Biomolecular Stratification of Standard Prostate Cell Lines and Primary Prostate Cell Cultures Using Confocal Raman Spectroscopy

MARCUS DUNCAN CAMERON

Doctor of Philosophy

The University of York

Physics

December 2021

Abstract

At present, there is no definitive test used for the diagnosis of prostate cancer. It is therefore essential that prostate cancer is not only fully characterised to aid in the detection of disease, but also to be able to monitor its progression. Raman spectroscopy is a non-destructive label-free method that can be used for the biomolecular characterisation of prostate cancer cells. This thesis aims to characterise and stratify standard prostate cell lines and primary cell cultures. The methods selected to perform both characterisation and stratification included principal component analysis (PCA), linear discriminant analysis (LDA) and peak intensity ratios (PIRs). Within this work, analysis of standard cell lines was performed on air-dried and live cells, determining PIR markers that could be used to discriminate between the live and dried states, as well as disease states, using the relative ratios of DNA/RNA, proteins and lipids. The set of live standard cell lines, representative of a comprehensive set of normal to diseased cells, was further able to be separated into specific disease states using PIR ratios. Additionally, a PCA-LDA model was also used to show that Raman spectroscopy could be used to classify primary prostate cell cultures. The PIR markers identified to discriminate the disease states of standard cell lines were applied to the primary cell cultures. From the results of this comparison, it was determined that standard cell lines are not able to model the complex normal and malignant heterogeneity associated with primary cell cultures.

ACKNOWLEDGMENTS

First and foremost, I would like to thank my supervisor Dr. Yvette Hancock. You have invested a lot of your time, patience and resources to make sure I progressed through every step throughout my PhD. I am also very grateful to Dr. Fiona Frame for your invaluable support and knowledge in the labs and in regards to the biological elements of the PhD.

I would also like to thank my family, specifically my mother Hazel Cameron and grandmother Dorothy Cameron for their financial support throughout my undergraduate and PhD studies. I would also like to thank my great aunt Margaret Wilson and great uncle Rodney Bambrick for their financial assistance while completing my studies.

I am especially grateful for all my friends support, both academically and personally. I would like to especially mention Sophie, Felix and Patrick for your motivation throughout my PhD.

AUTHOR'S DECLARATION

I declare that the work in this thesis was carried out in accordance with the requirements of the University's Regulations and Code of Practice for Research Degree Programmes and that it has not been submitted for any other academic award. Except where indicated by specific reference in the text, the work is the candidate's own work. Work done in collaboration with, or with the assistance of, others, is indicated as such. Any views expressed in the thesis are those of the author.

SIGNED: .....MARCUS D. CAMERON..... DATE: ...31/12/2021...

'I am the master of my failure... If I never fail how will I ever learn.'

Chandrasekhara Venkata Raman

CONTENTS

Page

Abstract **i**

Acknowledgments **ii**

Author's declaration **iii**

List of Tables **viii**

List of Figures **xi**

1 Introduction **1**

1.1 Overview 1

1.2 The prostate 8

1.2.1 Morphology of the prostate gland 8

1.3 Aetiology of prostate conditions 10

1.3.1 Benign prostate hyperplasia 10

1.3.2 Prostate cancer 10

1.3.3 Other conditions 11

1.4 Epidemiology of prostate cancer 11

1.5 Current prostate cancer models 12

1.5.1 Standard prostate epithelial cell lines 12

1.5.2 Primary cell cultures 13

1.5.3 Wider models 13

1.6 Raman spectroscopy 14

2 Theory **19**

2.1 Theory of Raman spectroscopy . . . 19
2.1.1 Overview . . . 19
2.1.2 Classical theory . . . 20
2.1.3 Quantum theory . . . 24
2.2 Multivariate analysis . . . 26
2.2.1 Pre-treatment of spectra for multivariate analysis . . . 27
2.2.2 Principal component analysis and linear discriminant analysis 28

3 Materials and Methods 32
3.1 Standard cell lines . . . 32
3.1.1 PNT2-C2 (normal prostate cell line) . . . 32
3.1.2 BPH-1 (benign hyperplasia prostate cell line) . . . 32
3.1.3 LNCaP (metastatic cell line) . . . 33
3.1.4 DU145 (metastatic cell line) . . . 33
3.1.5 PC3 (metastatic cell line) . . . 34
3.1.6 P4E6 (localised cancer cell line) . . . 34
3.1.7 Reasons for selecting these specific cell lines . . . 34
3.2 Primary cell cultures . . . 35
3.2.1 H646/17 . . . 36
3.2.2 H722/18 . . . 36
3.2.3 H794/18 . . . 36
3.3 Gleason grading . . . 37
3.4 Cell culturing methods and Raman sample preparation . . . 39
3.4.1 Culturing of the standard cell lines . . . 40
3.4.2 Culture of primary prostate epithelial cells derived from patient tissue . . . 40
3.4.3 Passaging of standard cell lines and primary cell cultures . 41
3.4.4 Counting cells . . . 41
3.4.5 Preparing cells for analysis using Raman spectroscopy . . . 42
3.5 Raman spectroscopy measurements . . . 43
3.5.1 Raman spectrometer set-up . . . 43
3.5.2 Parameters used in Raman spectra acquisition . . . 46

3.5.3 Confocal Raman spectroscopy and laser spot size 47
3.5.4 Assessing viability of cells 48
3.5.5 Acquisition of Raman spectra data on prostate cells 49
3.5.6 Removal of outliers from data set 49
3.6 Raman spectral analysis 50
3.6.1 Statistical convergence testing 50
3.6.2 Univariate analysis 51
3.6.3 PCA and LDA analysis 53

4 Live versus Dried, Study of PNT2-C2 55
4.1 Raman Spectroscopy: live versus dried, study of PNT2-C2 56
4.2 Convergence tests and removal of outliers 60
4.3 Raman spectroscopy of live and dried version of PNT2-C2 66
4.4 Principal component analysis and loading analysis 83
4.5 Comparison of peak intensity ratios 96
4.6 Conclusion 113

5 Raman spectroscopy of live, prostate standard cell lines 116
5.1 Overview of chapter 116
5.1.1 Statistical convergence testing of standard prostate cell lines 117
5.1.2 Raman spectroscopy of standard cell lines 124
5.2 BPH-1 137
5.2.1 PNT2-C2 vs BPH-1 mean profile subtraction 138
5.2.2 PNT2-C2 vs BPH-1 PCA comparison 141
5.2.3 Peak intensity ratio analysis PNT2-C2 vs BPH-1 150
5.3 LNCaP 165
5.3.1 PNT2-C2 vs LNCaP mean profile subtraction 167
5.3.2 PNT2-C2 vs LNCaP PCA Comparison 171
5.3.3 Peak intensity ratio analysis PNT2-C2 vs LNCaP 180
5.4 DU145 189
5.4.1 PNT2-C2 vs DU145 mean profile subtraction 191
5.4.2 PNT2-C2 vs DU145 PCA analysis 194
5.4.3 Peak intensity ratio analysis PNT2-C2 vs DU145 204

5.5 PC3 . . . 214
5.5.1 PNT2-C2 vs PC3 mean profile subtraction . . . 215
5.5.2 PNT2-C2 vs PC3 PCA comparison . . . 219
5.5.3 Peak intensity ratio analysis PNT2-C2 vs PC3 . . . 226
5.6 P4E6 . . . 234
5.6.1 PNT2-C2 vs P4E6 mean profile subtraction . . . 234
5.6.2 PNT2-C2 vs P4E6 PCA comparison . . . 237
5.6.3 Peak intensity ratio analysis PNT2-C2 vs P4E6 . . . 247
5.7 The use of Raman spectroscopy to differentiate between standard cell lines . . . 252
5.7.1 PCA/LDA analysis of prostate standard cell lines . . . 252
5.7.2 Identification of biomarkers correlating to disease state . . . 264
5.8 Conclusion . . . 271

6 Raman spectroscopy on patient primary cell cultures 275
6.1 Statistical convergence testing of patient primary cell cultures . . . 276
6.2 Raman spectroscopy of patient primary cell cultures . . . 285
6.3 PCA-LDA for the classification of patient primary cell cultures . . . 296
6.4 Comparison of patient primary cell cultures to previously identified biomarkers relating to disease state . . . 308
6.5 Conclusion . . . 318

7 Conclusion 321

A Appendix 326
A.1 List of Abbreviations . . . 326

B Raman Peak Positions 328

C Peak Intensity Ratio Calculation Tables 332

Bibliography 342

List of Tables

Table **Page**

3.1 Parameters used on the HORIBA XploRA Raman spectrometer for whilst taking Raman spectra of air-dried and live cells. 46

4.1 Convergence data for PNT2-C2 live and dried. Key for table: PNT2-C2 dried AVG FP 80 indicates that the convergence as per Methods Section 3.6.1 is seen to first occur at approximately 80 spectra for the PNT2-C2 profile average. Furthermore, 2SD denotes twice the normal standard deviation and SE denotes the standard error. 65

4.2 Biomolecule peak-assignments for PNT2-C2 live and dried. The average peak position has been between live and dried (Table B.1 in Appendix), general and detailed assignments were obtained from the literature. The symbol $\otimes$ indicates a peak was not present in PNT2-C2 dried. For the peak position a standard error of ± 3 cm^{-1} as an uncertainty is attached to all positions. 67

4.3 The peak intensity ratios as proposed by Nieva *et al.* (2012) and Potcoava *et al.* (2014) to measure lipid composition in the cells [64, 140]. The I1656/1448 ratio calculated between PNT2-C2 live and dried implies there is a statistical difference in the mass unsaturation scissor mode. However, there is no statistically significant separation between the ratio I1656/1306 which is attributed to the mass unsaturation twisting mode. Similarly, the ratio CH_2/CH_3 anti-symmetric stretching calculated using the peaks I2930/2959 shows no statistically significant separation. However, there is further statistically significant separation found using the ratio I3015/2850, a difference in the CH stretching and CH_2 symmetric stretching in the high wavenumber. . . 109

5.1 The convergence data for live prostate standard cell line is displayed. Key for table: PNT2-C2 2SD FP 80 represents that twice the standard deviation in PNT2-C2 converged at approximately 80 spectra in the fingerprint profile. 123
5.2 Biomolecular peak-assignments for standard cell lines. The average peak position is shown (calculated in Table B.2 in the appendix), general and detailed assignments) obtained from the literature. For the peak position a standard error of ± 3 cm^{-1} as an uncertainty is attached to all positions. 128
5.3 PIR table for PNT2-C2 vs LNCaP 186
5.4 PCA-LDA leave-one-out cross-validation is performed on standard cell lines in the fingerprint region. This used a total of 59 principal components, accounting for 78.9% total population variance. The corresponding results are shown in the table below. 257
5.5 PCA-LDA cross-validation for standard prostate cell lines in the high wavenumber region 263

6.1 The convergence data for primary prostate cell cultures is displayed. Key for table: H646/17R 2SD FP 80 represents that twice the standard deviation in H646/17R converged at approximately 80 spectra in the fingerprint profile. 284
6.2 Biomolecular peak-assignments for primary cell cultures. The average peak position is shown (calculated in Table B.3 in the appendix), general and detailed assignments obtained from the literature. For the peak position, a standard error of ± 3 cm^{-1} as an uncertainty is attached to all positions. 289
6.3 PCA-LDA cross-validation for primary patient cell cultures in the fingerprint region 301
6.4 PCA-LDA leave-one-out cross-validation is performed on primary patient cultures in the high wavenumber region. This used a total of 60 principal components accounting for 86.7% total population variance. The corresponding confusion matrix results are shown in the table below:306

B.1 This table contains the peak positions found for PNT2-C2 live and PNT2-C2 dried as well as the corresponding average. 329
B.2 This table contains the peak positions found for PNT2-C2 live, BPH-1, P4E6, LNCaP, DU145, PC3 as well as the corresponding average. . . . 330
B.3 This table contains the peak positions found for H646/17L, H646/17R, H722/18L, H722/18R, H794/18L, H794/18R as well as the corresponding average. 331

C.1 PIR table for PNT2-C2 live vs dried 333
C.2 PIR table for PNT2-C2 vs BPH-1 334
C.3 PIR table for PNT2-C2 vs LNCaP 335
C.4 PIR table for PNT2-C2 vs DU145 336
C.5 PIR table for PNT2-C2 vs DU145 337
C.6 PIR table for PNT2-C2 vs P4E6 338
C.7 PIR table for H646/17R vs H646/17L 339
C.8 PIR table for H722/18R vs H722/18L 340
C.9 PIR table for H794/18R vs H794/18L 341

List of Figures

Figure **Page**

1.1 A table explaining heterogeneity in prostate disease 6
1.2 A cross section of the male reproductive organs 8
1.3 Zonal diagram of prostate gland 9
1.4 A diagram depicting a cross section of the epithelial layer of the prostate gland and stroma 11
1.5 A diagram depicting the possible photon energy state transitions for both vibrational and virtual states for Rayleigh Scattering and Raman scattering. 15

3.1 The Gleason grading system 37
3.2 Image of Horiba Xplora Raman spectrometer 44
3.3 Simple schematic of Raman instrument 45
3.4 Diagram of microscope objective setup 45
3.5 Trypan blue assay test to confirm cell viability 48
3.6 A plot showing the fitting regions being mapped onto a normal PNT2-C2 cell line fingerprint region. 52
3.7 This image shows the fitting of region for the live standard cell line PNT2-C2. Specifically, the average of live the PNT2-C2 cell line. 53

4.1 Comparison of Raman signal from cytoplasm and nucleus 61
4.2 PNT2-C2 Live Convergence Tests 63
4.3 PNT2-C2 Dried Convergence Tests 64
4.4 The average spectrum from PNT2-C2 live and dried are examined in the fingerprint region 76
4.5 A mean profile subtraction is shown for PNT2-C2 live vs PNT2-C2 dried in the fingerprint region 79
4.6 The high wavenumber region profiles of PNT2-C2 dried and live are displayed 81
4.7 PNT2-C2 dried and live mean profile subtraction in the high wavenumber region 82
4.8 PNT2-C2 live vs dried fingerprint PCA comparison 1 84
4.9 PNT2-C2 live vs dried fingerprint PCA comparison 2 85
4.10 PNT2-C2 live vs dried fingerprint PCA comparison 3 85
4.11 PNT2-C2 live vs dried fingerprint PCA comparison 4 86
4.12 PNT2-C2 live vs dried fingerprint loading 1 87
4.13 PNT2-C2 live vs dried fingerprint loading 2 88
4.14 PNT2-C2 live vs dried high wavenumber PCA comparison 1 90
4.15 PNT2-C2 live vs dried high wavenumber PCA comparison 2 91
4.16 PNT2-C2 live vs dried high wavenumber PCA comparison 3 91
4.17 PNT2-C2 live vs dried high wavenumber loading 2 92
4.18 PNT2-C2 live vs dried high wavenumber loading 3 93

4.19 PNT2-C2 live vs dried high wavenumber PCA comparison 4 94
4.20 PNT2-C2 live vs dried high wavenumber loading 4 95
4.21 PNT2-C2 live vs dried 745 cm^{-1} PIR panel 101
4.22 PNT2-C2 live vs dried 1234 cm^{-1} PIR panel 104
4.23 PNT2-C2 live vs dried 1256 cm^{-1} PIR panel 105
4.24 PNT2-C2 live vs dried 1681 cm^{-1} PIR panel 107
4.25 PNT2-C2 live vs dried 3015 cm^{-1} PIR panel 110
4.26 PNT2-C2 dried image at x100 magnification 112

5.1 BPH-1 live convergence tests 118
5.2 LNCaP live convergence tests 119
5.3 PC3 live convergence tests 120
5.4 DU145 live convergence tests 121
5.5 P4E6 live convergence tests 122
5.6 The fingerprint region of the live standard cell lines PNT2-C2, BPH-1, LNCaP, PC3, DU145 and P4E6 are displayed on the same graph 125
5.7 The high wavenumber region of the live standard cell lines PNT2-C2, BPH-1, LNCaP, PC3, DU145 and P4E6 are displayed on the same graph 126
5.8 The mean profile subtraction of BPH-1 from PNT2-C2 is shown for the fingerprint region 139
5.9 The mean profile subtraction of BPH-1 from PNT2-C2 is shown for the high wavenumber region 140
5.10 PNT2-C2 vs BPH-1 fingerprint PCA comparison 1 142
5.11 PNT2-C2 vs BPH-1 fingerprint PCA comparison 2 142
5.12 PNT2-C2 vs BPH-1 fingerprint PCA comparison 3 143
5.13 PNT2-C2 vs BPH-1 fingerprint PCA loading 2 144
5.14 PNT2-C2 vs BPH-1 fingerprint PCA loading 3 145
5.15 PNT2-C2 vs BPH-1 high wavenumber PCA comparison 1 146
5.16 PNT2-C2 vs BPH-1 high wavenumber PCA comparison 2 147
5.17 PNT2-C2 vs BPH-1 high wavenumber PCA comparison 3 147
5.18 PNT2-C2 vs BPH-1 high wavenumber PCA loading 2 148
5.19 PNT2-C2 vs BPH-1 high wavenumber PCA loading 3 149

5.20 PNT2-C2 vs BPH-1 1448 cm^{-1} PIR panel 152
5.21 PNT2-C2 vs BPH-1 699 cm^{-1} PIR panel 154
5.22 PNT2-C2 vs BPH-1 715 cm^{-1} PIR panel 155
5.23 PNT2-C2 vs BPH-1 673 cm^{-1} PIR panel 157
5.24 PNT2-C2 vs BPH-1 1234 cm^{-1} PIR panel 159
5.25 PNT2-C2 vs BPH-1 1554 cm^{-1} PIR panel 161
5.26 PNT2-C2 vs BPH-1 2934 cm^{-1} PIR panel 163
5.27 The mean profile subtraction of LNCaP from PNT2-C2 in the fingerprint region 168
5.28 The mean profile subtraction of LNCaP from PNT2-C2 in the high wavenumber region 170
5.29 PNT2-C2 vs LNCaP fingerprint PCA comparison 1 172
5.30 PNT2-C2 vs LNCaP fingerprint PCA comparison 2 172
5.31 PNT2-C2 vs LNCaP fingerprint PCA comparison 3 173
5.32 PNT2-C2 vs LNCaP fingerprint PCA loading 1 174
5.33 PNT2-C2 vs LNCaP fingerprint PCA loading 2 174
5.34 PNT2-C2 vs LNCaP fingerprint PCA loading 3 175
5.35 PNT2-C2 vs LNCaP high wavenumber PCA comparison 1 177
5.36 PNT2-C2 vs LNCaP high wavenumber PCA comparison 2 177
5.37 PNT2-C2 vs LNCaP high wavenumber PCA comparison 3 178
5.38 PNT2-C2 vs LNCaP high wavenumber PCA loading 1 179
5.39 PNT2-C2 vs LNCaP high wavenumber PCA loading 3 179
5.40 PNT2-C2 vs LNCaP 976 cm^{-1} PIR panel 182
5.41 PNT2-C2 vs LNCaP 1554 cm^{-1} PIR panel 184
5.42 PNT2-C2 vs LNCaP 3015 cm^{-1} PIR panel 187
5.43 The mean profile subtraction of DU145 from PNT2-C2 is shown for the fingerprint region 191
5.44 The mean profile subtraction of DU145 from PNT2-C2 is shown for the high wavenumber region 193
5.45 PNT2-C2 vs DU145 PCA fingerprint comparison 1 195
5.46 PNT2-C2 vs DU145 PCA fingerprint comparison 2 196
5.47 PNT2-C2 vs DU145 PCA fingerprint comparison 3 196

5.48 PNT2-C2 vs DU145 PCA fingerprint loading 1 197
5.49 PNT2-C2 vs DU145 PCA fingerprint loading 2 198
5.50 PNT2-C2 vs DU145 PCA fingerprint loading 3 199
5.51 PNT2-C2 vs DU145 PCA high wavenumber comparison 1 200
5.52 PNT2-C2 vs DU145 PCA high wavenumber comparison 2 201
5.53 PNT2-C2 vs DU145 PCA high wavenumber comparison 3 201
5.54 PNT2-C2 vs DU145 PCA high wavenumber loading 2 202
5.55 PNT2-C2 vs DU145 PCA high wavenumber loading 4 203
5.56 PNT2-C2 vs DU145 619 cm^{-1} PIR panel 206
5.57 PNT2-C2 vs DU145 641 cm^{-1} PIR panel 207
5.58 PNT2-C2 vs DU145 825 cm^{-1} PIR panel 209
5.59 PNT2-C2 vs DU145 1061 cm^{-1} PIR panel 210
5.60 PNT2-C2 vs DU145 3015 cm^{-1} PIR panel 212
5.61 The mean profile subtraction of PC3 from PNT2-C2 is shown for the fingerprint region 216
5.62 The mean profile subtraction of PC3 from PNT2-C2 is shown for the high wavenumber region 218
5.63 PNT2-C2 vs PC3 fingerprint comparison 1 219
5.64 PNT2-C2 vs PC3 fingerprint comparison 2 220
5.65 PNT2-C2 vs PC3 fingerprint comparison 3 220
5.66 PNT2-C2 vs PC3 fingerprint loading 1 221
5.67 PNT2-C2 vs PC3 fingerprint loading 4 222
5.68 PNT2-C2 vs PC3 high wavenumber comparison 1 223
5.69 PNT2-C2 vs PC3 high wavenumber comparison 2 224
5.70 PNT2-C2 vs PC3 high wavenumber comparison 4 224
5.71 PNT2-C2 vs PC3 high wavenumber loading 2 225
5.72 PNT2-C2 vs PC3 high wavenumber loading 4 225
5.73 PNT2-C2 vs PC3 745 cm^{-1} PIR panel 228
5.74 PNT2-C2 vs PC3 1254 cm^{-1} PIR panel 230
5.75 PNT2-C2 vs PC3 2934 cm^{-1} PIR panel 232
5.76 The mean profile subtraction of P4E6 from PNT2-C2 is shown for the fingerprint region 235

5.77 The mean profile subtraction of P4E6 from PNT2-C2 is shown for the high wavenumber region 236
5.78 PNT2-C2 vs P4E6 fingerprint PCA comparison 1 238
5.79 PNT2-C2 vs P4E6 fingerprint PCA comparison 2 238
5.80 PNT2-C2 vs P4E6 fingerprint PCA comparison 3 239
5.81 PNT2-C2 vs P4E6 fingerprint PCA comparison 4 239
5.82 PNT2-C2 vs P4E6 fingerprint PCA loading 1 240
5.83 PNT2-C2 vs P4E6 fingerprint PCA loading 4 241
5.84 PNT2-C2 vs P4E6 high wavenumber PCA comparison 1 242
5.85 PNT2-C2 vs P4E6 high wavenumber PCA comparison 2 243
5.86 PNT2-C2 vs P4E6 high wavenumber PCA comparison 3 243
5.87 PNT2-C2 vs P4E6 high wavenumber PCA comparison 4 244
5.88 PNT2-C2 vs P4E6 high wavenumber PCA loading 1 245
5.89 Loading 2 of the high wavenumber region comparison of PNT2-C2 Live vs P4E6 is displayed, capturing 12.5% of the variance between samples. Key peaks attributed to biological assignments are found at 2942 cm^{-1} (lipids and proteins) , 3015 cm^{-1} (lipids) and 3065 cm^{-1} (lipids). 245
5.90 PNT2-C2 vs P4E6 high wavenumber PCA loading 3 246
5.91 PNT2-C2 vs P4E6 high wavenumber PCA loading 4 246
5.92 PNT2-C2 vs P4E6 715 cm^{-1} PIR panel 248
5.93 PNT2-C2 vs P4E6 745 cm^{-1} PIR panel 250
5.94 PNT2-C2 vs P4E6 3015 cm^{-1} PIR panel 251
5.95 PCA and LDA of the fingerprint region of standard cell lines 254
5.96 PCA loading 1 in the fingerprint region from prostate standard cell line comparison 255
5.97 PCA loading 2 in the fingerprint region from prostate standard cell line comparison 256
5.98 PCA loading 3 in the fingerprint region from prostate standard cell line comparison 256
5.99 PCA and LDA analysis of the high wavenumber region of standard cell lines in the fingerprint region 260

5.100PCA loading 1 in the high wavenumber region from prostate standard cell line comparison . . . 261
5.101PCA loading 2 in the high wavenumber region from prostate standard cell line comparison . . . 262
5.102PCA loading 3 in the high wavenumber region from prostate standard cell line comparison . . . 262
5.103Key peak intensity ratios from PCA analysis of prostate standard cell lines. . . . 266
5.104Key peak intensity ratios from PCA analysis of prostate standard cell lines part 2. . . . 267

6.1 H646/17R Primary Cell Culture Convergence Tests . . . 278
6.2 H646/17L Primary Cell Culture Convergence Tests . . . 279
6.3 H722/18R Primary Cell Culture Convergence Tests . . . 280
6.4 H722/18L Primary Cell Culture Convergence Tests . . . 281
6.5 H794/18R Primary Cell Culture Convergence Tests . . . 282
6.6 H794/18L Primary Cell Culture Convergence Tests . . . 283
6.7 All Raman profiles for the fingerprint region of the patient primary cell cultures are shown on the same graph . . . 286
6.8 All Raman profiles for the high wavenumber region of the patient primary cell cultures are shown on the same graph . . . 287
6.9 PCA and LDA of the fingerprint region of patient primary cell cultures 298
6.10 PCA loading 1 in the fingerprint region from patient primary cell culture299
6.11 PCA loading 2 in the fingerprint region from patient primary cell culture300
6.12 PCA loading 3 in the fingerprint region from patient primary cell culture300
6.13 PCA and LDA analysis of the high wavenumber region of primary prostate cell line cultures in the fingerprint region . . . 303
6.14 PCA loading 1 in the high wavenumber region from primary prostate cell culture comparison . . . 304
6.15 PCA loading 2 in the high wavenumber region from patient primary cell culture comparison . . . 305

6.16 PCA loading 3 in the high wavenumber region from patient primary cell culture comparison 305
6.17 Comparison of the standard cell lines and primary cell cultures using the PIR R1001/1125 309
6.18 Comparison of the standard cell lines and primary cell cultures using the PIR R1448/1125 310
6.19 Comparison of the standard cell lines and primary cell cultures using the PIR R715/1125 310
6.20 Comparison of the standard cell lines and primary cell cultures using the PIR R745/1254 312
6.21 Comparison of the standard cell lines and primary cell cultures using the PIR R1001/1254 313
6.22 Comparison of the standard cell lines and primary cell cultures using the PIR R2894/1254 313
6.23 Comparison of the standard cell lines and primary cell cultures using the PIR R745/2873 315
6.24 Comparison of the standard cell lines and primary cell cultures using the PIR R2934/2873 316
6.25 R3015/2851 317

CHAPTER 1

INTRODUCTION

1.1 Overview

In 2020, the Global Cancer Observatory website (GLOBOCAN) , covering 185 countries and 36 cancers, ranked prostate cancer as the fourth leading cause of mortality overall due to cancers, behind breast, lung and colorectal cancers; ranking second overall for people with prostates [1]. The global estimate of new cases in 2020 was 1,414,259, putting the incidence of new prostate cancer cases at 7.3% of all newly discovered cancer cases [1]. Using the same GLOBOCAN estimates from 2008 to 2020 it is possible to see continual increases in the incidence of prostate cancer. The number of new cases grew from 903,500 in 2008 to 1,414,259 in 2020, a 56.5% increase in 12 years [1–4]. This increase in incidence has not affected the rank of prostate cancer-related deaths among people with prostates, remaining sixth between 2018 and 2020 [1, 4]. Although the key drivers have not been identified, the global incidence of prostate cancer is only predicted to increase over the next decade, with an estimated new case rate of 1.7 million cases per year by 2030 [5].

At present there is no definitive test used for the diagnosis of prostate cancer. At the primary point of care the standard route to diagnosis is the use of a prostate-

specific antigen (PSA) test and a digital rectal examination (DRE) [6]. The DREs are carried out at a primary care point. A review across four studies involving 3225 patients indicated that DRE had a sensitivity of 28.6% and specificity of 90.7% when used as a predictive tool for prostate cancer [7]. However, some have argued, given the risk that aggressive prostate cancers may be present in what feels to be a normal prostate, other symptoms would mitigate the test and require urgent referral [8]. PSA, although discovered as early as 1960, was not discovered in blood till 1980, with clinical PSA testing beginning in 1987 [9–11]. The choice of a PSA threshold at which a patient is referred for further treatment is not a fixed value. A study by Osterling *et al.* recommended the total PSA level at which a patient is referred be adjusted by age [12]. The total PSA level deemed necessary for referral: 40 years of age, 2.5ng/mL; 50 years of age, 3.5ng/mL; 60 years of age (4.5ng/mL); 70 years of age (6.5ng/mL). PSA is not only used in the diagnosis stages of prostate cancer but also the monitoring of the cancer as well as benign prostate conditions [13–15]. One example of using PSA in the monitoring of prostate cancer is measuring PSA velocity (ng/ml/year) [16]. PSA doubling time is a further monitoring tool utilising PSA: the time taken for the PSA level to double, for example the time taken for a patient with 2.5ng/mL PSA level to reach a 5ng/mL PSA level [17]. The PSA velocity can also be used predicatively to indicate when the PSA doubling time might occur. PSA doubling is widely used as a predictive marker in indicating a patient's chance of surviving prostate cancer and effectiveness of treatment methods [18]. However, despite the long-term global use of PSA tests they remain controversial in the use of managing prostate conditions [19]. This is primarily because PSA alone cannot differentiate between prostate cancer and other conditions which can lead to complications [20, 21]. One complication that can occur from PSA tests is false positives. As PSA is organ and not cancer specific, levels can be elevated in people with prostate glands suffering other conditions [22, 23]. The most common cause of elevated PSA levels especially occurring in men of old age, is benign prostate hyperplasia (BPH) [24]. Further common conditions resulting in elevated PSA levels include prostatitis and inflammation of the prostate [25, 26]. Despite the pitfalls of PSA testing many reports suggest that PSA testing has significantly reduced mortality

from prostate cancer [27].

The primary care physicians may decide that after the DRE, PSA test or if other symptoms/comorbidities are present, that further investigation is needed for specific diagnosis. This involves specialist referral to hospital. One method of diagnosis is to use a transrectal ultrasonography-guided biopsy (TRUS) to take around 10 random cores from the prostate. The effectiveness of the TRUS technique of taking random cores does generate a false negative rate of 15% to 46% as well as a tumour undergrading rate up to 38% [6]. One issue surrounding TRUS is that there are a lot of patients who are subjected to it unnecessarily [28]. Biopsy, aside from the time and finances consumed, can itself lead to complications such as pain and risk of infection [29]. Furthermore, as the amount of biopsies increase with the incidence of prostate cancer, greater pressure is put on pathology labs [30]. To reduce unnecessary biopsy and misdiagnosis it is necessary that improved diagnostic methods are developed [31]. A proposed and practiced current gold standard of diagnosis is to scan patients using an multiparametric magnetic resonance imaging (MRI) machine before conducting a targeted biopsy of the prostate. Studies have shown this technique to be more effective in detecting clinically significant prostate cancers. Kasivisvanathan *et al.* found an increase in diagnosis of prostate cancer using MRI in combination with TRUS with a 95% confidence interval; MRI targeted biopsies yielding 38% positive diagnosis in patients and 26% in the non targeted patients using random TRUS method [32].

Globally, the mortality rate of prostate cancer accounts for 6.6% of all male cancer related deaths, translating to over 300,000 deaths [33, 34]. It is for this reason that development of new therapeutic agents to treat prostate cancers must take place. However, one of the main issues in anticancer drug development is that even after passing through the drug testing pipeline, only 5% of drugs developed at a clinical stage are licensed for treatment [35]. The classical development of anticancer drugs involves testing of the therapeutic agent on malignant cell lines before progressing to xenograft tumours in mice to understand the mechanism by which the cancer operates [36]. A further strategy is to directly explore chemical pathways associated with cancer proliferation and development [37]. Cancer cell

lines are commercially available cell cultures that have been immortalised by introducing a plasmid containing a genome with a defective replication origin, meaning they can be cultured indefinitely [38–41]. Cell lines are generally preferred as the first step of preclinical trials due to their pure populations of cancer cells, making it easier to detect changes in the genetic material [42]. Furthermore, cell lines are easy to handle and easily replaced if lost through contamination [43]. A positive factor associated with standard cell lines is the ability to maintain reproducibility of experiments on these types of cells in labs, e.g., from 2000 to 2001, there are 125 papers whose titles included the cell line LNCaP that were published [44]. However, it is emphasised that discrepancies in results are common between labs using the same cell lines [45]. Cross contamination is a serious issue affecting cell lines with one German cell line bank reporting up to 18% of stock containing contaminants [43, 46]. Interspecies cross-contamination has also been commonly found in original stocks of human cancer cell lines [47]. Additionally, several of the prostate cancer cell lines that are used in drug development research are outdated with the cell lines themselves being established in excess of 20 years ago, in some cases. For example, LNCaP, which is an androgen-sensitive (able to grow but affected by steroid hormones) prostate cancer cell line, was immortalised in 1980 [39]. Due to the age of certain cell lines, it can be concluded that by using these they will not represent current disease, nor the complex level of heterogeneity that is associated with real prostate cancer [48]. For example, the patient from which the LNCaP cell line was derived had failed on a combination of drugs related to castration resistant cancer of the 1980s, before the cell line was established. This cancer is now treated with new generation androgen therapies so the prostate that had undergone these drug therapies from which LNCaP is derived no longer represents a prostate that has undergone state of the art treatment [49]. At clinical trial testing, underlying these issues is the inability to characterise and quantify real patient heterogeneity through the use of cell lines [50]. Patient-derived xenograft models have proven to be immensely helpful in the discovery of new therapeutic agents in the preclinical stages [51]. In parallel to the use of cell lines there are also disadvantages to using xenograft models; they do not accurately represent the primary tumours from which they

were derived, proliferating and reacting more rapidly to therapeutic agents than human tumours [52]. The use of models that do not represent true tumours and their heterogeneity, is not only ineffective, but also costly in terms of financial investment and the lab-time involved [53].

The hypothesis of this thesis is that live primary tissue-cultures, i.e, tissue that is derived directly from the patient, will allow us to develop patient specific drugs that will have better chances of success through the drug development pipeline. This will allow us to speed up, reduce expenditure on further drug development and increase the chances of successful treatment in clinical trials. The reason for this advance is because primary samples would be duly representative of the variability of disease in each patient [54]. Figure 1.1 highlights heterogeneity of the patient cell-cultures used in this study. Patient primary cell cultures contain stem cells, transit amplifying cells, committed basal cells and luminal cells combining to give complete cellular heterogeneity unlike standard cell lines that have only one undifferentiated cell type. However, in order for this primary tissue and hence cells, to be assessed for effectiveness in drug development, a full spectrum (ranging from normal tissue, dysplastic tissue to varying levels of cancerous tissue) of diseased and healthy tissue must be fully characterised.

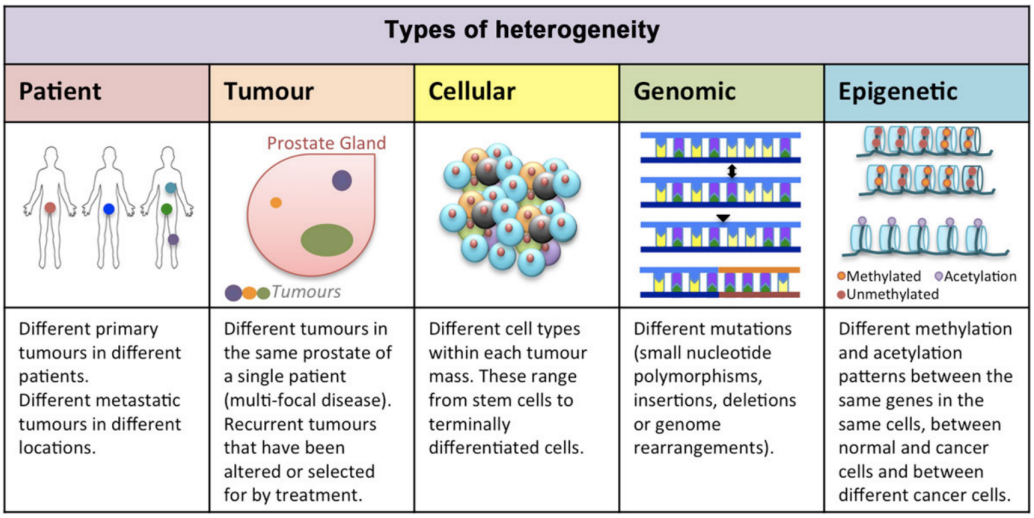


Figure 1.1: A table showing the complex nature of heterogeneity in prostate disease in humans. The different levels of diversity and ongoing variations are shown. Here, epigenetic refers to changes in cells caused by a change in the gene expression rather than the genetic code itself. Specifically, methylation prevents gene transcription and acetylation alters accessibility of chromatin to be read and expressed. Both methylation and acetylation lead to irregular functioning of the cell due to incorrect transcription. Figure obtained from Frame *et al.* (2017) [54].

It is therefore essential that prostate cancer is not only fully characterised to aid in the detection of disease but also able to monitor its progression [55]. Raman spectroscopy is a spectroscopic technique that was experimentally demonstrated in 1928 [56]. Despite this, it has only been in the past two decades that the development of instrumentation with the sensitivity to measure tissues and cells has been possible [57]. Raman spectroscopy utilises a laser to probe molecular vibrations by measuring the change in energy between the incident beam and inelastically scattered photons [58]. This is possible due to Stokes or anti-Stokes scattering; the incident source of energy of the photon is too small to excite an electron transition and energy is exchanged by the molecule as it interacts with by giving or receiving energy [59]. Stokes scattering is when the molecule, in a ground state, absorbs energy from the incident photon and now rests in a higher energy state. However, in anti-Stokes scattering, a molecule already exists in an excited state due to thermal energy and transfers energy to the incident photon before returning to a ground state, rarely happening at room temperature [60].

Raman spectroscopy at a single cell level can generate a detailed biochemical profile, with specific bands being able to be assigned to specific molecules such as proteins, lipids and carbohydrates [61]. This has generated a lot of interest and potential around the use of Raman for biochemical analysis [62]. The approach that this thesis has taken is to utilise Raman spectroscopy to achieve full biomolecular characterisations of standard cell lines, whilst also characterising tissue samples from patients. This would be a novel advance over previous Raman spectroscopy studies. Crow *et al.'s* (2005) study using Raman spectroscopy was able to discriminate between four prostate cancer standard cell lines representing several stages of androgen independent disease (tumours that are able to grow without androgen group hormones) [63]. Additionally Potcoava *et al.'s* (2014) study of hormone treated prostate cancer cell lines LNCaP and PC3 showed that Raman spectroscopy was able to differentiate between the treated and untreated samples with high accuracy, sensitivity and specificity [64].

This study aims to be a lot more comprehensive in terms of the number of standard cell lines and primary cell-cultures than that of previous studies that have looked at prostate cancer such as the previous studies mentioned [63, 64]. It includes the analysis of several standard prostate cell lines and primary cell cultures that are representative of the full spectrum of prostate disease, e.g., normal, dysplastic, localized cancer, metastatic cancer etc. Full biomolecular characterisation is used to probe differences between standard cell lines and patient derived cell lines using principal component analysis (PCA), linear discriminant analysis (LDA) and peak intensity ratios (PIR). Comparisons will also specifically address the use of a dried standard cell line against that of a standard live cell line. Analysis of the primary cell cultures will also demonstrate the separation of normal and cancer samples from the same patient and also from several other patients to observe quantitative differences in the variances between the cells. The results will also include comprehensive statistical convergence testing to find the necessary number of spectra samples for average spectrum and statistical quantities such as the standard error and twice the standard deviation to converge.

1.2 The prostate

1.2.1 Morphology of the prostate gland

The human prostate is an exocrine tubular-alveolar gland found at the base of the urinary bladder [65]. Its appearance is similar to that of a chestnut and in total it weighs roughly 30 grams [66]. The prostate is not encapsulated, but covered by a fibromuscular band that also forms an integral part of the gland [67]. The prostate and urinary bladder are separated posteriorly from the rectum by the Denonvilliers fascia, a series of adhered connective tissues. Furthermore, the space between the prostate, urinary bladder and seminal vessels is filled with continuous loose connective tissue.

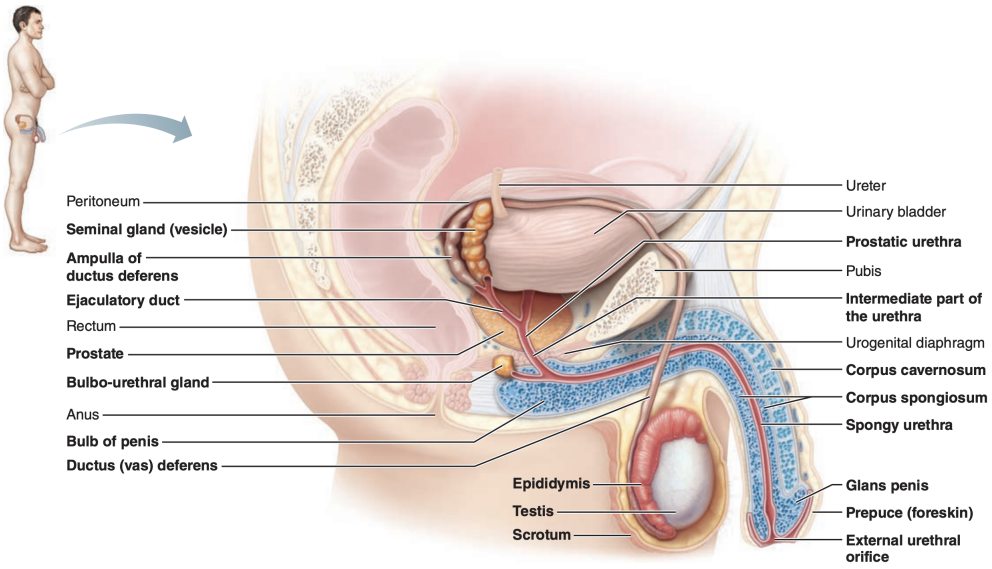


Figure 1.2: A longitudinal cross section of the male reproductive organs. Here, the prostate can be shown to be located below the urinary bladder and is positioned next to the rectum [68].

In 1981, John McNeal established the three-dimensional zonal anatomy of the human prostate through extensive analysis of cross sections from multiple planes from over 500 specimens [69]. The prostate itself can be described anatomically as having four distinct zones.

The peripheral zone is the largest part of the prostate consisting of over 70% of the organ. It is in the peripheral zone that nearly all carcinomas arise [71]. The

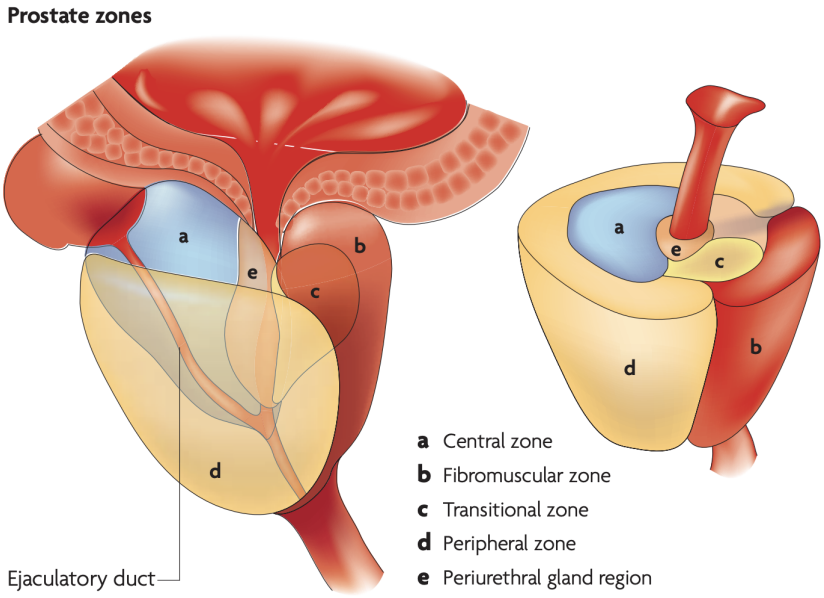


Figure 1.3: Zonal diagram of the prostate gland. The zones are as follows: a - central zone , b - fibromuscular zone, c - transitional zone, d -peripheral zone and e - periurethral gland region [70].

second largest zone is the central zone that constitutes approximately 25% of the prostate and contains the ejaculatory ducts. Its lateral border fuses with the proximal peripheral zone border, completing in continuity with the peripheral zone. The third glandular region accounts for only 5% of the prostate region, the prepostatic region [69]. Duct development is abandoned in the prepostatic region, creating a transition zone. It is in this transition zone that benign prostate hyperplasia (BPH) exclusively occurs [71]. Lastly, a non-glandular inseparable thick anterior fibromuscular stroma is fused onto the anterior of the prostate, protecting the three aforementioned glandular regions and covering them entirely.

The primary function of the prostate gland is to secrete prostatic fluid. The prostatic fluid is a milky fluid that is rich in sodium, potassium, calcium and enzymes [72]. The purpose of the prostatic fluid is to combine with sperm cells from the testis, fluid from the seminal vesicle and fluid secreted from the bulbourethral gland (a small gland that sits below the prostate) to form semen. The mixing of these fluids into semen occurs in the urethra. A secondary function of prostatic secretion is to mobilise the sperm to aid male fertility. This is due to the prostatic fluid containing the prostate specific antigen (PSA) which makes the semen

thinner.

1.3 Aetiology of prostate conditions

The aetiology of prostate cancer is not concrete and is described as having multiple contributing factors [73]. Diet and nutrition are thought to play a role in the development of prostate conditions as well as sexual and reproductive health factors. External environmental exposures such as certain chemicals have also been reported as risk factors in the development of abnormal prostate conditions [74]. The hereditary risk of developing prostate cancer doubles for men who have a father or sibling who have had prostate cancer, with increasing risk when it occurs in multiple relatives [75].

1.3.1 Benign prostate hyperplasia

Benign prostate hyperplasia (BPH) is a term used to describe the non-malignant enlargement of the prostate gland in histological diagnosis. BPH is considered to only occur in the epithelium and stromal layer of the transition and periurethral zones of the prostate [76]. Symptoms presented by those afflicted with BPH originate in the lower urinary tract, and include weak urination, incomplete bladder emptying and general irritation [77]. Statistically, almost all men will develop histologic BPH by the age of 90 but it is important to note that it is not considered a precursor to prostate cancer [78, 79].

1.3.2 Prostate cancer

It is well established that prostate cancer manifests out of the epithelial cell regions of the prostate [80]. There are two main cell types present in the human prostate; basal and luminal. Additionally, there is a small percentage of neuroendocrine cells [81].

.

There is still dispute over the specific type of cells in the prostate that account

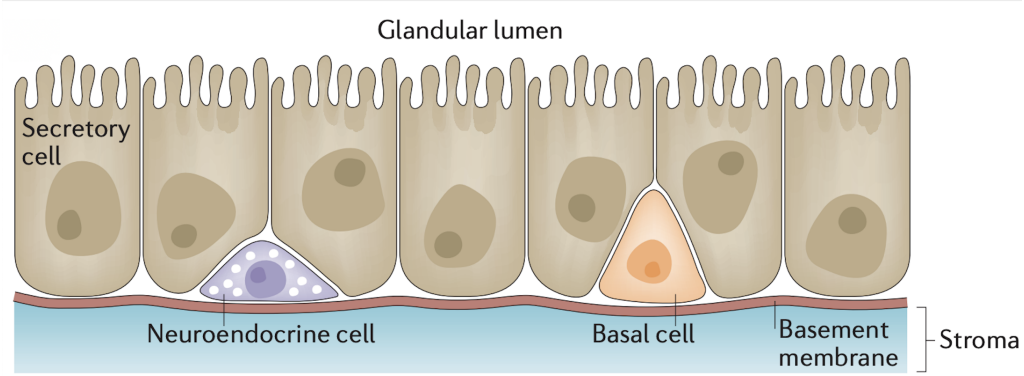


Figure 1.4: A diagram depicting a cross section of the epithelial layer of the prostate gland and stroma. It can be seen that the glandular epithelium is made up of two layers. The first is a secretory upper layer of luminar cells. Luminar cells secrete the seminal fluid, which includes prostate-specific antigen and prostatic acid phosphatase, both used as indicators in diagnosis of prostate cancer. Below the luminal layer sits a second layer including both neuroendocrine cells and basal cells. A basement membrane separates these cells from the stroma [81].

for the origin of the cancer [82]. One study, using genetically engineered mice, demonstrated that both basal and luminal cells can act as the point of proliferation for the disease [83]. However, it is generally assumed that prostate cancer predominantly originates from the luminal cell [84]. This assumption is due to prostate cancers generally displaying luminal cell phenotypes, as well as a loss of basal cells, suggesting that the latter may be the origin and fuel for prostate cancer development [85]. On the contrary, evidence has been presented suggesting that due to the recurrence and prevalence of castration-resistant prostate cancer after antihormonal therapy, that basal cells are an origin of prostate cancer [86, 87]. This assumption is in part due to basal cells expressing a phenotype of androgen-independence, which is common with advanced prostate cancers [88].

1.3.3 Other conditions

1.4 Epidemiology of prostate cancer

Using the 2018 GLOBOCAN estimates, a project run by the International Agency for Research on Cancer (IARC) that estimates the incidence and mortality worldwide for 36 cancers in 185 countries globally, using the best available data and

estimation models, it is observed that prostate cancer is only second to lung cancer in incidence rates among males, whilst also ranking the fifth in mortality [4, 89]. It is observed that the incidence rates of prostate cancer in African-Americans are higher than any other race, and a mortality rate double that of white men [90]. Reasons for the stark contrast in incidence and mortality rates between races have been hypothesized to not only genetic differences between races, but also social and environmental factors. [91]. A clear observation is that there is also a large deviation between countries in terms of incidence and mortality. It is observed that high income countries experience a higher incidence of prostate cancer, whereas the low to middle income countries experience a much higher mortality rate [92].

1.5 Current prostate cancer models

1.5.1 Standard prostate epithelial cell lines

Standard prostate epithelial cell lines are an example of an *in vitro* cell culturing technique [93]. In the literature these are referred to simply as "cell lines". Most cell lines are commercially available cell cultures that have been immortalised by introducing a plasmid containing a genome with a defective replication origin [94]. They are by far the most common *in vitro* model and widely used due to easy culturing, rapid proliferation, handleability and infinite reproduciblity [95]. Standard cell lines form a fundamental basis for cancer research, and the development of future drugs, as they are used as a primary step in testing new therapeutic agents [38–40].

Despite widespread use of prostate epithelial cell lines, there are issues pertaining to their use. Firstly, as the number of passages (the process of transferring the cells to new growth medium) increases, the risk of chromosome instability also increases [96, 97]. This will lead to issues such as tetraploidy, whereby cells exhibit twice the normal number of chromosomes, an early stage of tumorigenesis, which ultimately no longer accurately represents the initial epithelial cell culture [98]. Furthermore, due to the period in which several cell lines were created, it

can be concluded that by using these they will not represent current disease, nor the complex level of heterogeneity that is associated with real prostate cancer [48]. For example, the patient from which the cell line LNCaP was derived had failed on a combination of drugs in the 1980s, that are no longer used to treat castration resistant cancer in the present day. This cancer is now treated with new generation androgen therapies so the prostate that had undergone these drug therapies from which LNCaP is derived would not exist in 2018 [49].

1.5.2 Primary cell cultures

Primary cell cultures are derived directly from patients. In the case of this thesis, it is best to consider purely primary prostate epithelial cells. There are several ways in which the cells can be initially procured and derived. The most prevalent way of procuring the cells initially is through a radical prostatectomy, an operation to physically remove the prostate as well as surrounding tissue. However, other procedures such as needle biopsies and extraction of prostatic fluid have been detailed [99]. A key feature of primary prostate epithelial cells is that they maintain the cellular heterogeneity of the patients' prostate cells *in vitro*. This is because unlike standard cell lines that contain purely one cell type, a primary cell culture contains five cell types. These cell types are inter-related and specific to the sample. These 5 cell types can be broken down into basal cells, stem cells, neuroendocrine cells, transit amplifying cells and secretory luminal epithelial cells [99]. The two dominant cell types are the basal and secretory. The basal cells line the basement membrane of the prostatic epithelium while the secretory cells exist in the lumens of the epithelial layer. Neuroendocrine cells exist only in small quantities in the prostatic epithelium. These cell cultures are not immortalised by any plasmid, and thus have a limited lifespan.

1.5.3 Wider models

There are a number of other models used in cancer research and drug development that have not been included in this PhD project. One issue with using 2D cell cultures such as cell lines is that monolayer cultures overlook the extracellular

barriers that exist in three dimensional *in vivo* environments [100]. This has led to the development of 3D spheroid cell in vitro models. However, the reason for the exclusion of this model for this specific project was due to the labour intensive nature of procuring samples for the project [101, 102]. *In vivo* mouse xenograft models that allow the culturing of cell lines as well as primary tumours, have shown promising results, allowing discrimination of normal and cancerous tissue [103, 104]. Mouse xenograft models are not only expensive to maintain and require additional space but are never going to fully replicate an *in vivo* human environment.

1.6 Raman spectroscopy

Raman Spectroscopy is a vibrational laser based technique that enables the biochemical profiling of the molecules in a sample [105]. Raman scattering was postulated in 1923 by A. Smekal utilising the principles of classical quantum theory [106]. However, it was not until 1928 that it was observed experimentally by C. V. Raman and his student K. S. Krishnan [56, 107]. Raman spectroscopy exploits the phenomenon of inelastic scattering; the change in energy of a photon as it interacts with a molecule is measured using a charge coupled device (CCD) to generate a Raman profile. The energy change of the photon is described as a gain in energy (Stokes) or loss of energy (anti-Stokes). Raman spectroscopy is a weak type of scattering compared to Rayleigh scattering. Typically only one per million excitation photons from the laser source will generate a Raman photon [108]. Raman spectroscopy itself is a broad term used to describe over 25 different types of Raman spectroscopy [109]. In this thesis, spontaneous Raman spectroscopy is used to probe the molecular composition of cells as well as capturing spatial information in the commonly described fingerprint region (600-1800 cm^{-1}) and high wavenumber region (2700-3100 cm^{-1}) of the Raman spectrum [110].

One reason that Raman spectroscopy is suitable for the analysis of biological samples, as opposed to infrared spectroscopy, is due to cells consisting of water. The reason this is important is because the Raman activity from the water is

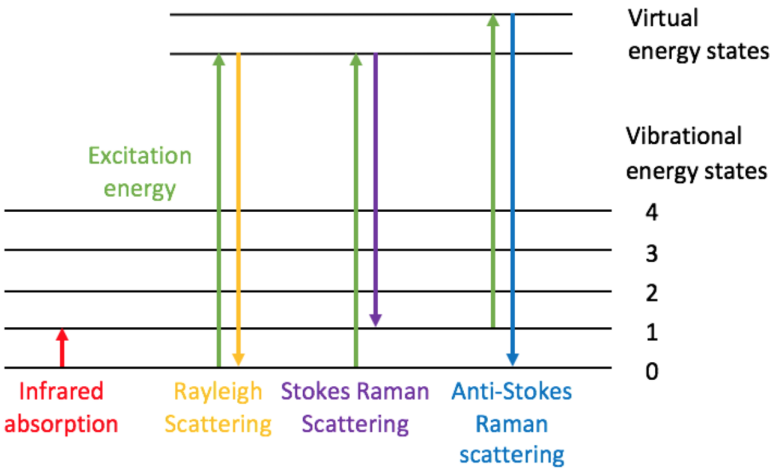


Figure 1.5: A diagram depicting the possible photon energy state transitions for both vibrational and virtual states for Rayleigh Scattering and Raman scattering.

very weak and therefore does not convolve with the biochemical Raman spectra generated by the cell, as opposed to infrared spectroscopy which generates a signal in key areas relating to biochemical composition [111]. Another benefit of Raman spectroscopy is that only a single cell is required to view biomolecular vibrations and their alterations, to differentiate samples. Furthermore, the non-destructive nature of Raman spectroscopy makes it suitable for analysis of living cells. At present, as far as the author is aware, there have not been any Raman studies carried out on the prostate gland *in vivo* . However, there have been exploratory *in vivo* Raman studies on other organs such as the cervix [112]. This would suggest that while it is not currently possible to carry out non-invasive *in vivo* diagnosis of prostate cancers, in the future non-invasive Raman spectroscopy could become a reality for *in vivo* diagnosis.

Rationally, it is important to consider other methods in the field of prostate cancer detection. Polymerase chain reaction (PCR) analysis is used to identify specific genetic alterations and drivers related to tumorigenesis in cells [113]. PCR is a destructive labelling technique that also requires the use of primers: single strands of DNA used to hybridise the original cell's DNA to amplify regions of interest to investigators [114]. It should be noted that the use of PCR is described as having extensive preparation procedures. This stems from the requirement of transportation of samples to a designated processing facility where staff have

specialised training and equipment necessary to carry out analysis [115]. Another destructive labelling technique is flow cytometry (FCM) whereby the measurements are made as the desired cells flow through a flow cytometer instrument [116]. Studies using flow cytometry have reported 5-10% of histograms being uninterpretable due to the possibility of autolytic changes in the cells during analysis [117, 118]. A further destructive technique used frequently is mass spectrometry. Mass spectrometry is a complex analysis method which generates ions through destruction of the organic or inorganic original compound and then differentiates and interprets constituents of the original sample through their mass-to-charge ratio and abundance [73]. In prostate cancer studies, mass spectrometry has been used to measure the abundance of molecular mass peptides and proteins between malignant and non malignant cells [119]. As a consequence of the aforementioned techniques being destructive and requiring interference of cellular processes through labelling, an alternative method is preferred for this study.

One of the aims of this study is the potential use of real time *in-vivo* diagnosis and therefore it is not suitable to use PCR, FCM or mass spectrometry as they do not provide results in real time [120–122]. Supplementary benefits to the avoidance of destructive and toxic labelling/tagging techniques also allows for additional use of samples post measurement [123]. Fourier transform infrared spectroscopy (FTIR) provides a non destructive, real time, non-labelling, high spatial resolution (1μm) and has been used as a method for the classification of benign and malignant prostate tissue [124]. Although FTIR has been used to probe and characterise bio-materials such as the prostate gland, it does suffer a major hindrance; the outputted spectrum is heavily convoluted by strong absorption of water in the infrared region [125]. Raman spectroscopy retains the properties of being non destructive, non-invasive, real time, non-labelling and high spatial resolution (1μm) whilst also possessing a very weak scattering cross section for water [126]. This makes Raman spectroscopy the best candidate for this PhD project due to the live cell measurements taking place in aqueous solutions. [127]

Raman spectroscopy allows for comprehensive scanning of a prostate cell line using the inelastic scattering of laser light. The unique molecular vibrations

from the biomolecular composition of the cell are then captured in a resulting spectrum. The Raman spectrum that is produced for each prostate cell line will allow us insight into what the constituent biochemicals of that specific cell are, providing a spectral fingerprint and allow us to discriminate between them through quantitative spectral assessment and comparison. A lot of previous work has been carried out to establish biochemical interpretations of Raman signatures [128, 129].

The reason that the characterisation of cell lines is so important for diagnostics is that to treat a cancerous cell, a complete analysis of the complex environment inside of it must be recorded. There is a large difference between these complex environments in cancerous prostate cells compared to normal prostate epithelial cells. A study highlighted this by investigating the difference in the Raman signatures generated by prostatic adenocarcinoma cell lines, androgen-sensitive cell lines (LNCaP and PCa 2b) and metastatic androgen-insensitive cell lines (DU145 and PC 3) [63]. Quantitatively it showed that a diagnostic algorithm was able to identify cell lines at a sensitivity of 98% and specificity of 99% [63]. Despite how accurately this study was able to discriminate between prostatic adenocarcinoma cell lines, it is also incomplete. To gain insight into what changes are taking place as cells become metastatic and proliferate, a complex biochemical understanding of the transitions taking place must be documented. This PhD is focused not only on prostatic adenocarcinoma cell lines but also the way that prostate cancer cells deviate from the normal cell lines and the changes that occur when transitioning to a dysplastic state (non-cancerous abnormal cells). It is important to expand on this work and to implement a diagnostic model that reflects the complex heterogeneity of a patient environment. A novel aspect of this PhD is the inclusion of patient cultures and not only characterising the patients cancerous prostate tissue but also the normal tissue to allow for insight into the biochemical changes. The reasoning behind the inclusion of primary patient cultures is to develop tailored treatment therapies in the future. Raman spectroscopy is used in this investigation as it is able to analyse these detailed intracellular changes without staining or labelling the cell [130]. Ultimately this

will allow Raman spectroscopy to be used *in vivo* in the future. However, the work carried out in this PhD will provide comprehensiveness into the biochemical differences in different prostate cancer cells and most importantly characterises the different standard prostate cell lines versus primary cell-cultures necessary to push Raman spectroscopy into more clinical applications.

CHAPTER 2

THEORY

2.1 Theory of Raman spectroscopy

2.1.1 Overview

Two landmark discoveries came in the late 19th and 20th centuries, the discovery of elastic scattering of light by small particles by John Strutt in 1871 and a further discovery of elastic scattering of light from atomic and molecular particles larger in diameter than that of the wavelength of the incident wave in 1908 by Gustav Mie [131, 132]. The former discovery by John Strutt would be known as Rayleigh Scattering and the latter discovery by Gustav Mie as Mie scattering [133, 134]. It was these discoveries that led to Adolf Smekal theoretically predicting the existence of inelastic scattering in 1923 [106]. It was this theoretical prediction that led to the discovery of inelastic scattering particles by Chandrasekhara Venkata Raman in 1928 [56]. In 1930 C.V. Raman won the Nobel prize for his discovery of the inelastic scattering of light and the phenomenon was subsequently named after him, the Raman effect. While the first Raman effects were observed on photographic paper a lot of progress scientifically has been made in the field. In particular the advent of lasers allowed for the control of a monochromatic coherent source of photons with the ability to control not only laser properties such as

wavelength, but also allow for integration into an optical microscope system. Advances in the field of optics, computing, sensors and integral components allowed for the transition from specialised research into much more commercial and widespread fields.

In this thesis, Raman spectroscopy is utilised to assess cellular information and fingerprint specific cell lines as well as primary cultures. This chapter explains the theoretical background of Raman spectroscopy. In this thesis, only spontaneous Raman spectroscopy is carried out. Whilst other types of Raman spectroscopy exist where the target system may vary, they are still all based on the interaction of a monochromatic radiation source inelastically interacting with the specific target system. The classical model of Raman spectroscopy still holds true; under specific circumstances it is necessary to apply the quantum theory to allow for complete analysis of the target system. For this reason an overview of the classical theory and quantum theory are given.

2.1.2 Classical theory

The classical theory is explained here using electromagnetic theory. In the classical theory for Raman spectroscopy it is considered that an electron cloud of a molecule can be distorted by an electric field **E**. This molecule then experiences an induced electric dipole moment **P**. This induced electric dipole moment is due to an oscillating electromagnetic wave interacting with the molecule, causing both electron and proton perturbation. The Raman signal can be received back from a sample, as different molecules in the system will emit photons of varying frequencies specific to each molecule. The electric field **E** and electrical dipole moment **P** can be mathematically expressed as a power series:

$$\mathbf{p} = \alpha \cdot \mathbf{E} + \frac{1}{2} \cdot \beta \cdot \mathbf{E}^2 + \frac{1}{6} \cdot \gamma \cdot \mathbf{E}^3 \tag{2.1}$$

In equation (2.1) α represents the polarisability tensor, β represents the 2^{nd} hyperpolarisabilty tensor and γ represents the 3^{rd} hyperpolarisability tensor. These tensors are second-ranked tensors and the units given to α, β and γ are

$CV^{-1}m^2$ [135]. Respective magnitudes are given to each polarisability tensor respectively: α, $10^{-40}CV^{-1}m^2$; β, $10^{-50}CV^{-1}m^2$; and γ, $10^{-61}CV^{-1}m^2$. It can be seen here that the tensors have decreasing magnitude of roughly 10^9. As the Raman spectra can be observed with relatively weak electric fields, β and γ can be treated as negligent and thus discarded. Therefore a directly proportional relationship is assumed between the induced dipole moment and the electric field, which can be reduced from equation (2.1):

$$\mathbf{p} = \alpha \cdot \mathbf{E} \tag{2.2}$$

In equation (2.2) α is used to notate the polarisability as a tensor. This tensor will vary as chemical bonds vary in size and shape. As previously stated an oscillating electric field can be used to describe the photon as it interacts with the molecule. In this case $\mathbf{E}$ is given as the electric field vector. The electric field strength $\mathbf{E}$ fluctuates with time t, it can be described as:

$$\mathbf{E} = \mathbf{E} \cdot \cos(2\pi \cdot \nu_0 \cdot t) \tag{2.3}$$

In equation (2.3) the expression ν_o represents the vibrational frequency of the electromagnetic radiation. The polarisability factor α as previously stated depends on the structure, shape and size of the bonds. The polarisability factor α also undergoes changes as the bonds interact with the excitation source. Using a Taylor series it is possible to express α in terms of the normal coordinates of vibration using coordinate Q:

$$\alpha = \alpha_0 + \sum_k \left(\frac{\partial \alpha}{\partial Q_k}\right)_0 \cdot \frac{1}{2} \sum_{k,l} \left(\frac{\partial^2 \alpha}{\partial Q_k \partial Q_l}\right)_0 \cdot Q_k \cdot Q_l \tag{2.4}$$

In equation (2.4) the normal coordinates are denoted by Q_k and Q_l. Furthermore, the normal coordinates Q_k and Q_l correspond to the normal vibration k^{th} and l^{th} and with the vibrational frequencies ν_k and ν_l. An approximation

can be made by considering the k^{th} and l^{th} normal vibrations independently. Consequently, the v^{th} normal vibration can be used to reduce equation (2.4) to:

$$\alpha_v = \alpha_0 + \alpha'_v \cdot Q_v \tag{2.5}$$

In equation (2.5) α'_v denotes the derivative of the polarisability tensor with Q_v representing the normal coordinate. In the classical theory the molecule consists of two or more atoms, modelled by visualising the atoms as balls with springs between them. It is then possible to consider that these molecules would obey Hooke's law. In this instance, the strong bonds with light molecules correlate to the highest frequencies and the heavy atoms with weak bonds have lower frequencies. An approximation for normal coordinate Q_v is to allow it to oscillate as a harmonic oscillator. In turn normal Q_v will vary with time t:

$$Qv = Q_{v0} \cdot \cos(2\pi \cdot \nu_v \cdot t \cdot \varphi_v) \tag{2.6}$$

Here, Q_{v_0} represents the amplitude of the normal vibration with the phase angle being denoted φ_v. It is therefore possible to substitute the expression for Q_v into equation (2.5):

$$\alpha_v = \alpha_0 + \alpha'_v \cdot Q_{v0} \cdot \cos(2\pi \cdot \nu_v \cdot t \cdot \varphi_v) \tag{2.7}$$

If the vibrational frequency of the normal coordinate is set to the polarisability factor, harmonically oscillating with frequency ν_v, it is possible to carry out a substitution of equations (2.3) and (2.7) into equation (2.2) which represents the dipole moment $\mathbf{p}$:

$$\mathbf{p} = \alpha_0 \cdot \mathbf{E}_0 \cdot \cos(2\pi \cdot \nu_0 \cdot t) + \alpha'_v \cdot \mathbf{E}_0 \cdot Q_{v0} \cdot \cos(2\pi \cdot \nu_0 \cdot t) \cdot cos(2\pi \nu_v t + \varphi_{nu}) \tag{2.8}$$

Equation (2.8) can be simplified by using the trigonometrical identity:

$$\cos A \cdot \cos B = \frac{1}{2}[\cos(A+B) + \cos(A-B)] \tag{2.9}$$

Allowing equation (2.8) to be simplified to:

$$\begin{aligned} \mathbf{p} &= \alpha \cdot \mathbf{E}_0 \cdot \cos(2\pi \cdot \nu_0 \cdot t) \\ &+ \frac{1}{2} \cdot \alpha'_\nu \cdot \mathbf{E}_0 \cdot Q_{\nu 0} \cdot \cos[2\pi \cdot (\nu_0 + \nu_\nu) \cdot t + \varphi\nu] \\ &+ \frac{1}{2} \cdot \alpha'_\nu \cdot \mathbf{E}_0 \cdot Q_{\nu 0} \cdot cos[2\pi \cdot (\nu_0 + \nu_\nu) \cdot t - \varphi\nu] \end{aligned} \tag{2.10}$$

Using equation (2.10) it is possible to rewrite it in terms of vibrational frequency ν_ν and the incident radiation ν_0.

$$\mathbf{p} = \mathbf{p}(\nu_0) + \mathbf{p}(\nu_0 + \nu_\nu) + \mathbf{p}(\nu_0 - \nu_\nu) \tag{2.11}$$

The induced dipole moment given by equation (2.11) can be understood better by splitting it into three parts. The first term of the ν_0 denotes the elastic scattering of radiation: Rayleigh scattering, no change in the frequency. The second and third terms both involve a change in frequency of the scattered radiation: inelastic scattering, Raman scattering. The term "$\nu_0 + \nu_m$" corresponds to the Raman anti-Stokes scattering. This occurs when the scattered beam of radiation has a higher frequency than that of the incoming beam. Finally, the term "$\nu_0 - \nu_m$" corresponds to the Raman Stokes scattering: a drop in the frequency of scattered radiation. When the rate of change of polarisability $(\partial_\alpha/\partial_Q)_0$ is zero, the vibration is therefore observed to be Raman-inactive. Therefore, to be a Raman active vibration mode, the rate of change of the polarisability must not be zero.

In this thesis, a monochromatic laser beam is used to measure a sample. To measure the inelastic scattering of the beam, its intensity is measured as a function of wavelength. The units of intensity are arbitrary and the wavelength is cm^{-1}. This is due to the relationship between the frequency of scattered and incoming beam, ν_m and ν_0 respectively. This can be show mathematically:

$$\omega = \bar{\nu_m} - \bar{\nu_0} = \frac{\nu_m}{c} = \frac{\nu_0}{c} \tag{2.12}$$

It can also be seen from equation (2.12) that positive and negative results are possible. Positive wavenumbers ($+cm^{-1}$) correlate to Stokes scattering and negative wavenumbers ($-cm^{-1}$) represent anti-Stokes scattering.

Whilst the classical theory does correctly tie together Rayleigh scattering and Raman scattering using frequency dependency as well as highlighting the dependency of the scattering on the polarisability factor, it does fall short of a complete understanding. In situations such as molecular bending and symmetric/ asymmetric the classical theory remains correct. However, it is in the case of molecular rotations that the classical theory falls short. This is due to its inability to describe the rotational frequencies of the molecule . Therefore, to fully extract all the data from Raman interactions we must treat it using a quantum theory, replacing the classical amplitude with a quantum mechanical amplitude [109, 136–139].

2.1.3 Quantum theory

From a quantum mechanical perspective Raman scattering is described as a transition of energy levels in molecules, from a ground to virtual state. A virtual energy state is when the incident photon interacts with the molecules, distorting the electron cloud surrounding the nuclei and creating a superposition state. This superposition is very brief and collapses, re-emitting a photon. If the virtual state collapses back to the ground state, a photon of the same wavelength as the incident photon will be emitted. This elastic scattering process is called Rayleigh scattering. However, if it collapses back to excited vibrational state a photon with an increased wavelength, redshift, is observed. The redshift is observed is Stokes Raman scattering. Furthermore, if the molecule existed in a vibrational state and collapses to a lower vibrational state a photon is emitted with a decreased wavelength, blue-shift, inelastic anti-Stokes scattering. It should be noted that it is not possible to measure the energy of a virtual state as it is a super position

of a state and not an eigenstate of the Hamiltonian itself. Furthermore, the transition of molecules in a material from the ground state to an eigenstate is impossible through simple photon absorption. Additionally, the quantum theory can provide an explanation as to why different types are of scattering are more dominant than others. Rayleigh scattering is the most common type of scattering and dominates the other two types of inelastic scattering. This can be explained using the Maxwell-Boltzmann distribution. The highest probability is that the particles exist in the ground state and return to the ground state after being in the virtual state. The least common type of scattering to take place is the inelastic anti-Stokes scattering. In this case, the molecule does not exist in a ground state but in a higher vibrational energy state. Although anti-Stokes Raman scattering would give the same results as Stokes Raman scattering, it is not used due to the Maxwell-Boltzmann distribution, stating that probability is greater that the molecules exist in the ground state 0, thus it is incredibly rare compared to Rayleigh scattering and inelastic Stokes scattering. In inelastic Stokes scattering when the virtual state collapses it returns to a higher vibrational energy state. Due to the Maxwell-Boltzmann distribution this is a lot more likely than anti-Stokes scattering as the initial molecule is more likely to exist in ground state 0. Therefore, we only consider Stokes Raman scattering when taking Raman measurements. Stokes scattering still has a very low probability of occurring and only accounts for 0.001% of light re-emitted from a sample of molecules [137].

Mathematically we can consider the quantum treatment of a molecule, treating it with quantised energy levels. The total energy of a molecule is given by:

$$E = E_e + E_v + E_r. \tag{2.13}$$

Where E_e represents the total electric energy of the molecule, E_v represents the vibrational energy and E_r is the rotational energy. When considering the E_v therefore only certain energy levels are permitted for the molecule

$$\mathbf{E}_v = (v + \frac{1}{2}) \cdot h\nu_v \quad (v = 0, 1, 2, 3, ...) \tag{2.14}$$

In Equation (2.13) v represents the vibrational quantum number. The term ν_v denotes the molecular vibrational frequency. The term h represents Planck's constant. As previously stated, the particle should exist in a state at 0 or raised energy state. This infers that the only transitions from original to final state should occur with change in energy 0 or ± 1. Rayleigh scattering accounting for no change in energy and Raman for a change in energy ± 1 from the original state. However, due to the Heisenberg uncertainty principle it is possible for other transitions to occur for very brief period in a virtual state:

$$\triangle E \triangle t \geq \frac{h}{4\pi} \tag{2.15}$$

This form of the Heisenberg uncertainty principle is given in terms of energy measurements of energy $\triangle E$ and time $\triangle t$. The difference in energy of the Raman interaction can measured by isolating the terms:

$$\pm h \cdot \nu_v \,, \qquad \text{where} \quad \nu_v = \frac{1}{2 \cdot \pi} \cdot \sqrt{\frac{\kappa}{\mu}} \tag{2.16}$$

Here, κ is the force constant of the bond between molecules and μ corresponds to the reduced mass. It should be noted that the reduced mass term can be expanded dependent on what type of molecule is interacting with the radiation beam. From equation (2.16) it is possible, using the force constant and the reduced mass, to determine the Raman position. The energy difference, similar to the classical theory, plots the energy difference in wavenumbers (cm^{-1}).

2.2 Multivariate analysis

This theory section covers principal component analysis (PCA) and linear discriminant analysis (LDA). These methods both work by reducing the dimensionality

of the data set and have been widely accepted as tools for the classification of Raman data [63, 64, 140–142].

2.2.1 Pre-treatment of spectra for multivariate analysis

Prior to PCA/LDA analysis the Raman data must be pre-processed. This is carried out using Raman Toolset, a piece of software commonly used in Raman analysis [143, 144]. The cell lines and primary cultures spectra are cut to the ranges 600-1800 cm^{-1} and 2700-3100 cm^{-1}. These ranges are selected due to them containing the necessary biochemical information to characterise the samples under the experimental design of this thesis. Cutting is essential so that separation of samples does not occur due to varying wave length. Furthermore, the range of waves may vary between spectra of the same sample due to the Raman instrument, so all waves of the same sample must be cut to the same length. To perform analysis between cell lines and primary cell cultures, an additional step may be necessary. This additional step is to interpolate the data set. Interpolation allows for two samples with a different number of points in their x and y coordinates to be compared. Interpolation works by constructing the Raman data a new artificial set of x and y coordinates within the range of predetermined data points specific to the Raman data set. This is done by considering the number of points in each Raman spectra and the exact start and end values of the x-wave. If the samples being compared do not have the same number of points in the x and y wave then PCA/LDA analysis can not be carried out. The linear interpolation was carried out in IGOR PRO 8 using group written code. The data is then baselined in Raman Tool Set. The fingerprint region 600-1800 cm^{-1} is baselined using a linear baseline. The high wavenumber region 2700-3100 cm^{-1} is baselined using a polynomial baseline. This is done to remove various background signals such as cosmic rays. The data is then normalised to the total spectrum area. Lastly, the data is smoothed using a cubic spline in Raman Tool Set to reduce noise in the spectra or any artefacts that may be present due to background interference.

2.2.2 Principal component analysis and linear discriminant analysis

Principal component analysis (PCA) is an unsupervised analysis technique using a mathematical algorithm to analyse a data table. There is some dispute over the origins of PCA, but historically accepted origins of the analysis technique trace back to Karl Pearson in 1901 as well as Harold Hotelling in 1933 [145, 146]. The primary purpose of PCA is to reduce the dimensionality of the data set. This allows for a large number of interconnected variables to be extrapolated from the data set whilst retaining the variance of the data set [147]. This is effectively done by creating a new set of correlated variables called principal components (PCs). PCs identify directions in the data that capture the most variance shared by the original data set [148]. This reduction will reduce the number of variables from large numbers in the original data set to a much lesser usable number of variables. It is then possible to plot different PCs against each other. Mathematically, PCA is an algorithm that treats variables as columns and observations as rows [149].

There are several steps that must be undertaken to perform PCA analysis. The first step is to subtract the mean values of the data set ($\bar{x}$, $\bar{y}$) in each data dimension from the x and y values. This leaves the mean of the data set at 0. The next phase of PCA analysis comprises of generate a covariance matrix. This is directly derived from the formula for covariance:

$$cov(x,y) = \frac{\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})}{(n-1)} \tag{2.17}$$

It should be noted that $cov(x,y)$ is the same as $cov(y,x)$. From here it is possible to generated the covariance matrix. The covariance matrix C accounts for all covariance values between all different dimensions:

$$C = \begin{pmatrix} cov(x,x) & cov(x,y) & cov(x,z) \\ cov(y,x) & cov(y,y) & cov(y,z) \\ cov(z,x) & cov(z,y) & cov(z,z) \end{pmatrix} \tag{2.18}$$

After the covariance matrix has been generated it is possible to extract both the eigenvectors and eigenvalues from it. In broad PCA terms the largest amount of variation occurs with the largest eigenvalue and corresponding eigenvector. It is possible to then plot these eigenvectors (PCA loadings). It should be noted however that the loadings' positive or negative values are arbitrary and only the magnitude must be taken into consideration. [150]. However, PCA alone can not be considered a complete classification tool for Raman data. This is due not only to class labels not existing in PCA but because PCA generates biases that are uncorrelated: for example experimental errors in the measurement of samples can increase variability between samples in PCA [64].

PCA is also used in conjunction with LDA analysis. This is an accepted form of analysis and has been used in multiple Raman cell studies [63, 64, 141].PCA and LDA are similar in that they both interpret the sample in terms of linear combinations of variables. However, unlike PCA, LDA is a supervised classification model. LDA aims to find boundaries that separate classes within the data. It is then possible to divide the space into these classes [151]. LDA also uses a covariance matrix for multivariate analysis.

Theoretically, linear discriminant analysis (LDA) can be described as projecting data from different dimensions onto an arbitrary line in an arbitrary space. In this thesis the data is considered to have d-dimensional data points that range from $X_1 ... X_n$ belonging to classes $C_{i=1,2...n}$ that each have $N_1, N_2 ... N_n$ samples. The classes in this thesis are represented by each standard prostate cell line and and primary cell culture. It is important to note that LDA makes two assumptions about the data being handled; the data has a Gaussian distribution and that the covariance matrices of each class are identical. To use LDA for classification it is necessary to model the probability density distribution of predictors X for the different categories that the Y variable takes, such that:

$$P(X = x|Y = k) \tag{2.19}$$

This can be further expanded by using Bayes theorem to model the posterior probability [152]. This allows the equation to be expressed in terms of posterior probability such that:

$$P(Y = k|X = x) = \frac{f_k(x)\pi_k}{\sum_{j=1}^{K} f_j(x)\pi_j} \tag{2.20}$$

The posterior probability of a given observation is now expressed on the left hand side of the equation where the k^{th} class can be expressed for X given a specific value of x. The right hand side of the equation can be described as the conditional distribution $f_k(x)$, allowing for calculation of observations in the k^{th} category. To be able to compute a linear score function for $f_k(x)$ it is necessary to assume that the probability density function of x is a multivariate Gaussian. It is also necessary that the probability density function has a class mean denoted m_k. The expression for multivariate Gaussian density function is denoted by the formula:

$$f(x|\pi_k) = \frac{1}{2(\pi)^{\frac{p}{2}}|\sum|^{\frac{1}{2}}} exp\Big(-\frac{1}{2}(x-m_k)^T|\sum|^{-1}(x-m_k)\Big) \tag{2.21}$$

In this equation 2.21 the term $|\sum|$ denotes the determinant of the covariance matrix for all classes. Furthermore the term T denotes the transpose matrix operator. It is then possible to combine the formula for posterior probability (eqn. 2.20) and the equation for Gaussian density function (eqn. 2.21) to derive the linear score function:

$$\delta_k(x) = x^T\sum^{-1}m_k - \frac{1}{2}{m_k}^T\sum^{-1}m_k + log\pi_k \tag{2.22}$$

This formula represents the linear discriminant function and as shown above depends on the x values linearly. This forms the theoretical basis and framework of linear discriminant analysis [153–155].

In this thesis, a cross-validation in the LDA-PCA is also used to evaluate the performance of the model. To make sure that no data is over-fitted, the number of principal components used in the cross-validation is dependent on Kaisers criteria, only PCS with variance higher or equal to or greater than one percent are used. A further classification model is tested using leave-one-out analysis of the PCA-LDA analysis space.

As LDA analysis is supervised, it is important that over-fitting of the data does not take place. For large data sets, to avoid over-fitting, a protocol has been determined. The established protocol restricts the number of principal components to half the number of the spectra in the smallest data set being included in the comparison. This criteria has been used successfully for larger data sets in multiple previous studies [156].

CHAPTER 3

MATERIALS AND METHODS

3.1 Standard cell lines

3.1.1 PNT2-C2 (normal prostate cell line)

PNT2-C2 is a name given to a normal adult prostate epithelial cell line immortalised by transfection with SV40 (simian virus 40) viral DNA, originating from a 33 year old male post mortem [94, 157]. Hierarchically it is the subculture of its parent cell line PNT2. The cells in vitro have an epithelial morphology expressing cytokeratin (CK) 8, 18, and 19 [158, 159]. Cytokeratins are types of keratin proteins that exist in the cytoskeleton and maintain the structure of epithelial tissue. They can indicate what type of cells are present. As cytokeratin 19 is present, PNT2-C2, is also described as having very well differentiated luminal epithelial [160]. Additionally, Cytokeratin 14, is not expressed meaning that basal cells are not present [161]. This cell line was acquired from Sigma-Aldrich, USA.

3.1.2 BPH-1 (benign hyperplasia prostate cell line)

BPH-1 is an immortalised benign prostatic hyperplasia (BPH) cell line originating from a 68 year old man and is not cancerous [162, 163]. Although not explored in this study, population-specific genetic variation is a factor for the development of

cancer and prostate conditions. BPH-1, although considered to be from a "Japanese donor" has been found to have a large European genetic component of 95% while only expressing a 4% Asian content [164]. BPH is the most common symptomatic neoplastic condition in humans [165]. The BPH-1 cells were immortalized with SV-40 large T-antigen [166]. BPH-1 expresses cytokeratin 7, 8, 18 and 19 consistent with well differentiated luminal cells. Additionally, Cytokeratin 14 is not present in BPH-1 cells, indicating that basal cells are not present. The BPH-1 cells used in this study were obtained by Professor Norman Maitland at the Cancer Research Unit (CRU), University of York, York, UK, with kind permission from Simon Hayward, director of Cancer Biology, NorthShore University HealthSystem. BPH-1 cell lines are also commercially available.

3.1.3 LNCaP (metastatic cell line)

LNCaP is a cancerous prostate epithelial cell line established from a 50 year old Caucasian male in 1977 [49, 167]. Although the cell line is representative of cancerous prostate epithelial cells it is derived from a metastatic site, the left supraclavicular lymph node [168]. LNCaP is representative of androgen-sensitive human prostate adenocarcinoma cells [169]. Androgen-sensitive means that although the prostate cancer cells may react to the steroid hormone androgen they do not require it for growth [170]. LNCaP contains cytokeratins 8 and 18 which indicate that it consistent with luminal cells[171]. As cytokeratin 14 is not present it can be assumed no basal cells are present. LNCaP is a commercially available cell line and was purchased from the American Type Culture Collection (ATCC; Manassas, VA).

3.1.4 DU145 (metastatic cell line)

DU 145 is a cancerous prostate epithelial line originating from a 69 year old Caucasian male with a three year history of lymphocytic leukemia with widespread metastatic carcinoma of the prostate [172]. DU145 is characterised as being androgen receptor (AR) negative meaning it is not hormone sensitive [172, 173]. This human prostate cell line is derived from a metastatic site, being a brain

lesion [172]. DU145 is a commercially available cell line and was purchased from American Type Culture Collection (ATCC; Manassas, VA).

3.1.5 PC3 (metastatic cell line)

PC3, established in 1979, is a metastatic human prostatic adenocarcinoma cell line that is derived from the bone [174, 175]. The site of origin is bone metastasis from a 62-year old-caucasian man with undifferentiated adenocarcinoma, gleason grade IV (4) [174, 176]. PC3 does not express any androgen receptors, making it an androgen independent cell line [177]. Additionally, PC3 does not display prostate specific antigen [178]. PC3 exhibits cytokeratins 8 and 18 consistent with luminal cells and did not express any basal cell cytokeratins [179]. Immunohistochemistry (IHC) staining analysis of cytokeratin 8 showed focal positive staining that would indicate that PC3 is representative of small cell neuroendocrine carcinoma [177]. PC3 is a commercially available cell line and was purchased from American Type Culture Collection (ATCC; Manassas, VA).

3.1.6 P4E6 (localised cancer cell line)

P4E6 is an immortalised human epithelial prostate cell line established from well-differentiated prostate cancer (gleason 2+2).P4E6 was immortalised with human papillomavirus type E6 gene [180, 181]. After the cell line had been immortalised with the HPV E gene, it was designated PxE6; x represents the patient number thus becoming P4E6. Patient 4's primary culture was obtained using a transurethral resection of the prostate (TURP). The TURP acquired culture, representing a level 4 (2+2) Gleason cancer [182]. P4E6 retains both PSA and AR expression [183]. This cell cell line was derived at the Cancer Research Unit (CRU), University of York. It is now commercially available.

3.1.7 Reasons for selecting these specific cell lines

The main reason for selecting PNT2-C2, BPH-1, LNCaP, PC3, P4E6 and DU 145 is that they are used widely in prostate cancer research and related studies. They are representatives of normal prostates and prostates that cover a spectrum of

disease, from benign tumour to varying cancerous states. Further interest is seen in the volume of material surrounding certain cell lines; from 2000 to 2001, there 125 papers whose titles included the cell line LNCaP that were published [44]. The high volumes of studies utilising these cell lines does in turn mean that reproducibility is increased and maintained. The use of a SV40-immortalised cell line to represent the normal cells in this study were chosen as they are already regarded as being suitable for the study of carcinogenisis through previous studies of epithelial cell line comparisons [184]. However, several studies have warned that the standard cell lines immortalized with SV-40 large T-antigen could differentiate from the original sample. The SV40-induced immortalization could sequentially lead to loss of DNA damage response, instability, tetraploidy (doubling of DNA content) and in some cases tumorigenicity [180, 184]. The incorporation of P4E6 was therefore included due to the aforementioned issues highlighted by previous publications with respect to SV40-induced immortalization, P4E6 does not undergo SV40-induced immortalization.

3.2 Primary cell cultures

The primary cell-cultures that are analysed in this study come directly from patients. These cell cultures are not immortalised by any plasmid, and thus have a limited lifespan. They are fully representative of the patients' cells. Patient tissue samples from the prostate were collected with ethical permission from Castle Hill Hospital (Cottingham, Hull) (ethics number: 07/H1304/121). The surgeon collaborating in this study was Mr. Matt Simms. Tissue samples were received from him and from several other urological surgeons. The biopsies are based on previous MRI and diagnostic biopsy results. Mr Simms removes the whole prostate through a radical prostatectomy procedure. Mr Simms then palpates the tumour before a needle biopsy is taken from it. A needle biopsy is then taken from an area of the prostate that is tumour-free that is labelled as normal. A dedicated tissue procurement officer based at the hospital retrieves the samples for us directly during or after surgery.

3.2.1 H646/17

H646/17 originates from a biopsy taken from a 67 year old man. Through analysis of biopsy the malignant tissue was graded with gleason score of 7(3+4). This means that while most of the cancer cells in the biopsy have a slow rate of growth there is a percentage that have a moderate growth rate. Normal tissue was also procured from the biopsy. The malignant tissue was removed from the left side of the prostate and labelled "H646/17L". The non malignant normal tissue was removed from the right side of the biopsy and labelled "H646/17R". It should be noted that these scores are calculated from a biopsy sample and thus could miss a more aggressive growth rate that might be present in the sample.

3.2.2 H722/18

H722/18 originates from a biopsy taken from a 69 year old man. Through analysis of biopsy the malignant tissue was graded with gleason score of 7(3+4). This means that while most of the cancer cells in the biopsy have a slow rate of growth there is a percentage that have a moderate growth rate. Normal tissue was also procured from the biopsy. The malignant tissue was removed from the left side of the prostate and labelled "H722/18 L". The non malignant normal tissue was removed from the right side of the biopsy and labelled "H722/18 R". It should be noted that these scores are calculated from a biopsy sample and thus could miss a more aggressive growth rate that might be present in the total sample.

3.2.3 H794/18

H794/18 originates taken from a biopsy from a 71 year old man. Through analysis of biopsy the malignant tissue was graded with gleason score of 7(3+4). This means that while most of the cancer cells in the biopsy have a slow rate of growth there is a percentage that have a moderate growth rate. Normal tissue was also procured from the biopsy. The malignant tissue was removed from the left side of the prostate and labelled "H794/18 L". The non malignant normal tissue was removed from the right side of the biopsy and labelled "H794/18 R". It should be

noted that these scores are calculated from a biopsy sample and thus could miss a more aggressive growth rate that might be present in the total sample.

3.3 Gleason grading

The current standard used to effectively grade prostate cancers is called a Gleason score. This was pioneered by Donald F. Gleason in 1966 [185]. The Gleason system is constantly developing as more research is carried out on biopsies. Changes were made to the system in 2005 and further updated in 2010 by Epstein due to changes in the way urological pathologists analyse prostate cancer and the radically different techniques available to them [186, 187]. Following an international conference lead by Epstein in 2014, further updates were made discerning what specifically constitutes each grade [188]. It is worth noting that there is another metric used to grade prostate cancers, using Whitmore-Jewett stages, although this is not common practice [189, 190].

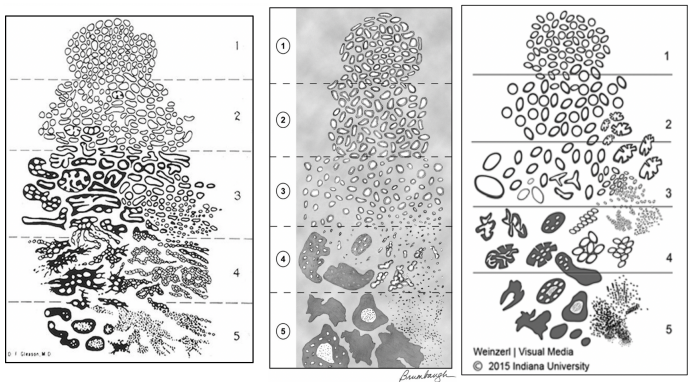


Figure 3.1: The left image represents the original morphological Gleason grading system for defining the appearance at the stage of prostate cancer defined by Donald F Gleason in 1966 [185]. The middle image corresponds to the 2010 update by Epstein [186]. The image on the right is respective of the last update by Epstein *et al.'s* study in 2015 [188]. However, these changes relate to the grading of the morphology of the prostate tissue itself, the grading is still graded 1 to 5.

