

Raman Spectroscopy Studies of Prostate Cancer and Streptomyces Bacteria

Charles Matthew Kershaw

MSc

University of York

Physics

January 2017

Abstract

This thesis will examine the use of Raman spectroscopy (RS) to probe biomolecular information from biological samples. The research is comprised of two main studies, a prostate cancer study and a *Streptomyces* bacteria study. The prostate cancer study used live cell samples from two different patients (H517–15 and H554–15), containing normal and Gleason seven cancer cells and two standard cell lines, PNT2–C2 (normal) and P4E6 (Gleason four cancer). The Raman spectra of each cancer/normal matched pair were analysed using principle component analysis (PCA). PCA was able to distinguish the normal cells from the cancer cells for the H517–15 and standard cell line samples. However, the H554–15 cancer and normal cells comparison proved to be too similar for PCA to separate. Peak intensity ratio (PIR) analyses were used to probe biomolecular differences between the normal and cancer cell samples. PIR results for the H554–15 sample showed that the cancer and normal cell samples were similar. PIR results for H517–15 revealed several biomarkers, which are consistent with the biological literature on cancer. The PIR results for the standard cell lines suggested that the samples may not be representative of patient biology due to the PIR biomarkers showing opposite trends to both the H517–15 primary patient sample, and the literature. The second aim of this research was to assess changes in lipid content between three samples of *Streptomyces* bacteria. The three bacteria samples comprised the wild type (J1929), a strain which had undergone a mutation causing a loss of intrinsic antibiotic resistance (DT3017), and a treated form of the mutated strain (PDT16). The Raman study in this work corroborated known biological information about the samples, showing an increase in unsaturated fatty acid (lipid) content from the J1929 sample to the DT3017 sample, followed by a recovery to the J1929 lipid content, upon treatment, for the PDT16 sample.

Contents

Abstract	i
Contents	i
List of figures	iii
List of tables	xi
Acknowledgements	xiv
Declaration	xv
1 Introduction	1
1.1 Raman Spectroscopy	1
1.2 Raman Spectroscopy as a Biological Tool	4
1.3 Prostate Cancer Cells	10
1.4 <i>Streptomyces coelicolor</i> Bacteria	13
1.5 Summary of Research Objectives	15
2 Methods	17
2.1 Prostate Cancer Cell Culture	17
2.2 <i>Streptomyces coelicolor</i> Cells Culture	19
2.3 Raman Spectra Acquisition	20
2.4 Trypan Blue Assay	23
2.5 Raman Measurements of Prostate Cancer Cells	23

2.6	Raman Measurements of <i>Streptomyces coelicolor</i> Bacteria	24
2.7	Prostate Cancer Spectra Analysis	24
2.8	<i>Streptomyces coelicolor</i> Spectra Analysis	26
3	Results and Discussion	28
3.1	Raman Spectroscopy Studies of Prostate Cancer and Normal Prostate Cells	28
3.1.1	Trypan Blue Assay	29
3.1.2	Raman Spectra	30
3.1.3	Principle Component Analysis (PCA) and Linear Discriminant Analysis (LDA) .	47
3.1.4	Peak Intensity Ratio (PIR) Comparisons	58
3.1.5	Prostate Cancer Summary	76
3.1.6	Further Work For Raman Spectroscopy Studies of Prostate Cancer	77
3.2	Raman Spectroscopy Studies of <i>Streptomyces coelicolor</i> Bacteria	78
3.2.1	Raman Spectra of <i>Streptomyces coelicolor</i> Bacteria	80
3.2.2	Peak Intensity Ratio Comparisons	88
3.2.3	Further work for Raman Spectroscopy Studies of <i>Streptomyces coelicolor</i> Bacteria	89
4	Conclusions	90
	References	94

List of Figures

1.1	Energy level diagram showing the Raman effect: Elastic, Rayleigh scattering, Stokes scattering, where a photon of light loses energy to the molecular bond and anti-Stokes scattering, where the photon gains energy from the molecular bond, Ferraro (1994).	2
1.2	Diagram of Raman spectra showing the difference between Stokes and anti-Stokes shifting, Ferraro (1994)	2
1.3	An example of typical prostate and prostate cancer Raman spectra taken from Crow <i>et al.</i> (2003). These spectra are averages taken from approximately 200 individual spectra from a benign prostate hyperplasia (BPH) and prostate cancer tissue samples. Both spectra have been normalised against the intensity of the 1443 cm^{-1} peak, which is related to proteins and lipids.	5
1.4	An example of Raman spectra of <i>Streptomyces</i> bacteria, from Walter <i>et al.</i> (2012). Four <i>S. coelicolor</i> Raman spectra are shown. From bottom to top, the un-contaminated sample and three samples with increasing levels of contamination with Cl ⁻ and Ni ²⁺	14
2.1	The Horiba XploRA instrument used for all of these experiments.	21
2.2	Basic flow schematic of a typical confocal Raman spectroscopy set-up, representative of this experimental set-up.	21
2.3	An example use of the multipeak fitting package to fit a highly convolved part of a spectrum. The example here is the H517–15 cancer sample between 2700 and 3100 cm^{-1} . This region contains the peaks, 2850 , 2871 , 2893 , 2937 , 2970 , 3008 and 3065 cm^{-1} . The figure shows the averaged spectrum, Gaussian peak profiles and cubic baseline.	25

3.1	Optical image obtained using a 63x dipping lens with a numerical aperture of one showing the P4E6 cell (inside the red square), which was analysed before laser exposure or the application of the trypan blue dye. The cell chosen is of a regular size with no obvious defects or contaminations.	31
3.2	Optical image obtained using a 63x dipping lens with a numerical aperture of one of the same P4E6 cell (inside the red square) after laser exposure and the addition of the trypan blue dye. Trypan blue is a dye which is only taken up by cells that have damaged outer membranes. In this case the dye has not been uptaken by the cell indicating that there has been no damage due to the laser exposure. The laser exposure used is the same applied for the Raman measurements in this study, hence verifying non-destructive testing of the cells.	32
3.3	To ensure that the trypan blue dye would function correctly in the event of cell damage, the sample was washed with ethanol to kill the P4E6 cells and then more trypan blue dye was added to test for cell damage. From this optical image, taken using a 63x dipping lens with a numerical aperture of one, it was concluded that if the cell had been damaged by the laser, then the trypan blue dye would have been uptaken by the cell therefore producing this result.	33
3.4	Averaged Raman spectra taken in the fingerprint region ($600\text{--}1800\text{ cm}^{-1}$) for the direct patient derived primary cell cultures, H517-15 cancer (n=164), H517-15 normal (n=161), H554-15 cancer (n=156), H554-15 normal (n=167), and the P4E6 cancer (n=149) and PNT2-C2 normal (n=154) standard cell lines. The biological relevance of the key peak assignments identified on the spectra are detailed in Table. 3.1. The six averaged spectra have been linear baselined and normalised to the 1447 cm^{-1} peak, which represents the CH_2 bending mode of proteins and lipids. peak. The spectra are vertically shifted using an arbitrary offset to aid visualisation.	35

-
- 3.5 Averaged Raman spectra taken in the high wavenumber region ($2700\text{--}3100\text{ cm}^{-1}$) for the direct patient derived primary cell cultures, H517-15 cancer ($n=164$), H517-15 normal ($n=161$), H554-15 cancer ($n=156$), H554-15 normal ($n=167$) and the P4E6 cancer ($n=149$) and PNT2-C2 normal ($n=154$) standard cell lines. The biological relevance of the key peak assignments identified on the spectra are detailed in Table. 3.1. The six average spectra have been linear baselined and normalised to the 1447 cm^{-1} peak which represents the CH_2 bending mode of proteins and lipids. The spectra are vertically shifted using an arbitrary offset to aid visualisation. 36
- 3.6 Averaged Raman spectra in the fingerprint region ($600\text{--}1800\text{ cm}^{-1}$) with associated standard error envelopes for the direct patient derived primary cell cultures, H517-15 cancer ($n=164$), H517-15 normal ($n=161$), H554-15 cancer ($n=156$), H554-15 normal ($n=167$), and the P4E6 cancer ($n=149$) and PNT2-C2 normal ($n=154$) standard cell lines. The six spectra have been linear baselined and vertically shifted using an arbitrary offset to aid visualisation. These spectra have not been normalised. 37
- 3.7 Averaged Raman spectra in the high wavenumber region ($2700\text{--}3100\text{ cm}^{-1}$) with associated standard error envelopes for the direct patient derived primary cell cultures, H517-15 cancer ($n=164$), H517-15 normal ($n=161$), H554-15 cancer ($n=156$), H554-15 normal ($n=167$), and the P4E6 cancer ($n=149$) and PNT2-C2 normal ($n=154$) standard cell lines. The six spectra have been linear baselined and vertically shifted using an arbitrary offset to aid visualisation. These spectra have not been normalised. 38
- 3.8 The convergence of the standard error of the mean as a function of the number of spectra in the average, as indicated by the legend, for the fingerprint region only showing the (a) H517-15 cancer ($n=164$), (b) H517-15 normal ($n=161$), (c) H554-15 cancer ($n=156$), (d) H554-15 normal ($n=167$), (e) P4E6 cancer ($n=149$) and (f) PNT2-C2 normal ($n=154$) cell lines. 39

3.9	The convergence of the standard error of the mean as a function of the number of spectra in the average, as indicated by the legend, for the high-wavenumber region only showing the (a) H517–15 cancer (b) H517–15 normal (c) H554–15 cancer (d) H554–15 normal (e) P4E6 cancer (f) PNT2–C2 normal cell lines.	40
3.10	The convergence of the percentage standard error of the mean, as a function of the number of spectra in the spectral average for the PIR I1447/I1653 measured randomly across the population of each cell line. The cell lines studied are (a) H517–15 cancer, (b) H517–15 normal, (c) H554–15 cancer, (d) H554–15 normal, (e) P4E6 cancer and (f) PNT2–C2 normal.	43
3.11	The convergence of the percentage standard error of the mean as a function of the number of spectra in the spectral average for the PIR I3008/I2850 measured randomly across the population of each cell line. The cell lines studied are (a) H517–15 cancer, (b) H517–15 normal, (c) H554–15 cancer, (d) H554–15 normal, (e) P4E6 cancer and (f) PNT2–C2 normal.	44
3.12	Results of PCA followed by LDA analysis for the direct patient derived primary cell cultures, H517–15 cancer (n=164), H517–15 normal (n=161), H554–15 cancer (n=156), H554–15 normal (n=167), and the P4E6 cancer (n=149) and PNT2–C2 normal (n=154) standard cell lines. Principle components one, two and three account for 43.4%, 19.6% and 4.9% of the total variance, respectively.	48
3.13	PCA results for the direct patient derived primary cell cultures, H554–15 cancer (n=156), H554–15 normal (n=167) using the first three principle components, accounting for 55.7%, 6.8% and 4.7% of the total variance, respectively. No separation is seen here, indicating that the two samples produce very similar spectra and any variance in the data cannot be attributed to differences between the normal and cancer cells, only to the heterogeneity of the population as a whole.	52

3.14	Loadings of the first three PCs for the H554–15 samples. These loadings suggest that areas of high variance include 745 cm^{-1} , 781 cm^{-1} , 850 cm^{-1} , 934 cm^{-1} , 1000 cm^{-1} , 1029 cm^{-1} , 1125 cm^{-1} and 1578 cm^{-1} for PC1, 1000 cm^{-1} and 1653 cm^{-1} for PC2 and 781 cm^{-1} , 827 cm^{-1} , 850 cm^{-1} , 1000 cm^{-1} , 1029 cm^{-1} , 1081 cm^{-1} , 1125 cm^{-1} , 1447 cm^{-1} and 1578 cm^{-1} for PC3. These markers however are not enough to discriminate between the 2 samples, as shown in Fig. 3.13.	53
3.15	PCA results for the P4E6 cancer (n=149) and PNT2–C2 normal (n=154) standard cell lines showing principle components one, two and four accounting for 36.5, 13.3 and 5.4% of the total variance, respectively. The separation seen here can be attributed entirely to PC4, with PC1 and PC2 accounting only for the variance within the two samples combined. PC3 was omitted here and PC4 included as PC3 contributed no separation between normal and cancer.	54
3.16	Loadings for principle components 1, 2 and 4 for the standard cell lines samples, P4E6 and PNT2–C2. Only PC4 accounts for variation between the normal and cancer samples, as shown in Fig. 3.15. The peaks which correspond to high variance in PC4 are 724 cm^{-1} , 745 cm^{-1} , 781 cm^{-1} , 850 cm^{-1} , 1000 cm^{-1} , 1029 cm^{-1} , 1096 cm^{-1} , 1125 cm^{-1} , 1205 cm^{-1} , 1600 cm^{-1} , 1613 cm^{-1} and 1653 cm^{-1}	56
3.17	PCA results for the direct patient derived primary cell cultures, H517–15 normal (n=161), H554–15 normal (n=167) and the PNT2–C2 normal (n=154) standard cell line.	57
3.18	PCA results for the direct patient derived primary cell cultures, H517–15 cancer (n=164), H554–15 cancer (n=156) and the P4E6 cancer (n=149) standard cell line.	57
3.19	PCA results for the direct patient derived primary cell cultures, H517–15 cancer (n=164), H517–15 normal (n=161) showing the first three principle components, accounting for 44.6, 23.1 and 4.1% of the total variance, respectively. The separation seen here can be attributed almost entirely to PC2, with a small contribution from 3. PC1 makes no contribution to the separation, indicating that PC1 accounts only for the heterogeneity of the H517–15 population as a whole.	59

3.20	Loadings of the first three principle components (PCs) for the H517–15 samples. Only PC2 accounts for variation between the normal and cancer samples, as shown in Fig. 3.19. The peaks which correspond to high variance in PC2 are 713 cm^{-1} , 745 cm^{-1} , 781 cm^{-1} , 934 cm^{-1} , 1000 cm^{-1} , 1029 cm^{-1} , 1096 cm^{-1} , 1125 cm^{-1} , 1205 cm^{-1} , 1578 cm^{-1} , 1600 cm^{-1} , 1613 cm^{-1} and 1653 cm^{-1}	60
3.21	PIR markers from the fingerprint peak intensity ratios against 781 cm^{-1} (DNA/RNA) for the H517–15 primary patient samples. Markers shown here are fully discriminatory for one or more of the cancer vs. normal pairs, non-discriminatory markers are not shown. .	64
3.22	PIR markers from the fingerprint peak intensity ratios against 781 cm^{-1} (DNA/RNA) for the H554–15 primary patient samples. Markers shown here are fully discriminatory for one or more of the cancer vs. normal pairs, non-discriminatory markers are not shown. .	65
3.23	PIR markers from the fingerprint peak intensity ratios against 781 cm^{-1} (DNA/RNA) for the P4E6 cancer and PNT2–C2 normal cell line samples. Markers shown here are fully discriminatory for one or more of the cancer vs. normal pairs, non-discriminatory markers are not shown.	66
3.24	Peak intensity ratio (PIR) markers from the fingerprint peak intensity ratios against 1447 cm^{-1} (lipids and proteins) for the H517–15 primary patient samples. Markers shown here are fully discriminatory for one or more of the cancer vs. normal pairs, non-discriminatory markers are not shown.	67
3.25	PIR markers from the fingerprint peak intensity ratios against 1447 cm^{-1} (lipids and proteins) for the H554–15 primary patient samples. Markers shown here are fully discriminatory for one or more of the cancer vs. normal pairs, non-discriminatory markers are not shown.	68
3.26	PIR markers from the fingerprint peak intensity ratios against 1447 cm^{-1} (lipids and proteins) for the P4E6 cancer and PNT2–C2 normal cell line samples. Markers shown here are fully discriminatory for one or more of the cancer vs. normal pairs, non-discriminatory markers are not shown.	69

3.27 Raman spectra for the H554–15 primary patient samples. The spectra have been averaged, baselined and normalised against 1447 cm^{-1} . The spectra have not been shifted so a direct comparison can be made.	75
3.28 Raman spectra for the P4E6 cancer and PNT2–C2 normal cell lines. The spectra have been averaged, baselined and normalised against 1447 cm^{-1} . The spectra have not been shifted so a direct comparison can be made.	75
3.29 Raman spectra for the H517–15 primary patient samples. The spectra have been averaged, baselined and normalised against 1447 cm^{-1} . The spectra have not been shifted so a direct comparison can be made.	76
3.30 An optical image, using a 100x objective microscope lens with a numerical aperture of 0.9, of the surface of the <i>S. coelicolor</i> sample J1929. Owing to the rough surface of the sample and the confocal nature of the microscope, the in focus area is very small, shown in the red box.	79
3.31 Averaged Raman spectra in the fingerprint region ($600\text{-}1800\text{ cm}^{-1}$) for the three <i>S. coelicolor</i> cell lines (J1929 wild-type, PDT16 treated and DT3017 untreated) showing the key peak assignments (Table 3.5). The spectra are vertically shifted using an arbitrary offset to aid visualisation.	82
3.32 Averaged Raman spectra in the high-wavenumber region ($2750\text{-}3110\text{cm}^{-1}$) for the three <i>S. coelicolor</i> cell lines (J1929 wild- type, PDT16 treated and DT3017 untreated) showing the key peak assignments (Table 3.5). The spectra are vertically shifted using an arbitrary offset to aid visualisation.	83
3.33 Convergence of the percentage standard error of the mean (SE) for the TUFA/TFA peak intensity ratio I3008/I2850 for the J1929 (wild) cell line as a function of the number of randomly chosen spectra comprising the spectral average. In total, 42 individual spectra were averaged.	85

3.34	Convergence of the percentage standard error of the mean (SE) for the TUFA/TFA peak intensity ratio I3008/I2850 for the PDT16 (treated) cell line as a function of the number of randomly chosen spectra comprising the spectral average. In total, 55 individual spectra were averaged.	86
3.35	Convergence of the percentage standard error of the mean (SE) for the TUFA/TFA peak intensity ratio I3008/I2850 for the DT3017 (untreated) cell line as a function of the number of randomly chosen spectra comprising the spectral average. In total, 40 individual spectra were averaged.	87

List of Tables

2.1	Table showing the three samples and their descriptions [1].	19
3.1	All peak positions identified in these experiments and their published assignments along with examples of their uses in the literature. All peak assignments correspond to Movasaghi <i>et al.</i> (2007). The peak positions in this study were identified by fitting and then averaging across the 6 samples, this returned an associated error in the peak locations of 1 cm^{-1}	47
3.2	The results of the leave-one-out cross validation, performed on the LDA result shown in Fig. 3.12. The total error in the LDA, calculated from the number of spectra that were grouped incorrectly, is 0.14196 (14%). The H517–15 normal, H554–15 normal and H554–15 cancer samples are spread across their relevant groups showing that the LDA was not able to tell these spectra apart as well as the other samples.	50
3.3	Table showing the mass unsaturation degree ratio (I1653/I1447) and TUFA/TFA ratio (I2850/I3008) for all samples investigated.	61
3.4	Table showing the P-values obtained from a T-test when comparing the ratios (I1653/I1447) and (I2850/I3008) between the normal and cancer samples for all cell types investigated.	62
3.5	All peak positions identified in these experiments and their published assignments along with examples of their uses in the literature. All peak assignments correspond to Movasaghi <i>et al.</i> (2007). These peak positions were identified by fitting the average spectra for each sample, then averaging the peak positions. The associated error in the peak locations is approximately 1 cm^{-1}	84

3.6	Table showing the number of spectra and TUFA/TFA ratios with associated errors for all samples	88
3.7	Table showing the P-values obtained from a T-test when comparing the TUFA/TFA ratio between the three different samples.	88

Acknowledgements

The author is grateful to Yvette Hancock, Raquel Ponzoni, Andre Ponzoni, Fiona Frame, Maggie Smith, Julie Wilson, Nicholas Read and Robert Howlett for their valued assistance in the realization of this thesis.

Declaration

I declare that this thesis is a presentation of original work and I am the sole author. This work has not previously been presented for an award at this, or any other, University. All sources are acknowledged as References

Chapter 1

Introduction

1.1 Raman Spectroscopy

Raman spectroscopy (RS) is a label free method for assessing molecular and structural information from a variety of materials. The Raman effect was discovered in 1928 by Indian physicist Sir Chandrasekhara Venkata Raman [2] and has been employed extensively across many areas of science. The Raman effect involves the inelastic scattering of monochromatic light, based on the polarizability of vibrating molecules. Small differences in the outgoing and incident light energies provides sensitive spectroscopic information and a molecular scale fingerprint of matter [3].

When light is scattered it is mostly at the same wavelength and frequency as the incident light. This is known as Rayleigh, (elastic) scattering [4]. A second type of scattering is Raman scattering, which is much weaker. Raman scattering contributes only 1 in 10^{10} of the scattered photons [5] with frequencies $\nu_0 \pm \nu_r$, where ν_0 is the frequency of the incident light and ν_r is the shift in frequency due to molecular interaction in the sample. This Raman shift is referred to as Stokes shift when energy is lost by the photon or an anti-Stokes shift when energy is gained (Fig. 1.1) [6]. Since the same information can be obtained from both Stokes and anti-Stokes scattering, it is convention to only measure the Stokes scattered photons. As the population of molecules in the vibrational energy ground state is greater than in the excited state according to the Maxwell Boltzmann distribution, Stokes scattering also results in a more intense spectra, (Fig. 1.2) [7].

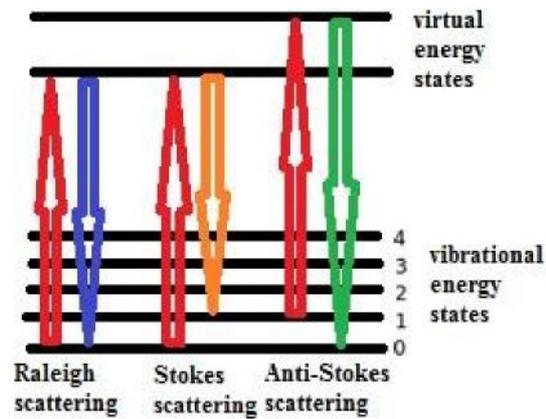


Figure 1.1: Energy level diagram showing the Raman effect: Elastic, Rayleigh scattering, Stokes scattering, where a photon of light loses energy to the molecular bond and anti-Stokes scattering, where the photon gains energy from the molecular bond, Ferraro (1994).

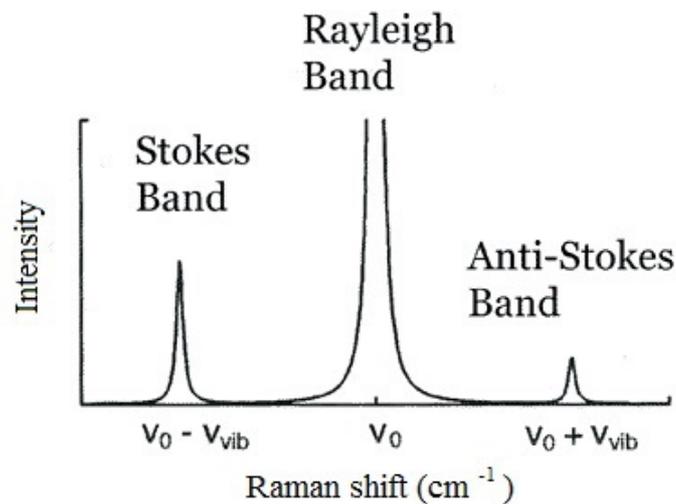


Figure 1.2: Diagram of Raman spectra showing the difference between Stokes and anti-Stokes shifting, Ferraro (1994)

A Raman shift occurs when the electron distribution surrounding a bond in a molecule is distorted by an incident photon. Following the scattering of the photon the bond returns to its natural state [8]. Classically, the electric field strength (E) of an electromagnetic wave, in this case, laser light, varies with time (t) according to

$$E = E_0 \cos(2\pi\nu_0 t), \quad (1.1)$$

where E_0 is the maximum vibrational amplitude and ν_0 is the laser frequency. For the case of a diatomic molecule, the photon re-emission causes the bond to become temporarily polarised, inducing a dipole moment (P), where

$$P = \alpha E = \alpha E_0 \cos(2\pi\nu_0 t) \quad (1.2)$$

Here, α is a constant of proportionality known as the *polarizability*. If a molecule is vibrating at frequency ν_m , then the nucleus is displaced by q , given by

$$q = q_0 \cos(2\pi\nu_m t), \quad (1.3)$$

where q_0 is the maximum vibrational displacement. The polarizability (α) can be written as a linear function of q , for small vibrational displacements. Thus,

$$\alpha = \alpha_0 + \left(\frac{\delta\alpha}{\delta q}\right)_0 q_0 \quad (1.4)$$

where α_0 is the polarizability at equilibrium and $(\frac{\delta\alpha}{\delta q})_0$ is the rate of change of the polarizability with respect to the change in nuclear displacement, at $q=0$. Hence, using Eq.1.4 in Eq.1.2 we can write

$$\begin{aligned} P &= \alpha E_0 \cos(2\pi\nu_0 t) \\ &= \alpha_0 E_0 \cos(2\pi\nu_0 t) + \left(\frac{\delta\alpha}{\delta q}\right)_0 q_0 E_0 \cos(2\pi\nu_0 t) \\ &= \alpha_0 E_0 \cos(2\pi\nu_0 t) + \left(\frac{\delta\alpha}{\delta q}\right)_0 q_0 E_0 \cos(2\pi\nu_0 t) \cos(2\pi\nu_m t) \\ &= \alpha_0 E_0 \cos(2\pi\nu_0 t) + \frac{1}{2} \left(\frac{\delta\alpha}{\delta q}\right)_0 q_0 E_0 (\cos(2\pi(\nu_0 + \nu_m)t) + \cos(2\pi(\nu_0 - \nu_m)t)). \end{aligned} \quad (1.5)$$

The first term in this expression represents Rayleigh scattering, or light radiated from a dipole with frequency ν_0 . The second term represents the Raman scattered frequency shifts $\nu_0 + \nu_m$ (anti-Stokes) and $\nu_0 - \nu_m$ (Stokes). From this expression, we can determine that a vibration is only Raman active when $(\frac{\delta\alpha}{\delta q})_0$ is not equal to zero, or in other words, the rate of change of polarizability must not be zero [6].

A Raman spectrum is a plot of the scattered light intensity, measured in counts per given time frame against the wavenumber shift of the scattered photons. The wavenumber ($\hat{\nu}$) is defined in units of (cm^{-1}) such that

$$\hat{\nu} = \frac{\nu}{c} = \frac{1}{\lambda} \quad (1.6)$$

where ν is the frequency and c is the speed of light and λ is the wavelength.

Raman spectra are therefore characterised by frequency differences between the incident light and the Raman scattered light, or equivalently, by the Raman shift. The peaks which appear in Raman spectra can therefore be associated with molecular bonds or functional groups in the sample [6] [9] [10]. As an example, Fig. 1.3 shows some typical Raman spectra for prostate cancer cells [11].

RS is fundamentally different to *infrared spectroscopy*, despite it also being a vibrational spectroscopy method. IR spectroscopy requires the molecular vibrations to be associated with changes in a dipole moment, as opposed to RS being associated with a change in polarizability [12]. One of the major disadvantages of IR spectroscopy is that simple symmetrical molecules, e.g. N_2 , are not IR active. However these molecules can still be probed using Raman spectroscopy due to the molecule having a single vibrational band [12].

RS has been used to characterise various sample types, both biological and non-biological. Examples of the use of RS include, the characterisation of carbon, [13] [14] [15], the forensic analysis of illegal drugs [16] [17] [18] and the characterisation of biological samples.

1.2 Raman Spectroscopy as a Biological Tool

The collection of Raman spectra can be performed *in vitro*, *ex vivo* or *in vivo* with minimal disruption to the biological sample. This is a major advantage of RS, as most biological assays use chemical biomarkers and usually require conditions non-native to the biological environment [3]. RS can also be used on living cells to produce spectra that contain information which is better representative of the cell's biology [19].

RS on biological samples produces a fingerprint containing detailed biomolecular information. For the majority of biological samples, RS spectra are usually split into two regions: the Raman finger-

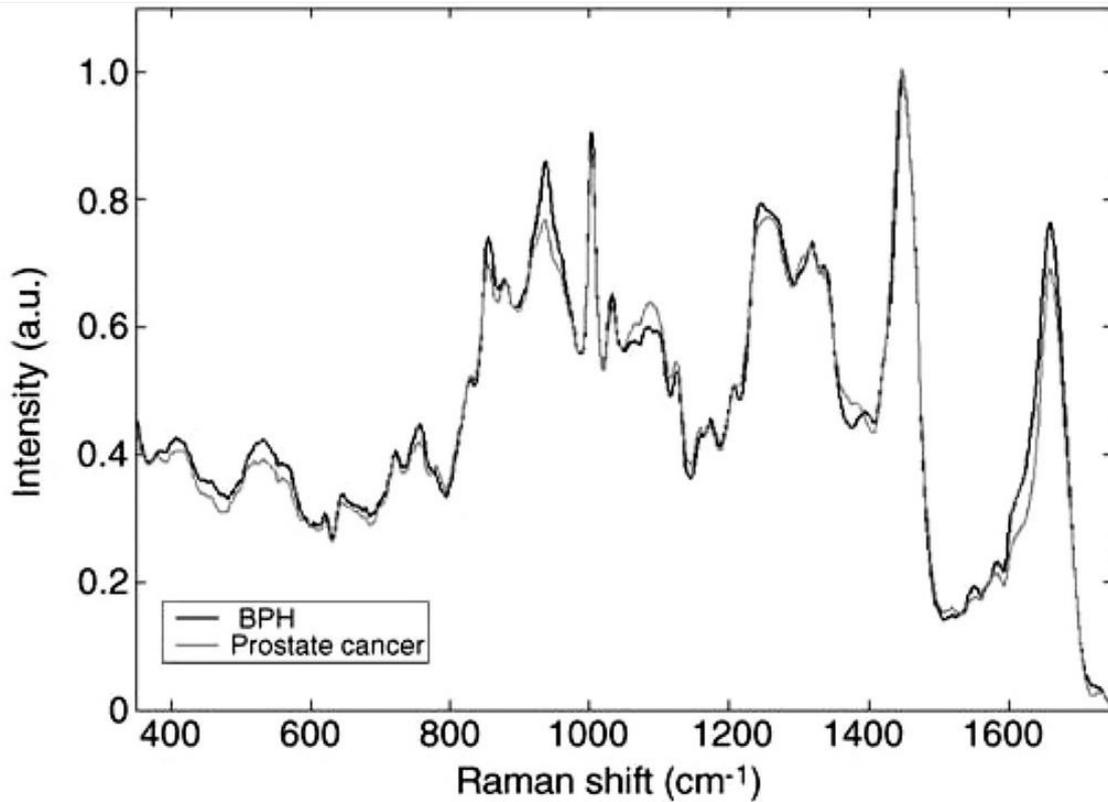


Figure 1.3: An example of typical prostate and prostate cancer Raman spectra taken from Crow *et al.* (2003). These spectra are averages taken from approximately 200 individual spectra from a benign prostate hyperplasia (BPH) and prostate cancer tissue samples. Both spectra have been normalised against the intensity of the 1443 cm⁻¹ peak, which is related to proteins and lipids.

print region, between 600 and 1800 cm^{-1} and the high wavenumber region from approximately 2700-3100 cm^{-1} . The fingerprint region comprises of a complex representation of molecular vibrations from DNA/RNA, proteins and lipids [10]. Due to the complex nature of this region, it is often difficult to extract information from spectra without precise analysis as many peaks overlap and convolve. The high-wavenumber region is dominated by signatures for various CH bonds commonly associated with fatty acids and lipids [10]. The portion of the spectra between the fingerprint and high wavenumber regions is void of biomolecular information and is often left out when acquiring data.

