 describing cytopathology and the principles of Raman spectroscopy. Chapter 2 provides a full description of all the methods used throughout the thesis. Translating Raman data to a diagnosis requires the application of pre-processing and chemometric techniques. The spectral pre-processing and chemometric techniques used throughout the thesis are also explained in Chapter 2. The third chapter is a published detailed literature review of the recent progress made using vibrational spectroscopy for NSCLC diagnosis. The review identified a variety of vibrational spectroscopic systems, biological sample types, and data mining methods that have been applied for Raman based NSCLC diagnosis to date. Based on the findings of previous studies, the potential benefits of applying Raman spectroscopy to bronchoscopy cytology samples for the future diagnosis of NSCLC are highlighted.

The work presented in Chapters 4 and 5 provide a body of evidence that demonstrates the suitability of using Raman spectroscopy on bronchoscopy cytology samples to detect and differentiate the NSCLC subtypes. In Chapter 4, Raman spectroscopy was applied to patients BAL and BW samples with the aim to accurately identify malignant cases, discriminate between the NSCLC subtypes, and establish the molecular differences influencing the discrimination. The results of this study indeed confirm the ability of Raman spectroscopy to discriminate normal lung cells from malignant lung cells with high diagnostic sensitivities, specificities, and accuracies. The Raman spectral profiles of the cells from negative, SCC and AC cases indicated many biochemical alterations between the SCC and AC subtypes.
The aim of Chapter 5 was to assess the potential of Raman spectroscopy to identify malignant signatures in NSCLC cases with clinical cytological analyses reported as diagnostic or no malignant cells seen (NMCS), and to evaluate the sensitivity and specificity of Raman spectroscopy on cytology in comparison to the final histology report. The recurrent patterns in Raman spectra of cells from NSCLC cases with differing cytology reports demonstrated that Raman spectroscopy may be an effective tool for detecting lung cancer on morphologically normal cytology samples, as detailed analysis of the Raman spectra revealed malignant signatures in the cases with NMCS on cytology. The evidence provided in Chapters 4 and 5 demonstrate the feasibility of using Raman spectroscopy for the detection of NSCLC on bronchoscopy cytology samples.

In Chapter 6, the thesis analysed the diagnostic effectiveness of applying Raman spectroscopy to lung tissue sections for NSCLC diagnosis, and compared the results to the efficacy achieved using cytology samples. In comparison to the negative tissue, malignant tissues showed significant variations in biomolecular content similar to those identified in Chapter 4. The PLSDA algorithm yielded good performance scores for the discrimination of the SCC and AC subtypes in Chapter 6. As broad spectral regions are significantly different between the negative and malignant lung tissue, applying Raman spectroscopy to cytology samples may be more suitable for identifying differences in individual biomolecules between the negative, SCC and AC subtypes.

The work presented in Chapters 7 and 8 demonstrate the feasibility of using Raman spectroscopy on FNAC samples to diagnose thyroid cancer. In Chapter 7, Raman spectroscopy was applied to thyroid cancer cell lines with the aim to accurately discriminate between the cell lines and establish the molecular differences influencing the discrimination. The results of this proof of concept study indeed confirm the ability of Raman spectroscopy to discriminate normal thyroid from thyroid cancer with high
diagnostic sensitivities, specificities, and accuracies. Analysis of the Raman spectral profiles of the cell lines also revealed molecular differences between the TC subtypes. As thyroid FNAC samples inevitably contain blood contamination, and blood interferes with the spectra, the aim of Chapter 8 was to establish a blood removal method, which is suitable for use on thyroid FNAC samples to allow Raman spectroscopic analysis. The initial protocol was developed using cell lines and required minor adjustments when applied to ‘mock’ thyroid FNAC samples taken from surgically removed specimens. Blood contamination was successfully removed from the ‘mock’ thyroid FNAC samples and diagnostic peaks discriminating the negative and malignant cases were identified. However, subsequent application of the protocol to clinical thyroid FNAC samples was unsuccessful in removing blood. With further optimisation, blood contamination was eradicated from the spectra, indicating that the method developed is suitable to prepare thyroid FNAC samples with low levels of blood for Raman spectroscopic analysis.