The updated Gleason grading system effectively combines clinical staging and histological grading to predict the malignancy of the tumour in its current state. The morphology of the cells in the prostate are observed through needle biopsy or radical prostatectomy (RP) (total removal of the prostate gland and attached

seminal vesicles). Here, Gleason grades according to morphological appearance of prostate cells: (1) The prostate glands appear normal in size with uniform distribution and are well differentiated; (2) Prostate glands are well formed but beginning to show signs of neoplastic behaviour. There may be an increase in stroma and connective tissue; (3) The tissue still has recognizable glands, but they may appear smaller or larger. They must not form chains or cords but may appear as irregular shapes with ill defined edges. No necrosis is present; (4) The tissue has few recognizable glands. Many cells are invading the surrounding tissue. The cells may appear as cords or chains. The cells can appear much larger than normal; (5) The tissue in the prostate glands is no longer recognisable. Ragged sheets of cells may span surrounding tissue [186, 188, 191, 192]. To summarise, Grades 1 and 2 are non-cancerous while 3,4,5 are malignant.The grades are added together for a final gleason score as it is normal for more than one stage to be present in the prostate. The higher the Gleason score, the more likely the cancer is to spread outside of the prostate [193]. The Gleason score is calculated using the equation:

$$\text{Gleason score} = \text{the most common grade} + \text{the highest other grade in the sample} \tag{3.1}$$

As seen in equation 3.1 the first number of the Gleason score represents the most common grade in the sample and the second number represents the highest grade of what is in the sample. The initial Gleason scores that are assigned in this project come from the diagnostic biopsies that are obtained a few months before surgery on the patients. They are assigned by the pathologist, Dr. Greta Rodriguez who is based in the hospital (Castle Hill Hospital, Cottingham, Hull).

3.4 Cell culturing methods and Raman sample preparation

There are many types of ways to prepare cells for Raman spectroscopy and these have been documented extensively in literature. When a tissue biopsy is taken it can be stored in a certain number of ways. Common procedure dictates that the tissue sample be flash frozen, formalin fixed or paraffin embedded [194–196]. However, it is accepted that chemical fixation methods such as using formaldehyde for preservation result in the structure of the DNA exhibiting sequence alteration [197]. Furthermore, formaldehyde fixed samples are not only shown to result in the disintegration of hydrogen bonds but the denaturing of intracellular proteins between the primary and secondary amine groups [198, 199]. This in turn can lead to the loss of integrity of the quaternary and tertiary structure of the proteins [200]. Techniques such as chemical fixation methods therefore could result in the perturbation of key bio markers such as established Raman vibrational correlating to DNA, specific proteins and lipids , which in studies have shown strong variations and perturbations distinguishing prostate cancer cells from normal cells, generating a false positive [201].

Therefore, it is essential that no damage or alteration is made to the DNA protein structure of the cells being analysed in this project due to the nature of it relying on fine changes in biochemical pathways to explain differences in tumoricity and morphology between cell lines and primary cultures. Specifically, as these biochemical pathways are being investigated for the use for diagnostic and prognostic bio markers. However, previous studies have shown that the least damaging and preferred method of tissue preservation be flash freezing [202]. The method of flash freezing has been used in previous studies and allowed the successful extraction of high quality cell line and primary cell tissue DNA [203]. It is for this reason that this project utilises cell lines and primary tissues that are flash frozen.

3.4.1 Culturing of the standard cell lines

All of the standard cell lines were grown in T75 tissue culture flasks. Each cell line grows in a specific cell culture media. Some of the cells were resuspended in R10 (RPMI plus 10 percent FCS) to inactivate the trypsin, counted and then spun down and resuspended in their specific growth media. It is important that the confluence is consistent in each cell sample, where confluence defines the percentage coverage of the cells on the calcium fluoride discs. A confluence of 50% therefore translates to 50% coverage [204]. Another variable that must be kept constant is the number of cell passages, defined as the amount of times a cell line has been re-plated and allowed to grow back to the required confluence. It is important that the number of passages that takes place is as small as possible as the more passages or re-culturing steps that take place, the further the cells are away from the original cells being examined. If these factors are not monitored, the cells will develop biochemical changes with the initial desired biochemical markers no longer able to be measured. The aim of this study is to measure cell cultures as close to the original patient as possible.

3.4.2 Culture of primary prostate epithelial cells derived from patient tissue

A full ethical agreement is signed off and permission from the patient is obtained before tissue is retrieved for experimental analysis. The full method of processing patient tissue to retrieve primary prostate epithelial cells is described in depth [205]. A summary is therefore provided here. The patient tissue is initially left in a collagenase enzyme overnight to break down the tissue. To achieve a single cell suspension, the tissue must also be broken down further in trypsin solution. The reason for adding trypsin is that cells have adherent molecules that stick to the surface and trypsin splits these connections. The single cells are then cultured on collagen I-coated tissue culture dishes mimicking the basement membrane. cells are grown with STO (mouse fibroblast) feeder cells as these simulate natural prostate stromal cells. The STO cells have been irradiated and thus will die out as the epithelial cells grow. The primary cell culture is grown in a different medium

than the standard cell lines. The patient cells are grown in a stem cell media (SCM), which is absent of any foetal calf serum that is used for the standard cell lines. When confluence is achieved, the cells are split, some are frozen down for later use and analysis, and the rest are put back in culture. It should be noted that all cells analysed in this study were between passage 0 and passage 4. This is important because the cells must remain as close to the original patient sample as possible. It is also noted that patient samples stop growing after passage 3-4.

3.4.3 Passaging of standard cell lines and primary cell cultures

Passages refer to the number of times a cell culture is subcultured, a process by which cells are taken from a parent sample and regrown in daughter cell culture flasks. It is accepted that when cells are trypsinized for freezing this also counts as a passage, due to the transfer of cells to a different culture dish. It should be noted that the passages do not only specifically attribute to the population density or growth of the sample, but can also be a measure of population doubling. However, it should be noted that cell lines that are subjected to higher split ratios will inherently be more generations away from the original sample. It is for this reason that there are strict protocols for this project and what confluence and cell density are allowed before each passage, as there is no way to measure what passage a cell line is at without such records. Visual inspection of the cells was also undertaken to make sure no morphological differences to the parent cells had taken place. Separately, all patient primary cell cultures were not passaged and directly cultured from a frozen tissue sample. Furthermore, the immortalised commercially available cell lines were produced specifically for each measurement and not cultivated after their use had expired. This, although labour intensive, was imperative so that the cell lines were representative of their parent sample.

3.4.4 Counting cells

All samples were assessed visually before Raman analysis, and throughout the culturing process. This was carried out using a haemocytometer, a specialised

microscope slide used to count cells. It has a thick base and special coverglass which is thick enough not to be distorted under the pull of the surface tension from the solution.

Cells were counted as resuspended cells after being trypsinised. This was done by adding 20 microliters of cell solution to haemocytometer before counting cells in each of the four corner squares. An average of the four squares was then taken. The average for the four squares was then multiplied by 10,000 to get the number of cells per ml.

3.4.5 Preparing cells for analysis using Raman spectroscopy

The initial step is to remove the media from the culture dish and to wash the cells in PBS (phosphate buffered saline) solution. To lift the cells off the plastic surface of the culture dish, trypsin must be added. The cells are then re-suspended in R10 and can then be counted. For live and dry cell Raman measurements, the cells need to be re-plated and analysed on a calcium fluoride disc. The CaF_2 discs are Raman grade of size 13mm diameter by 1mm thickness. They are used as CaF_2 does not have a Raman signature in any of the wavenumber regions of interest for cells, as well as their minimal background interference [111]. The CaF_2 is placed in a 35mm tissue culture disc with 50,000 cells (in 200ul) pipetted onto it. The CaF_2 was left for 10 minutes to allow the cells to semi-adhere, then 2.5ml of cell media was added to the dish together with 25ul of ABM (antibiotic, antimycotic) since the CaF_2 discs are not sterile.

After a 24-hour period, the media is removed and 2.5ml of starvation media plus ABM is added. Starvation media is the base media that the cells are grown in without any FCS. This is used to synchronise the cells. After another 24-hour period, the CaF_2 disc plus cells is then washed twice with HBSS. Following this, 2.5ml of fresh HBSS was added to the cells then taken to the Raman laboratory, for analysis. For dried cells, the CaF_2 disc and cells are washed with HBSS before removing the HBSS and leaving the disc to air-dry. For live cell measurements

in the Raman laboratory it was necessary to add more HBSS due to any spillage that could occur during transportation, as well as to have enough HBSS for the x60 dipping lens to operate in. The addition of HBSS to the 35mm tissue culture dish containing the CaF_2 slide, plus cells was achieved in the Raman laboratory and using sterile pipettes .

3.5 Raman spectroscopy measurements

3.5.1 Raman spectrometer set-up

Using HORIBA's LabSpec6 software, samples were first viewed using white light microscopy in video mode before switching to the Raman microscopy mode, to target specific constituents of cells, for example the nucleus. LabSpec6 software comes pre-loaded with the instrument and works in conjunction with the PC attached to the machine. The micro-Raman spectrometer that was used in this project is the HORIBA XploRA. The laser wavelength used in this study was the 532nm. An image of the HORIBA XploRA is shown in Figure 3.2. A further simplified schematic is provided to detail internal components of the Raman Spectrometer in Figure 3.3.

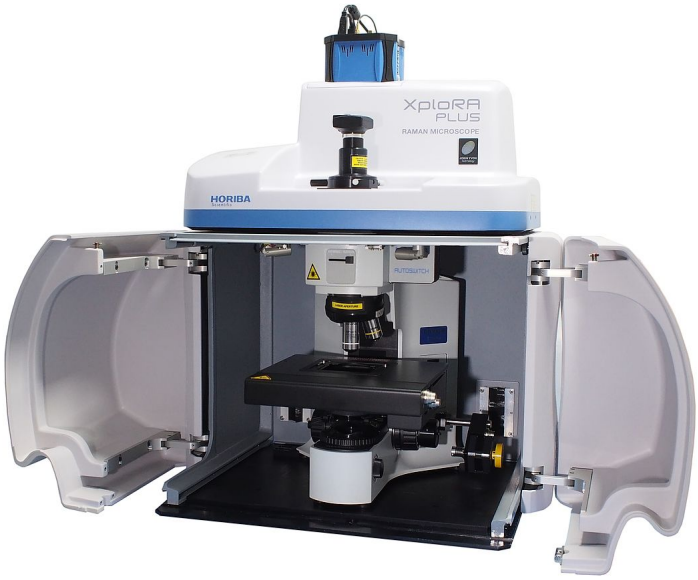


Figure 3.2: Horiba XploRA PLUS Raman spectrometer. As shown there are interchangeable lenses. The x100 Olympus Mplan lens with numerical aperture 0.9 was used to analyse dried cells and for calibration of the machine. A x63 Zeiss Optics dipping lens with numerical aperture 1 was used to measure cells in solution. Cell selection is possible due to a live video feed and internal lighting features. Focusing is achieved through use of the x-y-z motorised stage. The doors of the instrument are used to not only block out any cosmic rays but as a safety precaution due to the power of the laser; the laser will not operate unless the doors are closed. The instrument operates as a confocal microscope when the slit is set to $200\mu m$ and hole to $100\mu m$.

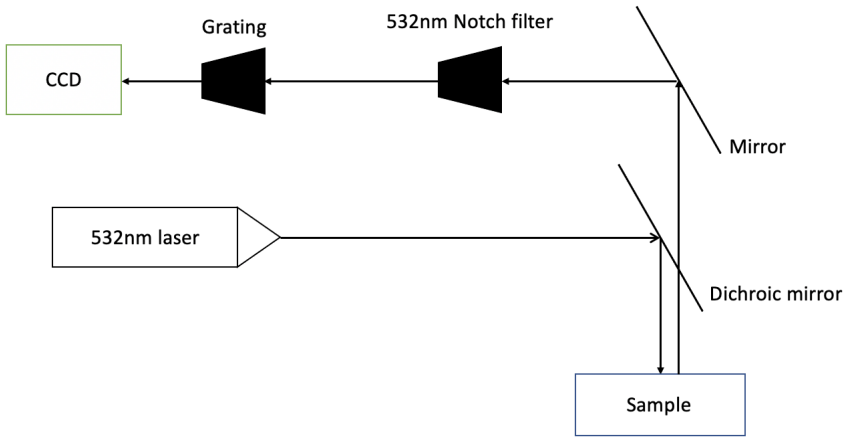


Figure 3.3: Simplified Schematic of Raman spectrometer used in this thesis. As shown, the laser is emitted before interacting with a dichroic mirror. In this thesis a density filter was also included to select the correct power setting. The laser interacts with the sample before passing back through the dichroic mirror and being reflected into the notch filter. The resulting laser beam then passes through a grating before reaching the CCD. This will generate a spectra on the computer using lab installed software.

The Raman spectrometer is calibrated daily using a silicon standard which has one very defined peak at 520 cm^{-1}, meaning that any bands that have shifted can be corrected easily [206]. For the calibration and dry cell measurements the x100 Olympus Mplan lens was used, which has a numerical aperture of 0.90. For live cell measurements a x63 Zeiss Optics dipping lens is used, which has a numerical aperture of 1. Figure 3.4 shows the experimental configuration for dry cell measurements versus the live cell measurements.

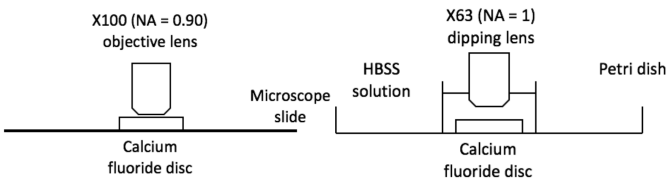


Figure 3.4: Dried cell measurement set-up (on the left) and the live cell measurement set-up (on the right. For the dried cell measurements a X100 lens is used. The cells are plated and then air-dried on the CaF_2 disc. In live cell measurements a X63 dipping lens is used. The cells are plated on a CaF_2 disc in a HBSS solution. The HBSS is a cell medium designed to keep the cells alive throughout the Raman experiment [207].

3.5.2 Parameters used in Raman spectra acquisition

The Raman spectroscopy measurement parameters for prostate cells has been established previously in group, due to Charles Kershaw's MSc thesis, in Dr. Y. Hancock's lab [208]. The parameters used for the collection of Raman spectral data are identical for the live standard cell lines, dried cell lines and primary cell cultures and are detailed in table 3.1. These parameters are set using LabSpec6 software which drives the instrument. Instead of taking a continuous spectrum from 0 to 3100 cm^{-1}, the spectra recording was split into two wavenumber regions to exclude unnecessary information and reduce acquisition time. Therefore, the two specified regions chosen were from 600 to 1800 cm^{-1} which is the fingerprint region, and from 2700 to 3100 cm^{-1}, which is the high wavenumber region. For live standard cell lines and primary cell cultures only one spectrum per cell is taken from a random position in the nucleus to limit cell damage. However, in dried cells five spectra are taken from random positions in the nucleus due to the larger number of artefacts caused by cell drying techniques.

Table 3.1: Parameters used on the HORIBA XploRA Raman spectrometer for whilst taking Raman spectra of air-dried and live cells.

Parameter	**value**
Acquisition time	45 seconds
Accumulations	2
Range Fingerprint	600-1800 cm^{-1}
Range High Wavenumber	2700-3100 cm^{-1}
Objectives	x100 or x63
Grating	2400T
Filter	50%
Slit correction	200 μm
Hole correction	100μm

The max laser power is stated as being 7.5 mW. At the sample the laser will have a max power of is 3.75 mW due to the 50% filter. The relatively long acquisition time of 45 seconds on the sample is due to a weak Raman signal from the prostate cells. Hence, the exposure time needs to be long enough to collect high quality spectra, but without causing photodegradation. The number of accumulations is set at 2 to take an average spectra ruling out any abnormalities. The LabSpec6 software is used to view the spectra after each accumulation, so

further observations can be made to ensure that the cells are not being damaged during measurements. The live video feed using white light microscopy can also be used to check that the appearance of the cell has not become deformed in any way during Raman measurement. It should also be noted when analysing live cells and primary cell cultures, both are only viable in the lab for 4 hours; multiple experiment days are required to determine a sufficient number of spectra.

3.5.3 Confocal Raman spectroscopy and laser spot size

Confocal spectroscopy rejects the residual background signals from outside the focus of the laser beam and the focal plane. This is done by using a pinhole in the spectrometer, reducing the collimation of the beam, allowing depth resolution [111]. The confocal setting on the Raman instrument can be achieved by using the LabSpec6 software to set the pinhole aperture to 100 μm and the slit size to 200 μm. The slit size controls the resolution of the image; the narrower the slit size the higher the resolution. The pinhole is used as a conjugate point with the focal point of the objective lens of the microscope, forming an image where the pinhole is. This blocks out fluoresce from the rest of the sample outside of the focal point. This means it is possible to measure certain characteristics of the cells whilst minimising background autofluorescence and reducing the risk of sample damage [209]. The diffraction-limited size of the laser spot can be calculated using:

$$laser\ spot\ size = \frac{1.22\lambda}{NA} = \frac{1.22 \times 532nm}{0.90} = 721nm \tag{3.2}$$

where λ is the wavelength of the laser and NA is the numerical aperture of the objective lens. Using the lens Na and the lenses wavelength in the experiment the laser spot is determined as 721nm in diameter. The benefits of confocal Raman spectroscopy can be summarised as having a high sensitivity to be able to measure the weak Raman signal and high lateral and depth resolution allowing the probing of the cells in the study [111]. One further consideration is that the spectral resolution is limited by the choice of diffraction grating. To ensure the highest resolution, all spectra were taken with the 2400T grooves density(lines/mm)

grating.

It has been confirmed by previous studies that confocal Raman spectroscopy has been able to successfully characterise both individual dried cells and living cells [210]. The technique allows depth measurements and the measurement of the chemical compounds inside the cell itself.

3.5.4 Assessing viability of cells

Even though Raman spectroscopy is labelled as a non-destructive method of analysing cells, it should be noted that this is only possible under the correct parameters. To establish that the Raman spectrometer settings are indeed non-destructive, a test must be done to ensure that the structure of the cell is not perturbed or perforated. We used trypan blue assay test to confirm cell viability. In a trypan blue assay test, any dead cells uptake the trypan blue and can be distinguished from the live cells which exhibit no colour [211, 212]. The cells viability and the validity of the raman acquisition parameters was carried out in our own lab by Charles Kershaw and Dr. Y Hancock [208].

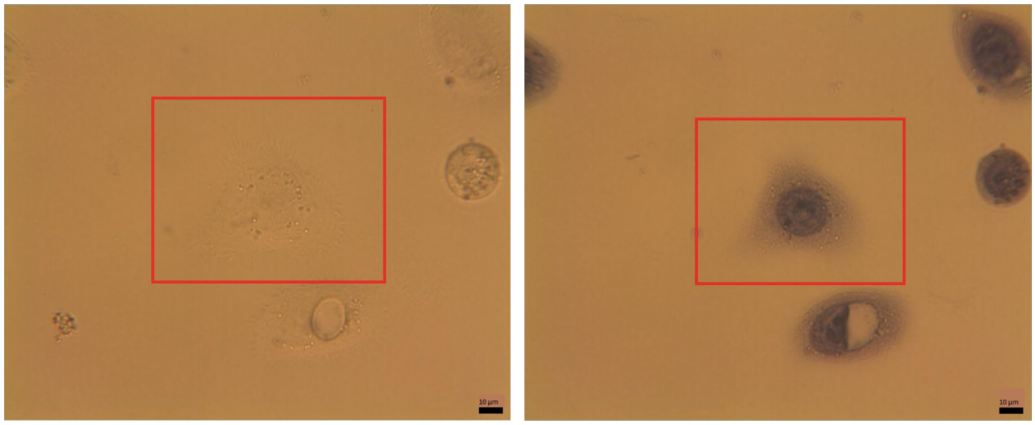


Figure 3.5: The left image represents a P4E6 prostate cell that has undergone Raman measurement using acquisition protocols in this project. The right image is a cell showing to have taken up trypan blue dye after the sample was washed with ethanol to kill the P4E6 cells cite [208].

3.5.5 Acquisition of Raman spectra data on prostate cells

Raman spectral data for the live cell lines PNT2-C2 cell and P4E6 were previously obtained by Charles Kershaw, a former MSc student in Dr. Y.Hancock's lab at the University of York [208]. Raman spectral data for the live cell BPH1 line and the dried PNT2-C2 cell line were obtained in collaboration with Gemma Davison, an MPhys student in Dr. Hancock's lab at the University of York. The Raman spectral data for PC3 was also individually taken by Gemma Davison. Raman spectral data for the live cell line LNCaP was also taken by myself in Dr. Hancock's lab, in the University of York. Furthermore, three patient cell-cultures H646/17, H722/18 and H794/18 were used in this study with these being sourced from the Cancer Research Unit (CRU) at the University of York. The Raman spectra of all patient derived primary cell-cultures were also taken by myself in Dr. Hancock's lab at the University of York.

3.5.6 Removal of outliers from data set

Outliers in statistical data are generally described as coming from two distinct sources. One is described as inherent variability whereby even in this case, if all Raman measurements were performed perfectly on the cell lines, there would still be variability of stature leading to what appear to be inconsistencies in the results [213]. The second source of error is instrument error [213]. In this project the Raman instrument has an associated error and mechanically decides where to start and end measurement, with a certain number of data points in each spectra. These should be consistent, but in fact there is variability among cell lines and even rare instances in cell lines themselves. Another factor that must be considered is outliers in sampling, whereby it is possible that some of the cells measured are not in fact the same population as the rest of the sample [214]. This is certainly more likely in the case of patient cell lines where there is an exponential increase in the heterogeneity of the sample. It is for this reason a distinct tangible criterion was established so that cell lines and further statistical tests would not be contaminated by extreme outliers [215]. The first reason for exclusion of data was if the length (number of data points) of the data was

different to that of the data set or if the instrument had inadvertently stopped mid measurement. The second reason for exclusion was if the spectra had no signal, producing a flat line with no features. The third reason was if the spectra resembled the signal of cytoplasm, low intensity and lack of peak information. The signature for Raman signals for cytoplasm was cross referenced with multiple studies [64, 216, 217]. If no clear reason could be found for excluding data the result was not excluded.

Whilst this study does exclude measurements of cytoplasm it should be noted that other studies in the field have included cytoplasm in Raman studies of cells. However, these studies were attempting to understand the whole cell through a Raman mapping technique or focusing on the lipid composition of cytoplasm [218]. It is not to say that no useful information can be extracted from the cytoplasm but that this study is only focused on the nucleus and its environment. This is due to the nucleus storing the information pertaining to regulation and proliferation [219]. Simply, the cytoplasm lacks the DNA and genetic information to probe these elements using our technique [212, 220]. The signal generated by cytoplasm can be found in Chapter 4 Figure 4.1 and is compared to the signal of the nucleus for reference. For cross comparison between cell lines it is therefore vital that only signals generated by the nucleus are included in comparisons. The identification of spectra representative of cytoplasm was through analysis of PCA and a visual profile comparison to other spectra generated by the cell line using the spectral profiles of each data point.

3.6 Raman spectral analysis

3.6.1 Statistical convergence testing

After collecting spectra for each cell sample it is necessary to make sure that the average spectrum and statistical quantities, twice the standard deviation and standard error, have converged, thus ensuring that the necessary number of spectra have been taken.

The convergence tests were carried out in multiple iterations. A random selection of 10 spectra from the data set is taken before being averaged, whilst at the same time calculating the standard error and twice the standard deviation. For example, the convergence test is carried out until the maximum number of spectra is being used e.g 10 spectra, 20 spectra, 30 spectra, 40 spectra, 50 spectra ... MAX spectra. The corresponding values for the standard error and twice the standard deviation are measured whilst the average is being taken.

It was decided that the convergence of the sample be determined by the local convergence at the peak position 1001 cm^{-1} for all statistical tests in the fingerprint region due to it being a very well defined sharp peak. The peak position 2934 cm^{-1} was used to determine convergence of the samples in the high wavenumber region due to it being the most relatively intense, well defined peak in the high wavenumber region. Convergence was achieved when the iterative steps of the increase spectra lead to no improvement in the associated error of the spectrum. This cut off, for determining a diminishing return, was deemed to be less than a 1% improvement in error between multiple iterations of profiles.

3.6.2 Univariate analysis

Univariate analysis, such as determining two peak intensity ratios (PIR), involves fitting of the average spectral profile for each sample thereby quantitatively characterising it. The first step in this characterisation is to fit each peak or band region of the statistically converged average spectrum, to obtain the peak positions. These peak positions are then compared to relevant literature to determine corresponding biomolecular information. There are several methods that can be used to fit Raman spectra. However, samples such a prostate cells will exhibit varying degrees of fluorescence and background signal. This fluorescence is countered by using a baseline subtraction. It should be noted that there is still sufficient Raman signal after the baseline to yield strong results. In the case of this thesis, a first order linear polynomial was used to baseline the fingerprint region 600-1800 cm^{-1} and a third degree cubic polynomial was used for the baseline in the high wavenumber region 2700-3100 cm^{-1}. There are many types of software

that can be used to fit Raman data. This project used a combination of IGORPRO7 and IGORPRO8 software to carry out peak fitting analysis. The fitting of peaks is carried out in piece-wise spectral regions that have been baselined; IGORPRO software allows for the user to input specific x-coordinates to allow automated fitting in the desired range. In these regions are wavenumber ranges associated with biomolecular information e.g., DNA, lipids or proteins (see Figure 6 for the fitting regions used in the study; spectra has not been baselined and was not used for analysis). The fitting criterion decided upon was using the minimum number of peaks fitted to obtain a good fit. This avoids over fitting the spectra and also assists with reproducibility.

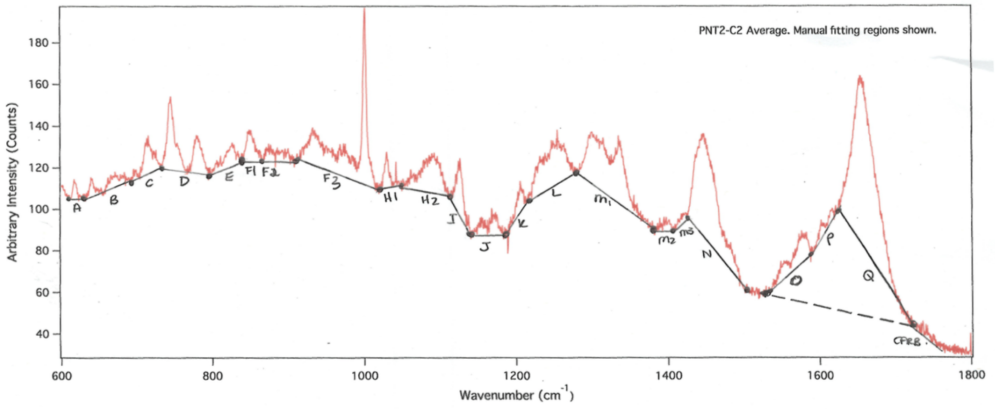


Figure 3.6: A plot showing the fitting regions being mapped onto a normal PNT2-C2 cell line fingerprint region.

Once the average spectrum of a cell line or primary cell culture has been broken down into these specific fitting regions, they are then fitted using the software IGOR PRO 8 using Gaussian profiles. This is done by the previously aforementioned auto-locate feature. The reason for using the auto-locate feature is to increase reproducibility in the peak fittings. When a band has been fitted correctly, it outputs the following information: peak location, peak width, peak height and peak area. An example of band fittings in IGOR PRO 8.

From the extracted data from the peak fittings it is possible to construct peak intensity ratio (PIR) graphs for each cell sample relative to key biological markers. An example of a marker would be the DNA peak 725cm^{-1} to which all other peak intensities can be compared. Changes in relative peak intensity

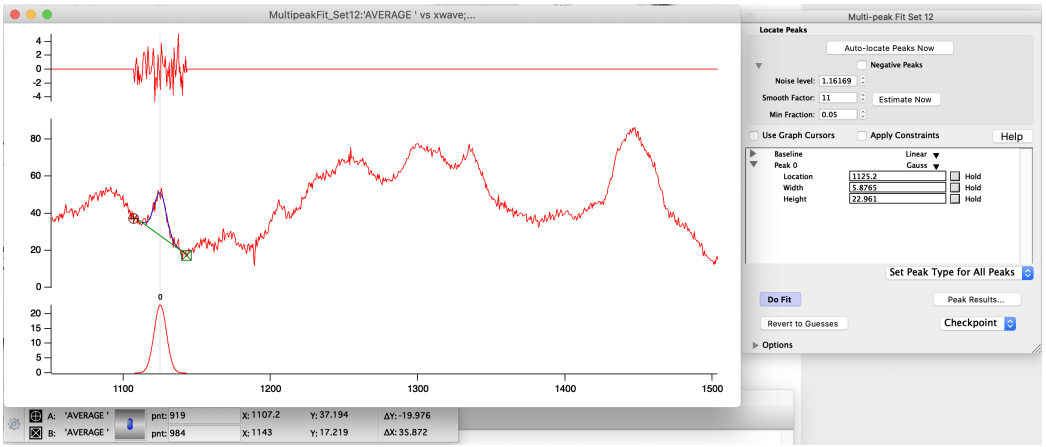


Figure 3.7: This image shows the fitting of region for the live standard cell line PNT2-C2. Specifically, the average of live the PNT2-C2 cell line.

ratios can then also be compared between cell lines [221]. The errors associated with the peak amplitude and subsequently the ratios calculated thereafter are deduced by adding and subtracting the standard error. This is done by fitting the sample from the average spectrum with the standard error of the spectra added or subtracted, before carrying out peak fitting to determine the intensity difference to the original average. The maximum error is then chosen as the error for that peak. These errors in amplitude are then propagated to determine peak intensity ratio uncertainties. The peak intensity ratios of each cell line are then calculated and are compared across other cell lines/ primary cell-cultures to assess what quantitative biomolecular differentiation can be made to discriminate between them. In this study peak intensity ratios of both the fingerprint and high wavenumber region are analysed.

3.6.3 PCA and LDA analysis

To perform principal component analysis (PCA) or linear discriminant analysis (LDA) on the Raman data sets, several pre-processing steps for the spectra must occur. The first step is to cut the spectra using the Raman Tool Set application [143]. This is done to make only the relevant parts of the spectra containing biological information are being analysed. For PCA and LDA analysis, all data sets must be contain the same number of data points which is ensured by interpolating the spectrum using IgorPro7. A linear baseline subtraction in the fingerprint

region or polynomial baseline subtraction in the high wavenumber region is then performed on each individual spectra using the Raman Tool Set, which allows removal of the background signal from the data [142]. A maximum area normalisation with respect to the total spectrum area is applied. The data is then smoothed using a cubic spline, again using Raman Tool Set application. The pre-processed data is then inserted into comma separated value tables in Excel, then loaded via PCA software into R studio [222] which is a statistical software. The PCA and LDA code outputs several graphs including 2D and 3D scatter plots. Both are shown in this work. The graphs were created using several R Studio packages [223].

As stated in Chapter 2 section 2.2.2 it was important to ensure that over-fitting did not occur during PCA-LDA analysis for large data sets. The established protocol was to restrict the number of principal components to half the number of the spectra in the smallest data set. This meant that for comparisons including the standard cell lines a maximum of 59 principal components can be used, the standard cell line with the least spectra is BPH-1, consisting of 118 spectra. For the patient primary cell culture PCA-LDA comparisons a maximum of 60 principal components can be used, the patient primary cell culture with the least spectra is H722/18L, consisting of 120 spectra.

CHAPTER 4

LIVE VERSUS DRIED, STUDY OF PNT2-C2

Raman spectroscopy can be performed on both live and dried prostate epithelial cell lines. In this chapter Raman spectroscopy has been carried out on a live and dried version of the same cell line, PNT2-C2. PNT2-C2 is a name given to a normal adult prostate epithelial cell line. To detail the differences between the Raman spectra of live and dried, this chapter has used Raman spectroscopy to fully characterise both the live and air-dried samples. Furthermore, convergence analyses was carried out to determine at what number of spectra the sample converged, namely: the spectral average, twice the standard deviation and the standard error. This Chapter further uses a mean profile subtraction and principal component analysis to determine which Raman bands and their respective biomolecular assignments capture the most variance between PNT2-C2 live and dried. In addition peak intensity ratio analyses are used to determine bands responsible for separation and for biomarkers that can differentiate the live and dried state outside of the propagated SE uncertainty relating to each peak intensity ratio.

4.1 Raman Spectroscopy: live versus dried, study of PNT2-C2

At present there are a high number Raman of studies that use prostate standard cell lines to carry out research [63, 64, 111, 141, 224]. However, there is a disparity in the way that standard cell lines are analysed in terms of their state; various cells are measured using different fixation methods [111]. Many studies, for reasons such as reduced costs or the lack of other resources, opt for the use of dried cells instead of analysing the cells in their live state [225]. When cells exist in a hydrated state, the water molecules are bound to other biomolecules, providing not only structural support for the cellular structures but are also imperative to cell processes [226]. In the case of prostate cells, the water molecules are bound to the DNA, proteins and lipids. There are still gaps in knowledge as to what effect the drying process has on the Raman spectra of prostate cells; some have argued that the drying process leads to conformational change in DNA . Multiple studies have given conflicting evidence on what happens not only morphologically but biochemically during the drying process of cells. This study explores key biochemical changes that occur during the drying process through comprehensive analysis of the cell line PNT2-C2 in its live and dried state.

A study carried out by Notinghger *et al.'s* (2002) investigated the difference between live and dead cells [212]. It compared the spectra generated by live and dead MLE-12 cells, immortalised lung epithelial cells derived from lung tumours in transgenic mice [227]. The study by Notingher *et al.* (2002) suggests that 785nm is the optimum laser wavelength for the purpose of measuring cells and suggests that using shorter wavelengths such as 488nm and 514nm leads to morphological changes in cells [212]. The conclusion statements made by Notingher *et al.* (2002) assume a laser power of 5 mW at the sample and an acquisition time of 5 minutes [212]. Conversely to Notingher *et al.* (2002), methodology used in this thesis demonstrated that under the correct conditions, a 532nm laser of 7.5 mW can be used with a 50% laser filter, generating 3.5 mW at the sample to probe cells without leaving damage [208]. The acquisition time of the 532nm laser is set to

45 seconds, much less than the 5 minute acquisition time used by Notingher *et al.* (2002) [212]. A trypan blue assay test was used on the cells post measurement, under the conditions stated for the thesis, which indicated the cells incurred no damage[208]. The principal limitation in using Notingher *et al.* (2002) as a comparable study is that the dead cells being measured were still in a medium and not air dried. Whilst air-drying is lethal to cells and damages cell constituents, this study is focused not only on the death of the cell but also the effect of dehydration on DNA, proteins and lipids.

A further study carried out by Notinghger *et al* in 2007 on human lung epithelial cells used Raman spectroscopy to measure the differences between living and dead cells. It found that the specific indicators of cell death are characterised by changes to the DNA bands 782 cm^{-1}, 788 cm^{-1}, and 1095 cm^{-1} [228]. These measurements were acquired with a laser wavelength of 785nm so it must be noted that these bands can not be compared directly to the results found in this study. However, relative changes to the biomolecules represented by these bands can be analysed using the associated peaks. As well as DNA differences, Notinghger *et al* (2007) also found differences between proteins lipids, and carbohydrates, between the live and dried sample. A drop in Raman intensity of phenylalanine at 1005 cm^{-1} was considered an indicator of cell death. Furthermore, in the amino acids a drop in Raman intensity at 1320 cm^{-1}, and 1342 cm^{-1} was also considered an indicator of cell death. Finally, a reduction in intensity of 1231 cm^{-1} indicated a change in structure of random DNA coils. However, there are critiques of the methodology and analysis of Notinghger *et al* (2007). The first arises from measuring cells in different stages of the cell cycle. A total of 20 live cells were used to generate the live average spectrum, made up of a subset of 10 isolated cells, 5 cells in close proximity to each other (designed to emulate a nutrient deficient environment) and a further 5 top layer round cells. The addition of nutrient deficient cells and top layer cells will lead to increases in the standard deviation of the live cell sample. The addition of higher density top layer cells will also contribute to an increased error in the average live spectrum. Furthermore, the nutrient deficient cells, although alive, may lead to false biochemical compar-

isons being drawn between the live and dead cells. This thesis, ultimately, uses a different methodology to Notinghger *et al* (2007) to prevent any false biochemical comparisons being drawn. As previously mentioned, this thesis focuses on the effects of air-drying as a fixation method and what the effects of drying are on the cell nucleus, biochemically. More importantly, the objective is to not only observe the differences in relative intensity between live and dried cells, but also to fully characterise each sample. This is done through a combination of peak intensity ratio analysis and principal component analysis.

There are positive aspects of using dried cells. One of the benefits of using dried cells for analysis is that the cells themselves are immobilised, allowing entire Raman mapping of the cell [229, 230]. This allows for multiple spectra from the whole cell to be measured, as opposed to live cells, where more than one measurement can lead to degradation of the cell. Another property of dried cells is that they are resistant to degradation over a prolonged period, of around 24hrs [231]. A prolonged period available for measurement reduces cost as the experiment will requires less samples for measurement and thus less lab resources dedicated to culturing multiple samples. This is in contrast to live cells that are only viable for a period of a few hours and require multiple samples to achieve a Raman spectra wholly representative of the sample [232]. There are conflicting studies describing the changes that occur during the air-drying process, morphologically and biochemically. Some report that air-drying cells does not alter the shape of the nucleus, retaining a nearly spherical shape [233]. However, further microspectroscopy studies have highlighted morphological changes to the shape of air dried cells. There is also evidence that there is a loss of biochemical information from the perinuclear region [234]. Additionally, studies have debated as to what effect the salt crystals, generated by the rinsing medium PBS that are left behind on the cells and substrate, have on Raman measurements. One study did suggest that drying artefacts, such as salt crystals, do not affect the biochemical signature of cells [235]. Nevertheless, despite no definitive approach, commonly, excessive salt crystals left by buffer solutions are removed from dried cells by rinsing with water [215]. While this thesis did observe key differences in

the morphology of the dried cells it is not restricted to microscopy only. This study seeks to understand the key biochemical differences between a live and dried cell line.

Research has already shown that the water content has a direct impact on the intensity of the cells Raman spectra, with an increase in water content being associated with a decrease in the peak intensity of the Raman signal [236]. The reported increase of Raman signal in dried cells also resulted in an increased baseline as the period of air-drying increases [232, 237]. It should also be considered that this increase in baseline has been attributed to heat denaturation by the excitation beam [238]. What is universally accepted in the field of Raman spectroscopy is that any method of air drying in cell preparation leads to a deviation in the Raman spectra of the biological subject, dependent on hydration state [232]. Specifically, changes in DNA conformation occur, transitioning from a B state to an A state. The B state is the most common state of DNA; a long narrow alpha-helix structure occurring at neutral pH and salt concentration. The A state is a more compact under-wound coil[239]. From a biochemical perspective, a change in conformation state is associated with a perturbation in DNA/protein vibrational modes, altering the Raman spectrum [237]. This in turn has led to air drying to be described as an inconsistent method for the preparation of cells undergoing Raman spectroscopy [235].

Despite the negative consequences, widespread use of dried cells in studies proceed, with an increase in relative signal-to-noise compared to other methods being cited, as well as the level of difficulty and challenges with using live cells being a deciding factor [225, 240]. As the use of Raman spectroscopy becomes prolific in the field of diagnostics it is essential to understand the key biochemical changes present when different methods are used, to ensure ex vivo environments accurately represent *in vivo* counterparts [241]. This study uses a normal prostate standard cell line to quantify the relationships between the DNA/RNA, proteins and lipids in a live and dried state. While there have been several studies that have included the cell line PNT2-C2 this will be the first study to compare it in its live and dried state, as far as the author is aware. The biochemical information

extracted from this will allow quantitative interpretation of the drying process and also act as a guide for further cell studies where dried cells may be used.

4.2 Convergence tests and removal of outliers

During Raman measurement of cells the laser is targeted at the nucleus. This is because the nucleus stores the genetic information necessary to control and regulate the cell. The nucleus is not only responsible for how the cell will proliferate and at what rate but also when the cell will undergo apoptosis [219]. The cytoplasm, being void of DNA in eukaryotes, however, lacks the genetic information relating to cell control. The cytoplasm itself will still generate a Raman signal and produce Raman bands, as observed with the nucleus, due to the presence of lipids and phosphates [220]. It should also be noted that the cytoplasm is used in other lipidomic studies, which form a key part of the total profile of the cell. However, for this study, to understand what key biological changes are taking place in the cell, it is necessary to make sure that the Raman spectra are being derived from the nucleus and not the cytoplasm. One of the key limitations of using standard cell lines in their dried or live form is that even if the nucleus is targeted there will be instances where the cytoplasm interferes to an extent that the resulting spectrum is of cytoplasm and not the intended DNA. This thesis aims to be consistent in terms of only obtaining information derived from the cell nucleus.

It is possible to identify which Raman spectra represent a measurement of cytoplasm rather than the nucleus. Generally, the Raman signal generated by the cytoplasm is much weaker than that generated by the nucleus. Another key difference observed in the Raman spectra taken from cytoplasm is that it is deficient in DNA Raman bands [212]. This can be illustrated by considering figure 4.1.

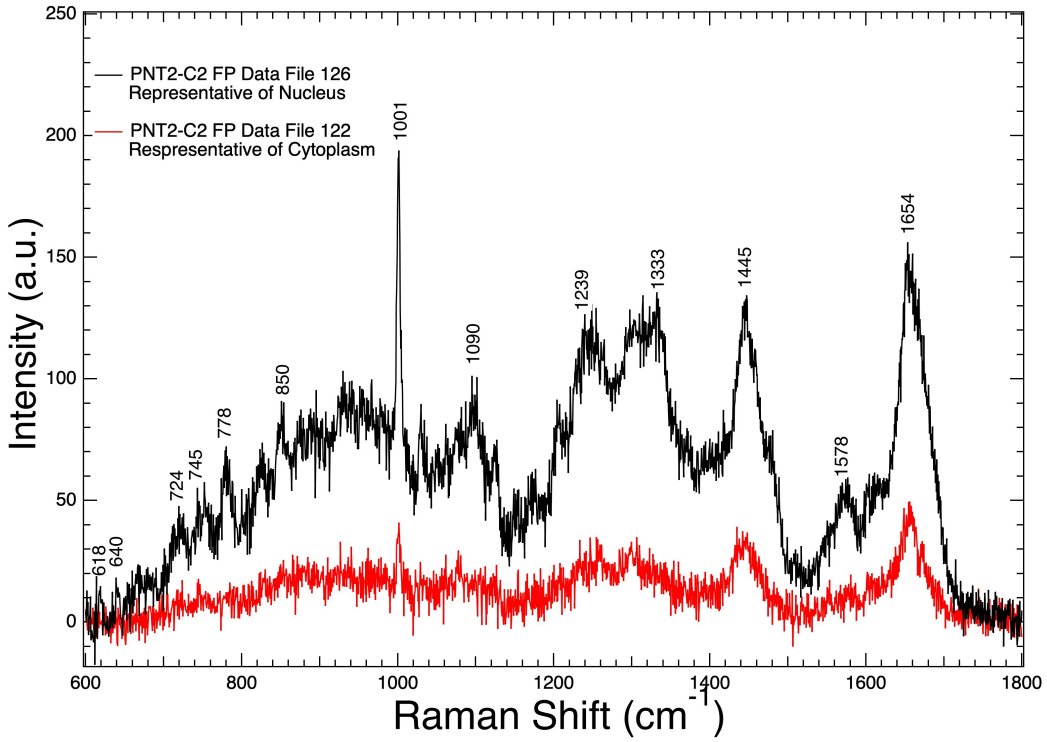
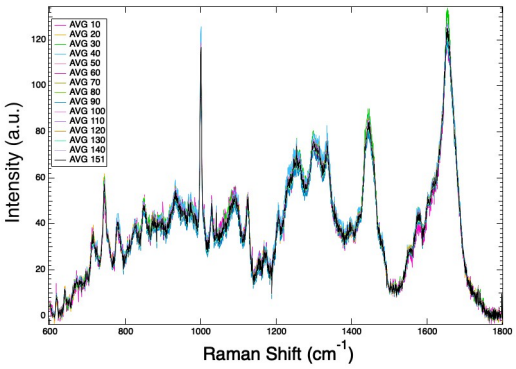


Figure 4.1: The Raman spectra from data files 122 and 126 from PNT2-C2 live are plotted on the same graph for the fingerprint region. Both spectra have been cut and baselined prior to plotting. These spectra both originate from the same population of cells. Data file 126 is representative of a Raman spectrum taken from the nucleus of a cell. It can be observed that the spectrum generated from the cytoplasm, spectra 122, is of significantly less intensity. Raman DNA bands 618 cm^{-1}, 640 cm^{-1}, 724 cm^{-1}, 745 cm^{-1}, 850 cm^{-1}, 1090 cm^{-1}, 1239 cm^{-1}, 1333 cm^{-1} and 1578 cm^{-1} are present in spectra 126 and not in 122. The absence of DNA bands in file 122 is indicative of cytoplasm [212]. There are bands present in file 122 although much less intense. These Raman bands 1445 cm^{-1} and 1654 cm^{-1} are attributed to the lipid and protein regions. It should be noted that these are only tentative labels and many more peaks are present in the spectra of file 126 that have not been labelled.

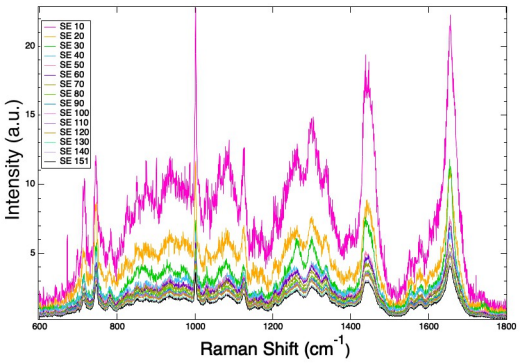
As seen in Figure 4.1 the signal generated from cytoplasm is very different to the spectrum generated from the nucleus. Multiple DNA bands are missing and the profile contains very low relative signal-to-noise. When selecting which spectra were to be excluded from the data sets, each spectra was considered individually before it was removed from the data set. A total of three spectra were removed from PNT2-C2 live as these were linked to cytoplasm/very low, relative signal-to-noise. A total of six spectra were removed from PNT2-C2 dried as these were linked to cytoplasm/very low, relative signal-to-noise.

An important part of the removal of outliers is that it allows for the capture of true population heterogeneity. Whilst the number of outliers removed is very small relative to the number of spectra taken they have been removed for completeness. In terms of PNT2-C2 for the live and dried counterparts, this allows us to capture the point at which the average Raman spectra is a true representation of the nucleus being measured. This is referred to as a statistically-converged average spectrum of the population. It is currently not clear in the current state-of-the-art use of Raman spectroscopy, at what point this statistical convergence is met. There is no standard number for statistical convergence and for each experiment this needs to be determined by a set process. In turn, this has led to a large variation in how many spectra are included in studies. Specifically for PNT2-C2 cells, Corsetti *et al.* (2018) measured 69 cells, Kerr *et al.* (2014) measured 110 and Harvey *et al.* (2009) measured 170. [224, 242, 243]. The issue, however, does not relate to the number of spectra taken but that the authors of the previously mentioned studies have not shown that these numbers lead to statistical convergence and population representation in their experiments. This is a very small sample of papers but it illustrates the wide variation in the number of cells being included in each study, to draw conclusions. Furthermore, it will also vary depending on the quality of spectra being collected and limitations to instrumentation.

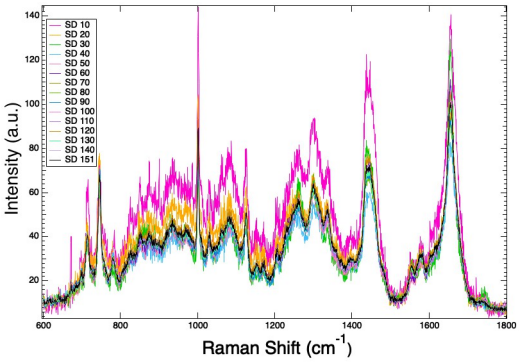
In order to make sure that the data sets used in this thesis have achieved convergence prior to analysis, statistical measures have been carried out. The statistical measures investigated were namely the spectral average, twice the standard deviation (2SD) and the standard error (SE). In total 151 spectra were collected for PNT2-C2 live. A total of 148 spectra were collected for PNT2-C2 dried. The convergence tests shown in Figure 4.2 are for PNT2-C2 live. Additionally, the convergence test results are shown for PNT2-C2 dried in Figure 4.3.



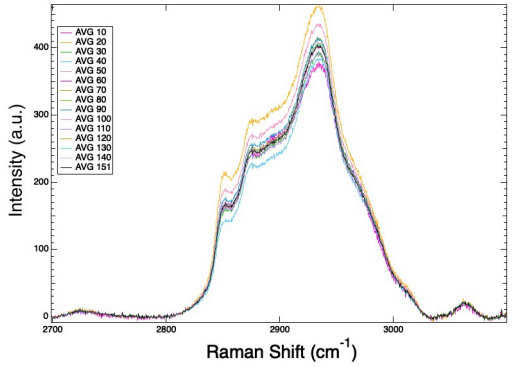
(a) PNT2-C2 LIVE FP AVG



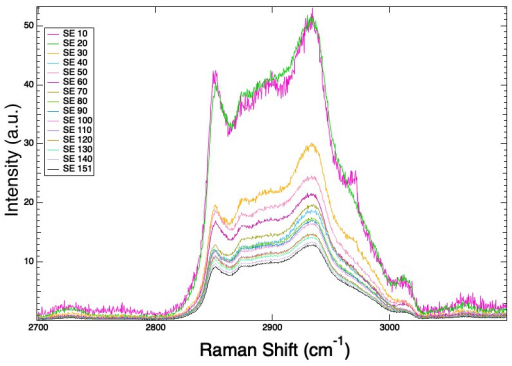
(b) PNT2-C2 LIVE FP SE



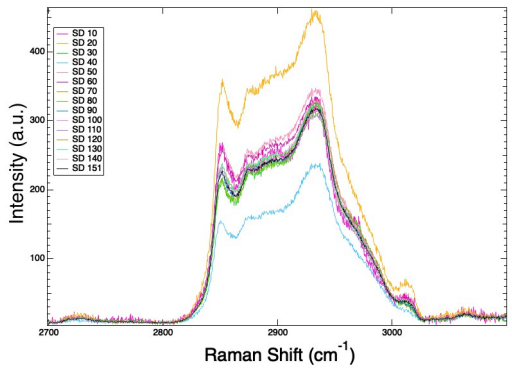
(c) PNT2-C2 LIVE FP 2SD



(d) PNT2-C2 LIVE HW AVG

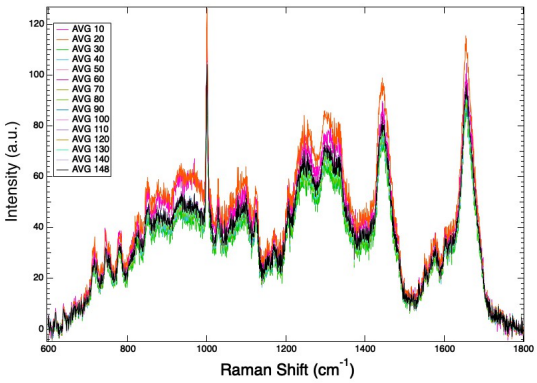


(e) PNT2-C2 LIVE HW SE

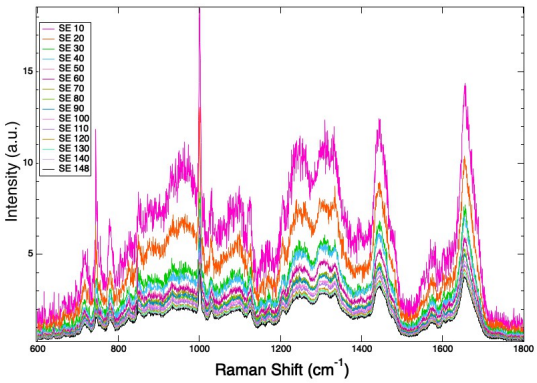


(f) PNT2-C2 LIVE HW 2SD

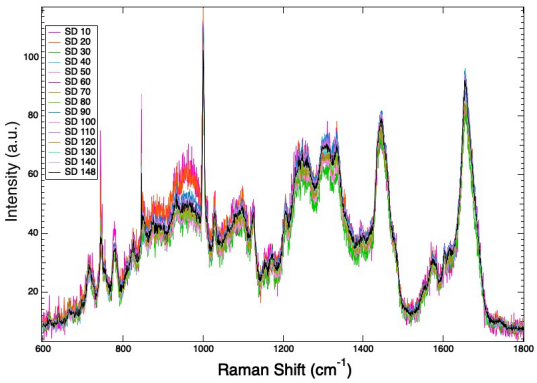
Figure 4.2: Convergence of the average (AVG),twice the standard deviation (2SD) and standard error of the mean (SE) as a function of the increasing number of spectra per spectral average for the PNT2-C2 live fingerprint (FP) and high wavenumber (HW). [Figs. (a) & (b), (c) & (d) and (e) & (f), respectively].



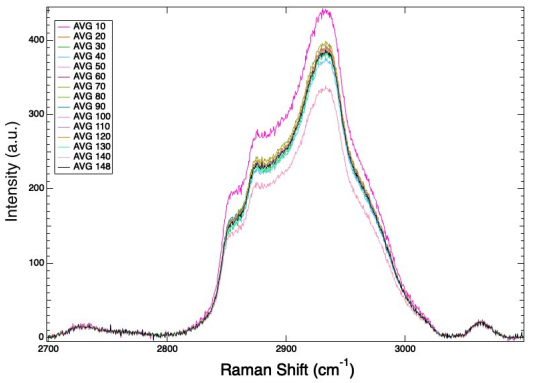
(a) PNT2-C2 DRIED FP AVG



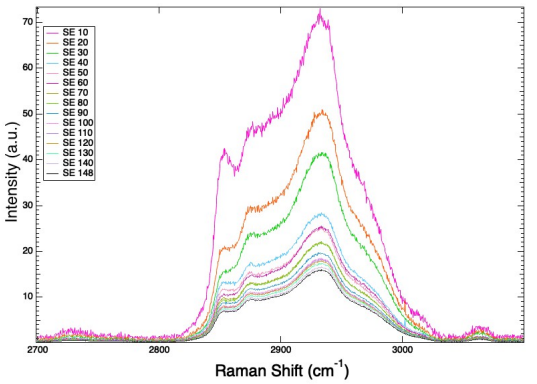
(b) PNT2-C2 DRIED FP SE



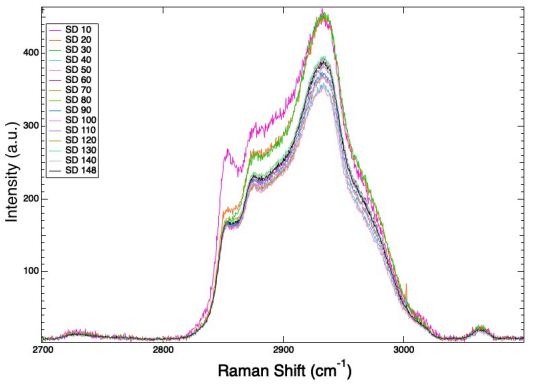
(c) PNT2-C2 DRIED FP 2SD



(d) PNT2-C2 DRIED HW AVG



(e) PNT2-C2 DRIED HW SE



(f) PNT2-C2 DRIED HW 2SD

Figure 4.3: Convergence of the average (AVG), twice the standard deviation (2SD) and standard error of the mean (SE) as a function of the increasing number of spectra per spectral average for the PNT2-C2 dried fingerprint (FP) and high wavenumber (HW). [Figs. (a) & (b), (c) & (d) and (e) & (f), respectively].

The results from the convergence tests of the spectral average, twice the standard deviation and the standard error for both the fingerprint and high wavenumber region of PNT2-C2 live and dried are shown in Table 4.1. The convergence in the fingerprint region was assessed using the 1001 cm^{-1} peak. The

convergence in the high wavenumber region was assessed using the 2934 cm^{-1} peak. The criteria for convergence is detailed as per the Methods Section 3.6.1.

Table 4.1: Convergence data for PNT2-C2 live and dried. Key for table: PNT2-C2 dried AVG FP 80 indicates that the convergence as per Methods Section 3.6.1 is seen to first occur at approximately 80 spectra for the PNT2-C2 profile average. Furthermore, 2SD denotes twice the normal standard deviation and SE denotes the standard error.

Cell line	**AVG FP**	**2SD FP**	**SE FP**	**AVG HW**	**2SD HW**	**SE HW**
PNT2-C2 live	60	80	70	60	60	70
PNT2-C2 dried	80	100	70	80	60	90

The convergence tests summarised in Table 4.1 show that the respective spectral profiles of both PNT2-C2 live and PNT2-C2 have converged. Another observation in Table 5.1 is the variation between cell lines in both the fingerprint and high wavenumber regions. This would suggest that even under the same experimental setup there needs to be convergence analysis for each region and each population rather than an arbitrary number of spectra needing to be taken.

While these convergence results are the case for this experimental setup using specific methodology outlined in Section 3, it is acknowledged that it is difficult to draw comparisons with other studies. This is primarily due to other studies not having a convergence protocol in place and the convergence depending on various experimental factors. Experimental factors may include varying laser wavelength, target sample, medium and substrate. It is therefore impossible to draw conclusion from previous studies such as Corsetti *et al.* (2018) using 69 cells against Harvey *et al.* (2009) taking 170 spectra and what might constitute a reasonable number of measurements [224, 243]. Clearly with such variation in the number of spectra being taken it is necessary for future studies to also publish their convergence results. This is currently not tracked within current research papers such as Corsetti *et al.* (2018) and more widely in the field [224]. The inclusion of converge tests would not only increase the robustness of Raman cell studies in terms of biomolecular characterisation and understanding of biomolecular heterogeneity but also allow for more accurate artificial intelligence (AI) models in the future. There is a large amount of interest in the use of AI in the future of medical imaging and it is therefore imperative that the data provided for machine learning is truly

representative of the sample [244].

4.3 Raman spectroscopy of live and dried version of PNT2-C2

To begin the process of quantitative analysis between the Raman profiles PNT2-C2 live and dried they have been characterised by fitting all the bands present in the profile. The method for fitting the Raman bands is described in Section 3.6.3. The mean Raman spectra from both PNT2-C2 live and dried for the fingerprint and high wavenumber region were analysed in the fingerprint region 600-1800 cm^{-1} and the high wavenumber region 2700 to 3100 cm^{-1}. As shown in Section 4.3, both data sets are fully converged. It can be seen, within an acceptable degree of uncertainty, that the cell lines PNT2-C2 live and PNT2-C2 dried share an almost identical band structure throughout their Raman profile. However, peak fitting of the PNT2-C2 live cell line and the PNT2-C2 dried cell line found that PNT2-C2 dried sample has absent bands at 698 cm^{-1} and 1396 cm^{-1} that are present in PNT2-C2 live. In total there were a total of 45 peaks identified in PNT2-C2 live, identified by being able to be fitted using the minimum number of Gaussian profiles to fit the band. The 1484 cm^{-1} band in PNT2-C2 dried was fitted by being held in place; all other bands were free-fitted. This is because during the fitting process the peak had to be held in place. This is against the protocols defined in Section 3.6.3. No other peaks were held in place. A list of all 45 peaks and their general assignments, as well as detailed biomolecular assignments, is presented in Table 4.2. The peak location difference between live and dried PNT2-C2 is given in 4.2. The average peak positions used for the mean profile subtraction, PCA analyses and peak intensity ratio analyses, are calculated by the average of the peak position range.The detailed assignments have been provided from literature and are widely accepted in the field of Raman spectroscopy of cells. Furthermore, the detailed assignments are gathered from studies that use Raman in the visible spectrum.

Table 4.2: Biomolecule peak-assignments for PNT2-C2 live and dried. The average peak position has been between live and dried (Table B.1 in Appendix), general and detailed assignments were obtained from the literature. The symbol ⊗ indicates a peak was not present in PNT2-C2 dried. For the peak position a standard error of ± 3 cm^{-1} as an uncertainty is attached to all positions.

Peak Position (cm^{-1})	**General assignment**	**Detailed biomolecule assignments**
618	Proteins	C-C twist of proteins ([128], [129]) Phenylalanine C-C twist ([128], [129], [224], [245])
640	Proteins	C-S stretch & C-C twist of tyrosine ([128], [129] C-C twist of tyrosine ([224], [245], [216], [142])
669	DNA/RNA & Proteins	Guanine and thymine ring breathing mode ([128]) Tyrosine-G backbone in RNA ([128]) Guanine ring breathing mode and C-S stretching mode of cystine ([245]) Guanine, tyrosine and cysteine ([142])
698 ⊗	Proteins	Amino acid methionine ν(C-S) *trans* ([128])
715	Proteins & Lipids	Amino acid methionine ν(C-S) *trans* ([128]) $CN^+(CH_3)_3$ stretch in lipids ([128]) C-N membrane phospholipid head stretch ([128], [246], [245]) Adenine nucleotide peak ([245], [128])
724	DNA	Adenine ([245], [142]) DNA ([247], [128], [224])

Continued on next page

TABLE 4.2 (continued): Biomolecule assignments

Peak Position (cm^{-1})	**General assignment**	**Detailed biomolecule assignments**
745	DNA	Thymine ring breathing mode ([248], [105], [128], [224])
755	Proteins	Symmetric breathing of tryptophan ([245]) Tryptophan ring breathing mode ([128], [246]) Tryptophan ([249], [245], [129])
781	DNA/RNA	Uracil, Cytosine and thymine ring breathing modes ([249] [249], [142] [142], [128], [224]) Nucleic acids ([63]) Uracil, cytosine, and thymine overlapping with O-P-O stretch in DNA ([105])
808	DNA/RNA	O-P-O stretch DNA ([249] [249]) and O-P-O stretch RNA ([128], [246])
825	Proteins & DNA/RNA	Tyrosine out-of-plane ring breathing and O-P-O stretch in DNA ([249] [249], [245], [142] [142], [105] [105], [128]) DNA O-P-O asymmetric stretch, tyrosine ([248]) Proline and hydroxyproline, out-of-plane tyrosine ring breathing, DNA/RNA asymmetric O-P-O stretch ([246]) DNA backbone (O-P-O) ([63]) Out-of-plane tyrosine ring breathing ([224])
850	Proteins	Tyrosine ring breathing ([249] [249], [105], [128] , [224]) C-C stretch in proline and tyrosine ring breathing mode ([245], [142])

Continued on next page

TABLE 4.2 (continued): Biomolecule assignments

Peak Position (cm^{-1})	**General assignment**	**Detailed biomolecule assignments**
		Proteins (C-C stretch in proline; tyrosine ring breathing); carbohydrates (glycogen, polysaccharides C-O-C stretch) ([246])
875	Proteins & Lipids	Proteins C-C stretch, antisymmetric vibrations of choline $N(CH_3)_3$ (phospholipids), hydroxyproline, tryptophan (proteins) ([128]) Hydroxyproline C-C stretch in collagen, tryptophan ring deformation, choline asymmetric stretch in phospholipids and C-O-C stretch in carbohydrates ([246]) Tryptophan ([224])
889	Carbohydrates & Proteins	Saccharide band, methylene and C-C skeletal, proteins (protein modes for tumours) ([128]) C-C skeletal ([105])
934	Proteins & Carbohydrates	Skeletal C-C stretch α-helix (protein), proline, valine ([245], [128]) C-C backbone stretch α-helix ([249], [63] [63], [142], ([246]) C-O-C glycogen (carbohydrate) ([246])
967	Lipids, Proteins & DNA/RNA	Lipids, phosphate monoester groups of phosphorylated proteins and cellular nucleic acids ([128])
1001	Proteins	Phenylalanine ring breathing ([248], [249], [245], [142], [105], [128], [246], [224])
1030	Proteins	Phenylalanine C-H in-plane bending mode ([249], [245], [142], [128], [224])

Continued on next page

TABLE 4.2 (continued): Biomolecule assignments

Peak Position (cm^{-1})	**General assignment**	**Detailed biomolecule assignments**
	Lipids & Carbohydrates	C-N stretch of proteins ([105] [105], [128]) Collagen, keratin, C-N stretch in proteins, Phe C-H in plane bending, phospholipids and polysaccharides ([246])
1061	Lipids & DNA/RNA	Lipids chain C-C stretch ([249] [249], [245], [128]) DNA/RNA PO_2^- backbone stretch ([128])
1089	DNA, Lipids & Proteins	DNA PO_2^- backbone stretch and C-C stretch (gauche) in lipids ([248], [245], [105], [128]) C-N of proteins ([105], [128]) Lipids ([63]) DNA PO_2^- stretch, lipids chain C-C stretch, carbohydrates C-O and C-C stretch [228, 249]
1125	Lipids, Proteins & Carbohydrates	Protein C-N stretch ([249], [142], [128]) C-C stretch in lipids ([249], [245], [142] [128]) C-C stretch in proteins ([245], [128]) C-N stretch (glucose), C-O & C-C stretch in carbohydrates ([245], [128]) phospholipids ([63])
1155	Proteins & Carbohydrates	C-C and C-N stretch in proteins ([249] [249], [105], [128], [224]) Carotenoids—correlated to disease ([128]) Glycogen ([128])

Continued on next page

TABLE 4.2 (continued): Biomolecule assignments

Peak Position (cm^{-1})	**General assignment**	**Detailed biomolecule assignments**
1171	Proteins	(CH) phenylalanine, tyrosine ([105], [128]) Bending mode (in-plane) tyrosine ([249], [245]) Protein-related peak (C-H bend) ([105])
1205	Proteins & DNA/RNA	Tyrosine and phenylalanine ([248], [128]) Phenylalanine and tryptophan ([249], [245], [128]) Phenylalanine ([224]) CH_2 wagging of glycine and proline ([128]) Amide III ([105], [128]) A,T ring breathing modes of DNA/RNA ([105], [128]) Hydroxyproline & tryrosine ([245], [105])
1234	Proteins & DNA/RNA	Amide III and CH_2 wagging of glycine and proline ([128]) PO_2^- antisymmetric stretch ([128]) Amide III ([249],[142], [224])
1256	Proteins	amide III : CH_2 wagging [245, 250]
1306–1307	Proteins, Lipids & Nucleic acids	CH_2 deformation in lipids, adenine and cytosine ([245], [128]) CH_3/CH_2 modes in lipids / collagen ([245], [128]) CH-N stretch in amines ([128]) Nucleic acids ([63]) Lipids; proteins (amide III), adenine, cytosine ([248])

Continued on next page

TABLE 4.2 (continued): Biomolecule assignments

Peak Position (cm^{-1})	**General assignment**	**Detailed biomolecule assignments**
1336	Proteins & DNA/RNA	CH_3CH_2 wagging, twisting &/or bending of collagen, lipids, tryptophan & nucleic acids (adenine and guanine)([128]) Amide III ([128]) CH_2 deformation in lipids, adenine and cytosine ([245]) Adenine and guanine; protein C-H deformation ([249], [105]) Adenine, phenylalanine and CH deformation ([248]) Adenine ([142])
1398 $\otimes$	Proteins	CH_2 deformation ([128, 129]
1448	Lipids, proteins & DNA/RNA	CH_2 deformation in lipids and proteins; at 1450 cm^{-1}, in malignant tissues ([248], [142], [128]) CH_2 bending mode of proteins ([245], [105], [224]) C-H deformation & vibration in lipids and proteins ([249], [128]) CH_2CH_3 deformation ([128]) Nuclear acids ([63]) CH_2 in-plane deformation (ring stretch) in DNA/RNA ([142])
1483	Protein & DNA/RNA	Amide II CN stretch & N-H bend ([128]) Guanine, adenine ring-breathing modes ([245], [105], [142], [128])
1552	Proteins	Tryptophan ([245], [128]) Amide II ([128])
1575	DNA/RNA	Guanine and adenine ring-breathing modes ([249], [245], [105], [142], [128]) Nucleic acids ([245], [63])

Continued on next page

TABLE 4.2 (continued): Biomolecule assignments

Peak Position (cm^{-1})	**General assignment**	**Detailed biomolecule assignments**
1603	Proteins	C=C bending mode of phenylalanine and tyrosine ([249], [245], [105], [128]) C-C ring stretch of phenyl ([128]) C-C and C-N stretch (proteins) ([224])
1614	Proteins	Tyrosine, tryptophan C=C stretch ([249], [245], [105], [142], [128])
1656	Proteins & lipids	Amide I C=O stretch, α-helix conformation; proteins ([249], [245], [128]) α-helix proteins & unordered proteins ([63]) protein (amide I) [105] C=C stretch of lipids ([249], [245], [248], [128], [64])
1681	Proteins	Amide I ([105], [128])
2851	Lipids	CH_2 symmetric stretch in lipids ([128], [64], [251], [252])
2872	Lipids & proteins	CH_2 asymmetric stretch in lipids and proteins ([128]) CH_2 symmetric stretch in lipids ([128]) CH_3 symmetric stretch in lipids & proteins ([64])
2895	Lipids & proteins	CH stretch of lipids and proteins (predominantly) and CH_3 symmetric stretch ([128]) CH_2 deformation ([251], [252])
2931	Proteins & lipids	CH_2 asymmetric stretch, CH_3 symmetric stretch (predominantly proteins) ([128])

Continued on next page

TABLE 4.2 (continued): Biomolecule assignments

Peak Position (cm^{-1})	**General assignment**	**Detailed biomolecule assignments**
		CH_2 asymmetric stretch of proteins and modes in lipids, CH modes in lipids and proteins ([251]) CH_3 symmetric stretch in proteins and lipids ([252])
2940	Lipids & proteins	C-H vibrations in lipids & proteins. CH_2 modes in lipids ([128])
2964	Lipids	CH_3 asymmetric stretch in lipids, cholesterol & cholesterol ester ([128])
3014	Lipids	Unsaturated =CH stretch in lipids ([128], [251], [252]) =CH stretch (unsaturated lipids, cholesterolester) ([64] [64])
3064–3064	Lipids	CH stretch in lipids ([128])

Figure 4.4 shows the average spectrum of both the PNT2-C2 live and dried fingerprint region. Here, they are presented after being cut, so both spectra are only viewed within the range (600-1800 cm^{-1}). The grey region on each average spectra is the standard error for the mean envelope. The method for calculating the mean envelope is detailed in Section 3.6.2 of the methods chapter. In the fingerprint region, the three most relatively intense Raman peaks are the same for both samples. These Raman peaks and their general biological assignments are: 1001 cm^{-1} (protein), 1445 cm^{-1} (lipids, proteins & DNA/RNA) and 1654cm^{-1} (proteins & lipids). It can be calculated, using Figure 4.4 as a reference, that the peak with larger relative intensity for PNT2-C2 dried is the 1001 cm^{-1} peak. This differs to PNT2-C2 live, where the 1656 cm^{-1} peak is the most relatively intense Raman band. Although there has not been a study on the intensity comparison of live and dried prostate cells, it has been discussed in depth in other biological systems. The study carried out by Notinghger *et al* in 2007 on human lung epithelial cells observed a relative drop in intensity at 1001 cm^{-1} in the Raman

signal of dead cells compared to live cells in a PBS medium. Whilst the observed relative drop in intensity may be an indicator of cell death at 1001 cm^{-1} this has to be measured relative to other intensity of peaks in the spectrum. It should be noted that the dead cells were not measured in a dehydrated state and are therefore not directly comparable. There is added difficulty to extrapolate the findings due to variations in methodology, instrumentation and laser wavelength between studies. It is known, however, that there is a relationship between the hydration state and intensity of the Raman signal [236]. A comparable study was carried out by Shim *et al* (1996) on cheek tissue cells, in which the dehydration of tissue caused a change to the relative intensities of protein vibrational modes. This study did not explore the relationship of overall intensity difference between fresh and dried samples. However, it did use the 930 cm^{-1} band to probe the conformational state of DNA between live and dried cells. Other studies, such as Ozaki *et al* (1992), have attributed an increase in relative intensity of peaks to heat denaturation from the laser beam [238]. The methodology described in Section 3.5.4 detailing Trypan blue assay test to confirm cell viability should indicate that heat denaturation is not responsible for the difference in relative intensities between PNT2-C2 live and dried. It should also be noted that the presence of HBSS as a medium for the cells during measurement does not exhibit fluorescence in the spectral range 600-1800 cm^{-1} or 2700-1800 cm^{-1} .

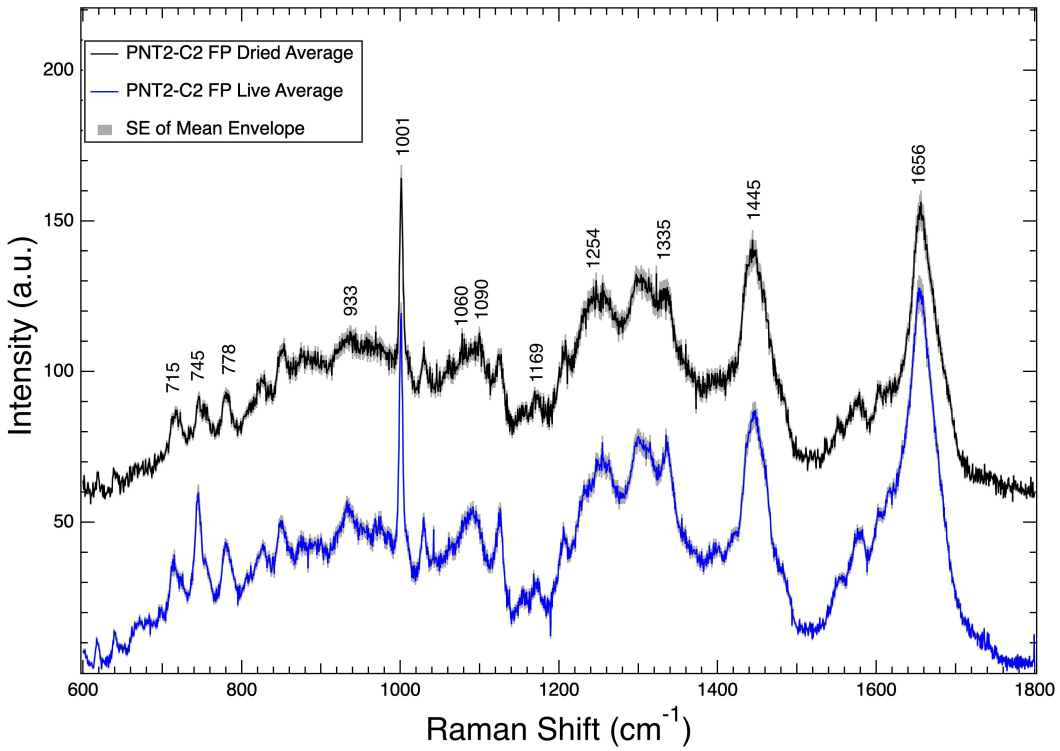


Figure 4.4: Fingerprint region of PNT2-C2 dried and live displayed on the same graph for direct spectral comparison. Both spectrum have been cut and baselined. The PNT2-C2 dried spectrum has been offset to allow for better comparison. The DNA Raman band 745 cm^{-1} is notably less pronounced relative to other bands in PNT2-C2 dried when compare to PNT2-C2 live. The protein/lipid band 715 cm^{-1} and DNA band 778 cm^{-1} adjacent to the 745 cm^{-1} band also experience a decrease in relative intensity compared to other bands in the Raman spectrum. However, this decrease is not as pronounced. Another observation is the change in profile shape occurring between the protein/carbohydrate band 933 cm^{-1} and amino acid (phenylalanine) band 1001 cm^{-1}. The 933 cm^{-1} band is less intense and less well defined in PNT2-C2 dried. The region between 933 cm^{-1} and 1001 cm^{-1} visually does not look representative of the live sample. This change in spectrum profile is also observed between the DNA/lipid bands 1060 cm^{-1} and 1090 cm^{-1}. Additionally, an increase in the standard error of the mean envelope and loss of intensity is observed at the 1335 cm^{-1} DNA/RNA band. There are visual similarities between the live and dried spectrums. The protein/lipid bands 1445 cm^{-1} and 1656 cm^{-1} appear similar in profile and with similar standard error mean envelopes present.

The largest difference observed in Figure 4.4 is the decrease in relative intensity of the 745 cm^{-1} (DNA) peak in PNT2-C2 dried. There is also a decrease in the adjacent peaks 715 cm^{-1} (protein/lipid band) and DNA band 778 cm^{-1} (DNA). The detailed biomolecular assignments for each of these peaks are: 715 cm^{-1}

(amino acid methionine, $CN^{+}(CH_3)_3$ stretch in lipids, C-N membrane phospholipid head stretch and adenine nucleotide peak [128, 245]), 745 cm^{-1} (thymine ring breathing mode) and 778 cm^{-1} (uracil, cytosine and thymine ring breathing modes, nucleic acids and uracil, cytosine, and thymine overlapping with O-P-O stretch in DNA ([128, 141, 142, 212, 224]). In multiple Raman cell studies an increase in intensity of the bands associated with thymine and other nucleotide chains has been described as a marker for cancerous cells, when compared to the normal [253, 254]. This decrease in intensity could potentially lead to false positives or negatives, depending on which cell lines are being used. It should be highlighted that if a study has compared a normal and cancerous cell line in a dried state that although there is consistency between samples the cells are still not representative of their live state and will still potentially lead to false positives or negatives.

The absence of the two bands 698 cm^{-1} and 1396 cm^{-1} demonstrate that there may be information loss from using air-dried cells. For these bands to qualify as absent, the background noise has to be be defined as greater than the amplitude of the peak. The 698 cm^{-1} is attributed to proteins and more specifically amino acid methionine ν(C-S). Methionine is an essential amino acid that is essential for growth and development of the cell. There are three other amino acids that are characterised in PNT2-C2: tyrosine, phenylalanine and tryptophan. However, of these three only phenylalanine and tryptophan are considered essential amino acids. It is clear from the general and detailed biological assignments of PNT2-C2 live and dried in Table 4.2 that methionine has the least contribution to the Raman spectrum in terms of the number of bands. There are 2 methionine bands, 17 tyrosine bands, 10 tryptophan bands and 10 phenylalanine bands. To elaborate, the disappearance of the 698 cm^{-1} band will reduce the reliability of the only other band attributed to methionine, 716 cm^{-1}, when using the peak as biomarker to assess the relative intensity of it to other biomolecules in the cell. Additionally, the 716 cm^{-1} band is convoluted with a contribution from the $CN^{+}(CH_3)_3$ stretch in lipids. A Raman study by Bhatia *et al.* (2010) which examined the presence of amino acids in monoclonal antibody producing mammalian cell lines, detected the

two methionine bands using 785nm laser but also noted that the intensity of the methionine peaks were much weaker than the other amino acids [255]. Although not a study of live and dried cultures, the Raman study carried out by Bhatia *et al.* (2010) also noted that methionine was the least represented amino acid in the Raman profile of the cell culture. Complementary to the study by Bhatia *et al.* (2010), this study also highlights that there is a danger that vital amino acid information loss may occur when using dried cell lines. The reason methionine is so important is that it plays a vital role in cell growth and apoptosis [256]. It can be argued that this is the first study detailing the loss of the methionine band at 698 cm^{-1} due to the drying process. The band 1398 $cm^{-1}$is attributed to CH_2 deformation. This loss is attributed to the loss in intensity observed in PNT2-C2 dried. It can be argued that the loss of the 1398 cm^{-1} band is not as important as the 698 cm^{-1} band due to there being several other dominant CH_2 deformation bands. Other CH_2 deformation bands occur at 1448 cm^{-1}, a very intense peak in the fingerprint region and band 2930 cm^{-1}, a very intense peak in the high wavenumber region. This very basic analysis shows nucleic acid, protein and lipid information can be lost using dried cells, which could have major knock-on effects if used in clinical applications. Separately, the loss of the bands 698 cm^{-1} and 1398 cm^{-1} could be indicative of the drying process in PNT2-C2. Work would have to be expanded to consider multiple cell lines and their dried cell counterparts to see if this loss of information is observed between other cell lines. It is important to also recognise that these bands, not present in PNT2-C2 dried, may be quantifiable under a different experimental setup.

To further analyse both samples in the fingerprint region it is possible to consider a mean profile subtraction. To allow for this the spectra must be baselined and then normalised. This allows for a more detailed side by side comparison of both profiles. The mean subtraction for the fingerprint region is shown in Figure 4.5.

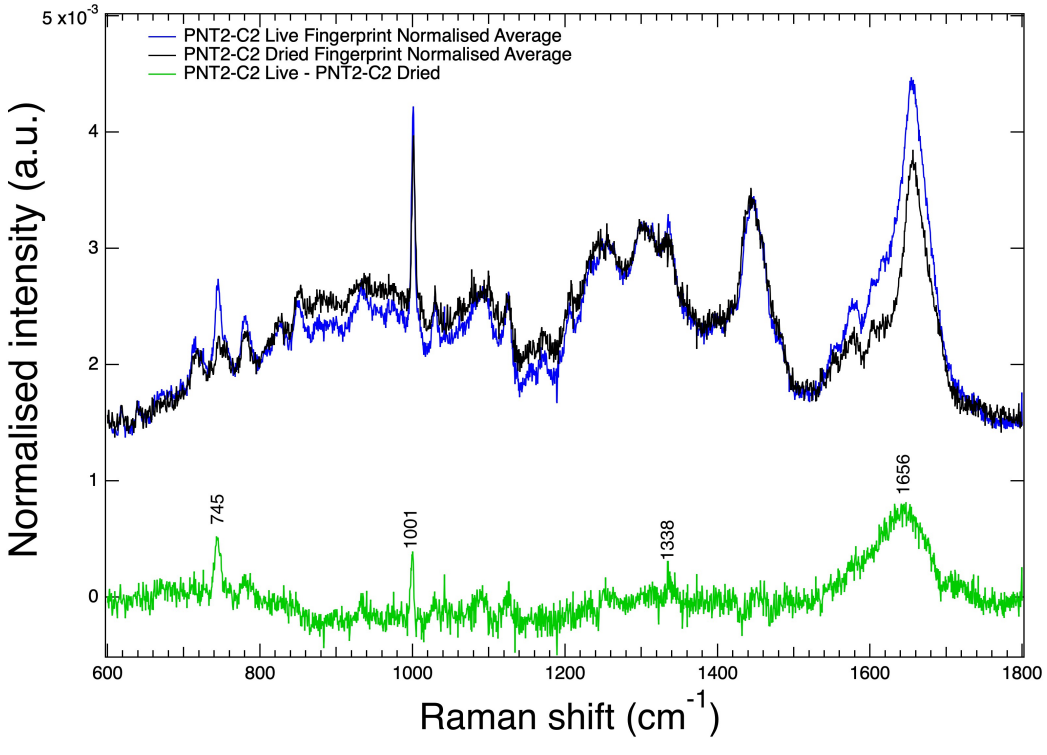


Figure 4.5: A mean profile subtraction is shown for PNT2-C2 live vs PNT2-C2 dried. Both spectra have been normalised using the total spectrum area before the subtraction. Here, the profile of PNT2-C2 dried has been subtracted from PNT2-C2 live. The resulting spectrum shown in green has been annotated to show key Raman peak locations. The key differences in relative intensity occur at the bands 745 cm^{-1} (DNA), 1001 cm^{-1} (phenylalanine) and 1656 cm^{-1} (lipids and proteins). The largest difference in relative intensity is observed around the 1656 cm^{-1} peak. Further analysis in this chapter will analyse the relative decrease of intensity of the 1656 cm^{-1} peak in PNT2-C2 dried.

The mean profile subtraction in Figure 4.5 strengthens evidence that there are significant changes in the Raman profiles between PNT2-C2 live and dried. Through the mean profile subtraction it can be seen that there are specific changes in the relative intensities of the peaks. In PNT2-C2 dried, the drop in intensity of the 1656 cm^{-1} peak relative to other peaks in the profile signifies key changes to the biochemical structure of the cell. As identified in the literature, the 1656 cm^{-1} band is representative of C=C bonds. As demonstrated by Potcoava *et al.* (2014) the 1656 cm^{-1} peak can be used for calculating mass unsaturation when used in a ratio against the peaks 1448 cm^{-1} and 1306 cm^{-1}. A large difference in the mean subtraction suggests that there may be changes to mass saturation

of the sample. Mass unsaturation, calculated by the ratio of C=C bonds present to the number of CH_2 bonds, allows for determination of changes to the lipid content of cells. This ratio can thus be calculated using the relative intensity of the 1448 cm^{-1} peak, which is attributed to CH_2 bonds. Similarly, the 1306 cm^{-1} peak is also representative of CH_2 bonds and the relative intensity can also be used to calculate mass unsaturation. Whilst visually there are observable differences in the mean subtraction of the fingerprint profiles it is not enough to draw a conclusion. It is necessary to carry out peak intensity ratio analysis in combination with PCA analysis to quantify changes not only to mass unsaturation but to find more subtle changes in the relative intensities of different Raman bands. This is carried out in Section 4.4.

The high wavenumber region (2700-3100 cm^{-1}) of PNT2-C2 live and dried is considered in Figure 4.6. The PNT2-C2 dried high wavenumber region has been offset to allow for easier comparison. They have both undergone a polynomial baseline prior to averaging and calculating the standard error as per Section 3.6.1 of the Methods chapter. The high wavenumber profile of PNT2-C2 dried, observationally, appears to be very similar to the profile of PNT2-C2 live. This could be attributed to the absence of DNA bands in the high wavenumber region, which have been found to be more sensitive to dehydration, although detailed quantitative analysis will investigate this further. There is a larger standard error of the mean envelope present in the region 2930-2940 cm^{-1} in PNT2-C2 dried. This region is mainly dominated by CH stretching of lipids and proteins found at the bands 2931 cm^{-1} and 2940 cm^{-1}. There is also a CH_3 symmetric stretching present.

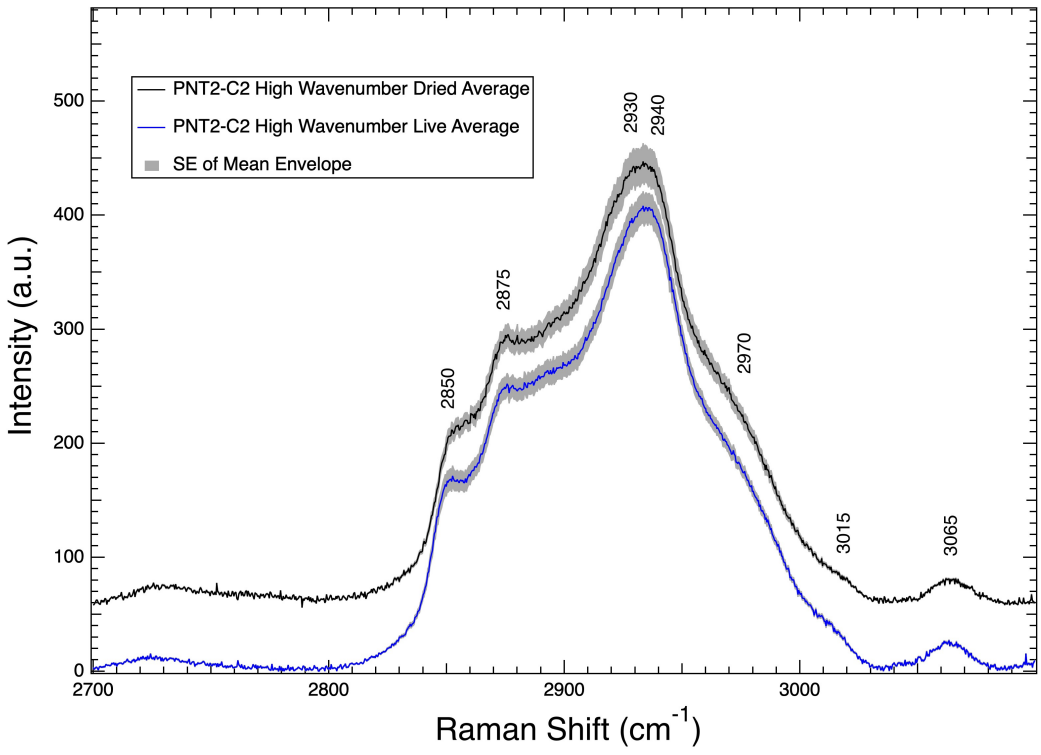


Figure 4.6: The high wavenumber Region of PNT2-C2 dried and live are displayed on the same graph for spectral comparison. These spectra have not been normalised. The standard error of the mean (SE) envelope is shown in grey around each spectrum. Both profiles appear to be very similar in structure and shape. However, it can be seen that the standard error (SE) of the mean envelope is larger for the region 2930-2940 cm^{-1}).

To further analyse the high wavenumber region it is possible to consider a mean profile subtraction. To allow for this, the spectra must have undergone a polynomial baseline subtraction and then normalisation. This allows for a more accurate comparison and overlay of profiles. The mean subtraction for the high wavenumber region is shown in Figure 4.7.

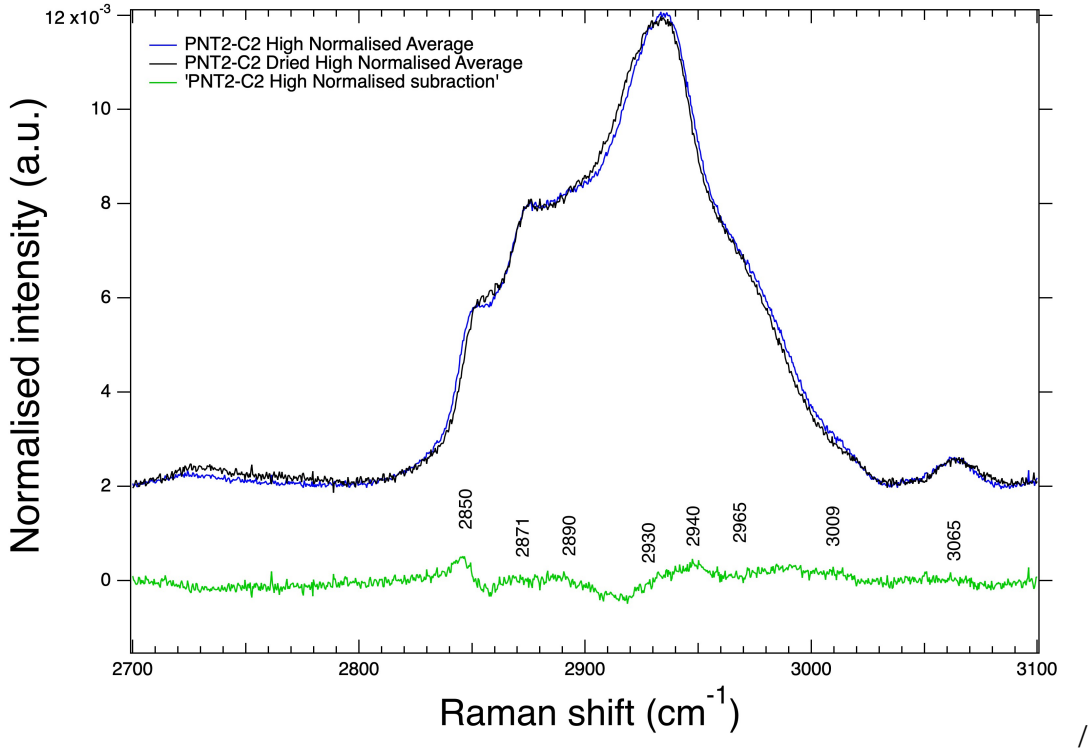


Figure 4.7: A mean profile subtraction has been carried out in the high wavenumber region of PNT2-C2 live and dried. The mean subtraction profile is shown in green. The peaks associated with the high wavenumber region have been marked on the mean subtraction profile. There were very minimal observational differences occurring between PNT2-C2 live and dried in Figure 4.6. However, there are large differences between the profiles emerging in the mean profile subtraction. This difference is characterised by peaks 2850 cm^{-1} and 2940 cm^{-1}. There is lower observed difference between the two populations at the 3015 cm^{-1} peak.