As well as analysing individual peaks within the Raman spectra, it is also common to use the spectra as a whole to discriminate between samples using multivariate analyses, such as a combination of Principle Component Analysis (PCA) and Linear Discriminant Analysis (LDA) [11] [20] [21]. These methods first require the raw Raman spectra to be pre-processed, undergoing baseline corrections, noise removal and normalisation, to ensure that the discrimination is due to the physical characteristics present in the spectra, not the noise or background profile. PCA is a statistical process whereby the data is transformed into a set of principle components (PCs) each expressing the maximum variance in the data set. This transformation is defined such that that the first PC has the greatest possible variance accounting for as much of the variability in the data as possible. Each successive component, in turn, then has the next greatest variance possible under the constraint that it is statistically independent to the previous components. The PCA identifies new variables, the principal components, which are linear combinations of the original variables. The samples can then be plotted in two or three dimensions using their projections onto the first two or three principle components, respectively [22]. Linear discriminant analysis (LDA) is a technique used in statistics to maximise the variance between two groups of variables. LDA can be combined with PCA, where the values of the PCs are in-putted as variables into the linear discriminant, in order to maximise any separation between groups of variables. Any analysis done using LDA must be cross validated, as in Stone *et al.* (2002) [23] by a method such as leave-one-out, where each spectrum is removed in turn, the LDA is performed to produce a model and the removed spectrum is classified using this model. LDA requires a greater level of validation than PCA as it functions to look explicitly for differences between classes, whereas PCA only looks for linear combinations of variables to represent the data as a whole [24].

Another technique which is used to analyse Raman spectra is peak intensity ratio (PIR) analysis. PIR analysis allows the relative intensities of peaks in the spectra to be compared across samples. This allows for the identification of specific biomolecular differences between samples or groups. Some research has suggested that, due to the complex nature of biological samples and their spectra, PIR analysis is inferior to PCA/LDA, owing to the need for individual peaks to be isolated and measured [25]. However, some considerations, such as the idea that the direction of greatest variance may be of no biological interest, must be made about PCA. While PCA is a good tool for discriminating variance between samples [20], PIR analysis has been used to elicit specific biomolecular markers for samples under observation [26] [27]. PCA can also be used to inform PIR analysis such that the principle component loadings can provide the peak locations, which have the greatest variance between samples and these can then be further interrogated by using the PIR method [28].

The interpretation of Raman spectra is well established. Movasaghi *et al.* (2007) published a review of Raman literature on biological samples [10], in which a database of peak assignments correlated to specific biomolecular bond vibrations was created. This database draws from literature on all kinds of biological tissues. The Movasaghi *et al.* review found assessed biomolecular peak assignments across the literature, allowing the creation of a reliable database from which research groups working on Raman spectroscopy studies of biological specimens could draw assistance [10].

Relevant to this thesis is a review of Raman spectroscopy studies of bacteria. RS has been used to monitor changes in populations of *E. coli* bacteria to great effect. Chan *et al.* [29] were able to monitor the proteins expressed by single bacteria cells in real-time by observing markers at 1257, 1340, 1453 and 1660 cm^{-1} , which are known to be characteristic of proteins [10].

A study by Munchberg *et al.* was able to not only use RS combined with PCA/LDA to differentiate between *E. coli* and *pseudomonas* bacteria, but also to identify samples treated with four different antibiotics [30]. This study was only able to provide very basic biomolecular explanations for the PCA/LDA result based on what was already known about the antibiotic treatments, e.g. one antibiotic is known to impair DNA replication in the cells, so a change in DNA content is expected in the spectra [30].

Rusciano *et al.* used RS combined with PCA to differentiate between sputum samples taken from cystic fibrosis patients that had been infected *p.aeruginosa* or *s.aurens* bacteria and those who were not

infected [31]. With some further analysis, PIR or similar, this study could begin to form a biomolecular picture of the difference between the Raman spectra of each sample. The loading presented for the first principle component (PC1) in this study has several very distinct features with biomolecular significance, e.g. lipids, phenylalanine and proteins [10].

Other methods have been applied to the RS study of bacteria and cells, in general. Huang *et al.* (2003) were able to perform single cell RS on bacteria samples using an optical trap [32]. This process involved exposing the sample to high laser power, potentially leading to a significant changes and damage to the cells [33].

A commonly used technique in the RS study of bacteria and biological systems, in general, is *surface enhanced Raman spectroscopy* (SERS) [34]. SERS is a technique where the Raman signal is enhanced by the plasmonic response of metallic substrates or nanostructures. Although SERS has been applied to the study of biological systems [35], it does not enhance the signal uniformly and is not considered to be a label free method.

An important component of any biological system is it's lipid content. In terms of quantitative Raman analyses of cells, Wu *et al.* used lipidomics to study lipid changes in algae cells [26]. They showed that the peak intensity ratio of the 1650 cm^{-1} C=C band to the 1440 cm^{-1} CH_2 band (I1650/I1440) correlates to the degree of lipid unsaturation in extracted lipids from the algae cells [26]. The degree of unsaturation ratios (I1655/I1440 and I1655/I1294), the lipid chain length ratio (I2930/I2959), and the total unsaturated fatty acids to total amount of fatty acids (i.e., TUFA/TFA ratio I3015/I2851), have also been used to study lipid droplets in treated and untreated breast and prostate cancer cells [27]. Although the peak assignments of other lipid-related peak intensity ratios (i.e., degree of lipid unsaturation and chain length) relate also to proteins, in the study of whole cells, the TUFA/TFA ratio considers specifically lipid vibrations [10], and thus can be used to interrogate the fatty acids in intact cells. The mass unsaturation ratio (I1653/I1447) can vary with both the degree of unsaturation and the chain length of lipids in the sample, making it a less precise measure of lipid content [27] [26]. The peaks at 1447 cm^{-1} and 1653 cm^{-1} are also convolved with various other peaks assigned to both lipids and proteins [10] furthering the case against its use as a measure of unsaturated lipid content in intact cells.

Lipid content within a sample is commonly analysed using gas chromatography/mass spectroscopy, but information such as the cellular dynamics and the lipids distribution [27] are lost during the homogenization process. These techniques cannot be used *in vivo* or for live cell studies, but with the latest techniques, Raman spectroscopy has become a tool for analysing the lipid content of biological samples.

One of the key characteristics of Raman spectroscopy is that it can be used non-destructively. It is therefore important to ensure that when performing live cell studies the cells themselves are not adversely affected by the exposure to laser light [36]. Hence it is necessary to carry out tests to ensure that the Raman experiment does not interfere with the results by changing or damaging the cells [37] [38]. A simple test involves taking multiple spectra from the same cell and analysing to determine if there is any deterioration in spectral quality over time. A more complete test is a ‘Trypan Blue Assay’ [39], Trypan blue is a dye which only permeates dead or damaged cells [40]. The trypan blue assay is used in biological research as a cell viability assay, to determine whether a cell culture contains dead cells, or if a particular procedure has had a damaging effect on a cell culture. It has been applied to live cell RS by Oshima *et al.* (2010) [38], who reported that for a 532nm laser with a power of 50mW and an exposure time of 60 seconds, the trypan blue assay was negative for lung cancer cells. Notingher *et al.* (2002) [41] also reported that for a 785nm laser with a power of 115mW and an exposure time of 120 seconds the assay also returned a negative result for lung epithelial cells taken from mice. Both studies used a trypan blue assay to confirm that RS could be used to study living cells, non-destructively.

Similarly, dried cell samples are also susceptible to the effects of long term exposure to laser light, although the laser is less likely to radically alter the biology of a dried sample compared to a live cell sample, there is still a significant risk of burning, characterised by obvious black/brown marks on the sample, over saturation of the CCD, or by the appearance of broad bands centered at 1500 cm^{-1} signifying amorphous carbon [42]. This type of damage can be avoided by lowering the intensity of the laser using a filter and shortening the exposure time of the sample, while still maintaining a strong enough signal to provide meaningful results [43].

1.3 Prostate Cancer Cells

Prostate cancer (PCa) is the most commonly diagnosed cancer in men [44]. There are approximately 30,000 new diagnoses every year in the UK alone, which accounts for 23% of male cancers [44]. PCa results in around 10,000 deaths in the UK every year, giving the disease a mortality rate second only to lung cancer amongst cancers [45]. While an early diagnosis can lead to a ten-year survival rate of around 85%, the life expectancy for a patient with advanced PCa can be as low as 12 months [36] [46]. In its early stages PCa is virtually symptomless and only becomes noticeable when the tumour is of a size large enough to affect the passage of urine to the penis [47].

The current diagnosis methods for PCa include a prostate specific antigen (PSA) test [48], a digital rectal examination (DRE) [49], or a prostate biopsy, where a physical tissue sample is taken from the patient for testing. A PSA test simply measures the amount of PSA present in a patients blood, as a tumour on the prostate causes an increase in its production. However, this test is considered unreliable as PSA is not specific to cancerous growth. Enlarging or swelling of the prostate can also cause an increase in PSA in the blood and, in turn, lead to a false positive diagnosis of a tumour. With 78% of test results later being determined to be false [48]. Physical examinations, as well as being time consuming and uncomfortable, can also lead to misdiagnoses due to the human element of the procedure [50]. These factors together provide significant justification to look for new, cost-effective, more discriminatory and patient friendly methods for interrogating prostate tissue.

The current standard of grading PCa is the Gleason grading system [51]. A Gleason score is given to a cancer based on the microscopic appearance of the prostate gland patterns, or histopathology, with a higher score equating to a more aggressive tumour [52]. In general, a lower grade corresponds to prostate tissue with small, uniform glands [51]. The grade increases as cells become more spread out and glands only occasionally form [51]. The Gleason grade is a sum of two Gleason scores assigned to the two most commonly observed gland patterns on an individual prostate by a pathologist [53]. Each score is from one to five, making the final Gleason grade a number between two and ten. Gleason grades are often grouped according to similar biological behaviour: (i) low-grade, or ‘well differentiated’, from two to six, (ii) intermediate-grade at seven and (iii) high-grade, from seven to ten [54].

Early cancer research has suggested that cell metabolism may provide many of the key features

exhibited by malignant cells [55]. One of the main hallmarks of a cancerous cell is rapid growth and proliferation. In order to grow and divide at a higher-rate, the cell must develop a degree of autonomy and does so by a process called *de novo lipogenesis* [56]. De novo lipogenesis is the conversion of excess carbohydrates into fatty acids. These fatty acids in turn support the increased need for membranes in accelerated cell growth and proliferation [57].

Past RS research has found that lipid markers are a valuable tool for discriminating cancer cells from non-cancerous cells [27] [58]. Relative peak intensities at 1450 cm^{-1} (CH proteins and lipids) and 1305 cm^{-1} (CH bending, CH twisting of protein) have been found to increase for various types of dysplasia [32] [58] [59]. Such markers indicate higher concentrations of lipid biomolecules within the cells with, lipid accumulation being thought to be a hallmark of aggressive cancer cells [60]. The total unsaturated fatty acid to total fatty acid TUFA/TFA ratio (I3015/I2845) was also used in the literature [27] [61]. A lower TUFA/TFA ratio, or relative unsaturation degree, is indicative of de novo lipogenesis, which is thought to be a hallmark of malignancy in cells [62] [63]. The unsaturation degree, however, is not always sensitive enough to provide a distinction between benign and malignant cells, as in Neiva *et al.* (2012) [61].

RS peaks at 1000 cm^{-1} and 1586 cm^{-1} , corresponding to phenylalanine have been proposed by Huang *et al.* and Li *et al.* as markers for malignancy [32] [64]. Both references showed that phenylalanine increases in cancer cells [32] [64]. Research by Fu *et al.* found that melanoma cells are dependent on phenylalanine as its restriction causes multiple changes in cell behaviour [65].

RS has been used to quantitatively analyse the DNA/RNA expression of malignant cells [66] [58]. The literature suggests that the Raman band at 1335 cm^{-1} (nucleic acids and proteins [10]) shows higher relative intensities in cancerous cells compared to normal cells when normalised to the 1447 cm^{-1} peak [58] [66] and when normalised to the total area under the curve [32]. Raman bands between 721 and 827 cm^{-1} representing nucleic acids and the DNA backbone are more broadly categorised as DNA [10]. This region in the Raman spectra shows a similar relationship to that of the 1335 cm^{-1} peak, in that higher intensities in this region generally correspond to increasingly aggressive tumours [20]. Specifically, the intensity of the DNA peak at 780 cm^{-1} (phosphate bond, cytosine, thymine) is said to correlate strongly with malignancy in breast tissue [3]. The peak around 780 cm^{-1} has also been shown

to decrease dramatically when living lung cancer cells undergo programmed cell death, or apoptosis [19].

Another area in which RS has proven to have significant discriminatory power is the study of proteins and individual amino acids. Studies have found that aggressive tumours, specifically metastatic tumours, express greater levels of various proteins [8] [67] shown by the increased relative intensities of peaks in the Raman spectra at 643, 760, 853, 937, 1000 and 1032 cm^{-1} [8] [10] [67]. Previous work has been able to use these bands amongst other protein signals, combined with other peaks in the fingerprint region, to fully discriminate between malignant and non-malignant cells and tissue samples [67].

Crow *et al.* [11] and Shetty *et al.* [68], respectively, found that glycogen, with peaks at 850 cm^{-1} , 934 cm^{-1} and 1029 cm^{-1} in the RS spectra, is a marker for malignancy in both prostate [11] cancer and oesophageal [68]. Both sets of research found that lower glycogen content is indicative of malignancy. Non-RS research, however, has found repeatedly that cancer cells exhibit higher glycogen concentrations [69] [70]. This disagreement highlights a key area for RS and this study in that the spectral results must represent the biology of the cells. Such differences may be caused by low RS sampling, inaccurate analysis or contaminated or damaged samples. More research is therefore needed to develop the RS as a biomolecular tool.

The above examples of RS as a means to interrogate cancer cells involve a variety of sample types. Most of the published research uses standard cell lines, which are immortalised cell lines originally derived from a specific organ or tumour site. For example, the LNCaP cell line used in some of the literature [11] [67] was derived originally from a lymph node metastasis of prostate adenocarcinoma (PCa) and have been used extensively in the field of oncology as a standard [71]. Another example is the standard cell line PNT2-C2 which comprises cells derived from normal prostate tissue that have been immortalised so as to act as a control against which malignant cells can be compared [72]. Although these immortalised cell lines provide a good standard and a high degree of reproducibility, they do not fully mimic clinical disease [73].

Primary, patient-derived cell lines are cell lines and have come directly from a patient and often come in matched pairs, i.e. from a tumour and from normal prostate tissue in the same patient. Primary samples are often more difficult to obtain and maintain, however, they have been shown to offer advantages over standard cell lines [74] [75]. For example, studies which use primary samples gathered

from multiple patients allow the researchers to capture the inherent individual and interindividual heterogeneity of biological tissue [11]. These samples also eliminate the likelihood that the results obtained are cell line specific as there have been several reports in the literature regarding the mis-identification and cross-contamination of cell lines [74] [75]. Primary patient-derived samples not only open the door for techniques such as RS to provide better diagnoses using multivariate analyses, as well as to provide patient specific information based on individual peak intensity ratio (PIR) analysis. An important next step for RS is to determine the difference in spectral information between biologically similar cell lines and primaries, and to ascertain whether, for use in RS, standard cell lines give a clinically accurate molecular fingerprint of PCa.

1.4 *Streptomyces coelicolor* Bacteria

As discussed earlier, RS is a powerful technique for analysing a variety of biological samples. One of the key areas of RS research is in the study of bacteria. *Streptomyces coelicolor* (*S. coelicolor*) is a bacteria belonging to the *Streptomyces* genus commonly found in soil [76]. This particular strain is responsible for the production of a large variety of antibiotics for both human and veterinary use [76]. The wider study, which this RS study of *S. coelicolor* is to be a part of refers to the development of antibiotic hypersensitive mutants of *S. coelicolor* [1]. In mycobacteria, polyphosphatase (*ppm1*) is an essential enzyme required for lipomannan biosynthesis where lipomannan is a major constituent of the cell wall [77]. Mutations in *ppm1* lead to increased susceptibilities to multiple antibiotics, most of which act at different stages of cell wall biogenesis. In the RS portion of this study, three different strains of *S. coelicolor* are analysed, a wild type bacteria (J1929), mutant type, that has a mutation in the *ppm1* gene (DT3017) and the mutant type treated with plasmids encoding the wild type *ppm1* gene (PDT16) [78] [79].

RS has been used in the fundamental characterisation of bacteria, for infection diagnostics and in antimicrobial treatment development [30] [80] [81] [82]. RS has also been used for the specific characterisation of *Streptomyces* bacteria [80] [83]. Walter *et al.* (2011) used RS to classify different species of *Streptomyces*, noting particularly high variance in the high wavenumber 2700–3100 cm^{-1} . Also concluding that due to its filamentous nature, classification of *Streptomyces* bacteria using a single cell approach

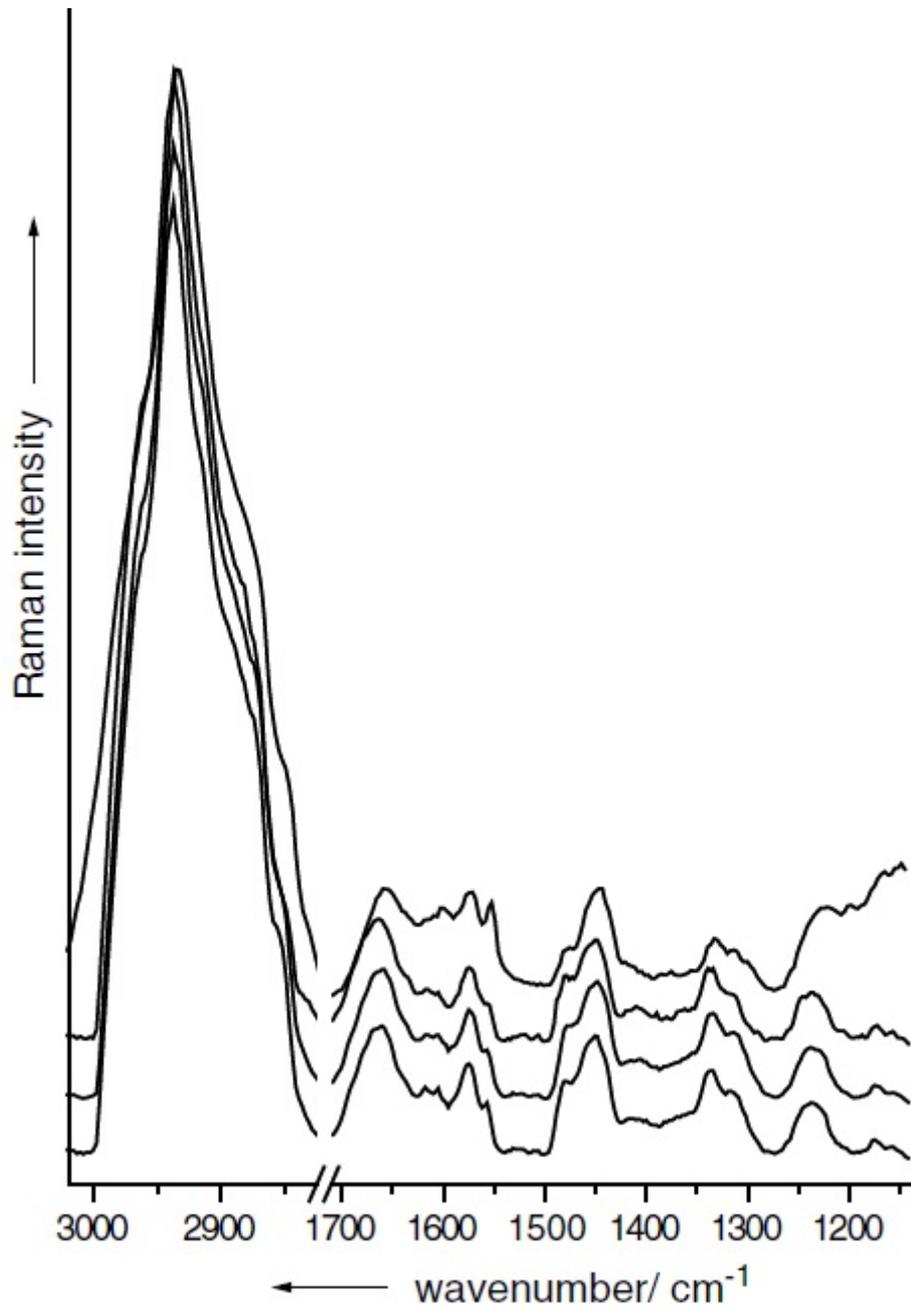


Figure 1.4: An example of Raman spectra of *Streptomyces* bacteria, from Walter *et al.* (2012).

Four *S. coelicolor* Raman spectra are shown. From bottom to top, the un-contaminated sample and three samples with increasing levels of contamination with Cl⁻ and Ni²⁺

cannot be expected to be reliable, hence their use of bulk samples [83]. Walter *et al.* (2012) used RS to classify samples of *Streptomyces* bacteria which had been exposed to different levels of heavy metal contamination [80]. Fig. 1.4 shows an example of Raman spectra taken from *Streptomyces* bacteria by Walter *et al.* [80]. This figure shows four *Streptomyces* Raman spectra. From bottom to top, the un-contaminated sample and three samples with increasing levels of contamination with Cl⁻ and Ni²⁺ [80].

Studies have verified Ramans ability to not only discriminate bacterial species, but also its sensitivity to distinguish at the sub-species level as well [83]. RS is especially useful as a tool for characterising bacteria as it requires no invasive preparation, which can lead to contamination and altered findings [84].

In this work, the TUFA/TFA lipidomic ratio is used for the first time to determine changes in the lipid content of J1929 (wild), PDT16 (treated) and DT3017 (untreated, *ppm1* mutant) *S. coelicolor* cell lines. These results will inform the larger biological study of any changes in lipid membrane composition (unsaturated to total fatty acids) between the three bacteria types. It is expected that the wild type and treated mutant type should have a similar lipid content, whereas the untreated mutant type will have undergone a change in membrane composition, therefore a change in lipid content and hence TUFA/TFA ratio is anticipated.

1.5 Summary of Research Objectives

This work will examine RS as a tool for interrogating biological cells. Drawing on previous work in the field of PCa research, this project will examine the use of both PIR analyses and PCA/LDA as tools for determining the molecular composition of prostate cancer and normal prostate samples. These data will be tested for their ability to discriminate not only tumour samples from normal tissue samples, but also to assess patient-specific information from patient derived primary cell culture samples. Alongside these investigations there will also be a critical analysis of the use of standard cell lines in RS for cancer research compared to the use of primary samples. The RS analyses will also be applied to the study of *S. coelicolor* bacteria. Considered by this work will be the use of RS as a tool for discriminatory characterisation of *S. coelicolor* bacteria to a sub-species level, specifically the ability of RS to discriminate between three strains of *S. coelicolor* bacteria having undergone a mutation and subsequent gene treatment. This is to

determine the effect the mutation and gene treatment have on the lipid composition of the *S. coelicolor* bacteria.

Chapter 2

Methods

2.1 Prostate Cancer Cell Culture

The two patient-derived primary cell culture samples used in this research are labelled H517-15 and H554-15. These samples are both matched pairs meaning that both normal and cancer samples were taken from the biopsies of each patient. As determined on histopathology, the normal samples are representative of normal prostate tissue and the cancer samples are Gleason seven graded meaning the cancer is of intermediate aggressiveness [51] [52]. In general, the location of prostate tumours is determined by MRI, with the pathologies diagnosed by needle biopsies prior to surgery. On the day of the surgery, either a transurethral resection of the prostate (TURP), or a radical prostatectomy is performed, the latter involves the removal of the entire prostate. The tissue procurement officer collects the prostate tissue from the surgeon in the operating theatre and takes two needle punch biopsies, one from a normal area and one from the tumour site. Only samples where the tumour is palpable are used. In addition to the pathologies on the needle punch biopsy samples, confirmatory pathology is also performed on the the sites surrounding the biopsies to ensure accuracy in the final diagnosis of the samples [85]. After a sample is collected from surgery it in is placed into a transport medium (RPMI Medium 1640 plus 10% fetal calf serum, 2 mM l-glutamine, and 1% ABM (antibiotic/antimycotic)). These samples are then delivered to the Cancer Research Unit (CRU) at the University of York, and then processed according to the Unversity of York CRU published protocols [85] [86] [87] before being

frozen in liquid nitrogen for storage.

The two standard cell lines used in this research are P4E6 and PNT2-C2, which are thought to be representative of cells from a Gleason four tumour and a normal prostate sample respectively [72] [88]. While the P4E6 cell line is of a lower Gleason grade than the primary samples used in this study, the consensus in the literature is that cancer specific biomarkers apply to a range of stages. Crow *et al.* (2003) observed that for all cancers (Gleason less than seven, Gleason seven and Gleason greater than seven), when compared to normal prostate cells, show the same trends [11]. It is therefore hypothesised that when compared to its normal match, the P4E6 cell line should match the trends seen for the primary sample pairs. The PNT2-C2 normal sample is taken from normal prostate tissue and is therefore most closely representative of the primary normal samples used in this study and can thus be used to compare direct patient-derived primary culture cells and standard cell lines.

All of the cell culturing in this study was performed in a sterile environment under a fume hood at the University of York CRU. Both the primary cell cultures and standard cell lines, were frozen in liquid nitrogen for storage.

Prior to RS analysis, the cell sample is mixed with growth media and put into a T25 flask to be split. The growth media is different depending on the cell line. The P4E6 cell line requires K2 media, which comprises KSFM, supplements, 2% FCS and L-glutamine. The PNT2-C2 cell line requires R10 media, which comprises RPMI, 10% FCS and L-glutamine. Finally, the primary patient samples are cultured in *stem cell media* (SCM), which comprises KSFM, supplements, GM-CSF, cholera toxin, SCF, LIF and L-glutamine. Half of the cell sample is then re-frozen to later use again, while the other half of the cell sample is placed back into a T25 flask together with fresh growth media and incubated at 37°C. The growth media for each cell sample is replaced with fresh media every two to three days. Cell preparation must begin at least two days prior to RS analysis. Two days prior to RS analysis the growth media is removed from the T25 flask containing the growing cells and the remaining cells are then washed with 3 ml of phosphate buffered saline (PBS). 1.5 ml of trypsin is then added to separate the cells from the flask and following this the flask is then incubated for a few minutes at 37°C. At this point the flask is then inspected to check that the cells are coming off it, and once achieved, they are then re-suspended in 5 ml of growth media.

From the cells and media solution, 10 μl of cells is then used in a haemocytometer to determine the number of cells per unit volume for plating. For a CaF_2 disc, 13 mm in diameter and 1 mm thick, placed into a 35mm petri dish, 50,000 cells had to be plated to achieve a good degree of cell confluence. Once plated onto the CaF_2 disc, an appropriate amount of growth media is then added to the petri dish containing the disc (2.5 ml for a 35 mm petri dish) together with antibiotic/antimycotic (ABM) to prevent cell contamination. The plated CaF_2 disc and petri dish are then incubated for 24 hours. The next day the media in the petri dish is changed to starvation media, which is the growth media minus the nutrient serum, i.e. K2 and SCM starvation medias are just KSFM, and the R10 starvation media is just RPMI. Starving the cells synchronises them into the G_0 (resting) state of the cell cycle as RS has been shown to be sensitive to cell cycle changes [89]. On the day of the Raman analysis the CaF_2 disc and petri dish are washed twice in Hanks balanced salt solution (HBSS), and fresh HBSS (2.5 ml) is added. HBSS is used as it keeps the cells from drying out, as well as adding minimal background signal to the Raman spectra [38].

2.2 *Streptomyces coelicolor* Cells Culture

<i>S. coelicolor</i> cell line	Description
J1929 (wild)	Wild type <i>S. coelicolor</i> bacteria.
DT3017 (untreated)	<i>S. coelicolor</i> bacteria with a mutation in the <i>ppm1</i> gene, causing a decrease in intrinsic antibiotic resistance.
PDT16 (treated)	The mutant type bacteria after treatment with plasmids encoding the <i>ppm1</i> gene. This treatment is designed to reverse the effects the mutation has on the antibiotic resistance of the <i>S. coelicolor</i> bacteria.

Table 2.1: Table showing the three samples and their descriptions [1].

All bacteria culturing methods for the *S. coelicolor* cells have been adapted from the procedure set out in Kieser *et al.* (2000) and summarised here [90]. All cell culturing here was performed in Prof. Maggie Smith's lab, at the University of York, by *Dr Nicholas Read* and *Dr Robert Howlett*. The following process is identical for each of the three strains to be analysed in this study, detailed in Table. 2.1. A dense bacteria spore suspension (500 μL c.1010 spores per ml), of one of J1929, PDT16 or DT3017, is heat-shocked at 50°C for ten minutes in 10 ml TES (0.5 M pH8), then cooled under cold tap water. The bacteria culture is then added to 10mL double-strength germination medium in a 250 ml baffled conical flask and shaken at 37°C at 180 rpm for 8 to 9 hours. The double-strength germination medium comprises Difco yeast extract, 1 %, Casaminoacids, 1 %, CaCl_2 , 0.01 M, and is prepared and autoclaved separately as a 5 M solution. The bacteria culture is then spun down and suspended in 10 ml Difco nutrient broth. The OD450nm is measured and the culture diluted in Difco nutrient broth to obtain 40 ml with an OD450nm of 0.05.

The 40 ml culture is then grown in a 250mL baffled flask at 30°C for 12 to 13 hours, after which 10 ml is taken for dry weight measurements. To dry the culture, it is spun down, washed and spun down again with the resulting pellet put into a pre-dried tube. The remaining 30 ml of the bacteria culture is spun down and washed with MilliQ water ready for plating.

Cell samples from each of the J1929 (wild), PDT16 (treated) and DT3017 (untreated, *ppm1* mutant) bacterial cultures were washed in high-purity deionised water to remove the bacterial culture media and then plated via pipette-deposit directly onto small CaF_2 disks (13 mm diameter and 1mm thick). Bacteria were then gently spread over the disk surface with excess water removed by pipette aspiration. The samples were air-dried in a fume cupboard in preparation for Raman analyses.

2.3 Raman Spectra Acquisition

The Raman laser spot size (d_l) for any optical system is given by

$$d_l = \frac{1.22\lambda}{NA} \quad (2.1)$$

where λ is the laser wavelength and NA is the numerical aperture of the objective lens [6] [7].

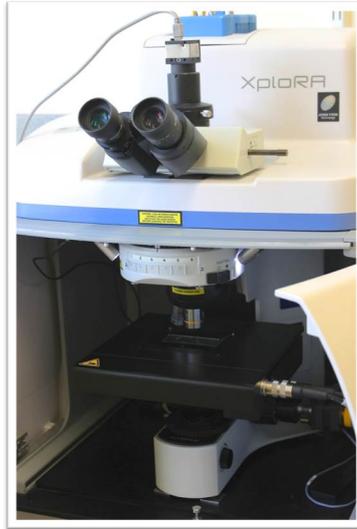


Figure 2.1: The Horiba XploRA instrument used for all of these experiments.

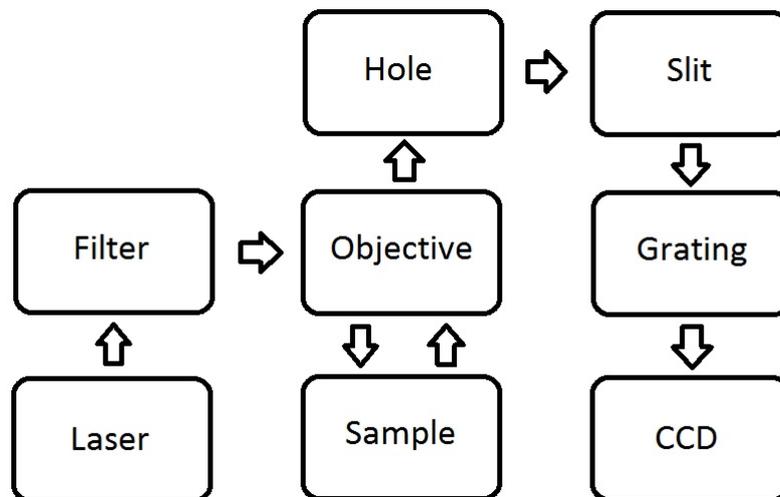


Figure 2.2: Basic flow schematic of a typical confocal Raman spectroscopy set-up, representative of this experimental set-up.