9.2 Clinical Relevance

The ability to obtain detailed chemical information label-free makes Raman spectroscopy an attractive method for clinical diagnostic cytology, although translating Raman spectroscopy to routine clinical use requires some factors to be addressed. Sample preparation is a fundamental step for preparing cytology specimens for spectroscopic analysis. Appropriate standard operating procedures should be designed for each diagnostic sample type, and implemented internationally. Spectroscopic analysis of ex vivo diagnostic samples also requires the sample to be fixed to a substrate. Glass is the most applicable substrate for clinical setting as it is cost effective, is already used in clinical laboratories, and signals derived from glass can be digitally removed from the spectrum. Standardised data analysis and interpretation is another factor that needs to be addressed. Many chemometric methods extract diagnostic data from discrete features and
so considerable care is needed to define diagnostically relevant features. For this reason, the method of validation of the classifier is crucial for determining the reliability of the model. Larger studies with large numbers of patients require cross validation methods such as leave one-patient-out-cross-validation (LOPOCV), as using leave-one-spectrum-out cross-validation LOSOCV on larger datasets can over-train the model. LOPOCV is therefore better for evaluating the performance characteristics of the classifier. Further, the classifier must be validated with an independent test set to determine how well it classifies unseen samples. The Raman instrument should be suitable for a diagnostic lab, and should therefore be easily calibrated and small enough to fit on a benchtop.

The body of work presented in this thesis shows that Raman spectroscopy can differentiate NSCLC subtypes using cytology specimens, and can even detect malignant signatures in cytology samples with NMCS reported on cytology. The findings of this thesis imply that having Raman spectroscopy as part of the diagnostic algorithm for NSCLC could provide a label free differential diagnosis and preserve the limited cell sample for subsequent molecular tests, saving patients an invasive surgical procedure to obtain more tissue for analysis. Although previous studies have explored the use of Raman spectroscopy on lung tissue sections and in vivo, this thesis shows that Raman spectroscopic analysis of cytology specimens is comparably effective for NSCLC diagnosis. Similarly in the case of TC, Raman spectroscopy has the potential to reduce the number of invasive biopsy procedures required to confirm malignancy. With adequate sample preparation, quality Raman spectra can be acquired from thyroid FNAC samples, and with continued research this technique may show the ability to differentiate malignant and benign thyroid lesions.

Bronchoscopy cytology samples are routinely utilised for first line diagnostics in cases of NSCLC. Further, FNAC samples are routinely utilised for first line diagnostics in cases
of TC, yet to our knowledge there is no research investigating the spectroscopic diagnosis of NSCLC and TC on these diagnostic sample types. The work presented in this thesis is unique compared to previous published studies as it examines the application of Raman spectroscopy for the diagnosis of NSCLC and TC on cytology samples prepared as ThinPrep®.

9.3 Future Work

There were insufficient patient samples obtained in this project to draw clinical conclusions about the diagnostic capabilities of Raman based NSCLC and TC diagnosis using cytology samples. The difficulty in obtaining consented samples from patients undergoing bronchoscopy for the investigation of lung cancer was a factor that was not fully anticipated. However, this thesis has demonstrated that diagnostically relevant Raman peaks can be acquired from cytological specimens, warranting a larger scale investigation.

The findings of this thesis contribute to the application of Raman spectroscopy as a diagnostic tool on cytology, and it is hoped that the work presented in this thesis will be a platform from which further studies can be developed. There are several interesting directions for future studies to build on the work we have done. To address the low number of patient samples included in this study, a larger patient cohort could be developed to increase the size of the Raman spectral datasets. As a result of the low number of patients samples, the venetian blinds cross validation method was used throughout this thesis to analyse the limited data, as the datasets were too small for LOPOCV, and LOSOCV would have over-trained the classifiers. With an increased patient sample number and larger datasets, a cross validation method suitable for larger datasets, such as LOPOCV, could be utilised to provide an accurate representation of the diagnostic efficacy of applying Raman spectroscopy to bronchoscopy cytology samples.
for NSCLC diagnosis. Further, with a larger patient cohort, future studies could build on the findings explored in this thesis by investigating the ability to correlate tumour mutation profiles with Raman spectra.