The high wavenumber region is mainly dominated by lipids [128, 129]. It is known that air-drying does have an effect on the DNA/protein vibrational modes [212, 225]. However, Notinghger *et al* (2002) and Mariani *et al* (2009) both expressed the opinion that there was a negligible effect on the lipid modes of the cell [212, 225]. Arguably, there are variations occurring in peaks 2850 cm^{-1} (CH_2 symmetric stretch in lipids) and 2940 cm^{-1} (C-H vibrations in lipids and lipids as well as CH_2 modes in lipids). The mean profile subtraction indicated that there was lower observed difference found at the peak 3015 cm^{-1} representative of unsaturated =CH stretch in lipids. There was no variation between samples detected at the peak 3065 cm^{-1}, representative of CH stretching in lipids. It is thought that differences in variation are generated through a change in state of

vibrational protein modes that overlap with the lipid bands. This will be further examined through the use of PCA analysis and peak intensity ratio analysis. It is thought that this thesis is the first time a mean profile subtraction has been carried out for the high wavenumber region between PNT2-C2 live and dried.

4.4 Principal component analysis and loading analysis

Prior to this study, Mariani *et al.* (2009) carried out PCA analysis to measure the impact of fixation on in vitro standard cell lines using Raman spectroscopy. They analysed two types of cells using three fixation methods. One type of cell analysed was a human dermal derived keratinocyte (HaCaT; ECACC, UK) which is a skin cell found in the epidermis [257]. The second was a cell line that represents human peripheral macrophages (MM6; DSMZ, Germany), a specialised type of white blood cell [258]. The three fixation methods analysed were the use of formalin, air-drying and desiccation. Desiccation, as not previously defined, is the removal of moisture thoroughly from the cell. This is done in a desiccator using pressure and silica beads [225]. It is considered to be more controlled than air-drying. Based on the findings of Mariani *et al.* (2009) it can be argued that air-drying is the most inconsistent method of cell fixation due to spectral reproducibility [225]. The study by Mariani *et al.* (2009) would have been improved by the addition of analysis of both cell types in their live state. Mariani *et al.* (2009) expressed the added level of difficulty in live cell culturing as a distinct challenge. Whilst this thesis does not examine desiccated or formalin fixed sample preparation, it does consider the differences between the live and dried state of the same cell line. This will not only compliment the work done by others in the field, but will also allow researchers limited to air-dried cell lines to understand some of the key differences that may occur in PCA and loading analysis between live and dried cell lines. The results for PCA analysis and loading analysis are considered for PNT2-C2 live and dried below in this section. They are then compared to the PIR analysis results in Section 4.5 for cross-validation.

PCA analysis between the PNT2-C2 live and dried fingerprint region is able to capture the variance of the first six principal components (PCs): PC1 = 31.2%, PC2 = 12.9%, PC3 = 7.4%, PC4 = 3.7%, PC5 = 2.8% and PC6 = 2.3%. This in turn means that the first three principal components capture 51.5% and the first six capture 60.4% of the total variance of the sample. Cumulatively, 299 principal components are required to capture total variance of the sample. Three rotations of PCA space using the first three principal components are displayed in Figures 4.8, 4.9 and 4.11.

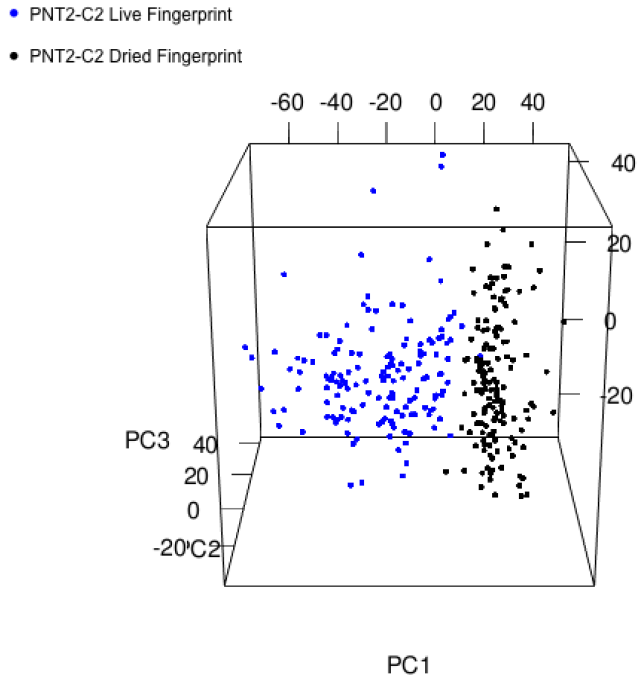


Figure 4.8: PCA analysis of the fingerprint region of PNT2-C2 dried and live in 3D space shows that separation is occurring along PC1. From this angle no separation is occurring along PC3.

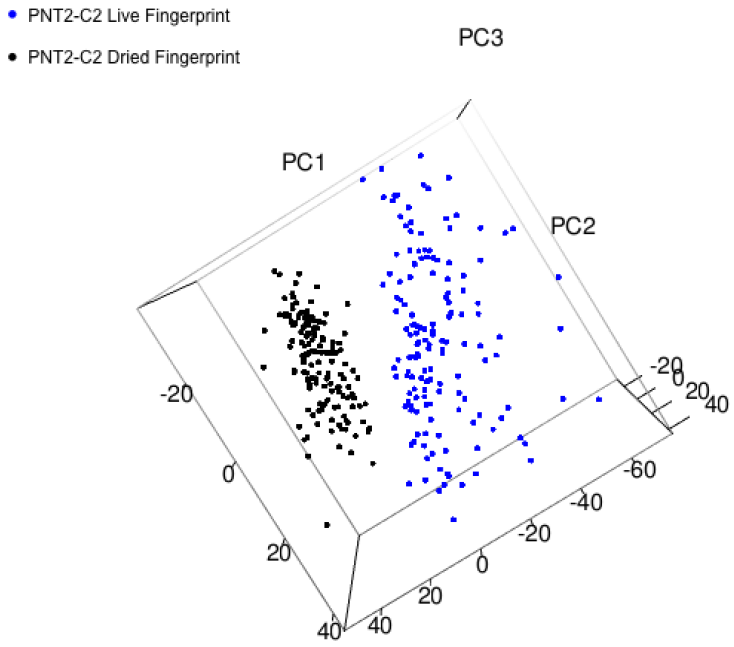


Figure 4.9: Here, a further rotation in PCA space is used to understand the contribution of PC2 in the fingerprint region of PNT2-C2 live and dried. Separation of samples is observed as happening at an angle along PC1 and PC2.

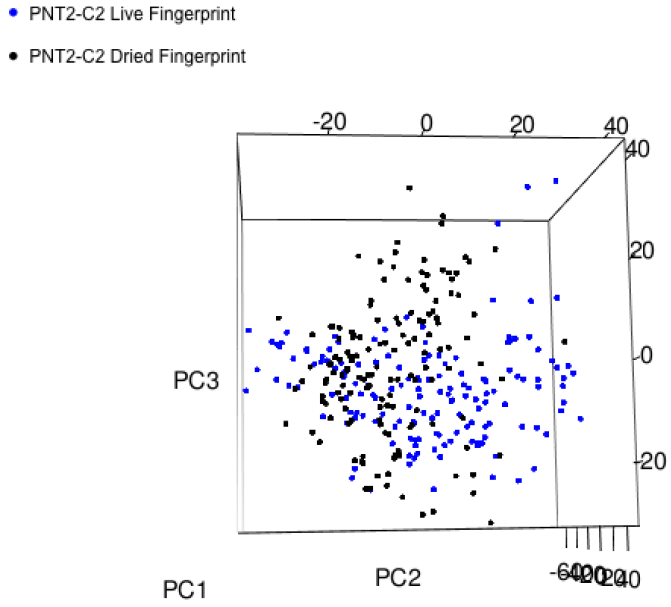


Figure 4.10: A further rotation in PCA space between the fingerprint region of PNT2-C2 dried and live is displayed using the first three principal components. The results indicate that there is no separation occurring between PC2 and PC3. This conclusion is reached due to complete overlap of points contributed by each spectrum. In conjunction with Figure 4.12 it can be seen that no separation is occurring along PC3 with PC1 or PC2

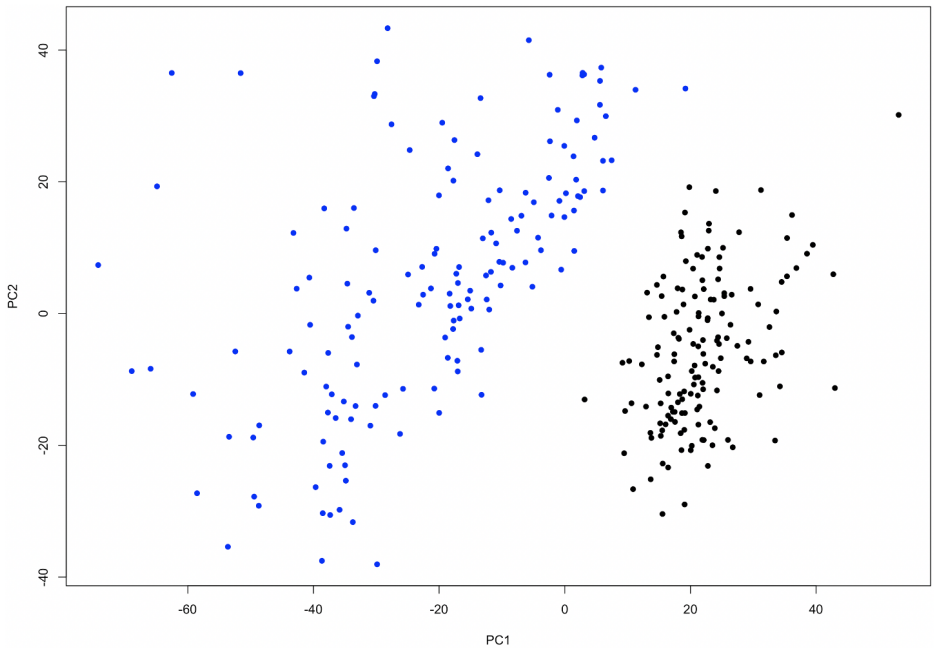


Figure 4.11: A 2D comparison is shown to highlight the angle along PC1 and PC2 is able to generate 100% separation between PNT2-C2 live and PNT2-C2 dried.

Figure 4.8 observes the fingerprint PCA space using the first three principal components. From this orientation it is observed that there is clear separation between PNT2-C2 live and dried. This separation is generated along PC1. There is no separation occurring along PC3 at this angle. Further rotations of PCA space were observed to investigate this. The 3D PCA space is then rotated for observation between PC1 and PC2. It is observed in Figure 4.9 that separation of the fingerprint region of the two populations is occurring along the axis of PC1 and PC2. Figure 4.10 considers the impact of PC3 further. It can be seen, however, that whilst there is some separation along PC2, separation is not occurring along PC3. It is for this reason that the biochemical components of PC3 will not be explored. This in turn means that the loading corresponding to PC3 is not considered as no separation is observed in PC3. A 2D plot of PC1 and PC2 is shown in 4.11 showing 100% separation is occurring between the fingerprint regions of PNT2-C2 live and dried in PCA space. The separation in PC1 can be analysed by plotting this principal component in the form of loading 1. The spectra for loading 1 is shown in 4.12. The loadings corresponding to PC2 are plotted in the form of loading 2 in Figure 4.13.

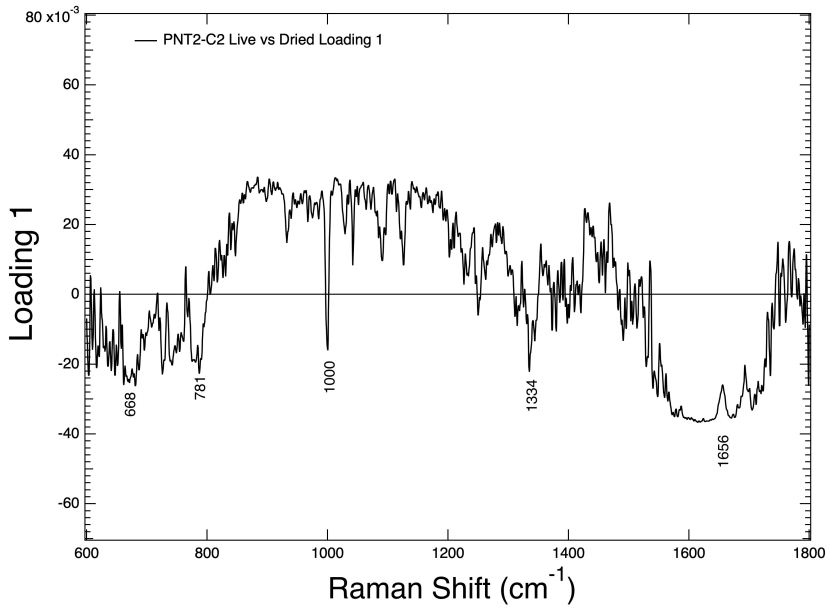


Figure 4.12: PCA informed analysis indicates that separation between PNT2-C2 live and dried in the fingerprint region is occurring in PC1. Here, loading 1 of the fingerprint region comparison of PNT2-C2 live vs dried is displayed. Key peaks attributed to biological assignments have been found 668 cm^{-1} (DNA/RNA and proteins), 781 cm^{-1} (DNA/RNA), 1001 cm^{-1} (proteins), 1334 cm^{-1} (DNA/RNA and proteins) and 1656 cm^{-1} (proteins and lipids). Careful attention should be paid to the shape of the loading profile. The large relative variance attributed to the regions in between identifiable peaks suggests variance of the background region of both PNT2-C2 live and dried.

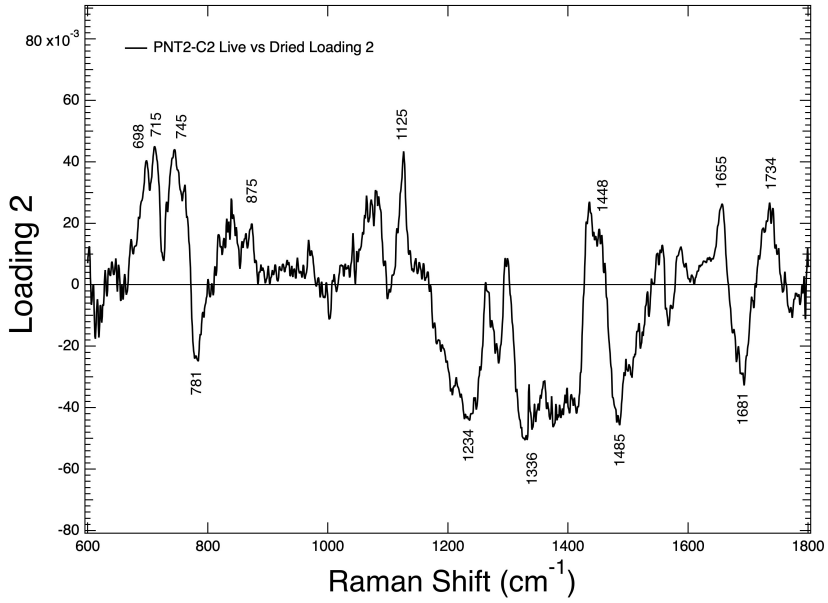


Figure 4.13: Loading 2 has been plotted and annotated to detail the biomolecular assignments that contribute maximum variances derived from PC2. This loading is derived from the result of PC2 which can be seen in Figures 4.9 and 4.11. The peaks found to majorly contribute to PC2 are as follows: 698 cm^{-1} (proteins), 715 cm^{-1} (proteins and lipids), 745 cm^{-1} (DNA), 781 cm^{-1} (DNA/RNA), 875 cm^{-1} (proteins and lipids), 1080 cm^{-1} (DNA. lipids and proteins), 1234 cm^{-1} (DNA/RNA and proteins), 1336 cm^{-1} (DNA/RNA and proteins), 1448 cm^{-1} (lipids, proteins and DNA/RNA), 1485 cm^{-1} (protein and DNA/RNA), 1656 cm^{-1} (proteins and lipids), 1681 cm^{-1} (proteins), 1734 cm^{-1} (lipids).

It can be seen that separation is occurring in PC1, which, through observation of loading 1 in Figure 4.12, shows variation between samples at the following peaks: 668 cm^{-1} (DNA/RNA and proteins), 781 cm^{-1} (DNA/RNA), 1001 cm^{-1} (proteins), 1334 cm^{-1} (DNA/RNA and proteins) and 1656 cm^{-1} (proteins and lipids). The DNA/RNA and protein band 668 cm^{-1} is specifically representative of a C-C twist of tyrosine and cystine. The DNA/RNA and protein band 781 cm^{-1} is specifically attributed to uracil, cytosine and thymine ring breathing modes. The 1001 cm^{-1} peak, which specifically represents the aromatic amino acid phenylalanine is also a well defined band in loading 1. Variations around the peak 1334 cm^{-1} also suggest contributions from amide III and protein/lipid modes [128]. Finally, there is a large broad difference observed around the 1656 cm^{-1} peak representative of lipids and amide I. It is apparent from Figure 4.12 that loading 1 has a relatively high level of variance attributed to the background

between between the two sets of spectra. It is a particularly noisy spectrum with few defined peaks. The strongest evidence for this is the very wide profile dip around the 1656 cm^{-1} peak position. The peaks identified in loading 1 will be further evaluated using peak intensity ratio analysis in Section 4.5.

Loading 2 corresponding to PC2 is now considered in figure 4.13. Loading 2 shows points of variance occur at positions 698 cm^{-1} (proteins), 715 cm^{-1} (proteins and lipids), 745 cm^{-1} (DNA), 781 cm^{-1} (DNA/RNA), 875 cm^{-1} (proteins and lipids), 1080 cm^{-1} (DNA. lipids and proteins), 1234 cm^{-1} (DNA/RNA and proteins), 1336 cm^{-1} (DNA/RNA and proteins), 1448 cm^{-1} (lipids, proteins and DNA/RNA), 1485 cm^{-1} (protein and DNA/RNA), 1656 cm^{-1} (proteins and lipids), 1681 cm^{-1} (proteins), 1734 cm^{-1} (lipids). Particular attention will be paid to the peaks in loading 1 and 2 at the PIR stage in 4.5 as both of these are present in loadings 2 and 3.

The last peak present in loading 2 was annotated although it does not appear in any peak fitting analysis or mean subtraction. The reason for its inclusion in loading 2, Figure 4.13 is that several other studies have mentioned this band as a biomarker [249, 259]. The band 1734 cm^{-1} is biochemically representative of the C=O ester group in lipids [249, 259]. It was reported by Notingher *et al.* (2003) that this peak, although small in intensity, could be used to confirm the oxidative damage that occurs during cell death. It is not unreasonable to assume that the presence of this peak in loading 2 signifies that PNT2-C2 dried exists in an oxidised state. However, as it is not identifiable at the mean profile subtraction stage or during peak fitting, it is not considered as a biomarker in this study.

The high wavenumber region between PNT2-C2 live and dried has also been analysed in PCA space. As the fingerprint region corresponds to the high wavenumber, in that both are generated from each measurement, the same number of spectra are selected to undergo PCA analysis for each region. This means that live N = 151, Dried N = 148 for both the fingerprint region and the high wavenumber region. The first six principal components in the high wave number region cumulatively account for 81% of the variance in the sample, capturing

28.5% more variance of the total sample when compared to the first six principal components in the fingerprint region. The first six principal components associated with the high wavenumber region are respectively weighted: PC1 = 38.8%, PC2 = 26.4%, PC3 = 7.4%, PC4 = 4.1%, PC5 = 2.8%, PC6 = 1.6%. Furthermore, the first three principal components for the high wavenumber capture 73% of the cumulative variance, as opposed to 52% of the first three PCs of the fingerprint region. The PCA results of the high wavenumber region are visualised using the first three principal components in Figures 4.14, 4.15 and 4.16.

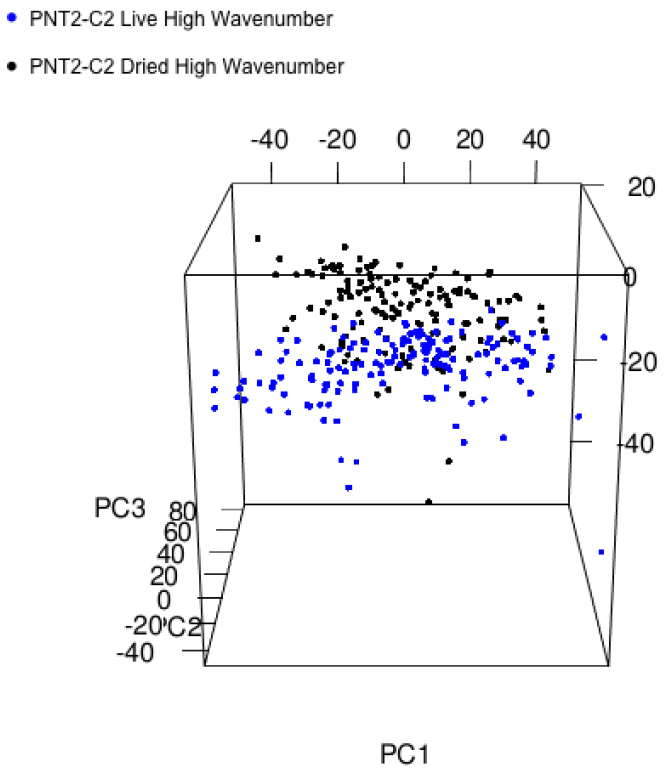


Figure 4.14: PCA analysis is displayed in a 3D format for comparison of the high wavenumber region of PNT2-C2 dried and live. It can be seen that there is no separation occurring along PC1. It can be seen that separation between the samples is occurring along PC3.

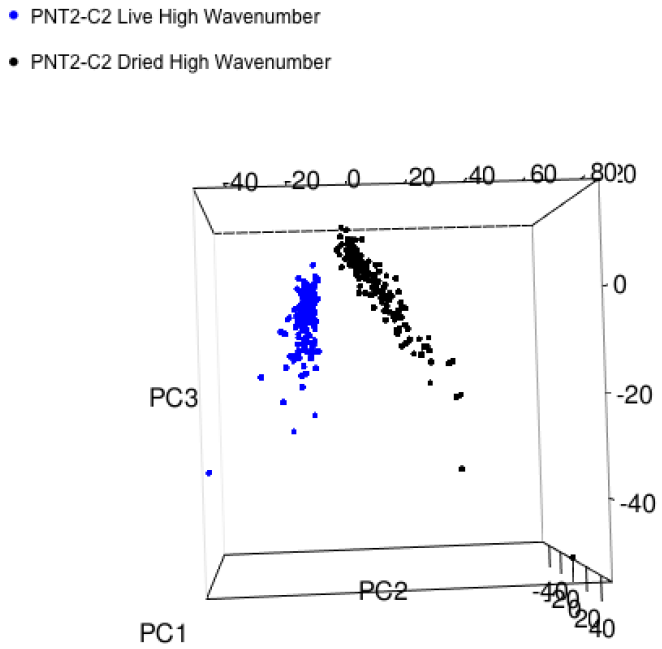


Figure 4.15: The second rotation of PCA space comparing the high wavenumber region of PNT2-C2 live and dried details the separation occurring at an angle along PC2 and PC3. Using this relationship it can be seen that 100% separation of the samples can be seen.

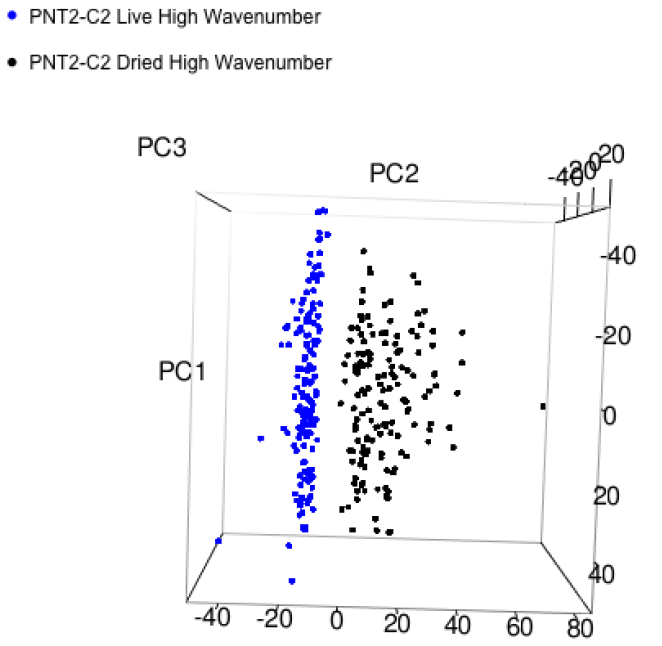


Figure 4.16: A final rotation of PCA space comparing the high wavenumber region of PNT2-C2 dried and live details the separation occurring between PC1 and PC2. It is seen that 100% separation of the samples can be seen along the PC2 axis. No separation is observed as happening along PC1.

The first visualisation of the 3D PCA plot of PNT2-C2 live vs dried in the high wavenumber region is shown in 4.14. This visualisation shows that there is no separation occurring in PC1. There does appear to be separation occurring in PC3 although it is not enough to completely separate both cell populations. Further analyses of Figures 4.15 and 4.16 show that 100% separation of the samples is possible using PC2. It also shows that there is some separation occurring in PC3 due to the angle of separation observed in Figure 4.15. However, no separation was seen to be occurring along PC1 in Figure 4.16. It is for this reason that the corresponding loading and biomolecular profile of PC1 is not considered. To better understand the biomolecular changes taking place in the process of air-drying, the loading corresponding to PC2 has been plotted with the corresponding peak positions in Figure4.17.

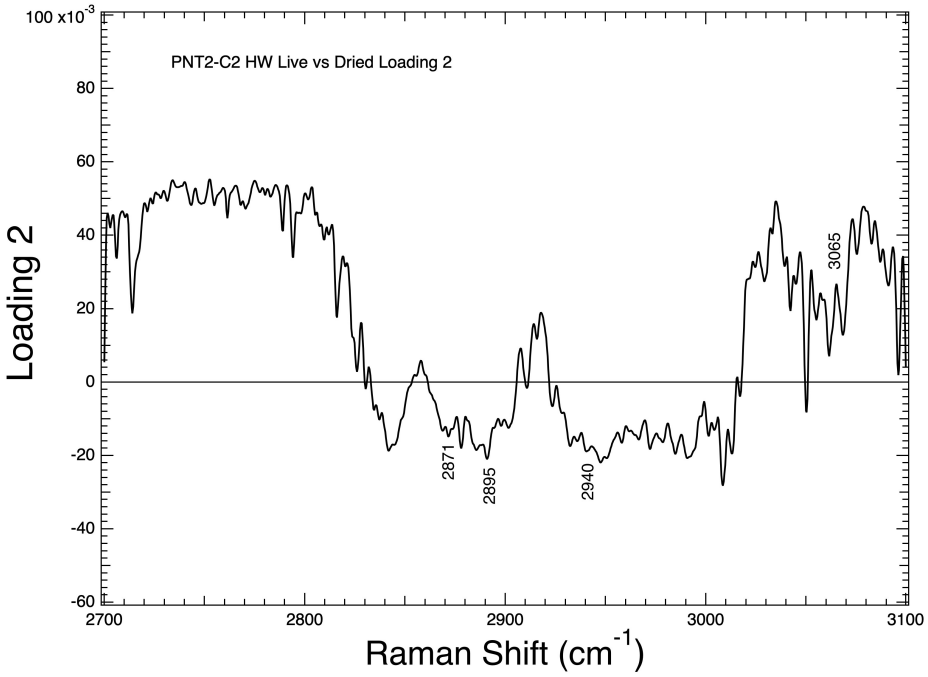


Figure 4.17: Here, loading 2 of high wavenumber comparison of PNT2-C2 live vs dried is displayed. It is apparent that the major reason for separation lies in the background profile of both peaks. The region 2700-2800 cm^{-1} is by far the most dominant part of the profile. Tentative labeling of peaks attributed to biological assignments have been highlighted. These are 2871 cm^{-1} (lipids and proteins), 2895 cm^{-1} (lipids and proteins), 2940 cm^{-1} (lipids and proteins) and 3065 cm^{-1} (lipids). Whilst there is a lot of variance attributed to the background variation, the biomolecular information related to the peak assignments overlaps with this.

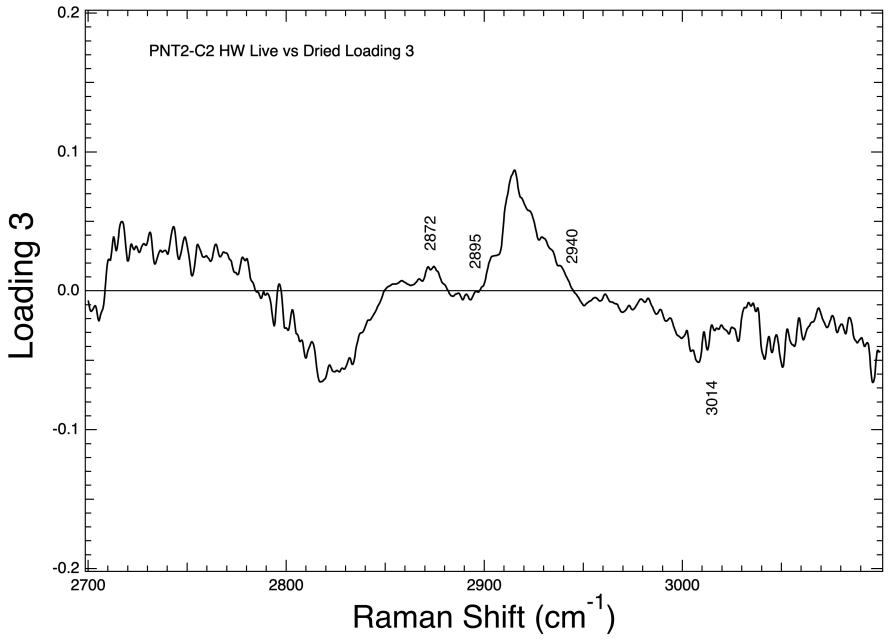


Figure 4.18: Loading 3 of high wavenumber comparison of PNT2-C2 live vs dried is displayed. Following the same trend as loading 2 the spectrum appears to be dominated by background signal. This does indicate that the separation being observed is due to background signal only. Tentative labelling of peaks has been done. These peaks are 2872 cm^{-1} (lipids and proteins), 2895 cm^{-1} (lipids and proteins), 2940 cm^{-1} (lipids and proteins) and 3014 cm^{-1} (lipids). It should be stated that these peaks have been added as reference to highlight the presence of background signal. There is no obvious biomolecular contribution to PC3 from any of these peaks.

There is a clear issue present in loading 2. The region 2700-2800 cm^{-1} is by far the most dominant part of the profile. This region is void of any peaks attributed to biomolecular assignments. Thus, it is appropriate to assume that the 100% classification is attributed to background profile and not key biomarkers present in the profile. Very tentative peak assignments were added for reference. The peaks 2871 cm^{-1} (lipids and proteins), 2895 cm^{-1} (lipids and proteins), 2940 cm^{-1} (lipids and proteins) and 3065 cm^{-1} (lipids), although labelled, are not well defined and should only be considered for reference. It is necessary to move to PC3 to see if molecular information can be extracted.

It can be seen in Figures 4.17 and 4.18 that there are strong influences from background signal contributing to the separation of PNT2-C2 live and dried. It is possible to extend the use of principal components further in the high wavenumber

region to try and capture biochemical interactions. As stated previously, as you choose higher principal components (PC4,PC5, PC6...) the corresponding variance weighting of these principal components decreases. It is important, however, to still try and capture significant interaction. It was therefore decided to examine PC4. The weighting associated with PC4 is 4.1%. The results of this are due to the large change in variance capture weightings between PC1 to PC4, where the high weightings of PC1 and PC2 would drown out results from PC4, mean that it is better to carry out this analysis in 2D space. The PCA plot in 2D space of PC3 against PC4 is shown in Figure 4.19

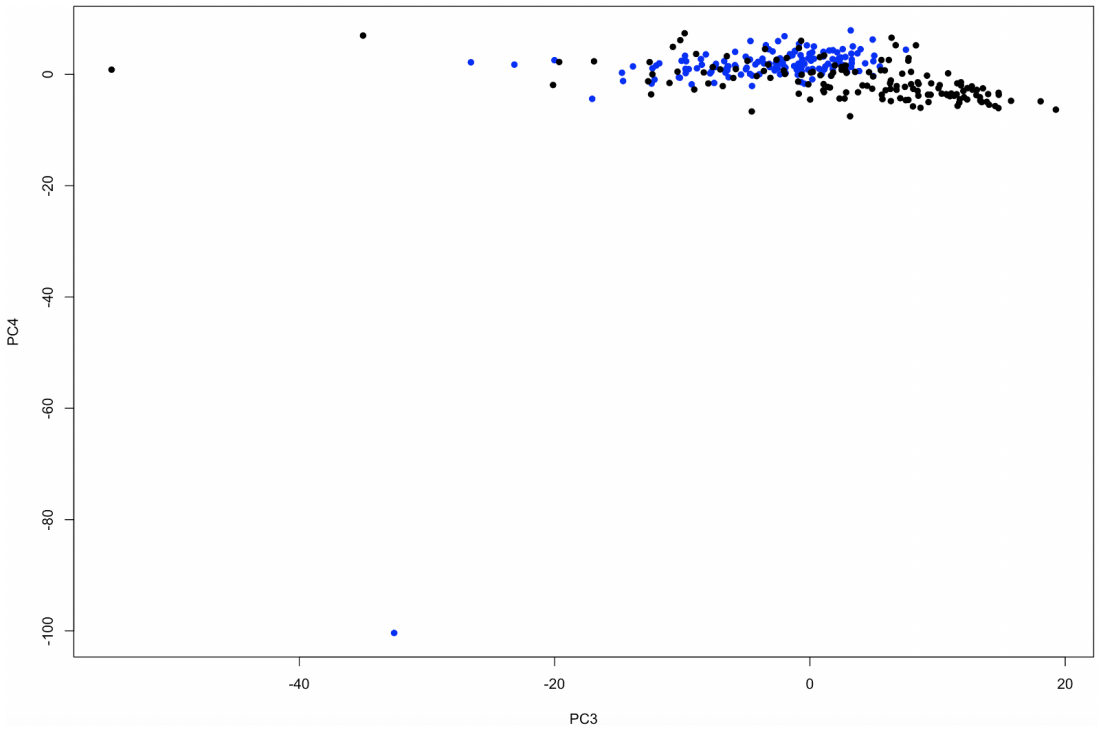


Figure 4.19: A 2D comparison is shown to highlight the angle of separation occurring between PC3 and PC4 in PNT2-C2 live vs dried. PC3 has a weighting of 7.4% and achieves some separation between PNT2-C2 live and dried. However, through analysis of its loading profile in 4.18 it is attributed to background signal. The principal component being examined is PC4, which has a weighting of 4.1%. It can be seen that there is separation in PC4. Whilst PC4 does not completely separate the samples, it is by enough of a margin to investigate further through loading analysis.

The 2D comparison of PC3 against PC4 shows that there is separation occurring along PC4. Whilst there are a number of points that overlap in each sample

when using this principal component, there is enough of a separation to consider the loading profile to extract biomolecular information. The loading for PC4 is considered in Figure 4.20

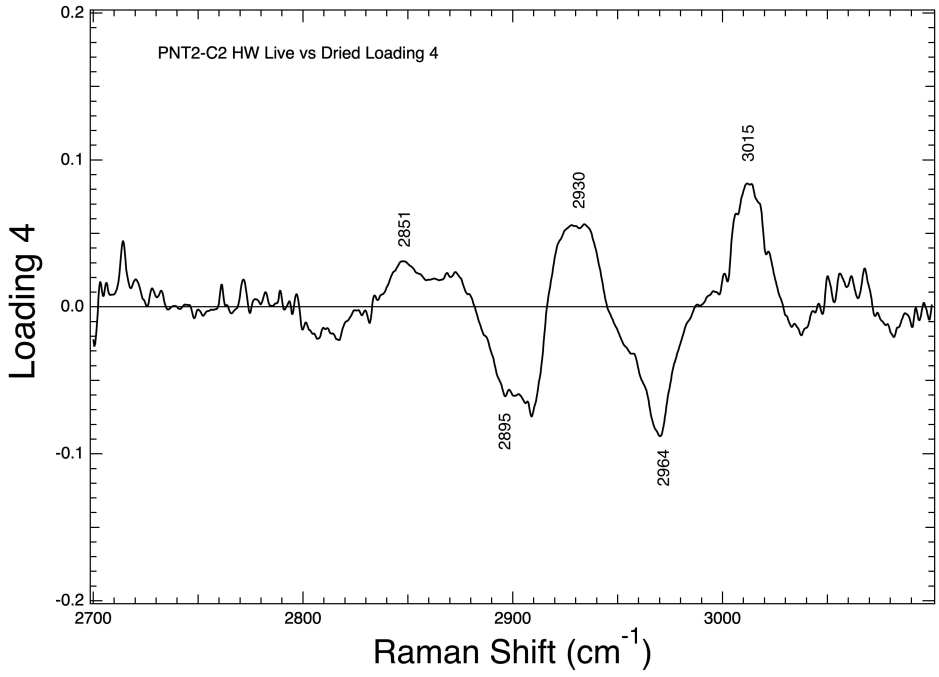


Figure 4.20: It can be seen in Loading 4 that the emergence of peaks takes place out of the background variance profile. There are clear defined peaks at 2851 cm^{-1} (lipids), 2895 cm^{-1} (lipids and proteins), 2930 cm^{-1} (proteins and lipids), 2964 cm^{-1} (lipids) and 3015 cm^{-1} (lipids). The peaks 2964 and 3015 cm^{-1} are the most dominant in the profile of loading 4.

It is observed in loading 4 (Figure 4.20) that there are clear, well defined peaks. These peaks are 2851 cm^{-1} (lipids), 2895 cm^{-1} (lipids and proteins), 2930 cm^{-1} (proteins and lipids), 2964 cm^{-1} (lipids) and 3015 cm^{-1} (lipids). This finding infers that principal components PC1, PC2 and PC3, although achieving strong separation, it is not primarily as the result of interactions between peaks, but background signal. The most dominant peak found in loading 4 was 2964 cm^{-1}, specifically attributed to a CH_3 stretch in lipids as well as cholesterol. This peak will be further analysed using peak intensity ratio analysis in Section 4.4. The second most dominant peak in loading 4 is the 3015 cm^{-1} peak, specifically attributed to unsaturated =CH stretch in lipids. This peak will also be further analysed using peak intensity ratio analysis in Section 4.4.

It is not necessary to carry out training and cross-validation on the PCA results between PNT2-C2 live and dried, as 100% separation is observed in both the fingerprint and high wavenumber region. It is noted that the separation that occurs in the high wavenumber region is the product of background signal. Nevertheless, as a classification tool it still groups PNT2-C2 live and dried correctly. With the evidence presented in terms of convergence of data in Section 4.2 there is capability for this classification to be considered for AI classification.

4.5 Comparison of peak intensity ratios

Peak intensity ratio analysis has already been carried out by others in the field of Raman spectroscopy on single cells. There has been significant work detailing how the relative intensities of DNA/RNA, protein and lipid peak bands can be exploited to find key biological differences between cells. More importantly, each Raman band is generated by specific molecules, but they also operate relative to each other, as part of a much more complex interconnected environment. It is important to fully understand this interconnected environment in terms of the relationships between bands, as even small changes in the Raman spectrum can signify complex changes in the biochemical environment of the cell [260]. As identified in the literature, there are some bands that can be used as biomarkers to quantify the complexities and changes that are occurring in cells. One such study by Potcoava *et al.* (2014) investigated changes in lipid content and composition between two cancerous cell lines in both the fingerprint and high wavenumber region [64]. The lipid composition was able to be analysed by comparing lipid peak intensities across the fingerprint and high wavenumber region. They were able to correlate the relationship between the 1656 cm^{-1} peak intensity, representative of C=C bonds and changes to mass saturation of the sample. Mass unsaturation, calculated by the ratio of C=C bonds present to the number of CH_2 bonds, allows for determination of changes to the lipid content of cells. This ratio can thus be calculated using the relative intensity of the 1448 cm^{-1} peak, which is attributed to CH_2 bonds. Similarly, the 1306 cm^{-1} peak is also representative of CH_2 bonds and the relative intensity can also be used to calculate mass unsaturation. This

is possible because the 1656 cm^{-1} band is described as varying linearly with the number of C=C bonds [64]. Potcoava *et al.* (2002) also goes on to state that it is possible to calculate the ratio of total unsaturated fatty acids to total fatty acids. This ratio is coined TUFA/TFA and is done by using the ratio of the 3015 peak cm^{-1} intensity (Unsaturated =CH stretch in lipids) and the 2850 cm^{-1} peak intensity(CH_2 symmetric stretch in lipids) [64, 140].

It should be noted that whilst Potcoava *et al.* (2002) compared cancerous prostate and breast cell lines in their study, they did not include a normal cell line to compare the results against. This study will replicate their analysis, but for the first time will use these ratio relationships to probe the differences between live and dried cells. The same methodology was used to study DNA conformation by Draux *et al.* (2009) [220]. The Raman bands analysed are not comparable due to their use of a 785nm laser. They measured the non-small cell lung cancer cell line in both a live and dried cell state. A focal point of their analysis was determining whether there was a relationship between guanine and thymine attributed to a peak found at 670 cm^{-1} and the guanine peak found at 680cm^{-1}. They postulated that whether the ratio I670/680 was less than (B-type) or greater than (A-type) would indicate whether the DNA was in a B or A-type conformation state, [220]. The results from Draux *et al.* (2009) are not directly comparable for this study. The Raman band 680cm^{-1} does not exist in our spectrum. However, as mentioned in Section 4.3 there is evidence suggesting that the 745 cm^{-1} peak intensity can be used to probe the conformation state. Every other Raman band has undergone peak intensity ratio analysis against the band 745 cm^{-1} to try and find a new relationship that could be used as a test for conformation state of DNA. This is found in table C.1.

It was decided to measure all ratios against each other to better understand the complex environments. There are other studies that although have not quantified ratios specifically, have alluded to relationships between certain Raman bands. Both Shim *et al.* (1996) and referenced therein Tu, A. T. (1986) discussed a change in the secondary protein structures as a function of drying time, reporting a decrease in the intensity of the band 930 cm^{-1} (skeletal C-C stretch α-helix of

proline and valine) relative to the 1655cm^{-1} and 1450 cm^{-1} [237, 261]. However, there was no comparison to a live cell during the investigation by Shim *et al.* (1996). This thesis will cover these Raman bands and their relationships and also measure them against a wider array of Raman bands. The peak intensity ratios have been calculated using the methods described in Section 3.6.3 for PNT2-C2 live vs dried. The detailed analysis of separation between relative peak intensities is found in Table C.1 in the Appendix.

Table C.1 shows the peak intensity ratio analyses used to determine the amount of separation between all identified Raman bands in PNT2-C2 live and dried. This is done to determine the largest separation of Raman peaks outside of the propagated SE uncertainty relating to each peak intensity ratio between PNT2-C2 live and dried. There were 5 relative peak intensities identified that achieved separation of 90% or greater between Raman bands in PNT2-C2 live and dried. The Raman bands and the percentage of separation are as follows: 724 cm^{-1} (98%), 745 cm^{-1} (98%), 1256 cm^{-1} (95%), 1681 cm^{-1} (90%) and 2940 cm^{-1} (90%). It is also important to consider the intensity of 3014 cm^{-1} peak which generated 67% separation between all relative peak intensity rations, however, generated 100% separation in the high wavenumber region. It is important to understand why these peaks are so dominant in terms of separation to probe the biochemical changes happening in air-drying.

PCA analysis between PNT2-C2 live and PNT2-C2 dried in section 4.4 was able to discriminate between samples in both the fingerprint and high wavenumber regions. To further interpret the differences in variance captured by the PCs, biomolecular information was extracted from the principal components that generated the separation. In the fingerprint region loading 1, Figure 4.12 shows variation between samples at the following peaks: 668 cm^{-1} (DNA/RNA and proteins), 781 cm^{-1} (DNA/RNA), 1001 cm^{-1} (proteins), 1334 cm^{-1} (DNA/RNA and proteins) and 1656 cm^{-1} (proteins and lipids). Further biomolecular information was extracted from the fingerprint region from loading 2, Figure 4.13, capturing variance at the peak positions: 698 cm^{-1} (proteins), 715 cm^{-1} (proteins and lipids), 745 cm^{-1} (DNA), 781 cm^{-1} (DNA/RNA), 875 cm^{-1} (proteins and lipids), 1080

cm^{-1} (DNA. lipids and proteins), 1234 cm^{-1} (DNA/RNA and proteins), 1336 cm^{-1} (DNA/RNA and proteins), 1448 cm^{-1} (lipids, proteins and DNA/RNA), 1485 cm^{-1} (protein and DNA/RNA), 1656 cm^{-1} (proteins and lipids), 1681 cm^{-1} (proteins), 1734 cm^{-1} (lipids). PCA results from the high wavenumber region also demonstrated that there was 100% separation between samples. Figure 4.17 shows that key points of variances in loading 2 occurred at the peak positions: peaks 2871 cm^{-1} (lipids and proteins), 2895 cm^{-1} (lipids and proteins), 2940 cm^{-1} (lipids and proteins) and 3065 cm^{-1} (lipids). Furthermore, loading 3, as shown in Figure 4.18 was able to capture variance at the peaks: 2871 cm^{-1} (lipids and proteins), 2895 cm^{-1} (lipids and proteins), 2940 cm^{-1} (lipids and proteins) and 3065 cm^{-1} (lipids). Lastly, it is observed in loading 4 (Figure 4.20) that variance was captured at the peaks: 2851 cm^{-1} (lipids), 2895 cm^{-1} (lipids and proteins), 2930 cm^{-1} (proteins and lipids), 2964 cm^{-1} (lipids) and 3015 cm^{-1} (lipids). To link the PCA and PIR analysis, only the peaks responsible for capturing the most variance at both stage are considered.

It should be highlighted that it is not possible to compare the 698 cm^{-1} peak, representative of amino acid methionine ν(C-S) *trans*, using peak intensity ratio analysis due to it being absent from PNT2-C2 dried. This band was identified as being absent at the characterisation stage in Section 4.3. Furthermore, at the PCA stage, the 698 cm^{-1} was identified in loading 2 as capturing significant variance between samples. This absence has had a knock-on effect because no peak intensity ratio analysis can be carried out on it. Thus, although PCA analysis has determined that it is a key point of variance between samples, the changes observed in the band can not be quantified using PIR analysis. It is for this reason that even though there is strong variance captured in loading 2 at the PCA stage, the absence of the 698 cm^{-1} band is not recommended as being a robust biomarker of separation between the samples.

The first band that is examined from the PIR stage is the 724 cm^{-1} band. The band 724 cm^{-1} has a detailed biomolecular assignment of DNA and Adenine ([128, 142, 224, 245, 247]). There were significant changes expected in all DNA bands as the DNA changes conformation state as it enters a dried state. However,

the 98% degree of separation between other bands was not expected. There was no obvious emergence of a 724 cm^{-1} in the mean subtraction in Section 4.3. This is important to note as it shows that simple visual analysis of Raman profiles and basic analysis such as baseline subtractions may miss important biochemical changes that are taking place. Furthermore, the 724 cm^{-1} band was not identified at the PCA stage in Section 4.4 as being a peak associated with any significant variance. To better understand the 724 cm^{-1} peak intensity it is important to consider what other bands exert influence on it. The most dominant band in the fitting region used to fit the 724 cm^{-1} band is the 715 band cm^{-1}. Furthermore, the 724 cm^{-1} band profile appears in the tail of the 715 cm^{-1} band profile. Additionally, the 715 cm^{-1} peak was identified as capturing very high variance between samples in loading 2. At the PIR stage the 715 cm^{-1} peak intensity, representative of the amino acid methionine and $CN^{+}(CH_3)_3$ stretching in lipids, has observable separation between 55% of other Raman bands. The lower observable separation is thought to be caused by the contribution of the lipid contributions; lipids are observed as not being as affected by the air-drying process [225]. Despite the lower separation observed when using the 715 peak cm^{-1} it does dominate the peak region in which the 724 cm^{-1} band exists and this convolution in turn makes the 724 cm^{-1} peak an unreliable biomarker.

The 745 cm^{-1} peak was also able to generate separation between 98% of the other Raman bands using PIR analysis. The 745 cm^{-1} peak was also identified at the PCA stage in loading 2, Figure 5.97, capturing significant variance between PNT2-C2 live and dried. The 745 cm^{-1} peak is generated by a pure thymine ring breathing mode in DNA. This band is in a different fitting region to 724 cm^{-1} but is present next to it in the spectrum. To further draw comparison to the 724 cm^{-1} peak, the 745 cm^{-1} peak was very pronounced in the mean subtraction profile in 4.3. The separation resulting from the 745 cm^{-1} peak PIR ratios is in line with expectation found from the mean subtraction in Section 4.3. The hypothesised reason for the large degree of separation is due to the conformation state change of DNA from a B to an A state [225, 237, 261]. This is expected as the hydration shell of water molecules structurally supporting the DNA is no longer present in

its dried state [262]. The looser unraveled A state will not generate as an intense Raman signal. In an effort to better understand the relationships between other biochemical changes taking place, the 745 cm^{-1} PIR panel detailing all ratios examined is shown in Figure 4.21.

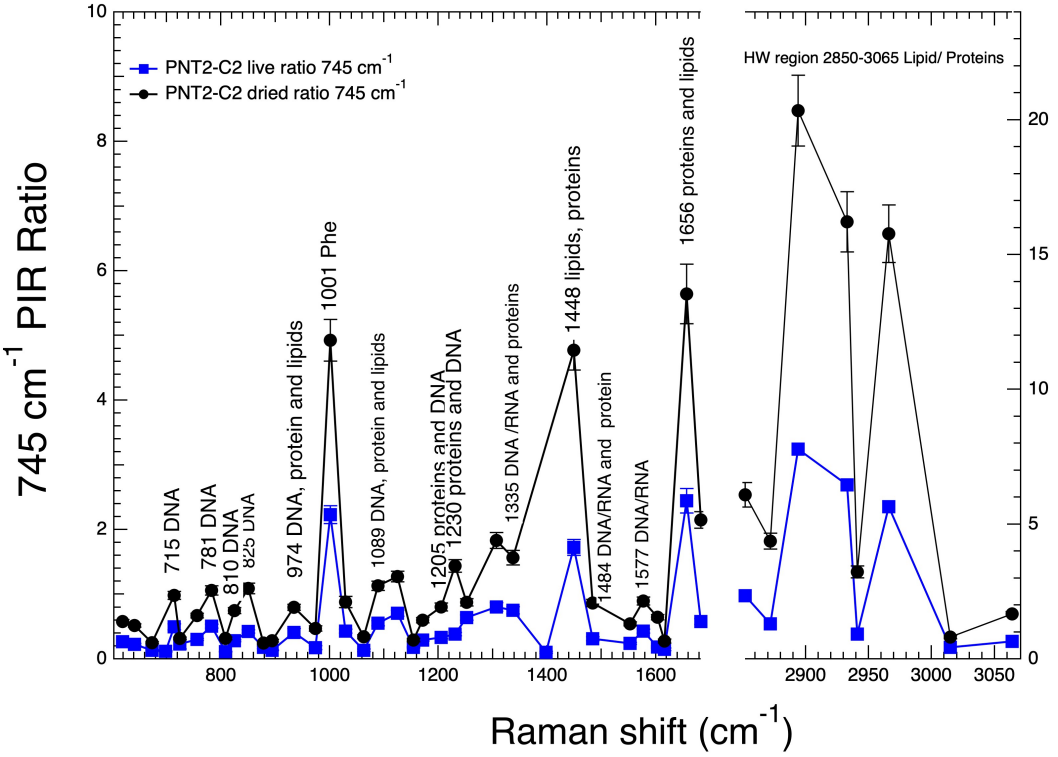


Figure 4.21: The peak intensity ratio panel for 745 cm^{-1} peak intensity, demonstrating statistically significant separation between 42 out of 43 Raman bands across the fingerprint region and high wavenumber region (HW region), of 98% separation. It is hypothesised that because the 745 cm^{-1} peak is the most relatively intense DNA peak that is not convoluted with other contributions from lipids and proteins that it could be used as a tool to measure whether DNA is in B or A type state. The 745 cm^{-1} peak, specifically, is representative of the thymine ring breathing mode. The largest separation in the fingerprint region outside of uncertainty is between the 1001 cm^{-1} cm^{-1} (phenylalanine), 1448 cm^{-1} (CH_2 deformation) and 1656 cm^{-1} (protein and lipids). This demonstrates, using the 745 cm^{-1} peak, that there is a higher relative ratio of thymine to proteins and lipids in PNT2-C2 live than PNT2-C2 dried. It is also noted that the 745 cm^{-1} peak generates high separation outside of uncertainty in the high wavenumber region. This further illustrates the increased relative ratio of DNA to lipids and proteins in PNT2-C2 live.

The 745 cm^{-1} peak intensity is identified as suitable to establish key ratios that could be used to measure DNA conformation state change for future studies.

This is necessary because it is a very exhaustive process to calculate PIR ratios for every peak against every other peak. A proposed method to identify specific ratios is to consider the largest separation between peaks in the fingerprint region outside of uncertainty. As shown in Figure 4.21 there is large separation outside of uncertainty in the fingerprint region between the peaks at 1001 cm^{-1} (phenylalanine), 1448 cm^{-1} (CH_2 deformation in lipids and proteins) and 1656 cm^{-1} (Amide I C=O stretch and C=C stretch of lipids). The peak positions at 1001 cm^{-1}, 1448 cm^{-1} and 1656 cm^{-1} were identified in the mean profile subtraction in Figure 4.5 as being key points of variance, as well as the 745 cm^{-1} peak. Moreover, all peaks were identified in the PCA loadings between PNT2-C2 live and dried as capturing significant variance: 745 cm^{-1} (loading 2, Figure 4.13), 1001 cm^{-1} (loading 1, Figure 4.12), 1448 cm^{-1} (loading 2, Figure 4.13) and 1656 cm^{-1} (loading 2, Figure 4.13). Another reason for selecting these peaks for use as PIR biomarkers is that they are all relatively intense bands and relatively well defined meaning that the bands are easily identified, increasing the reputability factor. The reason behind not using the 745 cm^{-1} peak as a PIR biomarker in conjunction with peaks in the high wavenumber region is not only that many studies only consider the fingerprint region but that the 1448 cm^{-1} (CH_2 deformation in lipids and proteins) peak already captures the separation between DNA, lipids and proteins[220, 224, 237, 261].

The 1256 cm^{-1} peak intensity separated 95% of the bands at the PIR stage. The 1256 cm^{-1} is generated by amide III : CH_2 wagging. To consider the 1256 cm^{-1} peak and its biomolecular assignment of amide III it is also necessary to consider other peaks in the region that are all generated wholly or partially by amide III 1205-1337 cm^{-1}. Although the peak 1256 cm^{-1} was not identified at the mean profile subtraction stage in Section 4.3, the peak 1338 cm^{-1}, representative of proteins, lipids, nucleic acids and amide III was identified. This would suggest that there is a difference between relative amide III intensities between PNT2-C2 live and dried. Additionally, the fitting region that the 1256 cm^{-1} peak exists in is also strongly dependent on the presence of the 1234 cm^{-1} peak in the same region. The 1234 cm^{-1} peak is specifically representative of amide III, CH_2 wagging of

glycine and proline as well as a PO_2^- antisymmetric stretch. Furthermore, the 1234 cm^{-1} band achieved 83% separation at the PIR stage. It is important to note that the 1256 cm^{-1} peak was not identified at the PCA stage of analysis as capturing any variance. However, the 1234 cm^{-1} band was identified as capturing significant variance in loading 2, Figure 4.13. It is also seen in Figure 4.13 that variance associated to the 1234 cm^{-1} peak position is a relatively wide band. Thus it is possible that the 1256 cm^{-1} peak, although not obviously present, is having an effect on the background variance between PNT2-C2 live and dried. The PIR panels generated for the 1234 cm^{-1} and 1256 cm^{-1} peaks are shown in Figures 4.22 and Figures 4.23.

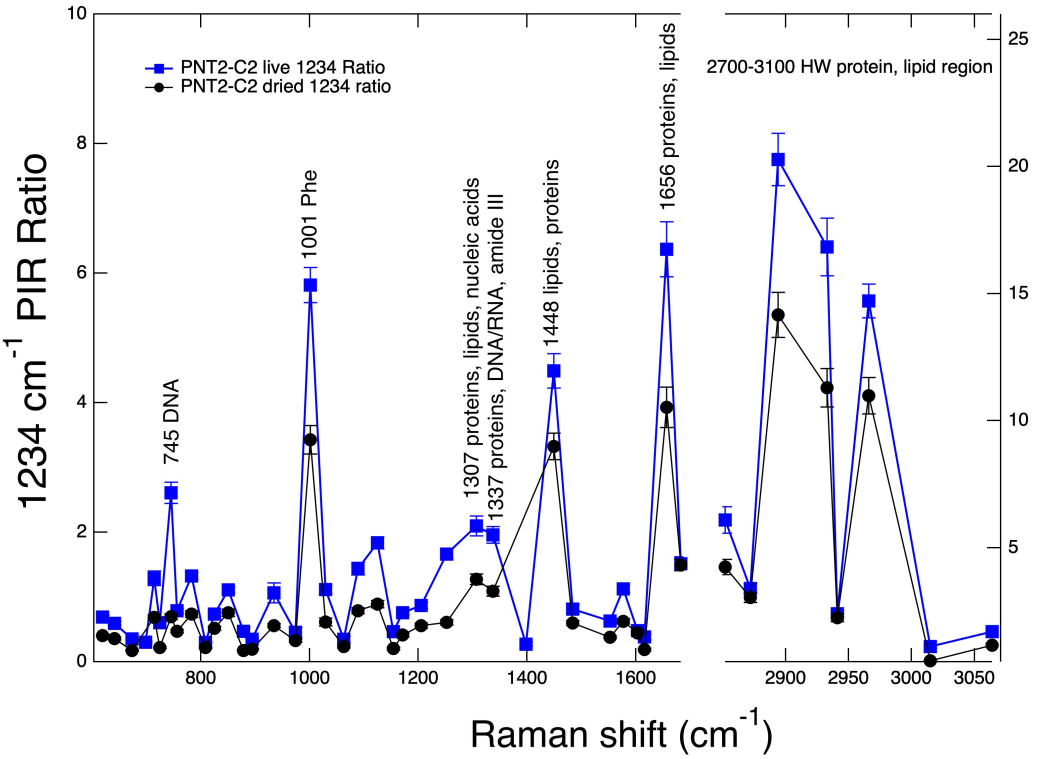


Figure 4.22: The PIR panel for the 1234 cm^{-1} peak intensity ratio demonstrating statistically significant separation across the fingerprint and high wavenumber region, achieving 83% total separation. The 1234 cm^{-1} peak is attributed, specifically, to amide III, CH_2 wagging of glycine and proline as well as a PO_2^- antisymmetric stretch. It is not possible to quantify what contribution amide III, CH_2 wagging of glycine and proline as well as a PO_2^- antisymmetric stretch have on the relative intensity of the band. If the 1234 cm^{-1} peak is used as a biomarker it would infer that there is a higher relative presence of amide III in PNT2-C2 dried. This observation is made due to the decrease in relative intensities of the peaks 745 cm^{-1}, 1001 cm^{-1}, 1448 cm^{-1} and 1656 cm^{-1}. This statement is in direct conflict with the PIR panel associated to the 1256 cm^{-1} peak (amide III : CH_2 wagging), Figure 4.23. It is also not possible to quantify the changes in relative intensities of the peaks 1307 cm^{-1} and 1337 cm^{-1} as these bands are made up from multiple biomolecular contributions, including amide III. It is for this reason that both the 1234 cm^{-1} and 1256 cm^{-1} peaks are not considered robust biomarkers.

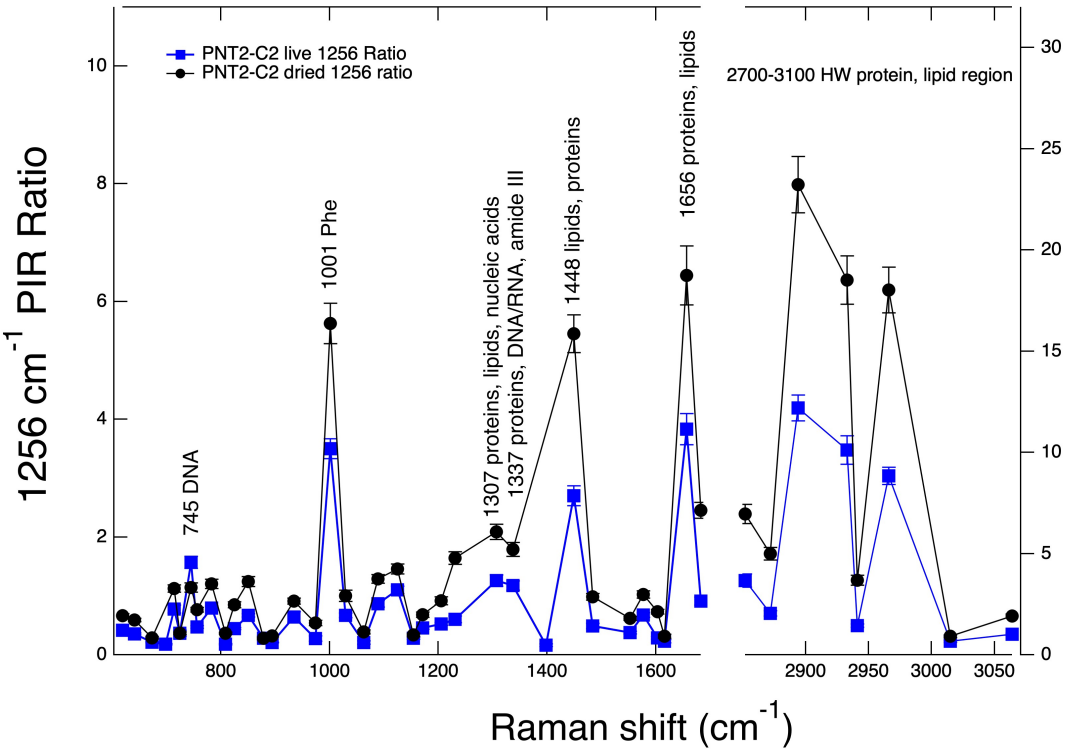


Figure 4.23: The PIR panel for the 1256 cm^{-1} peak intensity ratio demonstrating statistically significant separation across the fingerprint and high wavenumber region, achieving 95% total separation. The 1234 cm^{-1} peak is attributed, specifically, to amide III : CH_2 wagging. If the 1256 cm^{-1} peak is used as a biomarker it would infer that there is a higher relative presence of amide III in PNT2-C2 live. This observation is made due to the increase in relative intensities of the peaks 745 cm^{-1}, 1001 cm^{-1}, 1448 cm^{-1} and 1656 cm^{-1}. This statement is in direct conflict with the PIR panel associated to the 1234 cm^{-1} peak (amide III : CH_2 wagging), Figure 4.22. It is also not possible to quantify the changes in relative intensities of the peaks 1307 cm^{-1} and 1337 cm^{-1} as these bands are made up from multiple biomolecular contributions, including amide III. It is for this reason that both the 1234 cm^{-1} and 1256 cm^{-1} peaks are not considered robust biomarkers.

Figure 4.22 shows the PIR panel associated with the 1234 cm^{-1} peak intensity ratio. If the 1234 cm^{-1} peak is used as a biomarker it would infer that there is a higher relative presence of amide III in PNT2-C2 dried. This observation is made due to the decrease in relative intensities of the peaks 745 cm^{-1}, 1001 cm^{-1}, 1448 cm^{-1} and 1656 cm^{-1}. However, there is conflict with the 1256 cm^{-1} PIR panel. If the 1256 cm^{-1} peak is used as a biomarker it would infer that there is a higher relative presence of amide III in PNT2-C2 live. This observation is made due to the increase in relative intensities of the peaks 745 cm^{-1}, 1001

cm^{-1}, 1448 cm^{-1} and 1656 cm^{-1}. This obviously can not be the case as these propose two conflicting conclusions. In the peak 1234 cm^{-1} it is not possible to quantify what the individual contributions of amide III, CH_2 wagging of glycine and proline as well as a PO_2^- antisymmetric stretch have on the relative intensity of the band. That being said, the contribution of amide III of the 1234 cm^{-1} can not be ignored. Work carried out by Mariani *et al.* (2009) investigating human dermal derived keratinocyte found that there was a detected difference in the amide III peak occurring between various fixation methods, formalin fixation, air-drying and desiccation [225]. However, the study by Mariani *et al.* (2009) did not involve any live cell lines. Furthermore, their hypothesised reason for the use of air-drying leading to an increased keratin presence would not be applicable to this study. It is hypothesised in this thesis that there is overall structural change to amide III during the drying process, however, there is no biomarker in this comparison that can definitively determine the relative intensity ratios of the amide III contribution. It is for this reason that both the 1234 cm^{-1} and 1256 cm^{-1} peaks are not considered robust biomarkers.

The peak position 1336 cm^{-1} was identified at the PCA stage in loading 1 as contributing a relatively large source of separation between PNT2-C2 live and dried. The 1336 cm^{-1} peak itself, at the PIR stage only achieves 50% separation between all ratios between markers. However, it was identified in the mean profile subtraction in Figure 4.5 as being a point of variance between the samples. The 1336 cm^{-1} band is assigned to CH_3CH_2 wagging and twisting, lipids, tryptophan, adenine, guanine, amide III, protein C-H deformation, phenylalanine and CH_2 deformation in lipids. Due to the very high number of biomolecular contributions it is decided to not explore the 1336 cm^{-1} peak as general biomarker.

The 1681 cm^{-1} peak intensity was also able to distinguish between (90%) of the other Raman peaks in the fingerprint and high wavenumber region using PIR analysis. The 1681 cm^{-1} peak was also identified at the PCA stage in loading 2, Figure 5.97, capturing significant variance between PNT2-C2 live and dried. The 1681 cm^{-1} peak is generated by amide I. Although the 1681 cm^{-1}peak was not annotated during the mean profile subtraction in Section 4.3 it is considered

as contributing to the difference as observed through the annotation of the 1656 cm^{-1} peak; the 1681 cm^{-1} peak is located in the peak tailing of the 1656 cm^{-1} peak region, identified in the mean subtraction. The PIR panel associated with the 1681 cm^{-1} peak is shown in Figure 4.24.

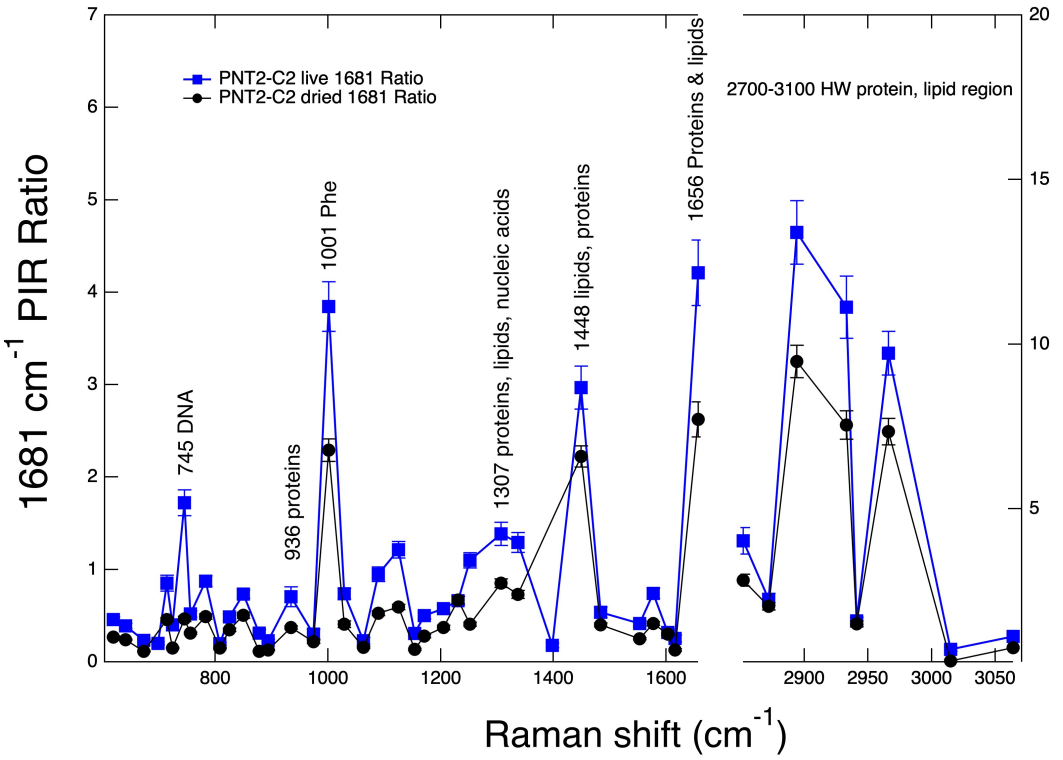


Figure 4.24: The PIR panel for the 1681 cm^{-1} peak intensity ratio demonstrating statistically significant separation across the fingerprint and high wavenumber region, achieving 90% total separation. The 1681 cm^{-1} peak is attributed, specifically, to Amide I. The PIR using the 1681 cm^{-1} peak shows that there is a relative increase of amide I in PNT2-C2 dried. This is observed as a relative decrease intensities of all peaks in both the fingerprint and high wavenumber region across DNA, proteins, lipids and nucleic acids.

The PIR panel for the 1681 cm^{-1} amide I peak intensity ratio as shown in Figure 4.24 demonstrates that the there is a relative increase of amide I in PNT2-C2 dried compared to PNT2-C2 live. This marked difference is observed throughout out the PIR profile of the 1681 cm^{-1} peak in PNT2-C2 dried, a relative decrease intensities of all peaks in both the fingerprint and high wavenumber region across DNA, proteins, lipids and nucleic acids. This demonstrates that there is a change in the secondary protein structure between PNT2-C2 live and

PNT2-C2 dried. In the study by Shim *et al.* (1996) and referenced therein Tu *et al.* (1986) using near-infrared Raman spectroscopy they reported a decrease in the intensity of the band 930 cm^{-1} peak (skeletal C-C stretch α-helix of proline and valine) relative to the 1655 cm^{-1} peak (amide I) [237, 261]. It must be noted though that these results were derived from measurements on hamster tissues [237]. Furthermore, the study by Shim *et al.* (1996) did not measure any live cells [237]. However, this same relationship, a decrease in the intensity of the band 936 cm^{-1} peak (skeletal C-C stretch α-helix of proline and valine) relative to the 1681 cm^{-1} peak (amide I) moving from a live to dried state is exhibited by PNT2-C2 live and PNT2-C2 dried. The results from both the PCA analysis in section 4.3 and in the PIR analysis in Figure 4.24 demonstrate that the 1681 cm^{-1} peak is a robust biomarker to discriminate between PNT2-C2 live and dried.

In the highwavenumber region, the 3015 cm^{-1} peak intensity ratio, attributed to the unsaturated =CH stretch in lipids, is able to generate 100% separation between all other peaks in the highwavenumber region. The 3015 cm^{-1} peak was identified in the mean profile subtraction in section 4.3 as a point of variance between PNT2-C2 live and dried. The 3015 cm^{-1} peak was also identified as capturing the most variance in loading 4 at the PCA stage in Figure 4.20. The 3015 cm^{-1} peak has been used previously in the literature to differentiate between lipids in prostate and breast cancer by [64, 140] Nieva *et al.* (2012) and Potcoava *et al.* (2014). [64, 140]. Both authors have stated that it is possible to calculate the ratio of total unsaturated fatty acids to total fatty acids . This ratio is named TUFA/TFA and is done by using the ratio of the 3015 peak cm^{-1} (unsaturated =CH stretch in lipids) and the 2850 cm^{-1} peak (CH_2 symmetric stretch in lipids) [64, 140]. There were further peak intensity ratios proposed by Potcoava *et al.* (2014) for analysis of the high wavenumber region; for a sense of completeness, these ratios are also considered in Table 4.3

Table: Peak intensity Ratios (PIRs) from literature examined				
Cell line	I1656/1448 [64]	I1656/1306 [64]	I2930/2959 [64]	I3015/2851 [64, 140]
PNT2-C2 Live	1.4±0.1	3.0±0.3	1.1±0.1	0.18±0.02
PNT2-C2 Dried	1.2±0.1	3.1±0.2	1.0±0.1	0.13±0.01

Table 4.3: The peak intensity ratios as proposed by Nieva *et al.* (2012) and Potcoava *et al.* (2014) to measure lipid composition in the cells [64, 140]. The I1656/1448 ratio calculated between PNT2-C2 live and dried implies there is a statistical difference in the mass unsaturation scissor mode. However, there is no statistically significant separation between the ratio I1656/1306 which is attributed to the mass unsaturation twisting mode. Similarly, the ratio CH_2/CH_3 anti-symmetric stretching calculated using the peaks I2930/2959 shows no statistically significant separation. However, there is further statistically significant separation found using the ratio I3015/2850, a difference in the CH stretching and CH_2 symmetric stretching in the high wavenumber.

Peak intensity ratios proposed by Nieva *et al.* (2012) and Potcoava *et al.* (2014) in the fingerprint region are considered in the table 4.3. As shown in Table 4.3 the ratio I1656/1306, measuring the degree of mass unsaturation twisting mode is not able to statistically differentiate between the two samples. Furthermore, there is no statistically significant separation between PNT2-C2 live and dried using the ratio I1656/1448, investigating the mass unsaturation twisting mode. Additionally, analysis of ratios in the high wavenumber region, in Table 4.3, shows that there is no statistical significance found between the ratio I2930/2959. This ratio calculates the peak intensity ratio between the 2930 cm^{-1} peak (CH_2 asymmetric stretch) and the 2959 cm^{-1} peak (CH_3 asymmetric stretch cm^{-1}). However, there is statistically significant separation occurring in the ratio I3015/2851. This ratio, as previously discussed, relates to the total unsaturated fatty acids to total fatty acids (TUFA/TFA). The peak 3015 cm^{-1} is a pure lipid peak with no other contributions generated by unsaturated =CH stretch in lipids and cholesterolester. Additionally, the band 3015 cm^{-1} is also a pure lipid peak, CH_2 symmetric stretch in lipids. The PIR panel associated to the 3015 cm^{-1} is considered in Figure 4.25.

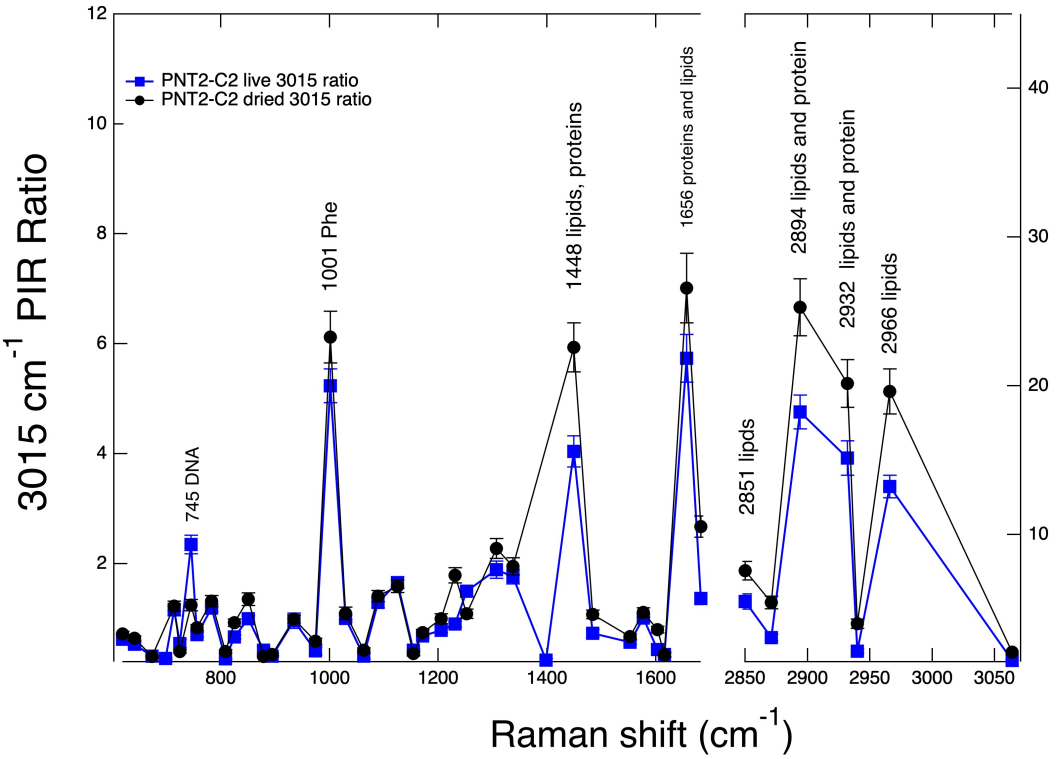


Figure 4.25: The PIR panel for the 3015 cm^{-1} peak intensity ratio demonstrating statistically significant separation across the fingerprint and high wavenumber region, achieving 67% total separation. It should be noted that the 3015 cm^{-1} peak intensity ratio was the only peak to achieve 100% separation in the high wavenumber region. The 3015 cm^{-1} peak is attributed, specifically, to unsaturated =CH stretch in lipids. The PIR using the 2015 cm^{-1} peak shows a lower relative presence of unsaturated fatty acids in PNT2-C2 dried against DNA, protein, lipids and nucleic acids. It is also noted that there is strong separation with the 2851 cm^{-1} peak outside of uncertainty. The 2851 cm^{-1} peak is representative of a CH_2 asymmetric stretch in lipids. The specific ratio I3015/2851 can be used to determine the ratio of total unsaturated fatty acids to total fatty acids in a sample. It is shown in the graph above that the ratio of total unsaturated fatty acids to total fatty acids demonstrates that there is a decrease in the relative intensity of unsaturated fatty acids in PNT2-C2 dried.

The 3015 cm^{-1} peak PIR panel in Figure 4.25 shows separation occurring in DNA/RNA, lipids and proteins between PNT2-C2 live and dried. Furthermore, by identifying separation with the 2851 cm^{-1} peak, it is possible to discriminate between PNT2-C2 live and dried based on saturated lipid content [64, 140]. This is due to the difference in the TUFA/TFA ratio mentioned prior. Figure 4.25 shows that the ratio of total unsaturated fatty acids to total fatty acids demonstrates that there is a decrease in the relative intensity of unsaturated fatty acids in PNT2-C2

dried outside of uncertainty. It should be noted that this ratio was devised as a molecular marker to distinguish between cancerous tumour cells and therefore any results of the TUFA/TFA ratio must be interpreted carefully [64, 140]. This is because previous studies examining this ratio have not included a normal cell line [64]. It is important to highlight that studies that have used dried standard cell lines for comparing different states of cancer using the TUFA/TFA ratio may have come to false conclusions. This is due to the evidence presented here in this thesis, suggesting that there is a specific effect that air-drying has on the unsaturated =CH stretch in lipids. It could be argued that the 2851 cm^{-1} is also a biomarker, but through analysis of the loadings in Figure 4.20 it can be seen that the variance captured by the 2851 cm^{-1} peak is considerably less than that of the 3015 cm^{-1} peak. Thus between between PNT2-C2 live and dried it is thought that the 3015 cm^{-1} peak biomarker is having a larger influence on this ratio. There has not been significant studies into the autolytic fate of lipids in cells for fixation methods such as air-drying [263]. It has been reported in Infrared spectroscopy that air-drying and formalin fixation both preserve cellular lipid much better than other fixation techniques such ethanol fixation, but will experience changes in lipid profile as a function of increasing time[264]. It is proposed to further investigate the air-drying effect on the autolysis of lipids that Raman measurements occur over a set time period in increasing amounts. For example, measurement of the same sample in a 0-1 hour window before returning to measure it again after in the 4-5 hour window post initial measurement.

Much debate has centered around the effects of salt crystals deposited during the drying process [235]. Morphologically, alterations in cells occur when salt crystals are present during the drying process. This can be seen in Figure 4.26 where there is a clear presence of salt crystals left by the HBSS solution used to rinse the substrate prior to Raman analysis.

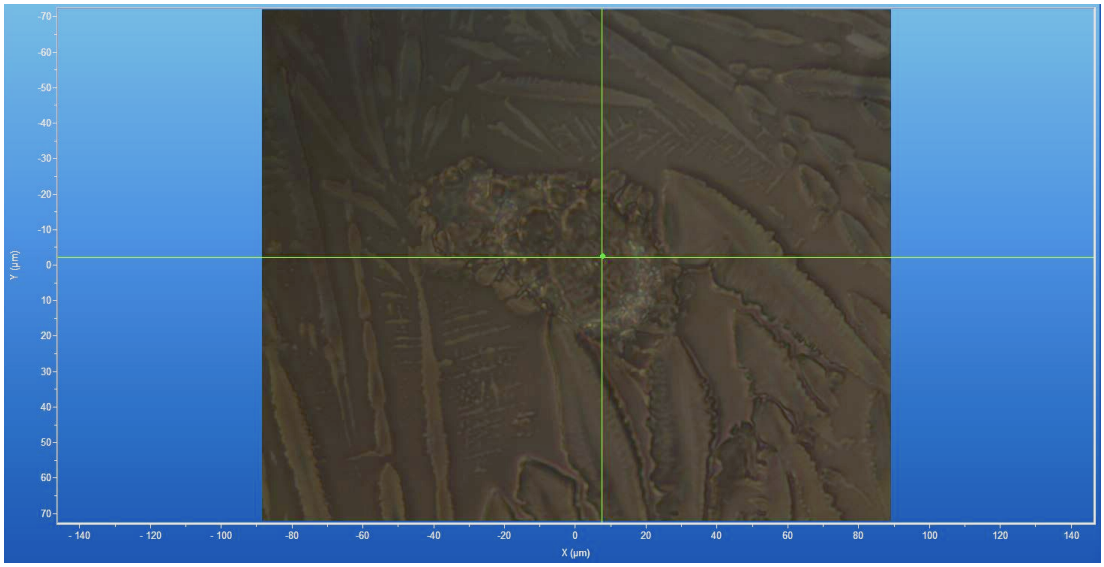


Figure 4.26: Image taken of dried PNT2-C2 cell at x100 magnification showing that salt crystals have been produced through drying as a result of washing with HBSS solution on the CaF_2 substrate. It can be seen that salt crystals do not form in a uniform pattern or size on the substrate. There are a number of crystals/drying artefacts that make contact with the cell. Moreover there is a risk that smaller particles will be present atop the cell. Subsequently, it is presumed that these salt crystals will alter the intensity of the Raman spectrum and cause a deviation from the normal. However, careful cell selection protocols as well as focusing protocols are thought to mitigate this risk. Furthermore, removal of outliers by observation and tools such as PCA as detailed in Methods Section 3.5.6 will remove the risk of salt crystals interfering with associated Raman spectrum of the sample.

There is limited evidence or data in previous studies to know what effect the presence of salt crystals will have on the dehydrated state of PNT2-C2 dried and the generated Raman signal. While it is thought that the excess salt crystals deposited on cells during the drying process has the potential to interact with the laser beam and interfere with results, much effort has been taken to mitigate this risk. Furthermore, removal of outliers using different analysis tools such as PCA as well as observation as detailed in Methods Section 3.5.6 will remove the risk of salt crystals interfering with associated Raman spectrum of the sample. It is for these reasons that the mean profile subtraction, PCA analysis and peak intensity ratio analysis are thought to be truly representative of the biomolecular differences between PNT2-C2 in a live and dried state.

4.6 Conclusion

This study considered the live and dried version of a normal prostate standard cell line PNT2-C2. It was found that the live and dried versions of PNT2-C2 have fundamental biomolecular differences.

The results from the convergence tests of the spectral average, twice the standard deviation and the standard error for both the fingerprint and high wavenumber region of PNT2-C2 live and dried, showed both data sets were fully converged. The variation between cell lines in both the fingerprint and high wavenumber regions demonstrates that even under the same experimental setup there needs to be convergence analysis for each region and each population rather than an arbitrary number of spectra needing to be taken.

A mean profile subtraction highlighted several differences between PNT2-C2 live and PNT2-C2 dried. Key points of difference in the fingerprint region were identified at 745 cm^{-1} (thymine ring breathing mode), 1001 cm^{-1} (phenylalanine), 1338 cm^{-1} (proteins and DNA) and 1656 cm^{-1} (Amide I C=O stretch, α-helix conformation and C=C stretch of lipids). In the High wavenumber profile subtraction the key peaks were identified as generating the spectral difference between cell populations, 2850 cm^{-1} (CH_2 symmetric stretch in lipids) and 2940 cm^{-1} (C-H vibrations in proteins and lipids), as well as CH_2 modes in lipids and 3015 cm^{-1} representative of Unsaturated =CH stretch in lipids. All peaks identified at this stage were used for informed analysis at the PCA and PIR stage.

Multivariate PCA analysis has been able to discriminate between PNT2-C2 live and PNT2-C2 dried. PCA analysis was able to achieve 100% separation in both the fingerprint region 600-1800 cm^{-1} and high wavenumber region 2700-3100 cm^{-1}. There were a number of peaks associated with variance leading to separation in the fingerprint region: 668 cm^{-1} (DNA/RNA and proteins), 698 cm^{-1} (proteins), 715 cm^{-1} (proteins and lipids), 745 cm^{-1} (DNA), 781 cm^{-1} (DNA/RNA), 875 cm^{-1} (proteins and lipids), 1001 cm^{-1} (proteins), 1080 cm^{-1} (DNA. lipids and proteins), 1234 cm^{-1} (DNA/RNA and proteins), 1334 cm^{-1} (DNA/RNA and

proteins), 1448 cm^{-1} (lipids, proteins and DNA/RNA), 1485 cm^{-1} (protein and DNA/RNA), 1656 cm^{-1} (proteins and lipids) and 1681 cm^{-1} (proteins). Similarly there were a number of peaks associated with variance leading to separation in the high wavenumber region: 2851 cm^{-1} (lipids), 2871 cm^{-1} (lipids and proteins), 2895 cm^{-1} (lipids and proteins), 2930 cm^{-1} (proteins and lipids), 2940 cm^{-1} (lipids and proteins), 2964 cm^{-1} (lipids), 3015 cm^{-1} (lipids) and 3065 cm^{-1} (lipids) although labelled, are not well defined. The peaks associated with contributing to separation at the PCA stage were used to inform the PIR analysis to detail relative biomolecular differences between PNT2-C2 in its live and dried state.

Peak intensity ratio analyses, used to determine the amount of separation between all identified Raman peaks in PNT2-C2 live and dried, determined peaks associated with capturing more than 90% separation. These peaks are the 724 cm^{-1} (DNA and proteins), 745 cm^{-1} (DNA), 1256 cm^{-1} (proteins), 1681 cm^{-1} (proteins) and 2940 cm^{-1} (lipids and proteins), 3014 cm^{-1} (lipids). At the PIR stage it was noted that the 698 cm^{-1} peak (protein), identified in loading 2, could not be fitted in PNT2-C2 dried. This resulted in no corresponding PIR profile. This illustrates that there is potential for information loss when using a cell line in its dried state. The first peak examined at PIR stage, the 724 cm^{-1} band, assigned to DNA and adenine, was not considered a biomarker due to several reasons. Mainly, there was no obvious emergence of a 724 cm^{-1} in the mean profile subtraction or at the PCA analysis stage. Furthermore, strong dependence on the 715 peak cm^{-1} peak made it unsuitable as a robust biomarker. The 745 cm^{-1} peak was identified as being a robust biomarker between PNT2-C2 live and dried. The 745 cm^{-1} peak is generated by a pure thymine ring breathing mode in DNA. The 745 cm^{-1} peak, identified in the mean profile subtraction, PCA stage and at the PIR stage, showed that there was a lower relative intensity of thymine in the PNT2-C2 dried sample. This was attributed to a change in DNA conformation state. The 1256 cm^{-1} peak (proteins) and 1234 cm^{-1} peaks (proteins and DNA) were consider as being potential biomarkers. However, conflicting results observed at the PIR stage ruled them out. The 1336 cm^{-1} (DNA, lipids and protein) peak was identified as being a source of variance at all stages of analysis. However, the 1336 cm^{-1}

band contained such a large amount of biomolecular contributions that make it unsuitable as a biomarker. A further band, 1681 cm^{-1} (amide I), was identified at the PCA stage and PIR stage as being a contributing factor in separation between samples. The 1681 cm^{-1} amide I peak showed at the PIR stage that there was a relatively higher intensity of amide I contribution in PNT2-C2 dried. This was characterised by a relative decrease in intensities of all peaks in both the fingerprint and high wavenumber region. For this reason the 1681 cm^{-1} was attributed to being a robust biomarker. The final biomarker identified was the 3015 cm^{-1} peak (unsaturated fatty acids), which was able to show, using the TUFA/TFA peak intensity ratio, that there is a decrease in the relative intensity of unsaturated fatty acids in PNT2-C2 dried.

In summary, Raman spectroscopy has been used to analyse the normal cell line PNT2-C2 in a live and dried state. Multivariate PCA analysis was able to discriminate between the live and dried cell populations with 100% accuracy. It was also able to detail the fundamental biomolecular differences between PNT2-C2 in its live and dried state through the identification of key biomarkers at the PIR stage. This information provides significant insight into the biomolecular changes that occur during the air-drying cell preparation technique.

CHAPTER 5

RAMAN SPECTROSCOPY OF LIVE, PROSTATE STANDARD CELL LINES

5.1 Overview of chapter

This chapter presents an investigation of Raman analysis for a complete set of standard cell lines. It covers a wide spectrum of prostate disease. The spectrum of disease covered in this study investigates everything from normal prostate cells and benign prostate cells to metastatic and local cancerous cells. It is hypothesised that this would make it possible to mimic conditions that would occur in prostate disease of various patients. A full characterisation of each cell line has been carried out using Gaussian peak fitting to identify peak positions. These peaks have been assigned biochemical assignments through detailed analysis of current literature. Principal component analysis (PCA) and linear discriminant analysis (LDA) have been used to separate samples, with corresponding loadings being used to highlight biochemical differences that capture a high degree of variance for further investigation. This allows not only for the stratification of prostate standard cell lines but allows in-depth further analysis of biochemical assignments to be made to each specific cell line. This was further investigated using peak intensity ratio (PIR) analysis to determine what specific relative changes in

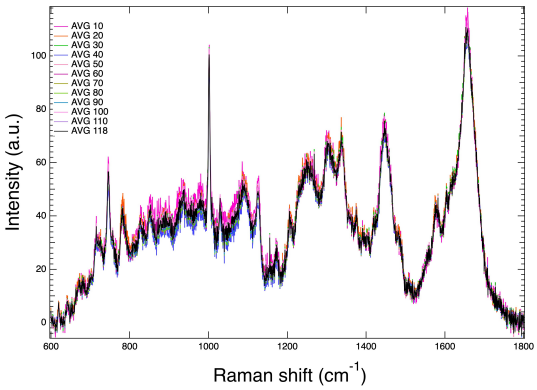
peak intensities were occurring between the normal PNT2-C2 sample and the respective cancerous sample. Additionally, a cross-comparison is also carried out between the standard cell lines.

5.1.1 Statistical convergence testing of standard prostate cell lines

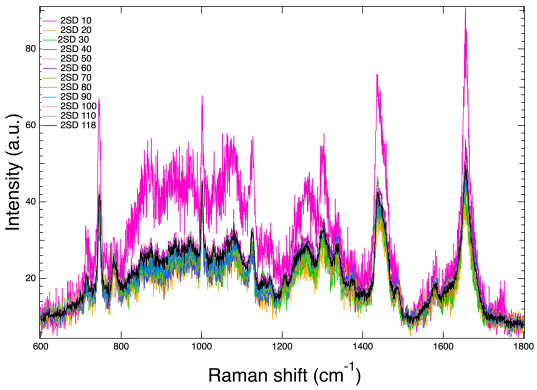
A total of three spectra were removed from PNT2-C2 live, seven from BPH-1, seven from LNCaP, four from DU145, eight from PC3 and twelve from P4E6. These spectra were removed due to falling into the category of being relatively low signal-to-noise or being representative of cytoplasm as per the criteria and justification outlined in the Methods Section 3.5.6. Although it is thought that there is enough spectra in all data sets that outliers would not have an effect on population-level statistics, they are removed for a sense of completeness.