From the equation for lateral resolution,

$$\Delta x = \frac{0.61\lambda}{NA} \quad (2.2)$$

it can be seen that the shorter the wavelength of the laser (λ), the higher the resolution will be, where Δx denotes the smallest resolvable distance between two points and NA is the numerical aperture of the objective lens [6].

The values for the following parameters on the Horiba XploRA Raman instrument (Fig. 2.1) which were used in these studies will be given in the relevant sections. Fig. 2.2 shows a flow schematic of the Raman instrument used in this study. The grating in the spectrometer disperses the Raman signal onto the CCD detector by deflecting each wavelength at a different angle. A higher number of grooves/mm in the spectral grating results in a greater spectral resolution [91]. The *hole* setting on the Raman refers to the confocal hole, which is used to define the resolution depth of the measurement. A larger hole allows more than one focal plane of the sample to be measured, resulting in a greater intensity in the signal [91]. However, for thin samples this may result in unwanted readings from other materials such as the microscope slide. For the Raman experiments in this thesis, only confocal settings were used, meaning spectra were only taken from single focal planes. The ‘slit’ setting is used to change the spectral resolution of the measurement where a greater slit width results in a decrease in spectral resolution [6] [91].

Two types of corrections were performed as part of Raman measurement acquisition. When the *intensity correction system* (ICS) is turned on, a correction factor is applied to the spectrum so that measurements taken with different Raman acquisition settings yield the same results. When the *dark correction* is switched on, it subtracts a pre-determined *dark file* is subtracted from the spectrum correcting for the detector response and pixel variation.

To control the incident laser power at the sample and therefore avoid sample damage a neutral density filter was used. Raman measurement accumulations could also be set to acquire multiple accumulations of spectra data and to average them in order to improve the final spectral quality.

Before spectra were acquired a wavenumber calibration using a silicon reference sample was performed. After bringing the silicon reference into focus under the 100x objective lens (numerical aperture = 0.9), Labspec calibrated all lasers and gratings automatically [91]. To check the alignment of the

laser, an ink spot was drawn on to a glass microscope slide using a dry wipe marker pen. The laser was then used to burn a small part of the ink away. This burn mark was then viewed optically and aligned with the cursor on the video display.

2.4 Trypan Blue Assay

In order to confirm that extended exposure to the laser has not damaged the cells a, trypan blue assay was performed [38]. An image was taken of three individual cells before exposure to any laser light with the cell locations recorded using the x-y-z motorised stage. Raman spectra were then collected on each of the three cells and after this optical images were then taken. The HBSS was then removed and a 0.4% solution of trypan blue dye was added to the disc and allowing time to enter any damaged cells. The trypan blue was then gently washed away using HBSS before more HBSS was added back to the plate. Optical images were then taken of each of the three cells to confirm that they were undamaged. As a control, the cells were then killed with ethanol and a second trypan assay performed to ensure that the trypan was functioning as intended [39].

2.5 Raman Measurements of Prostate Cancer Cells

All live cell experiments on the H517–15 and H554–15 primary patient derived cancer and normal pairs, PNT2–C2 normal and P4E6 cancer standard cell lines were performed under the same conditions using the same acquisition parameters. Individual cells were measured using the HORIBA XploRA micro-Raman instrument with 532 nm laser-light, 63x dipping lens (NA=1) and 2400 lines/mm spectral grating. All measurements were made in the confocal mode using a slit of $200\mu\text{m}$ and a hole of $100\mu\text{m}$. A resulting laser spot size of approximately $1\mu\text{m}$ was used with a 50% filter, resulting in 3.5mW laser power at the sample. As the laser light permeates HBSS, it is likely that the power at the sample is even further reduced. With these equipment settings the Raman microscope has a spectral resolution of 1 cm^{-1} and lateral resolution of $0.3\mu\text{m}$.

Spectra were acquired in the $600\text{--}1800\text{ cm}^{-1}$ (fingerprint) region and the $2700\text{--}3100\text{ cm}^{-1}$ (high-wavenumber) region, using HORIBA LabSpec 6 software, with an acquisition time of 45 seconds averaged

over two spectral accumulations. Between 150 and 170 spectra were acquired per cell type, with these spectra then averaged to obtain a single representative spectrum for each cell sample. Each single spectrum was taken from a new, randomly selected individual cell nucleus, with great care taken not to analyse the same cell more than once to avoid over exposure to the laser.

2.6 Raman Measurements of *Streptomyces coelicolor* Bacteria

Raman point measurements were performed on randomly selected *S. coelicolor* cells in the J1929 (wild), PDT16 (treated) and DT3017 (untreated) populations using a HORIBA XploRA micro-Raman instrument with 532 nm laser-light, x100 objective (0.9 NA) and 2400 lines/mm spectral grating. All measurements were made in the confocal mode using a slit of 200 μm and a hole of 100 μm . The laser spot size was approximately 1 μm , and with a 50% filter, this resulted in a 3.5 mW laser power at the sample. Using these equipment settings the Raman microscope has a spectral resolution of 1 cm^{-1} and a lateral resolution of 0.4 μm .

Spectra were collected using the HORIBA LabSpec 6 software in the 600–1800 cm^{-1} fingerprint region and the 2750–3110 cm^{-1} high-wavenumber region. Each individual spectrum was obtained using a 90 second acquisition time averaged over two spectral repetitions. Between 40 to 55 spectra were randomly collected across the population of each cell line. With these spectra then averaged to obtain a single representative spectrum of each sample.

2.7 Prostate Cancer Spectra Analysis

All the spectra analysis was performed using IGOR Pro 6.35. The *Average waves* package in IGOR Pro was used to create an average spectrum and standard error of the mean (SE) wave from each data for each cell line. All of the peaks in the average spectrum and SE wave were then fitted using the package *multipeak fitting 2* (MF2). The fitting using the MF2 package involved first performing a linear or cubic baseline correction to the individual peak or cluster of peaks (if convolved) before assigning a Gaussian profile to the peak locations Fig. 2.3. From this baseline and fitted Gaussian peak(s), the peak heights were then determined. For display purposes, a linear baseline correction was also performed across the

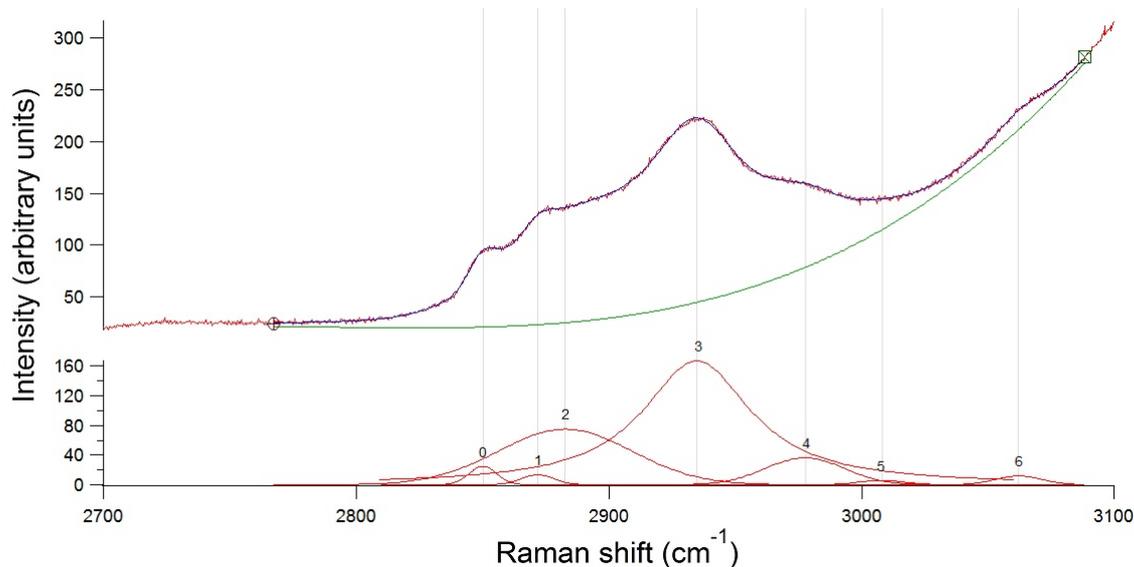


Figure 2.3: An example use of the multippeak fitting package to fit a highly convolved part of a spectrum. The example here is the H517–15 cancer sample between 2700 and 3100 cm^{-1} . This region contains the peaks, 2850, 2871, 2893, 2937, 2970, 3008 and 3065 cm^{-1} . The figure shows the averaged spectrum, Gaussian peak profiles and cubic baseline.

whole spectra using *The Raman Toolset* software [92]. Normalisation was also performed against the 1447 cm^{-1} peak for direct comparison of the spectra [11].

Principle component analysis (PCA) followed by linear discriminant analysis (LDA) analysis was performed on all six samples using the statistical package R v3.3.0. LDA was only used on the six samples combined to maximise the variance and achieve separation between the cell types. The cell separations were validated using leave one out cross validation [11], whereby each individual spectrum is left out of the LDA model in turn, before being classified using the groups returned by the model. This process is repeated for each and every spectrum in order to generate a picture of the overall accuracy of the LDA model on these data sets. Kaiser’s criterion, which states that on a standardised dataset only principle components (PCs) with a variance greater than one are significant [93], was used to determine the number of PCs for LDA. For the sample pair comparisons, only PCA was required to achieve the maximum possible separation of the samples.

Before loading the spectra into R, each individual spectrum was first linear baselined and normalised

to the total area under the curve, using the Raman toolset, to ensure that the variances picked up by the algorithm correspond to biomolecular differences and not to artefacts from the spectral acquisition. The loadings of the principle components responsible for separating the samples were analysed to determine regions in the spectra of greatest variance. These results could then be used to inform a deeper analysis of the peak intensity ratios of those areas [28].

To ensure sufficient data has been collected, SE waves were produced each time five more spectra were added to the average. As the number of spectra approached the maximum number collected for each cell line (between 150 and 170 spectra) it was expected that the SE wave would converge. This process of testing the SE convergence was repeated for both the fingerprint and high wavenumber regions. To ensure solid statistics in the analyses, peak intensity ratio (PIR) convergences were also performed as a function of the number of spectra in the spectral average for each cell line.

The unsaturation ratio (I1447/I1663) and the TUFA/TFA, or total unsaturated fatty acid to total fatty acid ratio (I2850/I3008) were selected to show the statistical convergence as they have been previously shown in the literature to differentiate between normal and cancer cell samples [27] [61]. The convergence of the percentage SE of these ratios was also calculated as a function of the number of spectra in each spectral average as a means of establishing data convergence. When directly comparing PIRs a T-test was performed to ensure validity, as in Oshima *et al.* 2010 [38].

The SE in the PIRs were calculated using the values from the SE wave, according to

$$dz = z \sqrt{\left(\frac{dx}{x}\right)^2 + \left(\frac{dy}{y}\right)^2} \quad (2.3)$$

where x and y are the peak intensities, z is the PIR and dx , dy and dz are their associated standard errors [7].

2.8 *Streptomyces coelicolor* Spectra Analysis

The averaged spectrum for each cell line was baseline-corrected from the first to the last spectral point in the fingerprint and high-wavenumber ranges using the ‘Raman tool set package [92]. Peak regions in the acquired spectral ranges were then linear baseline-corrected and Gaussian peak-fitted using the software package IGOR Pro 6.35. The peak fittings were used to obtain the average peak positions

measured across the cell-lines. Individual peak identifications were made by correlating the average peak positions with literature references [10] [80]. The total unsaturated fatty acid to total fatty acid (TUFA/TFA) peak intensity ratio (I2850/I3008) for each cell line was determined from each averaged spectrum, with its uncertainty calculated from the SE uncertainties in the measured TUFA and TFA peak intensities. The SE in the ratio was then calculated using Eqn.2.3. To ensure sufficient number of spectra was collected to represent each sample population, the convergence of the percentage standard error of the mean (SE) of the TUFA/TFA ratio was calculated as a function of the number of spectra in each spectral average.

Chapter 3

Results and Discussion

3.1 Raman Spectroscopy Studies of Prostate Cancer and Normal Prostate Cells

This section of the study aims to use Raman spectroscopy to analyse biomolecular differences between normal and cancerous prostate cells. Firstly, principle component analysis and linear discriminant analysis were used to separate the samples and provide a basis for deeper spectral analysis via the loadings of the principle components. Using examples from the literature and information from the loadings of the principle components, peak intensity ratio analysis was used to gather the information on the biomolecular differences between the normal and cancer samples presented in their spectra. Aside from critically analysing the use of RS in cancer research, this study will also observe and critique the use of immortalised standard cell lines to represent patient biology by comparing the biomolecular information presented by patient samples and standard cell lines.

Six cell line samples were used in this study comprising three normal and three cancer samples. These samples consist of the H517–15 and H554–15 primary patient samples (Gleason 7 cancer and normal matches), the P4E6, Gleason 4, prostate cancer cell line and the PNT2–C2 normal prostate standard cell line. The H517–15 and H554–15 primary patient cancer samples were both graded Gleason 7 at first histopathology.

Using principle component analysis (PCA) this study was able to differentiate normal from cancer

cells for both the H517–15 primary and P4E6/PNT2–C2 standard cell line pairs. However, this study was not able to differentiate normal from cancer for the H554–15 primary patient samples.

Peak intensity ratio (PIR) analysis undertaken by this study was able to identify several cancer biomarkers for the H517–15 primary sample, consistent with the literature on cancer and normal cells. Some of these biomarkers include lipids, glycogen, phenylalanine and DNA/RNA. PIR analysis corroborated the finding from PCA in that the H554–15 cancer and normal samples are biomolecularly similar. Differences in these samples were found in the heavily convolved peaks related to DNA, lipids and proteins with these being the only discriminatory markers for this sample. Although separated by PCA, the picture with the standard cell lines is less clear. Some markers for these cell lines (e.g. total unsaturated fatty acids against total fatty acids (I2850/I3008)) are corroborated by the literature, some are not (e.g. phenylalanine (I1000/I1447)), bringing forth the question as to whether these cell lines provide an accurate representation of patient biology.

3.1.1 Trypan Blue Assay

In order to prove that exposure to laser light does not cause damage to these cells, a trypan blue assay was performed [39]. Notingher *et al.* (2002) performed a trypan blue assay as part of the first ever live cell RS study [41]. The trypan blue assay in this study was performed on the P4E6 standard cell line, which was chosen due to the limited availability of primary samples for this study and the relative abundance of standard cell lines. RS is well documented in the literature as being non-destructive to single cells [26] [94], at much greater laser exposures than that used in this study. Hence, a trypan blue assay would be a definitive check on the non-destructive testing used in this work.

The results of the trypan blue assay are shown in Figs. 3.1–3.3. The optical images show the single cell that was tested using this procedure (in the red square). This test was also repeated for a further two cells yielding the same result. In Fig. 3.1 the first cell is shown before any exposure of it to laser light or dye. The cells that were selected for this assay display no visual abnormalities or contaminations as per the selection criteria for all cells in this study. Fig. 3.2 shows the same cell after exposure to laser light under the same conditions as used for Raman measurements, as outlined in the methods section, and the addition of trypan blue dye. In the case that the exposure had caused damage to the

cell, the dye would have been uptaken by the cell. As there was no uptake of the dye by the irradiated cell, Fig. 3.2 therefore confirms that laser exposure, at the power and acquisition time used for Raman measurement in this study, does not cause damage to the cells.

Fig. 3.3 shows the result of a second trypan blue assay where the cells have first been killed with ethanol before the dye is added. In this case, the dye was uptaken by the damaged cells, therefore confirming the result of the first assay where no cell damage occurred. Based on these results, it can therefore be concluded that the Raman spectra obtained in this work provide a fingerprint of unmodified living cells. This result is supported by literature on the use of trypan blue assays to determine cell viability, for example, Strober (2001) [40]. Similar results were shown in other RS studies which also used trypan blue to assess cell viability after exposure to laser light. Oshima *et al.* (2010) [38] reported that for a 532nm laser with a power of 50mW and an exposure time of 60 seconds, the trypan blue assay was negative. Notingher *et al.* (2002) [41] also reported that for a 785nm laser with a power of 115mW and an exposure time of 120 seconds the assay also returned a negative result.

Fig. 3.1 shows a cell in the lower right of the image which disappears in Figs. 3.2 and 3.3. It is likely that this cell was not properly attached to the surface of the disc, and therefore was easily washed away when the trypan dye was removed. It is unlikely that its disappearance is a result of any laser exposure, as the laser spot size is approximately $1\mu\text{m}$.

3.1.2 Raman Spectra

The averaged Raman spectra for the direct patient derived primary cell cultures, H517-15 cancer (n=164), H517-15 normal (n=161), H554-15 cancer (n=156), H554-15 normal (n=167) and the P4E6 cancer (n=149) and PNT2-C2 normal (n=154) standard cell lines (Figs. 3.4 and 3.5) are consistent with previously published spectra of prostate cancer and normal prostate cells [11] [20] [27] (see also Fig. 1.3), as well as human cancer cells [27] [95], in general, where the spectra have been taken from the cell nuclei. Here n refers to the total number of single spectra obtained for each cell line. Figs. 3.4 and 3.5 show the averaged spectra for each cell line after being linear baselined and normalised to the peak at 1447 cm^{-1} . The normalisation to the 1447 cm^{-1} peak follows Crow *et al.* (2003) [11], Chan *et al.* (2009) [96] and Krafft *et al.* (2012), where the peak at 1447 cm^{-1} was used to normalise prostate

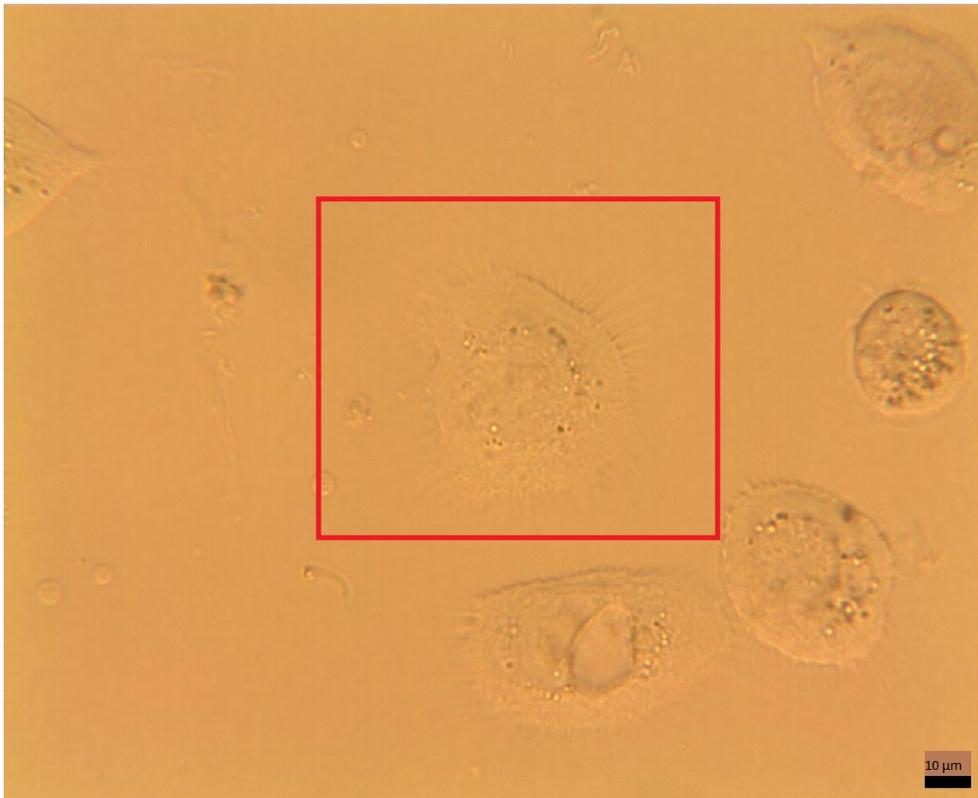


Figure 3.1: Optical image obtained using a 63x dipping lens with a numerical aperture of one showing the P4E6 cell (inside the red square), which was analysed before laser exposure or the application of the trypan blue dye. The cell chosen is of a regular size with no obvious defects or contaminations.

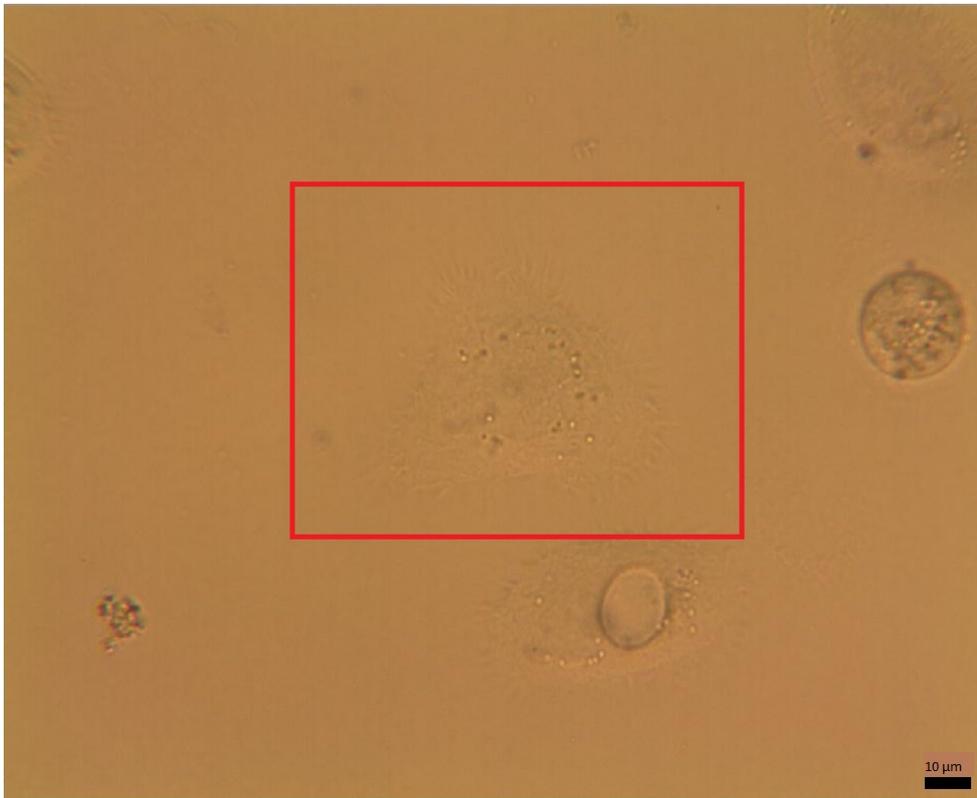


Figure 3.2: Optical image obtained using a 63x dipping lens with a numerical aperture of one of the same P4E6 cell (inside the red square) after laser exposure and the addition of the trypan blue dye. Trypan blue is a dye which is only taken up by cells that have damaged outer membranes. In this case the dye has not been uptaken by the cell indicating that there has been no damage due to the laser exposure. The laser exposure used is the same applied for the Raman measurements in this study, hence verifying non-destructive testing of the cells.

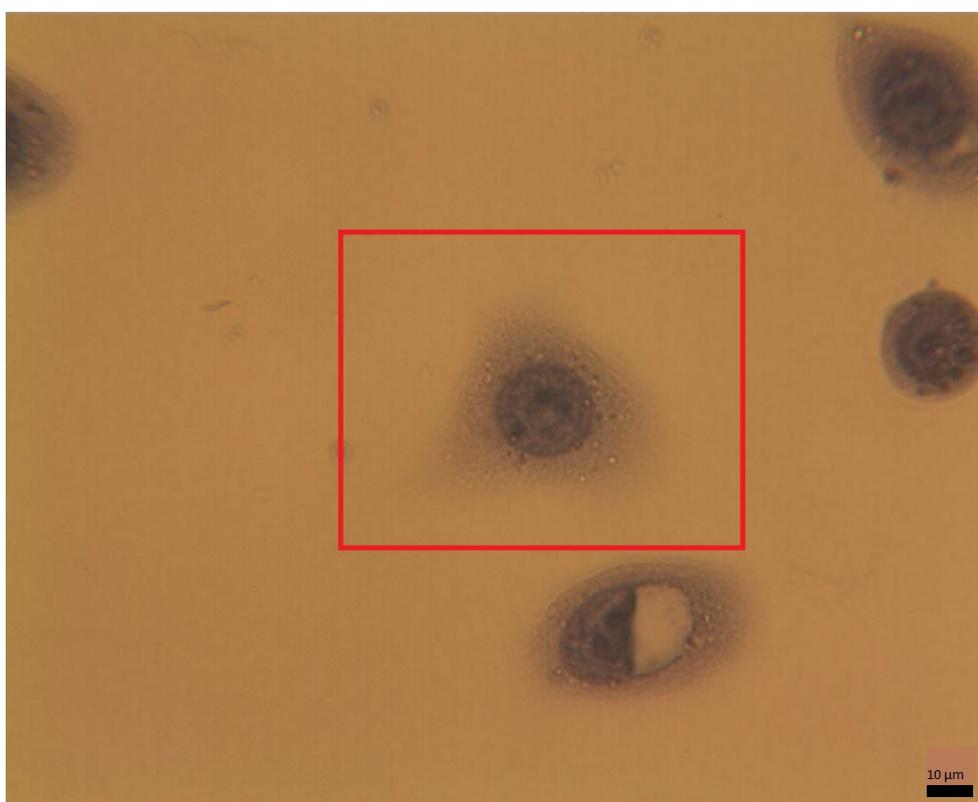


Figure 3.3: To ensure that the trypan blue dye would function correctly in the event of cell damage, the sample was washed with ethanol to kill the P4E6 cells and then more trypan blue dye was added to test for cell damage. From this optical image, taken using a 63x dipping lens with a numerical aperture of one, it was concluded that if the cell had been damaged by the laser, then the trypan blue dye would have been uptaken by the cell therefore producing this result.

3.1 Raman Spectroscopy Studies of Prostate Cancer and Normal Prostate Cells

cancer, leukaemia and brain tumour cell spectra, respectively, prior to analysis. The 1447 cm^{-1} peak is often used for normalisation in Raman cancer studies due to its small variance, making it ideal for observing the relative variation in other peaks.

Figs. 3.6 and 3.7 show the same averaged spectra without normalisation together with the associated standard error of the mean envelopes. The relatively small SE of the mean envelopes displayed in Figs. 3.6 and 3.7, especially in areas of the spectrum where there are prominent peaks, suggest that these spectra are of a consistently high quality.

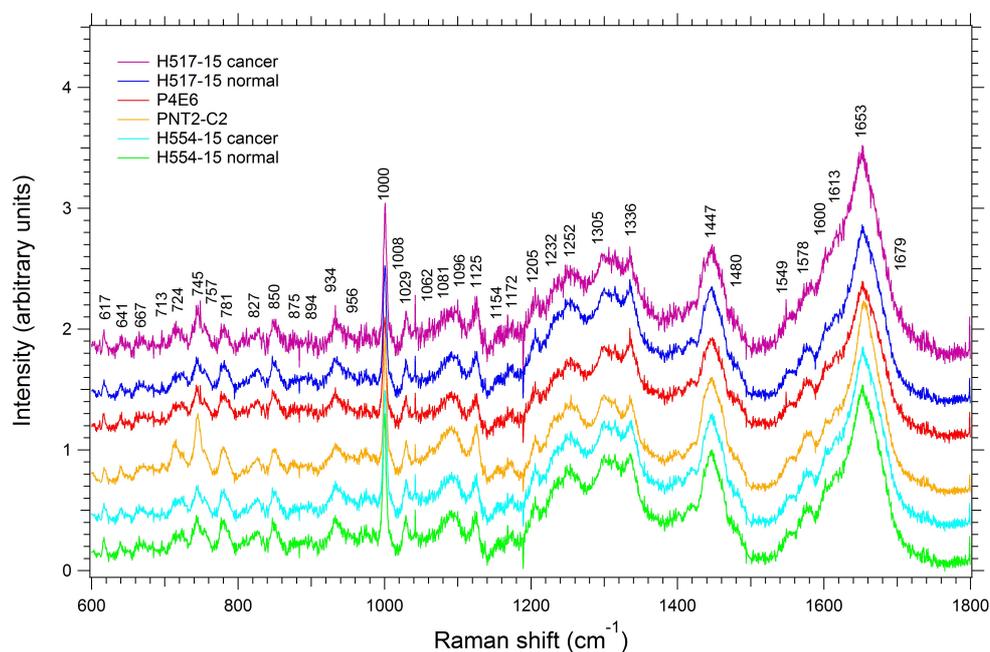


Figure 3.4: Averaged Raman spectra taken in the fingerprint region ($600\text{--}1800\text{ cm}^{-1}$) for the direct patient derived primary cell cultures, H517–15 cancer ($n=164$), H517–15 normal ($n=161$), H554–15 cancer ($n=156$), H554–15 normal ($n=167$), and the P4E6 cancer ($n=149$) and PNT2–C2 normal ($n=154$) standard cell lines. The biological relevance of the key peak assignments identified on the spectra are detailed in Table. 3.1. The six averaged spectra have been linear baselined and normalised to the 1447 cm^{-1} peak, which represents the CH_2 bending mode of proteins and lipids. peak. The spectra are vertically shifted using an arbitrary offset to aid visualisation.

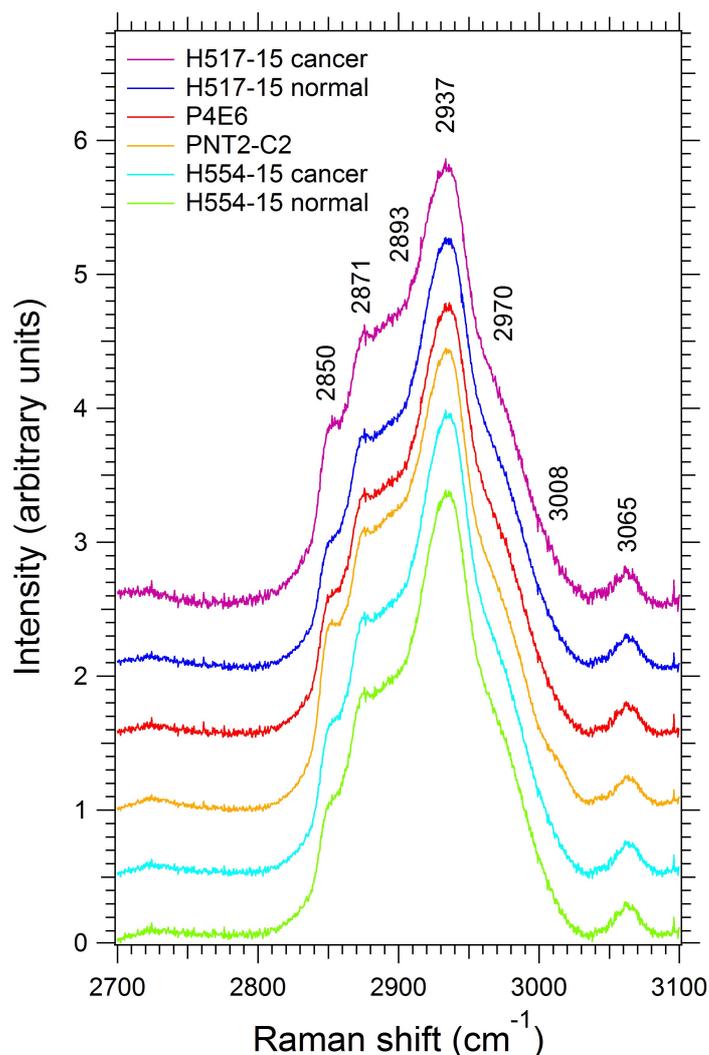


Figure 3.5: Averaged Raman spectra taken in the high wavenumber region ($2700\text{--}3100\text{ cm}^{-1}$) for the direct patient derived primary cell cultures, H517–15 cancer ($n=164$), H517–15 normal ($n=161$), H554–15 cancer ($n=156$), H554–15 normal ($n=167$) and the P4E6 cancer ($n=149$) and PNT2–C2 normal ($n=154$) standard cell lines. The biological relevance of the key peak assignments identified on the spectra are detailed in Table. 3.1. The six average spectra have been linear baselined and normalised to the 1447 cm^{-1} peak which represents the CH_2 bending mode of proteins and lipids. The spectra are vertically shifted using an arbitrary offset to aid visualisation.