Another interesting direction for further work would be to confirm lymphocyte identity. The results presented in chapter 4 indicate that the differences in the spectra taken from lymphocytes of NSCLC cases and lymphocytes of negative cases may be spectral signatures for the immune biochemical responses to the presence of tumour. However the data is limited in this study, and further identification of lymphocytes on bronchoscopy cytology samples is required. Continued research could confirm lymphocyte identity with immunohistochemistry and explore the use of Raman spectroscopy to further understand the role and diagnostic relevance of lymphocytes in NSCLC.

Studies building from this work could also explore the significance of the collagen related peaks observed in the epithelial cells in bronchoscopy cytology samples taken from NSCLC cases. While collagen is associated with EMT and invasion, further investigations should establish if this could lead to potential false positive diagnosis in the presence of fibrosis, for example in non-malignant interstitial lung diseases.

With increased sample number future work could also explore the possibility of using Raman spectroscopy to detect false positives on cytology, as the work presented in this thesis shows that malignancy can be detected in false negative bronchoscopy cytology samples, and reducing the rate of false diagnoses would be of great benefit to both patients and the healthcare system.

A limitation to take into consideration for future studies is the low amount of cell material in residual samples. Due to the low number of cells retained in residual sample vials, obtaining sufficient spectra for reliable statistical analysis is difficult. With the
appropriate ethical approval, acquiring a second sample for research purposes would overcome this limitation.

The thyroid FNAC sample preparation method developed in this thesis was demonstrated to be an effective method for preparing the samples for Raman spectroscopic analysis, yet it needs to be validated on a larger group of patient samples. With the development of an effective sample preparation method which allows spectroscopic analysis of thyroid FNAC samples, future research could conduct a clinical study with thyroid FNAC samples obtained from patients under investigation for thyroid cancer. Using the blood removal protocol devised in this thesis, the ability of Raman spectroscopy to differentiate benign and malignant thyroid lesions could be evaluated. The investigation could include a component for cellular recognition on thyroid FNAC samples. Similar to the method utilised in this thesis, slides could be Pap stained subsequent to Raman spectroscopic analysis to identify the targeted cell types marked with XY coordinates.

In the long term, if large sample studies confirm the sensitivity and specificity of Raman spectroscopy in the diagnosis of NSCLC and TC, there is potential for the development of an automated Raman system that could identify cell types and acquire and store spectral data. This could optimise the current time expensive method included in identifying target cells, recording spectra, staining, cell identification, and developing distinct databases. A more advanced fully automated system could even potentially apply data mining techniques to classify the spectra in real time.

List of Publications

Modules completed:

- Discipline specific
  - Data Mining for Biomedical Applications ENEH1004 – 5 ECTS
  - Applied Spectroscopy NMAD10XX - 5 ECTS
  - Biophotonics and Imaging Graduate Summer School - 5 ECTS
  - Advanced Diagnostic Methods BIOL9220 – 5 ECTS

- Employability skills
  - Introduction to statistics GRSO1005 – 5 ECTS
  - Project Management PRJM2000 – 5 ECTS
  - Introduction to Pedagogy GRSO1010 – 5 ECTS
  - Introduction to Probability and Data – 5 ECTS

Conference presentations:

- 2015 – DIT Annual Graduate Research Symposium (Poster)
- 2016 - Raman4Clinics (Poster)
- 2016 - 40th annual symposium of the Microscopy Society of Ireland (Poster)
- 2017 – DIT Annual Graduate Research Symposium (Poster)
- 2017 - CLIRCON in Manchester University (Poster)
- 2018 – SPEC2018 Glasgow