In this study the number of spectra remaining after removal of outliers are as follows: PNT2-c2 (151), BPH-1 (118), LNCaP (123), DU145 (130) and PC3 (122). BPH-1 contains the lowest amount of spectra contributing to its profile and PCA results.

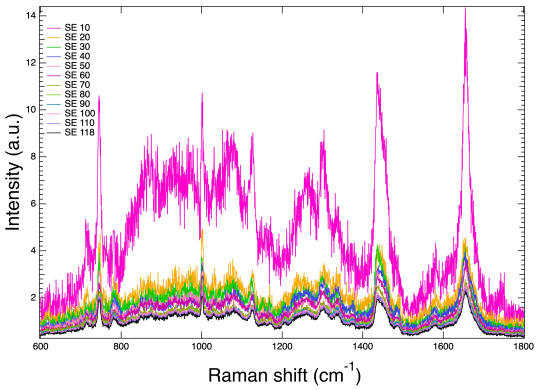
It is possible, now that outliers have been removed, to analyse the convergence of each standard cell line individually for the fingerprint (600-1800 cm^{-1}) and high wavenumber (2700-3100 cm^{-1}) region. Quantitative analysis as detailed in Methods Section 3.6.1 is used to measure the convergence of each prostate standard cell line, namely; the spectral average, twice the standard deviation and the standard error. These graphs associated with these convergence tests are given in Figures 5.1, 5.2, 5.3, 5.4 and 5.5. As PNT2-C2 live was used in Chapter 4, the graphs associated with its convergence test are given in Chapter 4 Figure 4.2.



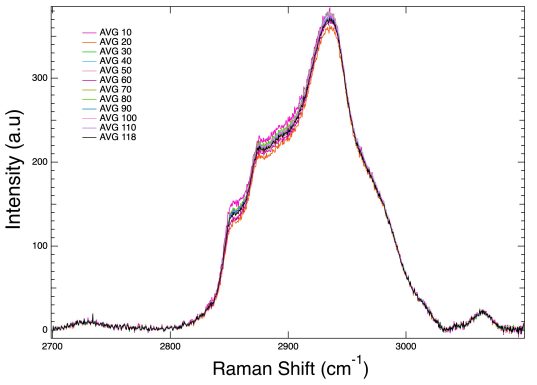
(a) BPH-1 LIVE FP AVG



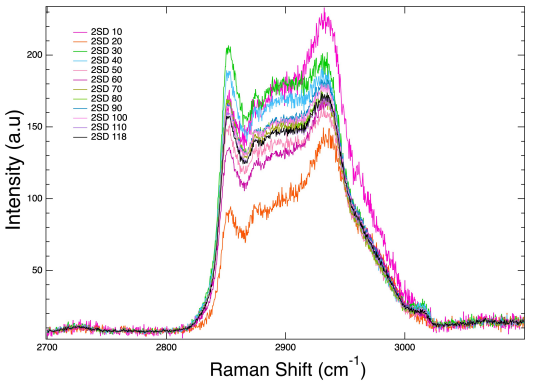
(b) BPH-1 LIVE FP 2SD



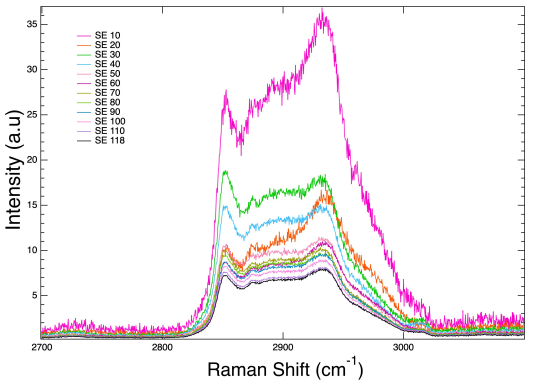
(c) BPH-1 LIVE FP SE



(d) BPH-1 LIVE HW AVG

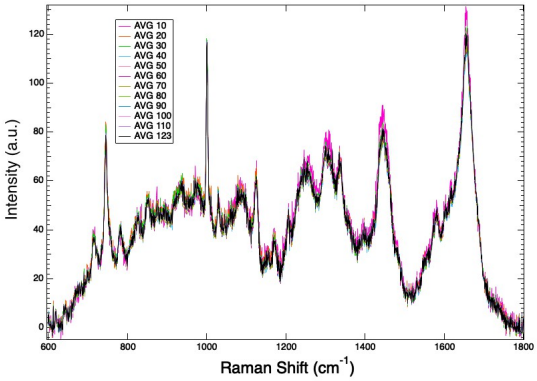


(e) BPH-1 LIVE HW 2SD

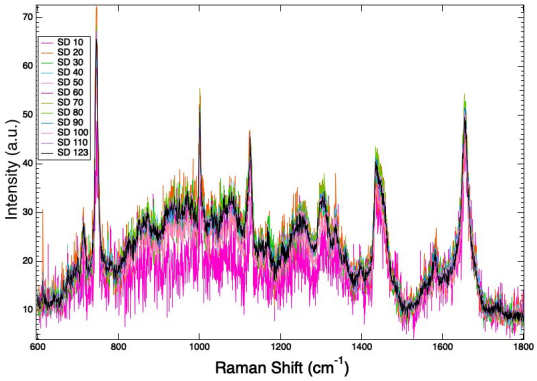


(f) BPH-1 LIVE HW SE

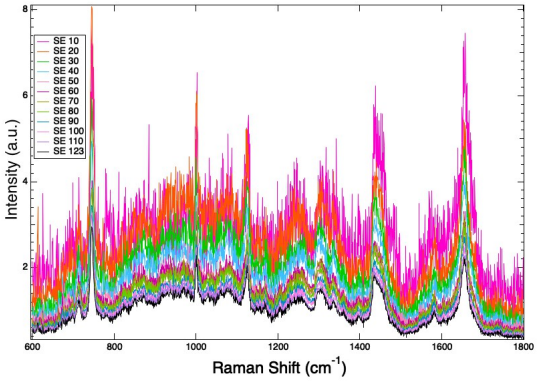
Figure 5.1: Convergence of the average (AVG), twice the standard deviation (2SD) and standard error of the mean (SE) as a function of the increasing number of spectra per spectral average for the BPH1 live fingerprint (FP) and high wavenumber (HW) region. [Figs. (a) & (b), (c) & (d), (e) & (f), respectively].



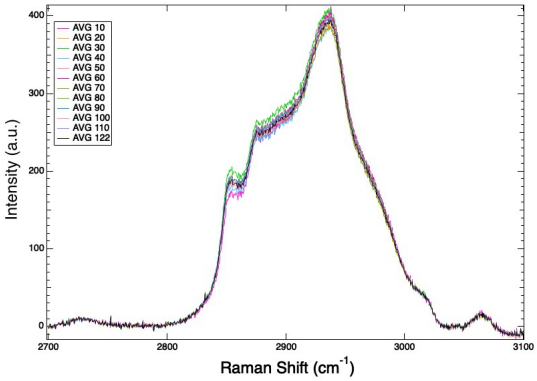
(a) LNCaP FP AVG



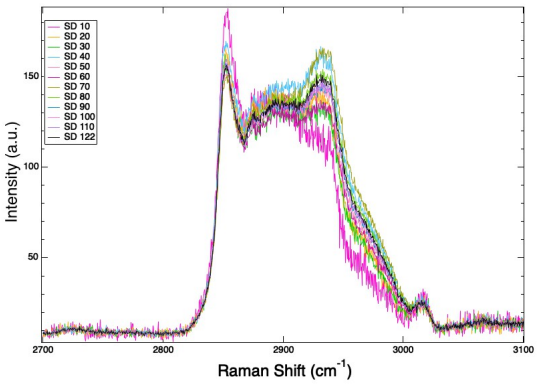
(b) LNCaP FP 2SD



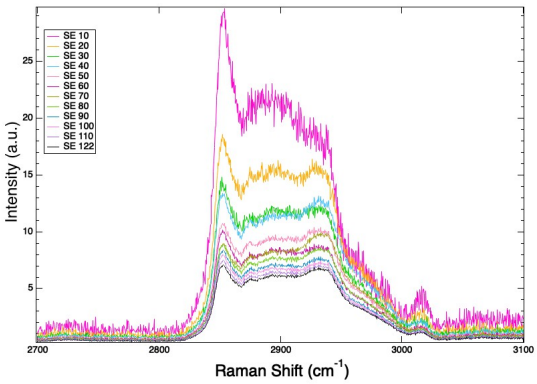
(c) LNCaP FP SE



(d) LNCaP HW AVG

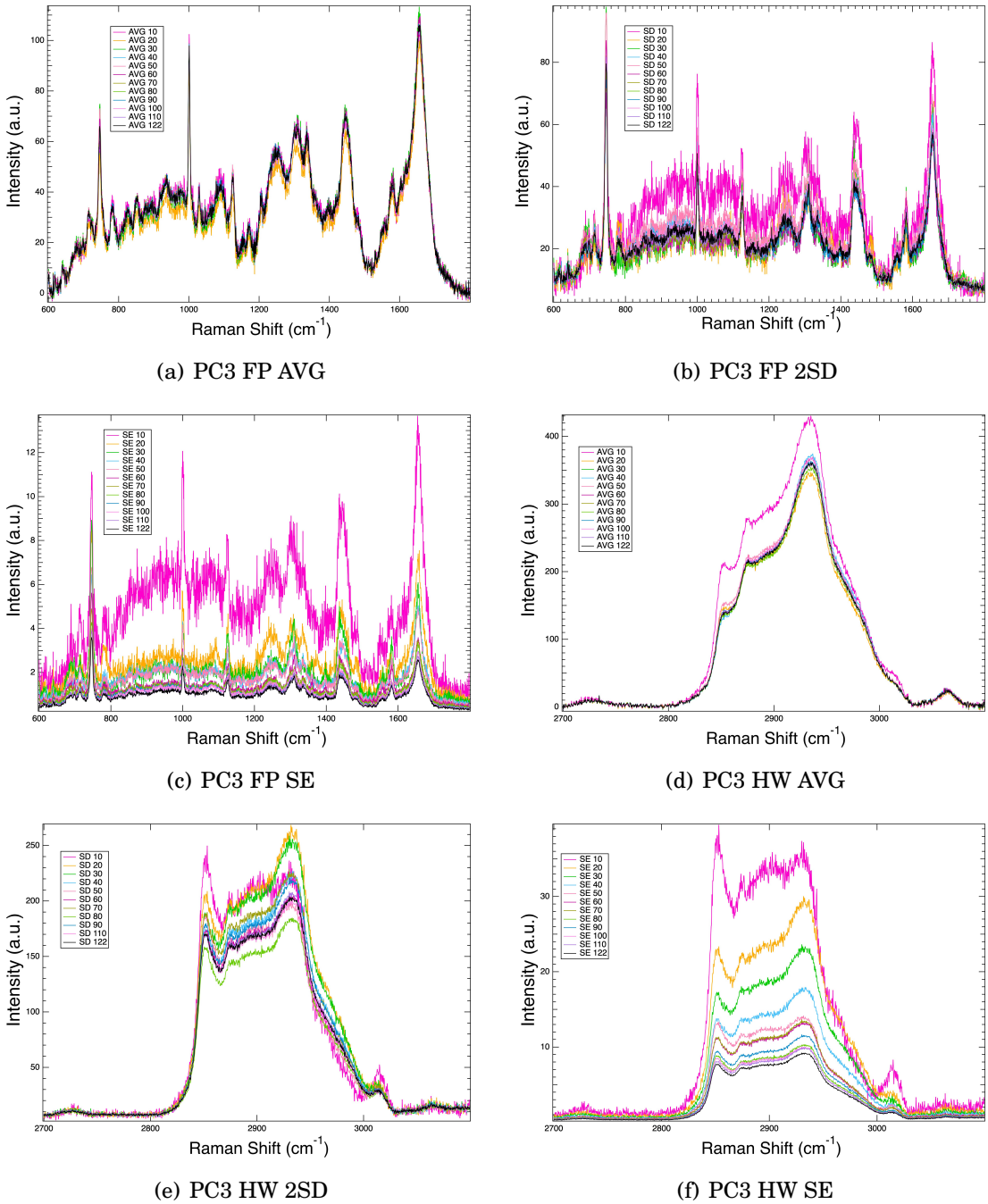


(e) LNCaP HW 2SD



(f) LNCaP HW SE

Figure 5.2: Convergence of the average (AVG), twice the standard deviation (2SD) and standard error of the mean (SE) as a function of the increasing number of spectra per spectral average for the LNCaP live fingerprint (FP) and high wavenumber (HW). [Figs. (a) & (b), (c) & (d) and (e) & (f), respectively].



(a) PC3 FP AVG

(b) PC3 FP 2SD

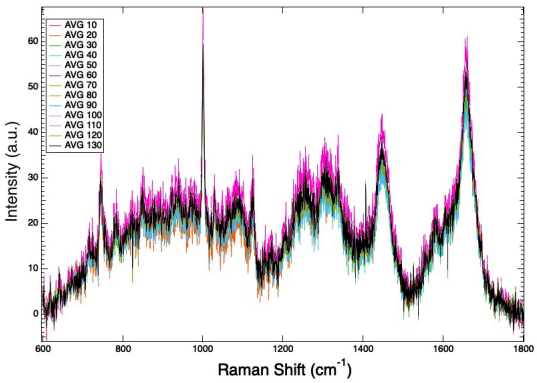
(c) PC3 FP SE

(d) PC3 HW AVG

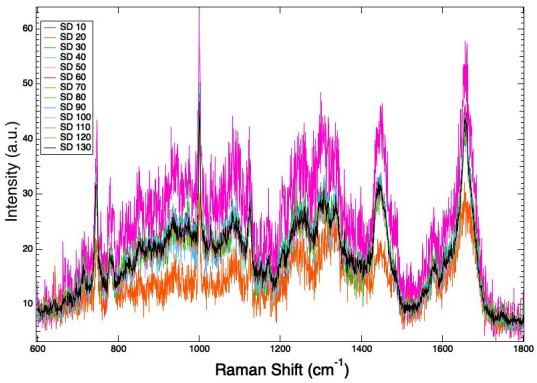
(e) PC3 HW 2SD

(f) PC3 HW SE

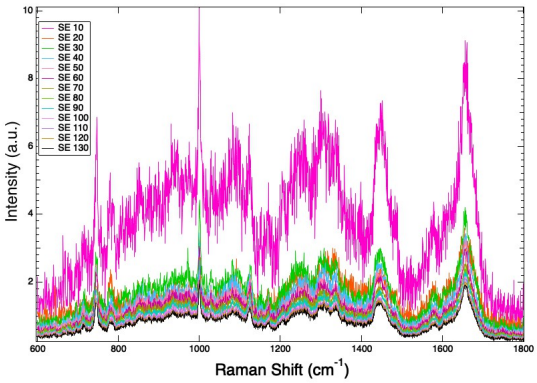
Figure 5.3: Convergence of the average (AVG), twice the standard deviation (2SD) and standard error of the mean (SE) as a function of the increasing number of spectra per spectral average for the PC3 live fingerprint (FP) and high wavenumber (HW). [Figs. (a) & (b), (c) & (d) and (e) & (f), respectively].



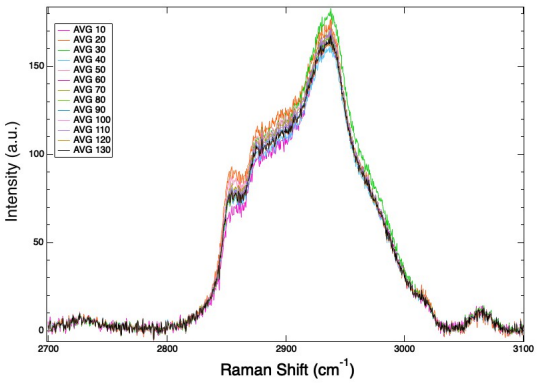
(a) DU145 FP AVG



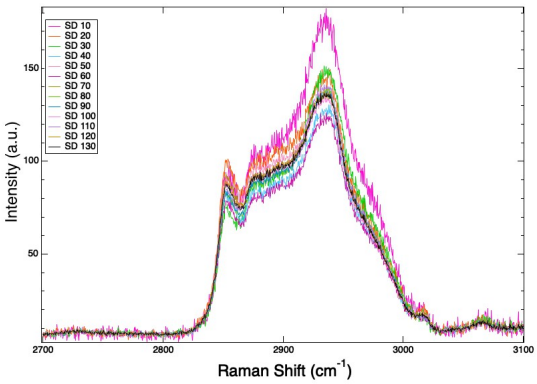
(b) DU145 FP 2SD



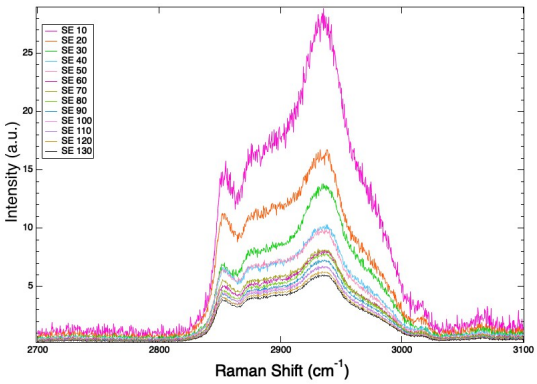
(c) DU145 FP 2SE



(d) DU145 HW AVG

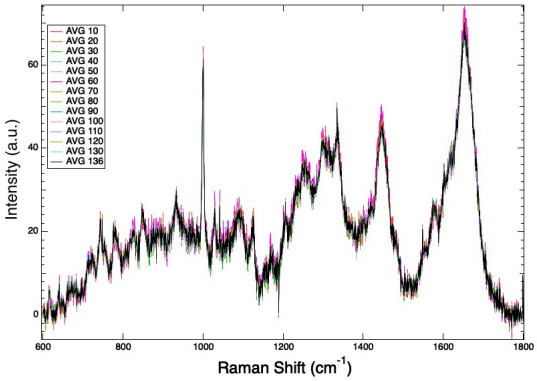


(e) DU145 HW 2SD

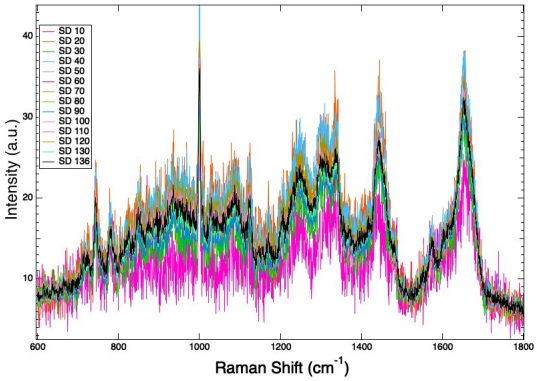


(f) DU145 HW SE

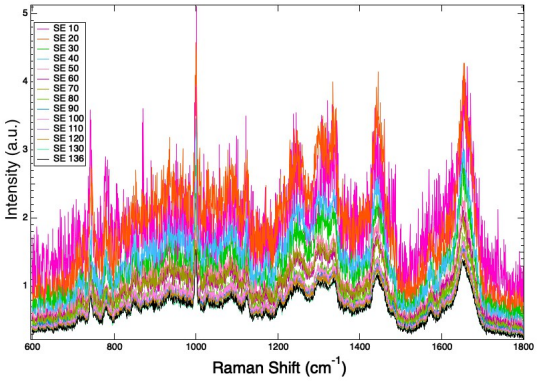
Figure 5.4: Convergence of the average (AVG), twice the standard deviation (2SD) and standard error of the mean (SE) as a function of the increasing number of spectra per spectral average for the DU145 live fingerprint (FP) and high wavenumber (HW). [Figs. (a) & (b), (c) & (d) and (e) & (f), respectively].



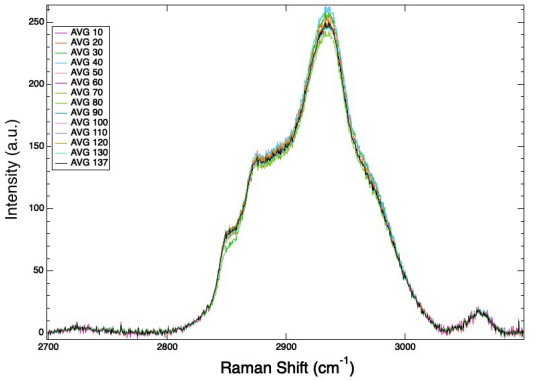
(a) P4E6 FP AVG



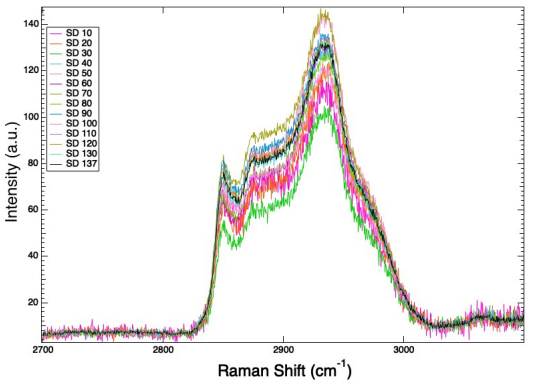
(b) P4E6 FP 2SD



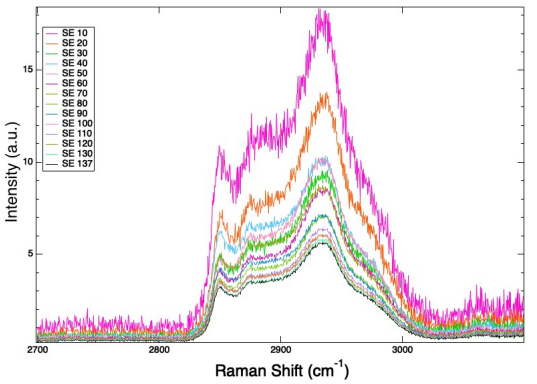
(c) P4E6 FP SE



(d) P4E6 HW AVG



(e) P4E6 HW 2SD



(f) P4E6 HW SE

Figure 5.5: Convergence of the average (AVG), twice the standard deviation (2SD) and standard error of the mean (SE) as a function of the increasing number of spectra per spectral average for the P4E6 live fingerprint (FP) and high wavenumber (HW). [Figs. (a) & (b), (c) & (d) and (e) & (f), respectively].

The convergence of the spectral average, twice the standard deviation and the standard error for each cell line for the fingerprint and high wavenumber region is detailed in Table 5.1. The convergence in the fingerprint region was assessed using the 1001 cm^{-1} peak. The convergence in the high wavenumber region was assessed using the 2934 cm^{-1} peak. The criteria for convergence is detailed as per the Methods Section 3.6.1.

Table 5.1: The convergence data for live prostate standard cell line is displayed. Key for table: PNT2-C2 2SD FP 80 represents that twice the standard deviation in PNT2-C2 converged at approximately 80 spectra in the fingerprint profile.

Cell line	**AVG FP**	**2SD FP**	**SE FP**	**AVG HW**	**2SD HW**	**SE HW**
PNT2-C2	60	80	70	60	60	70
BPH-1	60	90	70	50	60	50
LNCaP	50	80	50	40	50	60
DU145	80	60	70	50	70	60
PC3	40	60	60	50	70	60
P4E6	60	60	80	50	80	80

Table 5.1 shows that the respective spectral profiles of the cell lines being used for analysis in this study have converged. Another observation in Table 5.1 is the wide variation between cell lines in both the fingerprint and high wavenumber regions. This would suggest that even under the same experimental setup, there needs to be convergence analysis for each region and each population rather than an arbitrary number being taken.

It is hard when drawing comparison with other studies, as the convergence for data is frequently not mentioned in multiple studies [63, 64, 141, 224]. For example, the study by Potcoava *et al.* (2014) that analysed both a version of LNCaP and PC3, stated that for "each cell sample 28 to 32 individual cells were measured with 10 to 20 Raman spectra acquired per cell" [64]. This statement gives a range from 280 spectra to 640 which is large range in comparison to the size of the data set. It is acknowledged that there is no definitive number of spectra required for convergence due to varying experimental parameters. However, in the case of Potcoava *et al.* (2014) no attempt is made to find the number of spectra that lead to statistical convergence and population representation [64]. In contrast, others, such as Corsetti *et al.* (2018) state explicitly how many spectra have been measured. In their comparison between PNT2 and DU145 there were 69 spectra

taken from 69 PNT2 cells and 74 spectra taken from 74 individual cells [224]. However, despite stating exact amounts of spectra included, there is failure to report at which point convergence of statistical quantities is observed. There are also studies that do not examine the spectrum of the cell lines being measured, electing only to examine the cells through PCA/LDA analysis, using principal components and the corresponding loadings. The analysis by Crow *et al.* (2005) falls into this class. Crow *et al.* (2005) utilised LNCaP, DU145 and PC3 in their study; 50 spectra were taken from LNCaP and Du145 while 51 spectra were taken from PC3 [63]. The primary objective in the study by Crow *et al.* (2005) is to create a diagnostic algorithm. However, no convergence data is presented as part of this study. They do state that the diagnostic has sensitivity of 98% and a specificity of 99%, however, without the addition of convergence data it is hard to measure at which point this algorithm becomes effective. It should also be highlighted that there was no measure of associated errors in this study. Clearly, as biomedical applications move toward using AI and seek to build rapid diagnostic models or probing tools, there needs to be more inclusion of convergence analysis. This is because the data being provided for machine learning must be truly representative of the sample [244].

5.1.2 Raman spectroscopy of standard cell lines

The average profile of each standard cell line has been calculated. The profile for each cell line consists of the fingerprint region (600-1800 cm^{-1}) and high wavenumber region (2700-3100 cm^{-1}). These spectra have been pre-processed as per Section 3.6.2 of the Methods chapter. To allow for display of all standard cell lines on the same graph, they have been normalised using the total spectrum area and offset by an arbitrary amount to make visualisation between samples clearer. The graphs for both the fingerprint region and high wavenumber region are displayed below in Figures 5.6 and 5.7.

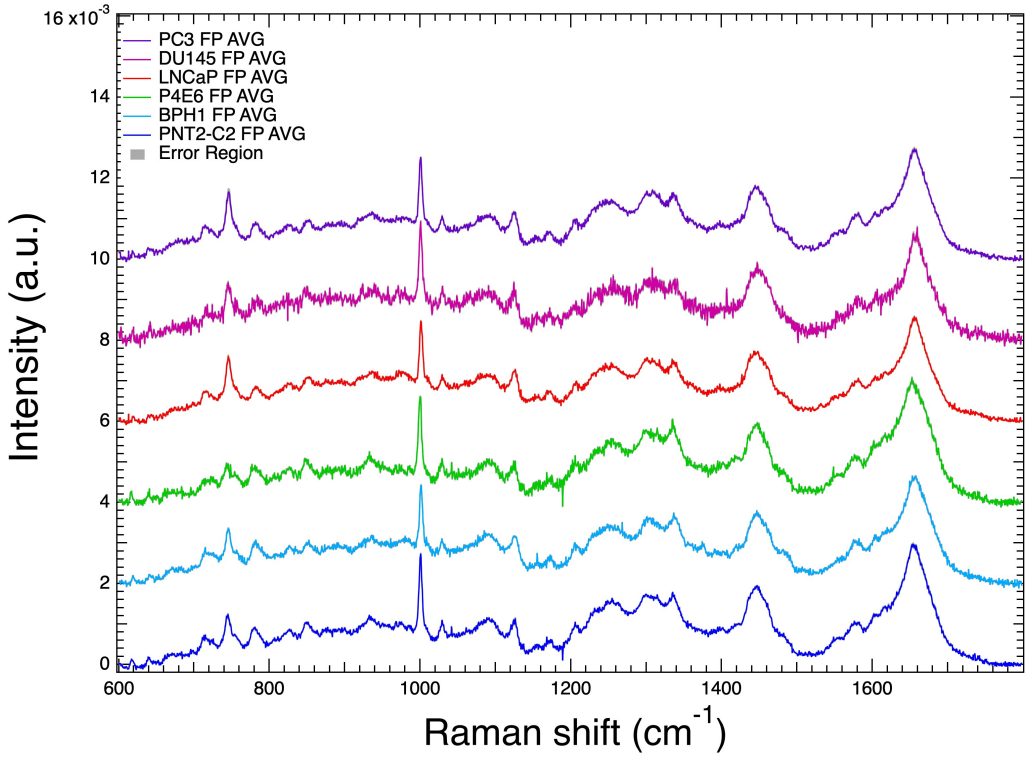


Figure 5.6: The fingerprint region of the live standard cell lines PNT2-C2, BPH-1, LNCaP, PC3, DU145 and P4E6 are displayed on the same graph. These spectra have been cut to the region 600-1800 cm^{-1}, undergone a linear baseline subtraction before being normalised to the total spectrum area. Additionally, each cell line has been offset by an arbitrary amount to make make it easier to view each cell line individually. The mean standard error envelope of each spectrum is displayed in grey around each profile. Observationally, at this stage, there are no significant spectral differences between the cell lines. There are differences in the spectrum that can be examined through other analysis techniques at a later stage. It is noted at this stage that there is significantly more noise observed in the profiles of P4E6 and DU145 than the other samples. DU145 has the lowest signal-to-noise ratio out of any cell line.

Figure 5.6 shows the fingerprint region of the live standard cell lines PNT2-C2, BPH-1, LNCaP, PC3, DU145 and P4E6 displayed on the same graph. The standard error region of each spectra is displayed in grey around each profile. It can be observed in 5.6 that the profiles of both P4E6 and DU145 have significantly more noise present than in the other samples. The exact cause of this is not clear. However, this should not be an issue for further analysis as both DU145 and P4E6 are converged spectral averages despite this noise. Whilst it is very difficult to interpret this increase in noise it is also noted that the presence of noise covers

the whole spectrum and is not limited to just a few bands. This would suggest that the noise is originating from the experiment itself. A drop in intensity and increase in noise could be the result of a decrease in laser power. However, to counteract differences in intensity the spectra are normalised with respect to the total spectrum area. The noise present is not a critically limiting factor in terms of analysis. It is not the intensity of the peaks that is being measured but the relative intensities. While the drop in intensity does indicate that there may be some bands lost to low-signal-to noise there are still a large number of bands that can be considered during analysis.

The average profiles for the high wavenumber region are shown in Figure 5.7.

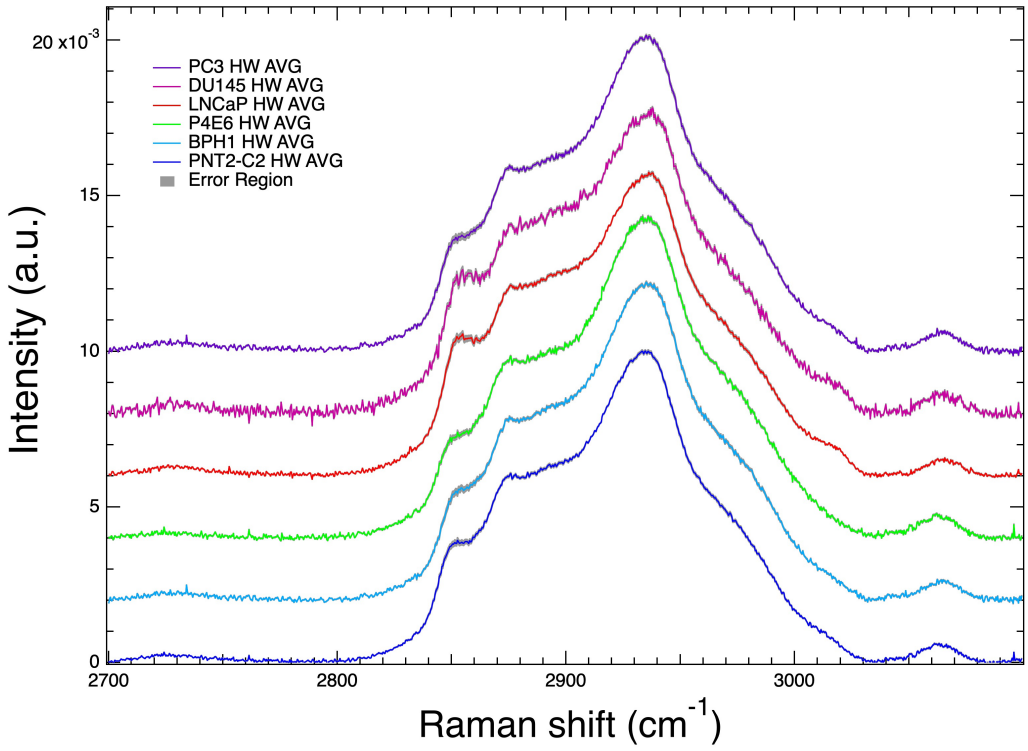


Figure 5.7: The high wavenumber region of the live standard cell lines PNT2-C2, BPH-1, LNCaP, PC3, DU145 and P4E6 are displayed on the same graph. These spectra have been cut to the region 2700-3100 cm^{-1}, undergone a polynomial baseline subtraction and normalisation with respect to the total spectrum area. The standard error mean envelope is displayed in grey around each spectrum. The spectra have been offset to allow for easier visual comparison. There is a noted increase in noise in the cell line DU145. However, the mean standard error envelope is similar for all standard cell lines.

Figure 5.7 shows observationally that the profiles of the standard cell lines are similar. It can also be seen in Figure 5.7 that there is less noise observed in the high wavenumber region of P4E6 than in the fingerprint region in Figure 5.6. In contrast to P4E6, there is still a relatively large amount of noise present in the cell line DU145. This, as previously mentioned, is likely to occur due to experimental setup. It should be reinforced, however, that the noise present in DU145 in the high wavenumber is not a limiting factor in terms of analysis as it is only relative intensities that are being studied between peaks.

The peak positions for each cell line have been located using the using Gaussian peak fitting. A detailed account of the fitting method is given in chapter 3 Section 3.6.3. The fitting of profiles has been carried out for the fingerprint region (600-1800 cm^{-1}) and high wavenumber region (2700-3100 cm^{-1}). The peak average is calculated using the range of the peaks found between all samples. The exact peak positions of all peaks can be viewed in Table B.2 in the Appendix. The general biomolecular assignments as well as specific assignments have been attributed to respective positions. This information is condensed in table 5.2.

Table 5.2: Biomolecular peak-assignments for standard cell lines. The average peak position is shown (calculated in Table B.2 in the appendix), general and detailed assignments) obtained from the literature. For the peak position a standard error of ± 3 cm^{-1} as an uncertainty is attached to all positions.

Peak Position (cm^{-1})	**General assignment**	**Detailed biomolecule assignments**
619	Proteins	C-C twist of proteins ([128], [129]) Phenylalanine C-C twist ([128], [129], [224], [245])
641	Proteins	C-S stretch & C-C twist of tyrosine ([128], [129] C-C twist of tyrosine ([224], [245], [216], [142])
673	DNA/RNA & Proteins	Guanine and thymine ring breathing mode ([128]) Tyrosine-G backbone in RNA ([128]) Guanine ring breathing mode and C-S stretching mode of cystine ([245]) Guanine, tyrosine and cysteine ([142])
698	Proteins	Aminoacid methionine ν(C-S) *trans* ([128])
715	Proteins & Lipids	Aminoacid methionine ν(C-S) *trans* ([128]) $CN^+(CH_3)_3$ stretch in lipids ([128]) C-N membrane phospholipid head stretch ([128], [246], [245]) Adenine nucleotide peak ([245], [128])
724	DNA	Adenine ([245], [142]) DNA ([105], [128], [224])
745	DNA	Thymine ring breathing mode ([248], [105], [128], [224])

Continued on next page

TABLE 6.2 (continued): Biomolecule assignments

Peak Position (cm^{-1})	**General assignment**	**Detailed biomolecule assignments**
756	Proteins	Symmetric breathing of tryptophan ([245]) Tryptophan ring breathing mode ([128], [246]) Tryptophan ([249], [245], [129])
781	DNA/RNA	Uracil, Cytosine and thymine ring breathing modes ([249] [249], [142] [142], [128], [224]) Nucleic acids ([63]) Uracil, cytosine, and thymine overlapping with O-P-O stretch in DNA ([105])
808	DNA/RNA	O-P-O stretch DNA ([249] [249]) and O-P-O stretch RNA ([128], [246])
825	Proteins & DNA/RNA	Tyrosine out-of-plane ring breathing and O-P-O stretch in DNA ([249], [245], [142], [105] [105], [128]) DNA O-P-O asymmetric stretch, tyrosine ([248]) Proline and hydroxyproline, out-of-plane tyrosine ring breathing, DNA/RNA asymmetric O-P-O stretch ([246]) DNA backbone (O-P-O) ([63]) Out-of-plane tyrosine ring breathing ([224])
850	Proteins	Tyrosine ring breathing ([249], [105], [128], [224]) C-C stretch in proline and tyrosine ring breathing mode ([245], [142])

Continued on next page

TABLE 6.2 (continued): Biomolecule assignments

Peak Position (cm^{-1})	**General assignment**	**Detailed biomolecule assignments**
		Proteins (C-C stretch in proline; tyrosine ring breathing); carbohydrates (glycogen, polysaccharides C-O-C stretch) ([246])
876	Proteins & Lipids	Proteins C-C stretch, antisymmetric vibrations of choline $N(CH_3)_3$ (phospholipids), hydroxyproline, tryptophan (proteins) ([128]) Hydroxyproline C-C stretch in collagen, tryptophan ring deformation, choline asymmetric stretch in phospholipids and C-O-C stretch in carbohydrates ([246]) Tryptophan ([224])
892	Carbohydrates & Proteins	Saccharide band, methylene and C-C skeletal, proteins (protein modes for tumours) ([128]) C-C skeletal ([105])
934	Proteins & Carbohydrates	Skeletal C-C stretch α-helix (protein), proline, valine ([245], [128]) C-C backbone stretch α-helix ([249], [63], [142], ([246]) C-O-C glycogen (carbohydrate) ([246])
974	Lipids, Proteins & DNA/RNA	Lipids, phosphate monoester groups of phosphorylated proteins and cellular nucleic acids ([128])
1001	Proteins	Phenylalanine ring breathing ([248], [249], [245], [142], [105], [128], [246], [224])
1029	Proteins (namely),	Phenylalanine C-H in-plane bending mode ([249], [245], [142], [128], [224])

Continued on next page

TABLE 6.2 (continued): Biomolecule assignments

Peak Position (cm^{-1})	**General assignment**	**Detailed biomolecule assignments**
	Lipids & Carbohydrates	C-N stretch of proteins ([105], [128]) Collagen, keratin, C-N stretch in proteins, Phe C-H in plane bending, phospholipids and polysaccharides ([246])
1063	Lipids & DNA/RNA	Lipids chain C-C stretch ([249], [245], [128]) DNA/RNA PO_2^- backbone stretch ([128])
1089	DNA, Lipids & Proteins	DNA PO_2^- backbone stretch and C-C stretch (gauche) in lipids ([248] [248], [245], [105], [128]) C-N of proteins ([105], [128]) Lipids ([63])
1125	Lipids, Proteins & Carbohydrates	Protein C-N stretch ([249], [142], [128]) C-C stretch in lipids ([249], [245], [142] [128]) C-C stretch in proteins ([245] , [128]) C-N stretch (glucose), C-O & C-C stretch in carbohydrates ([245], [128]) phospholipids ([63])
1155	Proteins & Carbohydrates	C-C and C-N stretch in proteins ([249], [105], [128], [224]) Carotenoids—correlated to disease ([128]) Glycogen ([128])
1171	Proteins	(CH) phenylalanine, tyrosine ([105], [128]) Bending mode (in-plane) tyrosine ([249], [245])

Continued on next page

TABLE 6.2 (continued): Biomolecule assignments

Peak Position (cm^{-1})	**General assignment**	**Detailed biomolecule assignments**
		Protein-related peak (C-H bend) ([105])
1205	Proteins & DNA/RNA	Tyrosine and phenylalanine ([248], [128]) Phenylalanine and tryptophan ([249], [245], [128]) Phenylalanine ([224]) CH_2 wagging of glycine and proline ([128]) Amide III ([105], [128]) A,T ring breathing modes of DNA/RNA ([105], [128]) Hydroxyproline & tryrosine ([245], [105])
1231	Proteins & DNA/RNA	Amide III and CH_2 wagging of glycine and proline ([128]) PO_2^- antisymmetric stretch ([128]) Amide III ([249],[142], [224])
1254	Proteins	amide III : CH_2 wagging [245, 250]
1307	Proteins, Lipids & Nucleic acids	CH_2 deformation in lipids, adenine and cytosine ([245], [128]) CH_3/CH_2 modes in lipids / collagen ([245], [128]) CH-N stretch in amines ([128]) Nucleic acids ([63]) Lipids; proteins (amide III), adenine, cytosine ([248])
1337	Proteins & DNA/RNA	CH_3CH_2 wagging, twisting &/or bending of collagen, lipids, tryptophan & nucleic acids (adenine and guanine)([128]) Amide III ([128])

Continued on next page

TABLE 6.2 (continued): Biomolecule assignments

Peak Position (cm^{-1})	**General assignment**	**Detailed biomolecule assignments**
		CH_2 deformation in lipids, adenine and cytosine ([245]) Adenine and guanine; protein C-H deformation ([249], [105]) Adenine, phenylalanine and CH deformation ([248]) Adenine ([142])
1398	Proteins	CH_2 deformation ([128, 129]
1448	Lipids, proteins & DNA/RNA	CH_2 deformation in lipids and proteins; at 1450 cm^{-1}, in malignant tissues ([248] [248], [142] [142], [128]) CH_2 bending mode of proteins ([245], [105], [224]) C-H deformation & vibration in lipids and proteins ([249], [128]) CH_2CH_3 deformation ([128]) Nuclear acids ([63]) CH_2 in-plane deformation (ring stretch) in DNA/RNA ([142])
1484	Protein & DNA/RNA	Amide II CN stretch & N-H bend ([128]) Guanine, adenine ring-breathing modes ([245], [105], [142], [128])
1554	Proteins	Tryptophan ([245], [128]) Amide II ([128])
1577	DNA/RNA	Guanine and adenine ring-breathing modes ([249], [245], [105], [142], [128]) Nucleic acids ([245], [63])
1603	Proteins	C=C bending mode of phenylalanine and tyrosine ([249], [245], [105], [128]) C-C ring stretch of phenyl ([128])

Continued on next page

TABLE 6.2 (continued): Biomolecule assignments

Peak Position (cm^{-1})	**General assignment**	**Detailed biomolecule assignments**
		C-C and C-N stretch (proteins) ([224])
1616	Proteins	Tyrosine, tryptophan C=C stretch ([249], [245], [105], [142], [128])
1657	Proteins & lipids	Amide I C=O stretch, α-helix conformation; proteins ([249], [245], [128]) α-helix proteins & unordered proteins ([63]) protein (amide I) [105] C=C stretch of lipids ([249], [245], [248], [128], [64])
1683	Proteins	Amide I ([105], [128])
2851	Lipids	CH_2 symmetric stretch in lipids ([128], [64], [251], [252])
2872	Lipids & proteins	CH_2 asymmetric stretch in lipids and proteins ([128]) CH_2 symmetric stretch in lipids ([128]) CH_3 symmetric stretch in lipids & proteins ([64])
2894	Lipids & proteins	CH stretch of lipids and proteins (predominantly) and CH_3 symmetric stretch ([128]) CH_2 deformation ([251], [252])
2933	Proteins & lipids	CH_2 asymmetric stretch, CH_3 symmetric stretch (predominantly proteins) ([128]) CH_2 asymmetric stretch of proteins and modes in lipids, CH modes in lipids and proteins ([251]) CH_3 symmetric stretch in proteins and lipids ([252])

Continued on next page

TABLE 6.2 (continued): Biomolecule assignments

Peak Position (cm^{-1})	**General assignment**	**Detailed biomolecule assignments**
2941	Lipids & proteins	C-H vibrations in lipids & proteins. CH_2 modes in lipids ([128])
2966	Lipids	CH_3 asymmetric stretch in lipids, cholesterol & cholesterol ester ([128])
3015	Lipids	Unsaturated =CH stretch in lipids ([128], [251], [252]) =CH stretch (unsaturated lipids, cholesterolester) ([64])
3064	Lipids	CH stretch in lipids ([128])

There is a total of 45 peaks that have been characterised across the 6 prostate standard cell lines measured. The standard cell lines PNT2-C2, LNCaP and PC3 were the only cell lines to have identified all 45 peaks in their respective samples. It should be noted that the baseline for the number of peaks represents the most peaks that can be resolved by peak fitting, with confidence.

In BPH-1 the 1396 cm^{-1} band could not be fitted. This band is a very small standalone peak. Its detailed biomolecular assignment is CH_2 deformation. This should not be an issue in capturing the role of CH_2 deformation, due to other peaks in the sample also being attributed to CH_2 deformation.

The cell line DU145 which was observed has having much more noise than the other cell lines by comparing the spectral averages in Figures 5.6 and 5.7. DU145 had significantly more absent bands than all other cell lines. The first band missing is found at 699 cm^{-1}. The 699 cm^{-1} band is attributed to amino acid methionine ν(C-S). This essential amino acid is essential for growth and development of the cell [265]. Unfortunately, it is the only band that is a pure contribution from methionine. Although this will have a knock-on effect in the interpretation of the biochemical changes in the cell, it will not affect comparisons to literature. The study by Corsetti *et al.* (2018) who also examined DU145 did not find any bands contributing to methionine. The 723 cm^{-1} DNA band is also absent

in DU145. The detailed assignment of the band is adenine. It is thought that the reason for the absence of this band is due to its band position. The 723 cm^{-1} band is found in the same region as the more intense 715 cm^{-1} band. It is thought that the combination of the intensity of the 715 cm^{-1} band being convolved with it, and the noise, means that the signal-to-noise makes the band unidentifiable. Two more additional bands 876 cm^{-1} and 893 cm^{-1} are also not present in DU145. Attributed to proteins/lipids and carbohydrates, both bands are not present due to low signal-to-noise in this region of the spectrum. Similarly to BPH-1, the 1396 cm^{-1} band is also absent. This is also due to low signal-to-noise. Lastly, the 1615 cm^{-1} band is missing. This protein band attributed to tyrosine, tryptophan C=C stretch is also a very weak peak in terms of relative intensity. The loss of the 1615 cm^{-1} band is also attributed to low signal-to-noise . In total there are six bands absent in DU145. This will have a knock-on effect when the cell line is used in comparisons. That said, there are still 39 bands to draw comparison against PNT2-C2 and from which to identify potential biomarkers. This will be explored later on through a combination of PCA/LDA analysis and peak intensity ratios.

P4E6 has only one band absent in its profile. This band is located at 976 cm^{-1} and is generated through contributions from lipid, proteins and DNA. However, unlike many other missing bands, this is provisionally thought to not be caused by low signal-to-noise. There appears to be no band present at 976 cm^{-1} that is able to be fitted using the set parameters established in the methodology. At this stage it is tentatively thought that this absence may be a biomarker of P4E6. This will be investigated at a later stage using a combination of PCA/LDA and peak intensity ratios.

5.2 BPH-1

There is not at present a detailed body of work comparing PNT2-C2 to BPH using Raman spectroscopy, as far as the author of this study is aware. A pilot study by Gazi *et al.* (2003) used Fourier transform infrared microspectroscopy to probe the differences between PNT2-C2, BPH-1 and other prostate cancer cell lines [124]. The study by Gazi *et al.* (2003) specifically focused on interpreting the ratio of glycogen, phosphodiester and RNA and their respective ratios as biomarkers of malignancy [124]. There are several criticisms of the Gazi *et al.* (2003) study that must be addressed. The analysis by Gazi *et al.* (2003) managed to separate benign and malignant cells using the ratio of glycogen to phosphodiester, however, this ratio also classified the PNT2-C2 cell line as being malignant [124]. This result, is obviously is not correct as PNT2-C2 is a well established normal cell line. Whilst it could be concluded that a different biomarker could be used to differentiate the PNT2-C2 cell line from the other cell lines, this work was not carried out. The study by Gazi *et al.* (2003) included only one biomarker, increasing the risk of errors such as false positives [124]. The peaks involved in that ratio were also fitted in the same region as other peaks, with overlapping profiles [124]. This increases the likelihood of variation and difficulty ascertaining the influence of each peak over each other. The final criticism is the use of paraffin-embedded cell samples [124]. In contrast this thesis did not not explore the use of paraffin-embedded samples, due to differences in biomolecular profile to what would be observed in a live state [225].

This study will investigate the relationship of glycogen, phosphodiester and RNA in PNT2-C2 and BPH-1 in a live cell state. This thesis will also expand the scope of comparison between PNT2-C2 and BPH-1 by considering nucleic acids, DNA/RNA, proteins and other lipid contributions. Furthermore, this thesis will expand on the analysis through the means of a mean profile subtraction, PCA/LDA analyses, and much more extensive peak intensity ratio analysis that has investigated the full range of spectral information.

Further use of benign hyperplasia cells in studies, although not specifically

BPH-1, have already been used to make comparisons against cancerous tissue in a previous Raman study by Crow *et al.* (2003), noting an increase in the levels of nucleic acids and reduced glycogen content in cancerous tissue compared to BPH [141]. In the study by Crow *et al.* (2003) the BPH tissue and cancerous tissue were not characterised using peak fitting methods [141]. This was due the primary objective of Crow *et al.* (2003) to construct a diagnostic algorithm to differentiate BPH from three different grades of prostatic adenocarcinoma in-vitro [141]. However, Crow *et al.* (2003) did not make any comparison to normal cells. The increased levels of nucleic acid in cancerous cells along with nucleic acids has also been found in comparative studies as being a marker of BPH-1 [266–269]. This thesis will expand on the work by Crow *et al.* (2003) by considering BPH relative to a normal cell line in a PCA setting. At a later stage it will also consider BPH-1 relative to not only prostate cancer metastatic cell lines, but also a representative cell line for local (prostate localised) cancer.

5.2.1 PNT2-C2 vs BPH-1 mean profile subtraction

A mean profile subtraction between PNT2-C2 and the BPH-1 will allow initial observation of differences between these cell lines. This will allow informed analysis of peaks at a later stage as well as visual consideration of changes to profile shape in respective profiles. The normalised Raman signal for PNT2-C2 and BPH-1 with respect to the total spectrum area is displayed for the fingerprint region in Figure 5.8 and the high wavenumber region is shown in Figure 5.9. The fingerprint region has undergone a linear baseline subtraction as well as being offset for an arbitrary amount. The high wavenumber region has undergone a polynomial baseline subtraction as well as being offset by an arbitrary amount for comparison purposes. The corresponding mean subtraction profile (PNT2-C2 minus BPH-1) is shown for the fingerprint region in Figure 5.8 and the high wavenumber region in Figure 5.9.

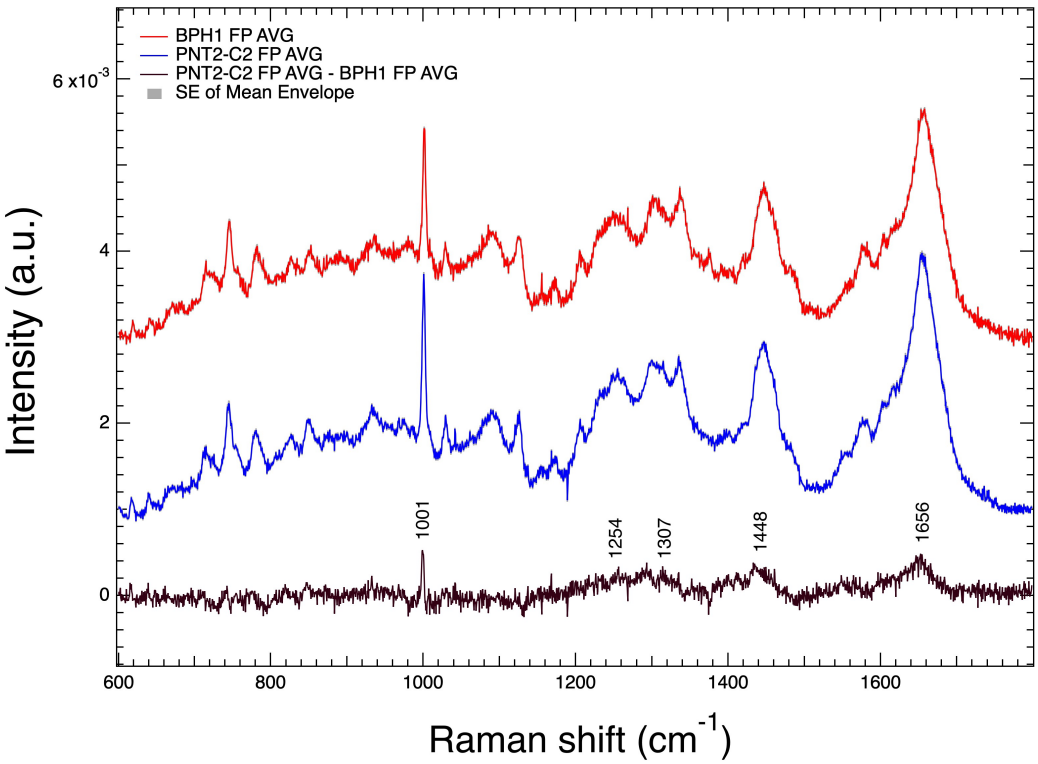


Figure 5.8: The mean profile subtraction of BPH-1 from PNT2-C2 is shown for the fingerprint region (600-1800 cm^{-1}). Pronounced differences are occurring at the peak positions 1001 cm^{-1} (proteins), 1254 cm^{-1} (proteins and carbohydrates), 1307 cm^{-1} (proteins, lipids and nucleic acids), 1448 cm^{-1} (lipids, proteins and DNA/RNA) and 1656 cm^{-1} (proteins and lipids). There is also a noticeable change in profile shape around 1254 cm^{-1} to 1506 cm^{-1} indicating that there may be differences in the secondary structure of proteins. The raised profile of the mean subtraction around 1656 cm^{-1} region also indicates changes in the protein and lipid structure. However, at this stage, both of these conclusions are tentative.

The mean profile subtraction in Figure 5.8 has been annotated to detail observed changes around five peaks. In PNT-C2 there is a relative increase in the peak 1001 cm^{-1} which is attributed to phenylalanine. As far as the author is aware, no previous comments have been made in the literature to suggest why this may be the case. In other prostate studies, phenylalanine has been used as a classification marker, with a noted increase in patients with prostate cancer [270]. It is therefore thought that it would not be unreasonable for a phenylalanine change to be occurring in BPH-1. At this stage of analysis, the 1001 cm^{-1} peak will be highlighted to be investigated further through PCA and peak intensity ratio analysis. The mean profile subtraction showed that variation was observed

around 1254 cm^{-1}, indicating that there may be a difference in the amide III band [128, 129]. Further variation observed around 1448 cm^{-1} indicated that there may be a difference in the CH_2 deformation in lipids and proteins [128, 129]. Furthermore, the change in the profile of the mean subtraction around 1657 cm^{-1} may also indicate a change in secondary protein structure, as well as lipid structure. This initial analysis suggests that there a number of differentiating biomolecular factors between PNT2-C2 and BPH1 [128, 129].

The high wavenumber region (2700-3100 cm^{-1}) of PNT2-C2 and BPH-1 is shown in Figure 5.9. It is thought that this is the first high wavenumber mean profile subtraction of PNT2-C2 and BPH-1 as far as this study is aware.

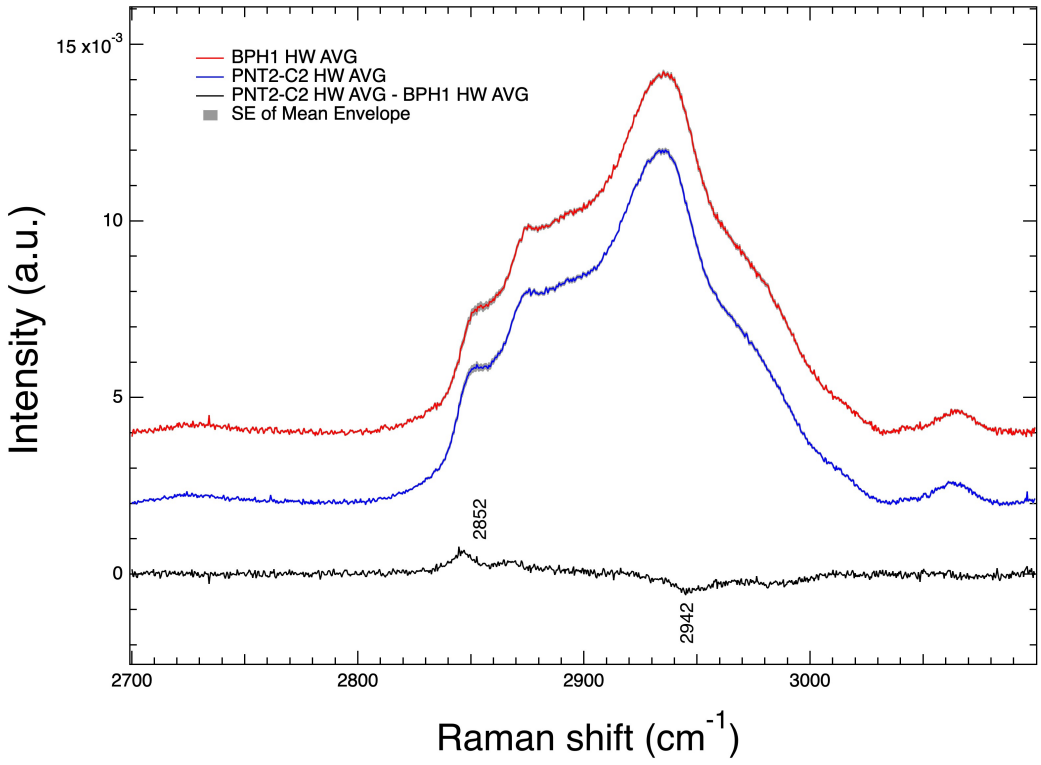


Figure 5.9: For the first time, as far as this study is aware, the differences in the high wavenumber region between PNT2-C2 and BPH-1 are reported. These spectra have undergone a polynomial baseline subtraction and normalised before being subtracted to calculate a mean profile of the subtraction. The peak 2852 cm^{-1} emerges from the mean profile subtraction. This difference around the 2852 cm^{-1} peak would suggest there may be a difference in the fatty acid content between PNT2-C2 and BPH-1. There may also be a change in the C-H vibrations in lipids and proteins based on the emergence of the peak at 2942 cm^{-1}. However, at this stage this analysis is only provisional.

As seen in Figure 5.9, there are biochemical differences occurring in the high wavenumber profile of PNT2-C2 and BPH-1, detailed by the emergence of the two peaks 2852 cm^{-1} and 2942 cm^{-1}. The 2852 cm^{-1} peak has previously been used as a marker to distinguish the relative amount of total fatty acids to unsaturated fatty acid between prostate and breast cancer cell lines[64, 140]. Furthermore, there may also be a change in the C-H vibrations in lipids and proteins based on the emergence of the peak at 2942 cm^{-1}. However, this preliminary analysis is not conclusive. To further examine these findings it is necessary to consider PCA and peak intensity ratio analysis. This is carried out in Section 5.2.3.

5.2.2 PNT2-C2 vs BPH-1 PCA comparison

PCA analysis between the PNT2-C2 Live and BPH-1 of the fingerprint region is able to ascertain that the variance captured by the first six principal components (PCs) are as follows: PC1 = 31.1%, PC2 = 11.9%, PC3 = 8.1%, PC4 = 4.7%, PC5 = 3.1% and PC6 = 2%. The first three PCs capture 51.1% and the first six PCs capture 61% of the total variance of the sample. Cumulatively, 268 principal components are required to capture total variance of the sample.

As far as the author of this thesis is aware, this is the first time PCA has been used to probe the differences between a normal standard cell line and BPH-1 in both the fingerprint and high wavenumber region. In reviewing the literature Crow *et al.* (2003), it was noted that there was an increase in the levels of nucleic acids and reduced glycogen content in cancerous cells compared to benign prostate hyperplasia in a study of tissues [141]. However, the study lacked detailed of biochemical peak assignment and also did not include a normal cell line. This study will expand on the work by Crow *et al.* (2003) by considering BPH-1 relative to a normal cell line [141]. Three rotations of PCA space using the first three principal components are displayed in Figures 5.10, 5.11 and 5.12.

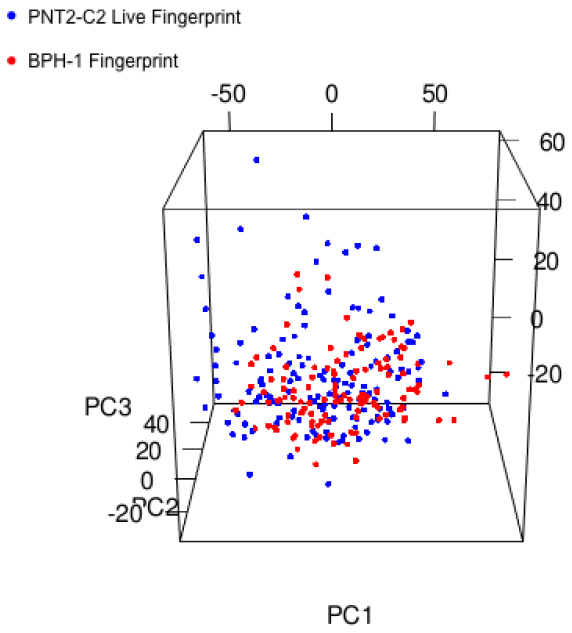


Figure 5.10: 3D PCA space is used to compare the fingerprint region of PNT2-C2 and BPH-1 using principal components 1, 2 and 3. From this angle no separation is observed along PC1 or PC3.

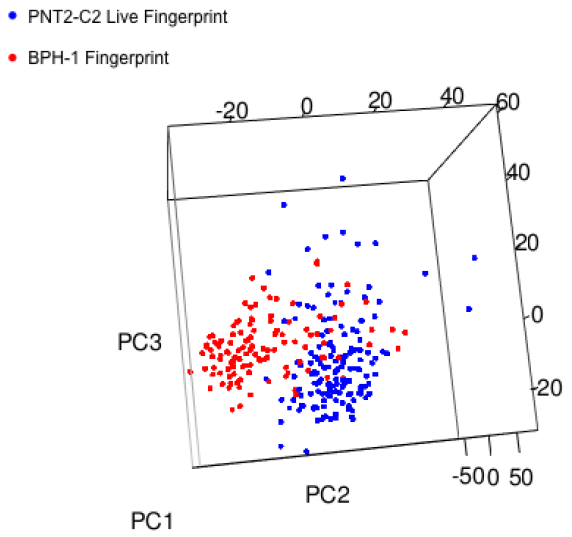


Figure 5.11: A further rotation in PCA space allows observation of separation between PNT2-C2 and BPH-1. It can be seen that the separation is occurring along a diagonal of PC2 and 3. It is also noted that there is an overlap of several points in PCA space.

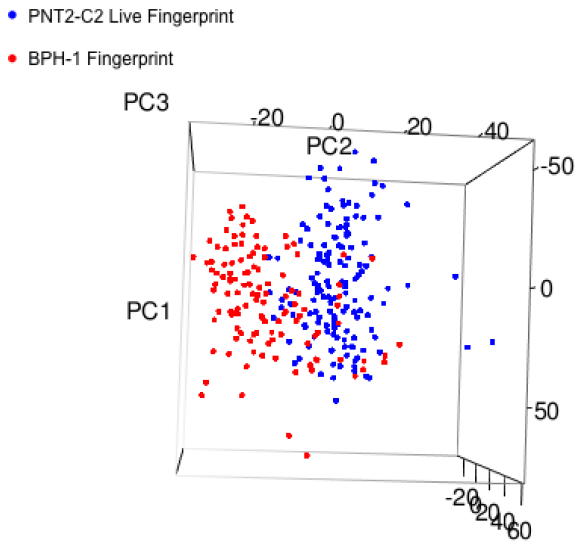


Figure 5.12: A final rotation in PCA space allows observation of separation between PNT2-C2 and BPH-1. It can be seen that the separation is only occurring along PC2 and no separation is occurring in PC1. It is also noted that there is an overlap of a several points in PCA space.

The first rotation of PCA space in Figure 5.10 shows that there is no separation occurring along the axis of PC1 and PC3. In the rotation observed in Figure 5.11 there is separation occurring at an angle along PC1 and PC2. This combination of principal components means the associated loadings will also be considered. A final rotation of PCA space in Figure 5.12 shows that there is no separation happening in PC1. It is for this reason that the associated loading to PC1 will not be considered. The loadings corresponding to PC2 and PC3 are shown in Figures 5.13 and 5.14.

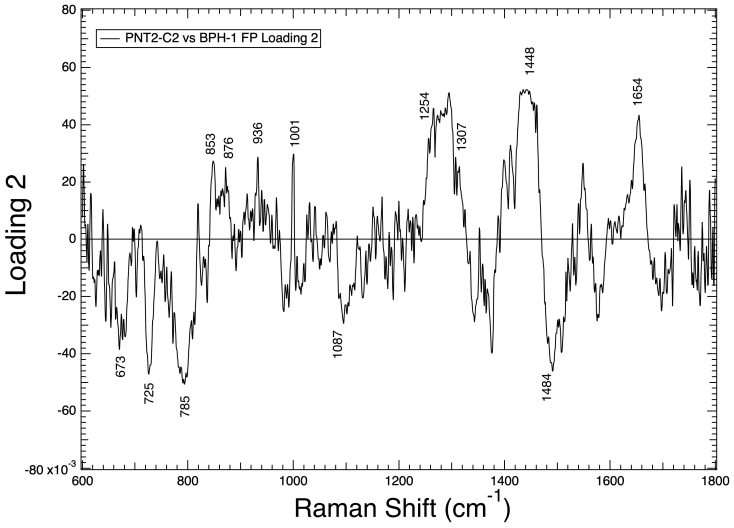


Figure 5.13: The associated loading for PC2 is displayed and annotated. Loading 2 of the fingerprint region comparison of PNT2-C2 Live vs BPH1 is displayed, capturing 11.9% of the variance between samples. Key peaks attributed to biological assignments spanning DNA/RNA, protein and lipids are found at: 673 cm^{-1} (proteins), 725 cm^{-1} (DNA), 785 cm^{-1} (DNA/RNA), 853 cm^{-1} (proteins), 876 cm^{-1} (proteins), 936 cm^{-1} (proteins and carbohydrates), 1001 cm^{-1} (proteins), 1087 cm^{-1} (DNA, lipids and proteins), 1254 cm^{-1} (proteins), 1307 cm^{-1} (proteins, lipids and nucleic acids), 1448 cm^{-1} (proteins, lipids and DNA), 1484 cm^{-1}, (protein and DNA/RNA), 1654 cm^{-1} (proteins and lipids).

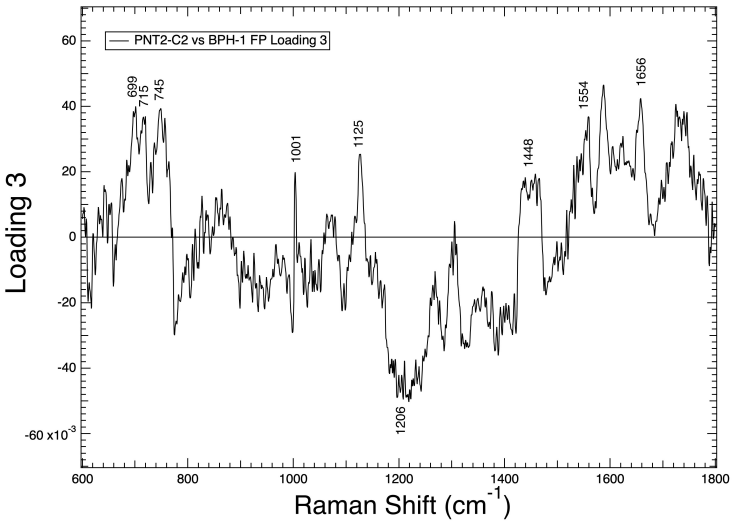


Figure 5.14: The associated loading for PC3 is displayed and annotated. Loading 3 of the fingerprint region comparison of PNT2-C2 Live vs BPH1 is displayed, capturing 8.1% of the variance between samples. Key peaks attributed to biological assignments are found at: 696 cm^{-1} (proteins), 715 cm^{-1} (proteins and lipids), 745 cm^{-1} (DNA), 1001 cm^{-1} (proteins), 1125 cm^{-1} (lipids, proteins and carbohydrates), 1206 cm^{-1} (proteins and DNA/RNA), 1448 cm^{-1} (proteins, lipids and DNA/RNA), 1554 cm^{-1} (proteins), 1656 cm^{-1} (proteins and lipids).

Key peak positions have been extracted from loading 2, Figure 5.13, as indicators of variance between PNT2-C2 and BPH-1. The peaks present in loading 2, shown with their general biomolecular assignment are: 673 cm^{-1} (proteins), 725 cm^{-1} (DNA), 785 cm^{-1} (DNA/RNA), 853 cm^{-1} (proteins), 876 cm^{-1} (proteins), 936 cm^{-1} (proteins and carbohydrates), 1001 cm^{-1} (proteins), 1087 cm^{-1} (DNA, lipids and proteins), 1254 cm^{-1} (proteins), 1307 cm^{-1} (proteins, lipids and nucleic acids), 1448 cm^{-1} (proteins, lipids and DNA), 1484 cm^{-1}, (protein and DNA/RNA), 1654 cm^{-1} (proteins and lipids). Key peak positions have also been taken from Loading 3, Figure 5.14, as indicators of variance between PNT2-C2 and BPH-1. The peaks present in loading 3 are: 696 cm^{-1} (proteins), 715 cm^{-1} (proteins and lipids), 745 cm^{-1} (DNA), 1001 cm^{-1} (proteins), 1125 cm^{-1} (lipids, proteins and carbohydrates), 1206 cm^{-1} (proteins and DNA/RNA), 1448 cm^{-1} (proteins, lipids and DNA/RNA), 1554 cm^{-1} (proteins), 1656 cm^{-1} (proteins and lipids). Particular attention will be paid to the peaks 1001 cm^{-1} and 1448 cm^{-1} at the PIR stage in 5.2.3 as both of these are present in loadings 2 and 3.

PCA analysis between the PNT2-C2 Live and BPH-1 of the high wavenumber region is able to ascertain that the variance captured by the first 6 principal components (PCs) are as follows: PC1 = 47.1%, PC2 = 8.9%, PC3 = 7.0%, PC4 = 3.7%, PC5 = 3.1% and PC6 = 2%. The first three PCs capture 63.0% and the first six PCs capture 71.7% of the total variance of the sample. Cumulatively, 262 principal components are required to capture total variance of the sample. The 3D PCA space of the high wavenumber region using the first three PCs is now considered in Figures 5.15, 5.16 and 5.17

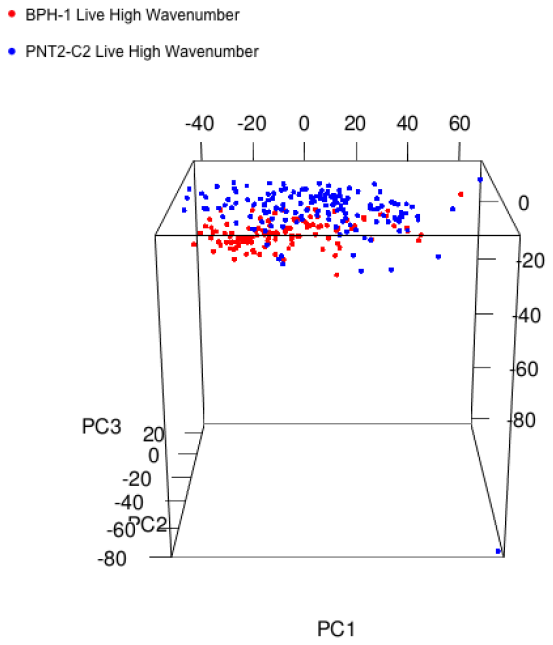


Figure 5.15: The high wavenumber comparison between PNT2-C2 and BPH-1 is displayed in 3D space using the first three principal components. No separation is observed along PC1. Separation between samples is observed along PC3.

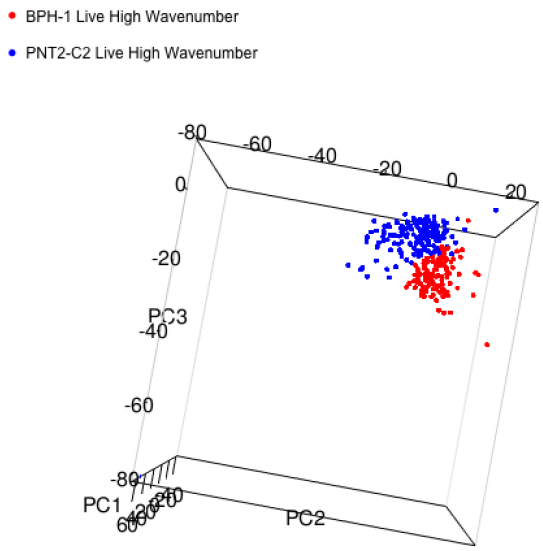


Figure 5.16: PCA analysis of the high wavenumber region of PNT2-C2 and BPH-1 using the first three principal components shows separation is occurring at an angle between PC2 and PC3.

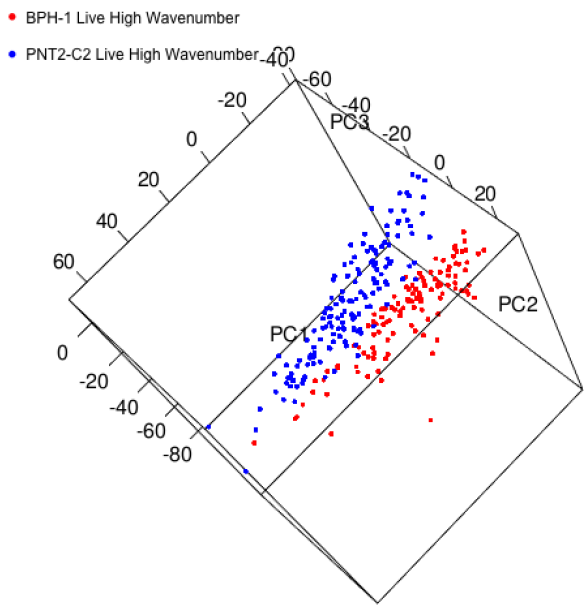


Figure 5.17: A final rotation of PCA space using the first three principal components for the high wavenumber region of PNT2-C2 and BPH-1. Separation is observed as happening only along PC2. There is no separation happening in PC1.

Through interpretation of Figures 5.15, 5.16 and 5.17 separation is observed along the axis of PC2 and PC3. No separation is observed along PC1. As PC1 was not identified as being responsible for separation its corresponding loading is not investigated. Loadings 2 and 3 are investigated further in Figures 5.18 and 5.19.

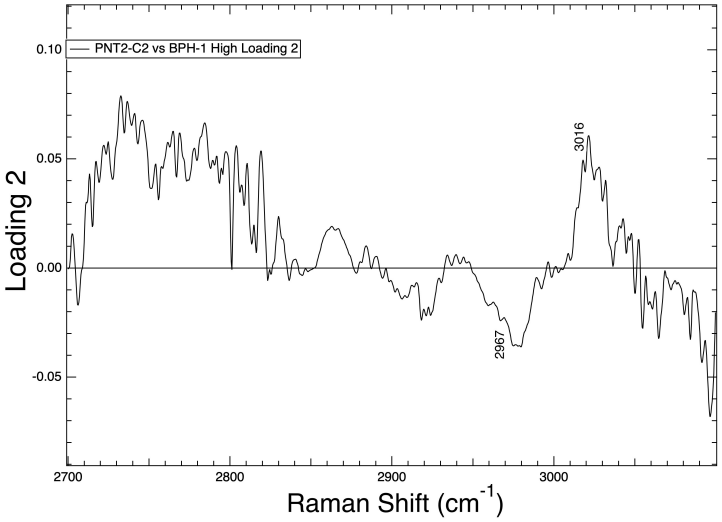


Figure 5.18: Loading 2 corresponding to PC2, capturing 8.9% of the total variance in the high wavenumber region is annotated to show key peak locations. The 3015 cm^{-1} peak (lipids) is annotated as being a point of variance between samples. The 2967 cm^{-1} (lipids) peak is also present but has much less contribution to variance than the 3015 cm^{-1} peak. The background variance present throughout the sample may be the product of a background difference between samples.

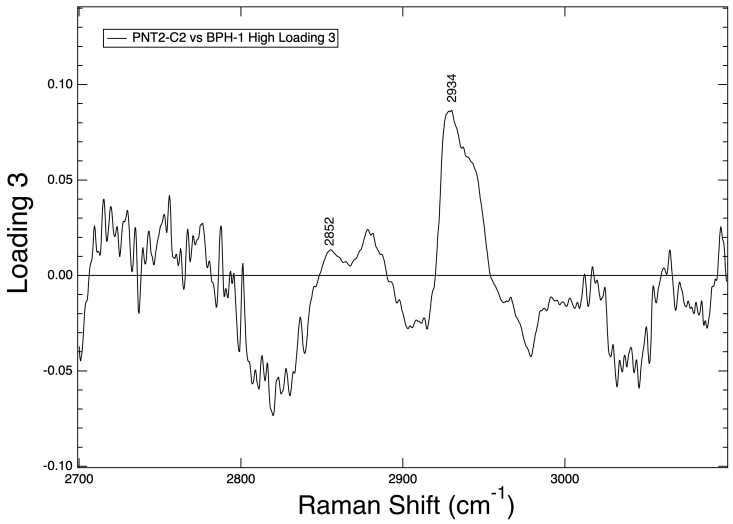


Figure 5.19: Loading 3 corresponding to PC3, capturing 7% of the total variance in the high wavenumber region is annotated to show key peak locations. The key peak emerging in loading 3 is the 2934 cm^{-1} peak (proteins and lipids). The 2852 cm^{-1} peak (lipids) has been annotated for reference but is not considered as generating any separation due to its low amplitude. The rest of the generated signal could be related to the background difference between PNT2-C2 and BPH-1.

In loading 2 and loading 3, Figures 5.18 and 5.19 respectively, there is a noticeable background present. Although biomolecular information may exist in these regions it was chosen not to deconvolve the peaks to such an extent. Loading 2 shows that variance is occurring at both the 2967 cm^{-1} (lipids) and 3015 cm^{-1} peak (lipids). The 2967 cm^{-1} peak has much less contribution to variance than the 3015 cm^{-1} peak in loading 2. It also determined using loading 3 that variance is occurring at the 2852 cm^{-1} peak (lipids) and 2934 cm^{-1} peak (proteins and lipids). The rest of the generated signal could be related to the background profile difference between PNT2-C2 and BPH-1. The peaks associated with the variance in loading 2 and 3 will be investigated further using PIR analysis.

Finally, a PCA-LDA leave-one-out cross-validation was used to determine the predictive classification of PNT2-C2 and BPH-1 cells. The number of principal components included was based on using Kaiser's criterion, which consists of using only principal components with variance greater than 1 percent. For PNT2-C2 vs BPH-1 the PCA-LDA classification tool correctly classified 99.6% cells in the fingerprint region using 9 PCs and 97.7% of cells in the high wavenumber

region using 8 PCs. This demonstrates a high classification accuracy between PNT2-C2 and BPH-1.

5.2.3 Peak intensity ratio analysis PNT2-C2 vs BPH-1

A previous study by Gazi *et al.* (2003) comparing PNT2-C2, BPH-1 and other prostate cancer cell lines using FTIR, only considers one peak intensity ratio to separate all cell lines; (1030 cm^{-1} /1080 cm^{-1}) [124]. Using this ratio, the authors were measuring the glycogen/phosphate ratio which is indicative of the metabolic turnover of the cell [124]. It should be strongly noted, however, that the ratio used by Gazi *et al.* (2003) classified PNT2-C2 as malignant. This is not the case and PNT2-C2 is stringently regarded as a normal cell line [94, 224].

Every peak intensity ratio has been fully analysed to make sure that it is possible to cross-compare between all peaks. This increases the likelihood of the identification of biomarkers. Moreover, it provides an aid to compare between multiple peak intensity ratios, in order to draw more robust conclusions. Small relationships between different peaks may signify significant structural changes in either cell line [260]. The table used to calculate the separation outside of the propagated SE uncertainty, relating to each PIR compared between PNT2-C2 and BPH-1, is given in the Appendix in Table C.2.

An important issue arising from table C.2 is that the 1396 cm^{-1} peak could not be fitted in BPH-1. However, the assignment for the 1396 cm^{-1} peak, CH_2 deformation, is very low in relative intensity and can also found at the much more intense 1448 cm^{-1} peak. It is therefore considered not a loss of vital information. There were 9 peaks associated with a separation of 90% outside of the propagated SE uncertainty, relating to each PIR compared between PNT2-C2 and BPH-1. In the order of peak location these were: 673 cm^{-1} (91%) (DNA/RNA and protein), 699 cm^{-1} (93%) (proteins), 715 cm^{-1} (91%) (proteins and lipids), 876 cm^{-1} (95%) (proteins and lipids), 893 cm^{-1} (93%) (proteins and carbohydrates), 1060 cm^{-1} (91%) (lipids and DNA/RNA), 1155 cm^{-1} (100%) (proteins and carbohydrates), 1234 cm^{-1} (95%) (proteins and DNA/RNA) and 1554 cm^{-1} (93%) (proteins). While

these PIRs indicate changes in the DNA/RNA, proteins, lipids and carbohydrates between PNT2-C2 and BPH-1 there is a higher number of peaks associated with proteins.

The PCA analysis between PNT2-C2 and BPH-1 was used to discriminate between these cell lines in Section 5.2.2. To interpret the differences in variance captured by peaks associated with biomolecular attributes, these are investigated as part of the PIR analysis. In interpretation of loading 2,Figure 5.13, the peaks 785 cm^{-1} (DNA/RNA), 1254 cm^{-1} (proteins), 1448 cm^{-1} (lipids, proteins and DNA/RNA) and 1657 cm^{-1} (proteins and lipids) were the most dominant in terms of contribution to the variance captured by loading 2.

As the largest source of variance at the PCA stage in loading 2, the PIR profile associated with the peak 1448 cm^{-1} (proteins and lipids) is considered first. This peak achieved 53% separation between samples at the PIR stage. The specific biomolecular assignment to the 1448 cm^{-1} peak is CH_2 deformation in lipids and proteins. The PIR panel associated with the 1448 cm^{-1} peak is shown in Figure 5.20.

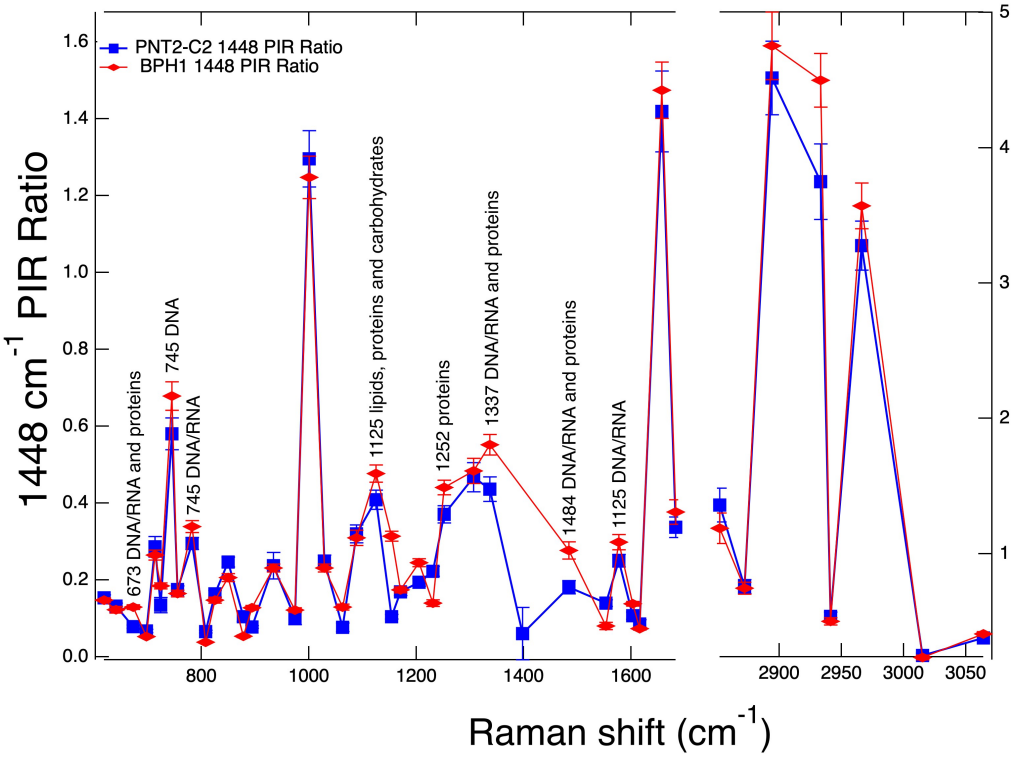


Figure 5.20: The PIR panel for 1448 cm^{-1} (CH_2 deformation in lipids and proteins) peak intensity demonstrating statistically significant separation across the fingerprint region and high wavenumber region, achieving 58% total separation. It can be seen that the 1448 cm^{-1} peak intensity ratio is able to discriminate between DNA/RNA, proteins, lipids and carbohydrates. The largest separation outside of uncertainty between the 1448 cm^{-1} peak, 1337 cm^{-1} peak and 1484 cm^{-1} peak.

As shown in Figure 5.20, the 1448 cm^{-1} peak is able to discriminate between DNA/RNA, proteins, lipids and carbohydrates using peak intensity ratio analysis. The largest separation between the 1448 cm^{-1} peak, 1337 cm^{-1} peak and 1484 cm^{-1} peak. The 1337 cm^{-1} peak intensity is generated by multiple contributions from CH_3 and CH_2 wagging, bending of collagen, lipids, tryptophan, as well as amino acids adenine and guanine. The 1337 cm^{-1} peak also falls into the region convoluted with contributions from the secondary protein structure amide III. The 1484 cm^{-1} peak is generated by a secondary protein amide II as well as Guanine, adenine ring-breathing modes. It is hard to quantify relative specific changes that are happening between the 1448 cm^{-1} peak and these two peaks. However, both of these peaks are associated secondary protein structures. As the 1448 cm^{-1} peak

itself is not representative of secondary protein structure, it is suggested that the 1448 cm^{-1} peak can be used as a biomarker to detect secondary protein structure changes between PNT2-C2 and BPH-1.

The 876 cm^{-1} peak intensity, generating (95%) separation at the PIR stage, was identified as being a source of separation in Loading 2. To consider the 876 cm^{-1} peak, it is also appropriate to consider the 893 cm^{-1} peak, generating (93%) separation respectively. This is because they both fall within the same fitting region. The band 893 cm^{-1} was not identified at any stage of the PCA. Neither peak was identified at the mean profile subtraction in Figure 5.8. The 876 cm^{-1} peak is representative of protein C-C stretching, antisymmetric vibrations of choline $N(CH_3)_3$ (phospholipids), hydroxyproline and tryptophan. Furthermore, the 893 cm^{-1} peak is representative of a saccharide band, methylene, and C-C skeletal stretch. As these bands are both very low signal-to-noise, they are very codependent on each other. It can only be determined that the high degree of separation observed in the 893 cm^{-1} peak PIR analysis is due to a relative change happening in the 876 cm^{-1} peak. It is for this reason that 893 cm^{-1} peak is not suggested as a robust biomarker between PNT2-C2 and BPH-1. It is also decided, due to this factor, that the 876 cm^{-1} peak intensity ratio not be used as a robust biomarker to differentiate samples. Further work could examine the deconvolved fitting of this region but it does not fall into the scope of this project.

The 699 cm^{-1} peak and 715 cm^{-1} peak, which were associated with a high degree of variance in loading 3 as shown in Figure 5.14 of the PCA, were also seen to have a large degree of separation at the PIR stage. The 699 cm^{-1} peak achieved 93% separation and the 715 cm^{-1} peak achieved 91% separation using PIR analysis. The 699 cm^{-1} peak is attributed to proteins and more specifically, amino acid methionine ν(C-S) *trans*. The 715 cm^{-1} peak is also attributed to amino acid methionine ν(C-S) *trans* as well as the $CN^{+}(CH_3)_3$ stretch in lipids. The PIR panel for the 699 cm^{-1} peak is shown in Figure 5.21 as well as the 715 cm^{-1} peak in Figure5.22.

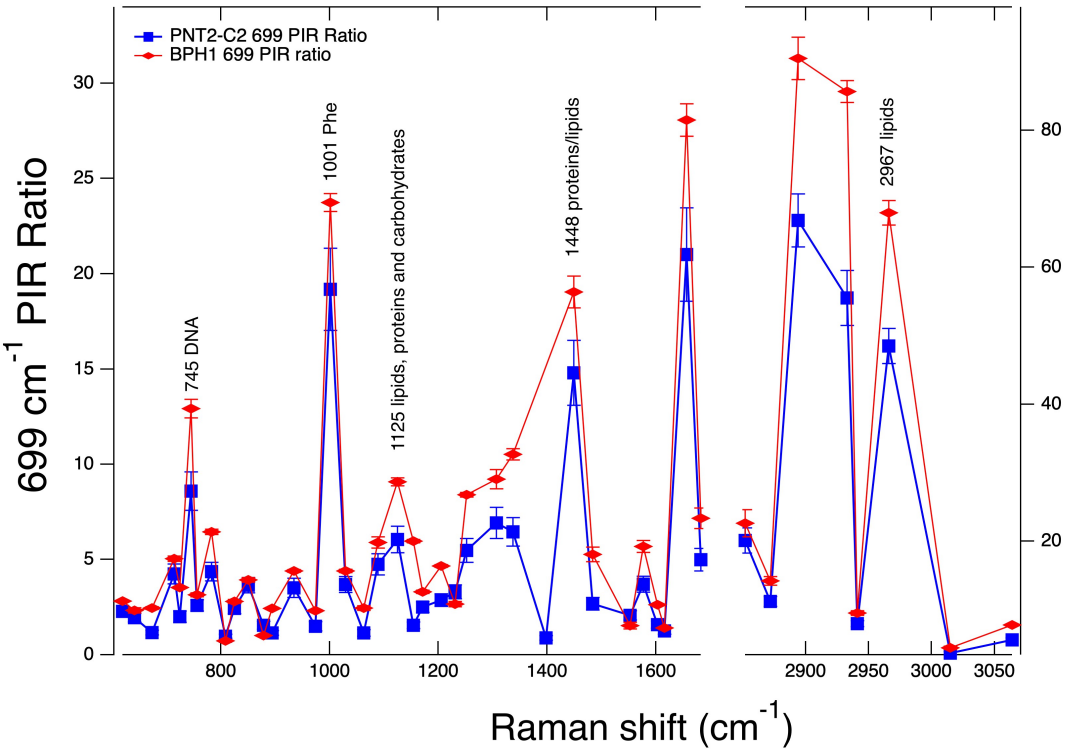


Figure 5.21: The PIR panel for the 699 cm^{-1} peak intensity ratio demonstrates statistically significant separation across the fingerprint region and high wavenumber region, achieving 93% total separation. The 699 cm^{-1} peak is attributed, specifically, to amino acid methionine ν(C-S) *trans*. It can be seen that by using the 699 cm^{-1} peak as peak intensity ratio marker that there is a lower relative content of amino acid methionine in BPH-1. This is concluded due to statistically separated values between DNA/RNA, proteins, lipids, carbohydrates, phenylalanine and nucleic acids.