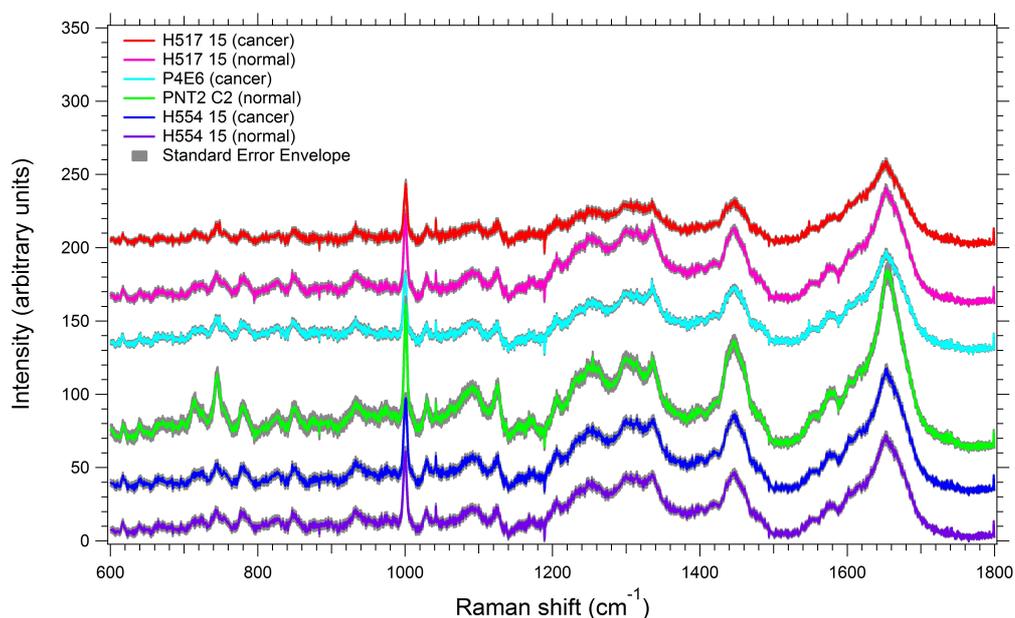


Figure 3.6: Averaged Raman spectra in the fingerprint region ($600\text{--}1800\text{ cm}^{-1}$) with associated standard error envelopes for the direct patient derived primary cell cultures, H517–15 cancer ($n=164$), H517–15 normal ($n=161$), H554–15 cancer ($n=156$), H554–15 normal ($n=167$), and the P4E6 cancer ($n=149$) and PNT2–C2 normal ($n=154$) standard cell lines. The six spectra have been linear baselined and vertically shifted using an arbitrary offset to aid visualisation. These spectra have not been normalised.

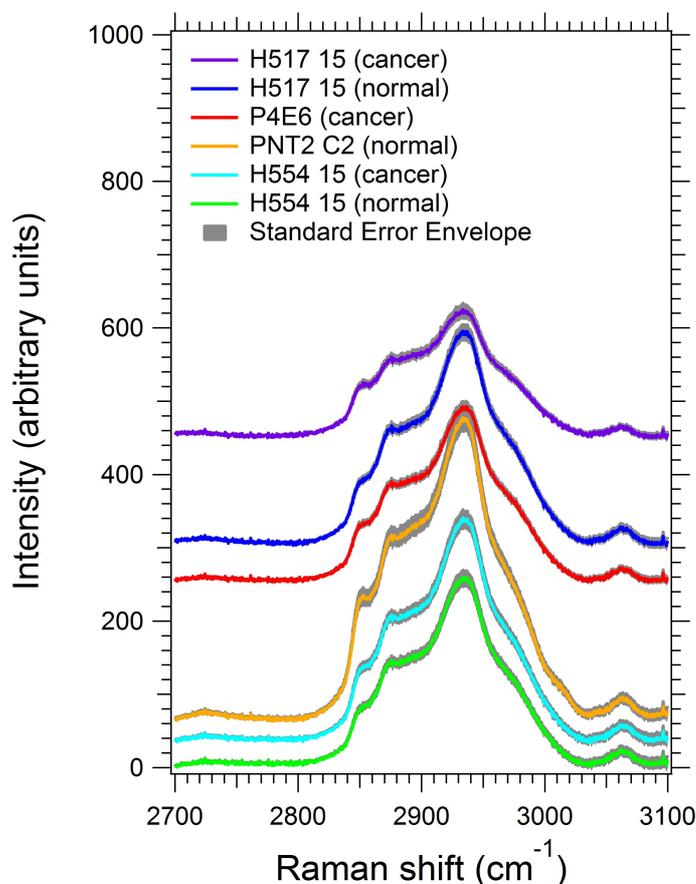


Figure 3.7: Averaged Raman spectra in the high wavenumber region ($2700\text{--}3100\text{ cm}^{-1}$) with associated standard error envelopes for the direct patient derived primary cell cultures, H517–15 cancer ($n=164$), H517–15 normal ($n=161$), H554–15 cancer ($n=156$), H554–15 normal ($n=167$), and the P4E6 cancer ($n=149$) and PNT2–C2 normal ($n=154$) standard cell lines. The six spectra have been linear baselined and vertically shifted using an arbitrary offset to aid visualisation. These spectra have not been normalised.

3.1 Raman Spectroscopy Studies of Prostate Cancer and Normal Prostate Cells

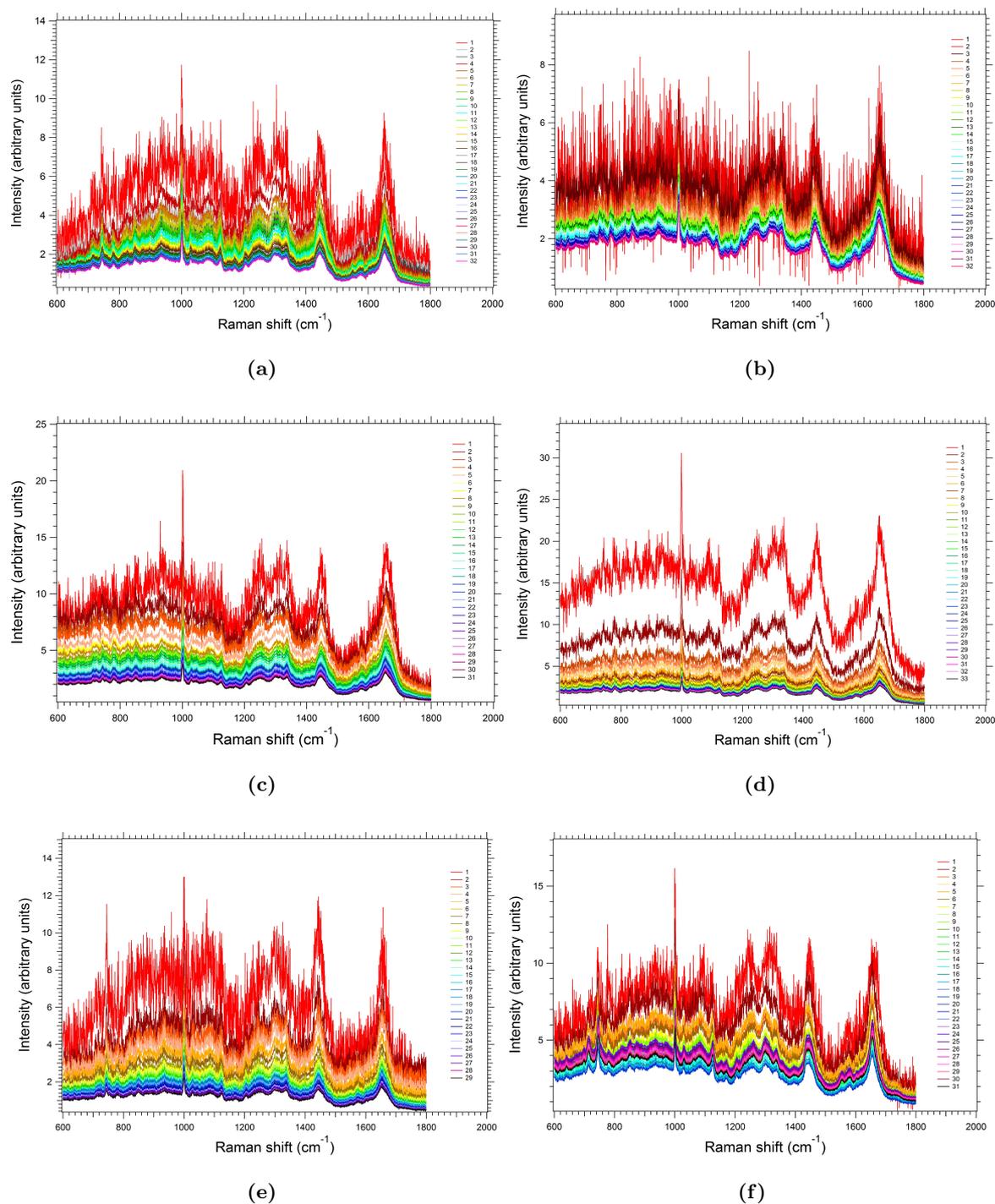


Figure 3.8: The convergence of the standard error of the mean as a function of the number of spectra in the average, as indicated by the legend, for the fingerprint region only showing the (a) H517-15 cancer (n=164), (b) H517-15 normal (n=161), (c) H554-15 cancer (n=156), (d) H554-15 normal (n=167), (e) P4E6 cancer (n=149) and (f) PNT2-C2 normal (n=154) cell lines.

3.1 Raman Spectroscopy Studies of Prostate Cancer and Normal Prostate Cells

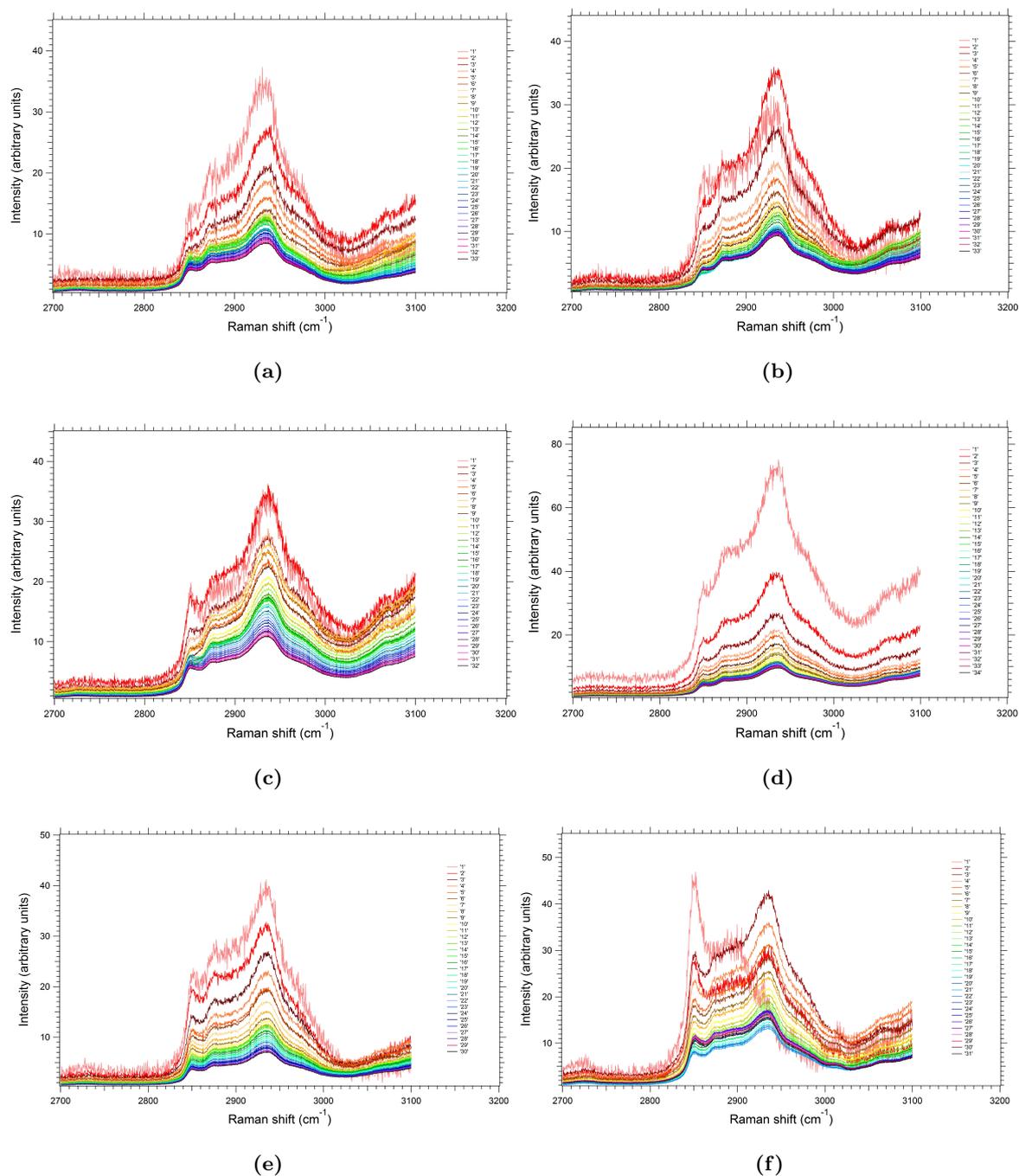


Figure 3.9: The convergence of the standard error of the mean as a function of the number of spectra in the average, as indicated by the legend, for the high-wavenumber region only showing the (a) H517-15 cancer (b) H517-15 normal (c) H554-15 cancer (d) H554-15 normal (e) P4E6 cancer (f) PNT2-C2 normal cell lines.

3.1 Raman Spectroscopy Studies of Prostate Cancer and Normal Prostate Cells

Figs. 3.8 and 3.9 show the convergence of the SE of the mean as a function of the number of spectra in the average, for the fingerprint and high-wavenumber regions of each sample. The spectra shown in these figures are plots of the standard error of the mean at each point on the spectrum, against wavenumber, not raw spectra. The converged SE waves shown provide the error envelopes shown in Figs. 3.6 and 3.7. As more spectra are added to the average, the SE of the mean can be seen to be decreasing, see figs. 3.8 and 3.9, up to the point where the data converges and the standard error is representative of the whole data set. This convergence is important as it shows that the averaged spectra are statistically representative of the whole population for each cell sample. Convergences showing almost identical trends were also obtained for the standard deviations (not shown).

Another way to show statistical convergence is using peak intensity ratios (PIRs) and their associated percentage standard errors. These types of convergence tests are especially important as PIRs will be used to later assess the biomolecular information contained within the spectra. Figs. 3.10 and 3.11 show the convergence of the percentage standard error of the mean for the peak intensity ratios I1447/I1653, where 1447 cm^{-1} is the CH_2 bending mode of proteins and lipids, and 1653 cm^{-1} is the C=C stretch of proteins and lipids, and the I3008/I2850 PIR where 3008 cm^{-1} is the CH stretching of lipids and 2850 cm^{-1} is the CH_2 stretching of lipids (see also Table. 3.1). These ratios were used to test for statistical convergence as they have been previously used in the literature to differentiate normal cells from cancer cells [27] [61]. These figures, together with the the convergence of the SE waves shown in Figs. 3.8 and 3.9, provide confidence in the statistical convergence of each data set, indicating that a sufficient number of individual spectra were collected to faithfully represent the population average of each cell line.

PIR convergences allow a fit to be drawn so that the rate of convergence can be observed. A faster convergence corresponds in this case to a more homogeneous sample population with respect to the PIR studied, i.e. there is less variation amongst the cells. Greater homogeneity is to be expected for the standard cell lines [72] [88]. Figs. 3.10 and 3.11 show that the sample which converged with the smallest number of averaged spectra was the P4E6 cell line, with the PNT2-C2 cell line converging at a similar rate to the primary patient samples. The faster convergence rate indicated that the P4E6 cell line has a more homogeneous distribution of cell characteristics with respect to these PIR markers compared to the other cell lines studied. It was also expected that the PNT2-C2 cell line would be more homogeneous than the primaries due to it being a standard cell line [72], although this was not the case for the particular PIRs studied. These figures show a similar convergence rate to Figs. 3.8 and 3.9, as Figs. 3.8 and 3.9 show the standard error of the means, from which the convergences in Figs. 3.10 and 3.11 were produced.

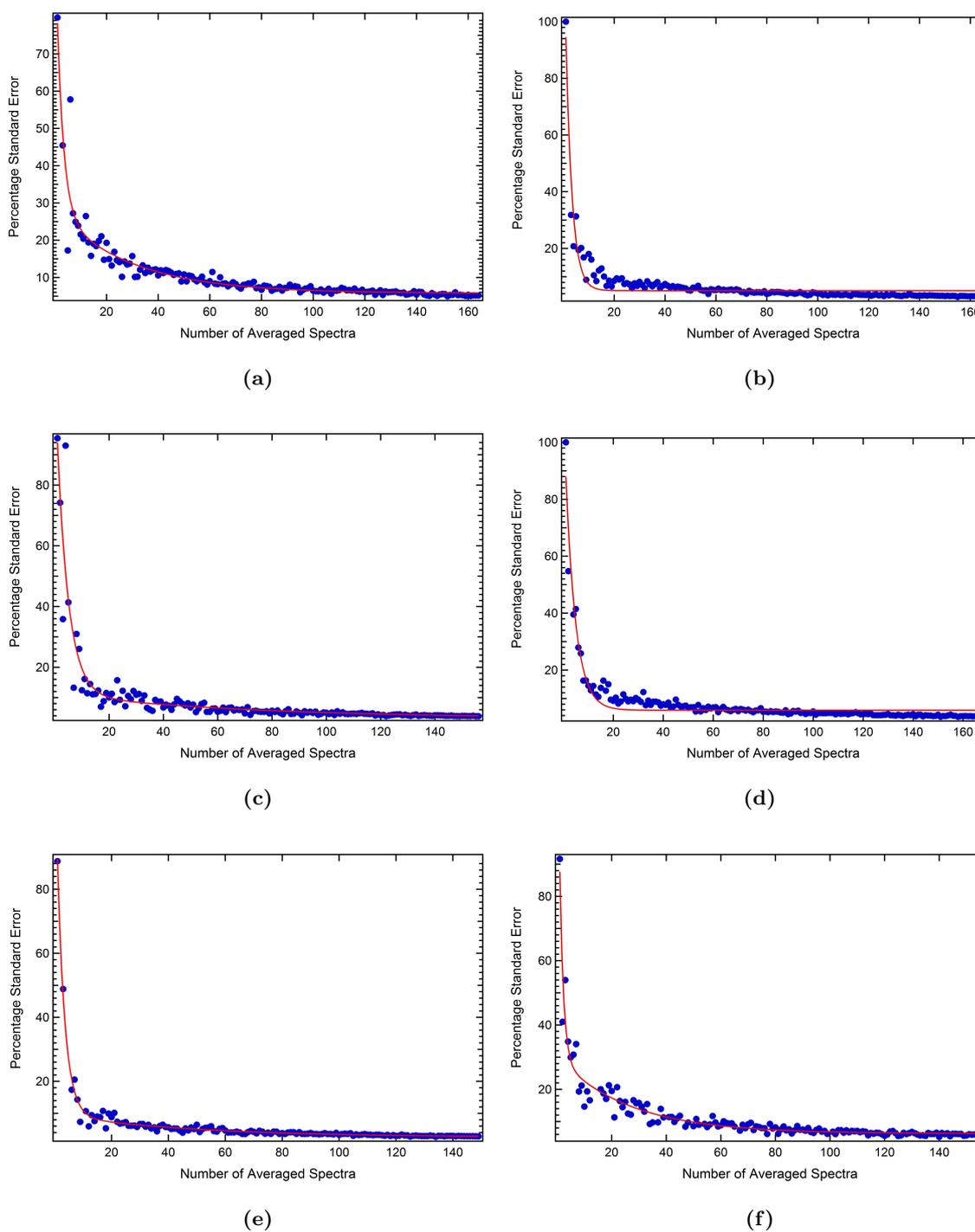


Figure 3.10: The convergence of the percentage standard error of the mean, as a function of the number of spectra in the spectral average for the PIR I1447/I1653 measured randomly across the population of each cell line. The cell lines studied are (a) H517-15 cancer, (b) H517-15 normal, (c) H554-15 cancer, (d) H554-15 normal, (e) P4E6 cancer and (f) PNT2-C2 normal.

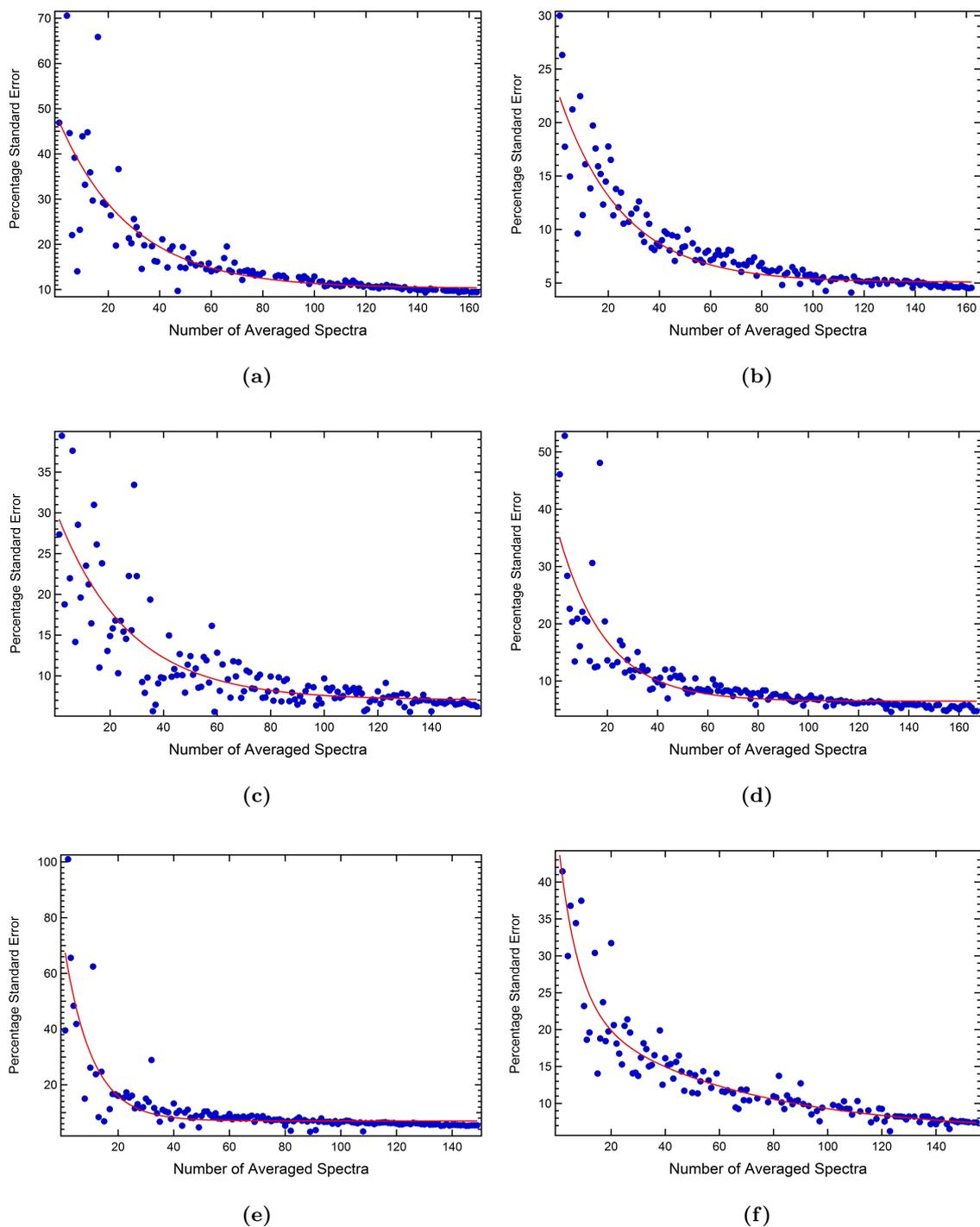


Figure 3.11: The convergence of the percentage standard error of the mean as a function of the number of spectra in the spectral average for the PIR I3008/I2850 measured randomly across the population of each cell line. The cell lines studied are (a) H517-15 cancer, (b) H517-15 normal, (c) H554-15 cancer, (d) H554-15 normal, (e) P4E6 cancer and (f) PNT2-C2 normal.

3.1 Raman Spectroscopy Studies of Prostate Cancer and Normal Prostate Cells

The peak assignments that have been identified by peak fitting and shown in Figs. 3.4 and 3.5 are also consistent with the previously published literature on Raman spectroscopy studies of biological samples [10] [11] [20] [27] (see also Table 3.1). All peaks positions obtained by fitting were averaged across the six spectra, returning an uncertainty of approximately 1 cm^{-1} in the peak position, which is also the same as the spectral uncertainty from the Raman instrument error.

Peak Position $\pm 1 \text{ (cm}^{-1}\text{)}$	Assignments
617	Proteins (CC twisting) [97]
641	Proteins (CC twisting) [97]
667	DNA/RNA [25]
713	Amino acids/methionine [10] [98]
724	DNA/RNA [99]
745	DNA/RNA [97]
757	Proteins, tryptophan (symmetric breathing) [66]
781	DNA/RNA [99]
827	DNA and proteins (proline, hydroxyproline, tyrosine) [100]
850	Proteins (single bond stretching of amino acids) [101]
875	Proteins and lipids (CC stretching) [102] [103]
894	Deoxyribose [104]
934	Proteins (CC, collagen) [102]
956	Cholesterol [66]
1000	Phenylalanine [105]
1008	Phenylalanine [106]
1029	Glycogen, collagen (OCH ₃ , CC) [100] [107]
1062	DNA/RNA and lipids (CC stretch) [25]
1081	DNA/RNA, phospholipids (phosphate vibration) and lipids (CC skeletal) [105] [100] [108]

3.1 Raman Spectroscopy Studies of Prostate Cancer and Normal Prostate Cells

1096	DNA/RNA, and lipids [105] [109]
1125	Proteins (CN stretching) and lipids (CC skeletal) [97] [100]
1154	Proteins (glycogen) [107]
1172	Proteins (CH, tyrosine) [32]
1205	Proteins, collagen [107]
1232	Proteins (Amide III) [107]
1252	Guanine, cytosine (NH ₂) and lipids [104] [107]
1305	DNA/RNA, proteins (CH ₂ , CH ₃) and lipids (CH ₂) [66] [102]
1336	DNA/RNA and proteins (CH ₂ , CH ₃) [32]
1447	CH ₂ bending mode of proteins and lipids [110]
1480	Proteins (Amide ii) [110]
1549	Proteins (Tryptophan) [66]
1578	Guanine, adenine [25]
1600	Proteins (Amide I, C=O stretching, phenylalanine) [107]
1613	Proteins (Tyrosine, tryptophan C=C) [97]
1653	Proteins and lipid (C=C stretch) [66]
1679	Proteins (Amide I) [111]
2850	CH ₂ stretching of lipids [112]
2871	CH ₂ stretching and CH stretch of lipids and proteins [112]
2893	CH ₂ stretching of lipids and proteins, CH stretching of lipids and proteins [112]
2937	Lipids (CH ₂ , CH ₃) [112]
2970	lipids and cholesterol (CH ₃) [103]
3008	Lipids (unsaturated, CH stretching) [103]
3065	CH stretching of Lipids [113]

Table 3.1: All peak positions identified in these experiments and their published assignments along with examples of their uses in the literature. All peak assignments correspond to Movasaghi *et al.* (2007). The peak positions in this study were identified by fitting and then averaging across the 6 samples, this returned an associated error in the peak locations of 1 cm^{-1} .

3.1.3 Principle Component Analysis (PCA) and Linear Discriminant Analysis (LDA)

Before any multivariate analyses were performed, every spectrum was first linear baselined and normalised to the area under the curve using the Raman toolkit package [92] then scaled so that the mean of each data point was zero, and the variance was one, using the statistical package, R v3.3.0.

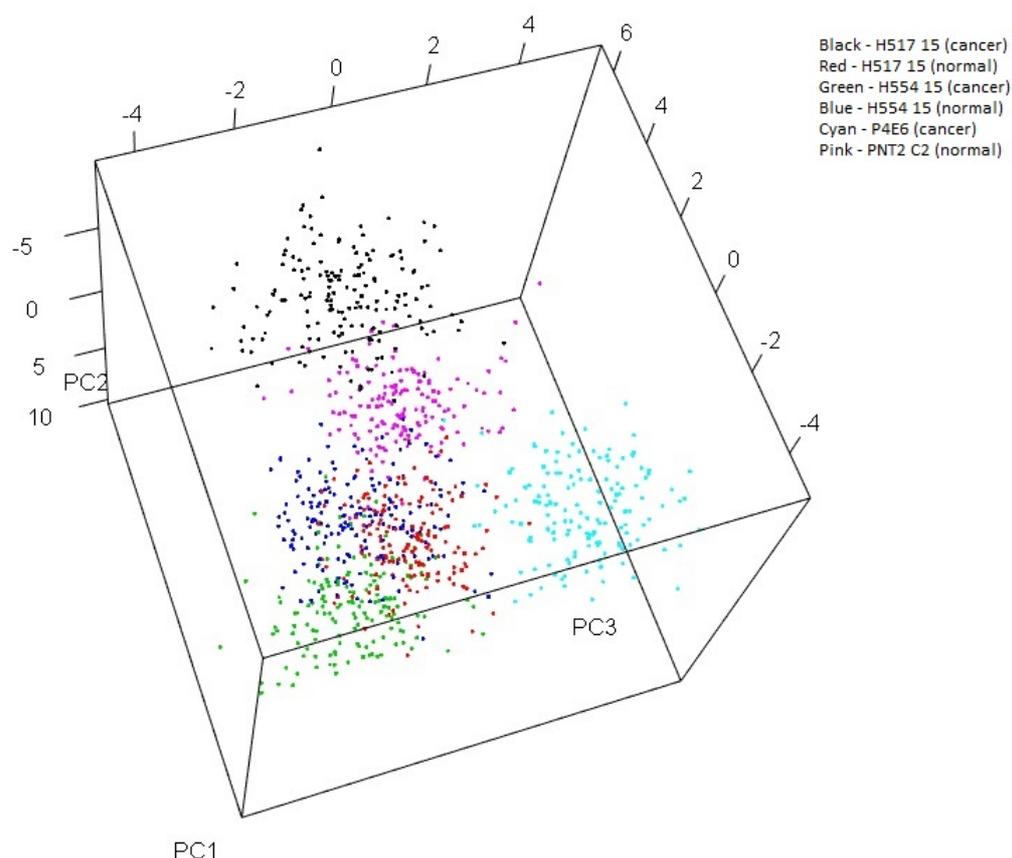


Figure 3.12: Results of PCA followed by LDA analysis for the direct patient derived primary cell cultures, H517–15 cancer (n=164), H517–15 normal (n=161), H554–15 cancer (n=156), H554–15 normal (n=167), and the P4E6 cancer (n=149) and PNT2–C2 normal (n=154) standard cell lines. Principle components one, two and three account for 43.4%, 19.6% and 4.9% of the total variance, respectively.

Fig. 3.12 shows the results of principal component analysis (PCA) followed by linear discriminant analysis (LDA) performed on the direct patient derived primary cell cultures, H517–15 cancer (n=164), H517–15 normal (n=161), H554–15 cancer (n=156), H554–15 normal (n=167) and the P4E6 cancer (n=149) and PNT2–C2 normal (n=154) standard cell lines. The LDA analysis was performed on the first 80 principle components as per the Kaiser criterion, whereby only PCs with variance greater than one in the standardised data are included [93]. Fig. 3.12 shows clear separation between most of the samples with the exception of the two normal patient–derived primary cell culture samples (H517–15 and

H554–15), which, as they are both samples from normal prostate tissue, are expected to be biologically similar [11]. Fig. 3.12 also shows only a small separation between the H554–15 patient cancer sample and the two normal primary culture cell samples. This lack of obvious separation indicates that the H554–15 patient cancer sample is more spectrally, therefore biologically, similar to the normal samples.

Both standard cell lines (P4E6 cancer and PNT2–C2 normal) are also separate from the primary cell culture samples (Fig. 3.12). This separation between direct patient derived normal cells and the normal standard cell line (PNT2–C2) casts doubt on the use of standard cell lines in research and their effectiveness in treatment development, as suggested in the literature [74] [75]. It has been suggested that some standard cell lines do not provide accurate representations of patient biology due to differences in the way immortalised cells react to hormones found in the body [114]. Capes–Davies *et al.* (2010) suggested that as many as 29% of certain standard cell lines may be contaminated by a foreign cell line or micro-organism [115]. Crow *et al.* (2003), using primary patient tissue samples, and Crow *et al.* (2005), using standard cell line samples, came to different conclusions about the change in DNA content between normal and cancer cells, using RS [11] [20]. It was reasonable to expect the P4E6 cancer cell line to be separate from the other cancer cell lines due to the difference in Gleason grade. This separation affirms the ability of RS to distinguish not only cancer cells from normal cells, but also its ability to separate cancers of different severity, as suggested by Crow *et al.* (2003) [11]. It should be noted however that this separation could also be attributed to the P4E6 cell line being an immortalised cell line, and not to differences due to malignancy.

The LDA process was verified using the leave-one-out cross validation method [11], whereby each spectrum is left out of the initial LDA in turn then grouped by the linear discriminant functions. This process returned an error of 14%, with the groupings shown in Table 3.2. A 0% error would show all the spectra in their correct corresponding groups. Although an error of 14% is seemingly high, it is apparent from the groupings that the majority of this uncertainty has arisen from the identification of the two normal patient derived primary cell culture samples, which are expected to be biologically similar [11], as well as the H554–15 cancer sample, which appears similar to the two normal samples.

	Group 1	Group 2	Group 3	Group 4	Group 5	Group 6
H517-15 (cancer)	157	0	1	3	2	1
H517-15 (normal)	0	116	17	26	2	0
H554-15 (cancer)	0	18	128	9	1	0
H554-15 (normal)	0	30	10	123	4	0
P4E6 (cancer)	0	7	0	4	138	0
PNT2-C2 (normal)	0	0	0	0	0	154

Table 3.2: The results of the leave-one-out cross validation, performed on the LDA result shown in Fig. 3.12. The total error in the LDA, calculated from the number of spectra that were grouped incorrectly, is 0.14196 (14%). The H517-15 normal, H554-15 normal and H554-15 cancer samples are spread across their relevant groups showing that the LDA was not able to tell these spectra apart as well as the other samples.

The primary H554-15 cancer sample is not fully separated from its normal-match using PCA. Fig 3.13 shows the result of PCA using the first three principle components accounting for 55.7%, 6.8% and 4.7% of the total variance, respectively. Although the two samples are not fully separate there is a small amount of separation in PC3. This result is corroborated by the PCA followed by LDA analysis shown in Fig. 3.12, as well as the groupings for the leave-one-out cross validation of the LDA (Table 3.2). Although Fig. 3.12 shows a small difference, there is little more separation between H554-15 cancer and

normal than between the two primary normal samples. Table 3.2 is especially telling as the H554–15 cancer sample has spectra grouped in both the H554–15 and H517–15 normal groups. This result could indicate that this sample is not in fact as high grade cancer (i.e. Gleason 7) as the first histopathology informed. We are now consulting with the hospital, which provided these samples, to ascertain this. If correct, this would then demonstrate Raman spectroscopy's potential to discriminate not only cancer cells from normal, as in much of the literature [95] but also to assess patient specific information, similar to the study by Crow *et al.* (2003) [11] where samples were separated by Gleason grade using PCA.

The loadings for the principle components, which show the areas of greatest variance, for the primary patient H554–15 cancer and normal cells (Fig. 3.14) show that what variance there is, primarily from PC1, comes from peaks corresponding to DNA/RNA (745, 781 and 1578 cm^{-1}), proteins (850 and 1125 cm^{-1}), phenylalanine (1000 cm^{-1}) and glycogen (934 and 1029 cm^{-1}). However, large amounts of the variance in PC1 can also be attributed to fluctuations in background between spectra, as shown by regions of large negative variance between 450–700 cm^{-1} and large positive variance from around 1200–1700 cm^{-1} . However as none of these principle components produce any separation between the two samples using PCA (Fig. 3.13) this variance can only be attributed to variations across both groups from the individual spectra in the entire H554–15 sample. LDA was not used here as in order to achieve any separation, over 120 PCs had to be included, this violates the Kaiser criterion.

The cell lines P4E6 (cancer) and PNT2–C2 (normal) are separated using principle components one, two and four accounting for 36.5, 13.3 and 5.4% of the total variance, respectively (Fig. 3.15). PC3 was replaced with PC4 in this case as no separation was seen using PCs one, two and three. From this figure we can see that the two samples are separated only by PC4. PC1 and PC2, while representing a greater share of the variance, only account for variance in the combined population of both samples, and not differences between the two cell lines. LDA was not used here as separation was achieved using only PCA.

Fig. 3.16 shows the corresponding loadings for the principle components in Fig. 3.15. The loading for the principle component, which separates the P4E6 cancer and PNT2–C2 normal standard cell lines, i.e. PC4, (Fig. 3.16) shows that the variance which separates the cancer from the normal sample comes from peaks corresponding to DNA/RNA (724, 745 and 781), proteins and lipids (850, 1096, 1125,

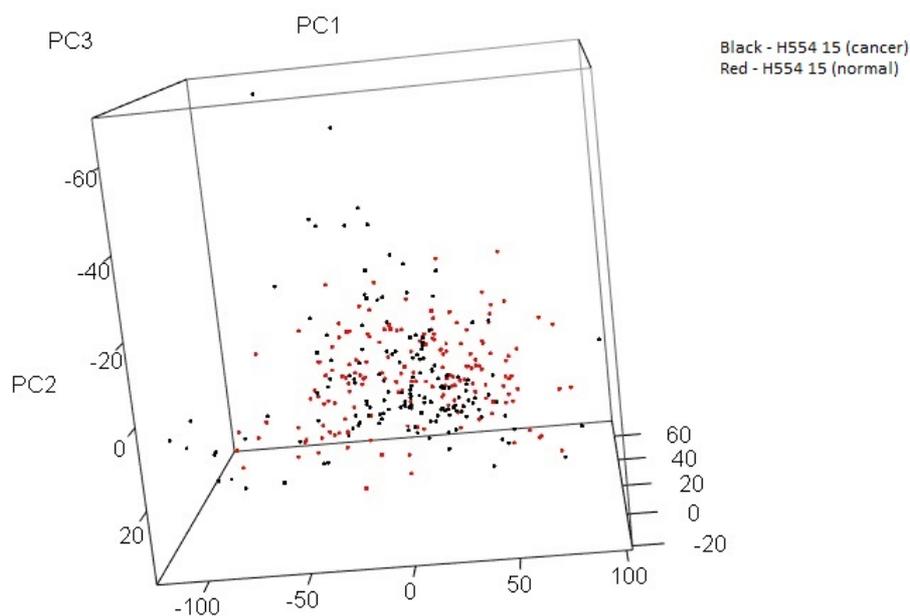
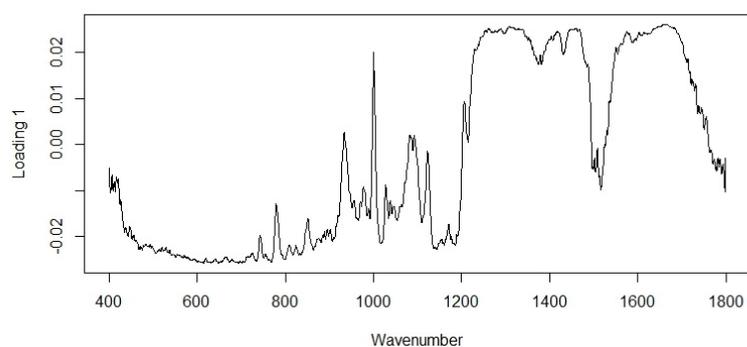
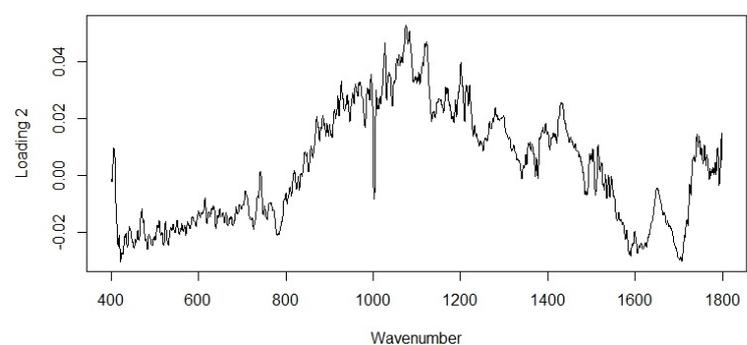


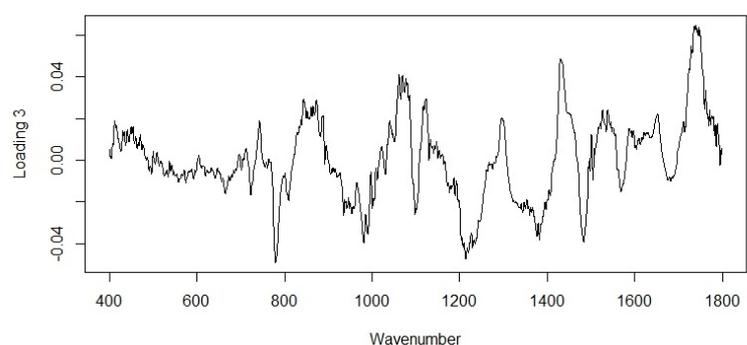
Figure 3.13: PCA results for the direct patient derived primary cell cultures, H554–15 cancer (n=156), H554–15 normal (n=167) using the first three principle components, accounting for 55.7%, 6.8% and 4.7% of the total variance, respectively. No separation is seen here, indicating that the two samples produce very similar spectra and any variance in the data cannot be attributed to differences between the normal and cancer cells, only to the heterogeneity of the population as a whole.



(a)



(b)



(c)

Figure 3.14: Loadings of the first three PCs for the H554–15 samples. These loadings suggest that areas of high variance include 745 cm^{-1} , 781 cm^{-1} , 850 cm^{-1} , 934 cm^{-1} , 1000 cm^{-1} , 1029 cm^{-1} , 1125 cm^{-1} and 1578 cm^{-1} for PC1, 1000 cm^{-1} and 1653 cm^{-1} for PC2 and 781 cm^{-1} , 827 cm^{-1} , 850 cm^{-1} , 1000 cm^{-1} , 1029 cm^{-1} , 1081 cm^{-1} , 1125 cm^{-1} , 1447 cm^{-1} and 1578 cm^{-1} for PC3. These markers however are not enough to discriminate between the 2 samples, as shown in Fig. 3.13.

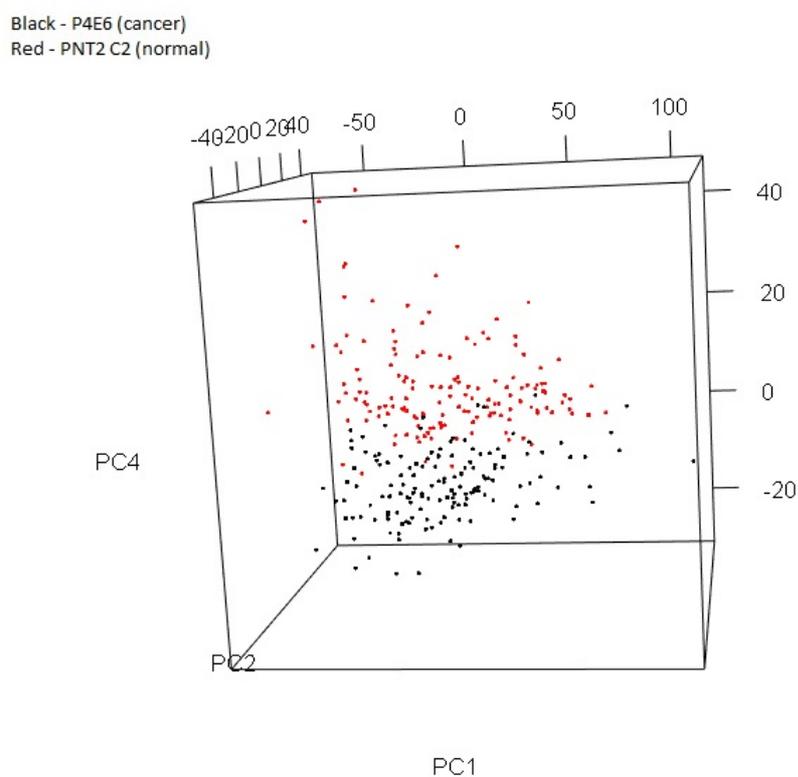


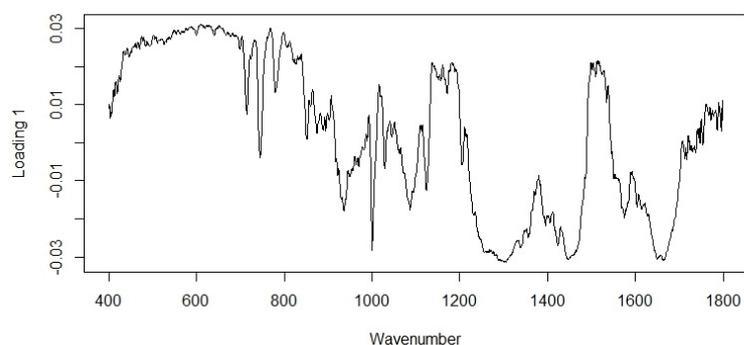
Figure 3.15: PCA results for the P4E6 cancer (n=149) and PNT2-C2 normal (n=154) standard cell lines showing principle components one, two and four accounting for 36.5, 13.3 and 5.4% of the total variance, respectively. The separation seen here can be attributed entirely to PC4, with PC1 and PC2 accounting only for the variance within the two samples combined. PC3 was omitted here and PC4 included as PC3 contributed no separation between normal and cancer.

3.1 Raman Spectroscopy Studies of Prostate Cancer and Normal Prostate Cells

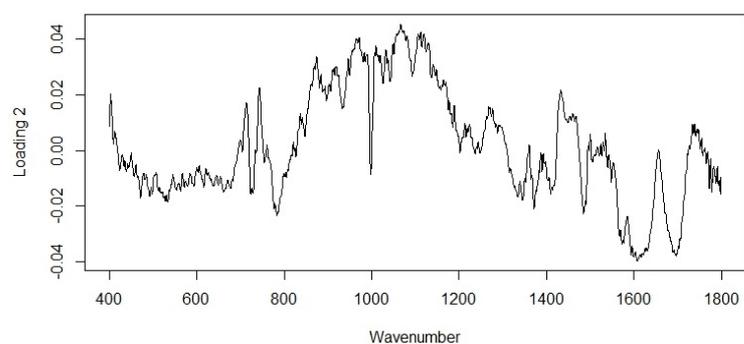
1205, 1600, 1613 and 1653 cm^{-1}), phenylalanine (1000 cm^{-1}) and glycogen (1029 cm^{-1}). All of the aforementioned peaks have been highlighted in the RS literature as possible markers for cancer, for example, Crow *et al.* (2005) produced loadings with similar areas of high variance when using PCA to separate normal prostate from prostate cancer samples [20]. The biomolecular assignments of these peaks also correspond to biomarkers suggested in the literature as markers for cancer, for example phenylalanine, as suggested by Huang *et al.* (2003) [32] and lipids, as suggested by Rysman *et al.* (2010) [62] and Calvisi *et al.* (2011) [63].

PNT2-C2 (normal) is separated partly from the other, primary, normal samples most likely due to the immortalisation process altering the biology of the cells. Fig. 3.17 shows this separation using the first three principle components accounting for 50.6, 8.8 and 4.8% of the total variance, for PC1, PC2 and PC3, respectively, where the separation arises almost entirely from PC3. Similarly, Fig. 3.18, shows that the P4E6 (cancer) sample is separate from the H517-15 cancer and slightly from the H554-15 cancer sample using principle components one, two and three, accounting for 43.1, 21.5 and 2.9% of the total variance, respectively. The separation between the H517-15 cancer and the H554-15 and P4E6 cancer samples is attributed to PC2 and the small amount of separation between the H554-15 and P4E6 cancer samples can be attributed to PCs one and two. This separation is to be expected due to the difference in gleason grade, however, the P4E6 and H554-15 cancer samples appearing to be similar indicates that the H554-15 cancer may have more in common with the lower grade P4E6 cell line than with the H517-15, Gleason seven, primary cell line.

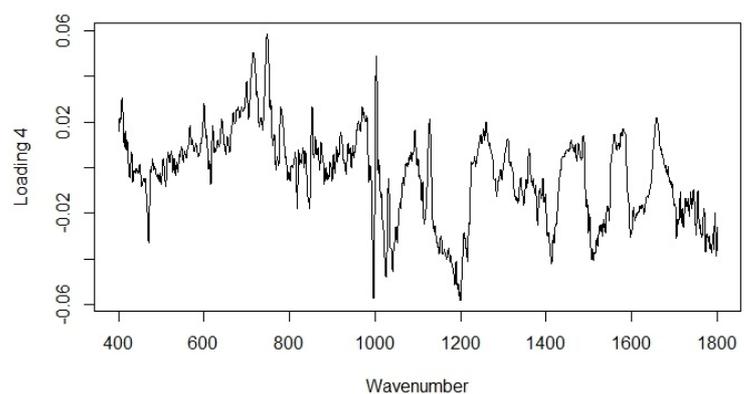
The differences seen here between the primary samples and representative standard cell lines (specifically for the normal samples) highlights one of the key aims of this research, to determine the accuracy of RS when used on immortalised cell lines and to assess how closely these cell lines represent the biology of actual patients. This is studied as previous work in other fields has cast doubt over the reliability of standard cell lines with several reports in the literature of mis-identification and cross-contamination [74] [75]. LDA was again not used here as increasing the separation shown in Figs. 3.17 and 3.18 required the inclusion of upwards of 100 PCs, in violation of Kaisers criterion.



(a)



(b)



(c)

Figure 3.16: Loadings for principle components 1, 2 and 4 for the standard cell lines samples, P4E6 and PNT2–C2. Only PC4 accounts for variation between the normal and cancer samples, as shown in Fig. 3.15. The peaks which correspond to high variance in PC4 are 724 cm^{-1} , 745 cm^{-1} , 781 cm^{-1} , 850 cm^{-1} , 1000 cm^{-1} , 1029 cm^{-1} , 1096 cm^{-1} , 1125 cm^{-1} , 1205 cm^{-1} , 1600 cm^{-1} , 1613 cm^{-1} and 1653 cm^{-1} .

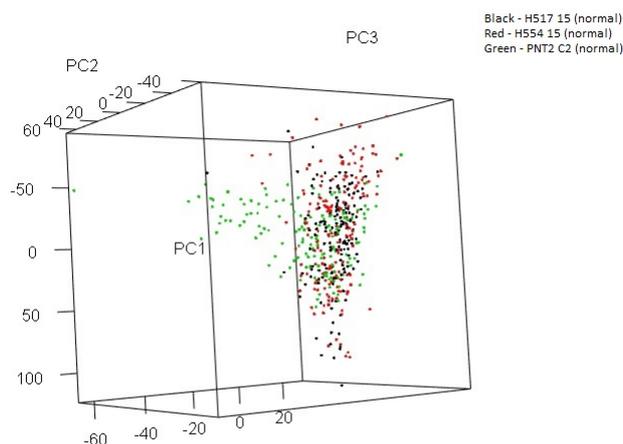


Figure 3.17: PCA results for the direct patient derived primary cell cultures, H517–15 normal (n=161), H554–15 normal (n=167) and the PNT2–C2 normal (n=154) standard cell line.

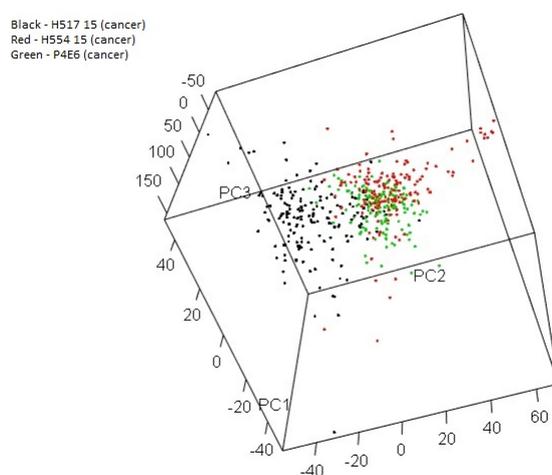


Figure 3.18: PCA results for the direct patient derived primary cell cultures, H517–15 cancer (n=164), H554–15 cancer (n=156) and the P4E6 cancer (n=149) standard cell line.

The PCA results in Fig. 3.19 shows that the H517–15 cancer and normal samples are fully separated using only the first three principle components, PC1, PC2 and PC3, accounting for 44.6, 23.1 and 4.1% of the total variance, respectively. Fig. 3.20 shows the loadings for the principle components used to separate the H517–15 cancer and normal samples with the first, second and third principle component

results. Here, PC2 separates the normal from cancer cells as seen in Fig. 3.19

The only PC loading which separated the cancer and normal cells for the H517–15 sample was PC2, as shown in Fig. 3.19. (Fig. 3.20) shows that the majority of the variance in PC2 comes from peaks corresponding to amino acids (713 cm^{-1} , DNA/RNA (745 , 781 and 1578 cm^{-1}), proteins and lipids (1096 cm^{-1} , 1125 cm^{-1} , 1205 cm^{-1} , 1600 cm^{-1} , 1613 cm^{-1} and 1653 cm^{-1}), Phenylalanine (1000 cm^{-1}) and glycogen (934 and 1029 cm^{-1}). These markers are corroborated in the literature by Crow *et al.* (2005) [20] (DNA/RNA), Huang *et al.* (2003) [32] (phenylalanine), Rysman *et al.* (2010) [62] (lipids) and Calvisi *et al.* (2011) [63] (lipids).

While the ability of RS to distinguish normal from cancer cells is not novel [11] [20] [27] the PCA/LDA results here can also be used to inform a more targeted analysis using peak intensity ratios [28]. By taking only the principle components which account for separation between normal and cancer samples and singling out the loadings which show the greatest variance, the peak intensity ratio biomarkers which correspond to malignancy can be identified. The loadings corresponding to the PCs which separate the cancer cells from the normal cells, for each cell line, serve to confirm the results of PIR analysis by showing the areas or individual peaks in the spectrum where there is variance between the two groups. This should match the peak intensity ratios which also show a separation.

3.1.4 Peak Intensity Ratio (PIR) Comparisons

Peak intensity ratios (PIRs) allow for a relative comparison of individual peak intensities, which in turn allows the biomolecular information associated with PIR changes to be extracted from the Raman spectra. Some ratios, such as I1653/I1447, have been attributed in the literature to physical characteristics of biological samples. This marker has been shown to differentiate cancer from treated cancer cells previously by Potcoava *et al.* [27] who found a decrease in this ratio upon treatment for all cell lines. In this case (I1653/I1447) is thought to represent the degree of mass unsaturation and the chain length of lipids in a sample [26]. Lipid changes are well documented in cancer studies, making PIRs which represent lipids an ideal candidate for discriminating cancer cell from normal [62] [63].

The mass unsaturation ratio (I1653/I1447), where 1447 cm^{-1} is the CH_2 bending mode of proteins and lipids and 1653 cm^{-1} is the $\text{C}=\text{C}$ stretch of proteins and lipids can vary with both the degree of

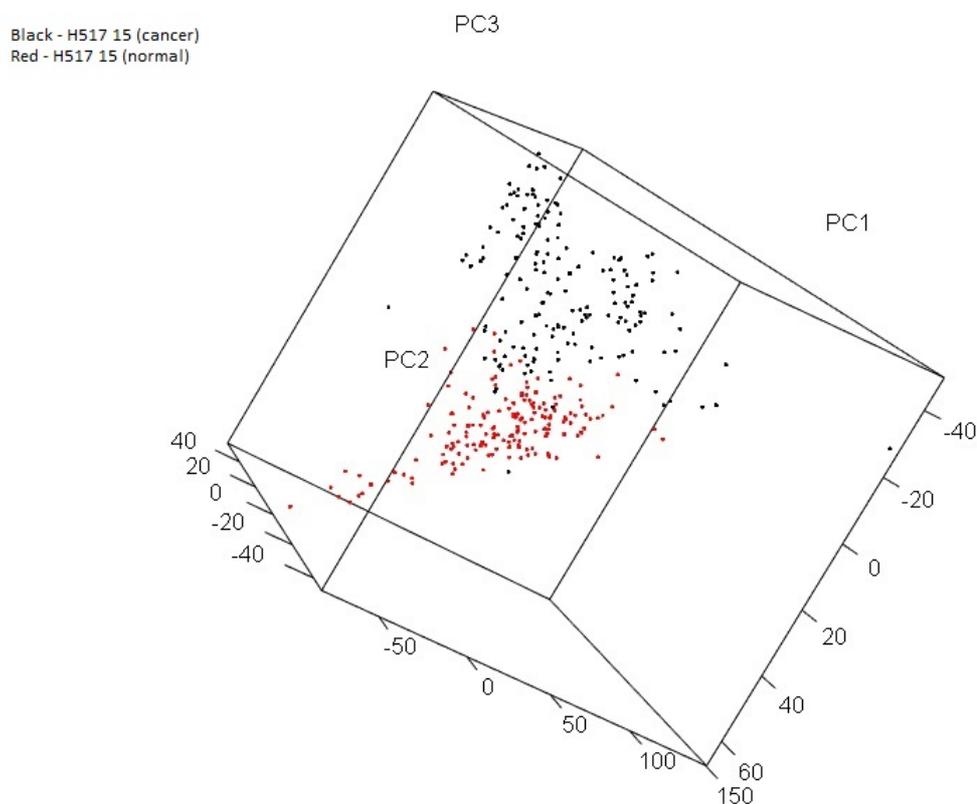
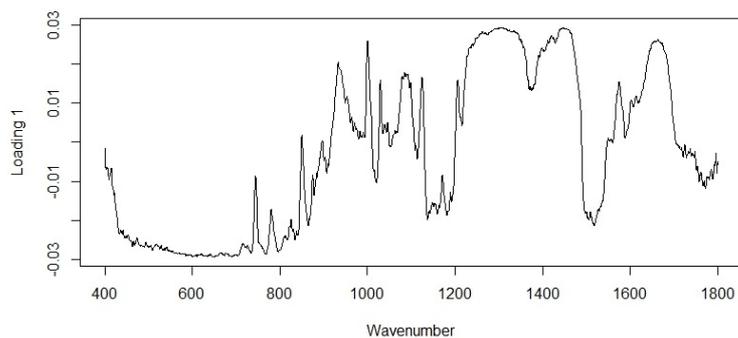
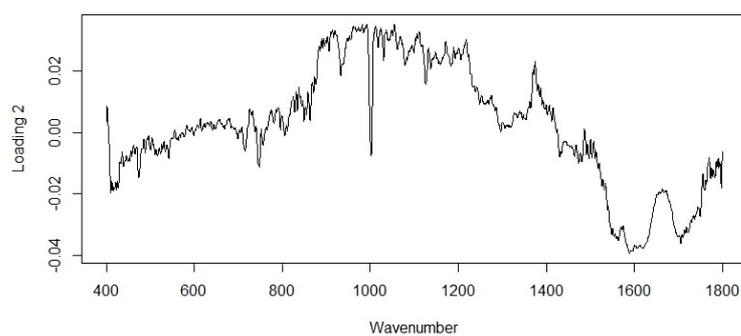


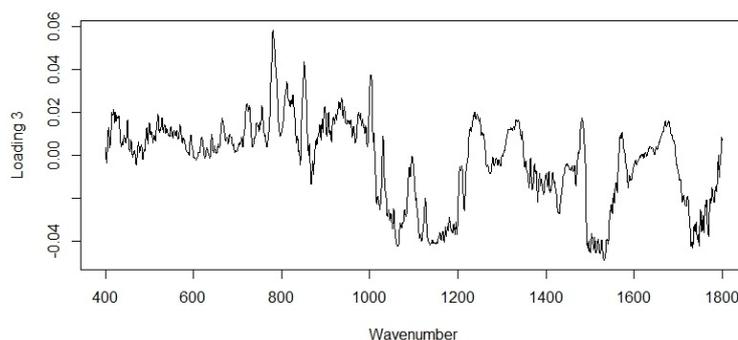
Figure 3.19: PCA results for the direct patient derived primary cell cultures, H517–15 cancer (n=164), H517–15 normal (n=161) showing the first three principle components, accounting for 44.6, 23.1 and 4.1% of the total variance, respectively. The separation seen here can be attributed almost entirely to PC2, with a small contribution from 3. PC1 makes no contribution to the separation, indicating that PC1 accounts only for the heterogeneity of the H517–15 population as a whole.



(a)



(b)



(c)

Figure 3.20: Loadings of the first three principle components (PCs) for the H517–15 samples. Only PC2 accounts for variation between the normal and cancer samples, as shown in Fig. 3.19. The peaks which correspond to high variance in PC2 are 713 cm^{-1} , 745 cm^{-1} , 781 cm^{-1} , 934 cm^{-1} , 1000 cm^{-1} , 1029 cm^{-1} , 1096 cm^{-1} , 1125 cm^{-1} , 1205 cm^{-1} , 1578 cm^{-1} , 1600 cm^{-1} , 1613 cm^{-1} and 1653 cm^{-1} .

3.1 Raman Spectroscopy Studies of Prostate Cancer and Normal Prostate Cells

mass unsaturation and the chain length of lipids in the sample [26]. Table 3.3 shows that this marker was only able to fully discriminate (i.e. outside of the SE uncertainty) normal cells from cancer cells for the H517–15 sample. For the H517–15 and standard cell line samples there is variance seen in PC2 and PC4, respectively, for 1653 cm^{-1} (Figs. 3.16 and 3.20), as these are the PCs which differentiate the normal from the cancer cells for each of these samples (Figs. 3.15 and 3.19) it is expected that the I1653/I1447 ratio may also provide some degree of separation. As the H554–15 sample shows very little separation under PCA, it was expected that the PIRs would reflect this and show little to no separation, as is the case here, shown in Table 3.3.

The results drawn from Table 3.3 are supported by the p-values from a T-test performed on the data. The p-values are shown in Table 3.4.

It should be noted that the exact biomolecular interpretation of this ratio is questionable in terms of being lipid specific, due to the peaks at 1447 cm^{-1} and 1653 cm^{-1} being convolved with other biomolecules containing CH_2 and double carbon bonds, such as amino acids [10].