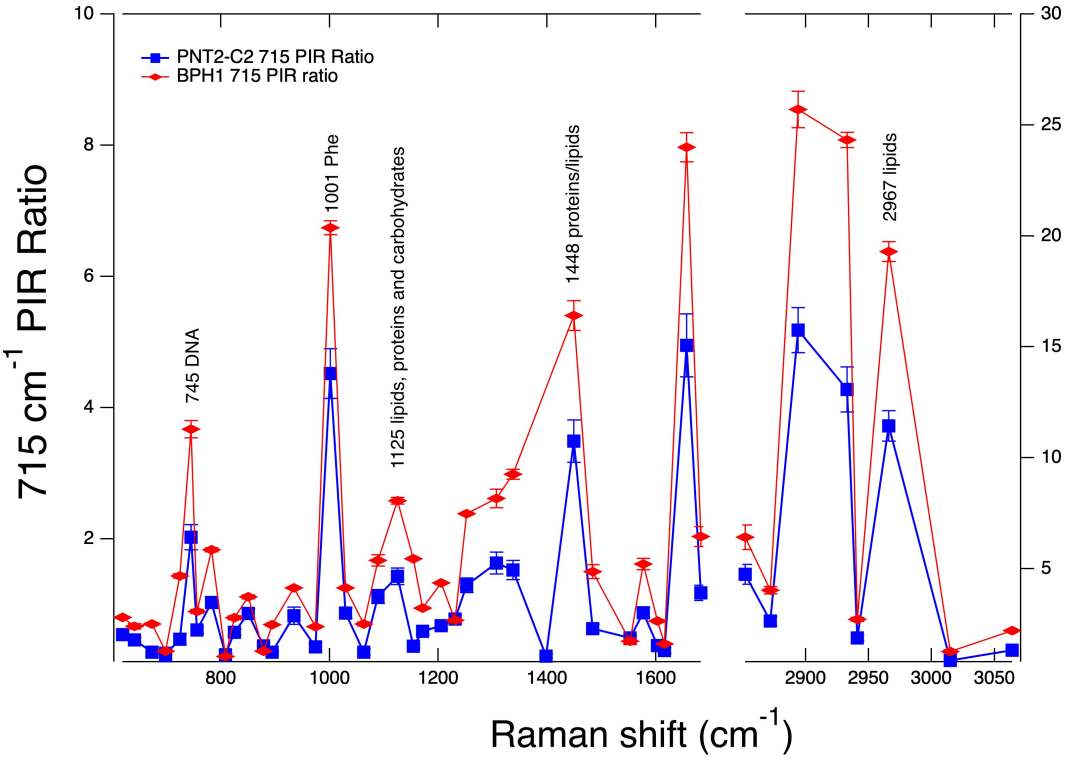


Figure 5.22: The PIR panel for the 715 cm^{-1} peak intensity ratio demonstrates statistically significant separation across the fingerprint region and high wavenumber region, achieving 91% total separation. The 715 cm^{-1} peak is attributed, specifically, to amino acid methionine ν(C-S) *trans* as well as the $CN^{+}(CH_3)_3$ stretching in lipids. It can be seen that by using the 715 cm^{-1} peak as peak intensity ratio marker that there is a lower relative content of amino acid methionine in BPH-1. This is concluded due to statistically separated values between DNA/RNA, proteins, lipids, carbohydrates, phenylalanine and nucleic acids.

From Figures 5.21 and 5.22 it is seen that there is a lower relative presence of amino acid methionine ν(C-S) *trans* in BPH-1 compared to PNT2-C2. Amino acid methionine is an essential amino acid that is essential for growth and development of the cell [265]. There is, however, a lack of research on methionine analysis as part of Raman studies on prostate cells and other cell systems. It is for this reason that results from other techniques are considered. There are several possible explanations for this result. For example, a study by Toth *et al.* (2005) focusing on methionine in prostate cells, notes that in BPH the expected amount of methionine is much lower than in tumorous lesions because of a lower amount of cells with high amino acid consumption rates [256]. However, the study by Toth *et al.* (2005) did not draw comparisons to normal cells. Another study by Lu *et al.*

(2000) showed that by using Western blot analysis that restriction of methionine inhibited the growth of the prostate cell lines PC3 and DU145 [265]. However, the study by Lu *et al.* (2000), did not consider a benign prostate hyperplasia cell line model in its comparisons. This thesis will also investigate the separation between cell lines study by Lu *et al.* (2000) as well as the addition of BPH-1, PNT2-C2 and P4E6 at a later stage in Section 5.7. A study of methionine on BPH-1 by Benavides *et al.* (2010) demonstrated using flow cytometry, cell cycle-related gene expression by microarray analysis and QRT-PCR, that methionine inhibition treatment of the standard cell line BPH-1 had no effect on proliferation [271]. Using the result derived from Benavides *et al.* (2010) and comparison to reduced relative intensity of methionine in BPH-1 in Figures 5.21 and 5.22, it could be suggested that BPH-1 growth and proliferation is methionine independent or is not using it as a main chemical pathway. Further work using different analysis techniques would have to be done to investigate this. Due to the strong separation between PNT2-C2 and BPH-1 using the 699 cm^{-1} and 715 cm^{-1} peaks in PIR analysis, as well as contributing to the variance in loading 3, they should also be considered a robust biomarker to differentiate between the two cell populations.

The peak 673 cm^{-1} was identified in loading 2 as contributing to the separation between PNT2-C2 and BPH-1. The 673 cm^{-1} peak also generated (91%) separation between PNT2-C2 and BPH-1 during PIR analysis. The specific biomolecular assignment of the 673 cm^{-1} peak is a protein contribution from the C-S stretching mode of cystine as well as a DNA/RNA contribution from Guanine and thymine ring breathing modes. It is also attributed to tyrosine G backbone in RNA. The PIR panel generated by the peak 673 cm^{-1} is displayed in Figure 5.23

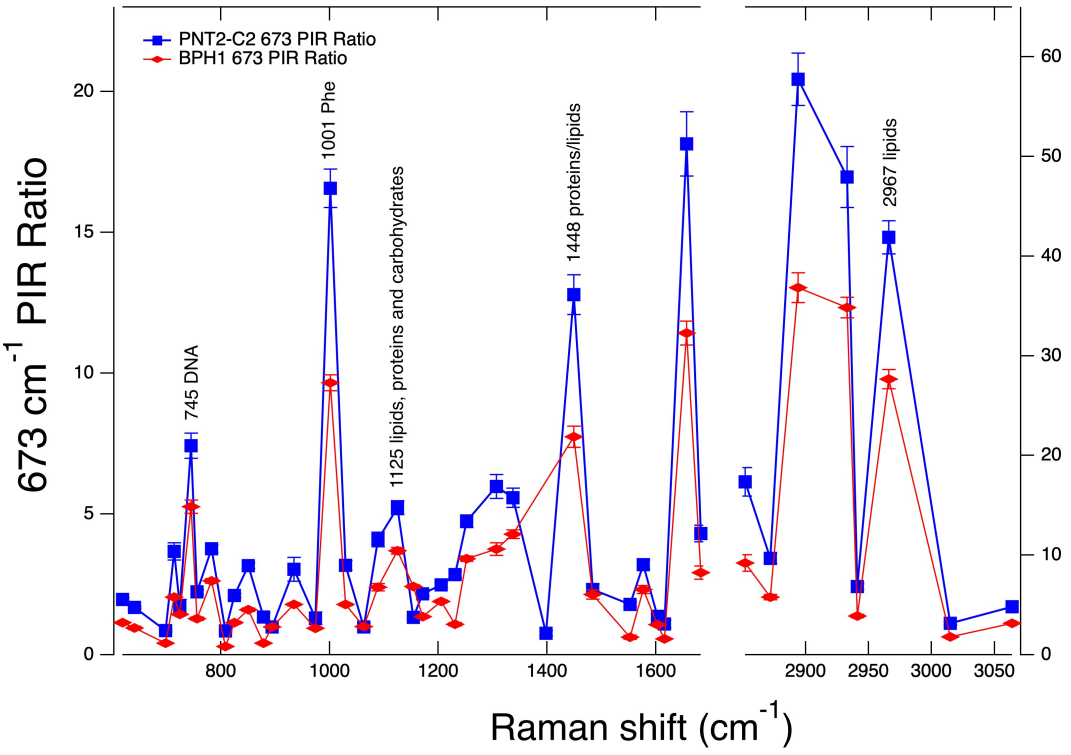


Figure 5.23: The PIR panel for the 673 cm^{-1} peak intensity ratio demonstrates statistically significant separation across the fingerprint region and high wavenumber region, achieving 91% total separation. The specific biomolecular assignment of the 673 cm^{-1} peak is a protein contribution from the C-S stretching mode of cystine as well as a DNA/RNA contribution from Guanine and thymine ring breathing modes. It is also attributed to tyrosine G backbone in RNA. It can be seen that by using the 673 cm^{-1} peak as peak intensity ratio marker that there is a higher relative content of cystine in BPH-1. This can only be said for assignment of cystine, as the other biomolecular contributions also exist at other peak locations that are not considered in this comparison. The relative intensity of cystine is lower relative to DNA/RNA, proteins, lipids, carbohydrates, phenylalanine and nucleic acids.

As described in Figure 5.23, the specific contribution from cystine is relevant due to the 673 cm^{-1} peak being the only location it is attributed to. The attribution to cystine is of particular relevance. Cystine C-S stretching represents an amino acid that consists of sulfur obtained from the oxidation of two cysteine molecules [272]. This bond is linked through a disulfide bond. The reason this difference is so vital to understanding BPH-1 is that cystine is used to produce glutathione (GSH), which protects against induced cell death in human prostate epithelial cell lines [273]. It is also known that GSH regulates cell growth and proliferation

[83]. As shown in Figure 5.23, the ratio of cystine is found to be significantly higher in BPH-1 than in PNT2-C2 when compared to DNA/RNA, proteins, lipids, carbohydrates, phenylalanine and nucleic acids across the cell profile. Due to the strong separation between PNT2-C2 and BPH-1 using the 673 cm^{-1} peak in PIR analysis, as well as it contributing to the variance in loading 3, it should be considered a robust biomarker to differentiate the respective samples.

It could be tentatively hypothesised, based on the analysis of comparing PNT2-C2 and BPH-1, using the results derived from the peaks 673 cm^{-1}, 699 cm^{-1} and 715 cm^{-1}, that the growth and proliferation of the cell line BPH-1 is being driven by cystine as opposed to methionine. This is thought to be the case due to the higher relative contribution from cystine and lower relative contribution of methionine in BPH-1 to PNT2-C2. Both cystine and methionine can be used to produce glutathione (GSH), which controls cell growth and proliferation [83]. However, it is not possible to form any conclusion to this hypothesis due to the limitations of the analysis techniques used in this thesis.

Another peak with considerable separation is the 1234 cm^{-1} peak, generating 95% separation. The 1234 cm^{-1} peak is generated by a contribution of Amide III and CH_2 wagging of glycine and proline as well as a PO_2^- antisymmetric stretch. It must be also noted that the peak intensity of the 1234 cm^{-1} band depends on the profile of the fitting region, that also includes the 1254 cm^{-1} band, which achieved 83% separation between samples. Although the 1234 cm^{-1} peak was not identified at the PCA stage as a source of variance, the 1254 cm^{-1} was identified as contributing to separation in loading 2. The 1254 cm^{-1} peak is also attributed to amide III : CH_2 wagging. The PIR panel associated with the 1234 cm^{-1} peak is show in Figure 5.24

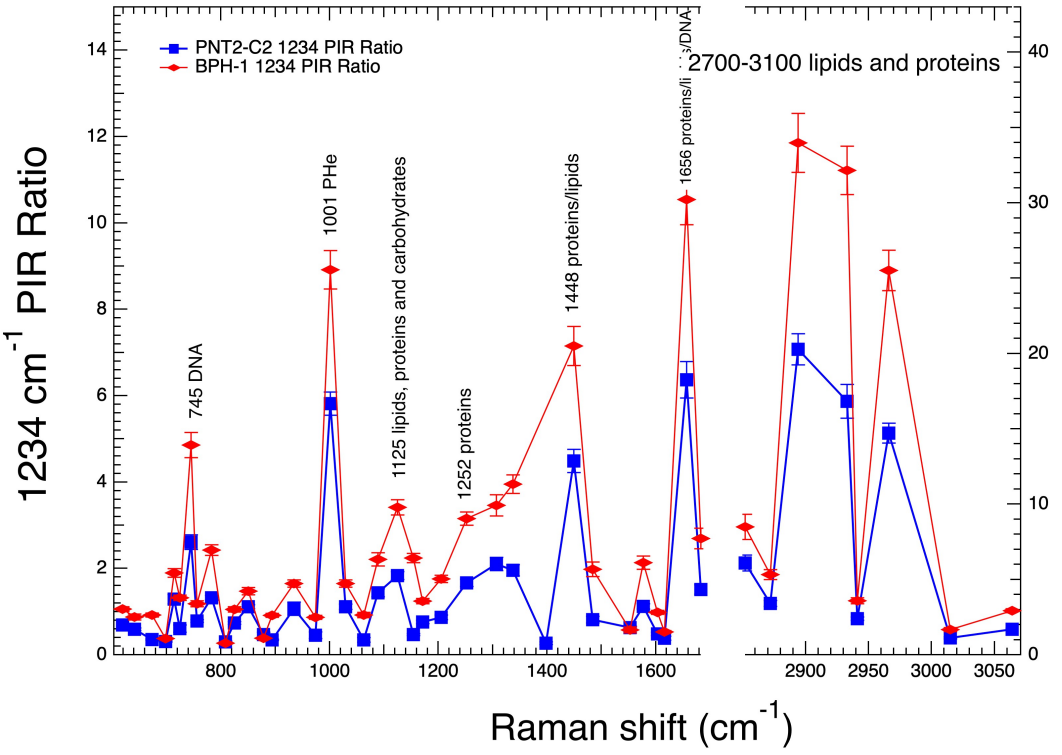


Figure 5.24: The PIR panel for the 1234 cm^{-1} peak intensity ratio demonstrates statistically significant separation across the fingerprint region and high wavenumber region, achieving 95% total separation. The 1234 cm^{-1} band is attributed to a contribution of Amide III and CH_2 wagging of glycine and proline as well as a PO_2^- antisymmetric stretch. It can be seen that by using the 1234 cm^{-1} peak intensity ratio marker that there is a lower relative intensity of amide III and PO_2^- antisymmetric stretching in BPH-1. This observation is made due to the high relative intensity of peak ratios relating to the 1234 cm^{-1} peak against DNA/RNA, lipids, proteins and carbohydrates.

The changes in relative ratios against the 1234 cm^{-1} peak in Figure 5.24 show that there is less relative intensity of amide III and PO_2^- antisymmetric stretching in BPH-1. This suggest that there is a difference in not only in the DNA structure of BPH-1 but in the secondary protein structure. This in turn suggests a greater change in the amide III region between 1205-1337 cm^{-1}, which would suggest that there may be changes occurring to the secondary amide III protein structure; specifically, to the coil state of amide III. However, due to the 1234 cm^{-1} peak having strong dependence on the 1254 cm^{-1} peak, which was seen as having significant contribution to separation in loading 2, it should not be considered a robust biomarker of discriminating between PNT2-C2 and BPH-1.

The 100 % separation that is occurring at the peak 1155 cm^{-1} should be interpreted with caution due to a very sharp narrow peak with high amplitude. The nature of this peak makes it very difficult to fit. This peak does, however, feature as part of a converged data set. Biochemically it is attributed to a C-C and C-N stretch in proteins, carotenoids correlated to disease and glycogen. This peak was not considered significant in the PCA analysis or mean profile subtraction. Whilst the findings at the PIR stage would substantiate claims made by Gazi *et al.* (2003) that glycogen may play an important role as a biomarker of benign prostate hyperplasia, there is not enough evidence to support its role as a biomarker in this thesis.

Lastly, in the fingerprint region, achieving separation of 93%, is the peak 1554 cm^{-1}, representative of Tryptophan and Amide II. This peak was not identified in the mean profile subtraction between PNT2-C2 and BPH-1. However, the peak 1554 cm^{-1} was also identified in loading 3 in Figure 5.14 as contributing to the separation between PNT2-C2 and BPH-1 in PCA space. The corresponding PIR panel below is shown to investigate the 1554 cm^{-1} peak intensity further.

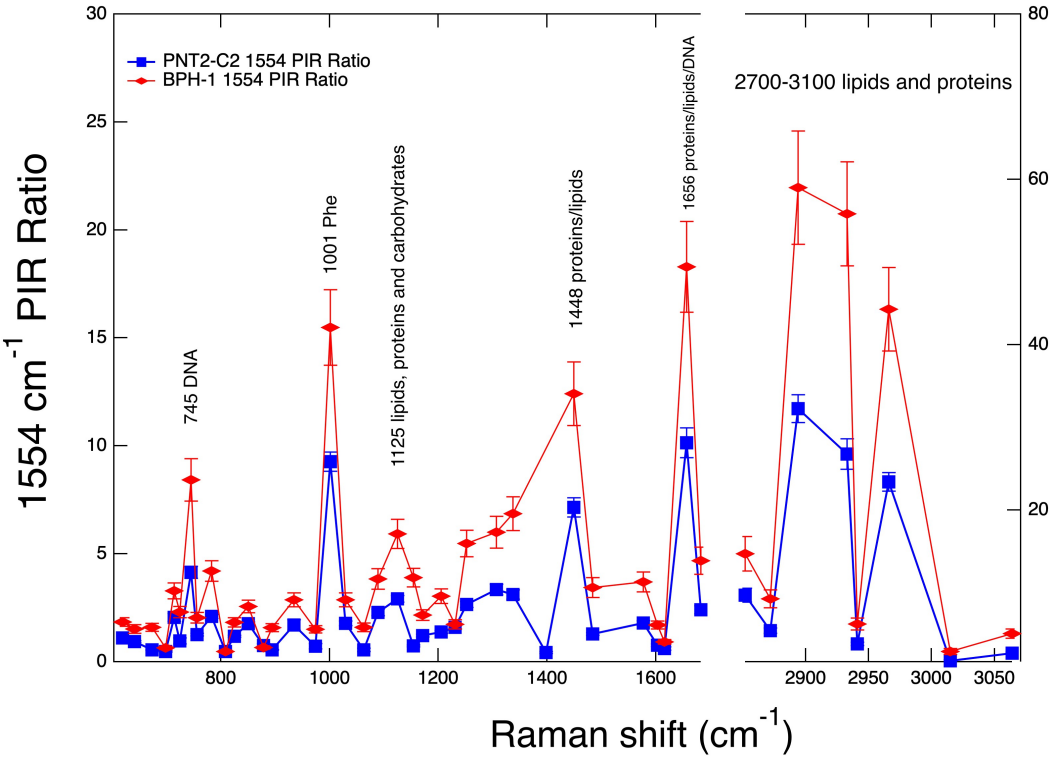


Figure 5.25: The PIR panel for the 1554 cm^{-1} peak intensity ratio demonstrates statistically significant separation across the fingerprint region and high wavenumber region, achieving 93% total separation. The 1554 cm^{-1} peak is biomolecular representative of Tryptophan and Amide II. It can be seen that PIR ratios using the 1554 cm^{-1} peak generated higher relative peak intensity ratios across the fingerprint and high wavenumber region for BPH-1. This observation indicates that there is a lower relative presence of amide II and tryptophan in BPH-1.

The PIR panel for the 1554 cm^{-1} amide II and tryptophan peak intensity ratio, as shown in Figure 5.25, demonstrates that there is a relative decrease of amide I in BPH-1 when compared to PNT2-C2. This is observed throughout out the PIR profile of the 1554 cm^{-1} peak in BPH-1, a relative increase in intensities of all peak intensity ratio in both the fingerprint and high wavenumber region across DNA, proteins, lipids and nucleic acids. This further demonstrates secondary protein structure changes between BPH-1 and PNT2-C2. Additionally, it suggests that there is a smaller contribution of tryptophan in BPH-1. Tryptophan is an important amino acid that plays a role in apoptosis and, during disease, inflammatory response [274, 275]. It is for this reason that the 1554 cm^{-1} amide II and tryptophan peak is determined as a robust biomarker.

In the high wavenumber region, no peaks were identified at the PIR stage as generating 90% separation outside of the propagated SE uncertainty, relating to each PIR compared between PNT2-C2 and BPH-1. To consider the high wavenumber region, the mean profile subtraction in Figure 5.9 is used for informed analysis, in conjunction with the PIR and PCA results. The 2934 cm^{-1} peak was selected due to the strong contribution to variance in loading 3, as displayed in Figure 5.19. Although this peak did not appear in the mean profile subtraction, the PCA analysis was determined to be sufficient for the selection of this peak. The 2934 cm^{-1} peak generated 70% separation outside of the propagated SE uncertainty in the fingerprint and high wavenumber region at the PIR stage. The 2934 cm^{-1} band is representative of CH_2 asymmetric stretching and CH_3 symmetric stretching. The PIR corresponding to the 2934 cm^{-1} band is shown in Figure 5.26

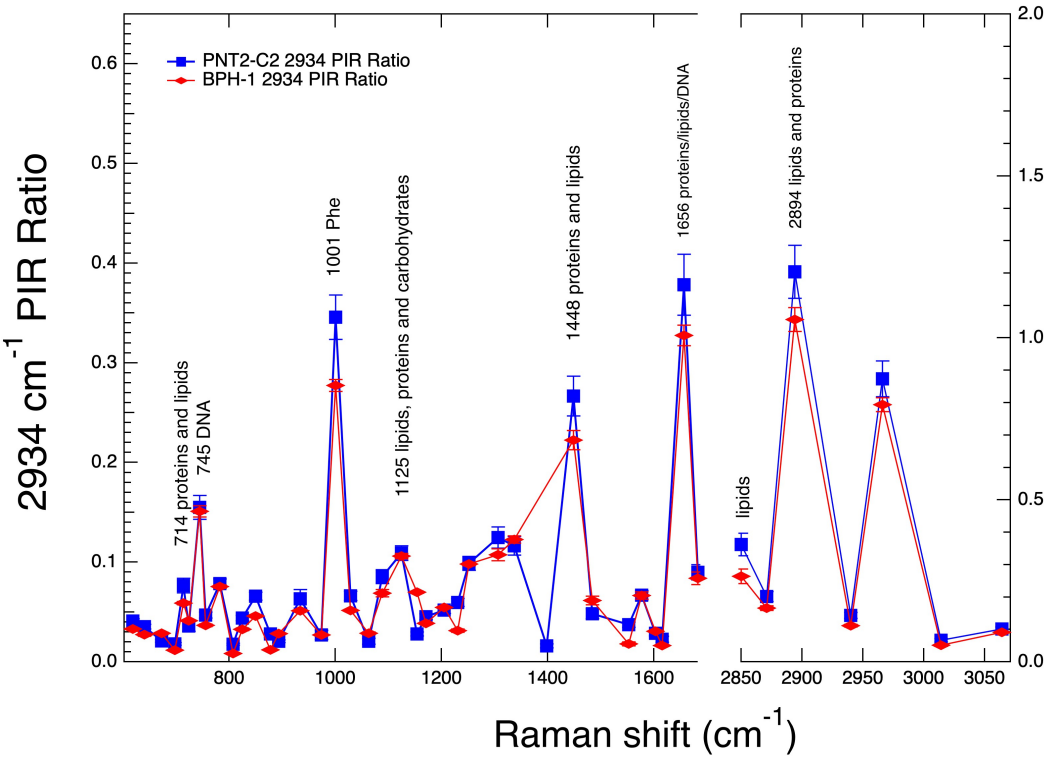


Figure 5.26: The PIR panel for the 2934 cm^{-1} peak intensity ratio demonstrates statistically significant separation across the fingerprint region and high wavenumber region, achieving 70% total separation. The 2934 cm^{-1} band is representative of CH_2 asymmetric stretching and CH_3 symmetric stretching. It can be observed at the I745/I2934 peak that no separation is occurring between the 745 cm^{-1} DNA peak and 2934 cm^{-1} peak. Similarly, no separation is observed at I1125/I2934 peak.The largest relative separation in the fingerprint region is observed at the 1001 cm^{-1} phenylalanine peak. In the high wavenumber region the largest relative separation is observed in the lipid peak 2852 cm^{-1}. The identified relative differences would suggest that there was a higher relative presence of CH_2 asymmetric stretching and CH_3 symmetric stretching in BPH-1.

The PIR panel for the 2934 cm^{-1} peak intensity ratio shown in Figure 5.26 demonstrates statistically significant separation across the fingerprint region and high wavenumber region, achieving 70% total separation. It can be observed in Figure 5.26 at the I745/I2934 peak that no separation is occurring between the 745 cm^{-1} DNA peak and 2934 cm^{-1} peak. Similarly, no separation is observed at I1125/I2934 peak.The largest relative separation in the fingerprint region is observed at the 1001 cm^{-1} phenylalanine peak. In the high wavenumber region the largest relative separation is observed in the lipid peak 2852 cm^{-1}. The identified relative differences would suggest that there was a higher relative

presence of CH_2 asymmetric stretching and CH_3 symmetric stretching in BPH-1. This peak was used by Potcoava *et al.* (2014) as a biomarker to distinguish between cancerous prostate and breast cancer cell lines. This is due to its relationship in determining the chain length of fatty acids [64]. This is not explored at this stage, but will be explored in Section 5.7.

5.3 LNCaP

There have been several papers that have used Raman spectroscopy to analyse LNCaP cells. A paper published by Potcoava *et al.* (2014) used micro-Raman spectroscopy and coherent anti-Stokes Raman scattering microscopy to characterise the lipid profile of LNCaP [64]. The main purpose of the study carried out by Potcoava *et al.* (2014) was to investigate the intracellular lipid droplets and their saturation in cells during hormone response therapy, as well as the process of lipogensis. Potcoava *et al.* (2014) was able to successfully observe higher numbers and concentration of intracellular lipid droplets in hormone treated LNCaP cultures compared to an untreated version. The experimental setup of Potcoava *et al.* (2014) varied with the use of a different laser and fixation methods to this thesis. The number of cells measured in their study was not an exact value, with a total of 28 to 32 LNCaP cells measured with 10 to 20 spectral measurements done per cell. As discussed in Section 5.1.2, this inaccuracy on reporting the number of cells measured will have an effect not only on convergence of the Raman profile cells, but also the training of diagnostic models, should it be required for future studies. Furthermore, the use of glass cover-slips by Potcoava *et al.* (2014) should also be highlighted. Whilst Potcoava *et al* (2014) did mention that the Raman spectra was convoluted with the Raman spectra of the glass cover slips, and that its signal was removed using post processing, it is still likely that the deconvolution process will result in data loss. Analysis by Potcoava *et al.* (2014) covered peak intensity ratio analysis as well as PCA/LDA analysis. One major criticism of the study carried out by Potcoava *et al.* (2014) is that there was no normal control sample in their study, electing to include two cancerous prostate cell lines. Whilst Potcoava *et al.* (2014) was able to differentiate between between treated and untreated samples, it is unclear what effect treatment would have had on a normal cell line. The results were cross referenced and validated using comparable results from a similar study which used chromatography in the analysis of breast cancers [64, 276]. This thesis will expand on the work carried out by Potcoava *et al.* (2014) by considering the DNA/RNA and protein assignments in the fingerprint and high wavenumber region for use as potential biomarkers.

A study that did include both a normal cell line and LNCaP was published by Taleb *et al.* (2006). Although only considering the fingerprint region (600-1800 cm^{-1}), Taleb *et al.* (2006) analysed PNT-1a (an immortalized normal cell line) and LNCaP [142]. The cell line PNT-1a is considered a comparable cell line to PNT-C2 but they can not be directly compared. Furthermore, there are notable differences in experimental setup between Taleb *et al.* (2006) and this thesis. The wavelength of laser used by Taleb *et al.* (2006) is 633 nm and the cells are measured after being removed from medium and dried on quartz [142]. There are not many comparative studies of PNT-1a against PNT2-C2, but a study by Dawson *et al.* (2006) found a difference in morphology and endothelin (ET-1) protein immunoreactivity, suggesting biochemical differences between the cell lines [277]. This protein is known to have an effect on the epithelial growth of cells [277]. However, the study by Dawson *et al.* (2006) does not consider the Raman profile of either PNT-1a or PNT2-C2 [277]. The difference in intracellular proteins between these two cell lines indicates that any results found in this thesis may not be directly comparable to the study by Taleb *et al.* (2006). Key results found by Taleb *et al.* (2006) include a higher presence of B conformation DNA in PNT-1a and higher presence of A conformation DNA in LNCaP [142]. Although Taleb *et al.* (2006) did not use peak intensity ratios, they did use PCA to separate the samples, achieving 90% classification. Taleb *et al.* (2006) also managed to achieve 100% separation of PNT-1a and LNCaP using adjacent band ratios, a technique not explored or considered by this thesis [142]. This thesis will add to the current body of knowledge through not only considering the high wavenumber region for analysis, but also carrying out this comparison using live cells for the first time, as far as the author is aware.

Another key publication using Raman spectroscopy of LNCaP was carried out by Crow *et al.* (2005). Crow *et al.* (2005) used principal component analysis and linear discriminant analysis to construct a diagnostic algorithm capable of separating four cell lines to a sensitivity of 98% and a specificity of 99% [63]. The four cell lines included were LNCaP, DU145, PC3 and MDAPCa 2b [63]. However, Crow *et al.* (2005) used an 830 nm laser meaning that it is not possible to directly

cross compare Raman profiles. The number of spectra collected for the cell line LNCaP was 49 spectra [63]. A key issue with this experiment was that there was no reporting on the convergence of data. As the experimental setup differs to this thesis, it is difficult to determine how many spectra are required for an accurate representation of a population. Nonetheless, it is particularly important that convergence is achieved in the case of Crow *et al.* (2005), as they are trying to establish a diagnostic algorithm. On closer inspection of methodology, the cell measurement and preparation techniques also expose the study by Crow *et al.* (2005) to potential issues. After cell preparation the calcium fluoride slide was kept moist before being taken back to the Raman spectrometer for measurement in a 30 minute window [63]. However, as shown in chapter 4 there is a difference in the Raman signature of dried cells. Crow *et al* (2005) have tried to minimise this with the short window of measurement but it is still expected that spectral deviation will occur as a function of time. A summary of results from the study by Crow *et al.* (2005) identified higher glycogen in LNCaP/MDApca 2b than in DU145 and PC3 [63]. The work in this thesis will build on their diagnostic model, with an increased number of converged spectra for the cell lines also present in this thesis. It will also consider the high wavenumber region. Most importantly it will include a normal prostate cell line, allowing comparison of glycogen content, not only to other cancerous cell lines, but a normal baseline. The use of LNCaP in a diagnostic algorithm, against a wider set of cell lines, is further explored in Section 5.7.

5.3.1 PNT2-C2 vs LNCaP mean profile subtraction

A mean profile subtraction has been carried out between PNT2-C2 and LNCaP to allow for initial observation of differences between the cell lines. The normalised Raman signal for PNT2-C2 and LNCaP, with respect to the total spectrum area, is displayed for the fingerprint region in Figure 5.27 and the high wavenumber region is shown in Figure 5.9. The fingerprint region has undergone a linear baseline subtraction as well as being offset for an arbitrary amount. The high wavenumber region has undergone a polynomial baseline subtraction as well as

being offset by an arbitrary amount for comparison purposes. The corresponding mean subtraction profile (PNT2-C2 minus LNCaP) is shown for the fingerprint region in Figure 5.27 and the high wavenumber region in Figure 5.28.

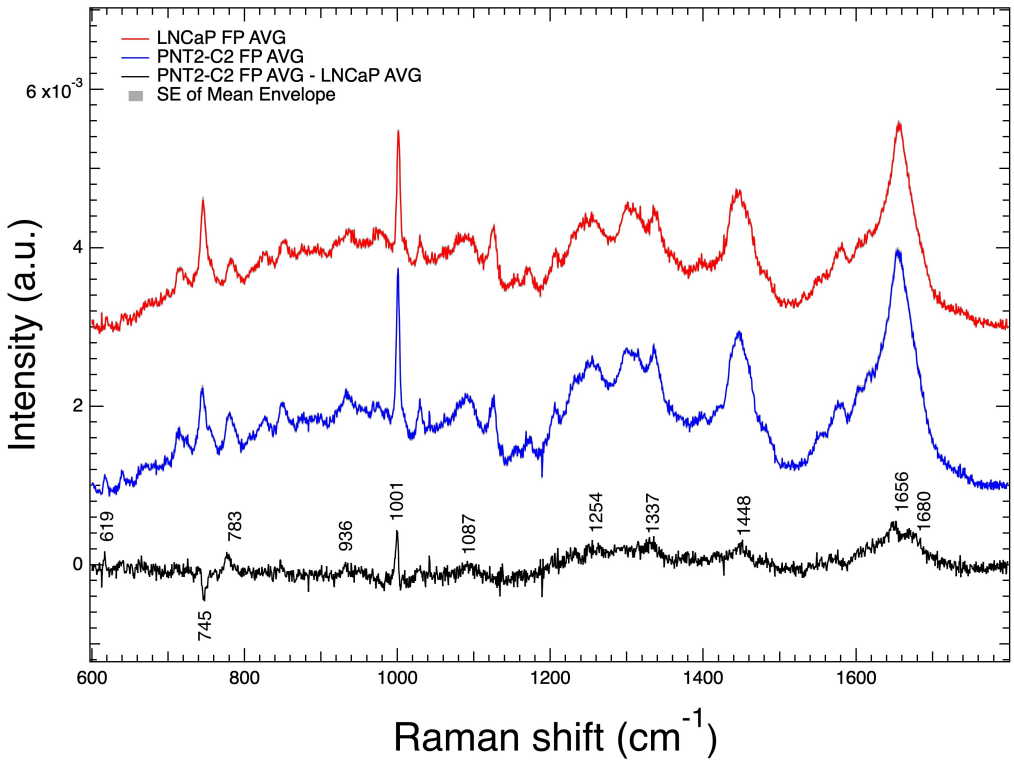


Figure 5.27: The mean subtraction between the fingerprint region of PNT2-C2 and LNCaP is shown. PNT2-C2 and LNCaP have been normalised with respect to the total spectrum area and the standard error of the mean envelope has been plotted to view the associated error. There is large variation between spectrum detailed by the annotated peaks: 619 cm^{-1} (proteins), 745 cm^{-1} (DNA), 783 cm^{-1} (DNA/RNA), 936 cm^{-1} (proteins and carbohydrates), 1001 cm^{-1} (proteins), 1087 cm^{-1} (DNA, lipids and proteins), 1254 cm^{-1} (proteins), 1337 cm^{-1} (proteins and DNA/RNA), 1448 cm^{-1} (proteins, lipids and DNA/RNA), 1656 cm^{-1} (proteins and lipids) and 1681 cm^{-1} (proteins). A large relative difference in the mean subtraction profile is observed at the 745 cm^{-1} (DNA) position, indicating a difference in the DNA between PNT2-C2 and LNCaP [128, 129]. The profile shape between 936 cm^{-1} and 1001 cm^{-1} appears to be noticeably different between cell lines. This change in profile is also observed at 1087 cm^{-1} and in the amide III region 1254 cm^{-1} to 1337 cm^{-1} [128, 129]. These changes in profile shape will be investigated further using PCA analysis then PIR analysis. Additionally, further variations in lipid and protein bands are found in the mean subtraction profile at 1448 cm^{-1}, 1656 cm^{-1} and 1680 cm^{-1}.

As seen in Figure 5.27, there are a total of eleven peaks in the fingerprint region that have been identified during the mean profile subtraction between

PNT2-C2 and LNCaP. The 11 peaks identified are: 619 cm^{-1} (proteins), 745 cm^{-1} (DNA), 783 cm^{-1} (DNA/RNA), 936 cm^{-1} (proteins and carbohydrates), 1001 cm^{-1} (proteins), 1087 cm^{-1} (DNA, lipids and proteins), 1254 cm^{-1} (proteins), 1337 cm^{-1} (proteins and DNA/RNA), 1448 cm^{-1} (proteins, lipids and DNA/RNA), 1656 cm^{-1} (proteins and lipids) and 1681 cm^{-1} (proteins). The 619 cm^{-1} peak is specifically attributed to a phenylalanine C-C twist, as well as a c-c twist in proteins. Further evidence of a potential difference is the relative phenylalanine content also found at the 1001 cm^{-1} peak. The study by Taleb *et al.* (2006) also found this difference between the normal PNT-1a cell line and LNCaP at the 1001 cm^{-1} phenylalanine peak using a mean profile subtraction. The evidence from the mean profile subtraction in Figure 5.27 and corresponding literature would infer there is a difference in phenylalanine between PNT2-C2 and LNCaP. Two large differences are occurring at the 745 cm^{-1} and 783 cm^{-1} peaks, representative of DNA and DNA/RNA respectively [128, 129]. The region 936 cm^{-1} to 1001 cm^{-1}, whilst emerging in the mean profile subtraction, takes on very different profile shape in LNCaP when compared to PNT-C2. It is not possible to determine the exact reason for the change in profile shape at this stage of analysis. Very tentatively, at this stage, it is suggested that the drop in intensity around 936 cm^{-1} which is attributed to a C-C backbone stretch in the α-helix, indicating that there is more B conformation DNA in PNT2-C2 [237, 261]. Further evidence for a change in DNA conformation will be explored using the 1656 cm^{-1} peak, representative of amide I and α-helix conformational proteins, which also is present in the mean profile subtraction. There is also another amide I band found at 1681 cm^{-1} which is convoluted with the 1656 cm^{-1} peak present in the mean subtraction. Figure 5.27 also demonstrates that a change in profile shape is occurring around the bands 1087 cm^{-1}, 1254 cm^{-1} and 1337 cm^{-1}. The relatively large intensity difference occurring between 1254 cm^{-1} and 1337 cm^{-1}, the region in which the amide III region spans, could be indicative of differences in secondary protein structure [128, 129]. Lastly, the 1448 peak cm^{-1}, generated primarily by a deformation in lipids and protein provides further evidence for variation between the lipid and protein structure of PNT2-C2 and LNCaP. These peaks will be further investigated using PCA analysis and peak intensity ratio analysis to

determine the largest biomolecular differences between PNT2-C2 and LNCaP.

The mean profile subtraction for PNT2-C2 and LNCaP for the high wavenumber region is now considered in Figure 5.28.

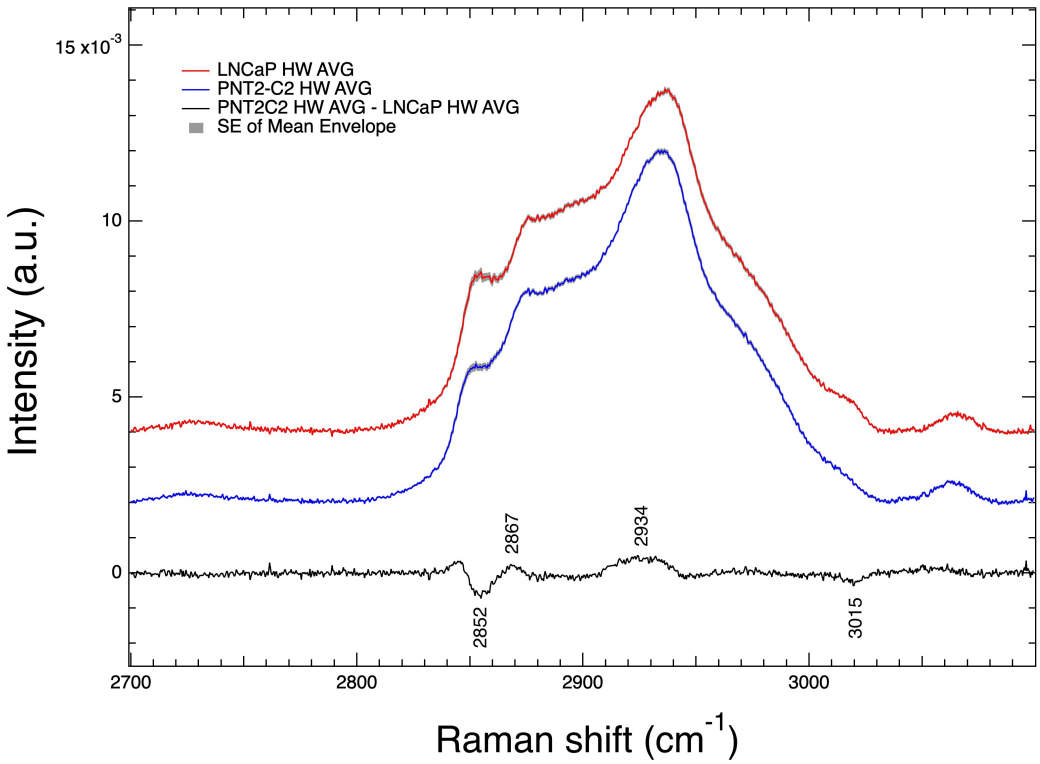


Figure 5.28: The mean subtraction between the high wavenumber region of PNT2-C2 and LNCaP has been carried out. There are four associated peaks identified that lead to separation, 2852 cm^{-1} (lipids), 2867 cm^{-1} (lipids and proteins), 2934 cm^{-1}(lipids and proteins) and 3015 cm^{-1} (lipids). This suggests that there is a difference in the ratio of total unsaturated fatty acids (3015 cm^{-1}) to fatty acids (2852 cm^{-1}). The observed difference in the four peaks, as well as the large rise in the mean subtraction profile around 2934 cm^{-1}, suggest that there is significant difference in the protein and lipid content between the cell lines.

The results obtained through analysis of the mean profile subtraction of PNT2-C2 and LNCaP in the high wavenumber region in Figure 5.28 show a variation in the intensity in both lipids and proteins. This is shown by the observed differences at the peak locations; 2852 cm^{-1} (lipids), 2867 cm^{-1} (lipids and proteins), 2934 cm^{-1}(lipids and proteins) and 3015 cm^{-1} (lipids). This suggests that there is a difference in the total unsaturated fatty acids (3015 cm^{-1}) to total fatty acids (2852 cm^{-1}). This ratio of total unsaturated fatty acids to total fatty acids has

previously been used as a marker of prostate and breast cancer [64, 140]. The bulge appearing in the mean subtraction profile around 2934 cm^{-1}, attributed to CH_2 asymmetric stretching and CH_3 symmetric stretching, suggests that there is a difference in the protein and lipid content between PNT2-C2 and LNCaP. It is thought that this is the first time a mean profile subtraction has been performed between PNT-C2 and LNCaP, to the best of the author's knowledge. These peaks will be further investigated using PCA analysis and peak intensity ratio analysis to determine the associated relative biomolecular differences between PNT2-C2 and LNCaP.

5.3.2 PNT2-C2 vs LNCaP PCA Comparison

As stated previously, a study by Taleb *et al.* (2006) did compare a normal cell line and LNCaP [142]. However, the study by Taleb *et al.* (2006) only considered the fingerprint region for analysis [142]. Key results found by Taleb *et al.* (2006) include a higher presence of B conformation DNA in PNT-1a and higher presence of A conformation DNA in LNCaP [142]. The classification tool developed by Taleb *et al.* (2006) using PCA/LDA was able to classify PNT2-C2 and LNCaP with 90% accuracy [142]. This thesis will add to the current body of knowledge through not only considering the high wavenumber region between LNCaP and a normal cell line, but also carrying out this comparison with live cells for the first time, as far as the author is aware.

PCA analysis between the PNT2-C2 and LNCaP is able to ascertain that the variance captured by the first six principal components are as follows: PC1 = 37.46%, PC2 = 9.25%, PC3 = 7.51%, PC4 = 3.97%, PC5 = 2.91% and PC6 = 2.33%. This in turn means that the first three principal components capture 54.22% and the first six capture 63.43% of the total variance of the sample. Cumulatively, 299 principal components are required to capture total variance of the sample. Three rotations of PCA space using the first three principal components are displayed in Figures 5.29, 5.30 and 5.31

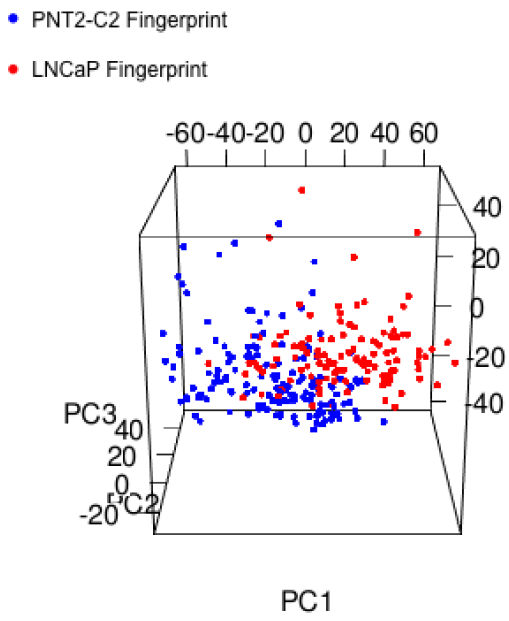


Figure 5.29: PNT2-C2 and LNCaP are viewed in PCA space using the first three principal components for the fingerprint region. At this angle of observation, although not clear, separation is occurring at an angle along PC1 and PC3.

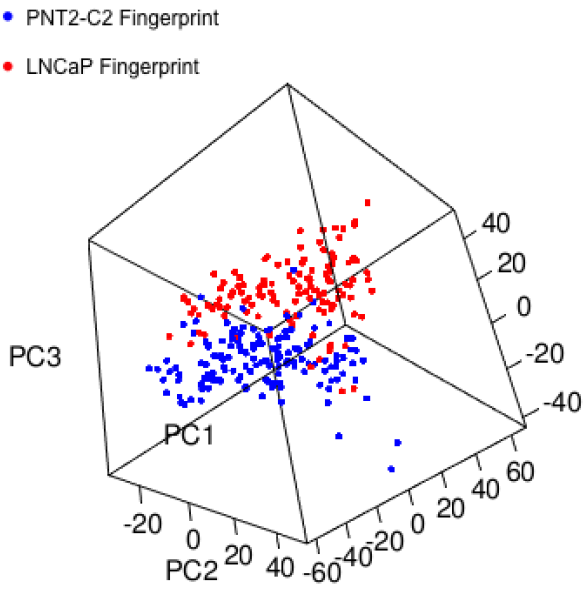


Figure 5.30: A second rotation of PCA space using the first three principal components for the fingerprint region of PNT2-C2 and LNCaP is displayed. From this angle clear separation between samples is occurring at an angle between PC1 and PC3.

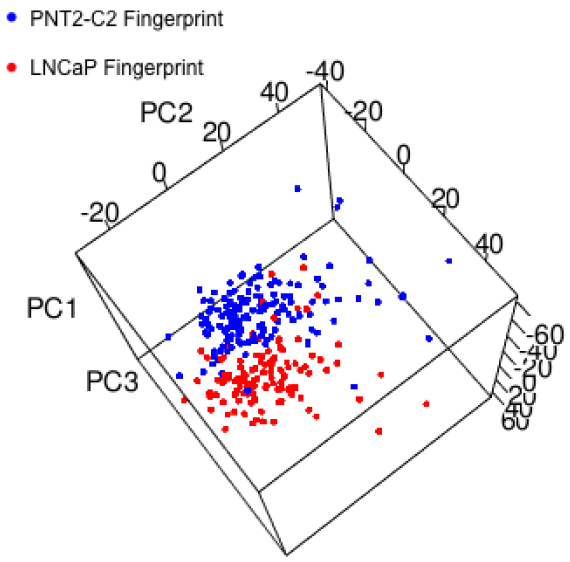


Figure 5.31: A final rotation of PCA space using the first three principal components for the fingerprint region of PNT2-C2 and LNCaP is shown. Separation between samples is occurring along PC2 and PC3.

The first observation of PCA space using the first three principal components in Figure 5.29 shows, although not clearly, separation happening at an angle between PC1 and PC3. This separation is investigated further with another rotation of PCA space, shown in Figure 5.30, showing clear separation along PC1 and PC3. The final observation of PCA space in 5.31 shows there is separation happening at an angle between PC2 and PC3. As separation is considered to occur in PC1, PC2 and PC3 their respective loadings are considered in Figures 5.32, 5.33 and 5.34. It should be noted that as separation was observed using the initial three principal components, no further principal components are considered. A future study could analyse further components but these are not explored in this thesis.

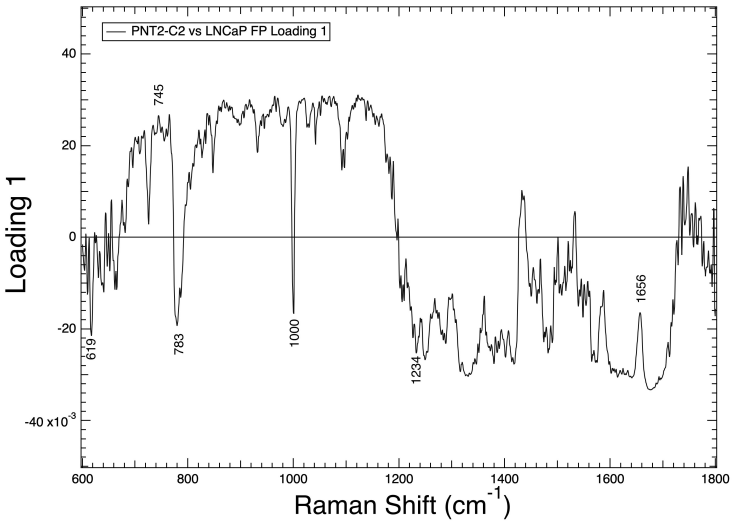


Figure 5.32: Here, Loading 1 of the fingerprint region comparison of PNT2-C2 Live vs LNCaP is displayed. Loading 1 captures 37.46% of the variance of the total sample population.Key peaks attributed to biological assignments have been highlighted: 619cm $^{-1}$ (proteins), 745cm $^{-1}$ (DNA), 783cm $^{-1}$ (DNA/RNA), 1001cm $^{-1}$ (Proteins), 1234cm $^{-1}$ (proteins and DNA/RNA), 1656cm $^{-1}$ (proteins and lipids).

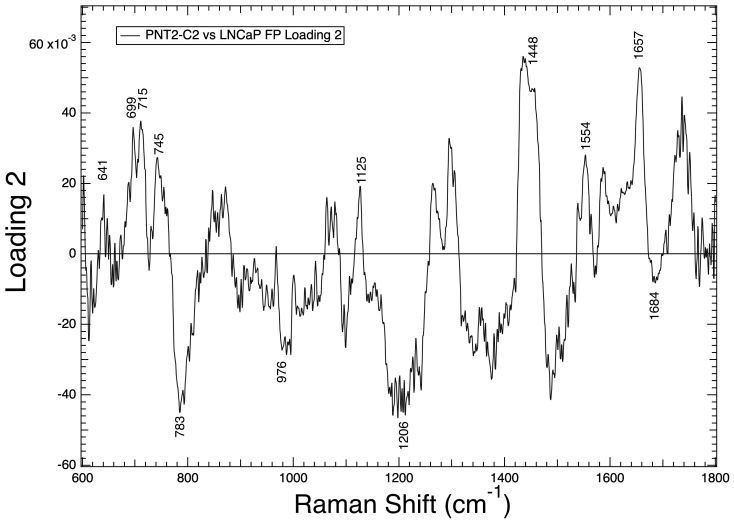


Figure 5.33: Loading 2, capturing 9.25% of the variance of the fingerprint region comparison of PNT2-C2 Live vs LNCaP, is displayed. Key peaks attributed to biological assignments have been highlighted: 641cm $^{-1}$ (proteins), 699cm $^{-1}$ (proteins), 715cm $^{-1}$ (proteins and lipids), 783cm $^{-1}$ (DNA/RNA), 976cm $^{-1}$ (lipids, proteins and DNA/RNA) , 1125cm $^{-1}$ (lipids, proteins and DNA), 1206cm $^{-1}$ (proteins and DNA/RNA), 1448cm $^{-1}$ (lipids, proteins and DNA/RNA), 1554cm $^{-1}$ (proteins), 1654cm $^{-1}$ (proteins and lipids) and 1681cm $^{-1}$ (proteins).

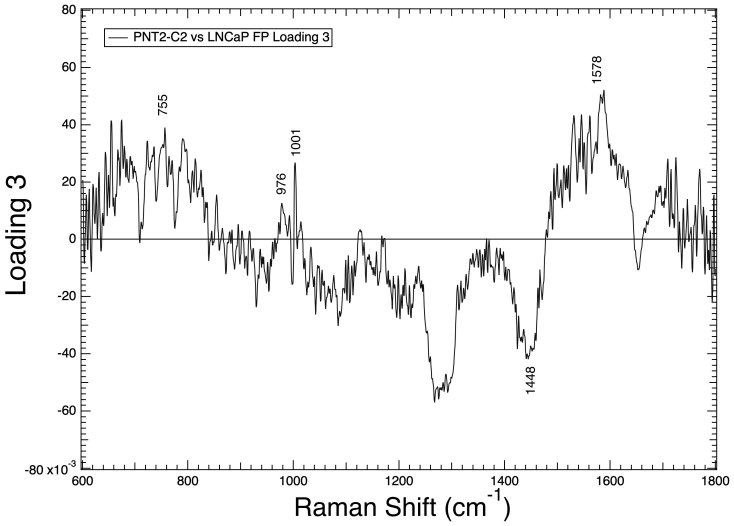


Figure 5.34: Loading 3, capturing 7.51% of the variance generated by the fingerprint region comparison of PNT2-C2 Live vs LNCaP, is displayed. Key peaks attributed to biological assignments have been highlighted: 755 cm^{-1} (DNA), 976 cm^{-1} (lipids, proteins and DNA/RNA), 1001 cm^{-1} (proteins), 1448 cm^{-1} (lipids, proteins and DNA/RNA) and 1578 cm^{-1} (DNA/RNA).

Loading 1, accounting for 37.46% of the variance between PNT2-C2 and LNCaP in the fingerprint region, is analysed in Figure 5.32. There are clear peaks that capture the variance in loading 1, such as the 619 cm^{-1} (proteins), 783 cm^{-1}(DNA/RNA) and 1001 cm^{-1} peaks (proteins). It is also noted, in general, the profile considers a lot of variance that is attributed to the background signal. The peaks identified, however, are well defined enough to be considered for further analysis at the PIR stage. The other peaks annotated such as 1234 cm^{-1} and 1656 cm^{-1} are not as well defined in loading 1 but are still able to be detected in the background signal and are thought to contribute to the variance. It can be seen through analysis of Figure 5.33 of loading 2 that there are multiple biochemical contributions. Key information can be extracted from the peaks at: 641 cm^{-1} (proteins), 699 cm^{-1} (proteins), 715 cm^{-1} (proteins and lipids), 783cm^{-1} (DNA/RNA), 976 cm^{-1} (lipids, proteins and DNA/RNA) , 1125 cm^{-1} (lipids, proteins and DNA), 1206cm^{-1} (proteins and DNA/RNA), 1448 cm^{-1} (lipids, proteins and DNA/RNA), 1554 cm^{-1} (proteins), 1654 cm^{-1} (proteins and lipids) and 1681 cm^{-1} (proteins). The peaks associated as generating variance in loading 2 will be investigated at the PIR stage to measure relative changes between

peaks. From Figure 5.34 it can be see that there is a large amount of noise present in loading 3. There were a number of peaks attributed to the separation in Figure 5.34. These peaks are the 755 cm^{-1} (DNA), 976 cm^{-1} (lipids, proteins and DNA/RNA), 1001 cm^{-1} (proteins), 1448 cm^{-1} (lipids, proteins and DNA/RNA) and 1578 cm^{-1} (DNA/RNA). The peaks 755 cm^{-1}, 1448 cm^{-1} and 1578 cm^{-1} should be interpreted carefully due to the large presence of noise in loading 3. Two clear peaks emerge at 976 cm^{-1} (lipids, proteins and DNA/RNA) and 1001 cm^{-1} (proteins). Both of these peaks have also been observed in loadings 2 and 1 respectively.

PCA analysis between the PNT2-C2 Live and LNCaP in the high wavenumber region is able to ascertain that the variance captured by the first 6 principal components (PCs) are as follows: PC1 = 46.3%, PC2 = 12.3%, PC3 = 7.4%, PC4 = 4.7%, PC5 = 2.8% and PC6 = 1.6%. The first three PCs capture 66.0% and the first six PCs capture 75.2% of the total variance of the sample. Cumulatively, 264 principal components are required to capture total variance of the sample. The results obtained from the PCA analysis of PNT2-C2 and LNCaP for the high wavenumber region are presented in Figures 5.35, 5.36 and 5.37.

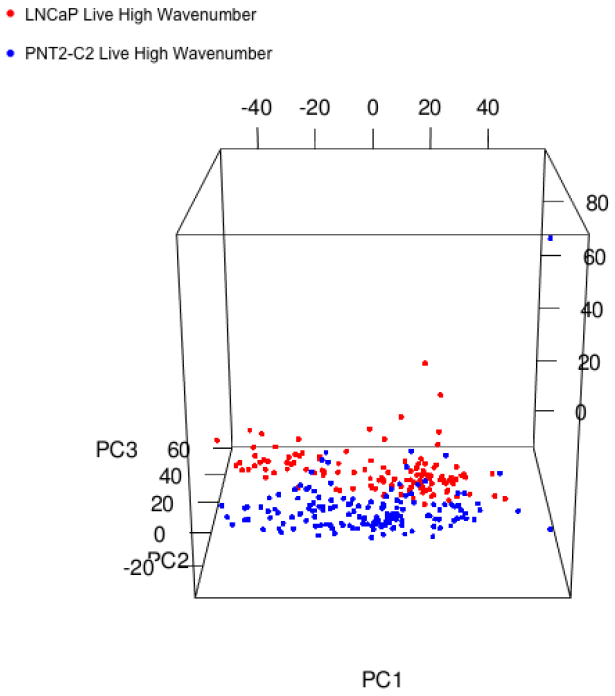


Figure 5.35: The first observation of PCA between PNT2-C2 and LNCaP in the high wavenumber region using the first three principal components is shown. This view of the 3D PCA space shows that there is no separation occurring along PC1. However, separation is occurring along PC3.

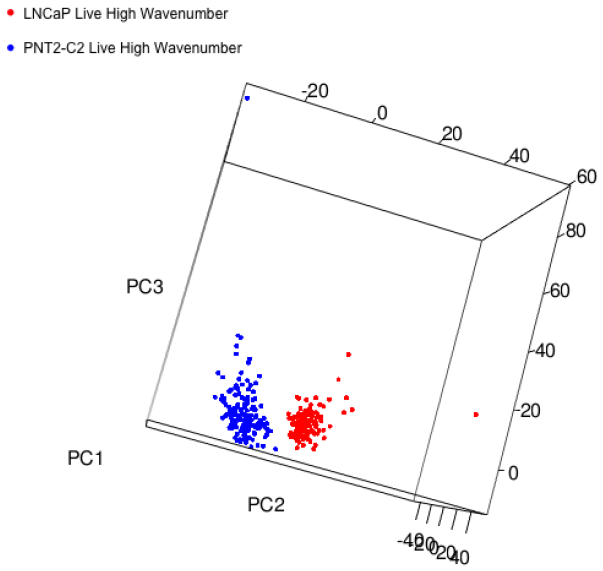


Figure 5.36: The second observation of PCA space between PNT2-C2 and LNCaP in the high wavenumber region, using the first three principal components, shows that there is 100% separation occurring at an angle along PC2 and PC3.

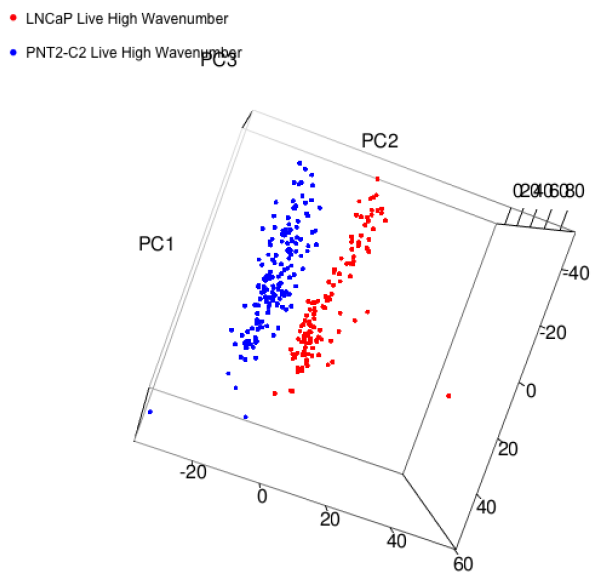


Figure 5.37: The final rotational observation of PCA space between PNT2-C2 and LNCaP in the high wavenumber region using the first three principal components. This orientation of 3D PCA space shows that there is no separation occurring along PC1. It is also seen from this view that 100% separation can be achieved along PC2.

The PCA results from the high wavenumber region using the first three principal components between PNT2-C2 and LNCaP in Figures 5.35, 5.36 and 5.37 show that no separation is occurring in PC1. It is observed in Figure 5.36 that separation is occurring at an angle between PC2 and PC3. On further investigation of the contribution of PC1 and PC2 in 5.37 it is also shown that PC2 alone can generated 100% separation of between PNT2-C2 and LNCaP. As no separation is happening in PC1 its corresponding loading is not investigated.

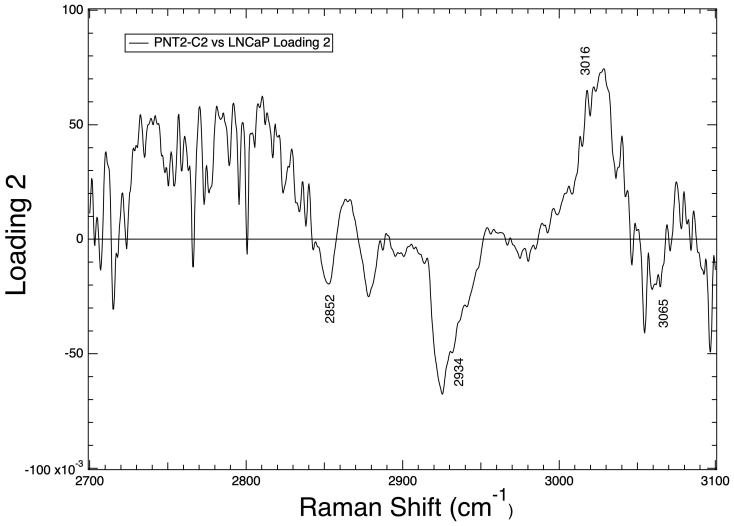


Figure 5.38: Loading 2, capturing 12.3% of the total population variance between the high wavenumber comparison of PNT2-C2 and LNCaP is displayed. The peaks: 2852 cm^{-1} (lipids), 2934 cm^{-1} (proteins, lipids), 3015 cm^{-1} (lipids) and 3065 cm^{-1} (lipids) have been identified in loading 2 as generating variance in the loading. A presence of background signal is noted in the region 2700-3100 cm^{-1}. However, the intensity of the peak at 3015 cm^{-1} exceeds the background variance.

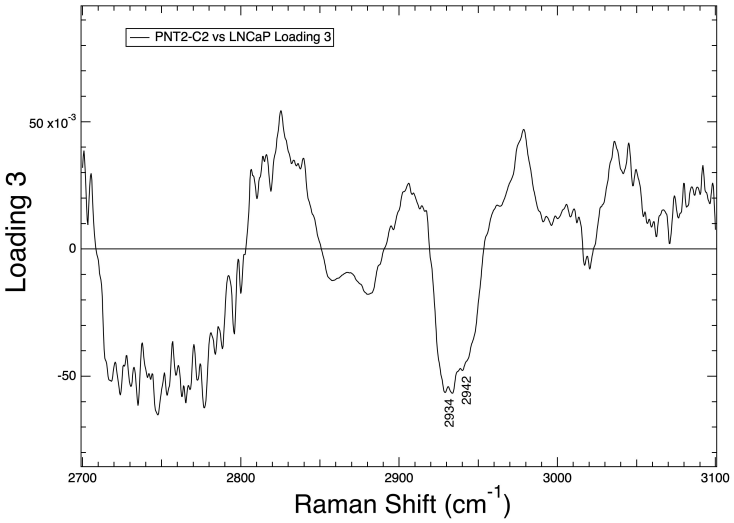


Figure 5.39: Loading 3 capturing 7.4% of the total population variance between the high wavenumber comparison of PNT2-C2 and LNCaP is displayed. The peaks 2934 cm^{-1} (proteins and lipids) and 2942 cm^{-1} (lipids and proteins) have been identified as contributing to variance in loading 3. In loading 3 there is a large influence from the background profiles of the PNT2-C2 and LNCaP high wavenumber spectra. This is particularly evident as there are large regions of intensity where no biochemical signatures could be found.

As shown in Figures 5.38 and 5.39, there is a considerable amount of background variance in regions that are not attributed to biochemical signatures, in both loading 2 and loading 3. In loading 2, the peaks that were identified as capturing variance were: 2852 cm^{-1} (lipids), 2934 cm^{-1} (proteins, lipids), 3015 cm^{-1} (lipids) and 3065 cm^{-1} (lipids). In loading 3 there were only two peaks that could be attributed to biomolecular assignments, the 2934 cm^{-1} (proteins and lipids) 2942 cm^{-1} (lipids and proteins). The contribution of these peaks to the biomolecular differences between PNT2-C2 and LNCaP will be investigated further through the use of PIR analysis.

Finally, a PCA-LDA leave-one-out cross-validation was used to determine the predictive classification of PNT2-C2 and LNCaP cells. The number of principal components included was based on using Kaiser's criterion, which consists of using only principal components with variance greater than 1 percent. For PNT2-C2 vs LNCaP, the PCA-LDA classification tool correctly classified 99.6% cells in the fingerprint region using 9 PCs and 99.3% of cells in the high wavenumber region using 7 PCs. This demonstrates a high classification accuracy between PNT2-C2 and LNCaP.

5.3.3 Peak intensity ratio analysis PNT2-C2 vs LNCaP

As previously discussed the study by Potcoava *et al.* (2014) has analysed the cell line LNCaP using peak intensity (PIR) analysis in a limited capacity [64]. The study by Potcoava *et al.* (2014) considered the lipid profile of LNCaP and utilised specific ratios to characterise the cell line in terms of lipid structure. The main purpose of the study carried out by Potcoava *et al.* (2014) was to investigate the intracellular lipid droplets and their saturation in cells during hormone response therapy, and the process of lipogensis [64]. The study by Potcoava *et al.* (2014), however, did not contain a normal cell line to compare ratios against [64]. This thesis will expand on the work carried out by Potcoava *et al.* (2014) through inclusion of a normal comparison and will also consider the DNA/RNA and protein assignments in the fingerprint and a high wavenumber region for use as potential biomarkers.

The complete set of results used to calculate the separation outside of the propagated SE uncertainty, relating to each PIR compared between PNT2-C2 and LNCaP for the fingerprint and high wavenumber region, is shown in Table C.3 in the Appendix. The results, as seen in table C.3, identify three peaks that achieve over 90% separation in the sample. These peaks and their corresponding degree of separation are the 976 cm^{-1} peak (98%), 1554 cm^{-1} peak (95%) and the 3015 cm^{-1} peak (91%). A further five peaks are able to achieve separation of 80% and above, the 618 peak cm^{-1} peak (84%), 745 cm^{-1} peak (87%), 781 cm^{-1} peak (84%), 1171 cm^{-1} peak (82%) and the 1396 cm^{-1} peak (84%).

The 976 cm^{-1} peak achieved the strongest separation, of 98%. This result was expected. As noted in 5.27 there was a major change in shape between the peaks 936 cm^{-1} and 1001 cm^{-1}, encompassing the 976 cm^{-1} peak. Furthermore, both loading 2 and loading 3 in Figures 5.33 and 5.34, the 976 cm^{-1} peak was identified as contributing to the variance. The specific biomolecular assignment of 976 cm^{-1} peak is lipids, phosphate monoester, groups of phosphorylated proteins and cellular nucleic acids [128]. The PIR panel for the 976 cm^{-1} peak is shown in Figure 5.40.

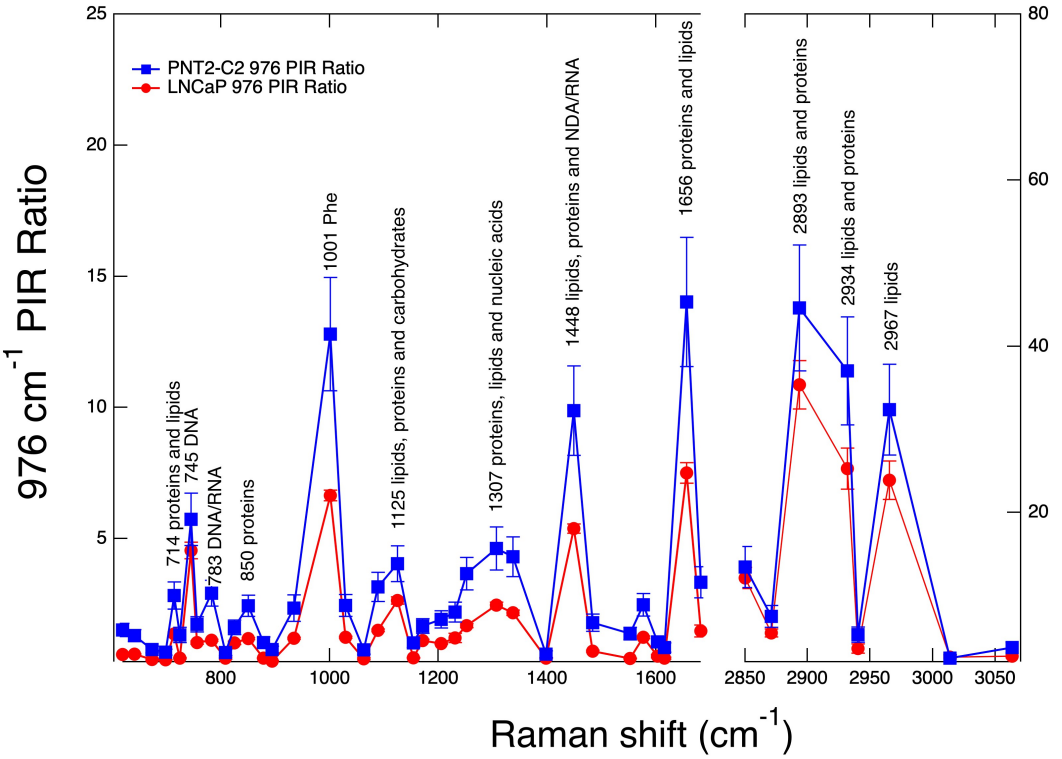


Figure 5.40: The PIR panel for the 976 cm^{-1} peak intensity ratio demonstrates statistically significant separation outside of uncertainty across the fingerprint region and high wavenumber region of PNT2-C2 and LNCaP, achieving 98% total separation. The 976 cm^{-1} band is representative of phosphate monoester, groups of phosphorylated proteins and cellular nucleic acids [128]. It can be seen that PIR ratios using the 976 cm^{-1} peak in LNCaP are higher than those generated for PNT2-C2 across the fingerprint and high wavenumber spectrum. This observation indicates that there is a lower relative presence of phosphate monoester, groups of phosphorylated proteins and cellular nucleic acids in LNCaP. It should be noted that due to the multiple biomolecular contributions represented by the 976 cm^{-1} peak, it is not possible to determine which biomolecular contribution is generating the most variance.

The PIR panel in Figure 5.40 shows statistically significant separation outside of uncertainty across the fingerprint region and high wavenumber region of PNT2-C2 and LNCaP. It can be seen that PIR ratios, using the 976 cm^{-1} peak in LNCaP, are higher than those generated for PNT2-C2 across the fingerprint and high wavenumber spectrum. This observation indicates that there is a lower relative presence of phosphate monoester, groups of phosphorylated proteins and cellular nucleic acids in LNCaP. It should be noted that due to the multiple biomolecular contributions represented by the 976 cm^{-1} peak, it is not possible to determine

which biomolecular contribution is generating the most variance. Although tentatively, a possible explanation for these results may be that there are groups of phosphorylated proteins present in LNCaP that are responsible for oncogenesis in the cell line. Although not a study considering Raman spectroscopy, the cell lines PC-3 and LNCaP were analysed in a study by Endoh *et al.* (2012) using GST-pull down analysis, a method of using protein markers to isolate a target protein from the rest the proteins present in a sample [278]. The study by Endoh *et al.* (2012) found multiple cancer-specific phosphorylated proteins in both LNCaP and PC3 [278]. It is suggested that the strong degree of separation observed at 976 cm^{-1} peak, as well as the difference in the mean profile subtraction from 936 cm^{-1} to 1001 cm^{-1}, may provide an initial insight of phosphorylated proteins driving oncogenesis in LNCaP. The study carried out between PNT2-1a and LNCaP by Taleb *et al.* (2006) did not note this observation [142]. However, observation of the loadings from PCA in the study by Taleb *et al.* (2006) does capture significant variance in the region region 936 cm^{-1} to 1001 cm^{-1} , although no discussion is made pertaining to it [142].

Achieving high discrimination between PNT2-C2 and LNCaP of 98%, the 1554 cm^{-1} peak intensity ratio is representative of tryptophan and Amide II. The relatively high degree of statistical separation observed at 1554 cm^{-1} indicates that there may be changes to the secondary protein structure between PNT2-C2 and LNCaP, as well as differences in tryptophan. The 1554 cm^{-1} peak was not identified in the mean profile subtraction as being a key peak of interest, but this is not a robust test for determining statistical significance. Furthermore, the 1554 cm^{-1} peak was identified at the PCA stage in loading 2, Figure 5.33. The PIRs corresponding to the 1554 cm^{-1} peak are investigated further through the use of the PIR panel in Figure 5.41.

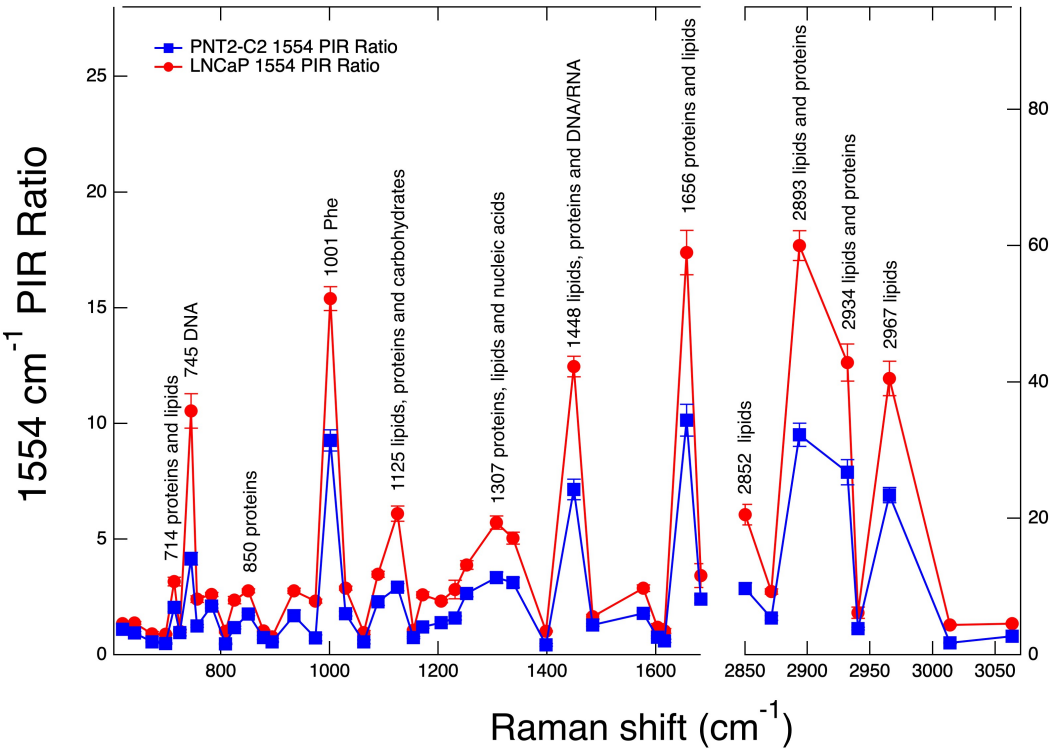


Figure 5.41: The PIR panel for PIRs of PNT2-C2 and LNCaP is shown for the 1554 cm^{-1} peak. It can be seen that there is statistically significant separation across the fingerprint region and high wavenumber region, achieving 93% total separation. The 1554 cm^{-1} band is representative of tryptophan and Amide II. It can be seen that PIRs generated using the 1554 cm^{-1} peak are relatively higher across the fingerprint and high wavenumber spectrum for LNCaP. This observation indicates that there is a lower relative presence of amide II and tryptophan in LNCaP.

The PIR panel for the 1554 cm^{-1} peak intensity ratio, as shown in Figure 5.41, demonstrates that there is a relative decrease of tryptophan and amide II in LNCaP when compared to PNT2-C2. This is observed throughout the PIR profile of the 1554 cm^{-1} peak in LNCaP, showing a relative increase in intensities of all peak intensity ratios in both the fingerprint and high wavenumber region across DNA, proteins, lipids and nucleic acids. Additionally, this could further signify secondary protein structure changes between PNT2-C2 and LNCaP. The reduced relative intensity of tryptophan is considered an important biomarker; tryptophan an important amino acid that plays a role in apoptosis, during disease and inflammatory response [274, 275]. It is for this reason that the 1554 cm^{-1} amide II and tryptophan peak is determined a robust biomarker to distinguish

between PNT2-C2 and LNCaP.

The 745 cm^{-1} peak demonstrates 87% separation outside of the propagated SE uncertainty relating to each PIR between PNT2-C2 and LNCaP. Whilst this is below the self determined cut off for investigation at 90%, strong evidence from the mean profile subtraction in Figure 5.27, as well as contribution to separation at the PCA stage in loading 2, Figure 5.33, highlight that key variance is occurring at the 745 cm^{-1} peak. The 745 cm^{-1} peak is attributed to the thymine ring breathing mode. Although achieving a lesser 84% degree of separation, the peak 783 cm^{-1} is also generated by thymine ring breathing as well as contributions from Uracil, cytosine, overlapping with an O-P-O stretch in DNA. The 783 cm^{-1} peak is also one of the most dominant peaks capturing variance in Figure 5.33. However, the peak associated with the 783 cm^{-1} peak in loading 2 also has a wide line width, indicating that other peaks may be contributing to the separation. The O-P-O stretch in DNA is also found at the locations 810 cm^{-1} peak (O-P-O stretch in DNA/RNA) and 825 cm^{-1} peak (DNA O-P-O asymmetric stretch, tyrosine and proline). It has been previously discussed that an increased relative intensity of the O-P-O stretch is an indicator of looser A state DNA [279, 280]. To investigate the peaks 745 cm^{-1}, 783 cm^{-1} 810 cm^{-1} and 823 cm^{-1} further they have been analysed using the 1448 cm^{-1} PIR.

Table: PIRs examining relative thymine and O-P-O intensity				
Cell line	I745/1448	I783/1448	I810/1448	I825/1448
PNT2-C2	0.58±0.04	0.29±0.02	0.066±0.004	0.16±0.01
LNCaP	0.85±0.06	0.21±0.01	0.081±0.006	0.19±0.01

Table 5.3: The peak intensity ratio I745/1448 is able to statistically differentiate between PNT2-C2 and LNCaP. The ratio I745/1448 determines a much higher relative intensity of thymine (DNA) to CH_2 deformation in lipids and proteins in LNCaP than in PNT-C2. The ratio I785/1448 shows that there is a higher relative intensity of uracil, cytosine and thymine ring breathing modes/O-P-O stretch in DNA relative to CH_2 deformation in lipids and proteins, in PNT2-C2 than LNCaP. As the I785/1448 ratio is the product of many contributions, it is difficult to assess what is the cause of this relationship. It is possible to assess the contribution of the O-P-O stretch further by considering the 810 cm^{-1} peak (O-P-O stretch in DNA/RNA) and 825 cm^{-1} peak (DNA O-P-O asymmetric stretch, tyrosine and proline). It can be seen that the band associated with a pure DNA O-P-O stretch, 810 cm^{-1} peak, has a higher relative ratio to the 1448 cm^{-1} peak (CH_2 deformation in lipids and proteins) in LNCaP. This relationship is also observed using the ratio 825/1448. This would suggest that there was a higher presence of A conformation DNA in LNCaP [279, 280].

As seen in Table 5.3, the peak intensity ratio I745/1448 is able to statistically differentiate between PNT2-C2 and LNCaP outside of uncertainty. The ratio I745/1448 determines a much higher relative intensity of thymine (DNA) to CH_2 deformation in lipids and proteins in LNCaP than in PNT-C2. This is expected, as multiple Raman studies have discussed an increased relative DNA content in metastatic cells compared to the normal [281]. Specifically, an increase in intensity of the thymine peak has been described as a marker for cancerous cells when compared to the normal [253, 254]. The ratio I785/1448 shows that there is a higher relative intensity of uracil, cytosine and thymine ring breathing modes/O-P-O stretch in DNA relative to CH_2 deformation in lipids and proteins, in PNT2-C2 than LNCaP. As the I785/1448 ratio is the product of many contributions it is difficult to assess what is the cause of this relationship. It is possible to asses the contribution of the O-P-O stretch further by considering the 810 cm^{-1} peak (O-P-O stretch in DNA/RNA) and 825 cm^{-1} peak (DNA O-P-O asymmetric stretch, tyrosine and proline). It can be seen that the band associated with a pure DNA O-P-O stretch, 810 cm^{-1} peak, has a higher relative ratio to the 1448 cm^{-1} peak (CH_2 deformation in lipids and proteins) in LNCaP. This relationship is also

observed using the ratio 825/1448. This would suggest that there was a higher presence of A conformation DNA in LNCaP [279, 280].

The 3015 cm^{-1} peak achieved 91% separation between PNT2-C2 and LNCaP at the PIR stage. This peak is biochemically representative of an unsaturated =CH stretch in lipids. The 3015 cm^{-1} peak was also identified in the mean profile subtraction in Figure 5.28. Furthermore, the 3015 cm^{-1} peak was also found to be a major source of variance in loading 2 at the PCA stage, Figure 5.38. To consider the peak intensity ratio relationships to the 3015 cm^{-1} peak more carefully, the associated PIR panel is examined in Figure 5.42.

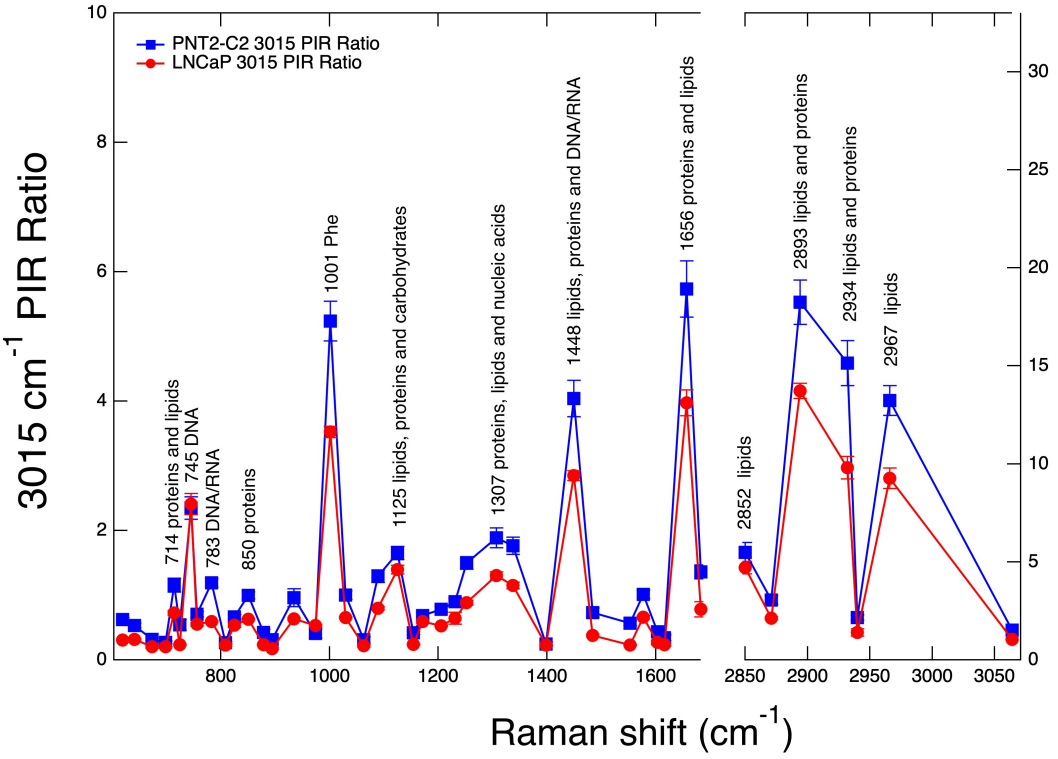


Figure 5.42: The PIR panel for the 3015 cm^{-1} peak intensity ratio demonstrates statistically significant separation across the fingerprint region and high wavenumber region, achieving 91% total separation. The 3015 cm^{-1} peak is attributed to an unsaturated =CH stretch in lipids. However, the 3015 cm^{-1} peak intensity ratio is not able to separate the whole spectrum. It can be seen that there is no statistical separation at several peak locations, particularly 745 cm^{-1} (DNA) and 2852 cm^{-1} (lipids). This shows that whilst generally there is an overall increase in relative intensity of the 3015 cm^{-1} peak in LNCaP, there is also an up-regulation of other peaks in LNCaP.

Figure 5.42 shows that there is higher general relative intensity of the 3015 cm^{-1} peak in LNCaP. This in turn signifies a higher relative intensity of unsaturated =CH stretch in lipids in LNCaP. However, there were two notable instances in which statistical separation was not observed. There was no separation between the peak intensity ratios I745/3015 and I2851/3015. The 745 cm^{-1} thymine peak was previously shown to have an increased relative intensity in LNCaP with respect to the 1448 cm^{-1} peak, in previous analysis in this Section. This would suggest that whilst there is no statistical difference in the ratio of thymine to the unsaturated =CH stretching in lipids, this is due to an up-regulation of the 745 cm^{-1} peak in LNCaP as well. It is also important to consider the 2852 cm^{-1} peak which represents CH_2 symmetric stretching in lipids. The 2852 cm^{-1} peak achieved 75% separation at the PIR stage. It is hypothesised that the resulting non-separation is also due to a higher relative intensity of the 2852 cm^{-1} peak. It is thought that the changes in lipid content are caused by the need for the growth of new cell membranes during cancerous growth and are brought about by metabolic changes in the cell [282].

It is important to note that the 3015 cm^{-1} peak intensity and 2852 cm^{-1} peak intensity have been previously identified by others in the field as a biomarker for malignancy, when used as a PIR [64, 140]. The study by Nieva *et al.* (2012) proposed the ratio of total unsaturated fatty acids to total fatty acids (TUFA/TFA) (I3015 cm^{-1} peak/I2851 cm^{-1} peak) as a molecular biomarker of malignancy in breast cancer tissue [140]. Nieva *et al.* (2012) used Raman spectroscopy to measure a metastatic breast cancer tissue cell line as well as a benign disease tissue cell line. They found that the TUFA/TFA band, itself, was not sensitive enough to differentiate benign disease tissue from malignant breast cancer tissue, with the inclusion of errors [140]. It should be noted that the study by Nieva *et al.* (2012) is also on biological tissue and not on a single cell level. The main limitation to the study carried out by Nieva *et al.* (2012) is that it is limited to only the high wavenumber region [140]. The use of TUFA/TFA as a biomarker will be explored further in Section 5.7

5.4 DU145

There has already been a study into the comparison of PNT2 and DU145. The study completed by Corsetti *et al.* (2018) has carried out detailed PCA analysis between these two cell lines [224]. This thesis varies in several ways to the techniques carried out by Corsetti *et al.* (2018). There is a change to the normal comparison, although subtle. The cell line PNT2 is not the same as the cell line PNT2-C2 used in this study. In terms of hierarchy PNT2 is the parent cell and PNT2-C2 a subculture of it [41, 94]. Both cell lines still represent well differentiated normal prostate cells and can be used in counter comparison.

Corsetti *et al.* (2018) studied the Raman profiles of 69 PNT2 and 74 DU145 cells using a 785nm diode laser. Their methodology varies by the use of the dried cells in analysis as well as a different laser wavelength. The use of dried cells in comparisons has been discussed extensively in chapter 4. This raises a lot of concerns regarding the degradation of DNA and proteins in the air-drying process. The use of dried cells also reinforces the previous comments made about the continuation of dried cell studies in the field of cancer research. The main barrier for counter comparison in this study, however, is the use of the 785nm laser. In terms of analysis methods Corsetti *et al.* (2018) did carry out a mean subtraction of the spectra between PNT2 and DU145 as well as PCA/LDA analysis. However, they only did this for the region 400-1300 cm^{-1}. There was some further consideration around the 1400-1800 cm^{-1} and 2700-3100 cm^{-1} regions but no mean profile subtraction was produced and the analysis remained only visual [224]. Corsetti *et al.* (2018) made several biochemical conclusions drawn from a mean subtraction and PCA analysis. A presence of increased intensity around the 1448 cm^{-1} peak in DU145 indicated that the cell line had an increased protein content compared to PNT2 [224]. Through analysis of the mean subtraction between both cell lines, it was determined that DU145 has a much higher DNA content due to an increased intensity at the 723 cm^{-1} DNA peak. Through PCA analysis it was found DU145 cells contain high concentrations of DNA, phenylalanine and Amide III [224]. The cell line PNT2 was found to have increased concentration in L-arginine and B conformation DNA [224]. PCA/LDA was carried out in the

region 400-1300 cm^{-1} but not in the other two regions. The PCA/LDA comparison in the range 400-1300 cm^{-1} was able to achieve separation between samples with a sensitivity of 95% and specificity 88%.

DU145 has appeared in further PCA/LDA analysis in the literature. Crow *et al.* (2005) used DU145 in a study to differentiate between androgen-sensitive and androgen independent prostate cancers. In their study four cell lines: two androgen-sensitive cell lines (DU145, PC3) and two androgen-sensitive cell lines (LNCaP, PCa2b) were compared using PCA/LDA analysis [63]. The experimental set up varies to the one used in this thesis. The laser used by Crow *et al.* (2005) consists of a near-infrared laser operating at 830 nm. Furthermore Crow *et al.* (2005) measured cells in a "moist" environment within a 30 minute window after the cells were removed from the medium. There are obvious downsides to this method, as while the cells are drying they are undergoing autolysis. This will not only affect the biochemical constituents of the cells but cause a change in the morphology as a function of time. Crow *et al.* (2005) was able to use PCA/LDA diagnostic algorithm to separate the cell populations with a sensitivity of 98% and a specificity of 99% [63].

In consideration of the studies by Corsetti *et al.* (2018) and Crow *et al.* (2005) there are still large areas to where this thesis will provide additional information. This study expands on the work already carried out, by exploring peak intensity ratio analysis, which was not carried out by Corsetti *et al.* (2018) or Crow *et al.* (2005). Furthermore, after showing that the drying of cell lines can affect the Raman spectra of cells, this study will also be able to compare results and any key differences in terms of protein, lipid and DNA structure, with both studies. These additions and comparisons will aid further in understanding the complex environment of androgen-independent prostate cancers. This is done through analysis of the mean subtraction, PCA and PIRs.

5.4.1 PNT2-C2 vs DU145 mean profile subtraction

The normalised Raman signal for PNT2-C2 and DU145 is displayed for the fingerprint region in Figure 5.43 and the high wavenumber region is shown in Figure 5.44. The normalisation has been carried out with respect to the total spectrum area. They have both undergone a linear baseline as well as being offset for an arbitrary amount. Neither the PNT2-C2 cell line nor DU145 have undergone smoothing. The mean subtraction is shown on the same graph as the Raman profiles for the fingerprint region in Figure 5.43 and the high wavenumber region 5.44.