	I1653/I1447 (unsaturation degree)	Standard Error	I3008/I2850 (TUFA/TFA)	Standard Error
H554–15 (normal)	1.98 (1.91–2.05)	0.07	0.35 (0.33–0.37)	0.02
H554–15 (cancer)	2.04 (1.97–2.11)	0.07	0.33 (0.31–0.35)	0.02
H517–15 (normal)	2.01 (1.94–2.07)	0.06	0.39 (0.37–0.41)	0.02
H517–15 (cancer)	2.5 (2.4–2.6)	0.1	0.24 (0.21–0.27)	0.03
PNT2–C2 (normal)	1.98 (1.89–2.07)	0.09	0.38 (0.36–0.40)	0.02
P4E6 (cancer)	2.07 (2.02–2.12)	0.05	0.29 (0.27–0.31)	0.02

Table 3.3: Table showing the mass unsaturation degree ratio (I1653/I1447) and TUFA/TFA ratio (I2850/I3008) for all samples investigated.

3.1 Raman Spectroscopy Studies of Prostate Cancer and Normal Prostate Cells

Cell type	P-value for I1653/I1447 (unsaturation degree)	P-value for I3008/I2850 (TUFA/TFA)
H554-15	0.5	0.5
H517-15	0.0001	0.0001
Standard cell lines	0.4	0.002

Table 3.4: Table showing the P-values obtained from a T-test when comparing the ratios (I1653/I1447) and (I2850/I3008) between the normal and cancer samples for all cell types investigated.

3.1 Raman Spectroscopy Studies of Prostate Cancer and Normal Prostate Cells

Figs. 3.21–3.23 and Figs. 3.24–3.26 show the discriminatory peaks, informed by both PCA and PIR analysis, against the 781 cm^{-1} (DNA/RNA) and 1447 cm^{-1} (lipids and proteins) peaks, respectively. Markers which are not fully discriminated for one or more of the cancer/normal pairs are not shown on these graphs. The peak at 1447 cm^{-1} was selected as a basis peak for the PIR analysis as it has been used extensively to normalise Raman cell spectra in the literature [11] [96] [116]. 1447 cm^{-1} is used in the literature to normalise as it often has a small variance, making it useful for studying the variance of other peaks, which can be seen here in the PCA loadings (Figs. 3.14–3.20). Note that the 1447 cm^{-1} peak does appear in some of the loadings, but not those which correspond to differences between normal and cancer cells. The peak at 781 cm^{-1} (DNA/RNA) [10] was used as it contributes one of the largest variances between cancer and normal, as seen in the PCA loadings that separate these, i.e. PC4 for the standard cell lines and PC2 for H517–15 (Figs. 3.16 and 3.20). Crow *et al.* (2005), for example, found large variances in the DNA/RNA peak at 781 cm^{-1} between prostate cancer samples at varying stages of differentiation [20].

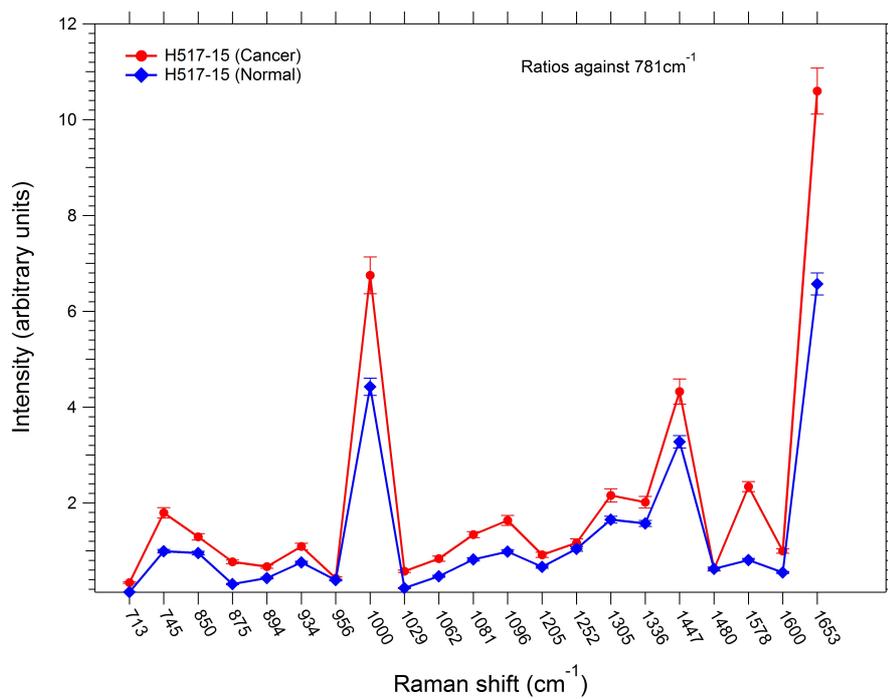


Figure 3.21: PIR markers from the fingerprint peak intensity ratios against 781 cm^{-1} (DNA/RNA) for the H517–15 primary patient samples. Markers shown here are fully discriminatory for one or more of the cancer vs. normal pairs, non-discriminatory markers are not shown.

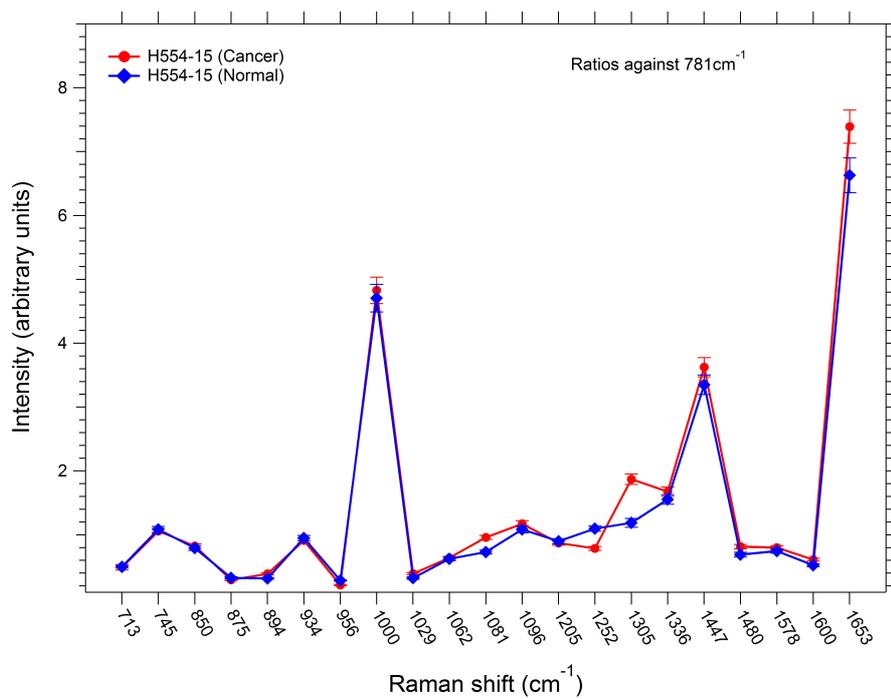


Figure 3.22: PIR markers from the fingerprint peak intensity ratios against 781 cm^{-1} (DNA/RNA) for the H554–15 primary patient samples. Markers shown here are fully discriminatory for one or more of the cancer vs. normal pairs, non-discriminatory markers are not shown.

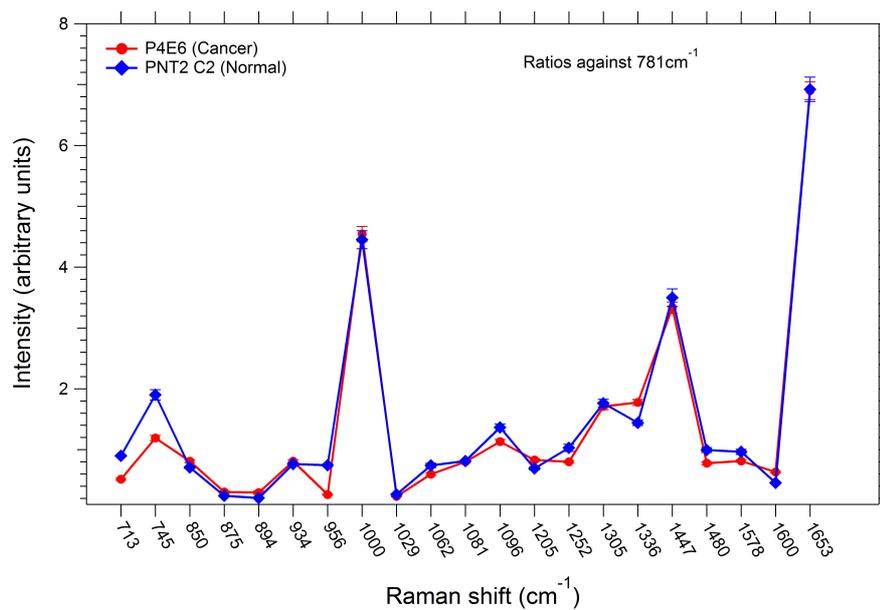


Figure 3.23: PIR markers from the fingerprint peak intensity ratios against 781 cm^{-1} (DNA/RNA) for the P4E6 cancer and PNT2-C2 normal cell line samples. Markers shown here are fully discriminatory for one or more of the cancer vs. normal pairs, non-discriminatory markers are not shown.

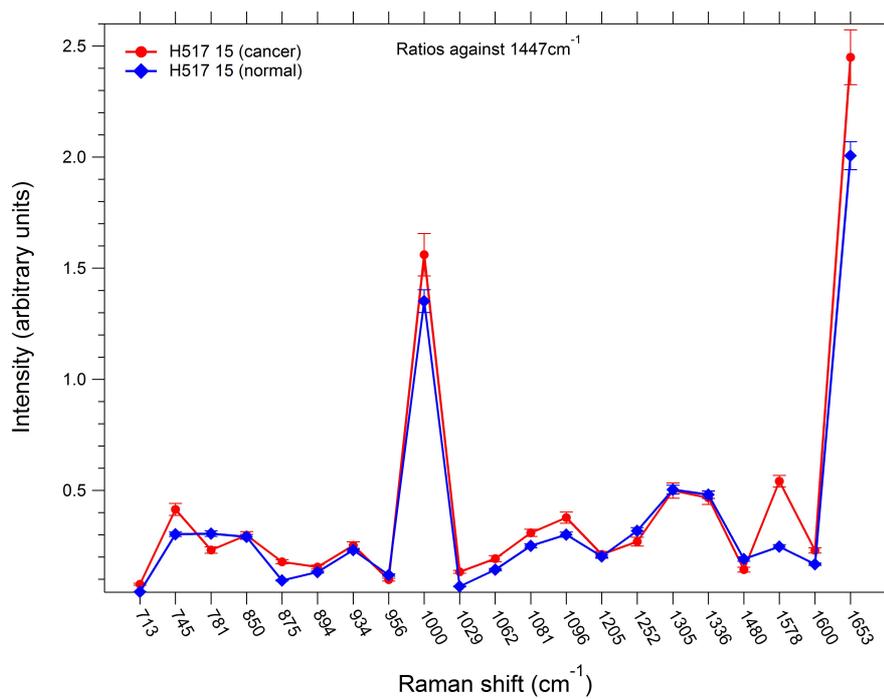


Figure 3.24: Peak intensity ratio (PIR) markers from the fingerprint peak intensity ratios against 1447 cm⁻¹(lipids and proteins) for the H517–15 primary patient samples. Markers shown here are fully discriminatory for one or more of the cancer vs. normal pairs, non-discriminatory markers are not shown.

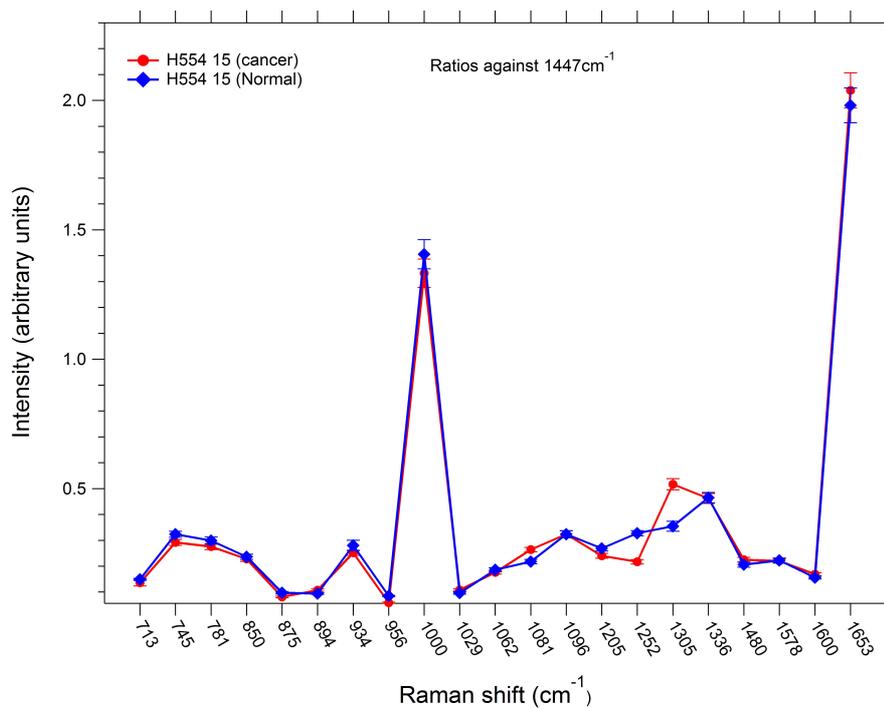


Figure 3.25: PIR markers from the fingerprint peak intensity ratios against 1447 cm⁻¹ (lipids and proteins) for the H554–15 primary patient samples. Markers shown here are fully discriminatory for one or more of the cancer vs. normal pairs, non-discriminatory markers are not shown.

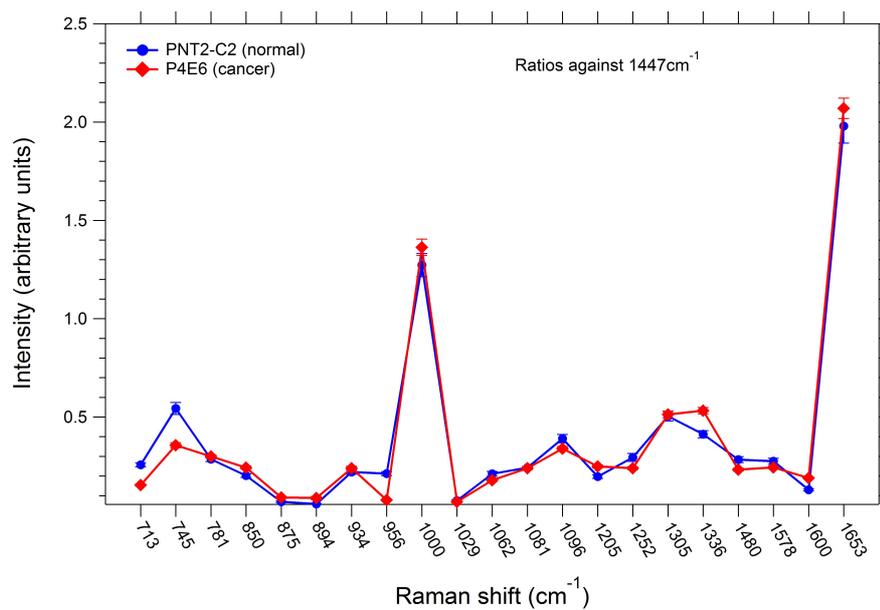


Figure 3.26: PIR markers from the fingerprint peak intensity ratios against 1447 cm⁻¹ (lipids and proteins) for the P4E6 cancer and PNT2-C2 normal cell line samples. Markers shown here are fully discriminatory for one or more of the cancer vs. normal pairs, non-discriminatory markers are not shown.

Phenylalanine has been linked to malignancy in prostate cells in RS studies using markers at 1000 cm^{-1} and 1600 cm^{-1} . Increases in phenylalanine peaks for lung cancer were reported by Huang *et al.* (2003) when the Raman spectra were normalised to the area under the curve [32]. The phenylalanine peak at 1600 cm^{-1} forms part of a shoulder to the larger lipid/protein peak at 1653 cm^{-1} and is therefore not an accurate representation of the phenylalanine content of the cells. The peak at 1000 cm^{-1} , however, is not convolved and can be used as a reliable estimate of phenylalanine content [10]. From the PCA loadings shown in Figs. 3.14, 3.14 and 3.20 we see that the peak at 1000 cm^{-1} is one of the largest sources of variance between the samples, specifically for loadings which account for separation between the normal and cancer samples, i.e. PC4 for the standard cell lines and PC2 for H517–15. Figs. 3.21 – 3.23 and Figs. 3.24 – 3.26 show the 1000 cm^{-1} phenylalanine peak against the 781 cm^{-1} and 1447 cm^{-1} peaks respectively. From these figures we see that the PIR for phenylalanine is increased relative to both the 781 cm^{-1} and 1447 cm^{-1} peaks for the H517–15 patient cancer sample, relative to its normal match. However, both the H554–15 patient and standard cell lines show no change in this PIR marker for cancer vs. normal.

The Raman peaks corresponding to DNA/RNA which are less convolved and therefore more representative of the DNA/RNA content of the cell are at 745 cm^{-1} and 781 cm^{-1} [10]. From the PCA loadings shown in Figs. 3.14, 3.14 and 3.20 we see that there is a large amount of variance between cancer and normal for the 745 cm^{-1} and 781 cm^{-1} DNA/RNA peaks in the H517–15 and standard cell line samples.

Fig. 3.24 shows the PIR at 781 cm^{-1} (DNA) to be decreased for H517–15 (cancer) relative to the 1447 cm^{-1} peak, whereas Figs. 3.25 and 3.26 show no discriminatory change in the I781/I1447 PIR for both H554–15 and the standard cell lines comparisons. The general increase seen in most PIRs when taken against the 781 cm^{-1} peak (Fig. 3.21) in the H517–15 comparison also suggests a marked sensitivity to the 781 cm^{-1} peak.

Figs. 3.21 and 3.24 show that two of the biggest discriminatory markers between the H517–15 cancer and normal samples occurs for the I1578/I781 and I1578/I1447 peak intensity ratios, with 1578 cm^{-1} corresponding to guanine and adenine [10], which are both components of DNA [117]. From the PCA loadings shown in Fig. 3.20 we see a high degree of variance in this peak in PC2 for the H517–15

3.1 Raman Spectroscopy Studies of Prostate Cancer and Normal Prostate Cells

sample suggesting it is a marker for cancer. This peak also appears in PC1 and PC3 for the H554–15 sample and PC3 for the standard cell line samples. These loadings, however, do not correspond to any difference between the normal and cancer cell spectra. The 1578 cm^{-1} PIR is increased relative to the 1447 cm^{-1} protein/lipid peak for cancer compared to its normal match in the H517–15 sample. This result can be compared to the other samples studied which did not show any relative difference for this marker between cancer and normal (Figs. 3.25 and 3.26).

There is some debate in the literature as to whether up-regulation or down regulation of DNA Raman markers is indicative of cell malignancy. Crow *et al.* 2003 suggested that malignancy can be represented in Raman spectra by higher intensities in peaks corresponding to DNA/RNA [11] when normalised to the 1447 cm^{-1} peak, which is agreement with Taleb *et al* [67]. However, in 2005, Crow *et al.* concluded that the poorly differentiated, or more aggressive prostate cancer cell line, DU145 had lower levels of nuclear materials and some proteins when compared the more well differentiated cell lines [20]. This lack of agreement raises the question as to whether DNA markers can be used to accurately discriminate between cancer and normal cells in the prostate using RS.

It is widely known that DNA is not uniformly distributed within the nucleus of the cell and that prostate cancer cells do express more DNA than non-cancerous cells [118]. This non-homogeneous distribution of DNA may lead to inaccuracies in spectra taken from points within a nucleus. This may have been a factor in previous literature therefore leading to this discrepancy. In this work, steps were taken to ensure that the Raman spectra collected were statistically and robustly representative of the average nucleus of each sample. For example, Figs. 3.8 and 3.9 show clearly that the standard error of the mean converged for all samples in both the fingerprint and high-wavenumber regions. The high number of spectra and good convergence obtained in our study indicate a good level of confidence that these results are indicative of the DNA/RNA content of the nucleus of the samples studied. Our conclusion from this therefore, is that markers relative to DNA/RNA are discriminatory in prostate cancer vs. normal cell line studies for both the H517–15 primary patient derived sample and the PNT2–C2 normal and P4E6 cancer standard cell lines.

The peak at 850 cm^{-1} (proteins) [10] appears in PC4 for the standard cell line PCA. The 850 cm^{-1} peak provides a small discrimination between normal and cancer for the P4E6 cancer and PNT2–C2

normal standard cell lines (Fig. 3.26) when taken against the peak at 1447 cm^{-1} .

The PIR H554–15 patient results in Figs. 3.22 and 3.25 show that, despite there being similarities between cancer vs. normal in most of the PIRs, and the PCA loadings (Fig. 3.13), there are still two large differences between the cancer and normal spectra. These differences in PIRs are apparent at 1252 cm^{-1} and 1305 cm^{-1} relative to both 781 cm^{-1} and 1447 cm^{-1} . The 1252 cm^{-1} and 1305 cm^{-1} peaks are heavily convolved with DNA, lipids and proteins [10]. These markers were not picked up by the PCA.

Other peaks are present in the Raman spectrum for the standard cell lines do provide discrimination between cancer and normal, yet were not picked up by the PCA loadings. This is possibly due to being convolved with other peaks which account for greater variances across the sample as a whole. These peaks are 956 cm^{-1} (cholesterol), which is down-regulated in cancer relative to 1447 cm^{-1} and 1096 cm^{-1} (DNA/RNA, proteins and lipids), which is down regulated in cancer relative to 1447 cm^{-1} (Fig. 3.26). Also fully separated are PIR markers for various protein peaks at 1205 cm^{-1} and 1600 cm^{-1} which are both up-regulated relative to 1447 cm^{-1} and 1480 cm^{-1} which is down-regulated relative to 1447 cm^{-1} (Fig. 3.26).

With the exception of the cholesterol peak at 956 cm^{-1} it is very difficult to extract any biomolecular information from these markers due to their high degree of convolution with other markers hence lack of distinct biological assignment (Table. 3.1). Many of these peaks were highlighted as discriminatory between different prostate cancer cell lines at various stages of differentiation, by Crow *et al.* 2005. PCA loadings in Crow *et al.* (2005), show variances between different stage cancers at 956 cm^{-1} , 1090 cm^{-1} and 1480 cm^{-1} in agreement with this study. The peak at 1205 cm^{-1} appears in the loadings in Crow *et al.* as a shoulder to a larger peak around 1190 cm^{-1} , it is difficult to ascertain whether this corresponds to any variance specific to 1205 cm^{-1} .

Fig. 3.24 shows further discriminatory markers between cancer vs normal cells for the H517–15 patient sample at both 1480 cm^{-1} (down-regulated in cancer) and 1600 cm^{-1} (up-regulated in cancer) both of which correspond to proteins and show similar results for the standard cell lines (Fig. 3.26). The 1600 cm^{-1} peak was picked up by PC2 for the H517–15 sample, however 1480 cm^{-1} was not (Fig. 3.20).

This lack of agreement between some of the PCA and PIR analysis results highlights a key flaw in PCA. Despite careful pre-processing to remove the background and normalise each individual spectra it is clear from PC1 for each sample that variances in background are still present (Figs. 3.14, 3.16 and 3.20). It was concluded that this lack of agreement between these two techniques stems from these difficulties. The accuracy of the PIR analysis in this study is corroborated by the convergences shown in Figs. 3.8 and 3.9 and the good agreement in the cancer biomarkers between the H517–15 primary patient sample, and the non-RS literature.

The loadings from the PCA show a large degree of variance in PC1, PC2 and PC3 for the H517–15 sample in the region between 1029 cm^{-1} and 1100 cm^{-1} , Fig. 3.24 shows that the peaks at 1029 , 1062 , 1081 and 1096 cm^{-1} are all up-regulated in the cancer sample relative to 1447 cm^{-1} . The peaks in this region all correspond to a variety of proteins, lipids, phospholipids and DNA/RNA [10] all convolved with one another, again making it difficult to extract any meaningful biomolecular information, however, showing there to be global biomolecular differences between the cancer vs normal samples for this patient. Also discriminatory is the ratio I1252/I1447, where 1252 cm^{-1} corresponds to guanine and lipids, which is down-regulated in cancer. Several of these markers were shown to discriminate between cancer vs normal cells by Crow *et al.* (2005) in the PCA loadings [20], specifically the region between 1029 cm^{-1} and 1100 cm^{-1} and the 1252 cm^{-1} peak.

One of the drawbacks of RS studies on complex biological systems, such as cells, is that there can be large, extensive convolution of different biomolecular signals, such as the $1029\text{--}1100\text{ cm}^{-1}$ region. While regions such as this can provide global markers for separating cancer cells from normal cells, as with the H517–15 and standard cell line samples in this study (Figs. 3.24 and 3.26) it is difficult to determine the biomolecular meaning behind these markers, also conferred by Notingher *et al.* 2004 [25].

The TUFA/TFA (total unsaturated fatty acids/total fatty acids) ratio (I3008/I2850) is directly proportional to the unsaturated fatty acid component of a cells membrane composition [27] [26] [61]. A lower TUFA/TFA ratio, or unsaturation degree, is indicative of de novo lipogenesis, the conversion of excess carbohydrates into fatty acids, which is thought to be a hallmark of malignancy in cells [62] [63]. Although no PCA/LDA was performed on the high-wavenumber region for this study, the TUFA/TFA marker is known in the RS literature as a marker for malignancy in cells.

3.1 Raman Spectroscopy Studies of Prostate Cancer and Normal Prostate Cells

For the H554–15 cancer and normal matched samples it was not possible to fully differentiate (i.e. outside of the SE uncertainty) between the normal and cancer cells using the TUFA/TFA marker, as shown in table. 3.3. This is because the values for the TUFA/TFA ratio for both normal and cancer fall within the SE range of the other (Table. 3.3). A similar finding was reported by Neiva *et.al*, who compared malignant and normal breast cancer cells [61] and found that the TUFA/TFA marker was not sensitive enough to fully differentiate the samples. However, the H517–15 patient samples and the standard cell lines (cancer vs normal) saw a decrease in this ratio for the cancer cells outside of the SE range of the normal cells (Table. 3.3). Neiva *et al.* predicted that cell malignancy would correspond to a decrease in this ratio due to an increase in the total number of lipids (TFA) [56], but could not acquire sensitive enough data to prove this hypothesis [61].

A previous RS study by Potcoava *et al.* found that the TUFA/TFA ratio can be used to discriminate hormone responsive prostate cancer cell lines, such as LNCaP [71], before and after treatment with medroxyprogesterone acetate, a hormone cancer treatment [27]. In this study the TUFA/TFA ratio separates both the P4E6 cancer cell line sample and H517–15 primary patient cancer sample from all three normal samples, which group together with a higher value (Table. 3.3). Also grouped with the three normal samples is the H554–15 cancer sample which serves to re-affirm the PCA/LDA findings, where the H554–15 cancer sample behaves more like normal prostate cells.

Figs. 3.27 – 3.29 show the averaged, baselined and normalised spectra, laid over one another for direct comparison. The spectra in these figures are normalised against 1447 cm^{-1} . These figures allow a direct comparison of the peak heights relative to the 1447 cm^{-1} peak. From these figures we can confirm the findings of the PIR analysis by inspecting the differences in intensity between particular peaks. This, however, is a far less quantitative measure of the relative intensities and should be used only as a visual aid.

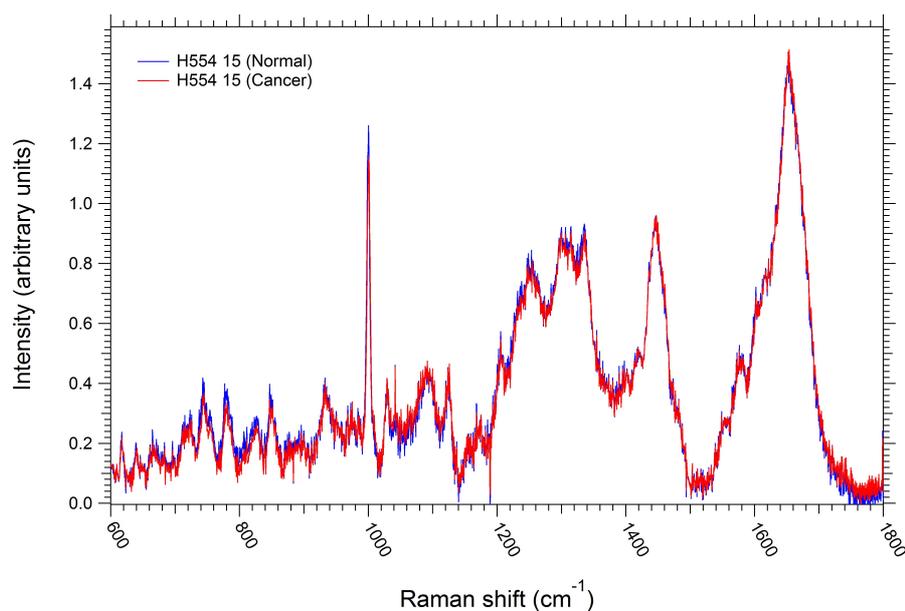


Figure 3.27: Raman spectra for the H554–15 primary patient samples. The spectra have been averaged, baselined and normalised against 1447 cm⁻¹. The spectra have not been shifted so a direct comparison can be made.

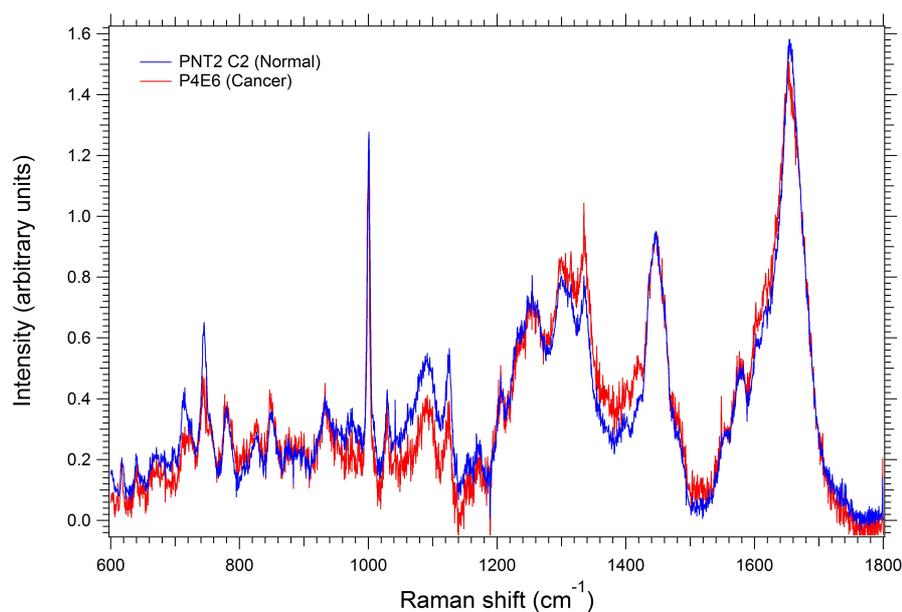


Figure 3.28: Raman spectra for the P4E6 cancer and PNT2–C2 normal cell lines. The spectra have been averaged, baselined and normalised against 1447 cm⁻¹. The spectra have not been shifted so a direct comparison can be made.

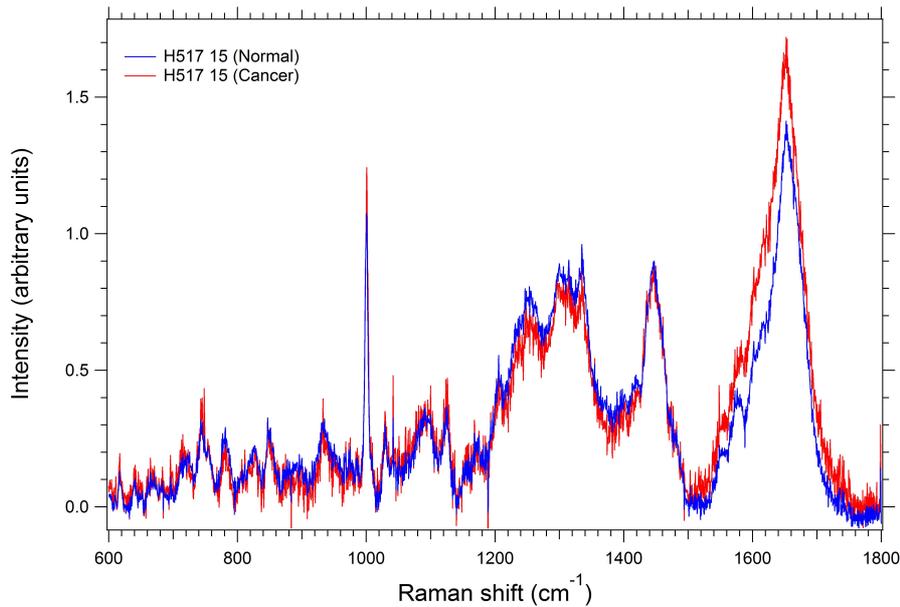


Figure 3.29: Raman spectra for the H517–15 primary patient samples. The spectra have been averaged, baselined and normalised against 1447 cm^{-1} . The spectra have not been shifted so a direct comparison can be made.

3.1.5 Prostate Cancer Summary

This work demonstrates the ability of RS to not only differentiate different standard cell lines, but also differentiate normal cells from cancer cells for a primary patient derived sample. For the primary patient sample, H517–15 RS was able to identify several biomarkers which are consistent with biological literature on cancer, such as the down regulation of the TUFA/TFA ratio in cancer, indicating *de novo* lipogenesis, a hallmark of cancer [62] [63]. Other markers include the increase in the 1000 cm^{-1} phenylalanine peak in cancer relative to the 1447 cm^{-1} peak, in agreement with Huang *et al.* (2003) [32] and the increase in the guanine and adenine peak at 1578 cm^{-1} in cancer, relative to the 1447 cm^{-1} peak, in agreement with Crow *et al.* (2003) [11].

These PIRs are, for the most part, also corroborated by PCA loadings which separate the normal cells from cancer cells. However, some of the discriminatory PIR markers are not present in the relevant PCs. Despite careful pre-processing it is apparent, especially so in the first PCs, that the spectral background contributed a large portion of the variance. This may have caused certain discriminatory peaks to become lost in the large background variance.

3.1 Raman Spectroscopy Studies of Prostate Cancer and Normal Prostate Cells

This study also highlighted the need for more work to be done using primary patient derived samples as opposed to standard cell lines. This outcome is made apparent throughout this chapter by the disagreements in the biomarkers between the H517–15 primary samples and P4E6 and PNT2–C2 standard cell line samples. The biomolecular information presented in the H517–15 data, as previously stated, closely matches what is known about prostate and other cancers, whereas there is considerable disagreement with the standard cell lines. This could lead to the development of treatments or diagnostic techniques which, while functional in a laboratory using cell lines, have little relevance to actual patient cancers.

This study found that a second primary patient derived sample, H554–15, displayed little difference between its cancer and normal matched spectra, despite having similar first histopathology. This result could serve to highlight the ability of RS to replace tissue histopathology as a diagnosis technique. The H554–15 sample has been sent back to the source hospital for further testing to ascertain the accuracy of the RS result.

3.1.6 Further Work For Raman Spectroscopy Studies of Prostate Cancer

To expand on the work in this study it would be useful to complete principle component analysis on the high wavenumber ($2700\text{--}3100\text{ cm}^{-1}$) region of the prostate cancer spectra. The graphs and PC loadings from this could inform a deeper lipidomic PIR analysis.

Also, future work could involve adding further primary patient samples and standard cell lines to this data set in order to understand fully where the inconsistencies lie between cancer vs. normal samples. As has been set out in this study and other studies in the literature [74] [75], standard cell lines which are currently used to represent patient biology for both cancer and normal prostate cells do not exhibit the matching Raman spectra which would be expected. Especially in the case of the PNT2–C2 normal prostate cell line, the biomolecular signatures in its Raman spectra are stark in contrast to those of the normal primary patient samples. This is something which must be addressed. Finally, adding further primary cancer samples and a Gleason seven standard cell line, would provide the opportunity to determine the ability of RS to provide patient specific biomolecular information pertaining to things such as Gleason grade, similar to work by Crow *et al.* (2003), using PIR analysis. Adding a standard cell

line which is thought to more closely represent the biology of the primary patient samples in this study would allow for a more direct comparison of the use of standard cell lines in prostate cancer research.

Aside from biomolecular analysis, Raman spectroscopy has proved that when combined with a multivariate analysis technique, such as PCA/LDA, it can achieve very low errors when diagnosing cancer from normal cells. The application of this in practice, in a hospital could drastically reduce both cost and error rates. Additionally, although great care was taken in the pre-processing of the spectra prior to PCA/LDA in this work, it is possible that manual baseline subtraction could improve the quality of the loadings. Better quality loadings would allow peaks which are currently lost in background to be identified. It was not possible for this thesis to perform manual background subtraction as it would have required the manual processing of over 1000 individual spectra.

3.2 Raman Spectroscopy Studies of *Streptomyces coelicolor* Bacteria

This section of the study was performed as part of a contribution to a paper entitled “Antibiotic hypersensitive mutants of *S. coelicolor*”, which aims to probe the underlying mechanisms behind a mutation in *S. coelicolor* bacteria which causes a loss of intrinsic antibiotic resistance [1]. It was proposed that Raman spectroscopy be used to analyse bulk *S. coelicolor* cultures to determine changes in lipid content between three different strains. The three different strains to be analysed are J1929, the ‘wild’ type with no mutations or treatments, DT3017, the strain with the mutation causing loss of antibiotic resistance and PDT16, the treated, mutated strain. The treatment is designed to reverse the effects of the mutation in PDT16 and return the bacteria to a state similar to that of the J1929 sample by supplementing the mutated *ppm1* gene in the DT3017 with *ppm1* encoding plasmids [1].

This study was able to identify changes in lipid content between the J1929 and DT3017 samples, as expected. This study was also able to show a return to the wild type lipid composition for the PDT16 treated sample, indicating that the treatment was effective.

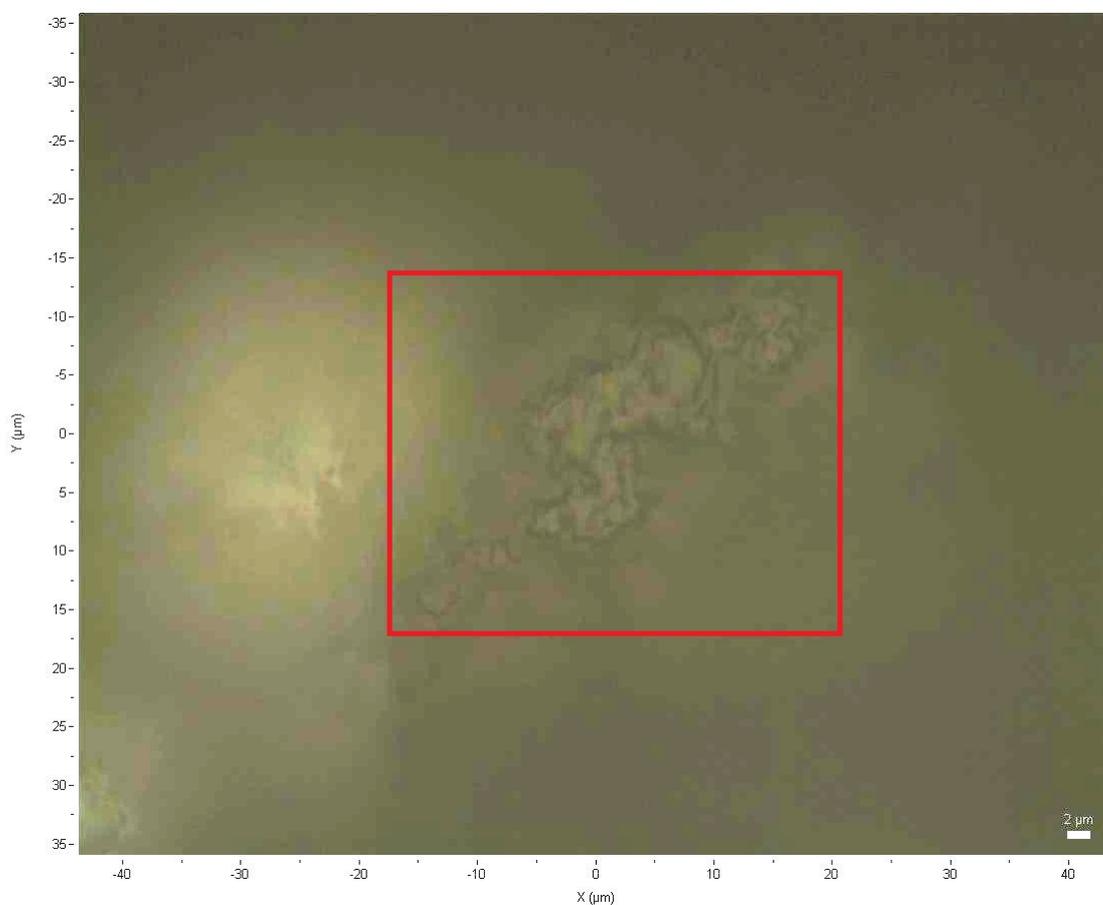


Figure 3.30: An optical image, using a 100x objective microscope lens with a numerical aperture of 0.9, of the surface of the *S. coelicolor* sample J1929. Owing to the rough surface of the sample and the confocal nature of the microscope, the in focus area is very small, shown in the red box.

3.2.1 Raman Spectra of *Streptomyces coelicolor* Bacteria

This study was performed as part of a collaboration on a paper in preparation with Prof. Maggie Smith's group at the University of York to look at the causes and effects of a mutation in the *ppm1* gene of *S. coelicolor* bacteria and its subsequent treatment. The three samples used in this study are J1929, the 'wild' type *S. coelicolor* with no mutations or treatments, DT3017, the strain with the mutation in the *ppm1* gene, and PDT16, the mutant strain which has been treated.

Differential RNA expression analysis performed by Prof. Smith's group found that the mutant strain displayed a switch from catabolism (the breaking down of molecules), to anabolism (the building of more complex molecules) [119]. This dramatic switch has led to a change in membrane composition within the cells, in turn altering the lipid content. This is thought to be causing the loss of intrinsic antibiotic resistance seen in the *ppm1* mutant [1]. Further to this, DESeq analysis of RNA sequencing data on the three strains showed that of the 658 genes altered between the J1929 and DT3017 strains, 574 of these changes showed some restoration toward J1929 in the PDT16 strain [1]. These results together suggest that the J1929 wild type and DT3017 treated types should have similar lipid compositions, different to the PDT16, mutant type.

Raman was used to assess the lipid content of three samples by means of the TUFA/TFA (Total Unsaturated Fatty Acid/Total Fatty Acid) ratio (I3008/I2850) [27] [26]. Both Potcoava *et al.* and Nieva *et al.* used this ratio and the mass unsaturation ratio (I1662/I1446) (C=C/CH₂) band to determine changes in the lipid content of human cancer cells [27] [61]. Potcoava *et al.* also used the I1662/I1294 ratio, which is also related to the unsaturated content of lipids [27]. Wu *et al.* performed RS lipidomics on single algae cells using the I1662/I1446 PIR. They showed that the peak intensity ratio (I1662/I1447) correlates to the degree of lipid unsaturation in extracted lipids from algae cells. Wu *et al.* however identified the 1446 cm⁻¹ and 1662 cm⁻¹ peaks as being related to lipids only, which in the case of Raman on pure lipids holds true [26]. In this study, the 1446 cm⁻¹ and 1662 cm⁻¹ peaks are identified as being convolved with proteins, making them unsuitable for performing lipidomics (see Table. 3.5).

The TUFA/TFA (I3008/I2850) ratio was used in this study as it relates only to pure lipid peaks [10]. Very little quantitative high wavenumber (2750–3110 cm⁻¹) Raman analysis has been published on bacteria, meaning the use of the TUFA/TFA ratio on *S. coelicolor* bacteria in this study is novel.

The use of the TUFA/TFA ratio is, however, justified by the similarities in spectral profile between *S. coelicolor* and other cell samples which have used the TUFA/TFA ratio [27] [61].

The averaged *S. coelicolor* Raman spectra for the three strains, J1929 (wild), PDT16 (treated) and DT3017 (untreated), (Figs. 3.31 and 3.32) and the peak position assignments in Table. 3.5 are consistent with previous RS literature on *Streptomyces* cells [80] and other published RS works on bacteria [84]. An example of previously published Raman spectra on *Streptomyces* bacteria can be seen in Fig. 1.4 [80], showing favourable comparison with the spectra in Figs. 3.31 and 3.32. All peak positions identified in Figs. 3.31 and 3.32 were obtained by fitting the average spectra for each sample, then averaging the peak positions, the peak positions and their assignment are shown in Table. 3.5. The associated uncertainty in the peak positions is 1cm^{-1} , which is the same as the uncertainty due to the error in the Raman equipment.

The convergence of the percentage SE of the TUFA/TFA peak intensity ratio (I3008/I2850) [27] as a function of the number of randomly selected individual spectra in the average indicates that sufficient spectra were collected from each strain to establish reliable statistics for the measurements (Figs. 3.33, 3.34 and 3.35). The convergence plots shown in Figs. 3.33, 3.34 and 3.35 give insight into the heterogeneity of the samples under analysis. As these are bulk samples (i.e. not sampled at the single cell level), as opposed to the single cell sampling used in the prostate cancer study, there is likely to be a greater variation in the spectra [80]. All of the percentage SE of the TUFA/TFA peak intensity ratios converge at roughly the same rate for this study, indicating a uniform heterogeneity across the three samples.

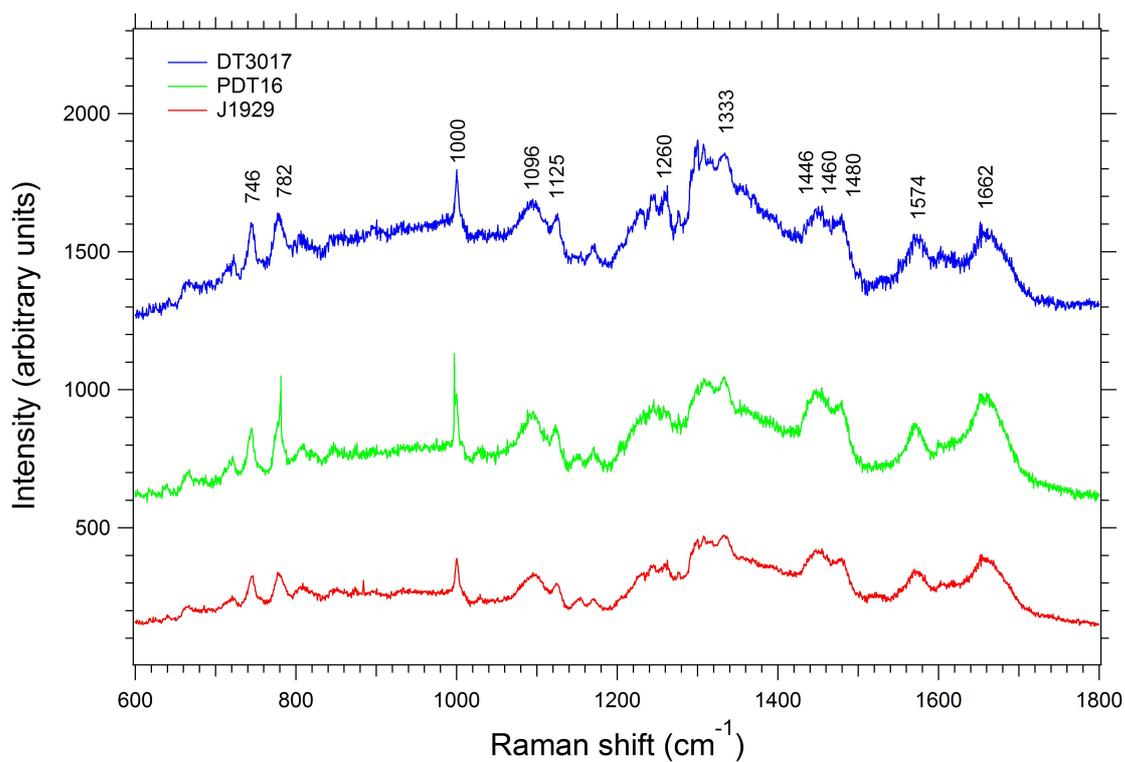


Figure 3.31: Averaged Raman spectra in the fingerprint region ($600\text{-}1800\text{ cm}^{-1}$) for the three *S. coelicolor* cell lines (J1929 wild-type, PDT16 treated and DT3017 untreated) showing the key peak assignments (Table 3.5). The spectra are vertically shifted using an arbitrary offset to aid visualisation.

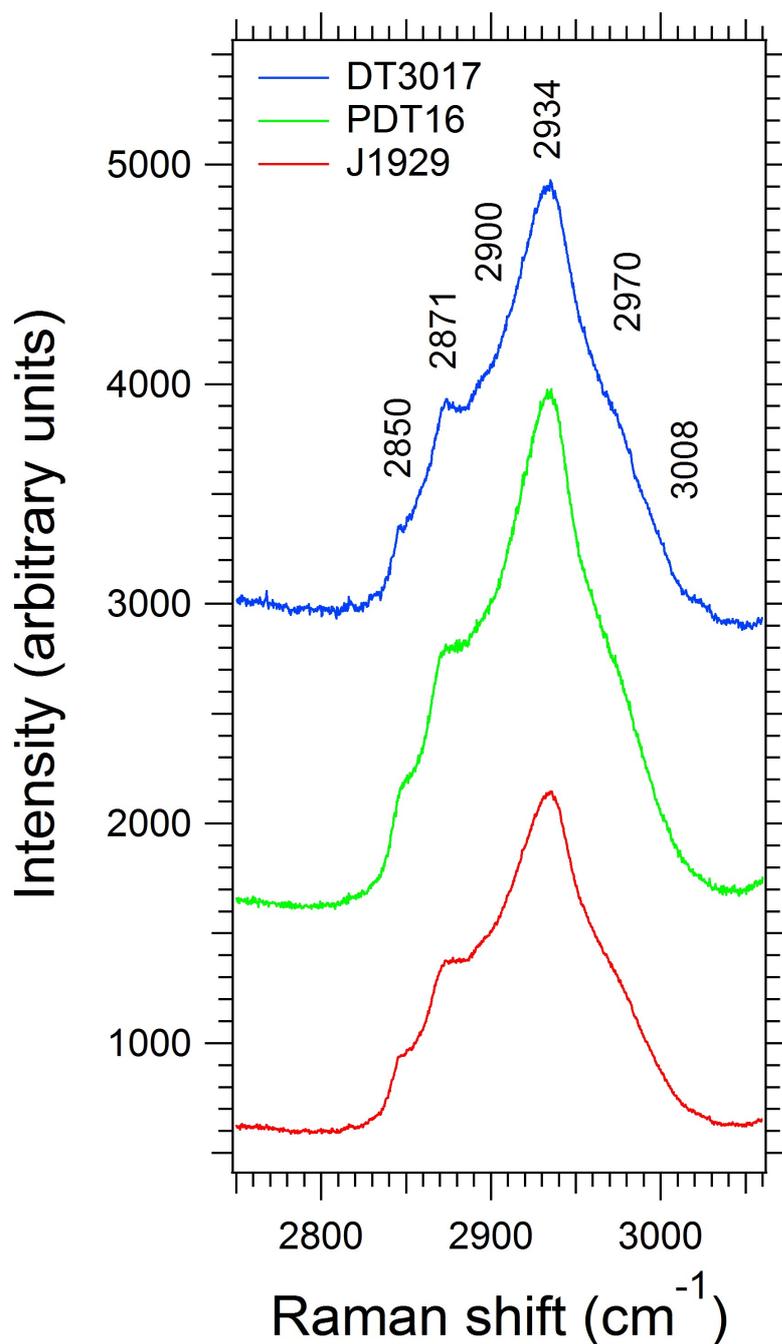


Figure 3.32: Averaged Raman spectra in the high-wavenumber region (2750-3110cm⁻¹) for the three *S. coelicolor* cell lines (J1929 wild- type, PDT16 treated and DT3017 untreated) showing the key peak assignments (Table 3.5). The spectra are vertically shifted using an arbitrary offset to aid visualisation.

Peak Position $\pm 1(\text{cm}^{-1})$	Assignments
746	DNA/RNA [97]
782	DNA/RNA [99]
1000	Phenylalanine [105]
1096	DNA/RNA, and lipids [105] [109]
1446	CH ₂ bending mode of proteins and lipids [110]
1460	Lipids (CH ₂ /CH ₃) [100]
1480	Proteins (Amide II) [110]
1578	Guanine, adenine [25]
1662	Proteins and lipid (C=C stretch) [66]
2850	CH ₂ stretching of lipids [112]
2871	CH ₂ stretching and CH stretch of lipids and proteins [112]
2900	CH ₂ stretching of lipids and proteins, CH stretching of lipids and proteins [112]
2934	Lipids (CH ₂ , CH ₃) [112]
2970	lipids and cholesterol (CH ₃) [103]
3008	Lipids (unsaturated, CH stretch) [103]

Table 3.5: All peak positions identified in these experiments and their published assignments along with examples of their uses in the literature. All peak assignments correspond to Movasaghi *et al.* (2007). These peak positions were identified by fitting the average spectra for each sample, then averaging the peak positions. The associated error in the peak locations is approximately 1cm^{-1} .

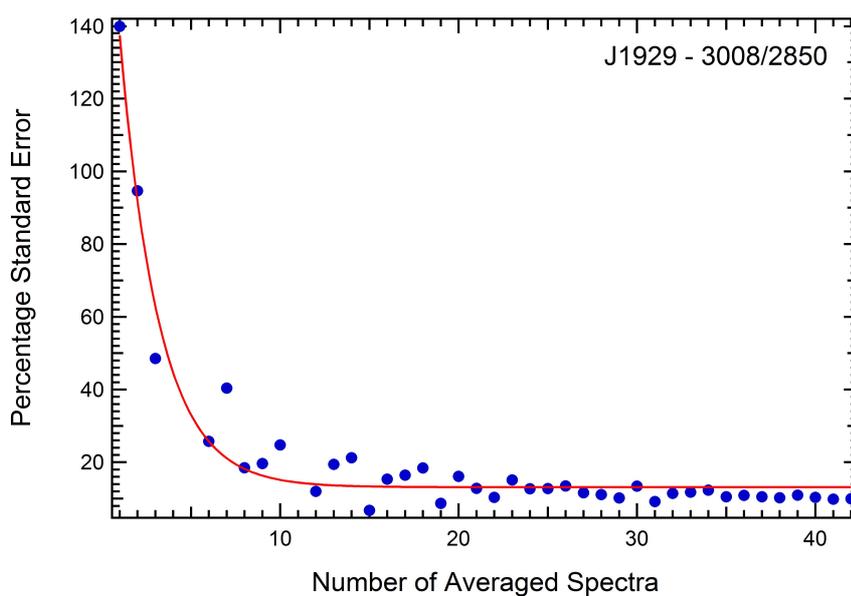


Figure 3.33: Convergence of the percentage standard error of the mean (SE) for the TUFA/TFA peak intensity ratio I_{3008}/I_{2850} for the J1929 (wild) cell line as a function of the number of randomly chosen spectra comprising the spectral average. In total, 42 individual spectra were averaged.

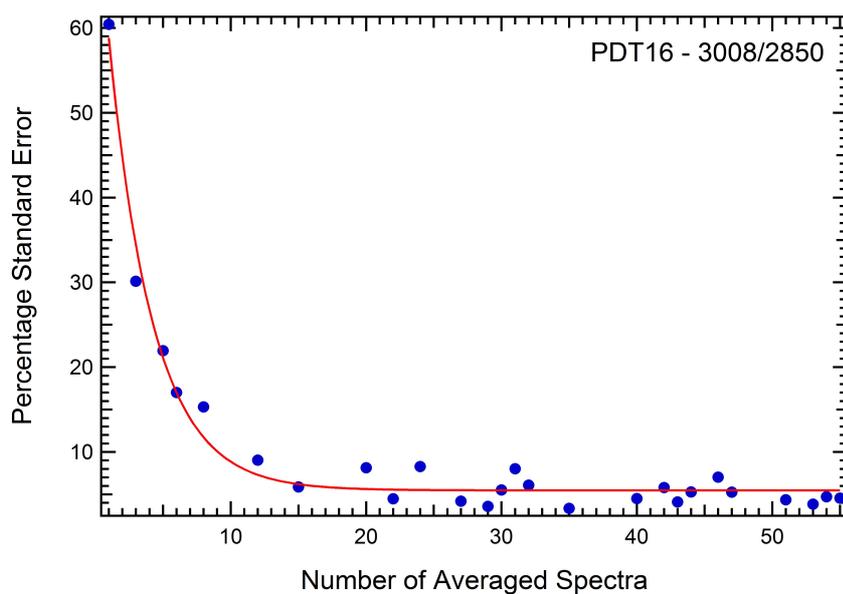


Figure 3.34: Convergence of the percentage standard error of the mean (SE) for the TUFA/TFA peak intensity ratio I3008/I2850 for the PDT16 (treated) cell line as a function of the number of randomly chosen spectra comprising the spectral average. In total, 55 individual spectra were averaged.

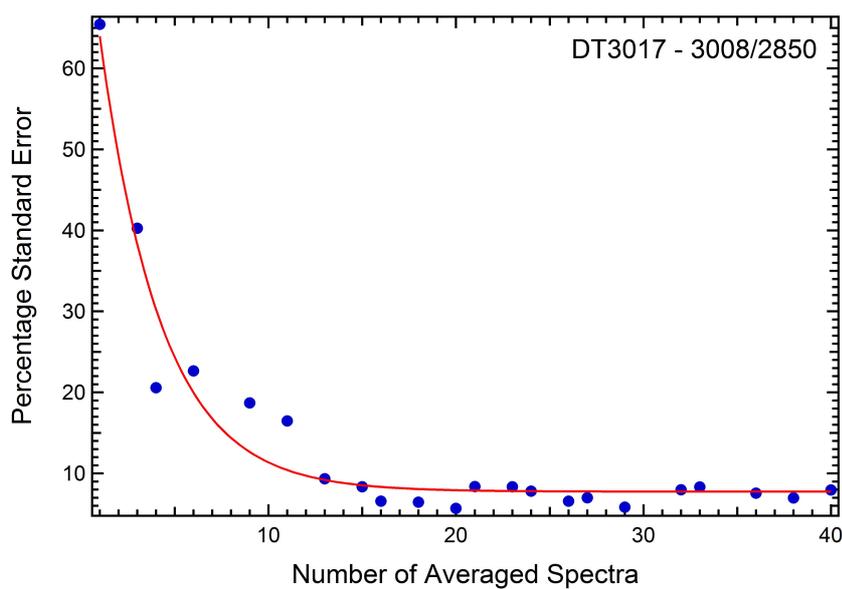


Figure 3.35: Convergence of the percentage standard error of the mean (SE) for the TUFA/TFA peak intensity ratio I3008/I2850 for the DT3017 (untreated) cell line as a function of the number of randomly chosen spectra comprising the spectral average. In total, 40 individual spectra were averaged.

3.2.2 Peak Intensity Ratio Comparisons

The TUFA/TFA marker shows that the treated PDT16 cells regain a similar lipid composition to that of the J1929 wild cells when comparing the TUFA/TFA marker mean values and the TUFA/TFA marker ranges that take into account the SE (Table 3.6). The TUFA/TFA marker for the DT3017 untreated cells lies outside of the range obtained for the J1929 wild and PDT16 cell lines confirming that it has a different lipid composition, specifically, a larger unsaturated fatty acid component.

<i>S. coelicolor</i> cell line	No. of spectra per spectral average	TUFA/TFA ratio \pm SE	TUFA/TFA range
J1929 (wild)	42	0.47 ± 0.05	0.42 - 0.52
PDT16 (treated)	55	0.51 ± 0.03	0.48 - 0.54
DT3017 (untreated)	40	0.60 ± 0.04	0.56 - 0.64

Table 3.6: Table showing the number of spectra and TUFA/TFA ratios with associated errors for all samples

<i>S. coelicolor</i> cell line comparison	P-value for TUFA/TFA ratio
J1929 (wild) vs. PDT16 (treated)	0.47
J1929 (wild) vs. DT3017 (untreated)	0.047
DT3017 (untreated) vs. PDT16 (treated)	0.049

Table 3.7: Table showing the P-values obtained from a T-test when comparing the TUFA/TFA ratio between the three different samples.

This result concurs with the findings of the differential RNA expression analysis whereby the mutant strain displayed a switch from catabolism (the breaking down of molecules), to anabolism (the building of more complex molecules) [1] [119]. This change led to a change in membrane composition within the cells, as corroborated by the TUFA/TFA PIRs (Table. 3.6). Further to this, the result shows a return to the wild type lipid composition after treatment, suggesting a restoration of the *ppm1* gene. This result is corroborated by DESeq analysis of the RNA sequencing data, whereby of the 658 genes altered between J1929 and DT3017, 574 of these changes showed some restoration toward J1929 in the PDT16 strain [1]. This agreement between RS and other established bioanalysis techniques re-affirms the technique and specifically the use of the TUFA/TFA ratio as an indication of lipid content in this study, thus indicating a broad range in its application in biological samples. These results are supported by the t-test results displayed in table 3.7. These p-values support the differences, or lack thereof, between the three samples discussed.

3.2.3 Further work for Raman Spectroscopy Studies of *Streptomyces coelicolor* Bacteria

This study consisted of only a small fraction of the analysis which RS is capable of for this type of sample. This was a side project to the main prostate cancer study which aimed to firstly provide data to a wider study by Prof. Maggie Smith's group at the University of York and to highlight the versatility of RS as a biomolecular analysis technique. Further work with these samples would be to complete both a full principle component analysis and peak intensity ratio analysis in both the high-wavenumber and fingerprint regions, similar to that carried out for the prostate cancer study. The same considerations that are being made for further work on the prostate cancer samples must also be made here with regards to the pre-processing of individual spectra, prior to PCA/LDA.

Chapter 4

Conclusions

This work demonstrates the ability of Raman spectroscopy (RS) to probe information relating to biomolecular changes between a range of samples, specifically, standard and primary prostate and prostate cancer cell lines, and bulk *S. coelicolor* bacteria.

Using principle component analysis and linear discriminant analysis (PCA/LDA) this study was able to identify differences between spectra taken from primary patient derived prostate and prostate cancer cell lines (H554-15 and H517-15). PCA determined that the H554-15 cancer and normal samples had similar spectra, therefore were biologically similar, but not identical. The H517-15 cancer and normal samples were fully separated using PCA. Examination of the loadings for the principle components (PCs) which separate the cancer cells from the normal cells allows the peaks, or groups of peaks, which vary between the two to be pin-pointed. As no PC separated the H554-5 sample, the loadings for the PCs showed only the variance between individual spectra across both the normal and cancer cells combined. The PC responsible for separating the H517-15 cancer and normal cell spectra (PC2) held information on which peaks varied between the two. The peaks identified in PC2 were the peaks corresponding to amino DNA/RNA (745, 781 and 1578 cm^{-1}), proteins and lipids (1096 cm^{-1} , 1125 cm^{-1} , 1205 cm^{-1} , 1600 cm^{-1} , 1613 cm^{-1} and 1653 cm^{-1}), Phenylalanine (1000 cm^{-1}) and glycogen (934 and 1029 cm^{-1}), these findings were corroborated in the literature.