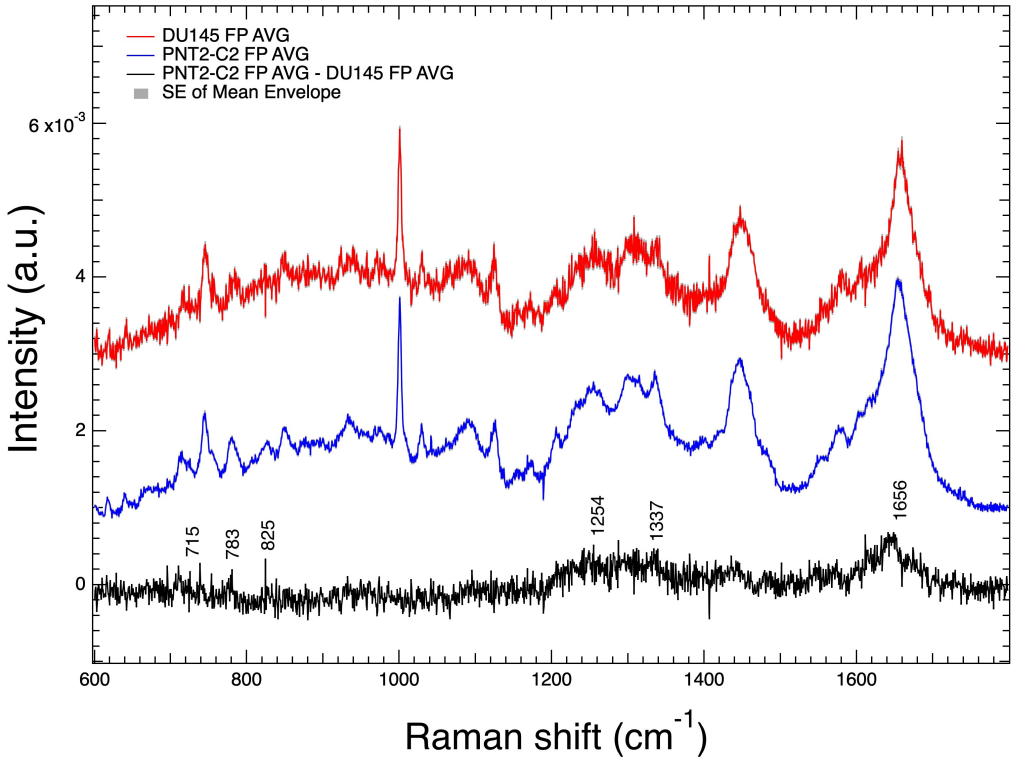


Figure 5.43: A mean subtraction is examined as the product of the fingerprint PNT2-C2 average profile minus the average fingerprint profile of DU145. The standard error of the mean envelope is also present on both of these profiles. The mean profile subtraction has identified several peaks of interest: 715 cm^{-1} (proteins and lipids), 783 cm^{-1} (DNA/RNA), 825 cm^{-1} (proteins and DNA/RNA), 1254 cm^{-1} (proteins), 1337 cm^{-1} (proteins and DNA/RNA) and 1656 cm^{-1} (proteins and lipids). It is however noted that these peaks are tentatively placed with a cursor and have not undergone any form of peak fitting. The mean subtraction allows for a visualisation of key differences only.

It can be seen visually in Figure 5.43 that there is a lower signal-to-noise ratio

in the Raman signal for DU145 in the fingerprint region. However, as the data is converged this is not thought to be an issue. Furthermore, this thesis is only measuring relative intensities between peaks, not absolute peak intensity.

The mean profile subtraction of PNT2-C2 and DU145 highlighted several peaks of interest in the fingerprint region. Key differences in the fingerprint region are found at 715 cm^{-1} (proteins and lipids), 783 cm^{-1} (DNA/RNA), 825 cm^{-1} (proteins and DNA/RNA), 1254 cm^{-1} (proteins), 1337 cm^{-1} (proteins and DNA/RNA) and 1656 cm^{-1} (proteins and lipids). It also observed that noise from DU145 features in the subtraction. This was expected, due to the noise observed in the DU145 fingerprint profile, Figure 5.43. The 715 cm^{-1} peak is attributed to amino acid methionine, $CN^{+}(CH_3)_3$ stretch in lipids, C-N membrane phospholipid head stretch and adenine nucleotide peak. The 783 cm^{-1} band consists of contributions from uracil, cytosine, thymine ring breathing modes, nucleic acids and thymine overlapping with O-P-O stretch in DNA which could imply differences in conformational state [224]. Further variance in spectra is observed at 825 cm^{-1}, indicating a potential change in the relative intensity of proteins and RNA. The sharp line in the profile at 825 cm^{-1} arises in the DU145 profile. This is not a peak and can not be fitted. There is no explanation to be found as to why this occurs. A similar line is also found in DU145 at 1400 cm^{-1} and in PNT2-C2 at 1200 cm^{-1}. At this stage it is not clear enough to draw a detailed biochemical comparison of the peak 825 cm^{-1}. The study by Corsetti *et al.* (2018) noted a large difference in the 1001 cm^{-1} peak when normalised profiles were subtracted. This is not observed in Figure 5.43. At this stage no conclusions can be made but it highlights the need for consideration in further analysis. There is a change in profile shape around the bands 1254 cm^{-1} and 1337 cm^{-1}. This region, although it has contributions from proteins and lipids, is dominated by the amide III protein. The relationship of these peaks and amide III will be probed using PCA and PIRs to provide more conclusive analysis. There is a considerable difference in the region surrounding the 1656 cm^{-1} peak. The 1656 cm^{-1} peak consists of contributions from amide I and α-helix conformations which will also be further investigated through a combination of PCA and PIR analysis.

The high wavenumber region mean subtraction for PNT2-C2 and DU145 has also been carried out and is shown in Figure 5.44.

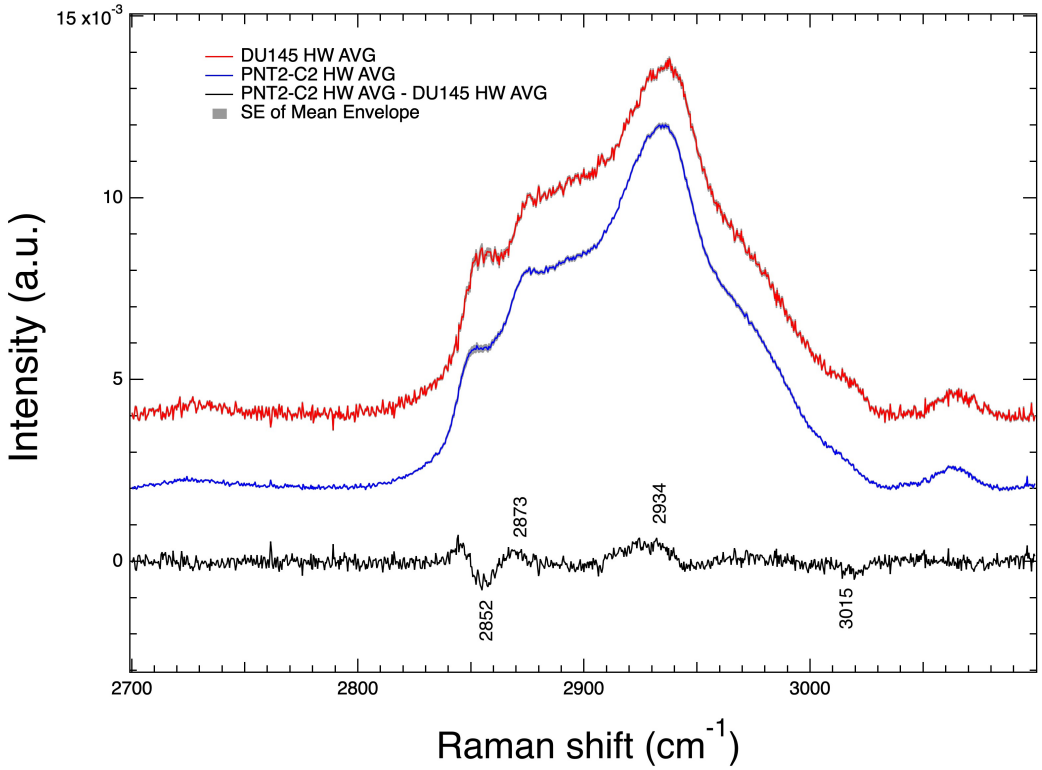


Figure 5.44: The mean subtraction between the high wavenumber region of PNT2-C2 and DU145 is carried out. The two original profiles have been normalised and the standard error of the mean envelope has been plotted to view the associated error. The presence of noise in the mean subtraction profile is generated by the DU145 high wavenumber region. There were four peaks identified that lead to separation; 2852 cm^{-1} (lipids), 2873 cm^{-1} (lipids and proteins), 2934 cm^{-1} (proteins and lipids), 3015 cm^{-1} (lipids). This suggests that there is a difference in the ratio of total unsaturated fatty acids (3015 cm^{-1}) to fatty acids (2852 cm^{-1}). The observed difference in the four peaks, as well as the large rise in the mean subtraction profile around 2934 cm^{-1}, suggests that there is significant difference in the protein and lipid content between the cell lines.

Similarly to the fingerprint region, the high wavenumber region in Figure 5.44 exhibits lower signal-to-noise ratio in the Raman signal for DU145. However, as the data is converged this is not thought to be an issue. Furthermore, as stated previously, this thesis is only measuring relative intensity changes between peaks, not absolute peak intensity. The decrease in intensity around the 2852 cm^{-1} peak (lipids) indicates a change in the CH_2 symmetric stretch in lipids, inferring that peak intensity ratio analysis may observe variation of the ratio of total

fatty acids (TFA) to total unsaturated fatty acids (TUFA) [64, 140]. This is done in conjunction with the 3015 cm^{-1} (lipids) peak, which represents unsaturated fatty acids (unsaturated =CH stretch in lipids). The ratio TUFA/TFA has been used as biomarker previously in Raman studies as a tool for the identification of malignant cells [64, 140]. At this stage it cannot be determined whether the observed difference in the mean subtraction will result in a statistically significant biomarker being found. The further variation observed at the locations 2873 cm^{-1} and 2934 cm^{-1} both indicate a difference in the CH_2 asymmetric stretch in lipids and proteins. The 2934 cm^{-1} band also represents a CH_3 symmetric stretch in lipids. All bands identified at this stage will be considered at the PCA and PIR stage.

Corsetti *et al.* (2018) made several biochemical conclusions drawn from a mean subtraction and PCA analysis. A presence of increased intensity around the 1448 cm^{-1} peak in DU145 indicated that the cell line had an increased protein content compared to PNT2 [224].

5.4.2 PNT2-C2 vs DU145 PCA analysis

Prior to undergoing PCA analysis, the spectra are subject to the necessary pre-processing detailed in methodology Section 3.1.7. Figure 5.43 shows the average spectrum of PNT2-C2 and DU145 prior to being smoothed for PCA analysis for the fingerprint region. Similarly, Figure 5.44 gives the two average spectrum of the high wavenumber region prior to smoothing for analysis. For this comparison both the fingerprint region and high wavenumber region are treated separately.

PCA analysis between the PNT2-C2 Live and DU145 fingerprint region is able to ascertain that the variance captured by the first 6 principal components are as follows: PC1 = 28.6%, PC2 = 9.9%, PC3 = 6.5%, PC4 = 4.3%, PC5 = 2.1% and PC6 = 2%. This in turn means that the first three principal components capture 45% and the first six capture 53.3% of the total variance of the sample. Cumulatively, 278 principal components are required to capture 100 % variance of the sample. Three rotations of PCA space using the first three principal components are displayed

in Figures 5.45, 5.46 and 5.47.

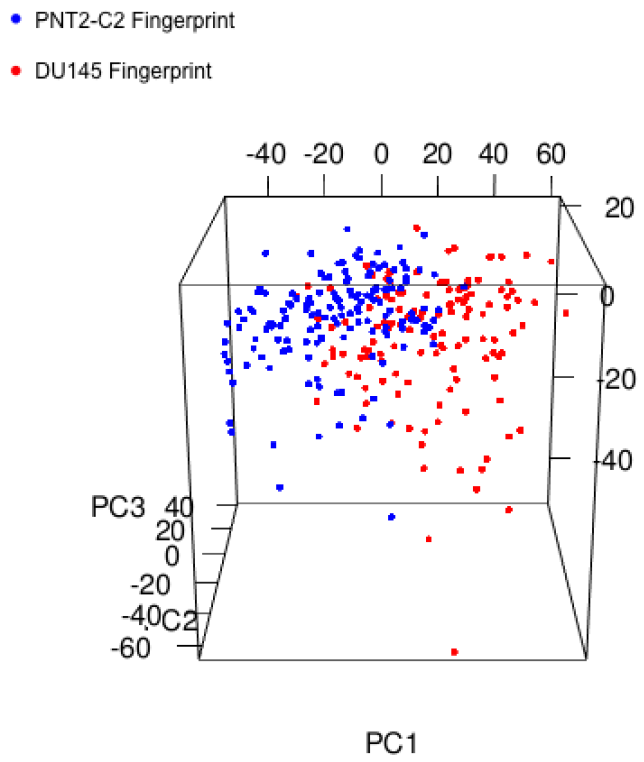


Figure 5.45: A 3D PCA comparison of the fingerprint region of PNT2-C2 and DU145 is showing separation occurring at an angle between PC1 and PC3. Although separation is observed there is an overlap of spectra when using these PCs.

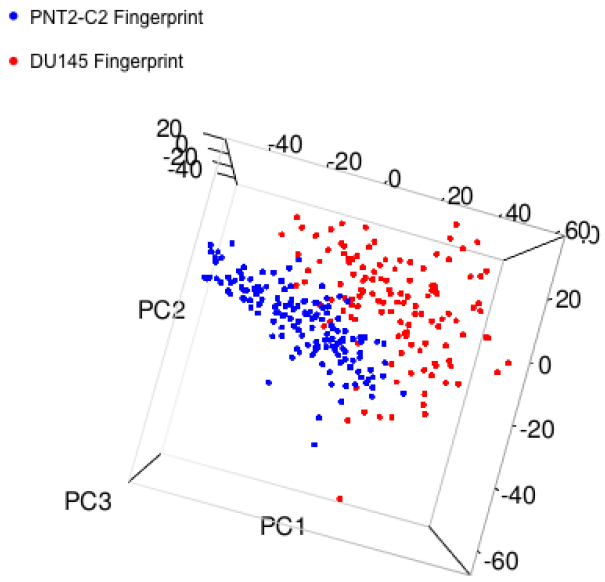


Figure 5.46: Another rotation of PNT2-C2 and DU145 in 3D space using the first three principal components. In this rotation separation is occurring at an angle along PC1 and PC2.

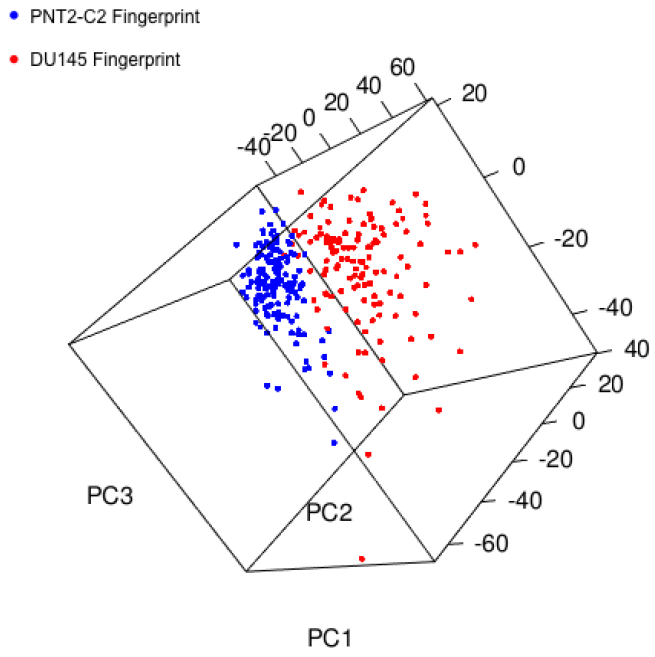


Figure 5.47: A final rotation of PNT2-C2 and DU145 in 3D space using the first three principal components. In this rotation there is separation occurring at an angle along PC2 and PC3.

Figures 5.45, 5.46 and 5.47 detail the PCA separation between PNT2-C2 and

DU145. Collectively, they show that the separation occurring between PNT2-C2 and DU145 is generated by contributions from PC1, PC2 and PC3. As there is separation occurring at in PC1, PC2 and PC3, the corresponding loadings are now investigated to determine which biomolecular contributions are responsible for variance between samples. The first three PCs are plotted in the form of loadings in Figures 5.48, 5.49 and 5.50.

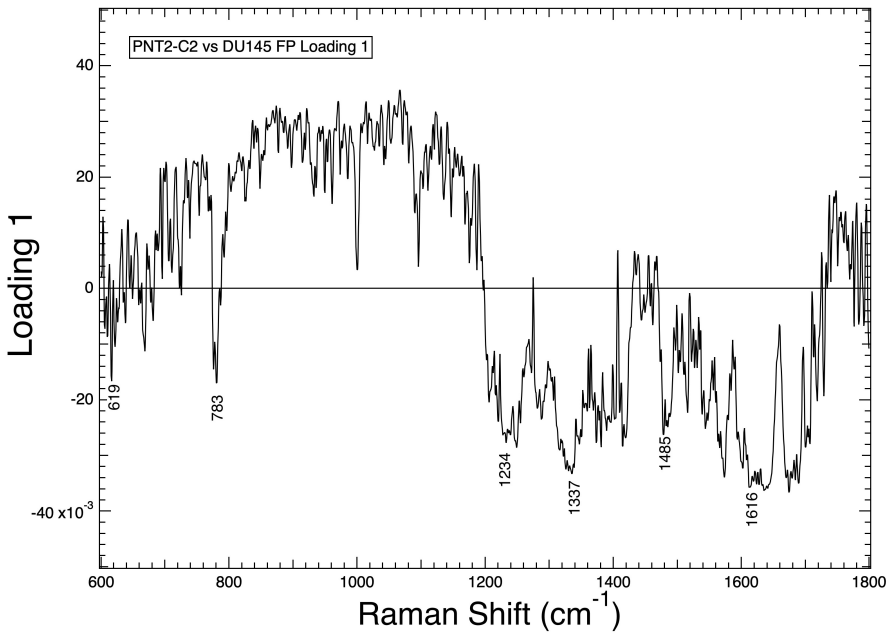


Figure 5.48: Loading 1 representative of PC1 and accounting for 28.6% of total variance shows multiple peak contributions in the fingerprint region between PNT2-C2 Live vs DU145. Key peaks attributed to biological assignments have been highlighted: 619 cm^{-1} (proteins), 783 cm^{-1} (DNA/RNA), 1234 cm^{-1} (proteins and DNA/RNA), 1337 cm^{-1} (proteins and DNA/RNA), 1485 cm^{-1} (proteins and DNA/RNA) and 1616 cm^{-1} (proteins). There is also a lot of background signal present in loading 1. There are distinguishable peaks, so it is thought that PC1 is both a combination of peaks and background signal.

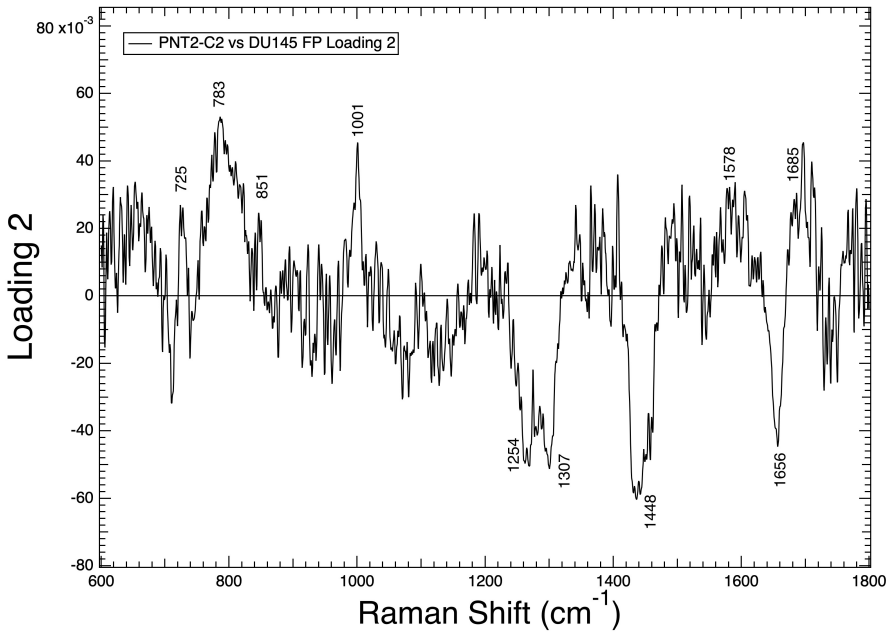


Figure 5.49: Loading 2, representative of PC2, responsible for 9.90% of the sample variation, shows multiple peak contributions in the fingerprint region between PNT2-C2 Live vs DU145. Key peaks attributed to biological assignments have been highlighted:725 cm^{-1} (DNA), 783 cm^{-1} (DNA/RNA), 851 cm^{-1} (proteins), 1001 cm^{-1} (proteins), 1254 cm^{-1} (proteins), 1307 cm^{-1} (proteins,lipids and nucleic acids), 1448 cm^{-1}, 1578 cm^{-1} (DNA/RNA), 1657 cm^{-1} (proteins and lipids) and 1685 cm^{-1} (proteins).

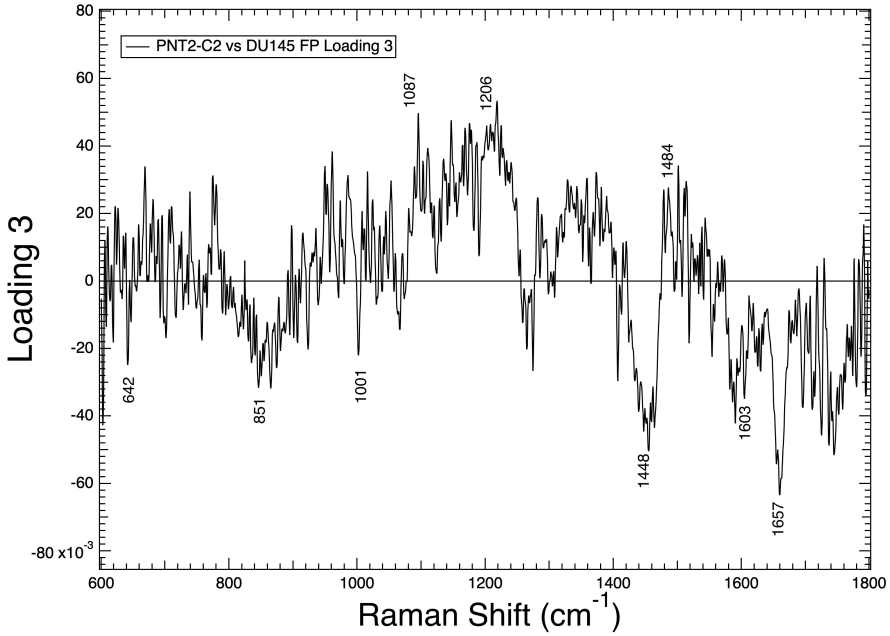


Figure 5.50: Loading 3 has been plotted for PC3 accounting for 6.5% of variance in the fingerprint region between PNT2-C2 Live vs DU145. Key peaks attributed to biological assignments have been labelled: 642 cm^{-1} (proteins), 851 cm^{-1} (proteins), 1087 cm^{-1} (DNA, lipids and proteins), 1206 cm^{-1} (proteins and DNA/RNA), 1448 cm^{-1} (lipids, proteins and DNA/RNA), 1484 cm^{-1} (proteins and DNA/RNA), 1604 cm^{-1} (proteins), 1656 cm^{-1} (proteins and lipids).

Through the analysis of loadings 1, 2 and 3, key peaks were annotated to represent where biomolecular assignments exist, that in turn are responsible for the separation in PCA space. In loading 1, capturing 28.6% of the sample variance, peaks identified and their general assignments are: 619 cm^{-1} (proteins), 783 cm^{-1} (DNA/RNA), 1234 cm^{-1} (proteins and DNA/RNA), 1337 cm^{-1} (proteins and DNA/RNA), 1485 cm^{-1} (proteins and DNA/RNA) and 1616 cm^{-1} (proteins). In loading 2, capturing 9.9% of the sample variance, peaks identified and their general assignments are: 725 cm^{-1} (DNA), 783 cm^{-1} (DNA/RNA), 851 cm^{-1} (proteins), 1001 cm^{-1} (proteins), 1254 cm^{-1} (proteins), 1307 cm^{-1} (proteins,lipids and nucleic acids), 1448 cm^{-1}, 1578 cm^{-1} (DNA/RNA), 1657 cm^{-1} (proteins and lipids) and 1685 cm^{-1} (proteins). Lastly, loading 3, capturing 6.5% of the sample variance, peaks identified and their general assignments are: 642 cm^{-1} (proteins), 851 cm^{-1} (proteins), 1087 cm^{-1} (DNA, lipids and proteins), 1206 cm^{-1} (proteins and DNA/RNA), 1448 cm^{-1} (lipids, proteins and DNA/RNA), 1484 cm^{-1} (proteins and

DNA/RNA), 1604 cm^{-1} (proteins), 1656 cm^{-1} (proteins and lipids). The three PCs are analysed and the corresponding loadings capture 45% of the total population variance. The identified peaks will be used in combination with PIR analysis to identify key biomarkers in Section 5.4.3.

PCA analysis between the PNT2-C2 Live and DU145 of the high wavenumber region is able to ascertain that the variance captured by the first 6 principal components (PCs) are as follows: PC1 = 40.7%, PC2 = 15.5%, PC3 = 5.0%, PC4 = 2.8%, PC5 = 2.4% and PC6 = 2.1%. The first three PCs capture 61.2% and the first six PCs capture 68.4% of the total variance of the sample. Cumulatively, 276 principal components are required to capture total variance of the sample. The PCA results for the high wavenumber region between PNT2-C2 and DU145, using the first three principal components, are shown shown in Figures 5.51 and 5.52. A further rotation of PCA space is shown using PC2, PC3 and PC4 5.53.

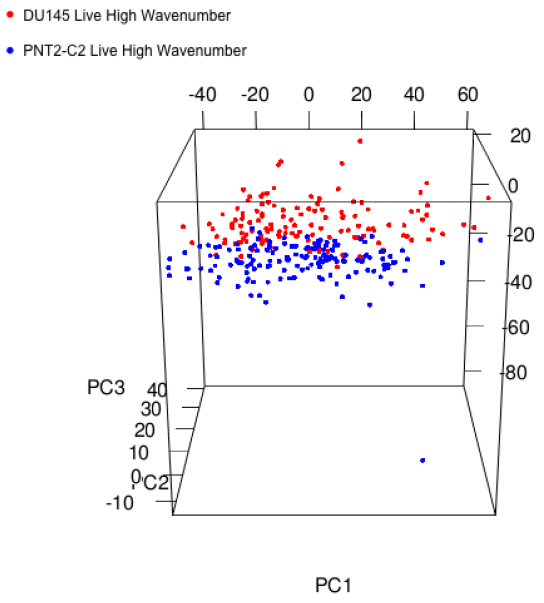


Figure 5.51: 3D PCA space is visualised for the high wavenumber region of PNT2-C2 and DU145 using the first three principal components. There is no separation visible along PC1 or PC3.

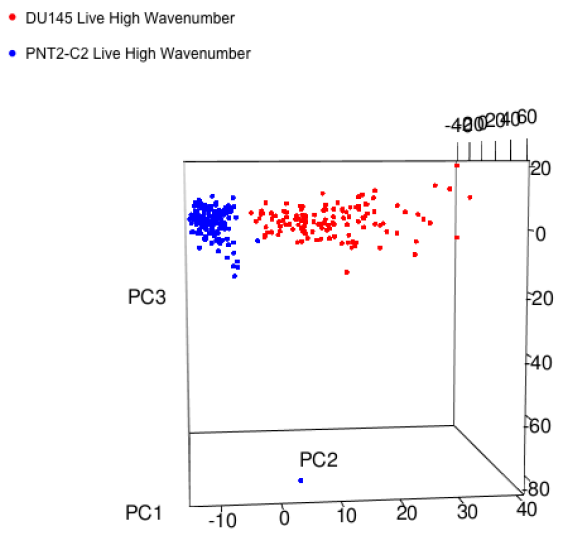


Figure 5.52: A further rotation of the PCA space using the first three principal components is shown to separate PNT2-C2 and DU145 in the high wavenumber region. This separation is observed as happening along PC2. No separation is observed along PC3

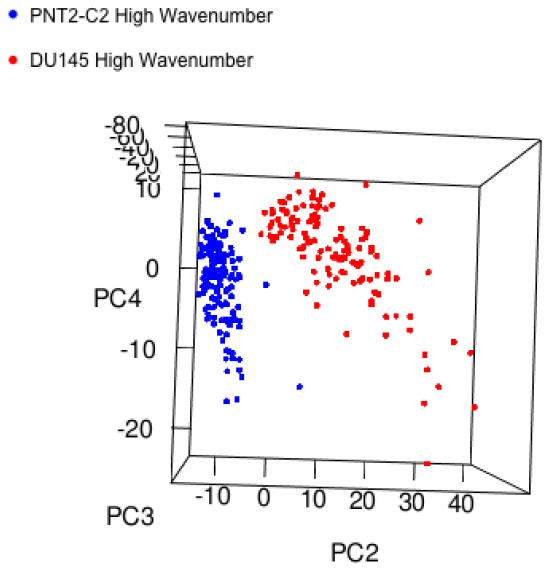


Figure 5.53: It is observed in a further rotation of the PCA space using PC2, PC3 and PC4 that further separation is observed in the PNT2-C2 vs DU145 high wavenumber comparison. There is clear separation occurring at an angle between principal components 2 and 4.

It can be seen in Figures 5.51, 5.52 and 5.53 that there is 100% separation

occurring in the high wavenumber region between PNT2-C2 to DU145. This is observed along PC2 and PC4 in the PCA comparison carried out in Figure 5.53. There was no separation observed in PC1 or PC3 at any stage. The corresponding loadings associated with PC2 and PC4 are examined in Figures 5.54 and 5.55.

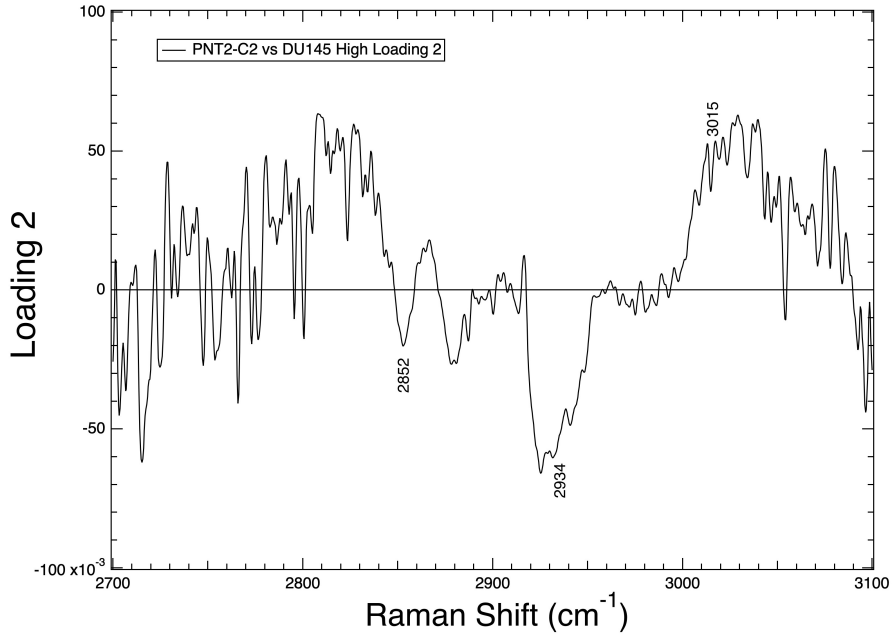


Figure 5.54: Loading 2 capturing 15.5% of the variance in the high wavenumber PCA comparison of PNT2-C2 and DU145 is displayed. The most well defined peak is at 2852 cm^{-1} (lipids). Further peaks annotated are the 2934 cm^{-1} (protein and lipids) and 3015 cm^{-1} (lipids). It is noted that the 3015 cm^{-1} peak exists in a region where there is a lot of background variance. However, it is thought that the biomolecular information is also contributing to the variance.

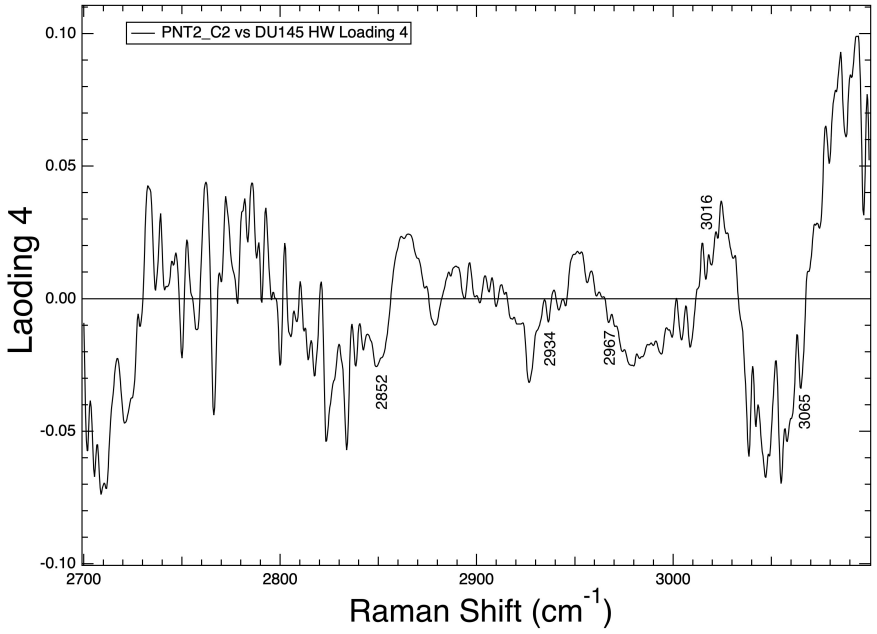


Figure 5.55: Loading 4 capturing 2.8% of the variance in the high wavenumber PCA comparison of PNT2-C2 and DU145 is shown. Key peaks relating to biomolecular assignments have been found at 2852 cm^{-1} (lipids), 2936 cm^{-1} (proteins and lipids), 2967 cm^{-1} (lipids), 3015 cm^{-1} (lipids) and 3065 cm^{-1} (lipids). These peaks will be investigated using PIR analysis at a later stage.

In total there were three peaks relating to biomolecular information identified in loading 2. Loading 2 captures 15.5% of the variance of the PCA in the high wavenumber region between PNT2-C2 and DU145. These peaks were 2852 cm^{-1} (lipids), 2934 cm^{-1} (protein and lipids) and 3015 cm^{-1} (lipids). Loading 3 corresponding to 2.8% of the variance of the sample was able to locate peaks at 2852 cm^{-1} (lipids), 2936 cm^{-1} (lipids). Loading 4 captures 2.8% of the variance in the high wavenumber PCA comparison. There were 5 peaks identified in loading 4 as having biomolecular assignments: 2852 cm^{-1} (lipids), 2934 cm^{-1} (protein and lipids), 2967 cm^{-1} (lipids), 3015 cm^{-1} (lipids) and 3065 cm^{-1} (lipids). It is noted that there is background signal contributing to variance in both loadings 2 and 4. However, it is thought that the biomolecular information is also contributing to the variance around peak locations. The peaks identified at the PCA stage will be used alongside PIR analysis to determine biomarkers to distinguish between PNT2-C2 and DU145.

Corsetti *et al.* (2018) made several biochemical conclusions drawn from a mean subtraction and PCA analysis, such as DU145 having an increased protein content compared to PNT2 [224]. Furthermore, Corsetti *et al.* (2018), through analysis of the mean subtraction and PCA, determined that DU145 has a much higher DNA content due to an increased intensity at the 723 cm^{-1} DNA peak. However, this analysis should not be considered robust enough to make these conclusions as PCA alone can not be considered a complete classification tool for Raman data. This is due not only to class labels not existing in PCA, but because PCA generates biases that are uncorrelated [64]. Furthermore, differences in the baseline subtraction of data could generate artificial differences at the mean subtraction stage. It is for this reason that PCA analysis is only used in tandem with PIR analysis to determine key sources of variance between cell lines.

Finally, a PCA-LDA leave-one-out cross-validation was used to determine the predictive classification of PNT2-C2 and DU145 cells. The number of principal components included was based on using Kaiser's criterion, which consists of using only principal components with variance greater than 1 percent. For PNT2-C2 vs DU145 the PCA-LDA classification tool correctly classified 99.6% cells in the fingerprint region using 11 PCs and 96.1% of cells in the high wavenumber region using 12 PCs. This demonstrates a high classification accuracy between PNT2-C2 and DU145.

5.4.3 Peak intensity ratio analysis PNT2-C2 vs DU145

There has not previously been a study considering peak intensity ratios between PNT2-C2 and DU145. The previously described study by Corsetti *et al.* (2018) did not use relative intensities for analysis, but relied on PCA loading analysis instead [224].

It should be highlighted that the 699 cm^{-1} peak, representative of aminoacid methionine ν(C-S) *trans*, was not able to be analysed due to it being absent from DU145. It is also highlighted that the 876 cm^{-1} and 893 cm^{-1} peak are missing from DU145 as well. These peaks have a biomolecular assignment of proteins,

lipids, carbohydrates and proteins. However, this is not thought to be detrimental to the PIR analysis between PNT2-C2 and DU145. The 699 cm^{-1}, 876 cm^{-1} and 893 cm^{-1} peaks were not identified in the mean profile subtraction or at the PCA stage as being a source of variation between samples.

The peak intensity ratios have been calculated using the methods described in Section 3.6.3 for PNT2-C2 live vs DU145. The complete set of results used to calculate the separation outside of the propagated SE uncertainty, relating to each PIR compared between PNT2-C2 and DU145 for the fingerprint and high wavenumber region, is shown in Table C.4 in the Appendix. There were only four peaks that achieve separation of 90% and above at the PIR stage, 619 cm^{-1} (90%), 641 cm^{-1} (95%), 825 cm^{-1} (95%) and 1060 cm^{-1} (95%). The highest separation observed in the high wavenumber region occurs at 3015 and achieves 82% separation between peaks across the whole spectrum and 100% separation between other peaks in the high wavenumber region.

The first peak intensity to be considered is the 619 cm^{-1} which generated 90% separation outside of the propagated SE uncertainty relating to each PIR compared between PNT2-C2 and DU145. The 619 cm^{-1} was not observed in the mean profile subtraction in Figure 5.43. However, the 619 cm^{-1} peak was observed at the PCA stage in loading 1 in Figure 5.48 as being a source of variance leading to separation between PNT2-C2 and DU145 in the fingerprint region. The specific biomolecular assignment of the 619 cm^{-1} peak is a C-C twist of proteins and a phenylalanine C-C twist. The associated PIR panel is given in Figure 5.56.

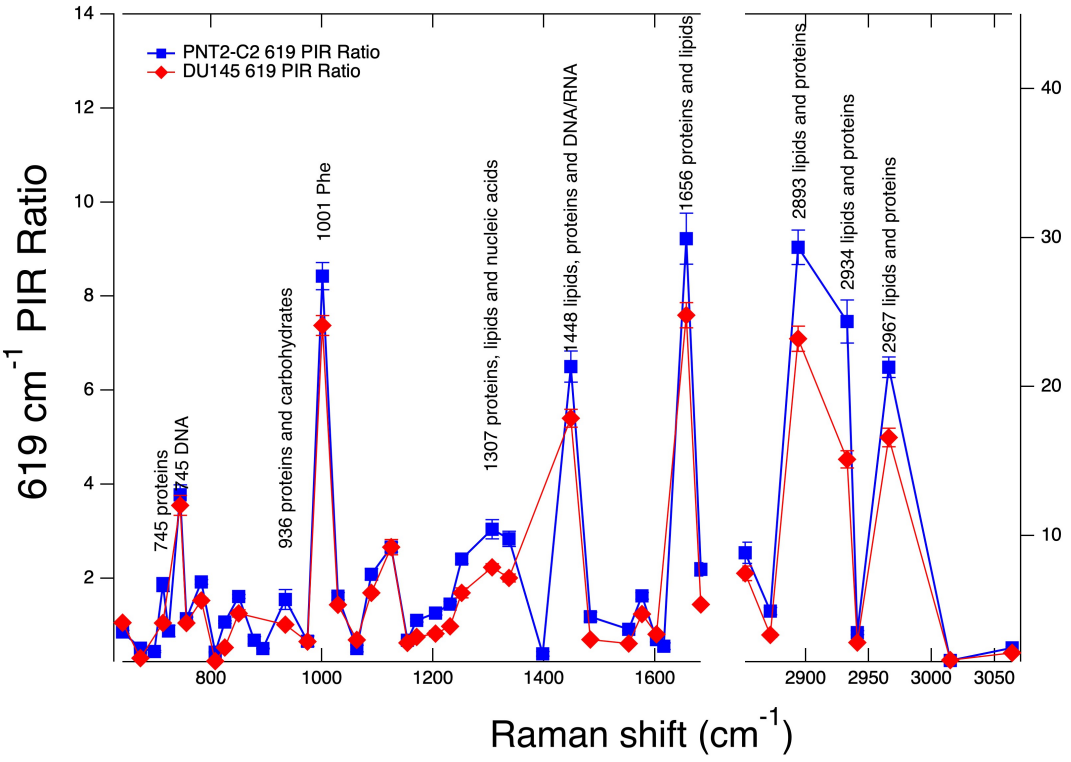


Figure 5.56: The PIR panel for the 619 cm^{-1} peak intensity ratio demonstrates statistically significant separation across the fingerprint region and high wavenumber region, achieving 90% total separation. The 619 cm^{-1} peak is attributed to a C-C twist of proteins and a phenylalanine C-C twist. It can be seen that the PIR profile of DU145 relative to the 619 cm^{-1} peak is lower than that of the PNT2-C2 PIR profile. This suggests that there is a higher relative intensity of the 619 cm^{-1} peak in DU145, attributed to C-C twisting of proteins and phenylalanine C-C twisting.

Figure 5.56 shows that the PIR profile of DU145 relative to the 641 cm^{-1} peak is lower than that of the PNT2-C2 PIR profile, in both the fingerprint and high wavenumber region. This suggests that there is a higher relative intensity of the 619 cm^{-1} peak in DU145, attributed to C-C twisting of proteins and phenylalanine C-C twisting. The study by Corsetti *et al.* (2018), although not examining peak intensity ratios, also determined that DU145 cells had a higher contribution of phenylalanine than PNT2-C2.

The 641 cm^{-1} peak intensity generated 95% separation outside of the propagated SE uncertainty relating to each PIR compared between PNT2-C2 and DU145. The 641 cm^{-1} peak was not observed in the mean profile subtraction

in Figure 5.43. It should be noted, however, that the noise present in the mean profile subtraction in PNT2-C2 made observational analysis less precise. The 641 cm^{-1} peak was observed at the PCA stage in loading 3, Figure 5.50 as being a source of variance leading to separation between PNT2-C2 and DU145. The specific biomolecular assignment of the 619 cm^{-1} peak is a C-S stretch and C-C twist of tyrosine. The Associated PIR panel is given in Figure 5.56.

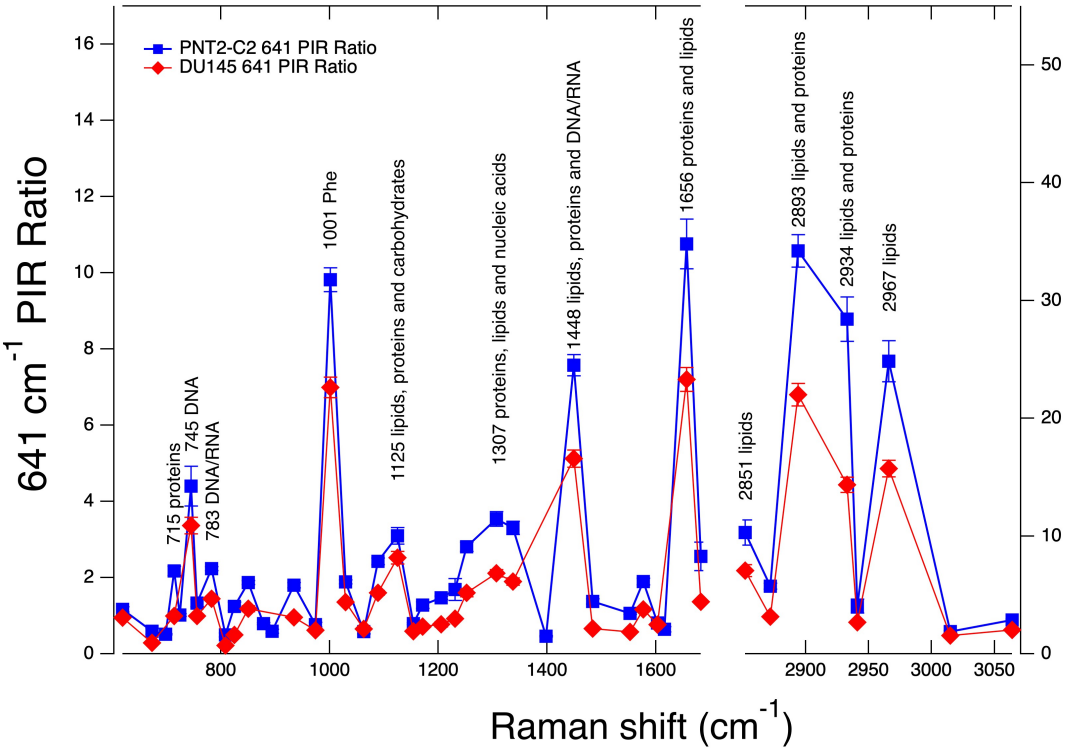


Figure 5.57: The PIR panel for the 641 cm^{-1} peak intensity ratio demonstrates statistically significant separation across the fingerprint region and high wavenumber region PNT2-C2 and DU145, achieving 95% total separation. The 641 cm^{-1} peak is attributed to a C-C twist of tyrosine. It can be seen that the PIR profile of DU145 relative to the 641 cm^{-1} peak is lower than that of the PNT2-C2 PIR profile in both the fingerprint and high wavenumber region. This suggests that there is a higher relative intensity of the C-C twist of tyrosine in DU145. However, no statistically significant separation is found against the 745 cm^{-1} peak, indicating that the relative ratio of the C-C twist of tyrosine to thymine is the same between PNT2-C2 and Du145.

Figure 5.57 shows that the PIR profile of DU145 relative to the 641 cm^{-1} peak is lower than that of the PNT2-C2 PIR profile in both the fingerprint and high wavenumber regions. This suggests that there is a higher relative intensity of the

C-C twist of tyrosine in DU145. An increase in the presence of tyrosine in cells is a feature of malignant cells, protecting the cell from undergoing apoptosis [283].

The 825 cm^{-1} peak intensity generated 95% separation outside of the propagated SE uncertainty relating to each PIR compared between PNT2-C2 and DU145. The 825 cm^{-1} peak was observed in the mean profile subtraction in Figure 5.43 as indicating a difference in the profile. However, the 825 cm^{-1} peak was not observed at the PCA stage. It is thought that the variance associated with the 825 cm^{-1} peak is contributing to the observed variance profile around the 850 cm^{-1} peak in loading 2 and 3.The specific biomolecular assignment of the 825 cm^{-1} peak is a tyrosine out-of-plane ring breathing and O-P-O stretch in DNA. It is also attributed to proline and hydroxyproline.The Associated PIR panel for the 825 cm^{-1} peak is given in Figure 5.58.

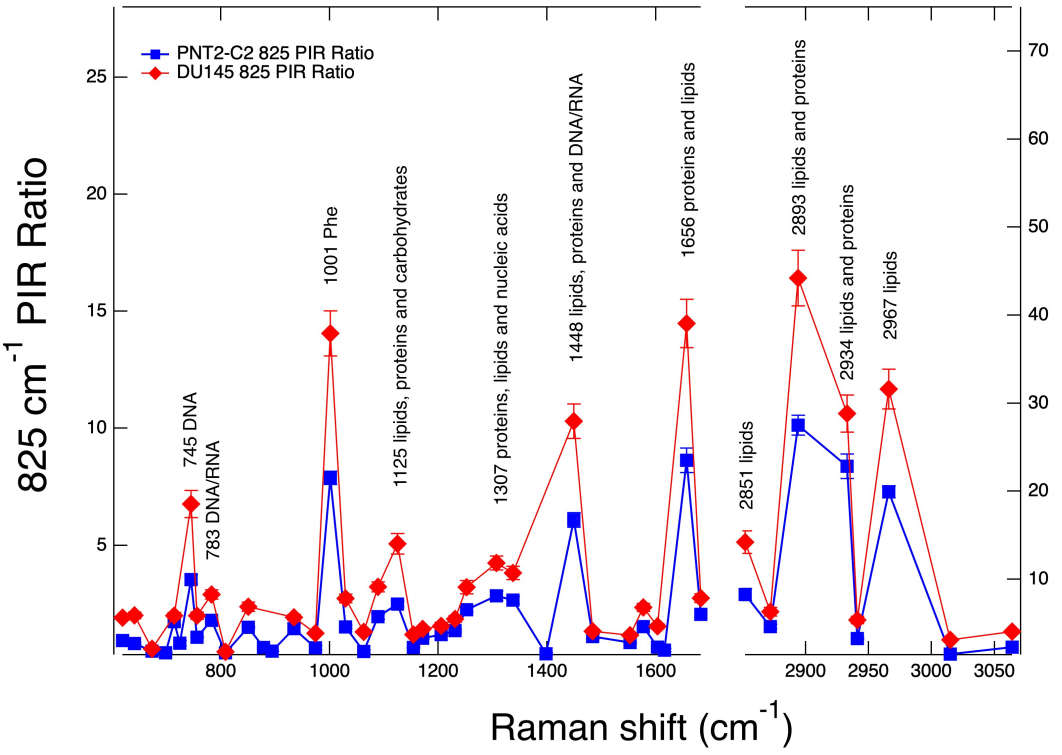


Figure 5.58: The PIR panel for the 825cm^{-1} peak demonstrates statistically significant separation across the fingerprint region and high wavenumber region of PNT2-C2 and DU145, achieving 95% total separation. The 825 cm^{-1} peak is attributed to tyrosine out-of-plane ring breathing and O-P-O stretch in DNA. It has a further biomolecular assignment of proline and hydroxyproline. However, the 825 cm^{-1} peak intensity ratio is not able to separate the whole spectrum. It can be seen that there is no statistical separation at 810 cm^{-1} (DNA) which is also representative of the O-P-O stretch in DNA. In general, the PIR profile of the 825cm^{-1} peak demonstrates that there is a lower relative intensity of the biomolecular constituents attributed to the peak in DU145. It is not possible to conclusively say which biomolecule is responsible for the lower relative intensity due to the high number of biomolecular contributions to the peak.

Figure 5.58 shows the PIR panel for the 825cm^{-1} peak, demonstrating that there is a lower relative intensity of the biomolecular constituents attributed to the peak in DU145. It is difficult to analyse which specific biomolecular assignment is responsible for this separation due to the high number of biomolecular contributions. It is for this reason, and also that the 825cm^{-1} peak did not appear in the loadings, that it should not be considered a robust biomarker.

The 1060 cm^{-1} peak associated with 95% separation outside of the propagated SE uncertainty is fitted as part of a two band region. The peak that it is convoluted

with, 1087 cm^{-1}, achieves 66% separation at the PIR stage as well as being one of the main contributors to the variance in loading 3, Figure 5.50. It should be noted that the 1060 cm^{-1} peak was not identified during the mean profile subtraction or in the loadings. The 1060 cm^{-1} peak is specifically attributed to a lipids (chain C-C stretch) and DNA/RNA PO_2^- back bone stretching. The PIR panel associated with the 1060 cm^{-1} peak is shown in Figure 5.59.

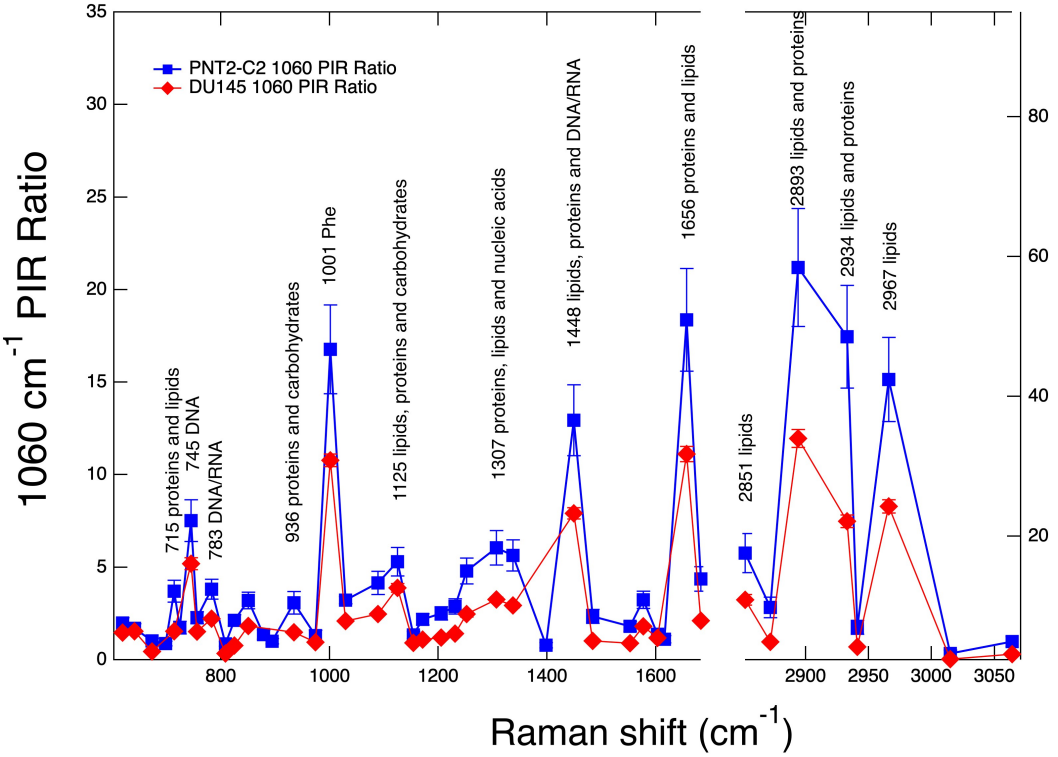


Figure 5.59: The PIR panel for the 1060 cm^{-1} PIR demonstrates statistically significant separation across the fingerprint region and high wavenumber region of PNT2-C2 and DU145, achieving 95% total separation. The 1060 cm^{-1} peak is specifically attributed to a lipids (chain C-C stretch) and DNA/RNA PO_2^- back bone stretching. The PIR profiles generated by the 1060 cm^{-1} peak are considerably lower in DU145 than PNT2-C2. This suggests that there is a much higher relative intensity of lipids (chain C-C stretch) and DNA/RNA PO_2^- back bone stretching in DU145than PNT2-C2.

The PIR panel for the 1060 cm^{-1} peak shown in Figure 5.59 shows that the PIR profiles generated by the 1060 cm^{-1} peak are considerably lower in DU145 than PNT2-C2. This suggests that there is a much higher relative intensity of lipids (chain C-C stretch) and DNA/RNA PO_2^- back bone stretching in DU145 than PNT2-C2. This relative increase in the PO_2^- back bone stretching of DNA/RNA is

consistent in the fact that a signature of metastatic cancer cells is an increased DNA/RNA content [281]. It also indicates that there is a difference in the lipid structure of DU145 when compared to PNT2-C2.

The 3015 cm^{-1} peak achieved 82% separation outside of the propagated SE uncertainty relating to each PIR compared between PNT2-C2 and DU145. However, the 3015 cm^{-1} peak generates 100% separation in the high wavenumber region. The 3015 cm^{-1} peak is specifically assigned to unsaturated =CH stretching in lipids. The 3015 cm^{-1} peak was also identified in the mean profile subtraction in Figure 5.44. Furthermore, the 3015 cm^{-1} peak was also found to be a major source of variance in both loading 2 (Figure 5.54) and loading 4(Figure 5.55)at the PCA stage. The 3015 cm^{-1} peak is explored using a PIR panel to explore relative peak relationships further in Figure 5.60.

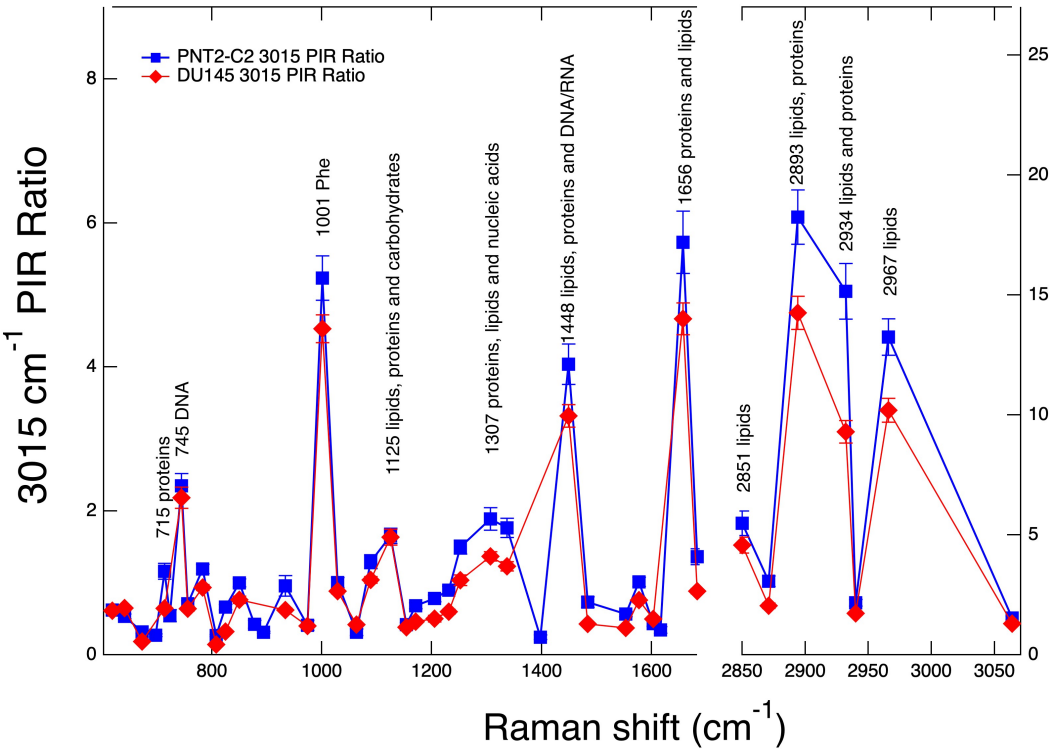


Figure 5.60: The PIR panel for the 3015 cm^{-1} peak intensity ratio demonstrates statistically significant separation across the fingerprint region and high wavenumber region, achieving 82% total separation. The 3015 cm^{-1} peak is attributed to an unsaturated =CH stretch in lipids, unsaturated fatty acids. However, the 3015 cm^{-1} peak intensity ratio is not able to separate the whole spectrum. It can be seen that there is no statistical separation at several peak locations, particularly 745 cm^{-1} (DNA) and 1125 cm^{-1} (lipids, proteins and carbohydrates). The general reduced relative intensity of the PIRs would suggest that there is a higher relative contribution of unsaturated fatty acids in DU145.

Figure 5.60 shows that there is higher general relative intensity of the 3015 cm^{-1} peak in DU145. This in turn signifies a higher relative intensity of unsaturated =CH stretch in lipids in DU145. This change in relative lipid content is indicative of cancerous growth and is brought about by metabolic changes in the cell [282]. Although not clear in Figure 5.60, there is separation occurring with the 2851 cm^{-1} lipid peak. The 2851 peak is specifically attributed to a CH_2 symmetric stretch in lipids. This means there is a relative difference in ratio of total unsaturated fatty acids to total fatty acids (TUFA/TFA) between PNT2-C2 and DU145. The (TUFA/TFA) ratio has been used as a marker for the malignant cells [64, 140]. Furthermore, Previous work by Hedegaard *et al.* (2010) comparing

metastatic and non metastatic mouse cell lines also noted that metastatic cells contained an increase in unsaturated fatty acids [284]. However, this study was not carried out in the high wavenumber region, using different peaks to generate this result. Furthermore, the study carried out by Hedegaard *et al.* (2010) did rely on the use of dried cells, which as shown in chapter 4 of this thesis, can lead to changes in the unsaturated fatty acid content and thus should be treated with caution.

5.5 PC3

A study by Potcoava *et al.* (2014) that analysed the cell line LNCaP, also considered PC3, generating a lipid profile for each [64]. The main purpose of the study carried out by Potcoava *et al.* (2014) was to investigate the intracellular lipid droplets and their saturation in cells during hormone response therapy, and the process of lipogensis. As previously stated, the experimental setup of this study varied with the use of a different laser and fixation methods. Another major source of uncertainty in the methodology employed by Potcoava *et al.* (2014) was an uncertain range of cells measured with regard to the convergence of spectra included in the spectrum of both LNCaP and PC3. Further issues associated with methodology of Potcoava *et al.* (2014) are reported in detail in Section 5.3. This thesis will expand on the work carried out by Potcoava *et al.* (2014) by considering the DNA/RNA and protein assignments in the fingerprint and high wavenumber region for the use as potential biomarkers.

In terms of PCA analysis, a further study using Raman spectroscopy by Crow *et al.* (2005) compared the androgen-sensitive cell lines (LNCaP and PCa 2b) and androgen-insensitive cell lines (DU145 and PC3). The study by Crow *et al.* (2005) used a 833nm laser and considered on average 50 spectra from each cell line; no exact amount is stated [63]. From a methodical perspective, it is not clear if the data has converged for each cell line population. There is no statement in the study by Crow *et al.* (2005) as to the convergence of the data. Due to the fact that the objective of the study by Crow *et al.* (2005) is to construct a diagnostic algorithm, this is highlighted as a major oversight.

In this thesis the convergence of each cell line is independently assessed, as different cell lines or cell populations will converge at different rates. One fundamental finding of Crow was a decreased glycogen content in androgen-independent cell lines [63]. This band is beyond the limits of this thesis, which only considers the range 600-1800 cm^{-1}. Additionally, Further differences were noted in the nucleic acid bands between the androgen-sensitive and androgen-insensitive cell lines. As the study by Crow *et al.* (2005) was carried out using cells

undergoing the drying process/autolysis, it is also highlighted that this thesis will be analysing PC3 in its live cell state.

The analysis presented in this thesis builds on the previous work carried out on PC3 by providing a full characterisation of PC3, using Raman spectroscopy in the fingerprint and high wavenumber region. Furthermore, as far as the author is aware, it will compare the cell line PC3 to a normal cell line for the first time, using a mean profile subtraction, principal component analysis and peak intensity ratio analysis.

5.5.1 PNT2-C2 vs PC3 mean profile subtraction

The normalised Raman signal for PNT2-C2 and PC3 is displayed for the fingerprint region in Figure 5.61 and the high wavenumber region is shown in Figure 5.62. The fingerprint region of both average spectrum has undergone a linear baseline, as well as being offset, for an arbitrary amount. The high wavenumber region has undergone a linear baseline as well as being offset for an arbitrary amount. Both the fingerprint and high wavenumber region have been normalised with respect to the total spectrum area.

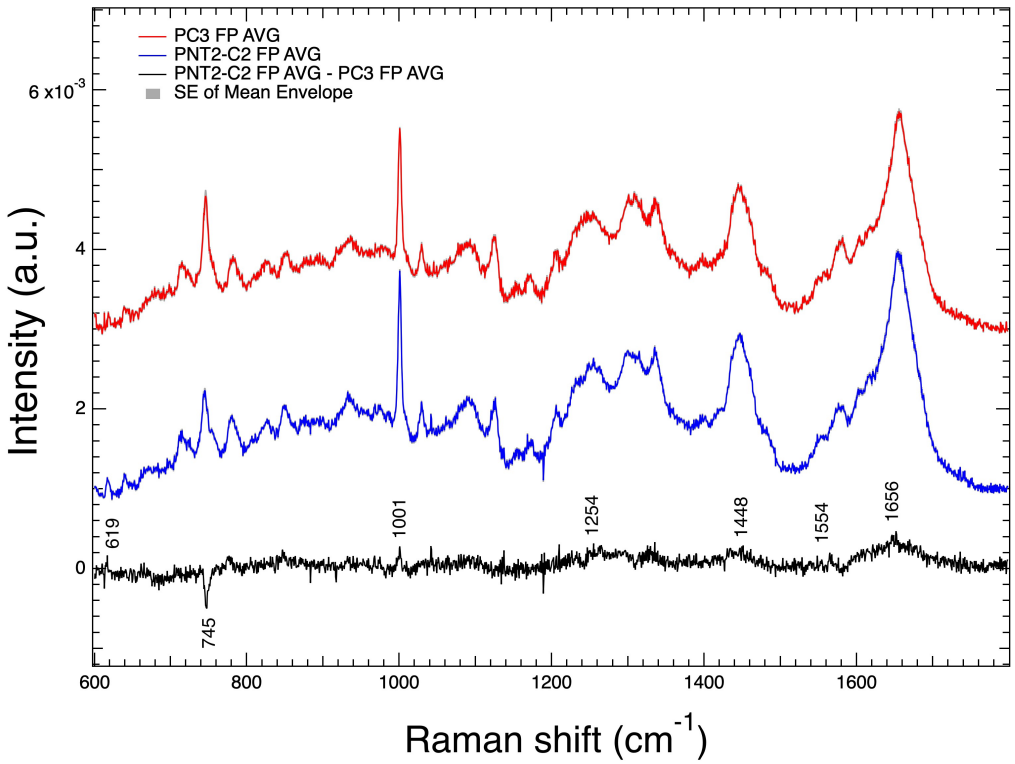


Figure 5.61: A mean profile subtraction is shown for PNT2-C2 minus PC3 for the fingerprint region (600-1800 cm^{-1}). The standard error of the mean envelope is shown by the grey region shaded around the average spectra. Key peaks of interest have been annotated at 619 cm^{-1} (proteins), 745 cm^{-1} (DNA), 1001 cm^{-1} (proteins), 1254 cm^{-1} (proteins), 1448 cm^{-1} (lipids, proteins and DNA), 1554 cm^{-1} (proteins) and 1656 cm^{-1} (proteins and lipids).

The mean fingerprint subtraction profile between PNT2-C2 and PC3 in Figure 5.61 shows that there are differences between the Raman profiles of PNT2-C2 and PC3. Key biochemical signatures have emerged at the locations; 619 cm^{-1} (proteins), 745 cm^{-1} (DNA), 1001 cm^{-1} (proteins), 1254 cm^{-1} (proteins), 1448 cm^{-1} (lipids, proteins and DNA), 1554 cm^{-1} (proteins) and 1656 cm^{-1} (proteins and lipids). The single largest difference in the mean comparison was the emergence of the 745 cm^{-1} peak, indicating that there may be differences in relative intensity of the thymine ring breathing mode between PNT2-C2 and PC3. The rest of the peaks found were not as pronounced in the mean profile subtraction. The sudden spike in both the positive and negative direction in the profile of the mean subtraction, at 619 cm^{-1}, may infer a difference in a C-C twist of proteins and the amino acid phenylalanine, or, that there are spectral artifacts artificially

generating the observed difference. Further analysis is needed to draw any conclusive remarks. There are noticeable visual differences in the profile at the peaks 1254 cm^{-1} and 1554 cm^{-1}. Whilst this will impact the fitting of peaks in these regions, it is not possible to say what effect it will have on the peak intensity ratios at this point. There are also broader differences observed at 1001 cm^{-1}, 1448 cm^{-1}, 1656 cm^{-1}. Similarly, these peaks require further analysis before any statements can be made. Tentatively, these peaks suggest there may be a difference in the secondary protein structure of the cell, due to the presence of amide III in the region 1254 cm^{-1} band and amide I at the 1656 cm^{-1} band. As far as the author is aware, this is the first time a mean subtraction has been carried out between PC3 and a normal cell line in the fingerprint region.

The mean high wavenumber mean subtraction profile between PNT2-C2 and PC3 is now considered in Figure 5.62.

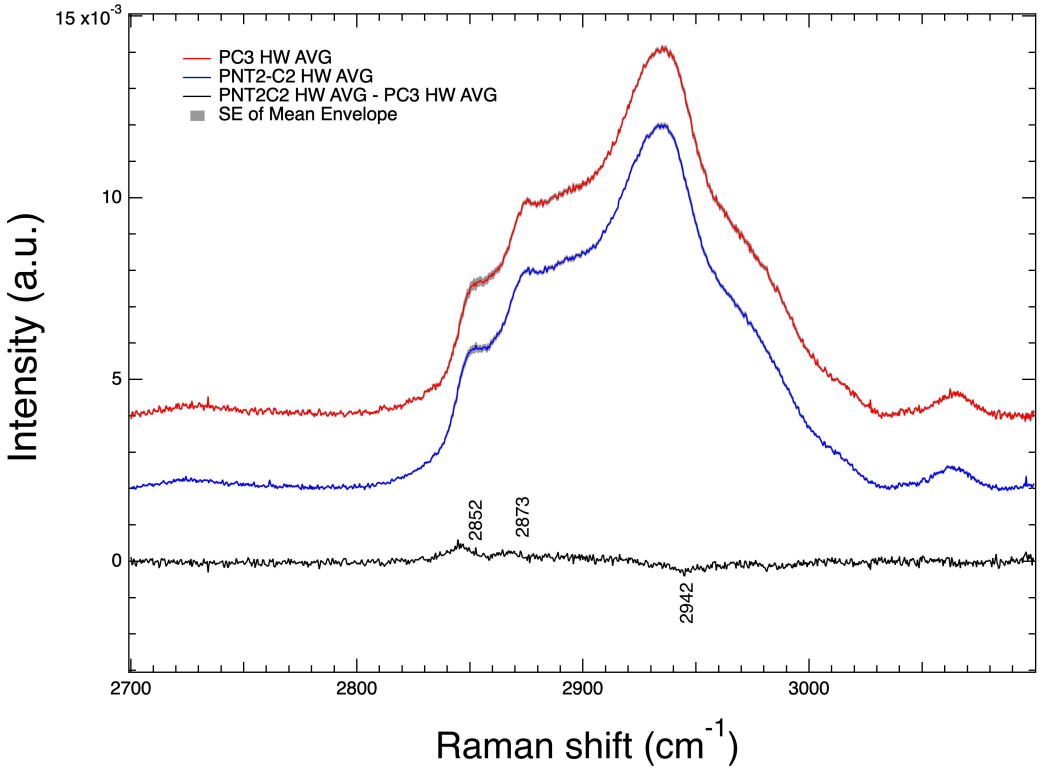


Figure 5.62: A mean profile subtraction is shown for PNT2-C2 minus PC3 for the high wavenumber region (2700-3100 cm^{-1}). The standard error of the mean envelope is shown by the grey region shaded around the average spectra. Key peaks of interest have been annotated at 2852 cm^{-1} (lipids), 2873 cm^{-1} (lipids and proteins) and 2942 cm^{-1} (lipids and proteins). It should be noted that the peak 2852 cm^{-1} is not at the local maximum of the peak in the mean subtraction profile but slightly to the right of it. The annotation of 2852 cm^{-1} was marked to show there is variation in the background of the profiles of PNT2-C2 and PC3 which may in turn influence peak fittings.

The high wavenumber mean profile subtraction between PNT2-C2 and PC3 in Figure 5.62 does not indicate there are any large differences in the respective profiles of the cell lines. The annotation of peaks 2852 cm^{-1} and 2873 cm^{-1} detail the respective background difference between them. The change in profile shape around these peaks may indicate that there is a change in the relative ratio of saturated fatty acids, attributed to the band at 2852 cm^{-1}. There was also a small peak that emerged at 2942 cm^{-1}, indicating further analysis may find differences associated with C-H vibrations in lipids and proteins as well as CH_2 modes in lipids. As far as the author is aware, this is the first time a mean subtraction has been carried out between PC3 and a normal cell line in the high wavenumber

region.

5.5.2 PNT2-C2 vs PC3 PCA comparison

PCA analysis between the PNT2-C2 Live and PC3 of the fingerprint region is able to ascertain that the variance captured by the first six principal components (PCs) are as follows: PC1 = 26.8%, PC2 = 16.7%, PC3 = 7.3%, PC4 = 5%, PC5 = 3.4% and PC6 = 2.2%. The first three PCs capture 50.8% and the first six PCs capture 61.4% of the total variance of the sample. Cumulatively, 271 principal components are required to capture total variance of the sample. Two rotations of PCA space, using the first three principal components, are displayed in Figures 5.63 and 5.64. A further rotation of PCA space is shown using principal components two, three and four in Figure 5.65.

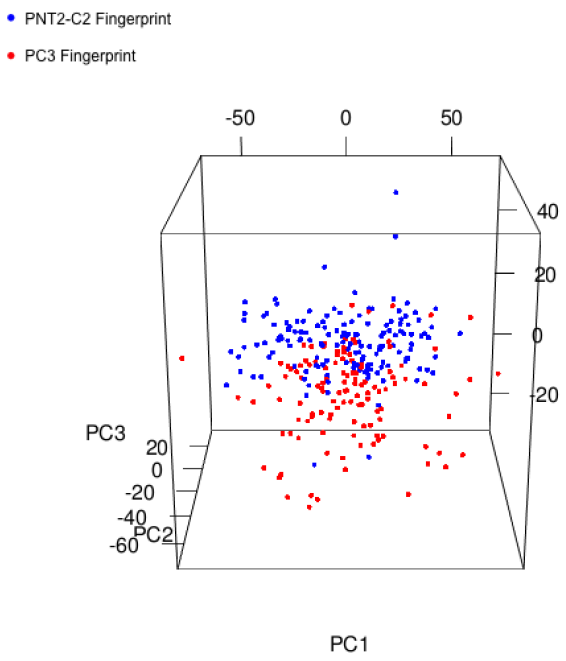


Figure 5.63: The 3D PCA space, using the first three principal components, compares the fingerprint region of PNT2-C2 and PC3. At this angle to is possible to see separation of samples occurring along PC3. There is no separation visible along PC1.

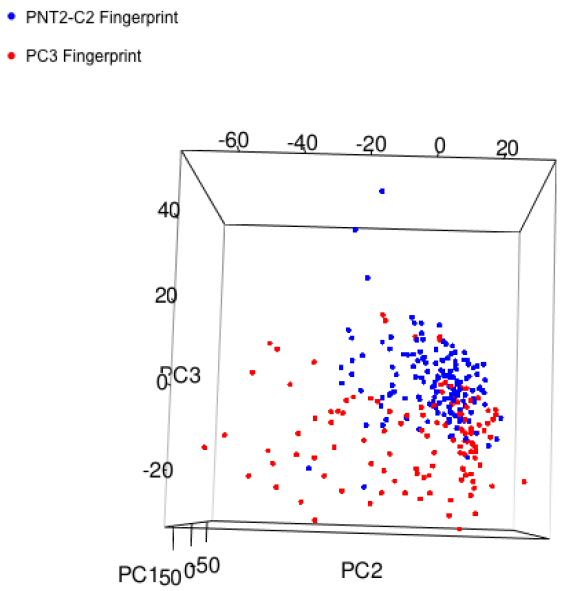


Figure 5.64: A further rotation of 3D PCA space using the first three principal components compare the fingerprint region of PNT2-C2 and PC3. This rotation shows separation of samples along PC3 but not along PC2.

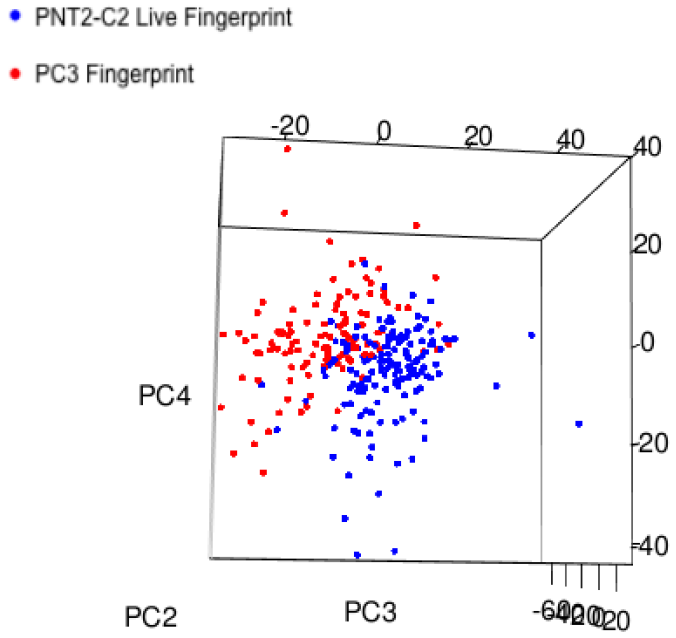


Figure 5.65: Here, PCA comparison of the fingerprint region of PNT2-C2 and PC3 is displayed in 3D space using principal components 2,3 and 4. Clear separation of samples is observed at an angle along PC3 and PC4.

The first two rotations of PCA space in Figures 5.63 and 5.64, using the first three principal components, showed that separation was only occurring in PC3. The 3D PCA space was further plotted using PC2, PC3 and PC4 in Figure 5.65, showing that separation was occurring along PC3 and PC4. As no separation was observed in PC1 and PC2, their corresponding loadings are not considered. The loadings corresponding to PC3 and PC4 are shown in Figures 5.66 and 5.67.

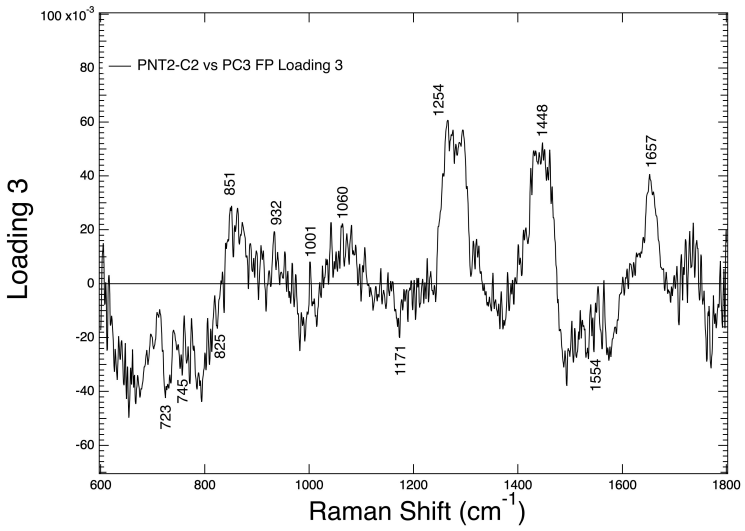


Figure 5.66: Loading 3 of the fingerprint region comparison of PNT2-C2 vs PC3 is displayed corresponding to 7.3% of the total variance captured. Key peaks attributed to biological assignments have been found at: 723 cm^{-1} (DNA), 745 cm^{-1} (DNA), 825 cm^{-1} (proteins and DNA/RNA), 851 cm^{-1} (proteins), 932 cm^{-1} (proteins and carbohydrates), 1001 cm^{-1} (proteins), 1060 cm^{-1} (lipids and DNA/RNA), 1171 cm^{-1} (proteins), 1254 cm^{-1} (proteins), 1448 cm^{-1} (lipids, proteins and DNA/RNA), 1554 cm^{-1} (proteins), 1657 cm^{-1} (proteins and lipids).

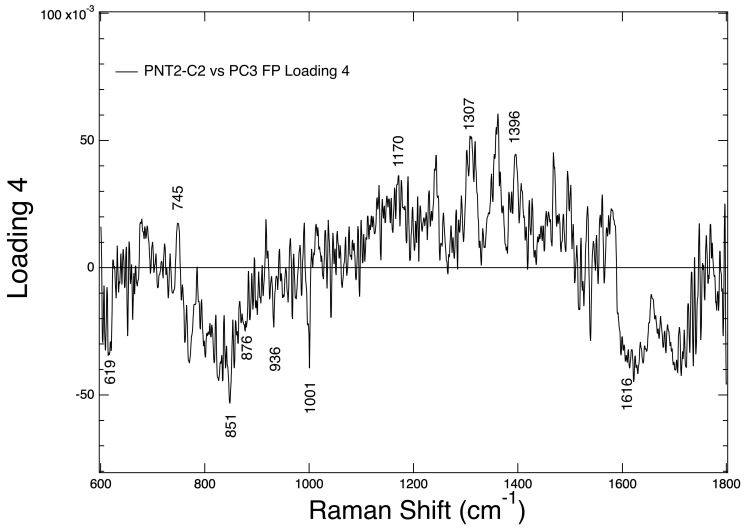


Figure 5.67: Loading 4, capturing 5% of the total variance of the fingerprint region comparison of PNT2-C2 vs PC3 is displayed. The peaks corresponding to biomolecular assignments have been annotated; 619 cm^{-1} (proteins), 745 cm^{-1} (DNA), 851 cm^{-1} (proteins), 876 cm^{-1} (proteins and lipids), 936 cm^{-1} (proteins and carbohydrates), 1001 cm^{-1} (proteins), 1170 cm^{-1} (proteins), 1307 cm^{-1} (proteins, lipids and nucleic acids), 1396 cm^{-1} (proteins), 1616 cm^{-1} (proteins).

Through analysis of Figures 5.66 and 5.67, it is possible to capture the peaks and their biological assignments that are contributing to the variance between PNT2-C2 and PC3 in the fingerprint region. In loading 3, which captured 7.3% of the variance of the sample, peaks were found at 723 cm^{-1} (DNA), 745 cm^{-1} (DNA), 825 cm^{-1} (proteins and DNA/RNA), 851 cm^{-1} (proteins), 932 cm^{-1} (proteins and carbohydrates), 1001 cm^{-1} (proteins), 1060 cm^{-1} (lipids and DNA/RNA), 1171 cm^{-1} (proteins), 1254 cm^{-1} (proteins), 1448 cm^{-1} (lipids, proteins and DNA/RNA), 1554 cm^{-1} (proteins), 1657 cm^{-1} (proteins and lipids). In loading 4, which captured 5% of the total variance of the samples, peaks were found at 619 cm^{-1} (proteins), 745 cm^{-1} (DNA), 851 cm^{-1} (proteins), 876 cm^{-1} (proteins and lipids), 936 cm^{-1} (proteins and carbohydrates), 1001 cm^{-1} (proteins), 1170 cm^{-1} (proteins), 1307 cm^{-1} (proteins, lipids and nucleic acids), 1396 cm^{-1} (proteins), 1616 cm^{-1} (proteins). The peaks identified at this stage will be used in tandem with peak intensity ratio analysis in Section 5.5.3 to identify potential biomarkers.

PCA analysis between the PNT2-C2 Live and PC3 of the high wavenumber region is able to ascertain that the variance captured by the first six principal

components (PCs) are as follows: PC1 = 47.2%, PC2 = 13.9%, PC3 = 6.8%, PC4 = 4.5%, PC5 = 2.5% and PC6 = 1.9%. The first three PCs capture 50.8% and the first six PCs capture 61.4% of the total variance of the sample. Cumulatively, 271 principal components are required to capture total variance of the sample. Three rotations of PCA space, using the first three principal components, are displayed in Figures 5.63 and 5.64. A further rotation of PCA space using PC2, PC3 and PC4 is shown in Figure 5.65.

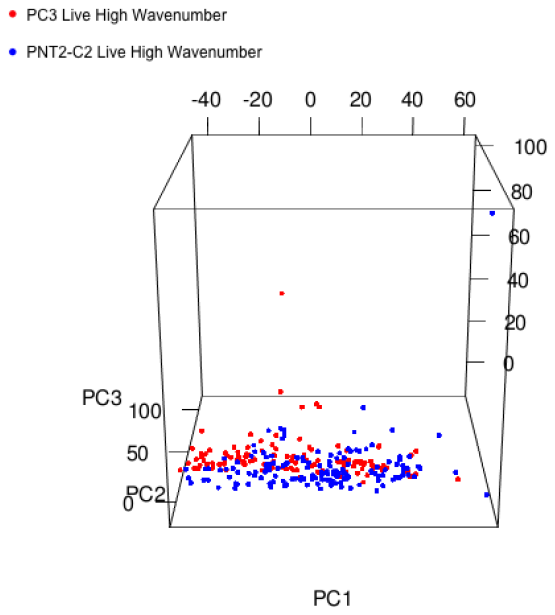


Figure 5.68: 3D PCA space shows the comparison of the high wavenumber region of PNT2-C2 and PC3 using the first three principal components. In this rotation of 3D space separation is not observed in PC1 or PC3. There is separation observed along PC2.

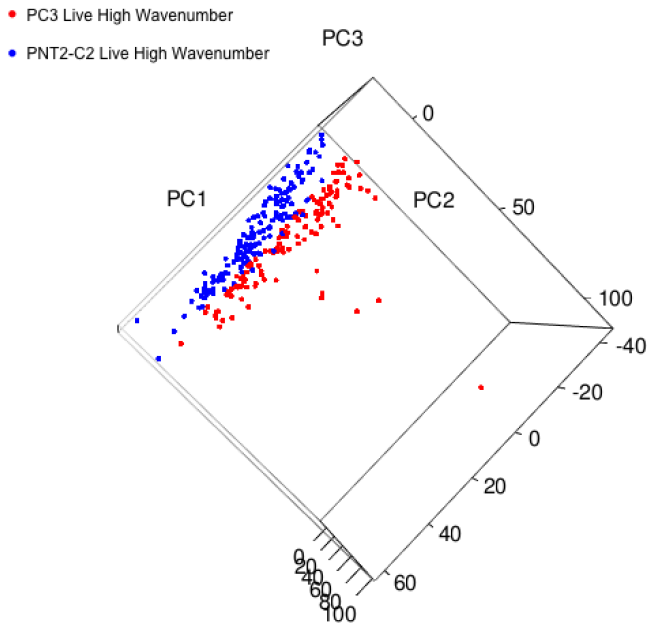


Figure 5.69: A further rotation of 3D PCA space shows the comparison of the high wavenumber region of PNT2-C2 and PC3 using the first three principal components. At this angle it can be seen that separation is occurring along PC2. No separation is observed along PC1.

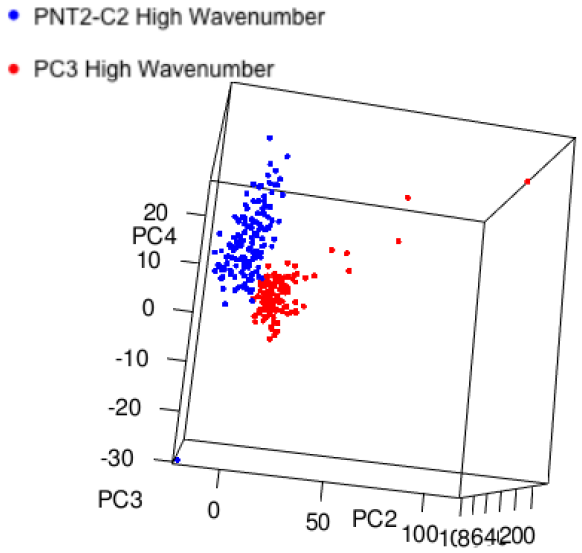


Figure 5.70: A final rotation of 3D PCA space shows the comparison of the high wavenumber region of PNT2-C2 and PC3 using PC2, PC3 and PC4. At this angle it can be seen that separation is occurring at an angle between PC2 and PC4.

Rotations of the 3D PCA space in Figures 5.68 and 5.69 show that PNT2-C2 and PC3 in the high wavenumber region can be separated using PC2. A further

consideration of 3D PCA space in Figure 5.70, using PC2, PC3 and PC4, is able to show that separation of PNT2-C2 and PC3 is also occurring at an angle along PC2 and PC4. No separation was found in PC1 and PC3, meaning that the corresponding loadings will not be examined. The loadings corresponding to PC2 (loading 2) and PC4 (loading 4) are examined in Figures 5.71 and 5.72

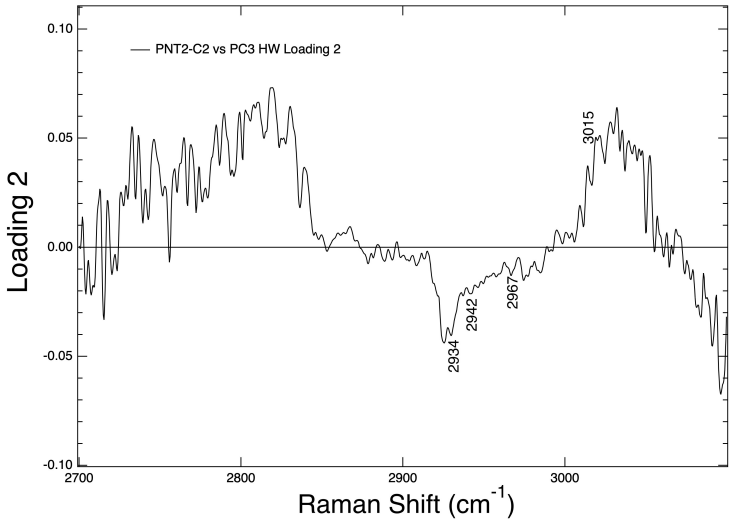


Figure 5.71: Here, Loading 2 of the fingerprint region comparison of PNT2-C2 Live vs PC3 is displayed. Key peaks attributed to biological assignments have been found at 2934 cm^{-1} (proteins and lipids), 2942 cm^{-1} (lipids and proteins), 2967 cm^{-1} (lipids), 3015 cm^{-1} (lipids).

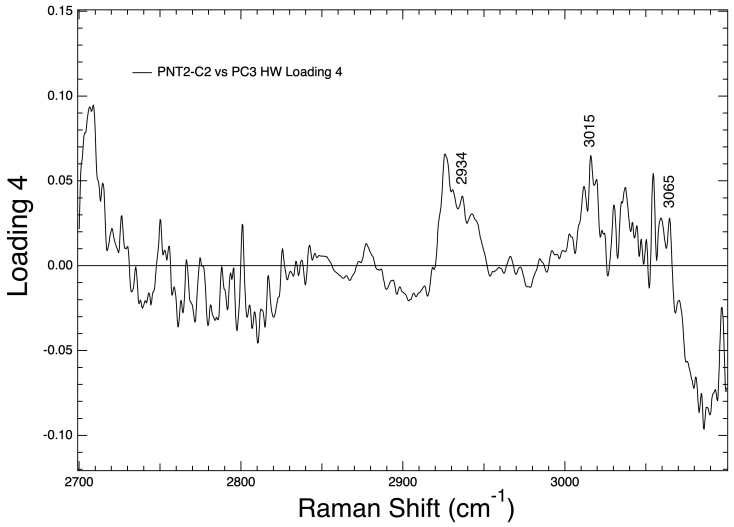


Figure 5.72: Here, Loading 4 of the fingerprint region comparison of PNT2-C2 Live vs PC3 is displayed. Key peaks attributed to biological assignments have been found at 2934 cm^{-1} (lipids and proteins), 3015 cm^{-1} (lipids) and 3065 cm^{-1} (lipids).

As shown in Figure 5.71, Loading 2 of the fingerprint region comparison of PNT2-C2 Live vs PC3 is displayed. Loading 2 is responsible for 13.9% of the variance of the total sample population. Key peaks attributed to biological assignments in loading 2 have been found at 2934 cm^{-1} (proteins and lipids), 2942 cm^{-1} (lipids and proteins), 2967 cm^{-1} (lipids), 3015 cm^{-1} (lipids). Similarly, loading 4, accounting for 4.5% of the total variance in the sample, is shown in Figure 5.72. In loading 4, key peaks were located at 2934 cm^{-1} (lipids and proteins), 3015 cm^{-1} (lipids) and 3065 cm^{-1} (lipids). The peaks identified in loading 2 and loading 4 will be combined with PIR analysis in Section 5.5.3 to detail robust biomarkers between PNT2-C2 and PC3 in the high wavenumber region.

Finally, a PCA-LDA leave-one-out cross-validation was used to determine the predictive classification of PNT2-C2 and PC3 cells. The number of principal components included was based on using Kaiser's criterion, which consists of using only principal components with variance greater than 1 percent. For PNT2-C2 vs PC3 the PCA-LDA classification tool correctly classified 99.3% cells in the fingerprint region, using 10 PCs and 96.0% of cells in the high wavenumber region, using 9 PCs. This demonstrates a high classification accuracy between PNT2-C2 and PC3.

5.5.3 Peak intensity ratio analysis PNT2-C2 vs PC3

A study by Potcoava *et al.* (2014) has previously carried out peak intensity ratio analysis on PC3 [64]. However, Potcoava *et al.* (2014) only considered the lipid profile of PC3 and utilised specific ratios to profile the cell line, to investigate lipid saturation in cells during hormone response therapy [64]. Furthermore, no comparison was made to normal cells. However, this thesis will expand on the work carried out by Potcoava *et al.* (2014) by considering the DNA/RNA, protein and lipid profiles to fully characterise PC3.

The peak intensity ratios have been calculated using the methods described in Section 3.6.3 for PNT2-C2 live vs PC3. The complete set of results used to

calculate the separation outside of the propagated SE uncertainty, relating to each PIR compared between PNT2-C2 and PC3 for the fingerprint and high wavenumber region, is shown in Table C.5 in the Appendix. There were four peaks identified that achieved separation of 90% or greater between the Raman peaks in PNT2-C2 and PC3. The Raman bands and the percentage of separation are as follows: 619 cm^{-1} (95%), 745 cm^{-1} (95%), 1234 cm^{-1} (98%), 1396 cm^{-1} (91%). A further four peaks were identified as achieving over 80% separation at 876 cm^{-1} (89%), 1125 cm^{-1} (95%), 1171 cm^{-1} (95%) and 1578 cm^{-1} (95%).