PCA was also able to separate the two standard cell lines (P4E6 cancer and PNT2-C2 normal). The loading for the PC which provided this separation (PC4) highlighted the points of variance between

the two samples to be the peaks corresponding to DNA/RNA (724, 745 and 781), proteins and lipids (850, 1096, 1125, 1205, 1600, 1613 and 1653 cm^{-1}), phenylalanine (1000 cm^{-1}) and glycogen (1029 cm^{-1}). Once again, variances between normal and cancer cells at these peaks was corroborated in the literature.

PCA also found separations between all three of the cancer cell samples. Given that the H554-15 sample showed very little separation between cancer and normal, and the P4E6 cancer and H517-15 cancer are of different Gleason grade, it is expected that the three samples would show differences. PCA also found that for the three normal samples, the two primary normal samples clustered together, this was expected as they have similar biologies, both being patient derived normal prostate tissue. Also expected was the PNT2-C2 (normal) sample to cluster with the primary normal samples. This was expected as the PNT2-C2 sample is supposed to be representative of normal prostate cells. This, however, was not the case, indicating that something in the processing of the PNT2-C2 cells from patient sample to immortalised standard cell line has altered the biology in such a way as to make them un-representative of patient biology. Further to this study it would be of use to add a Gleason 7 prostate cancer cell line to this data, to determine if similar changes occur for all cell lines. To bolster this data it would also be useful to add further primary patient samples with confirmed histopathology.

PCA has become a staple in RS literature. It does, however, have drawbacks. PCA requires a large amount of pre-processing in order to be able to isolate differences between the samples that are not related to background or spectral noise. This pre-processing, for studies such as this, which involved over 1000 individual spectra, each with subtly different noise and background characteristics, can be time and resource consuming.

The peak location highlighted in the PC loadings were used to inform targeted peak intensity ratio (PIR) analysis. As expected, PIR analysis found very little difference between the H554-15 cancer and normal cells. Only small differences are present in the PIR analysis of the H554-15 samples, in agreement with the PCA/LDA results for the sample. For the H517-15 samples the following PIR changes were noted: an increase in the 713 cm^{-1} amino acid/methionine peak in cancer relative to 1447 cm^{-1} , an increase in the 1000 cm^{-1} phenylalanine peak in cancer relative to the 1447 cm^{-1} peak, an increase in the guanine and adenine peak at 1578 cm^{-1} in cancer, relative to the 1447 cm^{-1} peak, and a decrease in the TUFA/TFA (I2850/I3008)ratio in cancer, amongst others. These biomarkers are all in agreement

with various sources in the literature, both RS and otherwise, affirming the use of Raman to probe accurate biomolecular information from live cell samples. Further to this study it would be useful to expand the PIR analysis to include the whole of the high-wavenumber region of the spectra. In future studies the addition of further primary patient samples would allow a comparison of PIRs, RS has proven here that the technique is sensitive enough to distinguish between different cancers. The next step would be to study large volumes of patient samples to build a reliable, multi-patient diagnosis method, likely involving a combination of PCA and PIR analysis.

The primary limitation of PIR analysis is the amount of information contained within a single spectra. While techniques such as PCA can provide regions of high variance within the spectra, a technique such as PIR is needed to extract meaningful biomolecular data. For a system as complex as a living cell, a large number of peaks are convolved with other peaks, making the process of extracting biomolecular data from them difficult. With developments in optical technologies increasing spectral intensity and reducing uncertainty in the measurements, and increased access to primary patient samples for research purposes it is possible for Raman to become a standard in cancer diagnosis, reinforcing or replacing the flawed current techniques.

The picture with the standard cell lines (P4E6 cancer and PNT2-C2 normal) is not as clear. While the samples show some biomolecular markers in agreement with the literature, such as a decrease in the TUFA/TFA ratio for cancer, other markers, such as phenylalanine (1000^{-1}) and guanine and adenine (1578cm^{-1}) show disagreement with both the H517-15 results and the wider literature on prostate and other cancers. This result further calls into question the use of standard cell lines in cancer research and indicates a move toward using primary patient derived cell lines exclusively else risking the development of ineffective technologies.

The second part of this study on the use of RS to probe biological samples involved the analysis of *S. coelicolor* bacteria. This was a small study with the aim of contributing lipidomics to a deeper study by Prof. Maggie Smith at the University of York. The aim of this study was to analyse the biological changes which occurred in the cells when a wild strain (J1929) underwent a mutation, causing it to lose intrinsic antibiotic resistance (DT3017), and when this mutant strain was treated to reverse this mutation (PDT16). RNA sequencing revealed a possible change in fatty acid content between the wild

type and mutant type and DESeq analysis revealed an 87% return to the genetic make up of the wild type after treatment.

The TUFA/TFA ratio was used to probe changes in lipid content between the three strains of *S. coelicolor* (J1929, DT3017 and PDT16). This ratio revealed similarities between the J1929 and PDT16 samples, as predicted by DESeq analysis, and differences between the J1929 and PDT16 samples, and the DT3017 sample, as predicted by RNA sequencing. The TUFA/TFA PIR showed an increase in unsaturated lipid content for the DT3017 sample. This confirmation by established biological techniques confirms the ability of RS to probe accurate lipidomics from bulk bacteria samples. Future work with these samples will involve a full Raman study, similar to that performed on the prostate cancer cells, including PCA/LDA and full PIR analysis, to provide a more detailed biomolecular view of the mutation and treatment described here.

These two studies taken together highlight the use of Raman spectroscopy to probe biomolecular information from a variety of samples. The use of the TUFA/TFA ratio across both the bacteria and prostate cancer experiments demonstrates the versatility of Raman and PIR analysis. The general agreement between the wider, non-RS, literature and the Raman biomolecular results for the H517–15 patient sample in this study confirms that RS can be used to probe biomarkers associated with prostate cancer to a high degree of accuracy. With the versatility and depth of analysis shown to be possible in this thesis, Raman has the potential to change the way researchers and medical professionals study diseases like cancer.

References

- [1] R. Howlett, N. Read, K. Anttonen, A. Varghese, C. Kershaw, Y. Hancock, and M.C.M. Smith. Antibiotic hypersensitive mutants of *s. coelicolor*. *In Preparation*, 2017.
- [2] C.V. Raman. A new radiation. *Indian Journal of Physics*, 2:387–398, 1928.
- [3] M.B. Fenn, P. Xanthopoulos, G. Pyrgiotakis, S.R. Grobmyer, P.M. Pardalos, and L.L. Hench. Raman spectroscopy for clinical oncology. *Advances in Optical Technologies*, page doi:10.1155/2011/213783, 2011.
- [4] J. Strutt. On the light from the sky, its polarization and colour. *Philosophical Magazine*, 41:107–120, 1871.
- [5] J. Conroy, A.G. Ryder, M.N. Leger, K. Hennessey, and M.G. Madden. Qualitative and quantitative analysis of chlorinated solvents using Raman spectroscopy and machine learning. *Proceedings of the SPIE - The International Society for Optical Engineering*, 5826:131–142, 2005.
- [6] J.R. Ferraro and K. Nakamoto. *Introductory Raman Spectroscopy*, volume 2. Academic Press, San Diego, USA, 1994.
- [7] H.D. Young and R.A. Freedman. *University Physics With Modern Physics (12th Edition)*. Addison Wesley, 2008.
- [8] R.E. Kast, S.C. Tucker, K. Killian, M. Trexler, K.V. Honn, and G.W. Auner. Emerging technology: Applications of Raman spectroscopy for prostate cancer. *Cancer Metastasis Reviews*, 33:673–693, 2014.

-
- [9] L.P. Choo-Smith, H.G.M. Edwards, H.P. Endtz, J.M. Kros, F. Heule, H. Barr, J.S. Robinson, H.A. Bruining, and G.J. Pupells. Medical applications of Raman spectroscopy: From proof of principle to clinical implementation. *Biopolymers (Biospectroscopy)*, 67:1–9, 2002.
- [10] Z. Movasaghi, S. Rehman, and I.U. Rehman. Raman spectroscopy of biological tissues. *Applied Spectroscopy Reviews*, 42:493–541, 2007.
- [11] P. Crow, N. Stone, C.A. Kendall, J.S. Uff, J.A.M. Farmer, H. Barr, and M.P.J. Wright. The use of Raman spectroscopy to identify and grade prostatic adenocarcinoma *in vitro*. *British Journal of Cancer*, 89:106–108, 2003.
- [12] P. Atkins and J. de Paula. *Elements of physical chemistry*, volume 5. Oxford University Press, 2009.
- [13] A.C. Ferrari. Raman spectroscopy of graphene and graphite: Disorder, electron-phonon coupling, doping and nonadiabatic effects. *Solid State Communications*, 143:47–57, 2007.
- [14] S. Costa, E. Borowiak Palen, M. Kruszynska, A. Bachmatiuk, and R.J. Kalenczuk. Characterization of carbon nanotubes by Raman spectroscopy. *Materials Science – Poland*, 26:433–441, 2008.
- [15] A.C. Ferrari and J. Robertson. Raman spectroscopy in carbons: from nanotubes to diamond. *Philos. Trans. R. Soc.*, 362:2267–2565, 2004.
- [16] C. Augusto Fernandes de Oliveira Penido, M. Tadeu Tavares Pacheco, I.K. Lednev, and L. Silveira Jr. Raman spectroscopy in forensic analysis: identification of cocaine and other illegal drugs of abuse. *Journal of Raman Spectroscopy*, 47:28–38, 2015.
- [17] J. Chance Carter, W.E. Brewer, and S. Michael Angel. Raman spectroscopy for the in situ identification of cocaine and selected adulterants. *Applied Spectroscopy*, 54:1876–1881, 2000.
- [18] J.S. Day, H.G. Edwards, S.A. Dobrowski, and A.M. Voice. The detection of drugs of abuse in fingerprints using Raman spectroscopy. *Spectrochim. Acta. A*, 60:563–568, 2004.
- [19] S. Verrier, I. Notingher, J.M. Polak, and L.L. Hench. In situ monitoring of cell death using Raman microspectroscopy. *Biopolymers*, 74:157–162, 2004.

-
- [20] P. Crow, B. Barrass, C. Kendall, M. Hart-Prieto, M. Wright, R. Persad, and N. Stone. The use of Raman spectroscopy to differentiate between different prostatic adenocarcinoma cell lines. *British Journal of Cancer*, 92:2166–2170, 2005.
- [21] S.K. Teh, W. Zheng, K.Y. Ho, M. Teh, K.G. Yeoh, and Z. Huang. Diagnostic potential of near-infrared Raman spectroscopy in the stomach: differentiating dysplasia from normal tissue. *British Journal of Cancer*, 98:457–465, 2008.
- [22] M. Ringner. What is principle component analysis. *Nature*, 26:303–304, 2008.
- [23] N. Stone *et al.* Raman spectroscopy for the classification of epithelial pre-cancers and cancers. *Journal of Raman Spectroscopy*, 33:564–573, 2002.
- [24] A. M. Martinez *et al.* PCA versus LDA. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 23:228–233, 2001.
- [25] I. Notingher, C. Green, C. Dyer, E. Perkins, N. Hopkins, C. Lindsay, and L.L. Hench. Discrimination between ricin and sulphur mustard toxicity *in vitro* using Raman spectroscopy. *Journal of the Royal Society Interface*, 1:79–90, 2004.
- [26] H. Wu, J.V. Volponi, A.E. Oliver, A.N. Parikh, B.A. Simmons, and S. Singh. *In vivo* lipidomics using single-cell Raman spectroscopy. *Proceedings of the National Academy of Sciences*, 108:3809–3814, 2011.
- [27] M. Potcoava, G. Futia, J. Aughenbaugh, I. Schlaepfer, and E. Gibson. Raman and coherent anti-stokes Raman scattering microscopy studies of changes in lipid content and composition in hormone treated breast and prostate cancer cells. *Journal of Biomedical Optics*, 19:1–10, 2014.
- [28] B. Davidson, A.A. Murray, A. Elfick, and N. Spears. Raman micro-spectroscopy can be used to investigate the developmental stage of the mouse oocyte. *PLoS ONE*, 8(7):e67972., 2013.
- [29] J. W. Chan. Monitoring dynamic protein expression in living *E.coli* bacterial cells by laser tweezers Raman spectroscopy. *Cytometry Part A*, 71A:468–474, 2007.
- [30] U. Münchberg, P. Rösch, M Bauer, and J. Popp. Raman spectroscopic identification of single bacterial cells under antibiotic influence. *Analytical and Bioanalytical Chemistry*, 406:3041–3050, 2014.

-
- [31] G. Rusciano, P. Capriglione, G. Pesce, P. Abete, V. Carnovale, and A. Sasso. Raman spectroscopy as a new tool for early detection of bacteria in patients with cystic fibrosis. *Laser Physics Letters*, 10:075603–075610, 2013.
- [32] Z. Huang, H. Lui, D.I. McLean, S. Lam, and H. Zeng. Near-infrared Raman spectroscopy for optical diagnosis of lung cancer. *The International Journal of Cancer*, 107:1047–1052, 2003.
- [33] S. Pahlow, S. Meisel, D. Cialla-May, K. Weber, P. Rosch, and J. Popp. Isolation and identification of bacteria by means of Raman spectroscopy. *Advanced Drug Delivery Reviews*, 89:105–120, 2015.
- [34] S. Nie and S.R. Emory. Probing single molecules and single nanoparticles by surface-enhanced Raman scattering. *Science*, 275:1102–1106, 1997.
- [35] H. Zhou *et al.* Surface-enhanced Raman scattering detection of bacteria on microarrays at single cell levels using silver nanoparticles. *Microchimica Acta*, 182:2259–2266, 2015.
- [36] T.J. Gardner, E.C. Faria, and A. Henderson. Spectral discrimination of live prostate and bladder cancer cell lines using Raman optical tweezers. *Journal of Biomedical Optics*, 13:064004(1–12), 2008.
- [37] G.J. Puppels *et al.* Studying single living cells and chromosomes by confocal Raman microspectroscopy. *Nature*, 347:301–303, 1990.
- [38] Y. Oshima, H. Shinzawa, T. Takenaka, and H. Sato. Spectral discrimination of live prostate and bladder cancer cell lines using Raman optical tweezers. *Journal of Biomedical Optics*, 15:017009(1–8), 2010.
- [39] M. Frei. Cell viability and proliferation. *Sigma-Aldrich: BioFiles*, 6:17–21, 2011.
- [40] W. Strober. Trypan blue exclusion test of cell viability. *Current Protocols in Immunology*, 111:A3, 2015.
- [41] I. Notingher, S. Verrier, H. Romanska, A.E. Bishop, J.M. Polak, and L.L. Hench. In situ characterisation of living cells by Raman spectroscopy. *Spectroscopy*, 16:43–51, 2002.
- [42] H.J. Butler *et al.* Using Raman spectroscopy to characterize biological materials. *Nature Protocols*, 11:664–687, 2016.

-
- [43] HORIBA Scientific. *How to Measure Samples Sensitive to Laser Power*, 2012.
- [44] R.L. Siegel, K.D. Miller, and A. Jemal. Cancer statistics, 2016. *CA: A Cancer Journal for Clinicians*, 66:7–30, 2016.
- [45] L.A. Torre, F. Bray, R.L. Siegel, J. Ferlay, J. Lortet-Tieulent, and A. Jemal. Global cancer statistics, 2012. *CA: A Cancer Journal for Clinicians*, 65:87–108, 2015.
- [46] M. Norgaard, A. Jensen, J.B. Jacobsen, K. Cetin, J.P. Fryzek, and H.T. Sorensen. Skeletal related events, bone metastasis and survival of prostate cancer: a population based cohort study in Denmark (1999 to 2007). *The Journal of Urology*, 184:162–167, 2010.
- [47] V.T. DeVita and T.S Lawrence S.R. Rosenberg *et al.* *Cancer: Principles and Practice of Oncology*. 9th ed. Lippincott Williams and Wilkins, 2011.
- [48] T.A. Stamey, N. Yang, A.R. Hay, J.E. McNeal, F.S. Freiha, and E. Redwine. Prostate-specific antigen as a serum marker for adenocarcinoma of the prostate. *New England Journal of Medicine*, 317:909–916, 1987.
- [49] M.J. Roobol *et al.* Prediction of prostate cancer risk: The role of prostate volume and digital rectal examination in the ERSPC risk calculators. *European Urology*, 61:577–583, 2012.
- [50] F. Rizzi *et al.* A novel gene signature for molecular diagnosis of human prostate cancer by rt-qpcr. *PLoS ONE*, 317:e3617, 2008.
- [51] B. Delahunt *et al.* Gleason grading: past, present and future. *Histopathology*, 60:75–86, 2012.
- [52] D.F. Gleason and G.T. Mellinger. Prediction of prognosis for prostatic adenocarcinoma by combined histological grading and clinical staging. *Journal of Urology*, 111:58–64, 1974.
- [53] J. Gordetsky and J. Epstein. Grading of prostatic adenocarcinoma: current state and prognostic implications. *Diagnostic Pathology*, 11:25, 2016.
- [54] J.I. Epstein, W.C. Allsbrook Jr, M.B. Amin, and L.L. Egevad. ISUP grading committee. The 2005 international society of urological pathology (ISUP) consensus conference on Gleason grading of prostatic carcinoma. *American Journal of Surgical Pathology*, 29:1228–1242, 2005.

-
- [55] V. Gupta, P. Gopinath, M.A. Iqbal, S. Mazurek, K.E. Wellen, and R.N. Bamezai. Interplay between epigenetics and cancer metabolism. *Current Pharmaceutical Design*, 20:1706–1714, 2013.
- [56] A. Vasquez-Martin, R. Colomer, J. Brunet, R. Lupu, and J.A. Menendez. Overexpression of fatty acid synthase gene activates HER1/HER2 tyrosine kinase receptors in human breast epithelial cells. *Cell Proliferation*, 41:59–85, 2008.
- [57] F. Ameer, L. Scandiuzzi, S. Hasnain, H. Kalbacher, and N. Zaidi. *De novo* lipogenesis in health and disease. *Metabolism*, 63:895–902, 2014.
- [58] N. Stone, P. Stavroulaki, C. Kendall, M. Birchall, and H. Barr. Raman spectroscopy for early detection of laryngeal malignancy: preliminary results. *Laryngoscope*, 110:1756–1763, 2000.
- [59] A. Mahadevan-Jansen and R. Richards-Kortum. Raman spectroscopy for the detection of cancers and precancers. *The Journal of Biomedical Optics*, 1:31–70, 1996.
- [60] C.J. Antalis, A. Uchida, K.K. Buhman, and R.A. Siddiqui. Migration of mda-mb-231 breast cancer cells depends on the availability of exogenous lipids and cholesterol esterification. *Clinical and Experimental Metastasis*, 28:733–741, 2011.
- [61] C. Nieva, M. Marro, N. Santana-Codina, S. Rao, D. Petrov, and A. Sierra. The lipid phenotype of breast cancer cells characterised by Raman microspectroscopy: towards a stratification of malignancy. *PLoS One*, 7:1–10, 2012.
- [62] E. Rysman, K. Brusselmans, K. Scheys, L. Timmermans, R. Derua, S. Munck, P.P. Van Veldhoven, D. Waltregny, V.W. Daniels, J. Machiels, F. Vanderhoydonc, K. Smans, E. Waelkens, G. Verhoeven, and J.V. Swinnen. *De novo* lipogenesis protects cancer cells from free radicals and chemotherapeutics by promoting membrane lipid saturation. *Cancer Research*, 70:8117–8126, 2010.
- [63] D.F. Calvisi, C. Wang, C. Ho, S. Ladu, S.A. Lee, S. Mattu, G. Destefanis, S. Delogu, A. Zimmermann, J. Ericsson, S. Brozzetti, T. Staniscia, X. Chen, F. Dombrowski, and M. Evert. Increased lipogenesis, induced by AKT-mTORC1-RPS6 signaling, promotes development of human hepatocellular carcinoma. *Gastroenterology*, 140:1071–1083, 2011.

-
- [64] J. Li *et al.* Surface enhanced Raman scattering detection of cancer biomarkers with bifunctional nanocomposite probes. *Analytical Chemistry*, 87:10698–10702, 2015.
- [65] Y.M. Fu and G.G. Meadows. Specific amino acid dependency regulates the cellular behavior of melanoma. *American Society for Nutrition*, 137:15915–15965, 2007.
- [66] N. Stone *et al.* Raman spectroscopy for identification of epithelial cancers. *Faraday Discussion*, 126:141–157, 2004.
- [67] A. Taleb, J. Diamond, J.J. McGarvey, J.R. Beattie, C. Toland, and P.W. Hamilton. Raman microscopy for the chemometric analysis of tumor cells. *The Journal of Physical Chemistry. B.*, 110:19625–19631, 2006.
- [68] G. Shetty, C. Kendall, N. Shepherd, N. Stone, and H. Barr. Raman spectroscopy elucidation of biochemical changes in carcinogenesis of oesophagus. *British Journal of Cancer*, 94:1460–1464, 2006.
- [69] M. Rousset, A. Zweibaum, and J. Fogh. Presence of glycogen and growth-related variations in 58 cultured human tumor cell lines of various tissue origins. *Cancer Research*, 41:1165–1170, 1981.
- [70] M. Rousset, G. Chevalier, J.P. Rousset, S. Robine-Leon, E. Dussaulx, and A. Zweibaum. Raman spectroscopy: elucidation of biochemical changes in carcinogenesis of oesophagus. *British Journal of Cancer*, 94:1460–1464, 2006.
- [71] J.S. Horoszewicz, S.S. Leong, E. Kawinski, J.P. Karr, H. Rosenthal, T.M. Chu, E.A. Mirand, and G.P. Murphy. LNCaP model of human prostatic carcinoma. *Cancer Research*, 43:1809–1818, 1983.
- [72] P. Berthon, O. Cussenot, L. Hopwood, A. Leduc, and N. Maitland. Functional expression of sv40 in normal human prostatic epithelial and fibroblastic cells - differentiation pattern of nontumorigenic cell-lines. *International Journal of Oncology*, 6:333–343, 1995.
- [73] D.L. Lin, C.P. Tarnowski, J. Zhang, J. Dai, E. Rohn, A.H. Patel, M.D. Morris, and E.T. Keller. Bone metastatic LNCaP-derivative c4 2b prostate cancer cell line mineralizes *in vitro*. *Prostate*, 47:212–221, 2001.
- [74] N.S. Blow. Right cell, wrong cell. *Biotechniques*, 51:75, 2011.

-
- [75] R.M. Nardone. Curbing rampant cross-contamination and misidentification of cell lines. *Biotechniques*, 45:221–227, 2008.
- [76] S.D. Bentley, K.F. Chater, A.M. Cerdeno-Tarranga, G.L. Challis, and N.R. Thompson *et al.* Complete genome sequence of the model actinomycete *Streptomyces coelicolor* a3(2). *Nature*, 417:141–147, 2002.
- [77] S.S. Gurcha, A.R. Baulard, L. Kremer, C. Locht, D.B. Moody, and W. Muhlecker. Ppm1, a novel polyprenol monophosphomannose synthase from *Mycobacterium tuberculosis*. *Biochemical Journal*, 365:441–450, 2002.
- [78] D.A. Cowlshaw and M.C. Smith. Glycosylation of a *Streptomyces coelicolor* a3(2) cell envelope protein is required for infection by bacteriophage phi c31. *Molecular Microbiology*, 41:601–610, 2001.
- [79] D.A. Cowlshaw and M.C. Smith. A gene encoding a homologue of dolichol phosphate-b-d mannose synthase is required for infection of *Streptomyces coelicolor* a3(2) by phage fc31. *Journal of Bacteriology*, 184:6081–6083, 2002.
- [80] A. Walter, S. Kuhri, M. Reinicke, T. Bocklitz, W. Schumacher, P. Rsch, D. Merten, G. Bchel, E. Kotheb, and J. Popp. Raman spectroscopic detection of nickel impact on single *Streptomyces* cells - possible bioindicators for heavy metal contamination. *Journal of Raman Spectroscopy*, 43:1058–1064, 2012.
- [81] K. Ramser, W. Wenseleers, S. Dewilde, S. Van Doorslaer, L. Moens, and D. Hanstorp. Microresonance Raman study of optically trapped *Escherichia coli* cells overexpressing human neuroglobin. *Journal of Biomedical Optics*, 12:1–9, 2007.
- [82] L. Ashton, K. Lau, C.L. Winder, and R. Goodacre. Lighting up the future of microbial identification. *Future Microbiology*, 6:991–997, 2011.
- [83] A. Walter *et al.* From bulk to single-cell classification of the filamentous growing *Streptomyces* bacteria by means of Raman spectroscopy. *Applied Spectroscopy*, 65:1116–1125, 2011.

-
- [84] S. Stökel, J. Kirchoff, U. Neugebauer, P. Rosch, and J. Popp. The application of Raman spectroscopy for the detection and identification of microorganisms. *Journal of Raman Spectroscopy*, 47:89109, 2016.
- [85] F.M. Frame, D. Pellacani, A.T. Collins, and N.J. Maitland. Harvesting human prostate tissue material and culturing primary prostate epithelial cells. *Methods in Molecular Biology*, 1443:181–201, 2016.
- [86] F.M. Frame, H. Savoie, F. Bryden, F. Giuntini, V.M. Mann, M.S. Simms, R.W. Boyle, and N.J. Maitland. Mechanisms of growth inhibition of primary prostate epithelial cells following gamma irradiation or photodynamic therapy include senescence, necrosis, and autophagy, but not apoptosis. *Cancer Medicine*, 5:61–73, 2016.
- [87] F.M. Frame, D. Pellacani, A.T. Collins, M.S. Simms, V.M. Mann, G.D.D. Jones, M. Meuth, R.G. Bristow, and N.J. Maitland. HDAC inhibitor confers radiosensitivity to prostate stem-like cells. *British Journal of Cancer*, 109:1–11, 2013.
- [88] N.J. Maitland, C.A. Macintosh, J. Hall, M. Sharrard, G. Quinn, and S. Lang. *In vitro* models to study cellular differentiation and function in human prostate cancers. *Radiation Research*, 155:133–142, 2001.
- [89] C. Matthaus, B. Bird, M. Miljkovic, T. Chernenko, M. Romeo, and M. Diem. *Methods in Cell Biology - Biophysical Tools for Biologists*, volume 89. Academic Press Elsevier, 2008.
- [90] T. Kieser, M.J. Bibb, M.J. Buttner, K.F. Chater, and D.A. Hopwood. *Practical Streptomycin Genetics*. John Innes Foundation, 2000.
- [91] HORIBA Scientific. *Labspec 5: User Manual*, 2010.
- [92] P. Candeloro, E. Grande, R. Raimondo, D.D. Mascolo, F. Gentile, M.L. Coluccio, G. Perozziello, N. Malara, M. Francardi, and E.D. Frabizio. Raman database of amino acid solutions: A critical study of extended multiplicative signal correction. *Analyst*, 138:7331–7340, 2012.
- [93] H.F. Kaiser. The application of electronic computers to factor analysis. *Educational and Psychological Measurement*, 20:141–151, 1960.

-
- [94] K. Chen *et al.* Diagnosis of colorectal cancer using raman spectroscopy of laser-trapped single living epithelial cells. *Optics Letters*, 31:2015–2017, 2006.
- [95] S. Manago, C. Valente, P. Mirabelli, D. Circolo, F. Basile, D. Corda, and A.C. De Luca. A reliable Raman-spectroscopy based approach for diagnosis, classification and follow-up of b-cell acute lymphoblastic leukemia. *Scientific Reports*, 6:24821, 2016.
- [96] J.W. Chan *et al.* Label-free separation of human embryonic stem cells and their cardiac derivatives using Raman spectroscopy. *Analytical Chemistry*, 81:1324–1331, 2009.
- [97] J.W. Chan, D.S. Taylor, T. Zwerdling, S.M. Lane, K. Lhara, and T. Huser. Micro-Raman spectroscopy detects individual neoplastic and normal hematopoietic cells. *Biophysical Journal*, 90:648–656, 2006.
- [98] R. C. Lord and N. T. Yu. Laser-excited raman spectroscopy of biomolecules: I. native lysozyme and its constituent amino acids. *Journal of Molecular Biology*, 50:509–524, 1970.
- [99] J. Binoy *et al.* NIR-FT Raman and FT-IR spectral studies and ab initio calculations of the anti-cancer drug combretastatin-A4. *Journal of Raman Spectroscopy*, 35:939–946, 2004.
- [100] W.T. Cheng *et al.* Micro-Raman spectroscopy used to identify and grade human skin pilomatricoma. *Microscopy Research and Techniques*, 68:75–79, 2005.
- [101] M. Gniadecka *et al.* Diagnosis of basal cell carcinoma by Raman spectroscopy. *Journal of Raman Spectroscopy*, 28:125–129, 1997.
- [102] C.J. Frank *et al.* Raman spectroscopy of normal and diseased human breast tissues. *Analytical Chemistry*, 67:777–783, 1995.
- [103] C. Krafft *et al.* Near infrared Raman spectra of human brain lipids. *Spectrochimica Acta, Part A*, 61:1529–1535, 2005.
- [104] A.J. Ruiz-Chica *et al.* Characterization by Raman spectroscopy of conformational changes on guanine cytosine and adenine-thymine oligonucleotides induced by aminoxy analogues of spermidine. *Journal of Raman Spectroscopy*, 35:93–100, 2004.

-
- [105] R. Malini *et al.* Discrimination of normal, inflammatory, premalignant, and malignant oral tissue: A Raman spectroscopy study. *Biopolymers*, 81:179–193, 2006.
- [106] D. Naumann. Infrared and NIR Raman spectroscopy in medical microbiology. *Proc. SPIE*, 3257:95–100, 1998.
- [107] R.K. Dukor. Vibrational spectroscopy in the detection of cancer. *Biomedical Applications*, 5:3335–3359, 2002.
- [108] L. Chiriboga *et al.* Infrared spectroscopy of human tissue. i. differentiation and maturation of epithelial cells in the human cervix. *Biospectroscopy*, 4:47–53, 1998.
- [109] S.M. Ronen *et al.* NMR studies of the lipid metabolism of T47D human breast cancer spheroids. *FEBS Letters*, 266:147–149, 1990.
- [110] E.O. Faolain *et al.* A study examining the effects of tissue processing on human tissue sections using vibrational spectroscopy. *Vibrational Spectroscopy*, 38:121–127, 2005.
- [111] R.J. Lakshimi *et al.* Tissue Raman spectroscopy for the study of radiation damage: Brain irradiation of mice. *Radiation Research*, 157:175–182, 2002.
- [112] S. Kolijenovic *et al.* Detection of meningioma in dura mater by Raman spectroscopy. *Analytical Chemistry*, 77:7958–7965, 2005.
- [113] K. Kachrimanis *et al.* Quantitative analysis of paracetamol polymorphs in powder mixtures by FT-Raman spectroscopy and PLS regression. *Journal of Pharmaceutical and Biomedical Analysis*, 43:407–412, 2007.
- [114] D. Cunningham and Z. You. *In vitro* and *in vivo* model systems used in prostate cancer research. *Journal of Biological Methods*, 2:e17, 2015.
- [115] A. Capes-Davies *et al.* Check your cultures! a list of cross-contaminated or misidentified cell lines. *International Journal of Cancer*, 127:1–8, 2010.
- [116] C. Krafft *et al.* Advances in optical biopsy - correlation of malignancy and cell density of primary brain tumors using Raman microspectroscopic imaging. *Analyst*, 137:5533–5537, 2012.

- [117] A.J.F. Griffiths, W.M. Gelbart, and J.H. Miller *et al.* *Modern Genetic Analysis*. W.H.Freeman, 1999.
- [118] D. Friedrich, C. Jin, Y. Zhang, C. Demin, L. Yuan, L. Berynsky, S. Biesterfeld, T. Aach, and A. Böcking. Identification of prostate cancer cell nuclei for DNA-grading of malignancy. *Bildverarbeitung für die Medizin*, pages 334–339, 2012.
- [119] M.W.G. De Bolster. Glossary of terms used in bioinorganic chemistry. *Pure and Applied Chemistry*, 6:1251–1303, 1997.