The 619 cm^{-1} peak is specifically attributed to a C-C twist of proteins and the amino acid phenylalanine. It was thought that there may have may been a large separation observed at the 619 cm^{-1} through observation of the mean profile subtraction between PNT2-C2 and PC3 in Figure 5.61. However, no conclusive remarks were made due to the noise obscuring detailed observation. The 619 cm^{-1} peak is also present in loading 4. Whilst the 619 cm^{-1} peak was identified in loading 4 there was also noise present. It is for these reasons that the 619 cm^{-1} peak, although exhibiting 95% separation outside of the propagated SE uncertainty relating to each PIR, is not investigated further or considered a reliable biomarker in this thesis.

The 745 cm^{-1} peak was also able to distinguish between 95% separation outside of the propagated SE uncertainty relating to each PIR between PNT2-C2 and PC3. The 745 cm^{-1} peak is generated by a pure Thymine ring breathing mode in DNA. The 745 cm^{-1} peak was very pronounced in the mean subtraction profile in Figure 5.61. Furthermore, the 745 cm^{-1} peak was also determined to be contributing to the separation in loadings 3 and 4, Figures 5.66 and 5.67 respectively. The PIR panel associated with the 745 cm^{-1} peak is shown in Figure 5.73.

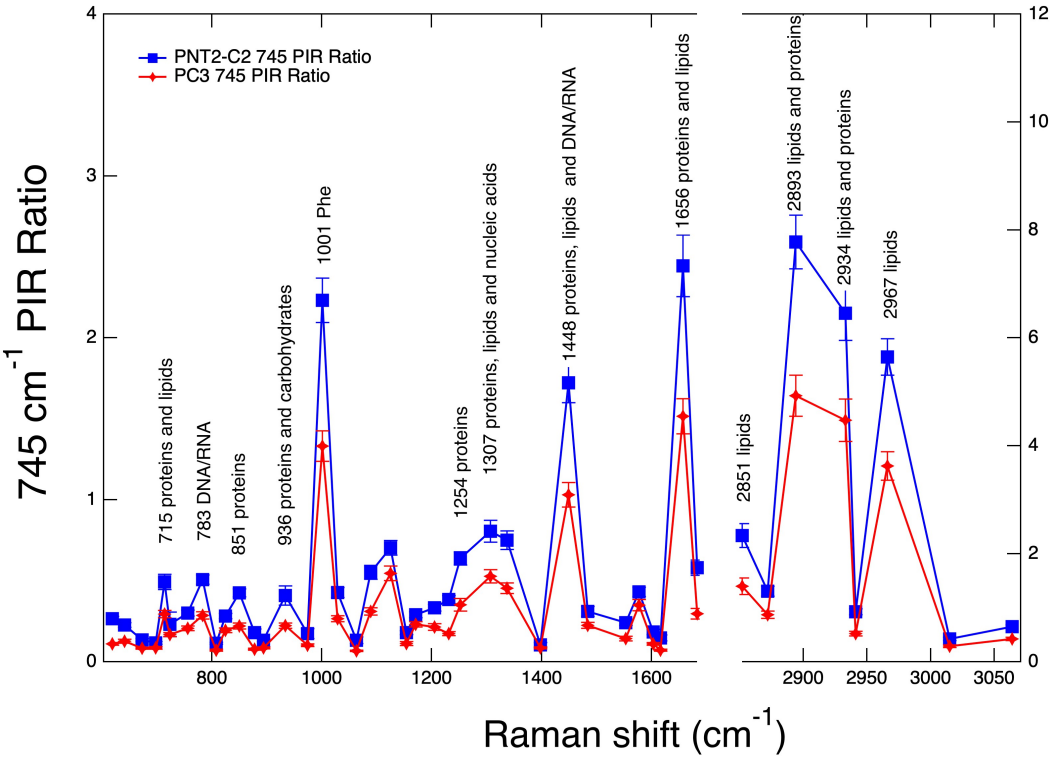


Figure 5.73: The PIR panel for the 745 cm^{-1} peak intensity ratio demonstrates statistically significant separation across the fingerprint region and high wavenumber region PNT2-C2 and PC3, achieving 95% total separation. The 745 cm^{-1} peak is attributed to thymine ring breathing mode. It can be seen that the PIR profile of PC3 relative to the 745 cm^{-1} peak is statistically lower than that of the PNT2-C2 PIR profile at every peak position in both the fingerprint and high wavenumber region. This suggests that there is a higher relative intensity of the thymine ring breathing mode in PC3. It can be seen in the graph that there is a significantly larger ratio of thymine to proteins, lipids, carbohydrates and nucleic acids. This evidence indicates that an increased relative thymine content at the 745 cm^{-1} peak is a marker of PC3.

As shown in Figure 5.73, the PIR profile of PC3 relative to the 745 cm^{-1} peak is statistically lower than that of the PNT2-C2 PIR profile at every peak position in both the fingerprint and high wavenumber region. This indicates a higher relative intensity of Thymine ring breathing mode in PC3 than in PNT2-C2. It is also possible to determine markers to illustrate this, using select peaks in the PNT2-C2 and PC3 profile. For example, a proposed biomarker of PC3 could be a lower ratio of the 1001 cm^{-1} phenylalanine ring breathing mode to the 745 cm^{-1} thymine ring breathing mode. It is widely accepted in literature that an increase in relative DNA content is representative of metastatic cells compared to that of

normal cells [281]. Specifically, thymine has also been described as a marker of metastatic cells [253, 254].

A high degree of separation is observed at the peak 1234 cm^{-1} (98%) outside of the propagated SE uncertainty relating to each PIR between PNT2-C2 and PC3. However, it must be noted that the 1234 cm^{-1} peak overlaps with the 1254 cm^{-1} peak. The 1254 cm^{-1} peak generated 41% separation outside of the propagated SE uncertainty, relating to each PIR between PNT2-C2 and PC3. It is thought that a change in one of these peaks will affect the other at peak fitting stage. It was seen in the mean profile subtraction, Figure 5.61, that there was an observed difference at the 1254 cm^{-1} peak but not at the 1234 cm^{-1} peak. Furthermore, at the PCA stage, the largest source of variation captured between PNT2-C2 and PC3 was at the 1254 cm^{-1} peak location, Figure5.66. In contrast, the 1234 cm^{-1} peak was not observed to be contributing to the separation at the PCA stage in loadings 3 or 4, Figures 5.66 and 5.67 respectively. It is for these reasons that the PIR analysis shall focus on the 1254 cm^{-1} peak and not the 1234 cm^{-1} peak, despite the higher degree of separation associated with it at the PIR stage. It should be noted that a high degree of separation is not always indicative of a biomarker or significant results, and that peaks with lower separations at the PIR stage still have important biomolecular information stored in them. The 1254 cm^{-1} band is representative of amide III : CH_2 wagging. The 1254 cm^{-1} peak PIR panel is investigated further in Figure 5.74

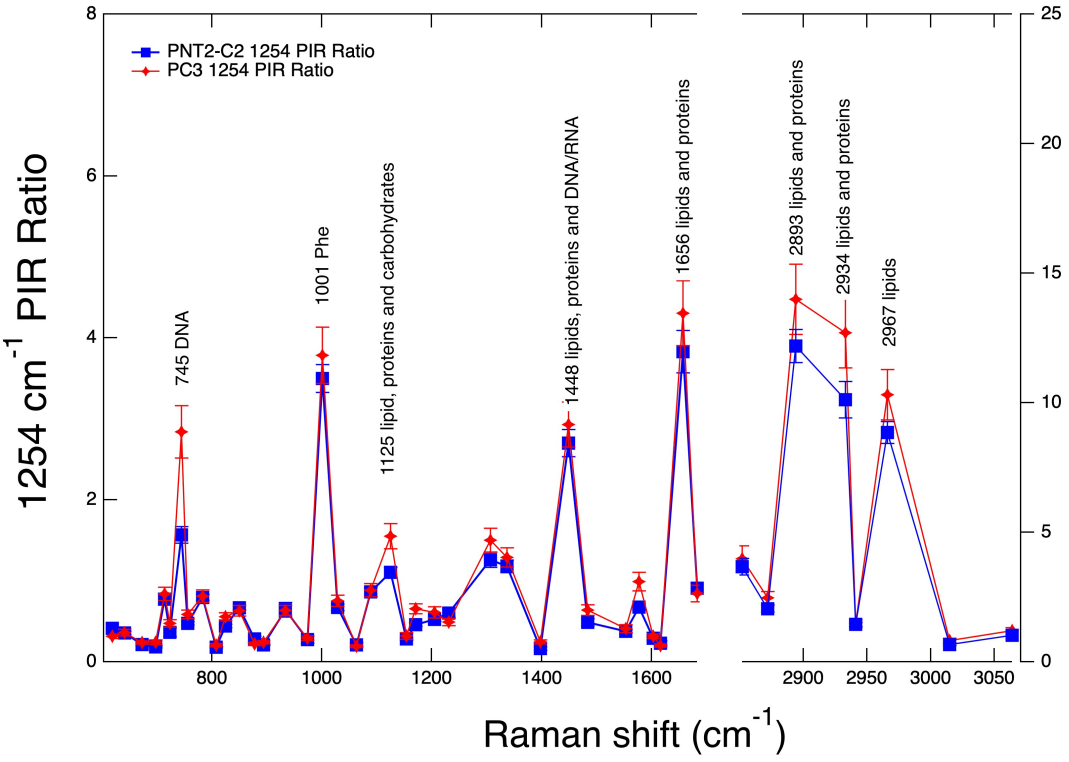


Figure 5.74: The PIR panel for the 1254 cm^{-1} peak intensity ratio demonstrates statistically significant separation across the fingerprint region and high wavenumber region PNT2-C2 and DU145, achieving 41% total separation. The 1254 cm^{-1} peak is attributed to amide III : CH_2 wagging. It can be seen that the PIR profile of PC3 relative to the 1254 cm^{-1} peak is very similar to that of the PNT2-C2 PIR profile in both the fingerprint and high wavenumber region but with a few exceptions. A relatively large separation outside of uncertainty is occurring at the 745 cm^{-1} peak, where the associated PIR is higher in PC3. This suggests that there is a lower relative intensity of the 1254 cm^{-1} peak in PC3 compared to the 745 cm^{-1} peak (thymine). Additionally, separation is occurring outside of uncertainty at the 1125 cm^{-1} peak, where the associated PIR is also higher in PC3. This suggests that there is a lower relative intensity of the 1254 cm^{-1} peak in PC3 compared to the 1125 cm^{-1} peak (protein C-N stretch, phospholipids).

Figure 5.74 shows the PIR panel for the 1254 cm^{-1} peak. It was determined that there were two main sources of separation occurring relative to the 1254 cm^{-1} peak between PNT2-C2 and PC3. As seen in Figure 5.74, the largest separation outside of uncertainty is occurring at the 745 cm^{-1} peak, where the associated PIR is higher in PC3. This suggests that there is a lower relative intensity of the 1254 cm^{-1} peak in PC3 compared to the 745 cm^{-1} thymine peak. Additionally, there is separation outside of uncertainty occurring at the 1125 cm^{-1} peak, where

the associated PIR is also higher in PC3. This suggests that there is a lower relative intensity of the 1254 cm^{-1} peak in PC3 compared to the 1125 cm^{-1} peak (protein C-N stretch, phospholipids). It is thought that both these relationships could be exploited as biomarkers, to differentiate between PNT2-C2 and PC3. The observed variance associated with the 1254 cm^{-1} peak in the mean subtraction, PCA and the identification of biomarkers suggest that there is a difference in the relative amide III contribution in PC3. Changes in secondary protein structure have been associated with malignancy in previous Raman studies [224].

The degree of separation of PIRs observed in the high wavenumber region is comparatively lower than the fingerprint region. It was seen at the mean profile subtraction stage in Figure 5.62 that there were not significant differences attributed to peaks in the high wavenumber region. Information gained from PCA analysis in loading 2 5.71, determined that the most variance was occurring at the 2934 cm^{-1} peak. Therefore, the 2934 cm^{-1} peak is investigated to consider the high wavenumber region between PNT2-C2 and PC3. The 2934 cm^{-1} peak generated 39% separation outside of the propagated SE uncertainty relating to each PIR between PNT2-C2 and PC3.

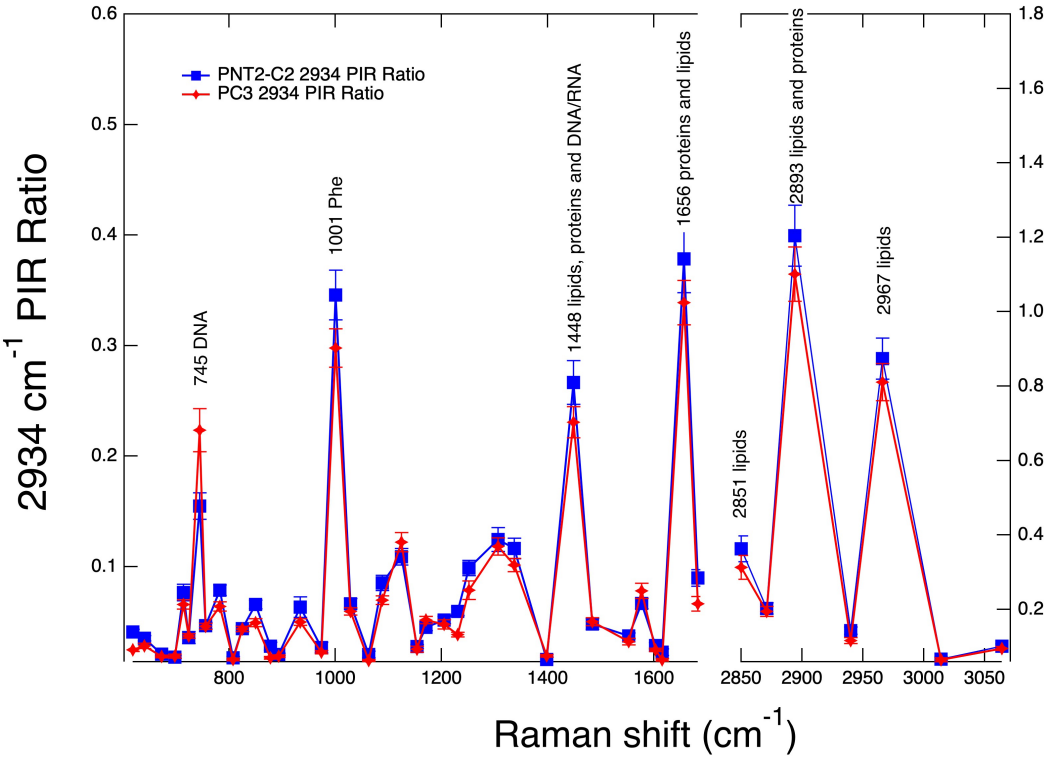


Figure 5.75: The PIR panel for the 2934 cm^{-1} peak intensity ratio demonstrates statistically significant separation across the fingerprint region and high wavenumber region, achieving 39% total separation. The 2934 cm^{-1} band is representative of CH_2 asymmetric stretching and CH_3 symmetric stretching. The largest relative separation is occurring in the fingerprint region, observed at the 745 cm^{-1} thymine ring breathing peak. There is a relative decrease in intensity of the 2934 cm^{-1} peak relative to the 745 cm^{-1}. Thus there is a lower relative ratio of CH_2 asymmetric stretching and CH_3 symmetric stretching to thymine in PC3. Conversely, there was separation identified at the 1001 cm^{-1} peak, however, the relative ratio of the 1001 cm^{-1} peak was higher in PNT2-C2. This evidence indicates a higher relative ratio of CH_2 asymmetric stretching and CH_3 symmetric stretching to phenylalanine in PNT2-C2. In the high wavenumber region no separation is observed using the 2934 cm^{-1} peak intensity ratio.

The PIR panel for the 2934 cm^{-1} peak intensity ratio shown in Figure 5.75 shows that the largest relative separation is occurring in the fingerprint region, observed at the 745 cm^{-1} thymine ring breathing peak. This indicates that there is a lower relative ratio of CH_2 asymmetric stretching and CH_3 symmetric stretching to thymine in PC3. Conversely, evidence also presented in Figure 5.75 indicates a higher relative ratio of CH_2 asymmetric stretching and CH_3 symmetric stretching to phenylalanine in PNT2-C2. At this point of analysis it should be noted that

the very strong relative intensity of the 745 cm^{-1} thymine peak is a defining characteristic of PC3.

5.6 P4E6

P4E6 has not been published as part of a Raman study in a formal journal, as far as the author is aware. There have been two masters by research projects that have been carried out in our own research group by Kershaw (2017) and Faulkner (2019) involving the Raman spectroscopy of P4E6 [208, 285]. However, as these are not published works they are not considered in this comparison.

5.6.1 PNT2-C2 vs P4E6 mean profile subtraction

The normalised and baselined Raman signal for PNT2-C2 and P4E6 are displayed for the fingerprint region in Figure 5.76 and the high wavenumber region is shown in Figure 5.77. Both cell lines have been normalised with respect to the total spectrum area. The corresponding mean subtraction profile (PNT2-C2 minus P4E6) is shown for the fingerprint region in Figure 5.76 and the high wavenumber region in Figure 5.77.

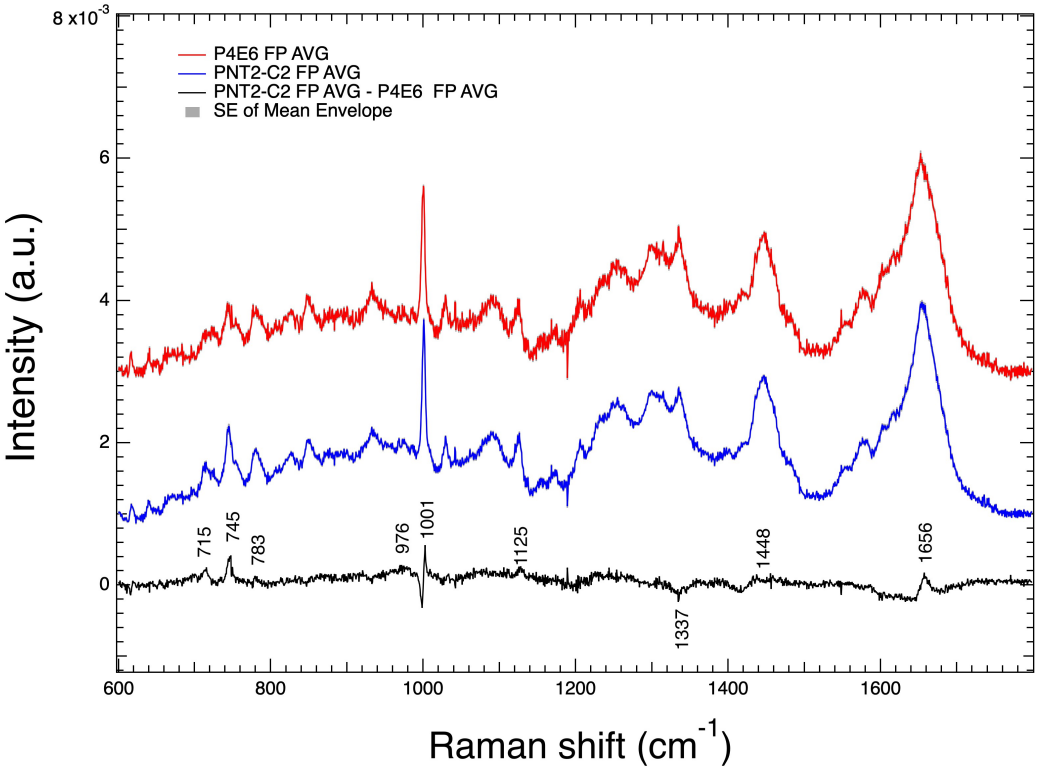


Figure 5.76: A mean subtraction of the normalised profiles of P4E6 fingerprint from PNT2-C2 fingerprint region has been carried out in the region 600-1800 cm^{-1}. There is an emergence of peaks 715 cm^{-1} (proteins and lipids), 745 cm^{-1} (DNA), 783 cm^{-1} (DNA/RNA), 1001 cm^{-1} (proteins), 1125 cm^{-1} (lipids, proteins and carbohydrates), 1337 cm^{-1} (proteins and DNA/RNA), 1448 cm^{-1} (lipids, proteins and DNA/RNA) and 1656 cm^{-1} (proteins and lipids) . The DNA Raman band 745 cm^{-1} is notably much less pronounced in P4E6 when compared to PNT2-C2. The band 715 cm^{-1} peak (proteins and lipids) also experienced a decrease in intensity in P4E6. Another observation is the changes in profile around the 976 cm^{-1} peak. This peak could not be fitted in P4E6 in Section 5.1 due to low signal-to-noise. Additionally, fluctuations in the mean subtraction around the peaks 1335 cm^{-1} and 1448 cm^{-1} suggest variation in lipid/protein structure. Finally, there is a marked difference in the 1656 cm^{-1} peak suggesting that there may be further variation to the secondary protein structure and DNA conformation. These results will be analysed in further detail using a combination of PCA/LDA and peak intensity ratios.

It can be seen in Figure 5.76 that there is an emergence of the peak 745 cm^{-1} in the mean subtraction profile. This peak is representative of the thymine breathing mode. Further variation around the peaks 715 cm^{-1} and 783 cm^{-1} indicated further variation in relative DNA differences between the samples. However, at this stage it is hard to say how these changes translate to relative

intensities between peaks. A considerable result from Figure 5.76 is the raised profile around the 976 cm^{-1} peak. The 976 cm^{-1} band could not be fitted, as detailed in Section 5.1, suggesting that there is a possibility that this could be a biomarker between PNT2-C2 and P4E6. Further variation is observed throughout the profile. Specific changes in the mean subtraction around the peaks 1335 cm^{-1} and 1448 cm^{-1} suggest variation in lipid/protein structure. Lastly, a marked difference in the 1656 cm^{-1} peak suggests that there may be further variation to secondary protein structure and DNA conformation.

The high wavenumber subtraction between PNT2-C2 and P4E6 is now considered in Figure 5.77.

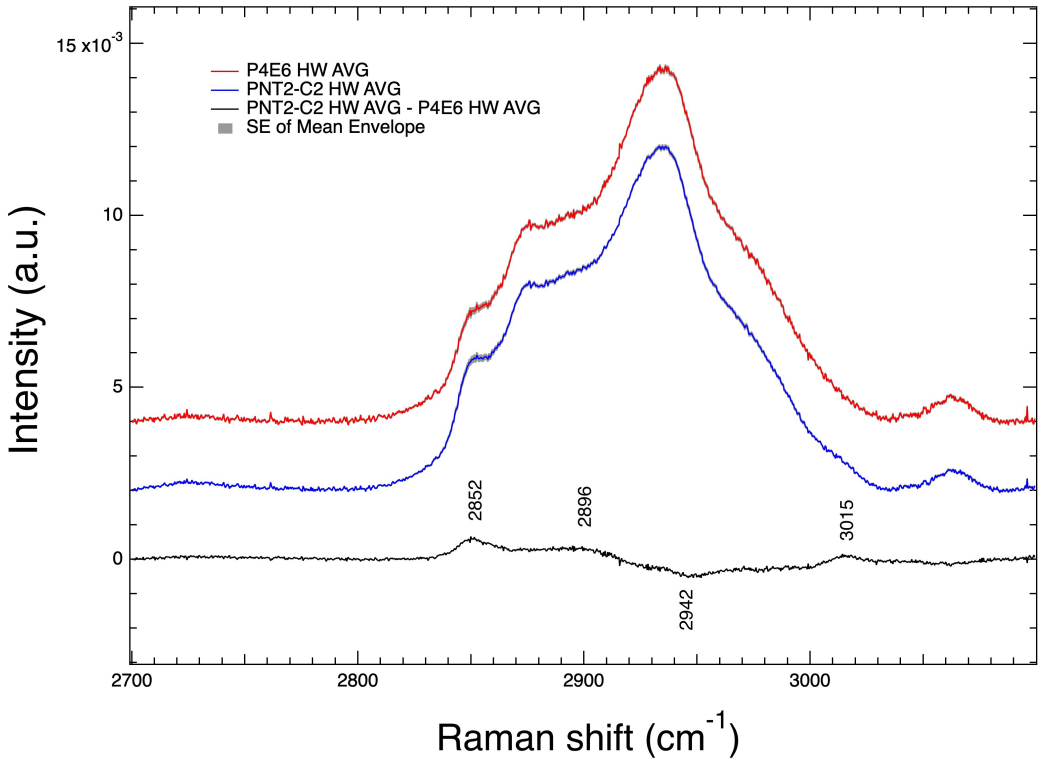


Figure 5.77: Differences in the mean profile subtraction between the normalised PNT2-C2 and P4E6 cell line are shown. There were four peaks identified in the mean profile subtraction: 2852 cm^{-1} (lipids), 2873 cm^{-1} (lipids and proteins), 2942 cm^{-1} (lipids and proteins) and 3015 cm^{-1} (lipids). While this is only preliminary analysis, it points to a difference in total unsaturated fatty acids (3015 cm^{-1}) and fatty acids (2852 cm^{-1}) as well as protein/lipid variation across the bands 2873 cm^{-1} and 2942 cm^{-1}. This must be validated using PCA/LDA and peak intensity ratio analysis before any conclusion can be made.

When considering the intensity profile of the mean subtraction in Figure 5.77, there is an increase in intensity at the 2852 cm^{-1} peak. This is indicative of a change in the CH_2 symmetric stretch in lipids inferring that peak intensity ratio analysis will potentially observe variation of the ratio of total fatty acids (TFA) to total unsaturated fatty acids (TUFA). This is dependent on the 3015 cm^{-1} peak which represents unsaturated fatty acids. At this stage it cannot be determined whether the changes observed in the mean subtraction profile of the high wavenumber region will result in a statistically significant biomarker being found.

5.6.2 PNT2-C2 vs P4E6 PCA comparison

PCA analysis between the PNT2-C2 and P4E6 fingerprint region is able to capture the variance of the first six principal components (PCs): PC1 = 31.1%, PC2 = 10.4%, PC3 = 5.9%, PC4 = 3.2%, PC5 = 2.1% and PC6 = 2.3%. This in turn means that the first three principal components capture 47.4% and the first six capture 54.6% of the total variance of the sample. Cumulatively, 285 principal components are required to capture total variance of the sample. The 3D PCA plots using the first three principal components associated with this comparison are shown in Figures 5.78, 5.79 and 5.80. A further comparison using PC2, PC3 and PC4 is shown in Figure 5.81.

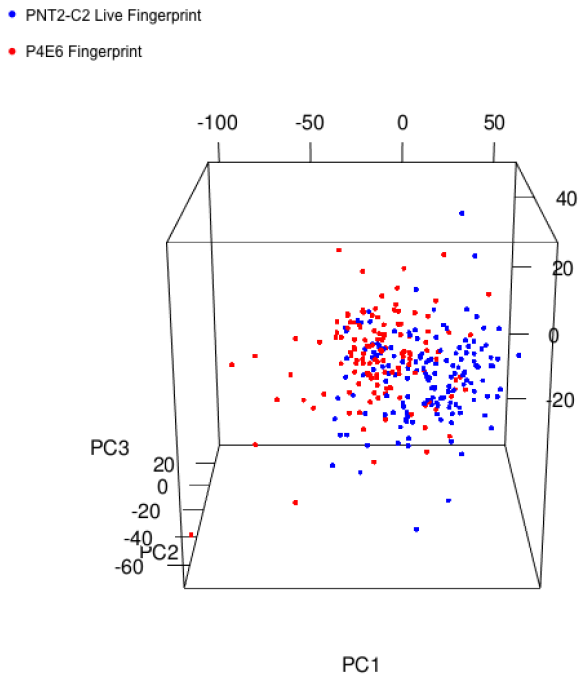


Figure 5.78: The PCA comparison of the fingerprint region of PNT2-C2 and P4E6 Live is displayed in 3D PCA space using the first three principal components. At this angle of observation there is separation occurring along PC1. There is no separation observed along PC3.

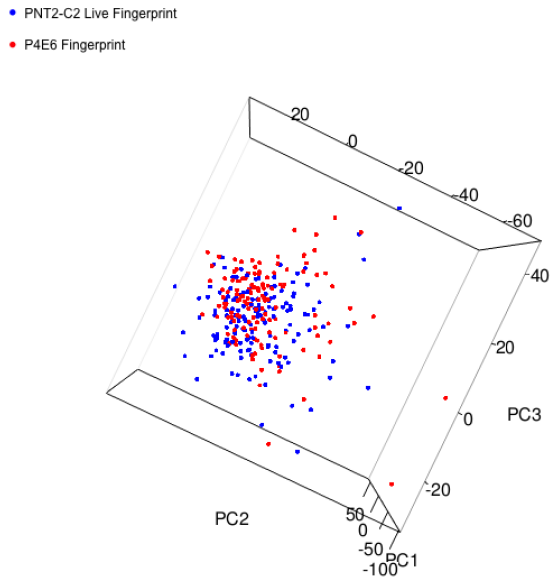


Figure 5.79: The PCA comparison of the fingerprint region of PNT2-C2 and P4E6 Live is displayed in 3D PCA space using the first three principal components. At this angle of observation there is no separation observed in PC2 or PC3

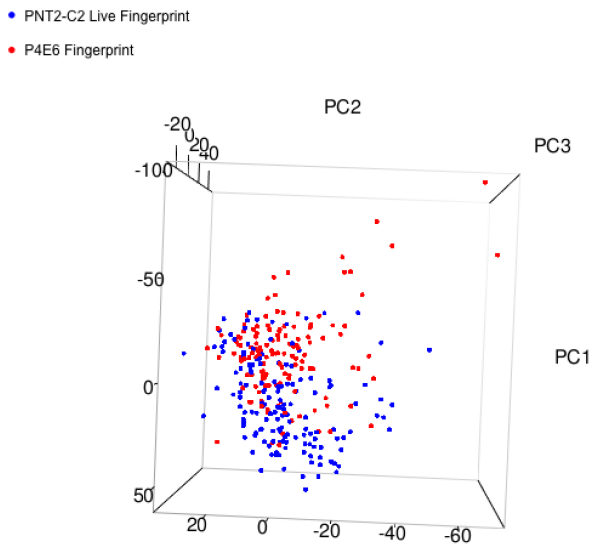


Figure 5.80: Another rotation of PCA space using the first three principal components allows comparison of the fingerprint region of PNT2-C2 and P4E6. At this angle there is no separation occurring along PC2. It can be seen that there is separation occurring along PC1.

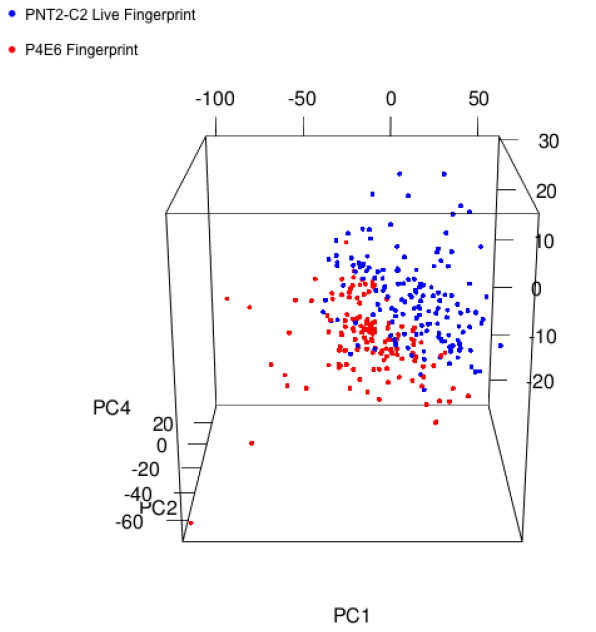


Figure 5.81: In this iteration of PCA analysis between PNT2-C2 and P4E6 the 3D PCA space is plotted using PC1, PC2 and PC4 for the fingerprint region. This is to investigate the influence of PC4 in terms of separation . It can be seen that the separation is occurring at an angle between PC1 and PC4.

Figure 5.78 demonstrates the fingerprint PCA space using the first three principal components. From this orientation it is observed that there is separation occurring along PC1 but not PC3. Analysis of two more rotations of PCA space, using the first three principal components, in Figures 5.79 and 5.80, showed that there was no separation occurring along PC2 or PC3. Additionally, the 3D space using PC1,PC2 and PC4, demonstrated that there was separation occurring at an angle along PC1 and PC4, in Figure 5.81. The separation demonstrated by PC1 and PC4 is analysed by plotting loading 1 and loading 4. The profile of loading 1 is shown in 5.82. The profile of loading 4 is shown in Figure 5.83.

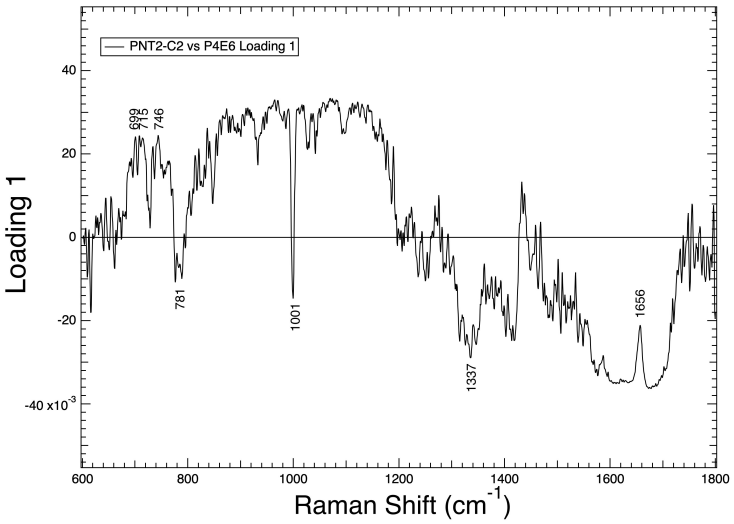


Figure 5.82: Loading 1 capturing 31.1% of total population variance in the fingerprint region where comparison of PNT2-C2 vs P4E6 is displayed. Key peaks attributed to biological assignments were located at 699 cm^{-1} (proteins), 715 cm^{-1} (proteins and lipids), 745 cm^{-1} (DNA), 781 cm^{-1} (DNA/DNA), 1001 cm^{-1}cm (proteins), 1337 cm^{-1} (proteins DNA/RNA), 1656 cm^{-1} (proteins and lipids).

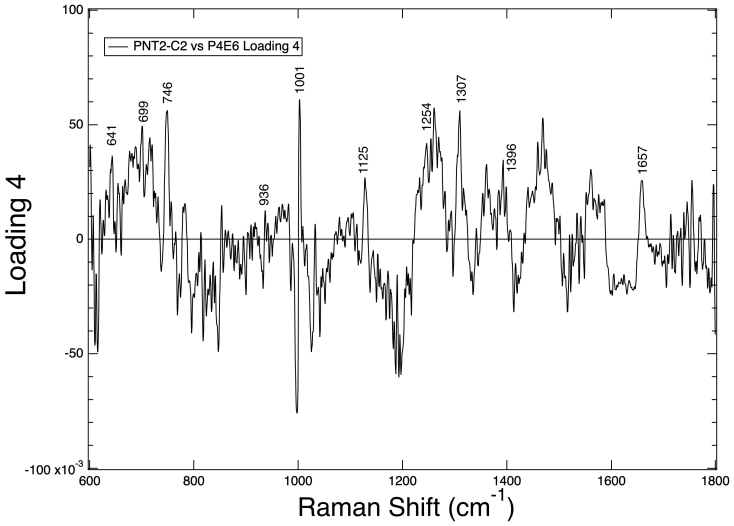


Figure 5.83: Loading 4, capturing 3.2% of total population variance, of the fingerprint region comparison between PNT2-C2 vs P4E6, is displayed. Key peaks attributed to biological assignments were located at 641 cm^{-1} (proteins), 699 cm^{-1} (proteins), 745 cm^{-1} (DNA), 936 cm^{-1} (proteins and carbohydrates), 1125 cm^{-1}cm (lipids, proteins and carbohydrates), 1396 cm^{-1} (proteins), 1657 cm^{-1} (proteins and lipids).

The Figures 5.88 and 5.91 show loadings 1 and 4 corresponding to the separation of PNT2-C2 and P4E5 in the fingerprint region. In loading 1, peaks linked to biomolecular signatures were found at 699 cm^{-1} (proteins), 715 cm^{-1} (proteins and lipids), 745 cm^{-1} (DNA), 781 cm^{-1} (DNA/DNA), 1001 cm^{-1}cm (proteins), 1337 cm^{-1} (proteins DNA/RNA), 1656 cm^{-1} (proteins and lipids). Furthermore, in loading 2, peaks linked to biomolecular assignments were found at 699 cm^{-1} (proteins), 715 cm^{-1} (proteins and lipids), 745 cm^{-1} (DNA), 781 cm^{-1} (DNA/DNA), 1001 cm^{-1}cm (proteins), 1337 cm^{-1} (proteins DNA/RNA), 1656 cm^{-1} (proteins and lipids). These peaks, responsible for biomolecular contribution to separation, will be investigated further using PIR ratios in Section 5.6.3.

The high wavenumber region is now considered between PNT2-C2 and P4E6. PCA analysis of the high wavenumber region between PNT2-C2 and P4E6 is able to capture the variance of the first six principal components (PCs): PC1 = 43.6%, PC2 = 12.5%, PC3 = 6.6%, PC4 = 4.1%, PC5 = 3.3% and PC6 = 2.3%. This in turn means that the first three principal components in the high wavenumber region capture 62.7% and the first six capture 71.5% of the total variance of the sample.

Cumulatively, 273 principal components are required to capture total variance of the sample. The 3D PCA plots, using the first three principal components associated with this comparison, are shown in Figures 5.84, 5.85 and 5.86. A further comparison using PC2, PC3 and PC4 is shown in Figure 5.87.

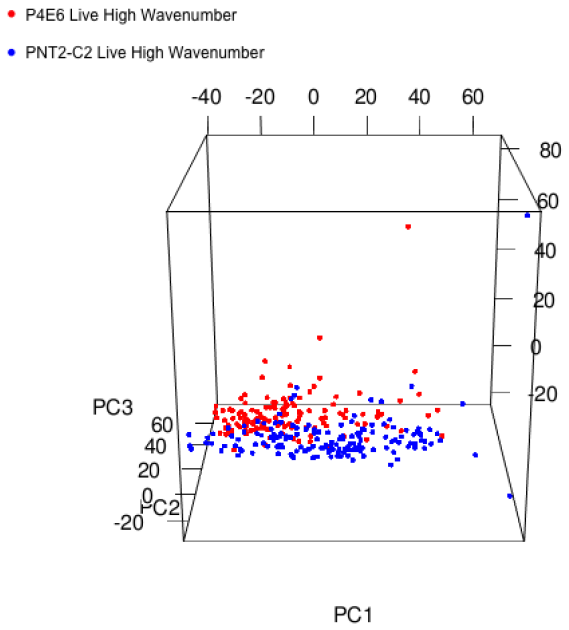


Figure 5.84: The PCA comparison of the high wavenumber region of PNT2-C2 and P4E6 Live is displayed in 3D PCA space using the first three principal components. At this angle of observation there is separation occurring along PC3. There is no separation observed along PC1.

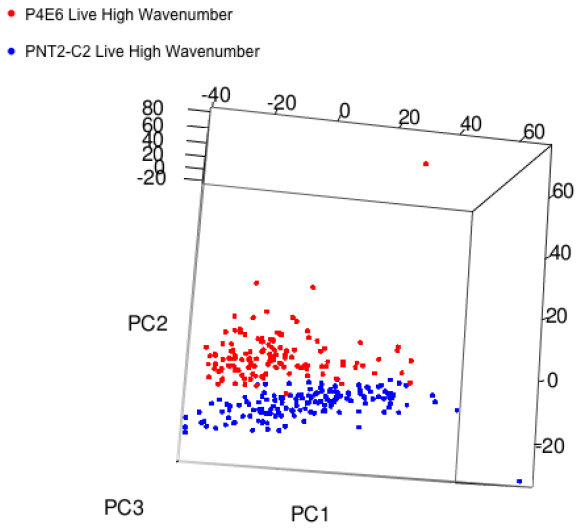


Figure 5.85: Another PCA comparison of the high wavenumber region of PNT2-C2 and P4E6 from a different angle is displayed in 3D PCA space using the first three principal components. At this angle of observation there is separation occurring at an angle along PC1 and PC2

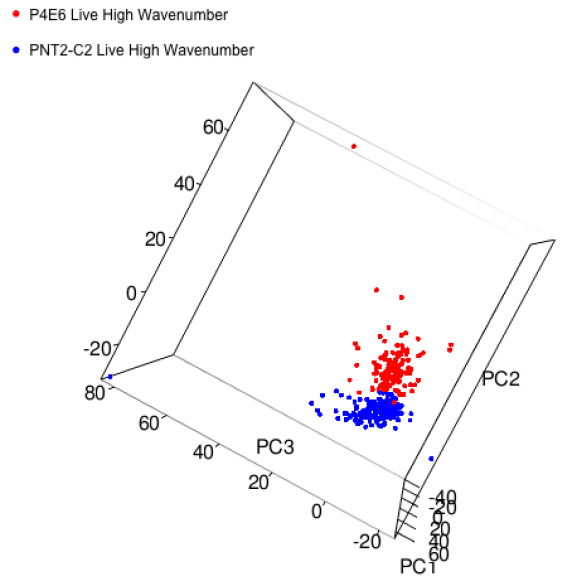


Figure 5.86: 3D PCA space, using the first three principal components in the high wavenumber region, is shown between PNT2-C2 and P4E6. At this angle of observation there is separation occurring at an angle along PC2 and PC3.

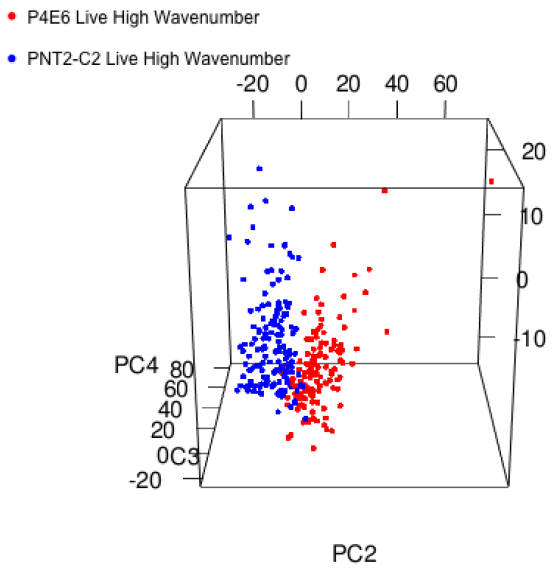


Figure 5.87: A final PCA comparison of the high wavenumber region of PNT2-C2 and P4E6. The 3D PCA space is now generated using PC2, PC3 and PC4. From this view of the 3D PCA space it is clear that separation is occurring at an angle along PC2 and PC4

Figures 5.84, 5.85 and 5.86 show multiple rotations of the PCA space using the first three principal components between PNT2-C2 and P4E6 in the high wavenumber region. It can be seen in these graphs that separation is occurring in all principal components. Furthermore, 5.87 shows that using a combination of PC2, PC3 and PC4, it is possible to show that separation is occurring at an angle between PC2 and PC4. Thus all four loadings corresponding to PC1, PC2, PC3 and PC4, capturing 66.8% of the total variance of the sample, are investigated below in Figures 5.88-5.91.

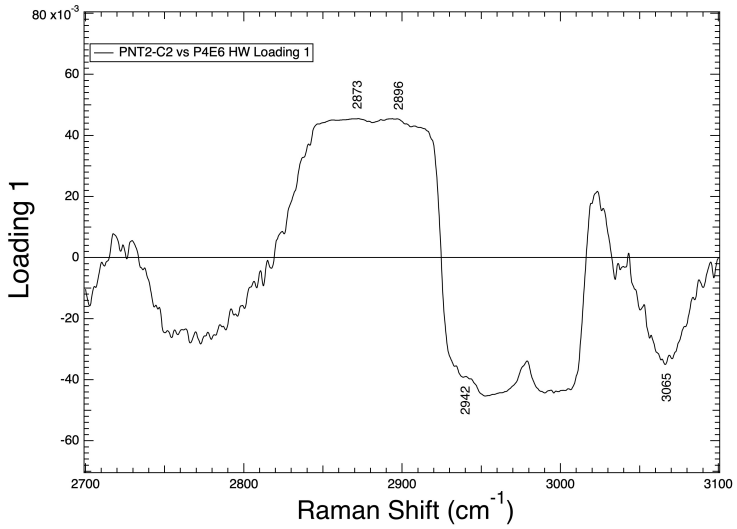


Figure 5.88: Loading 1 of the high wavenumber region comparison of PNT2-C2 Live vs P4E6 is displayed, capturing 43.6% of the variance between samples. Key peaks attributed to biological assignments spanning DNA/RNA, protein and lipids are found at: 2873 cm^{-1} (lipids and proteins), 2896 cm^{-1} (lipids and proteins), 2942 cm^{-1} (lipids and proteins) and 3065 cm^{-1} (lipids).

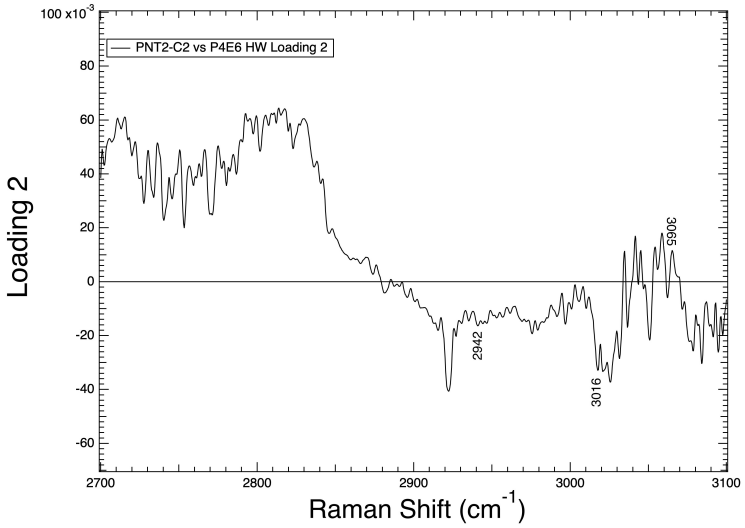


Figure 5.89: Loading 2 of the high wavenumber region comparison of PNT2-C2 Live vs P4E6 is displayed, capturing 12.5% of the variance between samples. Key peaks attributed to biological assignments are found at 2942 cm^{-1} (lipids and proteins) , 3015 cm^{-1} (lipids) and 3065 cm^{-1} (lipids).

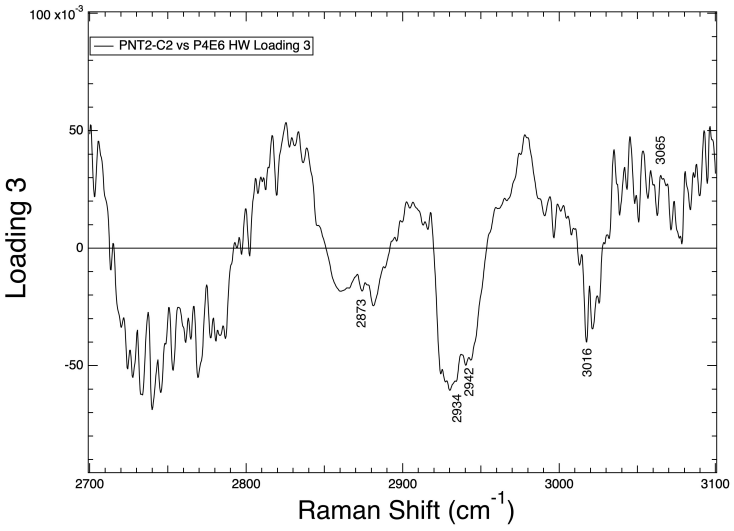


Figure 5.90: Loading 3 of the high wavenumber region comparison of PNT2-C2 Live vs P4E6 is displayed, capturing 6.6% of the variance between samples. Key peaks attributed to biological assignments are found at: 2873 cm^{-1} (lipids and proteins) , 2934 cm^{-1} (proteins and lipids), 2942 cm^{-1} (lipids and proteins), 3015 cm^{-1} (lipids) and 3065 cm^{-1} (lipids).

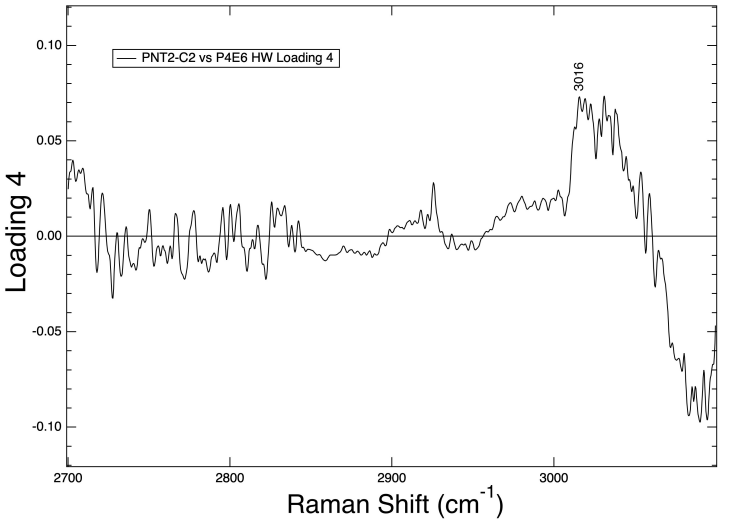


Figure 5.91: Loading 4 of the high wavenumber region comparison of PNT2-C2 Live vs P4E6 is displayed, capturing 4.1% of the variance between samples. Only one peak could be attributed to biological assignments found at 3015 cm^{-1} (lipids).

In Figures 5.88-5.91 there were 4 loadings analysed that generated 66.8% of the total population variance between the high wavenumber regions of PNT2-C2 and P4E6. In loading 1 there were 4 peaks linked to biomolecular signatures, 2873 cm^{-1} (lipids and proteins), 2896 cm^{-1} (lipids and proteins), 2942 cm^{-1} (lipids

and proteins) and 3065 cm^{-1} (lipids). However, whilst these peaks are thought to contribute to the variance between samples, the profile of loading 1 is indicative of background variance between the high wavenumber spectrums of PNT-C2 and P4E6. In loading 2, the peaks identified with biomolecular assignments were: 2942 cm^{-1} (lipids and proteins) , 3015 cm^{-1} (lipids) and 3065 cm^{-1} (lipids). Similar to loading 2, there was considerable background variance between profiles captured in the region 2700-2800 cm^{-1}. However, it is thought that the biomolecular assignments still contribute to the separation of the samples. Thirdly, loading 3 was analysed, finding peaks related to biochemical signatures at 2873 cm^{-1} (lipids and proteins) , 2934 cm^{-1} (proteins and lipids), 2942 cm^{-1} (lipids and proteins), 3015 cm^{-1} (lipids) and 3065 cm^{-1} (lipids). Lastly, in loading 4 the only peak attributed to a biomolecular signature was identified at 3015 cm^{-1}. These peaks, responsible for biomolecular contribution to separation, will be investigated further using PIR ratios in Section 5.6.3.

Finally, a PCA-LDA leave-one-out cross-validation was used to determine the predictive classification of PNT2-C2 and P4E6 cells. The number of principal components included was based on using Kaiser's criterion, which consists of using only principal components with variance greater than 1 percent. For PNT2-C2 vs P4E6 the PCA-LDA classification tool correctly classified 97.6% cells in the fingerprint region using 9 PCs and 94.4% of cells in the high wavenumber region using 10 PCs.

5.6.3 Peak intensity ratio analysis PNT2-C2 vs P4E6

The peak intensity ratios have been calculated using the methods described in Section 3.6.3 for PNT2-C2 live vs P4E6. The complete set of results used to calculate the separation outside of the propagated SE uncertainty, relating to each PIR compared between PNT2-C2 and P4E6 for the fingerprint and high wavenumber region is shown in Table C.6 in the Appendix. There were only three peaks that achieve separation of 90% and above at the PIR stage, 715 cm^{-1} (91%), 745 cm^{-1} (91%) and 3015 cm^{-1} (93%).

The 715 cm^{-1} peak demonstrates 91% separation outside of the propagated SE uncertainty relating to each PIR between PNT2-C2 and P4E6. In the initial mean profile subtraction, the 715 cm^{-1} peak was identified as a source of variance between PNT2-C2 and P4E65.76. The 715 cm^{-1} peak is also attributed to amino acid methionine ν(C-S) *trans*, as well as the $CN^{+}(CH_3)_3$ stretch in lipids. The PIR panel for the 699 cm^{-1} peak is shown in Figure 5.21 as well as the 715 cm^{-1} peak in Figure5.22.

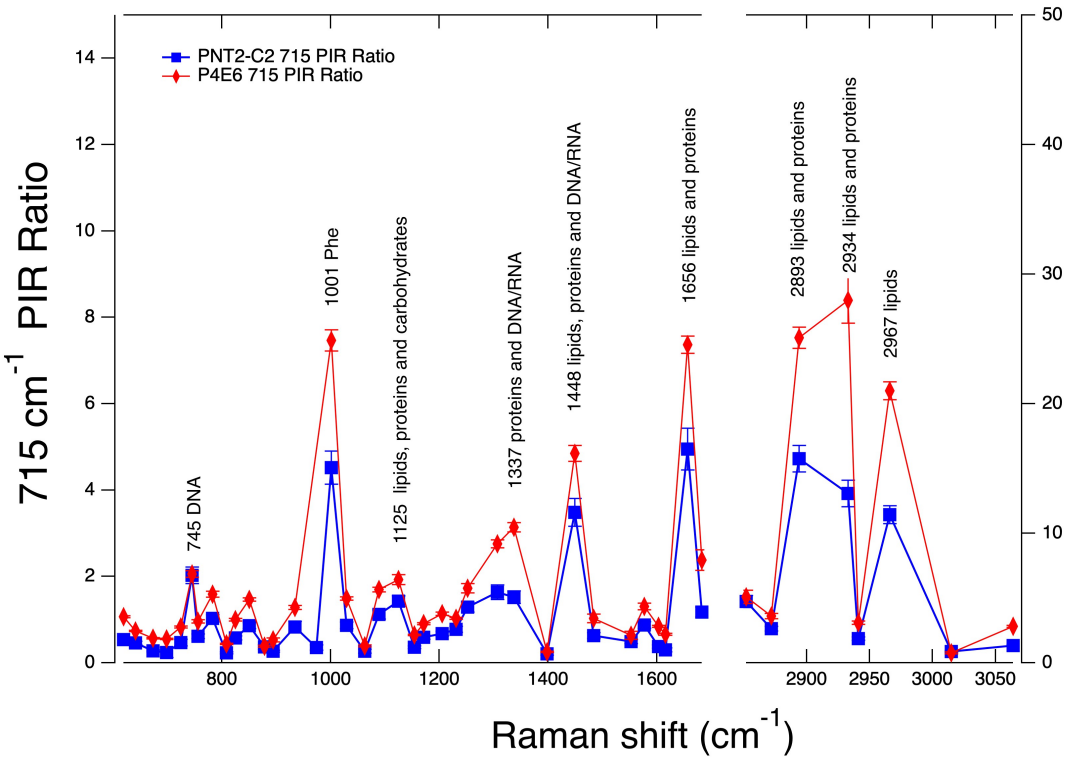


Figure 5.92: The PIR panel for the 715 cm^{-1} peak intensity ratio demonstrates statistically significant separation across the fingerprint region and high wavenumber region, achieving 91% total separation. The 715 cm^{-1} peak is attributed, specifically, to amino acid methionine ν(C-S) *trans* as well as the $CN^{+}(CH_3)_3$ stretching in lipids. It can be seen that by using the 715 cm^{-1} peak as a peak intensity ratio marker, that there is a lower relative content of amino acid methionine and $CN^{+}(CH_3)_3$ stretching in PNT2-C2.

From Figure 5.92, it is seen that there is a lower relative presence of amino acid methionine ν(C-S) *trans* and $CN^{+}(CH_3)_3$ stretching in lipids in P4E6 compared to PNT2-C2. It is also noted that there is no separation occurring between the 715 cm^{-1} peak intensity and 745 peak cm^{-1} intensity, another discriminating

marker investigated in this section. Nonetheless, the 715 cm^{-1} peak should be considered a robust marker for discrimination between PNT2-C2 and P4E6 for discriminating at a population level.

The 745 cm^{-1} peak was also able separate 91% of PIRs between PNT2-C2 and P4E6 outside of the propagated SE uncertainty. The 745 cm^{-1} peak is generated by a pure Thymine ring breathing mode in DNA. The 745 cm^{-1} peak was identified in the mean subtraction profile in Figure 5.76 as being a source of variation. Furthermore, the 745 cm^{-1} peak was also determined to be contributing to the separation in loadings 1 and 4, Figures 5.82 and 5.83. The PIR panel associated with the 745 cm^{-1} peak is shown in Figure 5.93.

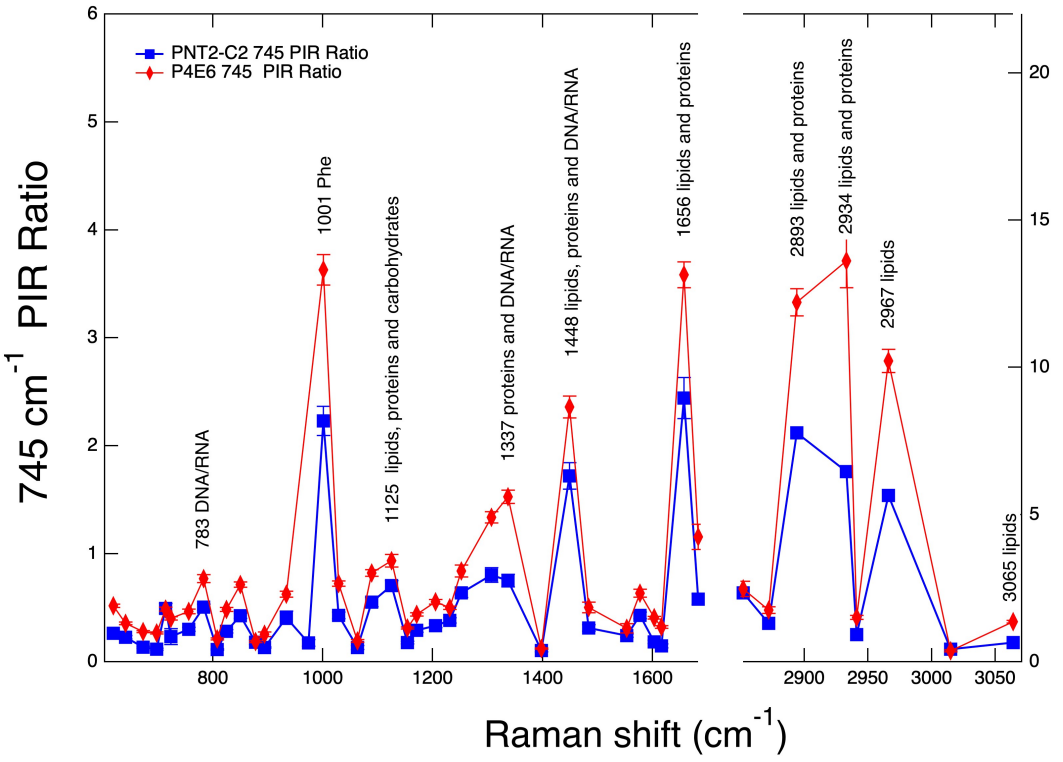


Figure 5.93: The PIR panel for the 745 cm^{-1} peak intensity ratio demonstrates statistically significant separation across the fingerprint region and high wavenumber region PNT2-C2 and P4E6, achieving 91% total separation. The 745 cm^{-1} peak is attributed to thymine ring breathing mode. It can be seen that the PIR profile of PNT2-C2 relative to the 745 cm^{-1} peak is statistically lower than that of the P4E6 PIR profile, with the exception of several peaks that are not statistically separated. These results show that there is a higher relative intensity of the thymine ring breathing mode in PNT2-C2 than P4E6. It can be seen in the graph that there is a significantly larger ratio of thymine to proteins, lipids, carbohydrates and nucleic acids. This evidence indicates that an increased relative thymine content at the 745 cm^{-1} peak is a marker of PNT2-C2

The PIR panel for the 745 cm^{-1} peak intensity ratio in Figure 5.93 demonstrates that there is a higher relative intensity of the thymine ring breathing mode (DNA) in PNT2-C2 than P4E6. It can be seen in the graph that there is a significantly larger ratio of thymine to proteins, lipids, carbohydrates and nucleic acids. This clear separation would suggest that the 745 cm^{-1} peak is a robust biomarker to differentiate between PNT2-C2 and P4E6. This result is not expected; usually an increase in relative DNA content, as seen at the 745 cm^{-1} peak in PNT2-C2, has been shown to be an indicator of malignant cells [253, 254, 281]. However, this does not infer that the P4E6 cell line, exhibiting a

lower relative DNA intensity, is not malignant. Further work could investigate this relationship but it is beyond the scope of this thesis. This thesis is focused on identifying discriminating markers between cell lines.

The 3015 cm^{-1} peak achieved 82% separation outside of the propagated SE uncertainty relating to each PIR compared between PNT2-C2 and P4E6. The 3015 cm^{-1} peak is specifically assigned to unsaturated =CH stretching in lipids. The 3015 cm^{-1} was identified in the mean profile subtraction in Figure 5.77. However, the 3015 cm^{-1} peak was also found to be a source of variance in the PCA stage in loadings 2,3 and 4, and Figures 5.89, 5.90 and 5.91 . The 3015 cm^{-1} peak intensity ratio is explored further using its associated PIR panel, in Figure 5.94.

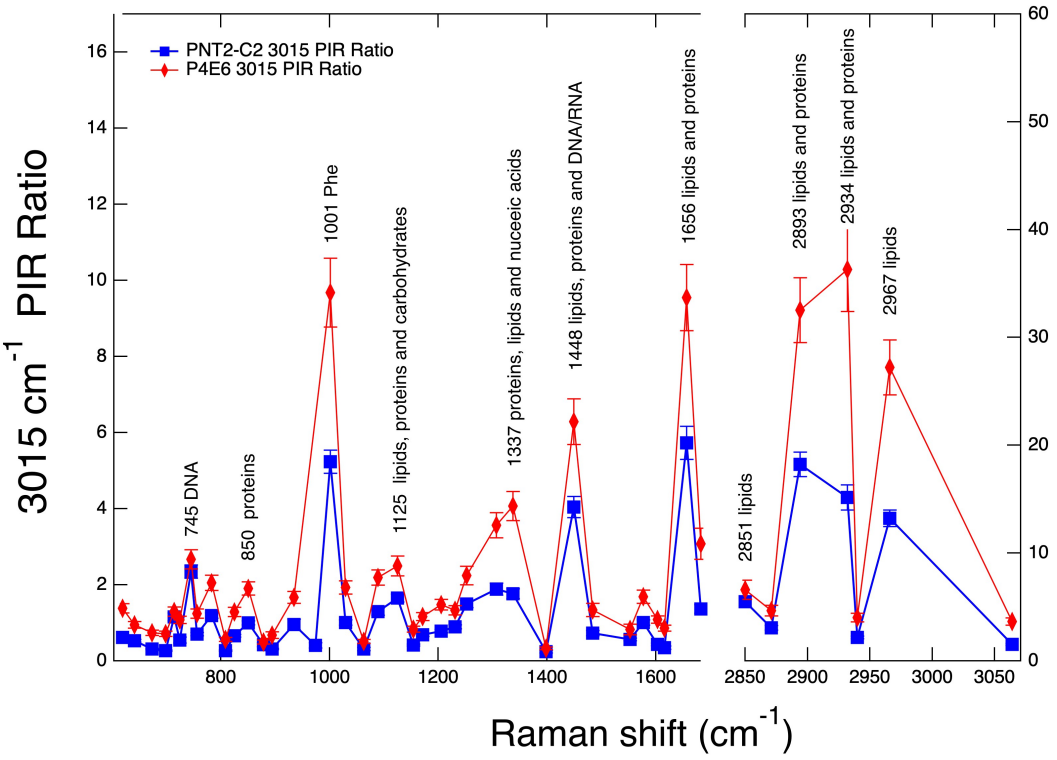


Figure 5.94: The PIR panel for the 3015 cm^{-1} peak intensity ratio demonstrates statistically significant separation across the fingerprint region and high wavenumber region, achieving 82% total separation. The 3015 cm^{-1} peak is attributed to an unsaturated =CH stretch in lipids, unsaturated fatty acids. In the PIR profile of PNT2-C2 and P4E6 it can be seen that the general reduced relative intensity of the PIRs would suggest that there is a higher relative contribution of unsaturated fatty acids in PNT2-C2.

It is determined from Figure 5.94 that there is higher general relative intensity

of the 3015 cm^{-1} peak in PNT2-C2 than in P4E6. A relative increase in the unsaturated lipid content has been previously indicative of malignancy brought about by metabolic changes in the cell [282, 284]. However, as PNT2-C2 is an established normal cell line, this argument is not investigated further.

5.7 The use of Raman spectroscopy to differentiate between standard cell lines

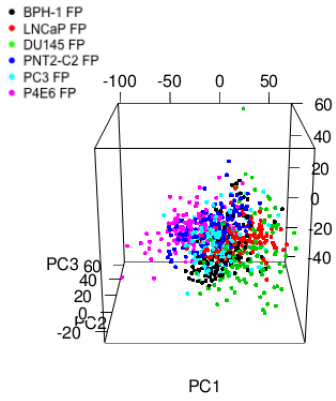
The specific aim of this comparison is to observe the variances between multiple live commercial prostate standard cell lines representative of disease states PNT2-C2 (normal), BPH-1 (benign prostate hyperplasia), LNCaP(metastatic androgen-sensitive), DU145(metastatic androgen-independent), PC3 (metastatic androgen-independent) and P4E6 (localised androgen-sensitive). The in-depth analysis of each cell line will provide the biomolecular detail to be able to determine which biomarkers and PIR panels evolve as a function of disease state, relative to the normal.

5.7.1 PCA/LDA analysis of prostate standard cell lines

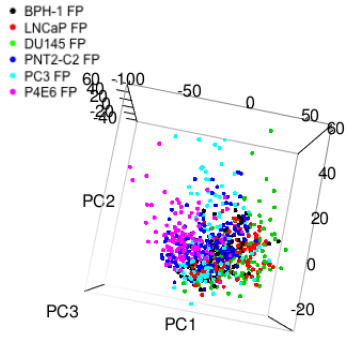
PCA and LDA analysis has been applied to all standard cell lines in the same comparison, to capture significant variance that is occurring in the total population.

The PCA analysis between all the standard cell lines in the fingerprint region is able to ascertain that the variance captured by the first 6 principal components is as follows: PC1 = 31.2%, PC2 = 8.6%, PC3 = 6.8%, PC4 = 5.5%, PC5 = 2.7% and PC6 = 2.6%. This in turn means that the first three principal components capture 46.6% and the first six capture 63.43% of the total variance of the sample. Cumulatively, 299 principal components are required to capture total variance of the sample. Three rotations of PCA space, using the first three principal components, are displayed in Figures 5.95(a), 5.95(b) and 5.95(c). Furthermore, three rotations of LDA space, using 148 principal components, capturing 90% of

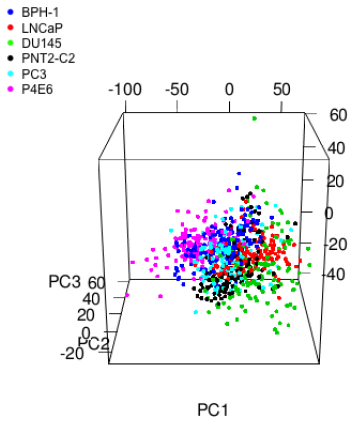
the variance in the sample, are displayed in Figures 5.95(d), 5.95(e) and 5.95(f)



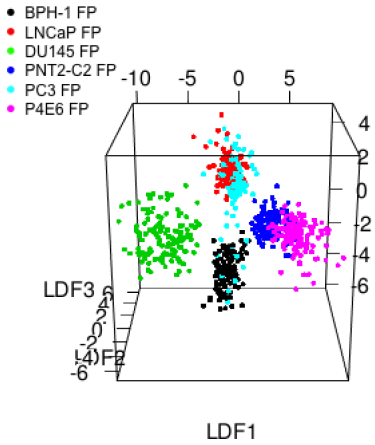
(a) PCA rotation 1



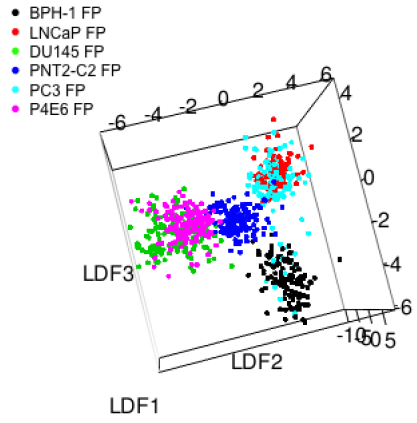
(b) PCA rotation 2



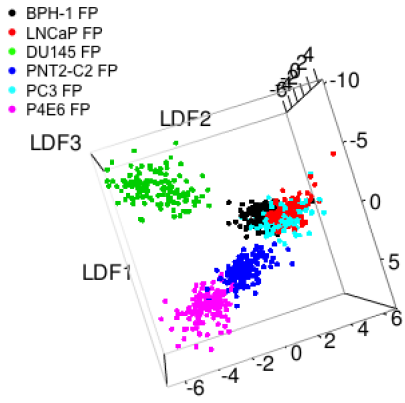
(c) PCA rotation 3



(d) LDA rotation 1



(e) LDA rotation 2



(f) LDA rotation 3

Figure 5.95: Comparison of all standard cell lines in the fingerprint region. PCA space, using the first three principal components, is used to compare all standard cell lines at different angles in Figs. (a), (b) & (c). LDA analysis, using 59 PCs (78.9% population variance capture)has been plotted using LDF1, LDF2 and LDF3, is shown (d), (e) & (f).

The 3D PCA comparison of PNT2-C2 , BPH-1 , LNCaP, DU145, PC3 and P4E6 for the fingerprint region, is shown in Figures 5.95(a), 5.95(b) and 5.95(c). However, although there is separation occurring, it is difficult to observe due to the high number of spectra involved in this comparison. It is determined that there is separation between individual cell lines occurring along PC1, PC2 and PC3. It is therefore possible to determine where the maximum variance of the sample is occurring, using the corresponding loadings. The corresponding loadings for PC1, PC2 and PC3 are shown in Figures 5.96, 5.97 and 5.98.

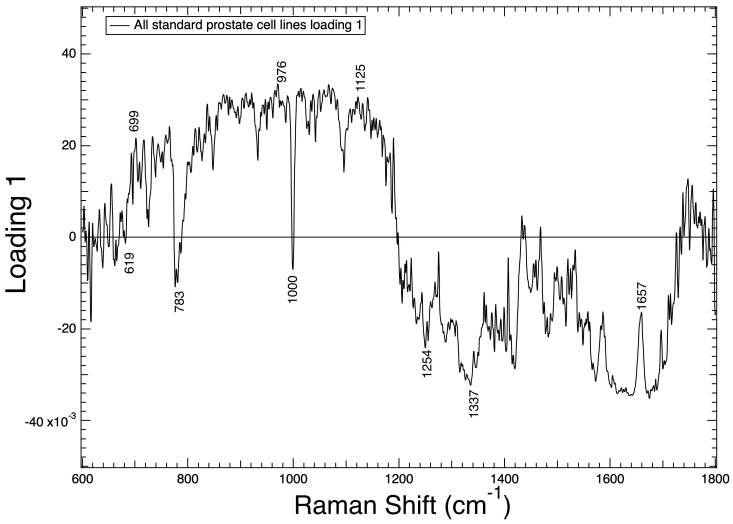


Figure 5.96: The associated loading for PC1, accounting for 31.2% variance, between all standard cell lines in the fingerprint region, is displayed and annotated. Loading 1 is capturing variance at 619 cm^{-1}, 699 cm^{-1} (proteins), 783 cm^{-1} (DNA/RNA), 976 cm^{-1} (lipids, proteins and DNA/RNA), 1000 cm^{-1} (proteins), 1125 cm^{-1} (lipids proteins and carbohydrates), 1254 cm^{-1} (proteins), 1337 cm^{-1} (proteins DNA/RNA), and 1657 (proteins and lipids).

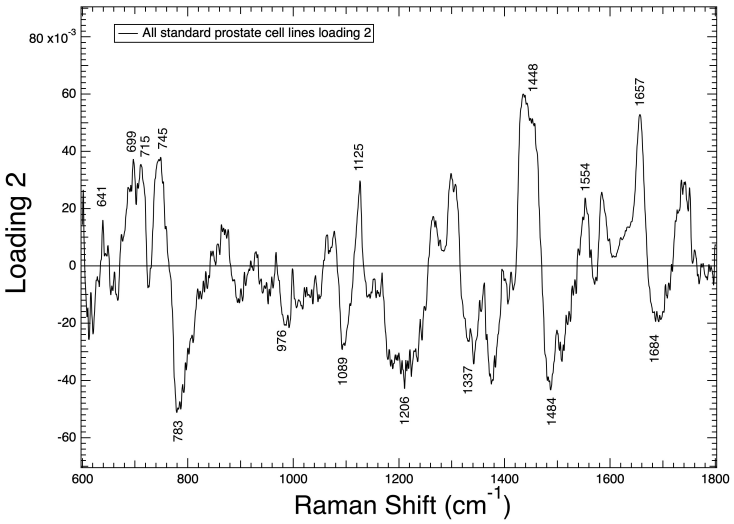


Figure 5.97: The associated loading for PC2, accounting for 8.6% variance, between all standard cell lines in the fingerprint region, is displayed and annotated. Loading 2 is capturing variance at 641 cm^{-1}, 699 cm^{-1} (proteins), 715 cm^{-1} (proteins and lipids), 745 cm^{-1} (DNA), 783 cm^{-1} (DNA/RNA), 976 cm^{-1} (lipids, proteins and DNA/RNA), 1125cm^{-1} (lipids, proteins and carbohydrates), 1206 cm^{-1} (proteins and DNA/RNA), 1337 cm^{-1} (proteins and DNA/RNA) and 1657 cm^{-1} (proteins and lipids).

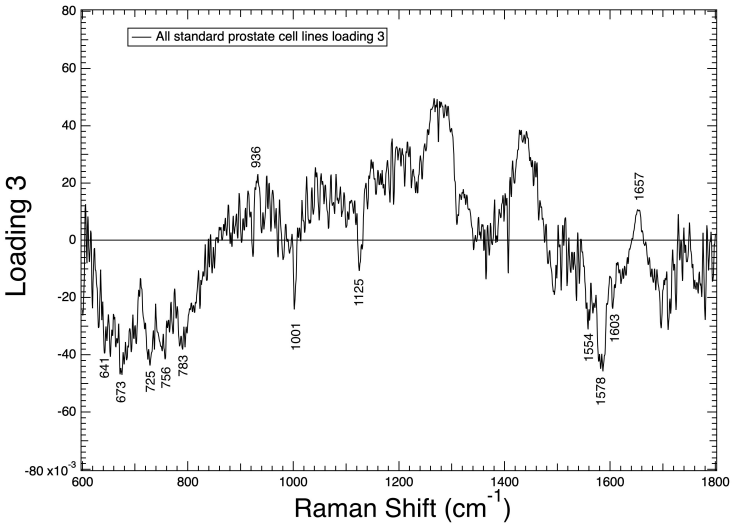


Figure 5.98: The associated loading for PC3, accounting for 6.8% variance, between all standard cell lines in the fingerprint region, is displayed and annotated. Loading 3 is capturing variance at 641 cm^{-1}, 673 cm^{-1} (proteins), 725 cm^{-1} (DNA), 755 cm^{-1} (proteins), 936 cm^{-1} (proteins and carbohydrates), 1001 cm^{-1} (proteins), 1125cm^{-1} (lipids, proteins and carbohydrates), 1554 cm^{-1} (proteins and carbohydrates), 1578 cm^{-1} (DNA/RNA), 1603 cm^{-1} (proteins) and 1657 cm^{-1} (proteins and lipids).

Key peak positions have been extracted from loadings 1, 2 and 3 in Figures 5.96, 5.97 and 5.98. Although there are a many peaks capturing variance, only the largest captures of variance are investigated, due to the number of cell lines at this stage. In loading 1, a relatively large capturing of variance was observed at: 699 cm^{-1} (proteins), 976 cm^{-1} (lipids, proteins and DNA/RNA), 1254 cm^{-1} (proteins) and 1337. In loading 2, a relatively large capturing of variance was observed at: 699 cm^{-1} (proteins), 715 cm^{-1} (proteins and lipids), 745 cm^{-1} (DNA), 783 cm^{-1}, 1089 (DNA/RNA), 1125cm^{-1} (lipids, proteins and carbohydrates), 1206 cm^{-1} (proteins and DNA/RNA), 1337 cm^{-1} (proteins and DNA/RNA) and 1657 cm^{-1} (proteins and lipids). Finally, in loading 3 a relatively large capturing of variance was observed at: 673 cm^{-1} (proteins), 725 cm^{-1} (DNA), 1554 cm^{-1} (proteins and carbohydrates) and 1578 cm^{-1} (DNA/RNA). These loadings will be investigated in Section 5.7.2 to try and measure a function of disease state relative to the normal, using peak intensity ratio analysis.

The separation observed in the LDA space in Figures 5.95(d), 5.95(e) and 5.95(f) show clear separation between the standard cell lines. It is noted that there is a clustering of cell lines based on which LDF in the 3D LDA space is being observed. It is observed, along LDF 1, that there is grouping of PNT2-C2 and P4E6, grouping of BPH-1, PC3 and LNCaP and total separation of DU145. It is observed along LDF 2 that PNT2-C2, P4E6 and DU145 are grouping together. Lastly, complete separation of BPH-1 is observed in LDF3. The relatively tight grouping is also noted between PC3 and LNCaP. The use of LDA as a prediction model is shown in Table 5.4

Table 5.4: PCA-LDA leave-one-out cross-validation is performed on standard cell lines in the fingerprint region. This used a total of 59 principal components, accounting for 78.9% total population variance. The corresponding results are shown in the table below.

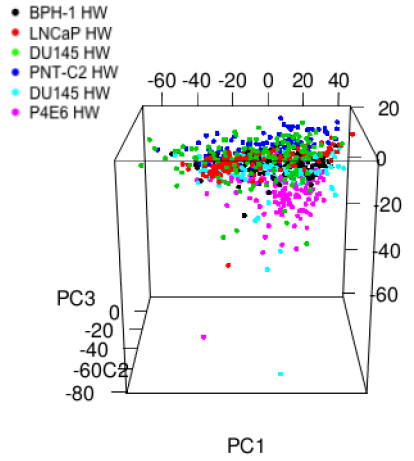
Cell line	**BPH-1**	**LNCaP**	**DU145**	**PNT2-C2**	**PC3**	**P4E6**
BPH-1	118	0	0	0	0	0
LNCaP	0	122	0	0	1	0
DU145	2	6	123	2	1	4
PNT2-C2	0	0	0	150	0	2
PC3	12	15	0	0	95	0
P4E6	0	0	1	4	0	132

The PCA-LDA cross-validation model was able to correctly classify 94.8% of the Raman spectra. The associated 5.2% prediction error in Table 5.4 is attributed to the misclassification of PC3, LNCaP and BPH-1. It should be noted that a study by Crow *et al.* (2005) carries out cross-validated PCA-LDA analysis between the fingerprint profiles of PC3, PCA2b, LNCaP and DU145, achieving 98% correct classification [63]. However, as previously mentioned, the study by Crow *et al.* (2005) used air-dried cells in analysis [63]. The use of air-dried cells was shown in chapter 4 to result in spectral deviations. This could result in artificial separation of cell lines. Furthermore, Crow *et al.* (2005) did not include a normal cell line or a benign hyperplasia cell line for comparison, as done in this thesis. It is thought that the increase in the number of prostate standard cell lines being examined has lead to a higher associated error in this thesis. That said, a correct classification rate of 94.8% is suitable for the classification of prostate standard cell lines.

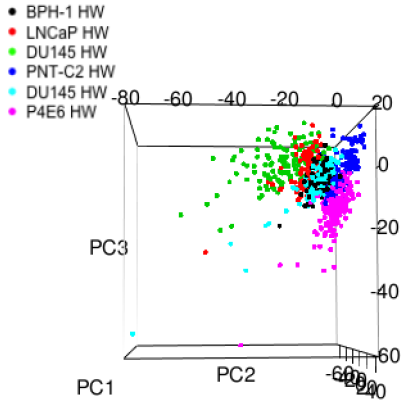
It is noted that in the PCA and LDA results of the fingerprint region in Figure 5.95, that although there is a separation between the different prostate standard cell lines, there is no clustering based upon disease state. The strongest grouping that occurred visually in LDA analysis, Figures 5.95(d), 5.95(d) and 5.95(d), showed that PC3 (metastatic androgen-independent) was grouping with LNCaP(metastatic androgen-sensitive). Whilst this does show that there is grouping of metastatic prostate cancer cell lines, it also show that there is limitations to this analysis. The PCA-LDA leave-one-out cross-validation, in Table 5.4, showed that although there was a high degree of correct classification, the misclassifications were not representative of specific traits, such as androgen-sensitivity. This was further demonstrated in Table 5.4, where the highest misclassification rate occurred between PC3 (metastatic androgen-independent) and BPH-1 (benign prostate hyperplasia).

In the high wavenumber region, PCA analysis between the standard cell lines is able to ascertain that the variance captured by the first 6 principal components is as follows: PC1 = 37.5%, PC2 = 9.3%, PC3 = 7.5%, PC4 = 4.0%, PC5 = 2.9% and PC6 = 2.3%. This in turn means that the first three principal components capture 54.2% and the first six capture 63.4% of the total variance

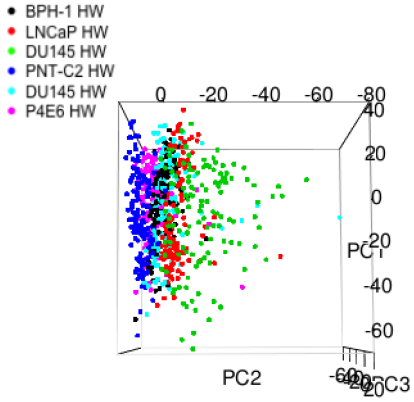
of the sample. Cumulatively, 299 principal components are required to capture total variance of the sample. Three rotations of PCA, space using the first three principal components, are displayed in Figures 5.29, 5.30 and 5.31



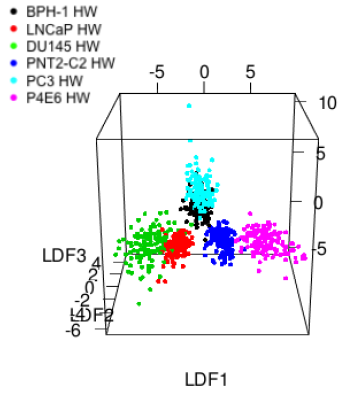
(a) PCA rotation 1



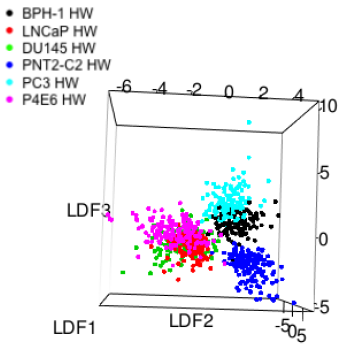
(b) PCA rotation 2



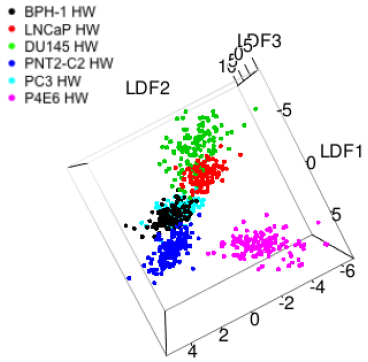
(c) PCA rotation 3



(d) LDA rotation 1



(e) LDA rotation 2



(f) LDA rotation 3

Figure 5.99: Comparison of all standard cell lines in the high wavenumber region. PCA space using the first three principal components is used to compare all standard cell lines at different angles in Figs. (a), (b) & (c). LDA analysis using 53 PCs (90% population variance capture) has been plotted using LDF1, LDF2 and LDF3, is shown (d), (e) & (f).].

The 3D PCA comparison of PNT2-C2 , BPH-1 , LNCaP, DU145, PC3 and P4E6 for the high wavenumber region is shown in Figures 5.99(a), 5.99(a) and 5.99(c). Although there is clearer separation occurring than in the fingerprint region, it is still difficult to observe due to the high number of spectra involved in this comparison. It is determined that there is separation between individual cell lines occurring along PC1, PC2 and PC3. It is therefore possible to determine where the maximum variance of the sample is occurring using the corresponding loadings. The corresponding loadings for PC1, PC2 and PC3 are shown in Figures 5.96, 5.97 and 5.98.

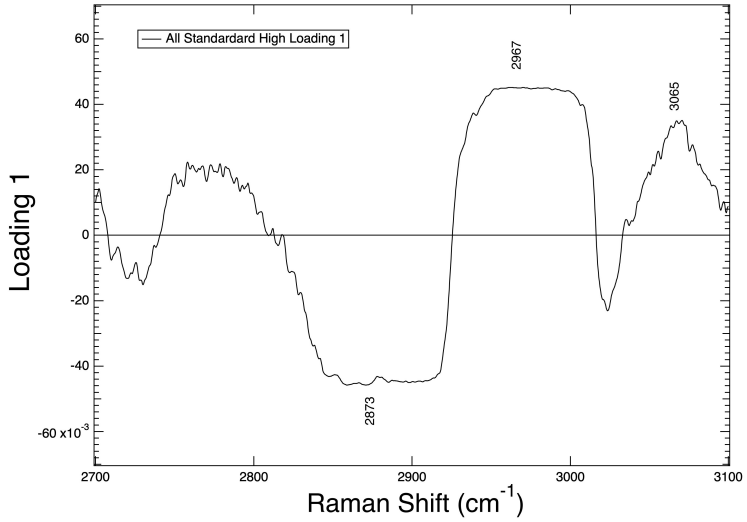


Figure 5.100: The associated loading for PC1, accounting for 37.5% of of the variance, between all standard cell lines in the fingerprint region is displayed and annotated. Loading 1 is capturing variance at 2873 cm^{-1} (lipids and proteins), 2967 cm^{-1} (lipids) and 3065 cm^{-1} (lipids). It is noted that the 2873 cm^{-1} and 2967 cm^{-1} peak locations are not well defined peak maximums.

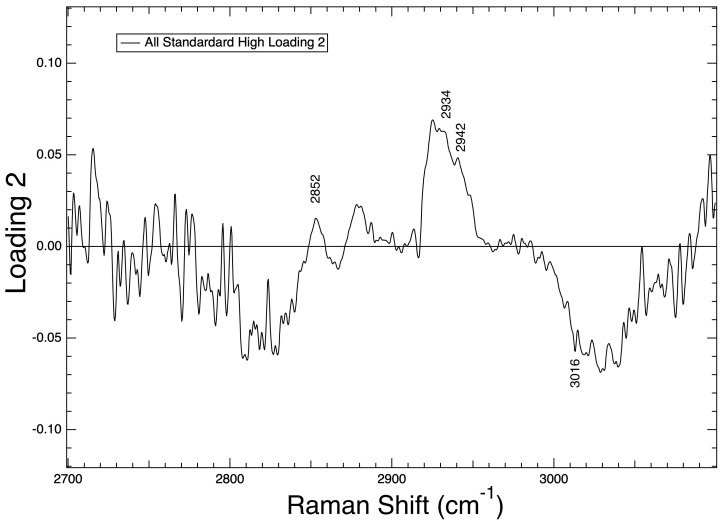


Figure 5.101: The associated loading for PC2, accounting for 9.3% of of the variance, between all standard cell lines in the fingerprint region is displayed and annotated. Loading 2 is capturing variance at 2852 cm^{-1} (lipids), 2934 cm^{-1} (protein and lipids), 2942 cm^{-1} (protein and lipids) and 3065 cm^{-1} (lipids).

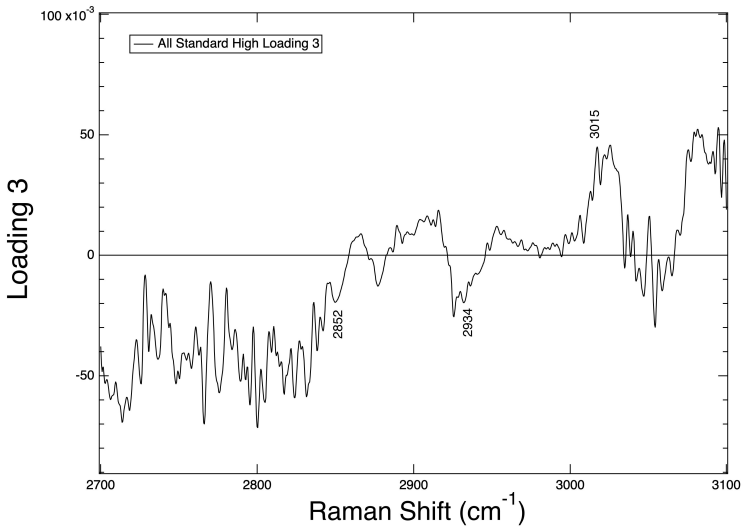


Figure 5.102: The associated loading for PC3, accounting for 7.5% of of the variance, between all standard cell lines in the fingerprint region is displayed and annotated. Loading 3 is capturing variance at 2852 cm^{-1} (lipids), 2934 cm^{-1} (lipids and proteins) and 3015 cm^{-1} (lipids).

In loading 1, figure 5.100, there was a large background variance present. However, the peaks: 2873 cm^{-1} (lipids and proteins), 2967 cm^{-1} (lipids) and 3065 cm^{-1} (lipids) were identified as contributing to the variance. Loading 2, as shown in 5.101, is capturing variance at 2852 cm^{-1} (lipids), 2934 cm^{-1} (protein

and lipids), 2942 cm^{-1} (protein and lipids) and 3065 cm^{-1} (lipids). Furthermore, as demonstrated by figure 5.102, loading 3 is capturing variance at 2852 cm^{-1} (lipids), 2934 cm^{-1} (lipids and proteins) and 3015 cm^{-1} (lipids).These loadings will be investigated in Section 5.7.2 in conjunction with the already identified peaks in the fingerprint region, to try and measure a function of disease state relative to the normal using peak intensity ratio analysis.

The separation observed in the LDA space of the high wavenumber region in Figures 5.95(d), 5.95(e) and 5.95(f) show clear separation between the standard cell lines. It is noted that there is a clustering of cell lines based on which LDF in 3D LDA space is being observed. It is observed along LDF 1 that there is clustering of PNT2-C2 and PC3. It is observed along LDF 2 that P4E6, DU145 and LNCaP are grouping together. Lastly, separation of PNT2-C2 is observed along LDF3. The use of PCA-LDA as a prediction model for the high wavenumber region is shown in Table 5.5.

Table 5.5: PCA-LDA leave-one-out cross-validation is performed on standard cell lines in the high wavenumber region. This used a total of 59 principal components accounting for 91.4% total population variance. The corresponding results are shown in the table below.

Cell line	**BPH-1**	**LNCaP**	**DU145**	**PNT2-C2**	**PC3**	**P4E6**
BPH-1	111	2	0	2	3	0
LNCaP	0	123	0	0	0	0
DU145	3	6	120	0	1	0
PNT2-C2	0	0	0	150	0	0
PC3	5	0	0	0	117	0
P4E6	1	0	1	1	0	134

The PC-LDA model was able to correctly classify 96.9% of the Raman spectra. The associated 3.1% prediction error in Table 5.4 is mainly attributed to the misclassification of PC3 and DU145. It is thought that this is the first time PCA-LDA analysis has been carried out on a complete set of disease state prostate standard cell lines in the high wavenumber region, as far as the author is aware. The previously mentioned study by Crow *et al.* (2005) does not include a normal or benign hyperplasia cell line in their comparison, limiting analysis only to the fingerprint region. The study by Potcoava *et al.* (2014) [64] did perform PCA-LDA analysis on treated and non treated versions of PC3 and LNCaP, however,

no cross-comparison occurred. The high prediction accuracy of the PCA-LDA cross-validation suggests that the high wavenumber region is suitable for the classification of prostate standard cell lines.

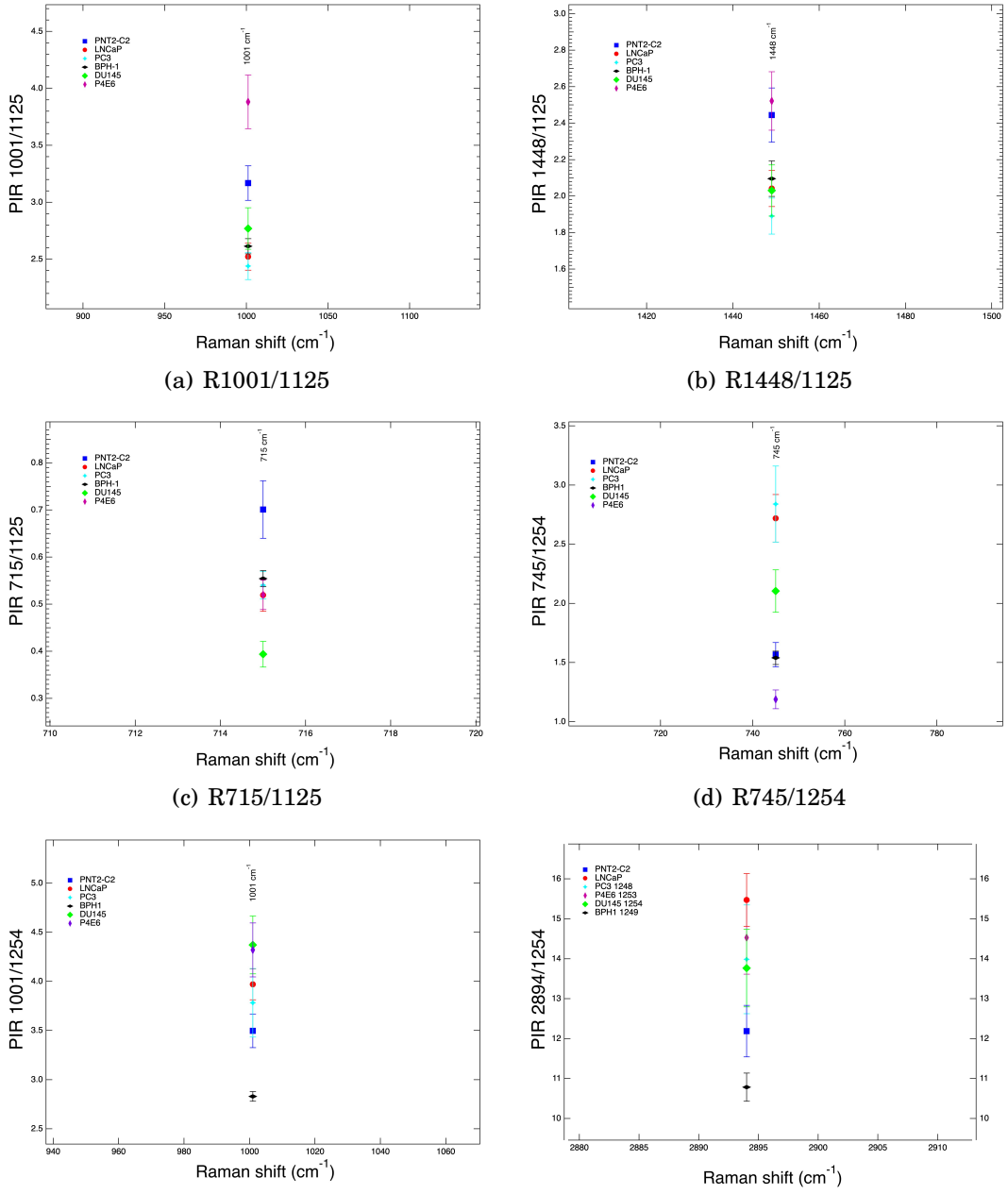
It is noted that in the PCA and LDA results of the high wavenumber region in Figure 5.99, that although there is separation between the different prostate standard cell lines, there is no clustering based upon disease state. The strongest grouping that occurred visually in LDA analysis, Figures 5.99(d), 5.99(e) and 5.99(f), showed that PC3 (metastatic androgen-independent) was grouping with BPH-1 (benign prostate hyperplasia). The grouping of DU145(metastatic androgen-independent) and LNCaP(metastatic androgen-sensitive) was also observed. This shows a similar trend to the LDA analysis of the fingerprint region, where no grouping was observed based on factors such as androgen sensitivity. Using the PCA-LDA leave-one-out cross-validation, in Table 5.5, it can be seen that the misclassifications are not representative of specific traits such as androgen-sensitivity. This was also noted in the fingerprint region. Additional evidence, detailed in Table 5.5, shows the highest misclassification rate occurred between DU145(metastatic androgen-independent) and LNCaP(metastatic androgen-sensitive). This analysis demonstrates that while PCA and LDA analysis can be used to classify prostate standard cell lines from a wide range of disease states, it is not able to group classes based factors such as androgen-sensitivity. The use of PIR analysis will further investigate the separation of cell lines collectively in Section 5.7.2

5.7.2 Identification of biomarkers correlating to disease state

In the previous section it was shown that PCA-LDA analysis could be used to separate the standard cell lines PNT2-C2 , BPH-1 , LNCaP, DU145, PC3 and P4E6 with a high classification rate. This section uses PIR analysis to investigate the key markers of variance observed in the loadings that were generated at the PCA stage.

In the fingerprint region the key peaks associated with variance were found at: 673 cm^{-1} (proteins), 699 cm^{-1} (proteins), 715 cm^{-1} (proteins and lipids), 725 cm^{-1} (DNA), 745 cm^{-1} (DNA),783 cm^{-1} (DNA/RNA) 976 cm^{-1} (lipids, proteins and DNA/RNA),1089 cm^{-1} (DNA/RNA), 1125cm^{-1}, 1206 cm^{-1} (proteins and DNA/RNA), (lipids, proteins and carbohydrates), 1254 cm^{-1} (proteins) and 1337 (proteins and DNA/RNA), 1554 cm^{-1} (proteins and carbohydrates), 1578 cm^{-1} (DNA/RNA) and 1657 cm^{-1} (proteins and lipids).Further variance was found in the high wavenumber region at the peaks ; 2852 cm^{-1} (lipids), 2873 cm^{-1} (lipids and proteins), 2934 cm^{-1}, 2942 cm^{-1} (protein and lipids) (protein and lipids), 2967 cm^{-1} (lipids), 3015 cm^{-1} (lipids) and 3065 cm^{-1} (lipids).

It is not within the scope of this thesis to investigate all peaks associated with variance at the PCA stage. This thesis will only examine the peaks that had the most associated variance in their respective loadings. These identified peak intensity ratios are shown in Figures 5.103 and 5.104.

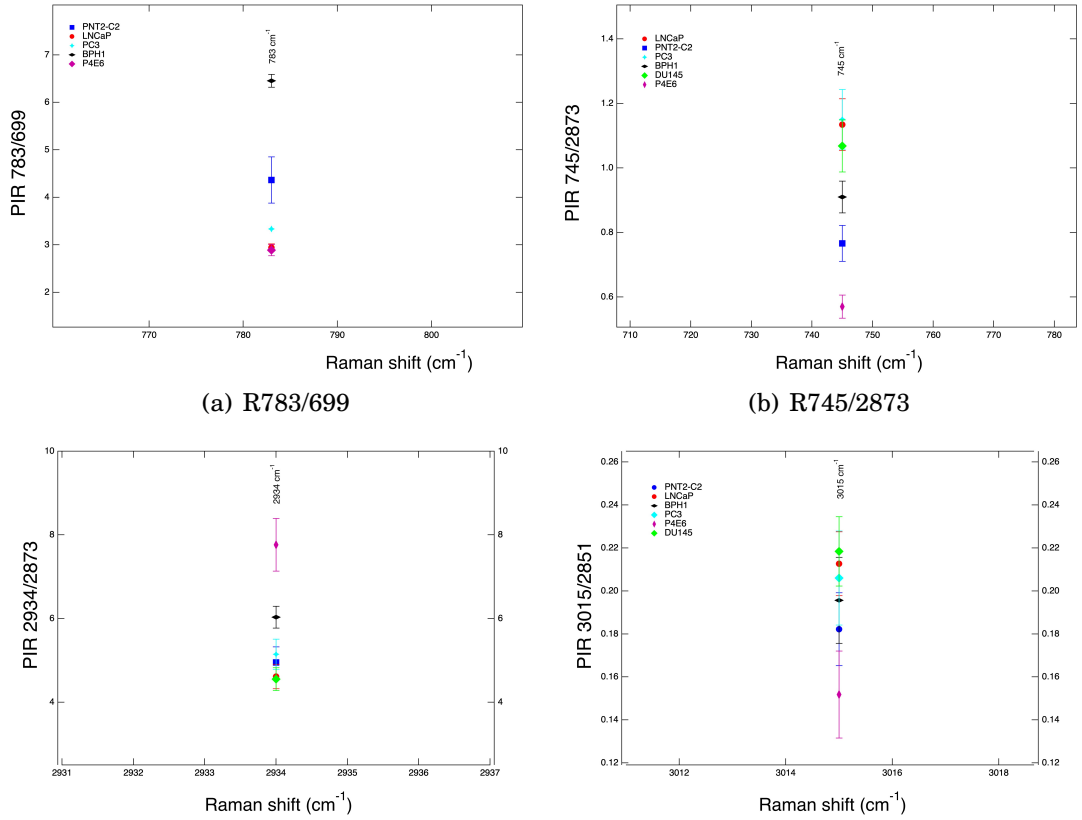


(a) R1001/1125 (b) R1448/1125

(c) R715/1125 (d) R745/1254

(e) R1001/1254 (f) R2894/1254

Figure 5.103: Each panel is displaying a visualisation of a specific peak intensity ratio to investigate changes in disease state in the prostate standard cell lines. [Figs. (a) & (b), (c) & (d) and (e) & (f), respectively].



(a) R783/699 (b) R745/2873

(c) R2934/2873 (d) R3015/2851

Figure 5.104: Each panel is displaying a visualisation of a specific peak intensity ratio to investigate changes in disease state in the prostate standard cell lines. [Figs. (a) & (b), (c) & (d), respectively].

Through analysis of loading 1 in Figure 5.96, it was determined that there was was a large amount of variance occurring at the 1125 cm^{-1} peak. Furthermore, the 1125 cm^{-1} peak was also present in loading 2 in Figure 5.97, capturing a significant amount of variance. It was for these reasons that the 1125 cm^{-1} peak was selected as a biomarker to discriminate between the prostate standard cell lines. The 1125 cm^{-1} peak is representative of lipids, proteins and carbohydrates, specifically, a protein C-N stretch, C-C stretch in lipids and C-C stretch in carbohydrates. The full 1125 cm^{-1} peak PIR panel was considered before selecting specific peak intensity ratios that were able to show the maximum variance between standard prostate cell lines. Figure 5.103(a) shows the peak intensity ratio between the 1001 cm^{-1} phenylalanine peak and 1125 cm^{-1} peak (R1001/1125). It can be seen using the ratio R1001/1125, that PNT2-C2 can be separated from the other cell lines outside of uncertainty. However, the R1001/1125 ratio is not able to

group the standard prostate cell lines based on malignancy, due to the P4E6 ratio also separating from the group. This does, however, show that P4E6 (localised cancer) can be separated from the metastatic groups. It was then decided to ratio the 1448 cm^{-1} peak (CH_2 deformation) against the 1125 cm^{-1} peak (R1448/1125). The ratio R1448/1125, in Figure 5.103(b), showed that PNT2-C2 and P4E6 were grouping with each other, away from the other prostate standard cell lines. It is concluded that the peak intensity ratio R1448/1125 is not able to discriminate between disease states. A third investigation measures the peak intensity ratio of 715 cm^{-1} (amino acid methionine ν(C-S) *trans* and $CN^{+}(CH_3)_3$) against the 1125 cm^{-1} peak (R715/1125). The ratio R715/1125, shown in Figure 5.103(b), is able to discriminate the PNT2-C2 normal from the rest of the prostate standard cell lines. However, this ratio would classify BPH-1 (benign hyperplasia) as malignant. It is also noted that the ratio R715/1125 classifies P4E6 amongst the diseased states, whereas, the ratios R1001/1125 and R1448/1225 do not.

Through further analysis of loading 1, in Figure 5.96, the 1254 cm^{-1} peak (amide III : CH_2 wagging) was shown to be contributing to the variance between all standard cell lines. The first ratio (R745/1254) is carried out using the 1254 cm^{-1} peak against the 745 cm^{-1} peak (thymine ring breathing mode). The ratio R745/1254, shown in Figure 5.103(d), was able to discriminate the standard prostate cell lines into three groups, P4E6 (localised non metastatic), PNT2-C2/BPH-1 (non-malignant), and DU145/PC3/LNCaP (metastatic). However, this ratio does not establish a clear pattern of evolution of disease state in the prostate cancer cell line because the cell line P4E6 has a lower relative thymine to amide III intensity than PNT2-C2 BPH-1; standard prostate cell lines DU145/LNCaP/PC3 have a higher thymine to amide III relative intensity. The 1254 cm^{-1} peak was also considered against the 1001 cm^{-1} phenylalanine peak. The ratio R1001/1254 in Figure 5.103(e), separated BPH-1 from the rest of the cell lines, indicating that a higher relative intensity amide III peak may be an indicator of benign hyperplasia. There was no separation observed between the other prostate standard cell lines using the ratio R1001/1254. The same relationship was observed when comparing the 1254 cm^{-1} peak to the 2894 cm^{-1} peak (CH stretch of lipids and proteins). As

shown in Figure 5.103(f), the ratio 2894/1254 indicates a higher relative intensity of amide III in BPH-1 to CH stretching in lipids and proteins. However, the ratio 2894/1254 is not able to discriminate between prostate disease states as a whole.

Further specific peak intensity ratios were carried out using peaks that were determined to capture high variance in the loadings. The ratio R783/699, using the 783 cm^{-1} peak (Uracil, Cytosine and thymine ring breathing modes/ O-P-O backbone stretch) and 699 cm^{-1} peak (amino acid methionine ν(C-S) *trans*), was able to separate the cell lines based on disease states. The R783/699 peak intensity ratio, shown in Figure 5.104(a) is able to separate the benign, normal and malignant states. However, this peak could not be fitted in DU145 due to low signal-to-noise. It is for this reason that this result is not considered, as either a false negative or positive result could occur. Another ratio analysed was the 745 cm^{-1} peak (thymine) to the 2873 cm^{-1} peak (CH asymmetric/symmetric stretch in lipids and proteins), R745/2873, shown in Figure 5.104(b). This peak intensity ratio was able to separate the standard cell lines, PNT2-C2 (normal), BPH-1(non-malignant) and P4E6 (localised metastatic) from the metastatic cell lines (LNCaP, DU145 and PC3). However, there was no clear pattern in disease state being observed.

There were PIRs that have been identified in the high wavenumber region using variance captured in loading 5.100 and 5.101. The first peak intensity ratio R2934/2873 using the 2934 cm^{-1} peak (CH_2 symmetric stretch) and 2873 cm^{-1} peak (CH_2 asymmetric/symmetric stretch in lipids and proteins). The ratio R2934/2873 was not able to separate the prostate cell lines based on disease state, but separation was observed in P4E6. The 2851 cm^{-1} peak (CH_2 symmetric stretch in lipids) was identified in loading 1 as capturing significant variance, in Figure 5.100. Secondly, the 3015 cm^{-1} peak has been used in previous studies of prostate cancer in tandem with the 2851 cm^{-1} peak (unsaturated =CH stretch in lipids) as an indicator of malignancy [64, 140]. The ratio R3015/2851, shown in Figure 5.104(d), demonstrates that in this thesis it was not possible to discriminate between disease states using this ratio. However it was possible to discriminate DU145. It is therefore determined that the ratio R3015/2851 (TUFA/TFA) can not

be used as a biomarker of malignancy using the prostate standard cell lines when comparing cell lines as a whole.

It was observed in the PIR ratios, that P4E6 (localised cancer) was separated from the metastatic cell lines (LNCaP, DU15 and PC3). The ratio R1001/1125 shows that there is a higher relative intensity of the 1001 cm^{-1} phenylalanine peak to 1125 cm^{-1} peak (lipids, proteins and carbohydrates). Further differences were found using the ratio R745/1254, which showed P4E6 has a much lower relative intensity in the 745 cm^{-1} DNA peak compared to the 1254 cm^{-1} amide III peak. Additionally, the ratio R745/2873 also showed a decreased relative intensity of the 745 cm^{-1} DNA peak compared to the 2873 cm^{-1} lipid peak. This would suggest that the discriminating marker between P4E6 (localised cancer) and the metastatic cell lines is that the metastatic cell lines have a much higher relative DNA content.

It was noted in the PCA and LDA results of the fingerprint region in Figure 5.95 that PC3 (metastatic androgen-independent) was grouping with LNCaP(metastatic androgen-sensitive). It is also seen in all of the PIRs established above, with the exception of Figure 5.103(f), that LNCaP and PC3 are not separated outside of statistical uncertainty. This was not expected as PC3 is androgen-independent and LNCaP is androgen-sensitive. Although not a Raman study, the study by Alimirah *et al.* (2006) used anti peptide antibodies and immunoblotting to show that there were low levels of androgen dependence expressed by PC3 and DU145 [173]. However, although there was some overlap with DU145 and LNCaP at the PIR stage this was not deemed to be significant. The Raman study by Potcoava *et al.* (2006) also analysed the lipids profiles of PC3 and LNCaP [64]. However, Potcoava *et al.* (2006) did not carry out PCA-LDA analysis or PIR analysis to compare these two cell lines; results were focused on the uptake of hormone treatment. As previously mentioned, the study by Crow *et al.* (2005) used PCA-LDA analysis to successfully classify the cell lines LNCaP, DU145, PC3 and pca2B. However, due to the study by Crow *et al.* (2005) using cells that are not in their live state, it is thought that the profiles measure will inherently be different due to the drying process.

To summarise, it was possible to discriminate prostate standard cell lines using PCA analysis and LDA analysis in both the fingerprint and high wavenumber region. It has also been shown that there are PIR markers that can identify differences between PNT2-C2 (normal), BPH-1 (benign), P4E6 (localised malignancy) and DU145/LNCaP/PC3 (metastatic malignancy). Future work should look to considering more peak intensity ratios that may be able to provide further insight into the separation between disease states. It is also possible to correctly classify the prostate standard cell lines with a high level of accuracy. However, neither the PCA-LDA analysis nor PIR ratios are able to group the diseased prostate standard cell lines based on severity of disease or evolution of disease.

5.8 Conclusion

In this chapter the cell lines PNT2-C2, BPH-1, LNCaP, DU145, PC3 and P4E6 have been both characterised and analysed using Raman spectroscopy. Each cell line was individually compared to the normal cell line PNT2-C2 to understand what discriminatory markers existed between the normal and the respective diseased state.

In the comparison of PNT2-C2 and BPH-1, a detection of a higher relative intensity of cystine in BPH-1 was found using the 673 cm^{-1} cystine peak as a biomarker. Furthermore, it was identified using the 699 cm^{-1} and 715 cm^{-1} peaks as biomarkers that there was a relative decrease in the intensity of amino acid methionine ν(C-S) *trans* in BPH-1. The 1448 cm^{-1} peak was also used to detail changes in secondary protein structure between PNT2-C2 and BPH-1. Additionally, in the fingerprint region, the 1554 cm^{-1} peak was used as a biomarker to show a lower relative intensity of amide II and tryptophan in BPH-1. In the high wavenumber region it was observed that by using the 2934 cm^{-1} peak as a biomarker, that there was a higher relative intensity of CH_2 asymmetric stretching and CH_3 symmetric stretching in BPH-1.

The analysis of PNT2-C2 and LNCaP was able to detect several key differences between the cell lines. Using the peak 976 cm^{-1} as a biomarker it was observed

that there was a higher relative intensity of phosphorylated proteins present in LNCaP. Furthermore, using the 1554 cm^{-1} peak as a biomarker it was shown that there is a relative decrease of tryptophan and amide II in LNCaP when compared to PNT2-C2. Using the peak intensity ratio (745 cm^{-1}/1448 cm^{-1}) it was determined that there is a higher relative intensity of thymine (DNA) to CH_2 deformation in lipids and proteins in LNCaP than in PNT-C2. Additionally, using the peaks 810 cm^{-1} and 815 cm^{-1} it was also possible to show that there is a relative increase in peak intensity ratio of the DNA O-P-O stretch in the 1448 cm^{-1} peak (CH_2 deformation in lipids and proteins), indicating a higher presence of A conformation DNA in LNCaP. In the high wavenumber region the 3015 cm^{-1} peak and 2851 cm^{-1} peak were able to demonstrate an up-regulation of lipids in total fatty lipids and total unsaturated fatty acids in relation to the fingerprint and high wavenumber profile in LNCaP, however, the ratio between these peaks remained the same in both PNT2-C2 and LNCaP.

Raman spectroscopy was also able to separate PNT2-C2 and DU145. It was possible to use the 619 cm^{-1} peak as a biomarker to show that there is a higher relative intensity C-C twisting of proteins and phenylalanine C-C twisting in DU145. The 641 cm^{-1} peak was also used as a biomarker to show that there is a higher relative intensity of C-C twisting of tyrosine in DU145. Additionally, the 1060 cm^{-1} peak (lipids chain C-C stretch and DNA/RNA (PO_2^- back bone stretching) was used to show that there was increased relative intensity of DNA/RNA in DU145 as well as differences in the lipid structure between PNT2-C2 and DU145. In the high wavenumber region the 3015 cm^{-1} peak was identified as a biomarker, providing evidence that an increased relative intensity of the unsaturated =CH stretch in lipids is present in DU145.

In the study of PNT2-C2 and PC3 the 745 cm^{-1} peak, representative of the thymine ring breathing mode, was used as a biomarker to show that there is a much higher relative intensity of DNA in PC3. The increased relative intensity of the 745 cm^{-1} peak was such a defining feature of PC3 that several proposed peak intensity ratios were hypothesised to discriminate between PC3 and PNT2-C2. The ratio between the 1254 cm^{-1} peak (amide III) and 1125 cm^{-1} peak (protein C-

N stretch, phospholipids) was used to show that there is a lower relative intensity of amide III in PC3.

In the last Raman characterisation, the study between PNT2-C2 and P4E6, three peaks were identified as being discriminatory markers; 715 cm^{-1}, 745 cm^{-1} and 3015 cm^{-1}. The 715 cm^{-1} showed that there is a lower relative presence of amino acid methionine ν(C-S) *trans* and $CN^{+}(CH_3)_3$ stretching in lipids in P4E6 compared to PNT2-C2. The 745 cm^{-1} biomarker discriminated between the PNT2-C2 and P4E6 cell line; there is a higher relative intensity of the thymine ring breathing mode in PNT2-C2 than P4E6. The 3015 cm^{-1} biomarker showed that there is a higher relative intensity of unsaturated fatty acids in PNT2-C2.

In all cases the diseased prostate standard cell lines were able to be discriminated from the normal PNT2-C2 using PCA analysis. Furthermore, the PCA-LDA leave-one-out cross-validation was able to classify each standard cell line with high classification accuracy between each disease state and PNT2-C2.

When comparing all prostate standard cell lines in the same PCA and LDA space, separation was observed. The PCA-LDA cross-validation model was able to correctly classify 94.8% of the Raman spectra in the fingerprint region. Furthermore, the PCA-LDA model was able to correctly classify 96.9% of the Raman spectra in the high wavenumber region. This demonstrates that Raman spectroscopy can be used as classification tool for standard prostate cell lines across a spectrum of normal and disease states. It was noted that misclassification was not occurring between cell lines with the same attributes such as androgen dependency.

It was possible to extract information from the loadings generated between all prostate standard cell lines, at the PCA stage. This was used to select specific PIR ratios to investigate relative intensity relationships between cell lines in the fingerprint and high wavenumber region. It has also been shown that there are PIR markers that can identify differences between PNT2-C2 (normal), BPH-1 (benign), P4E6 (localised malignancy) and DU145/LNCaP/PC3 (metastatic

malignancy). However, it was not possible to find a PIR that separated all of the prostate cell lines into disease states based on specific characteristics such as androgen-dependence; prostate cancers that require the hormone androgen to grow could not be separated from those that did not.

The biomolecular differences that have been identified between the normal and disease state prostate cells provide a basis for further investigation in fields such as molecular biology and oncology. This further work in neighbouring fields could add to and help improve models such as the Gleason scale in determining prostate cancer severity.

To summarise, through full Raman characterisation of each prostate standard cell line, it has been possible to build a model that can describe different disease states of the prostate. This has been carried out using convergence analysis to ensure that the data sets are converged before in depth PCA, LDA and PIR analysis to determine robust biomarkers of each disease state. A final PCA-LDA and PIR comparison of all standard cell lines showed that it was possible to discriminate between normal, benign hyperplasia, localised cancer and metastatic cancer.

CHAPTER 6

RAMAN SPECTROSCOPY ON PATIENT PRIMARY CELL CULTURES

In the previous chapter Raman spectroscopy was used to discriminate between standard prostate cell lines. It was possible to discriminate the prostate standard cell line PNT2-C2, a normal cell line, from BPH-1 (benign prostate hyperplasia), LNCaP(metastatic androgen-sensitive), DU145(metastatic androgen-independent), PC3 (metastatic androgen-independent) and P4E6 (localised androgen-sensitive). In the comparisons carried out there were also discriminatory markers located for each variation in disease state.

In this chapter, Raman spectroscopy is applied to live patient primary cell cultures. The comparisons carried out are between three sets of normal and malignant cells from three donor patients. The primary cell cultures corresponding to patients are labelled H646/17, H722/18 and H794/18. For each subset of primary cultures from the patients they are denoted with an L or R, depending on which side of the prostate they have been taken from in the biopsy.

Patient primary cell cultures contain stem cells, transit amplifying cells, committed basal cells and luminal cells combining to give complete cellular heterogeneity unlike standard cell lines that have only one undifferentiated cell type.

This increased heterogeneity will be examined through the use of convergence testing. The cancerous cells in the primary patient cell culture in this study all have the same Gleason grade of 7(3+4). PCA and LDA analysis will be used to find the biomolecular differences between patients and their derived prostate cell culture.

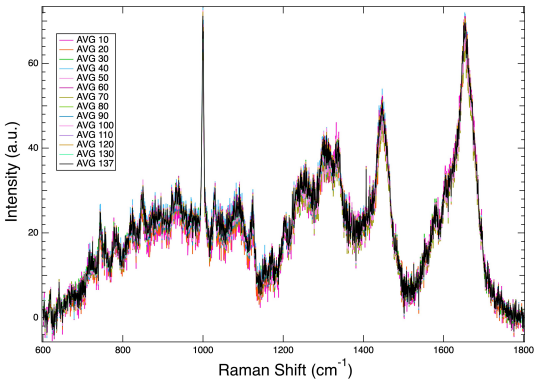
A final comparison will be made between the PIRs derived from the standard cell lines in Chapter 5 section 5.7.2 to the corresponding PIRs in patient primary cell cultures. This will detail how effective standard cell lines are in replicating patient cell cultures.

6.1 Statistical convergence testing of patient primary cell cultures

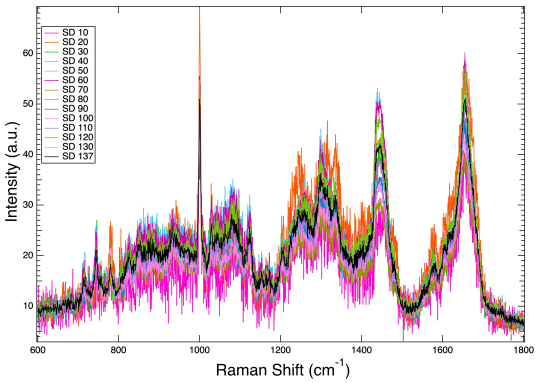
A total of 18 spectra were removed from H646/17R (normal), 15 from H646/17L (cancer), five spectra from H722/18R (normal), 11 from H722/18L (cancer), 17 from H794/18R (normal) and 10 from H794/18L(cancer). These spectra were removed due to falling into the category of being relatively low signal-to-noise or being representative of cytoplasm, as per the criteria and justification outlined in the Methods Section 3.5.6. However, it is thought that due to the number of spectra already in each data set, that the inclusion of this data would have no effect on population-level statistics. These have been removed for a sense of completeness in this study. The total number of spectra remaining after removal of outliers are as follows: H646/17R (137), H646/17L (138), H722/18R (123), H722/18L (120), H794/18R (135) and H794/18L (140).

It is possible, now that outliers have been removed, to analyse the convergence of each standard cell line individually for the fingerprint (600-1800 cm^{-1}) and high wavenumber (2700-3100 cm^{-1}) region. Quantitative analysis, as detailed in Methods Section 3.6.1, is used to measure the convergence of each patient primary cell culture, namely the spectral average, twice the standard deviation and the standard error. The graphs associated with these convergence tests are

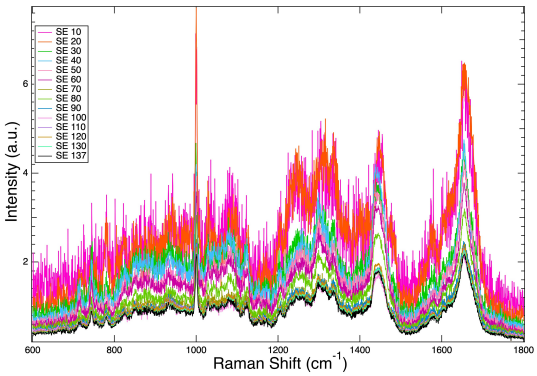
given in Figures; 6.1, 6.2, 6.3, 6.4, 6.5 and 6.6.



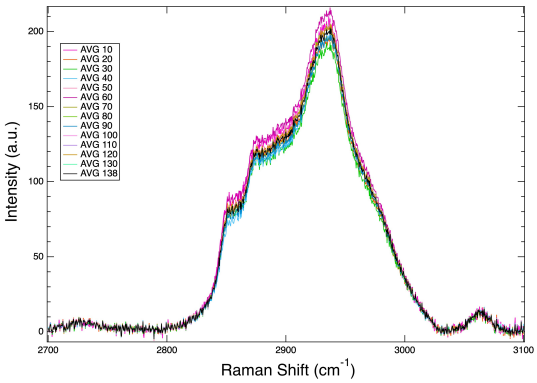
(a) H646/17R FP AVG



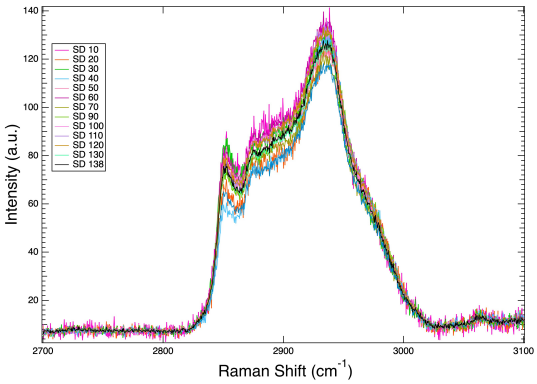
(b) H646/17R FP 2SD



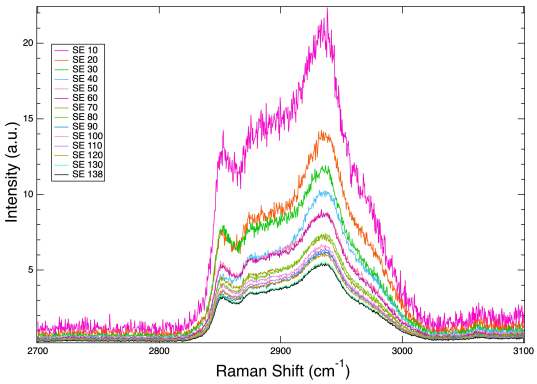
(c) H646/17R FP 2SE



(d) H646/17R HW AVG

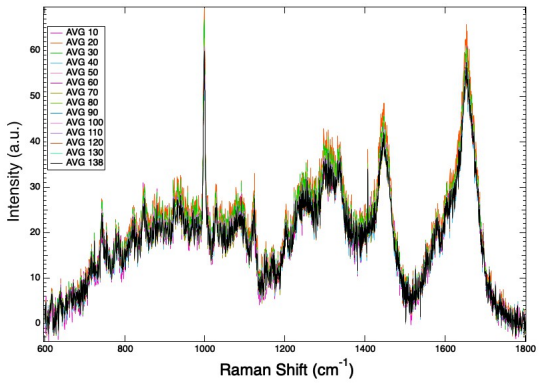


(e) H646/17R HW 2SD

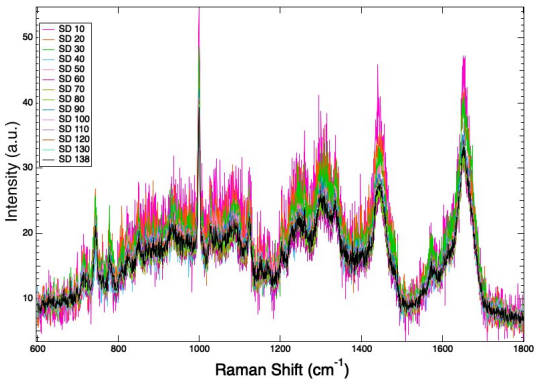


(f) H646/17R HW SE

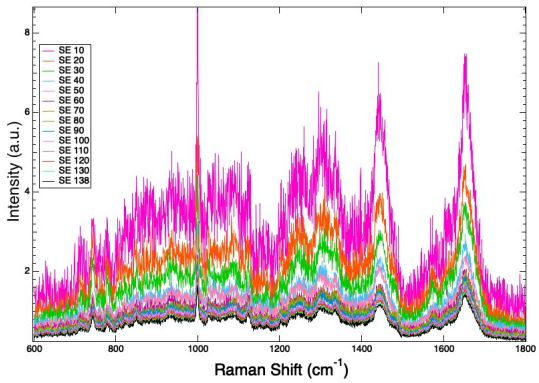
Figure 6.1: Convergence of the average (AVG), second-order standard deviation (2SD) and standard error of the mean (SE) as a function of the increasing number of spectra per spectral average for the H646/78 R primary culture fingerprint (FP) and high wavenumber (HW). [Figs. (a) & (b), (c) & (d) and (e) & (f), respectively].



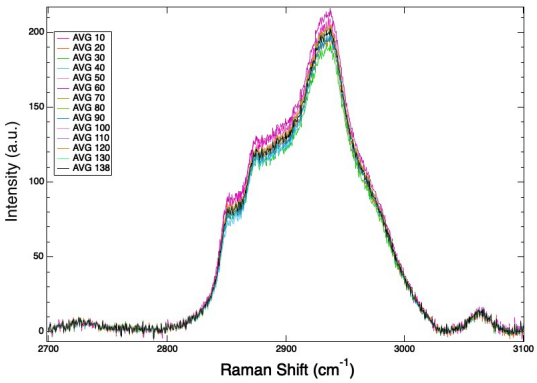
(a) 646L/17L FP AVG



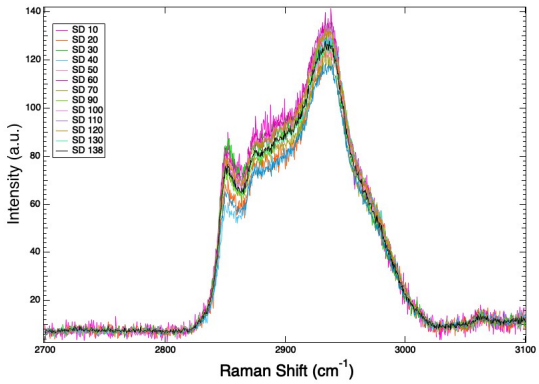
(b) 646L/17L FP 2SD



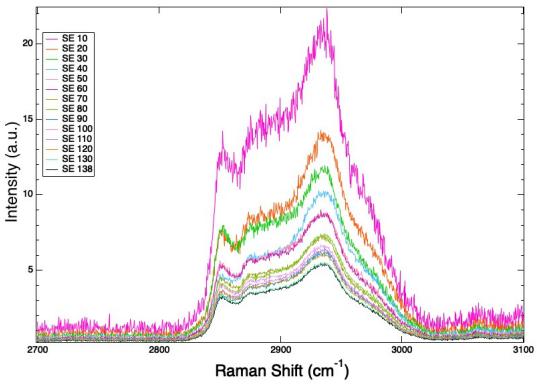
(c) 646L/17L FP SE



(d) 646L/17L HW AVG

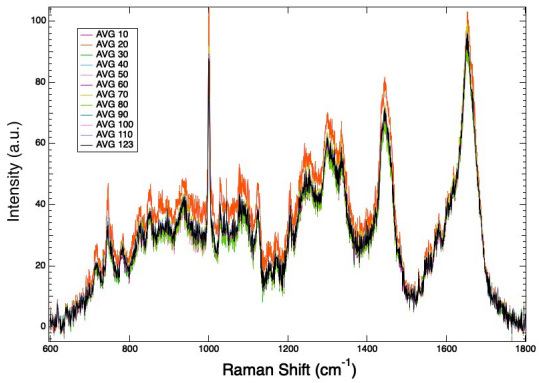


(e) 646L/17L HW 2SD

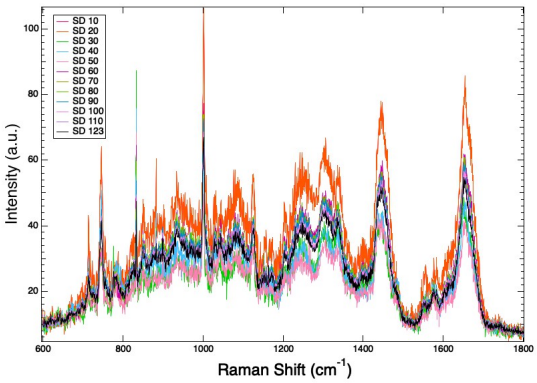


(f) 646L/17L HW SE

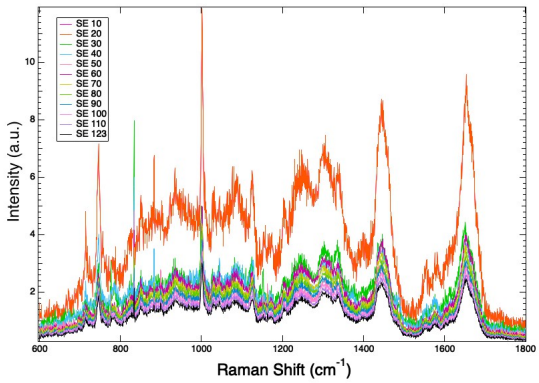
Figure 6.2: Convergence of the average (AVG), second-order standard deviation (2SD) and standard error of the mean (SE) as a function of the increasing number of spectra per spectral average for the H646/17 L primary culture fingerprint (FP) and high wavenumber (HW). [Figs. (a) & (b), (c) & (d) and (e) & (f), respectively].



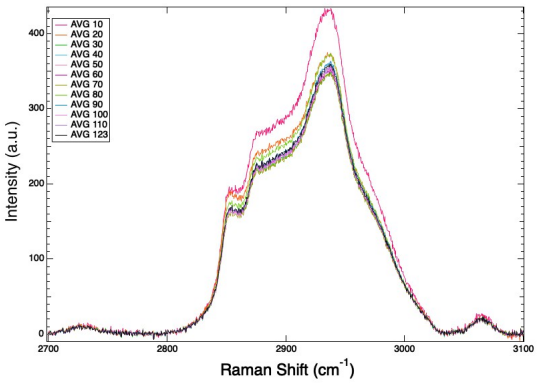
(a) H722/18R FP AVG



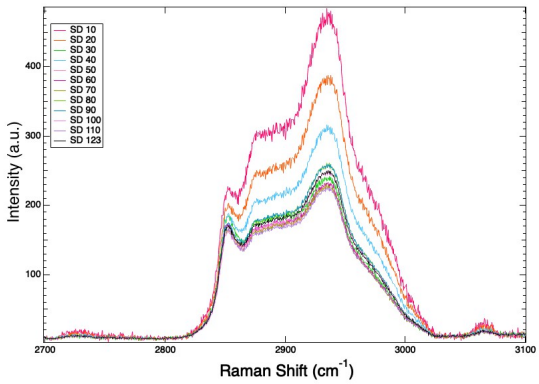
(b) H722/18R FP 2SD



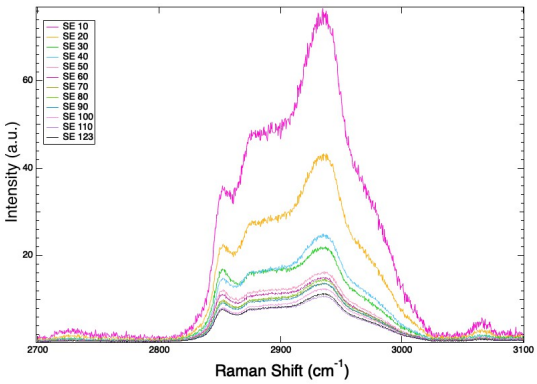
(c) H722/18R FP SE



(d) H722/18R HW AVG

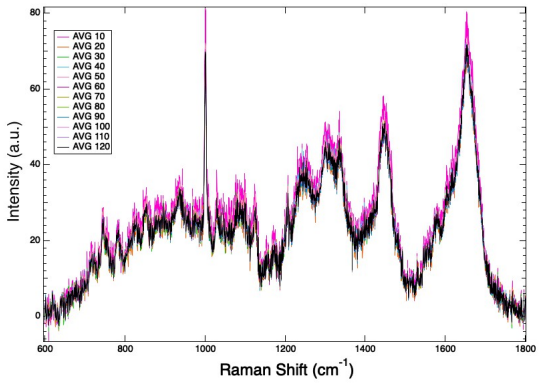


(e) H722/18R HW 2SD

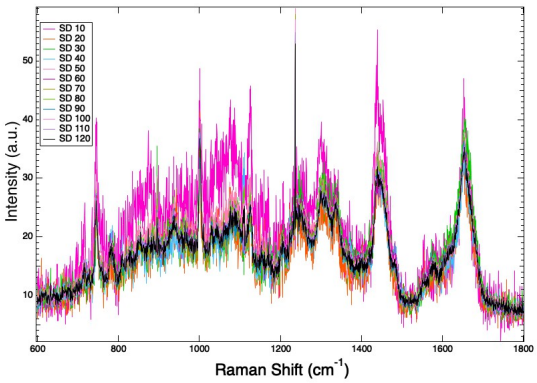


(f) H722/18R HW SE

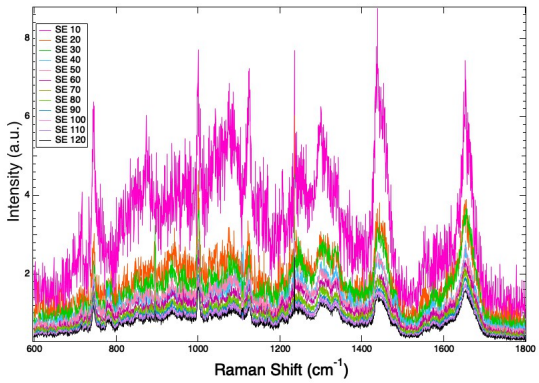
Figure 6.3: Convergence of the average (AVG), twice the standard deviation (2SD) and standard error of the mean (SE) as a function of the increasing number of spectra per spectral average for the H722/18 R primary culture fingerprint (FP) and high wavenumber (HW). [Figs. (a) & (b), (c) & (d) and (e) & (f), respectively].



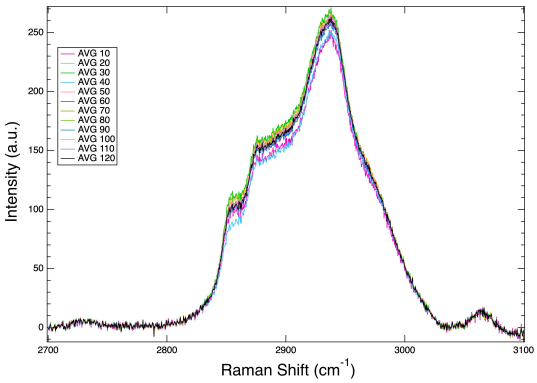
(a) H722/18L HW AVG



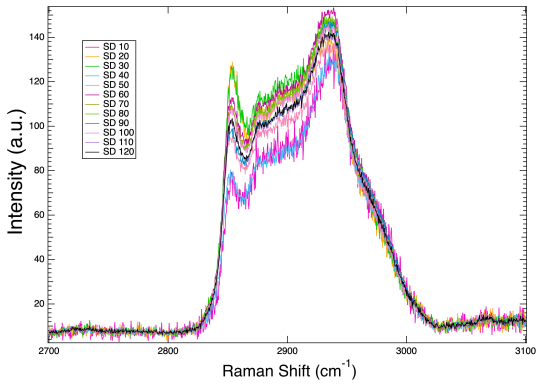
(b) H722/18L HW 2SD



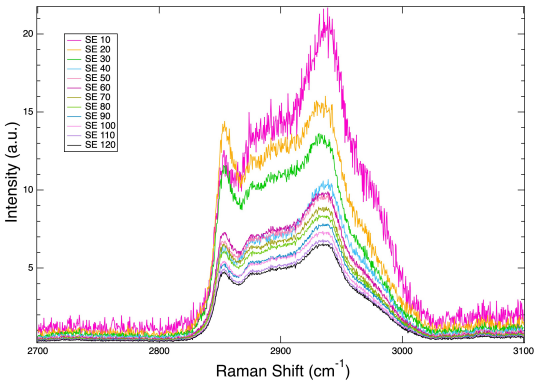
(c) H722/18L HW SE



(d) H722/18L HW AVG

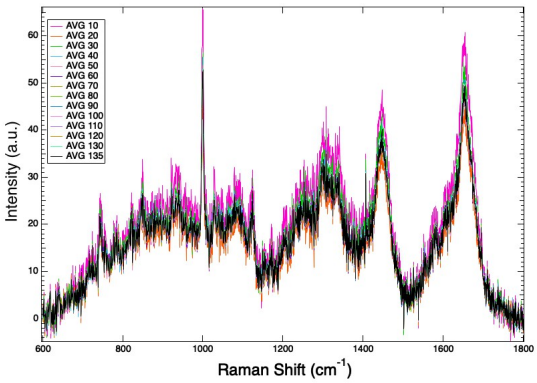


(e) H722/18L HW 2SD

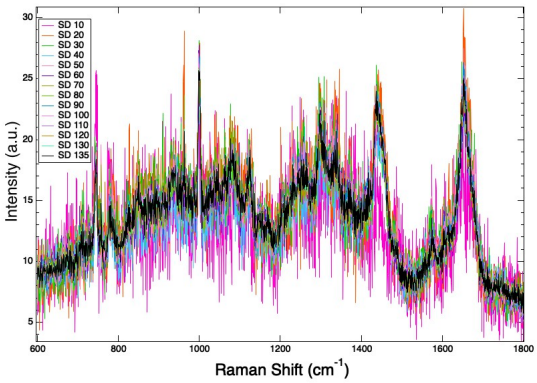


(f) H722/18L HW SE

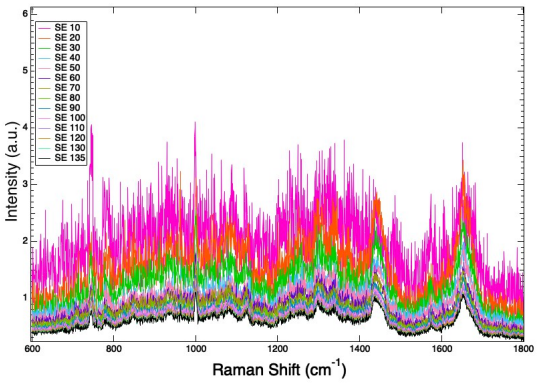
Figure 6.4: Convergence of the average (AVG), twice the standard deviation (2SD) and standard error of the mean (SE) as a function of the increasing number of spectra per spectral average for the H722/18 L primary culture fingerprint (FP) and high wavenumber (HW). [Figs. (a) & (b), (c) & (d) and (e) & (f), respectively].



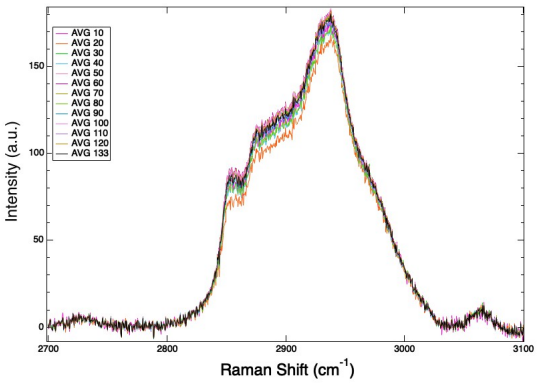
(a) H794/18R FP AVG



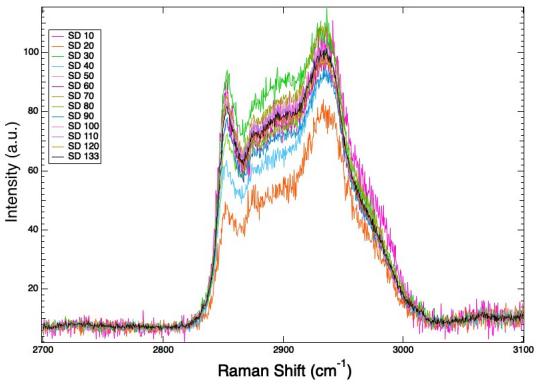
(b) H794/18R FP 2SD



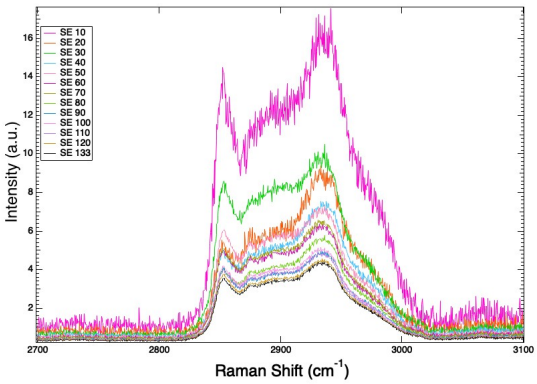
(c) H794/18R FP SE



(d) H794/18R HW AVG

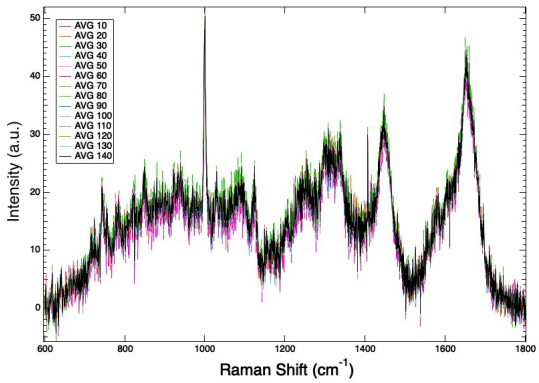


(e) H794/18R HW 2SD

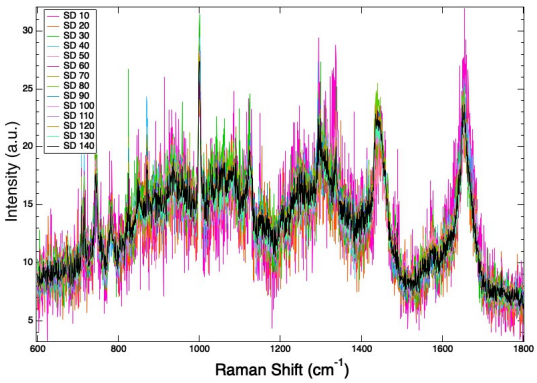


(f) H794/18R HW 2SE

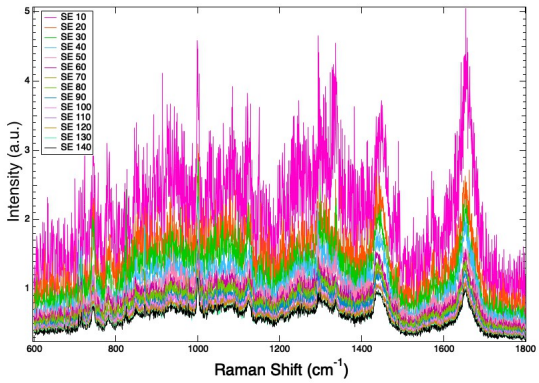
Figure 6.5: Convergence of the average (AVG), twice the deviation (2SD) and standard error of the mean (SE) as a function of the increasing number of spectra per spectral average for the H794/18 R primary culture fingerprint (FP) and high wavenumber (HW). [Figs. (a) & (b), (c) & (d) and (e) & (f), respectively].



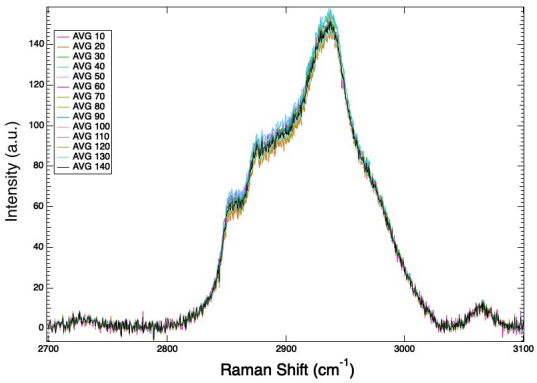
(a) 794/18L FP AVG



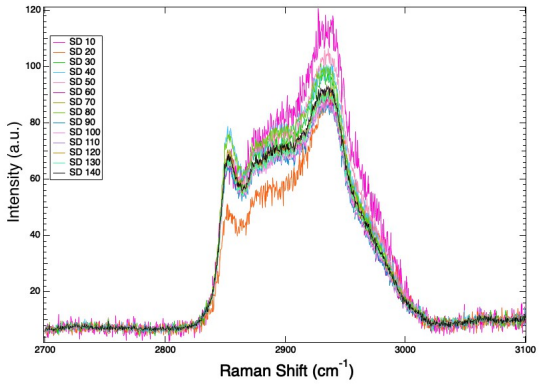
(b) 794L/18 FP 2SD



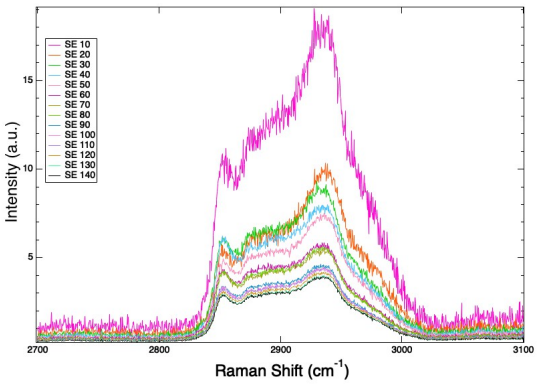
(c) 794L/18 FP 2SE



(d) 794/18L HW AVG



(e) 794/18L HW 2SD



(f) 794/18L HW SE

Figure 6.6: Convergence of the average (AVG), twice the standard deviation (2SD) and standard error of the mean (SE) as a function of the increasing number of spectra per spectral average for the H794/18 L primary culture fingerprint (FP) and high wavenumber (HW). [Figs. (a) & (b), (c) & (d) and (e) & (f), respectively].

The convergence of the spectral average, twice the standard deviation and the standard error for each patient primary cell culture, for the fingerprint and high wavenumber region, is detailed in Table 6.1. The convergence in the fingerprint region was assessed using the 1001 cm^{-1} peak. The convergence in high wavenum-

ber region was assessed using the 2934 cm^{-1} peak. The criteria for convergence is detailed as per the Methods Section 3.6.1.

Table 6.1: The convergence data for primary prostate cell cultures is displayed. Key for table: H646/17R 2SD FP 80 represents that twice the standard deviation in H646/17R converged at approximately 80 spectra in the fingerprint profile.

Cell line	AVG FP	2SD FP	SE FP	AVG HW	2SD HW	SE HW
H646/17R	80	80	70	50	60	70
H646/17L	50	60	60	80	70	70
H722/18R	90	90	90	50	50	60
H722/18L	60	70	70	60	70	70
H794/18R	50	60	80	60	70	80
H794/18L	70	70	60	80	70	90

As stated in the last chapter, there are a lack of studies that include convergence data as part of their results [63, 64, 141, 224]. It can be seen that all of the patient primary cell cultures are converged, from Table 6.1. In comparing the results from the primary patient cell culture to the convergence tests in Chapter 5 section 5.1.1 it is seen that no noticeable difference is occurring in the number of spectra it takes to achieve convergence. In both standard cell lines and patient primary cell cultures the convergence of data typically occurs between 50-80 spectra in the fingerprint and high wavenumber regions. This shows, that while patient primary cell cultures contain stem cells, transit amplifying cells, committed basal cells and luminal cells, combining to give complete cellular heterogeneity, the convergence of Raman spectra is similar to that of the standard cell lines, that have only one undifferentiated cell type. This would suggest that the increased heterogeneity in the patient primary cultures does not have an effect on the convergence. There have been no Raman studies on prostate patient primary cell cultures with which to compare this result, as far as the author is aware. One hyper-spectral Raman study by Oliveira *et al.* (2019), analysed patient derived human thyroid cells [286]. However, despite a large number of single cells being obtained from benign thyroid (n = 127) and thyroid carcinoma (n = 121) there was no convergence testing carried out [286]. Furthermore, preliminary analysis by Oliveira *et al.* (2019) on thyroid subtypes included far less cells; follicular adenoma (n = 20), ollicular carcinoma (n = 25) and papillary carcinoma (n = 18) [286]. As the study by Oliveira *et al.* (2019) also used a combination of PCA and

LDA analysis for discrimination and classification, there is a risk that data sets that are not converged could lead to false positives or negatives.

6.2 Raman spectroscopy of patient primary cell cultures

The average Raman profile for each patient primary cell culture in the fingerprint region has been plotted in Figure 6.7. Furthermore, the corresponding high wavenumber Raman profiles have been plotted in 6.8. In both the fingerprint and high wavenumber region the spectra have been pre-processed as per Section 3.6.2 of the Methods chapter. To allow for display of all primary cell cultures on the same graph, they have been normalised using the total spectrum area and offset by an arbitrary amount to make visualisation between samples clearer. The graphs for both the fingerprint region and high wavenumber region are displayed below in Figures 6.7 and 6.8.

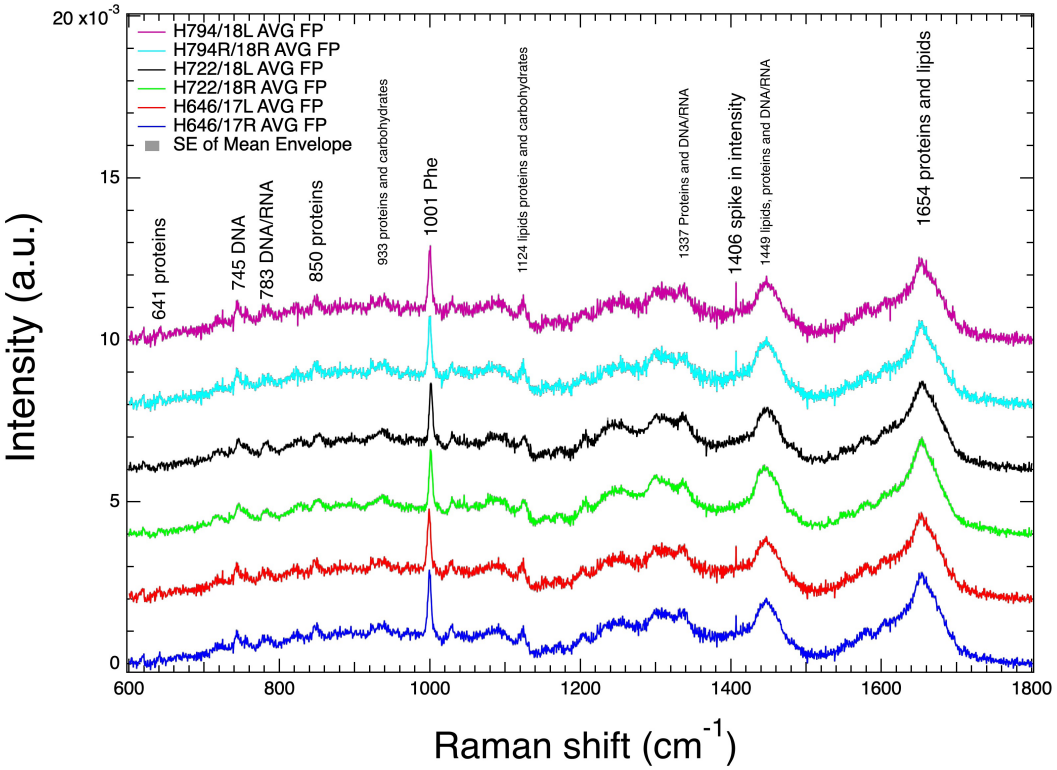


Figure 6.7: The fingerprint region of the primary cell cultures H646/17R, H646/17L, H722/18R, H722/18L, H794/18R and H794L/18 are displayed on the same graph. These spectra have been cut to the region 600-1800 cm^{-1}, undergone a linear baseline subtraction before being normalised to the total spectrum area. Additionally, the each cell line has been offset by an arbitrary amount to make make it easier to view each cell line individually. The mean standard error envelope of each spectrum is displayed in grey around each profile. Observationally, at this stage, it is noted that the patient primary cell cultures H646/17 and H794/18 have more noise present than the H722/18 sample. However, this will not be an issue as the data set is full converged and this thesis only investigates relative intensity, not absolute intensity. General peak assignments have been added for reference.

Figure 6.7 shows the fingerprint region of the primary cell cultures H646/17R, H646/17L, H722/18R, H722/18L, H794/18R and H794/18L displayed on the same graph. The standard error region of each spectra is displayed in grey around each profile. It can be observed in 6.7 that patient primary cell cultures H646/17 and H794/18 have more noise present than the H722/18 sample. The exact cause of this is not clear. However, this should not be an issue for further analysis as all the patient primary cultures are converged spectral averages despite this noise. It was also shown in Figure 6.7 that there was a spike in intensity occurring at the

location 1406 cm^{-1} in the primary patient cultures H794/18 and H646/17. The exact cause of this spike is not know but it was not considered a band during the fitting process in either patient primary cell culture.

The average profiles for the high wavenumber region of the primary patient cell cultures are shown in Figure 6.8.

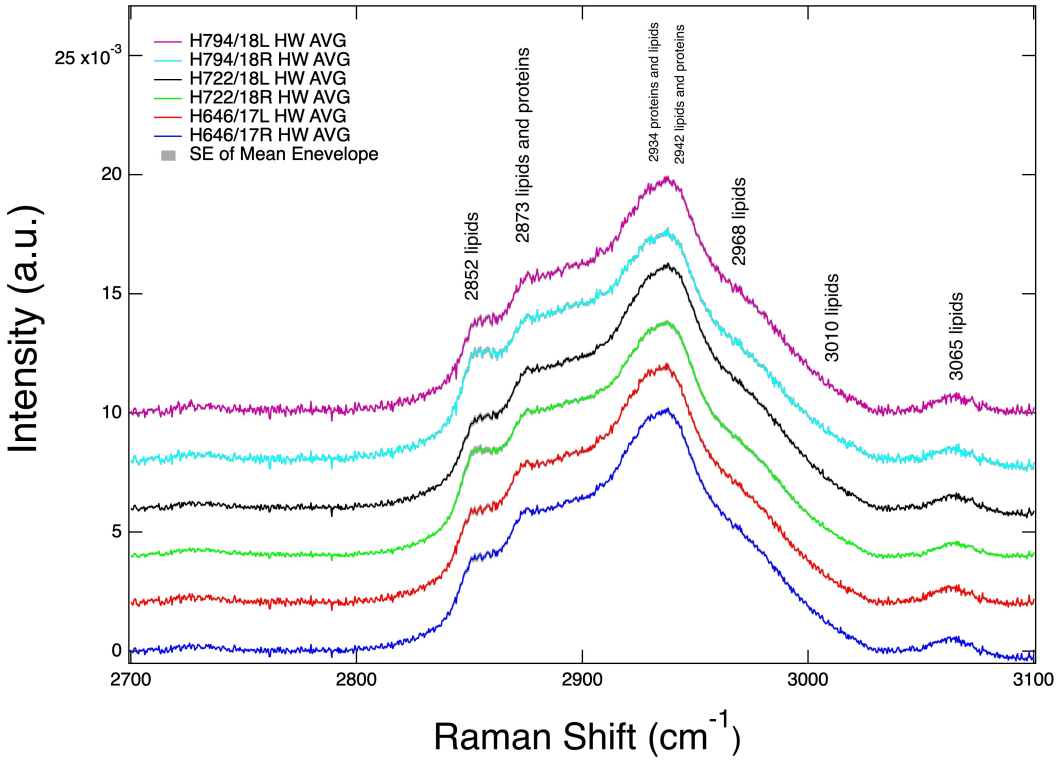


Figure 6.8: The fingerprint region of the primary cell cultures H646/17R, H646/17L, H722/18R, H722/18L, H794/18R and H794/18L are displayed on the same graph. These spectra have been cut to the region 2700-3100 cm^{-1}, undergone a polynomial baseline subtraction before being normalised to the total spectrum area. Additionally, the each cell line has been offset by an arbitrary amount to make make it easier to view each cell line individually. The mean standard error envelope of each spectrum is displayed in grey around each profile. General peak assignments have been added for reference.

Figure 6.8 shows observationally that the profiles of the patient primary cell cultures are similar. There is a slight increase in the noise observed in the H794/18 primary patient culture, although as for reasons stated previously, this will not affect analysis.

The peak positions for each primary patient cell culture have been located

using the using Gaussian peak fitting. A detailed account of the fitting method is given in chapter 3 Section 3.6.3. The fitting of profiles has been carried out for the fingerprint region (600-1800 cm^{-1}) and high wavenumber region (2700-3100 cm^{-1}). The peak average is calculated using the range of the peaks found between all samples. The exact peak positions of all bands can be viewed in Table B.3 in the Appendix. The general biomolecular assignments, as well as specific assignments, have been attributed to respective positions. This information is condensed in table 6.2.

Table 6.2: Biomolecular peak-assignments for primary cell cultures. The average peak position is shown (calculated in Table B.3 in the appendix), general and detailed assignments obtained from the literature. For the peak position, a standard error of ± 3 cm^{-1} as an uncertainty is attached to all positions.

Peak position (cm^{-1})	**General assignment**	**Detailed biomolecule assignments**
619	Proteins	C-C twist of proteins ([128], [129]) Phenylalanine C-C twist ([128], [129], [224], [245])
641	Proteins	C-S stretch & C-C twist of tyrosine ([128], [129] C-C twist of tyrosine ([224], [245], [216], [142])
668	DNA/RNA & Proteins	Guanine and thymine ring breathing mode ([128]) Tyrosine-G backbone in RNA ([128]) Guanine ring breathing mode and C-S stretching mode of cystine ([245]) Guanine, tyrosine and cysteine ([142])
697	Proteins	Aminoacid methionine ν(C-S) *trans* ([128])
718	Proteins & Lipids	Aminoacid methionine ν(C-S) *trans* ([128]) $CN^+(CH_3)_3$ stretch in lipids ([128]) C-N membrane phospholipid head stretch ([128], [246], [245]) Adenine nucleotide peak ([245], [128])
744	DNA	Thymine ring breathing mode ([248], [105], [128], [224])
755	Proteins	Symmetric breathing of tryptophan ([245])

Continued on next page

TABLE 6.2 (continued): Biomolecule assignments

Peak position (cm^{-1})	**General assignment**	**Detailed biomolecule assignments**
		Tryptophan ring breathing mode ([128], [246]) Tryptophan ([249], [245], [129])
783	DNA/RNA	Uracil, Cytosine and thymine ring breathing modes ([249], [142], [128], [224]) Nucleic acids ([63]) Uracil, cytosine, and thymine overlapping with O-P-O stretch in DNA ([105])
808	DNA/RNA	O-P-O stretch DNA ([249] [249]) and O-P-O stretch RNA ([128], [246])
822	Proteins & DNA/RNA	Tyrosine out-of-plane ring breathing and O-P-O stretch in DNA ([249], [245], [142], [105] [105], [128]) DNA O-P-O asymmetric stretch, tyrosine ([248]) Proline and hydroxyproline, out-of-plane tyrosine ring breathing, DNA/RNA asymmetric O-P-O stretch ([246]) DNA backbone (O-P-O) ([63]) Out-of-plane tyrosine ring breathing ([224])
850	Proteins	Tyrosine ring breathing ([249], [105], [128], [224]) C-C stretch in proline and tyrosine ring breathing mode ([245], [142]) Proteins (C-C stretch in proline; tyrosine ring breathing); carbohydrates (glycogen, polysaccharides C-O-C stretch) ([246])

Continued on next page

TABLE 6.2 (continued): Biomolecule assignments

Peak position (cm^{-1})	**General assignment**	**Detailed biomolecule assignments**
879	Proteins & Lipids	Proteins C-C stretch, antisymmetric vibrations of choline $N(CH_3)_3$ (phospholipids), hydroxyproline, tryptophan (proteins) ([128]) Hydroxyproline C-C stretch in collagen, tryptophan ring deformation, choline asymmetric stretch in phospholipids and C-O-C stretch in carbohydrates ([246]) Tryptophan ([224])
896	Carbohydrates & Proteins	Saccharide band, methylene and C-C skeletal, proteins (protein modes for tumours) ([128]) C-C skeletal ([105])
935	Proteins & Carbohydrates	Skeletal C-C stretch α-helix (protein), proline, valine ([245], [128]) C-C backbone stretch α-helix ([249], [63], [142], ([246]) C-O-C glycogen (carbohydrate) ([246])
1000	Proteins	Phenylalanine ring breathing ([248], [249], [245], [142], [105], [128], [246], [224])
1029	Proteins (namely), Lipids & Carbohydrates	Phenylalanine C-H in-plane bending mode ([249], [245], [142], [128], [224]) C-N stretch of proteins ([105], [128]) Collagen, keratin, C-N stretch in proteins, Phe C-H in plane bending, phospholipids and polysaccharides ([246])

Continued on next page

TABLE 6.2 (continued): Biomolecule assignments

Peak position (cm^{-1})	**General assignment**	**Detailed biomolecule assignments**
1063	Lipids & DNA/RNA	Lipids chain C-C stretch ([249], [245], [128]) DNA/RNA PO_2^- backbone stretch ([128])
1088	DNA, Lipids & Proteins	DNA PO_2^- backbone stretch and C-C stretch (gauche) in lipids ([248] [248], [245], [105], [128]) C-N of proteins ([105], [128]) Lipids ([63])
1124	Lipids, Proteins & Carbohydrates	Protein C-N stretch ([249], [142], [128]) C-C stretch in lipids ([249], [245], [142] [128]) C-C stretch in proteins ([245] , [128]) C-N stretch (glucose), C-O & C-C stretch in carbohydrates ([245], [128]) phospholipids ([63])
1154	Proteins & Carbohydrates	C-C and C-N stretch in proteins ([249], [105], [128], [224]) Carotenoids—correlated to disease ([128]) Glycogen ([128])
1170	Proteins	(CH) phenylalanine, tyrosine ([105], [128]) Bending mode (in-plane) tyrosine ([249], [245]) Protein-related peak (C-H bend) ([105])
1204	Proteins & DNA/RNA	Tyrosine and phenylalanine ([248], [128]) Phenylalanine and tryptophan ([249], [245], [128])

Continued on next page

TABLE 6.2 (continued): Biomolecule assignments

Peak position (cm^{-1})	**General assignment**	**Detailed biomolecule assignments**
		Phenylalanine ([224]) CH_2 wagging of glycine and proline ([128]) Amide III ([105], [128]) A,T ring breathing modes of DNA/RNA ([105], [128]) Hydroxyproline & tryrosine ([245], [105])
1232	Proteins & DNA/RNA	Amide III and CH_2 wagging of glycine and proline ([128]) PO_2^- antisymmetric stretch ([128]) Amide III ([249],[142], [224])
1251	Proteins	amide III : CH_2 wagging [245, 250]
1307	Proteins, Lipids & Nucleic acids	CH_2 deformation in lipids, adenine and cytosine ([245], [128]) CH_3/CH_2 modes in lipids / collagen ([245], [128]) CH-N stretch in amines ([128]) Nucleic acids ([63]) Lipids; proteins (amide III), adenine, cytosine ([248])
1337	Proteins & DNA/RNA	CH_3CH_2 wagging, twisting &/or bending of collagen, lipids, tryptophan & nucleic acids (adenine and guanine)([128]) Amide III ([128]) CH_2 deformation in lipids, adenine and cytosine ([245]) Adenine and guanine; protein C-H deformation ([249], [105]) Adenine, phenylalanine and CH deformation ([248]) Adenine ([142])

Continued on next page

TABLE 6.2 (continued): Biomolecule assignments

Peak position (cm^{-1})	**General assignment**	**Detailed biomolecule assignments**
1449	Lipids, proteins & DNA/RNA	CH_2 deformation in lipids and proteins; at 1450 cm^{-1}, in malignant tissues ([248] [248], [142] [142], [128]) CH_2 bending mode of proteins ([245], [105], [224]) C-H deformation & vibration in lipids and proteins ([249], [128]) CH_2CH_3 deformation ([128]) Nuclear acids ([63]) CH_2 in-plane deformation (ring stretch) in DNA/RNA ([142])
1484	Protein & DNA/RNA	Amide II CN stretch & N-H bend ([128]) Guanine, adenine ring-breathing modes ([245], [105], [142], [128])
1553	Proteins	Tryptophan ([245], [128]) Amide II ([128])
1577	DNA/RNA	Guanine and adenine ring-breathing modes ([249], [245], [105], [142], [128]) Nucleic acids ([245], [63])
1603	Proteins	C=C bending mode of phenylalanine and tyrosine ([249], [245], [105], [128]) C-C ring stretch of phenyl ([128]) C-C and C-N stretch (proteins) ([224])
1615	Proteins	Tyrosine, tryptophan C=C stretch ([249], [245], [105], [142], [128])
1654	Proteins & lipids	Amide I C=O stretch, α-helix conformation; proteins ([249], [245], [128])

Continued on next page

TABLE 6.2 (continued): Biomolecule assignments

Peak position (cm^{-1})	**General assignment**	**Detailed biomolecule assignments**
		α-helix proteins & unordered proteins ([63]) protein (amide I) [105] C=C stretch of lipids ([249], [245], [248], [128], [64])
1679	Proteins	Amide I ([105], [128])
2852	Lipids	CH_2 symmetric stretch in lipids ([128], [64], [251], [252])
2873	Lipids & proteins	CH_2 asymmetric stretch in lipids and proteins ([128]) CH_2 symmetric stretch in lipids ([128]) CH_3 symmetric stretch in lipids & proteins ([64])
2894	Lipids & proteins	CH stretch of lipids and proteins (predominantly) and CH_3 symmetric stretch ([128]) CH_2 deformation ([251], [252])
2934	Proteins & lipids	CH_2 asymmetric stretch, CH_3 symmetric stretch (predominantly proteins) ([128]) CH_2 asymmetric stretch of proteins and modes in lipids, CH modes in lipids and proteins ([251]) CH_3 symmetric stretch in proteins and lipids ([252])
2942	Lipids & proteins	C-H vibrations in lipids & proteins. CH_2 modes in lipids ([128])
2968	Lipids	CH_3 asymmetric stretch in lipids, cholesterol & cholesterol ester ([128])
3010	Lipids	Unsaturated =CH stretch in lipids ([128], [251], [252])

Continued on next page

TABLE 6.2 (continued): Biomolecule assignments

Peak position (cm^{-1})	**General assignment**	**Detailed biomolecule assignments**
		=CH stretch (unsaturated lipids, cholesterolester) ([64])
3064	Lipids	CH stretch in lipids ([128])

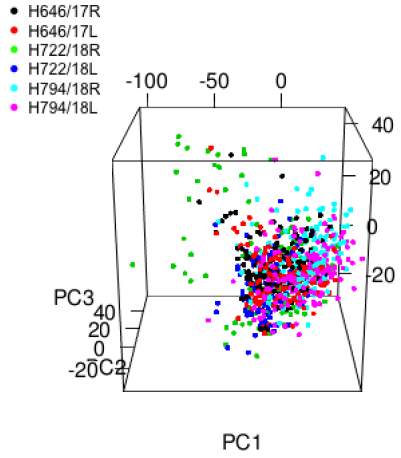
There is a total of 42 peaks that have been characterised across the three patient primary culture sets. This is 3 bands less than observed in the standard cell lines in chapter 5. The 723 cm^{-1} band (DNA and adenine) was not present in any of the primary patient cell cultures. Furthermore, the 974 cm^{-1} band (lipids proteins and carbohydrates) was not present in any of the primary patient cell cultures. The last of the three bands missing was the 1398 cm^{-1} band (lipids). However, it should be noted that the 1398 cm^{-1} band was also not present in the standard cell lines BPH-1 and DU145. It should be highlighted that there are also further bands absent in specific patient primary cultures, between patients and between the normal and malignant cells. It is thought that while the loss of these bands will have a knock-on effect on analysis, that this information loss is not detrimental to the understanding of the patient primary cell cultures, due to the large number of bands that can still be analysed.

6.3 PCA-LDA for the classification of patient primary cell cultures

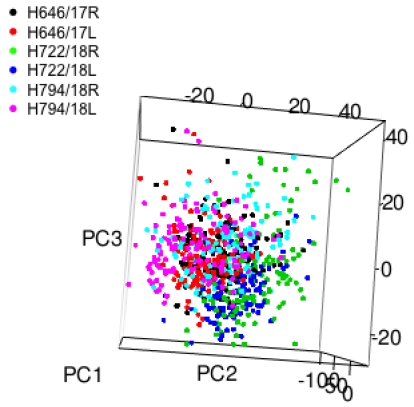
In this section, all of the patient primary cell cultures are examined together in PCA and LDA space. The specific aim of this comparison is to observe the patient primary cell cultures relative to each other in PCA and LDA space. PCA-LDA classification using the leave-one-out methodology will be applied to evaluate if the Raman methods used in this thesis can be used to accurately classify patient primary cell cultures.

The PCA analysis between all the primary patient cultures in the fingerprint region is able to ascertain that the variance captured by the first 6 principal

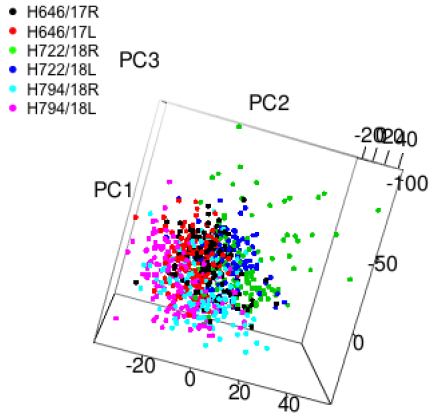
components is as follows: PC1 = 22.2%, PC2 = 9.7%, PC3 = 6%, PC4 = 3.3%, PC5 = 2.5% and PC6 = 1.9%. This in turn means that the first three principal components capture 37.8% and the first six capture 46.9% of the total variance of the sample. Cumulatively, 782 principal components are required to capture total variance of the sample. Three rotations in PCA space, using the first three principal components, are displayed in Figures 6.9(a), 6.9(b) and 6.9(c). Additionally, three rotations in LDA space using 60 principal components, capturing 67.7% of the variance in the sample, are displayed in Figures 6.9(d), 6.9(e) and 6.9(f)



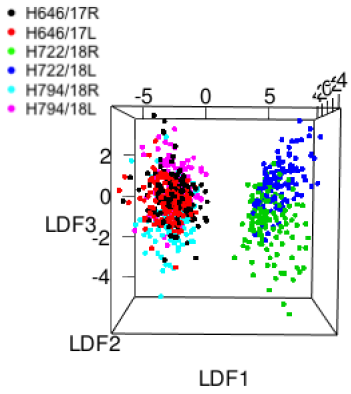
(a) PCA rotation 1



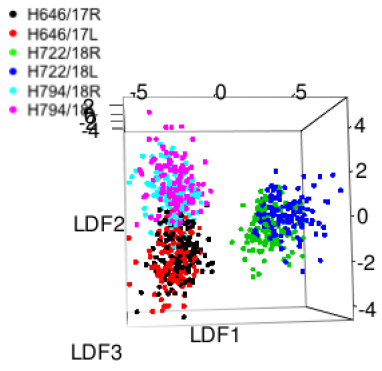
(b) PCA rotation 2



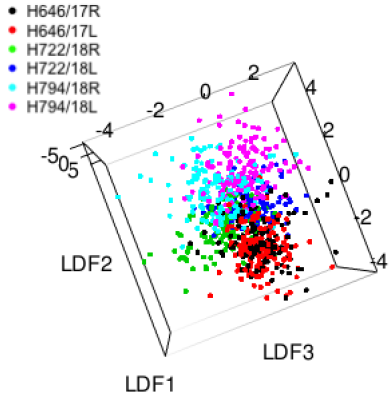
(c) PCA rotation 3



(d) LDA rotation 1



(e) LDA rotation 2



(f) LDA rotation 3

Figure 6.9: Comparison of all patient primary cell cultures in the fingerprint region. PCA space using the first three principal components is used to compare all standard cell lines at different angles in Figs. (a), (b) & (c). LDA analysis using 60 PCs (67.7% population variance capture) has been plotted using LDF1, LDF2 and LDF3, is shown (d), (e) & (f).

The 3D PCA comparison of H646/17R, H646/17L, H722/18R, H722/18L, H794/18R, H794/18L is shown for the fingerprint region in Figures 6.9(a), 6.9(b) and 6.9(c). It can be seen in all three rotations of PCA space that there is separation happening between different patient primary cell cultures. Therefore, the corresponding PCA loadings 1,2 and 3 are investigated to determine where separation is occurring, in Figures 6.10, 6.11 and 6.12.

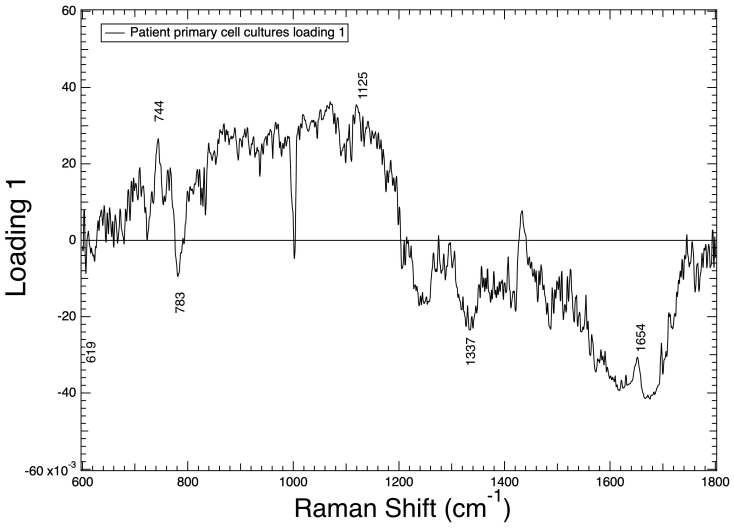


Figure 6.10: The associated loading for PC1, capturing 22.2% of the variance between all patient primary cultures in the fingerprint region, is displayed and annotated. Loading 1 is capturing variance at 744 cm^{-1} (DNA), 783 cm^{-1} (DNA/RNA), 1125 cm^{-1} (lipids proteins and carbohydrates), 1337 cm^{-1} (proteins and DNA/RNA), and 1654 cm^{-1} (proteins and lipids).

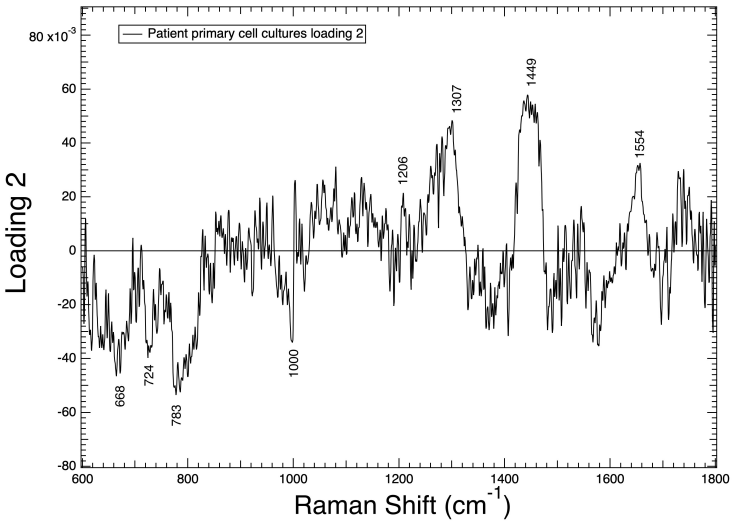


Figure 6.11: The associated loading for PC2, capturing 9.7% of the variance between all patient primary cell cultures in the fingerprint region, is displayed and annotated. Loading 2 is capturing variance at 668 cm^{-1} (proteins), 724 cm^{-1} (DNA), 783 cm^{-1} (DNA/RNA), 1000 cm^{-1} (proteins), 1232 cm^{-1} (protein, DNA/RNA), 1307 cm^{-1} (proteins, lipids and nucleic acids), 1449 cm^{-1} (lipids, proteins and DNA/RNA), 1554 cm^{-1} (tryptophan) and 1657 cm^{-1} (proteins and lipids).

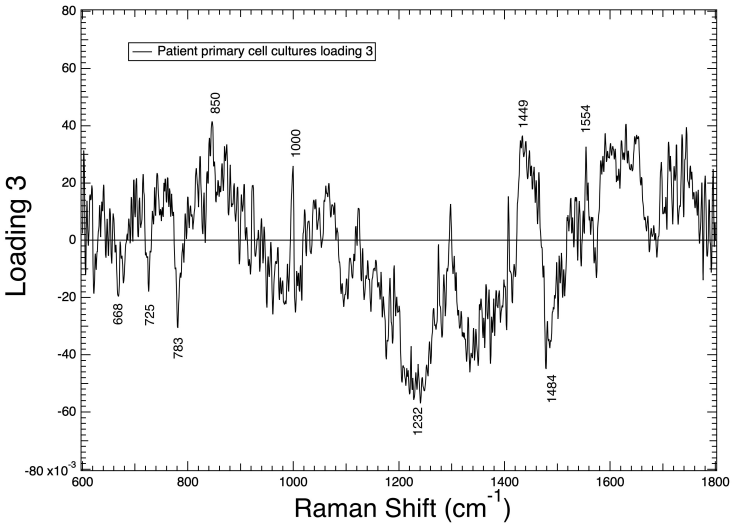


Figure 6.12: The associated loading for PC3, capturing 9.7% of the variance between all patient primary cell cultures in the fingerprint region, is displayed and annotated. Loading 3 is capturing variance at 783 cm^{-1} (DNA/RNA), 850 cm^{-1} (proteins), 1000 cm^{-1} (proteins), 1125cm^{-1} (lipids, proteins and carbohydrates), 1449 cm^{-1} (lipids, proteins and DNA/RNA), 1484 cm^{-1} (protein and DNA/RNA) and 1554 cm^{-1}.

Key peak positions have been extracted from loadings 1, 2 and 3 in Figures 6.10, 6.11 and 6.12. In loading 1, variance was observed at: 744 cm^{-1} (DNA), 783 cm^{-1} (DNA/RNA), 1125 cm^{-1} (lipids proteins and carbohydrates), 1337 cm^{-1} (proteins and DNA/RNA), and 1654 cm^{-1} (proteins and lipids). In loading 2 variance was observed at: 668 cm^{-1} (proteins), 724 cm^{-1} (DNA), 783 cm^{-1} (DNA/RNA), 1000 cm^{-1} (proteins), 1232 cm^{-1} (protein, DNA/RNA), 1307 cm^{-1} (proteins, lipids and nucleic acids), 1449 cm^{-1} (lipids, proteins and DNA/RNA), 1554 cm^{-1} (tryptophan) and 1657 cm^{-1} (proteins and lipids). Lastly, in loading 3, peaks were located at 668 cm^{-1} (proteins), 783 cm^{-1} (DNA/RNA), 850 cm^{-1} (proteins), 1000 cm^{-1} (proteins), 1125cm^{-1} (lipids, proteins and carbohydrates), 1449 cm^{-1} (lipids, proteins and DNA/RNA), 1484 cm^{-1} (protein and DNA/RNA) and 1554 cm^{-1}. These loadings will be considered for PIR analysis in Section 6.4.

The separation observed in the LDA space in Figures 6.9(d), 6.9(e) and 6.9(f) shows that there is clustering happening between all patient primary cell cultures. It is noted in all diagrams that the H794/18 primary cell culture is separating along LDF3. Whilst not as clear, it can be seen that, although H646/17 and H722/18 are overlapping, separation is occurring. The use of PCA-LDA as a prediction model for primary patient cell cultures in the fingerprint region is shown in Table 6.3.

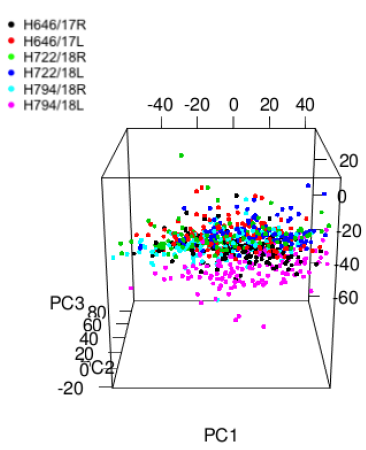
Table 6.3: PCA-LDA leave-one-out cross-validation is performed on primary patient cultures in the fingerprint region. This used a total of 60 principal components accounting for 67.7% total population variance. The corresponding confusion matrix results are shown in the table below

Cell line	**H646/17R**	**H646/17L**	**H722/18R**	**H722/18L**	**H794/18R**	**H794/18L**
H646/17R	102	18	0	0	9	9
H646/17L	31	102	0	0	4	1
H722/18R	0	0	97	26	0	0
H722/18L	0	0	26	95	0	0
H794/18R	7	5	0	0	83	36
H794/18L	5	4	0	0	37	94

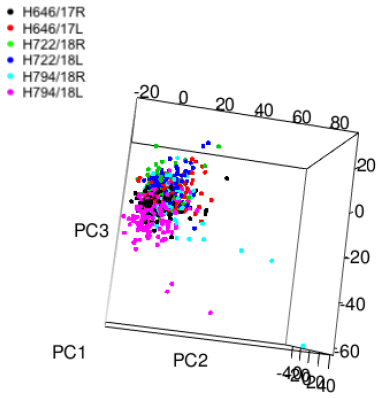
The PCA-LDA cross-validation model was able to correctly classify 72.4% of the Raman spectra. The associated misclassification of 27.6% in Table 6.3 is attributed to the misclassification between patients normal and malignant cells. Further contributions are found in misclassifications between patients. However, some of these are expected; the largest misclassification between patients occurred

between H646/17R (normal) and H794/18R (normal). This shows that whilst most of the misclassification is attributed to patients' normal and cancerous cells, there is misclassification happening due to the overlap between normal cells between patients. This PCA-LDA comparison of the fingerprint region confirms that Raman spectroscopy can classify the patient primary cell cultures in the fingerprint region with reasonable accuracy.

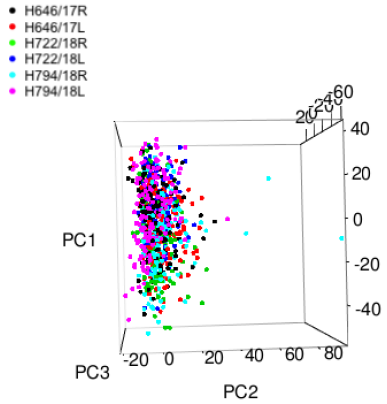
The PCA analysis between all the primary patient cultures in the high wavenumber region is able to ascertain that the variance captured by the first 6 principal components is as follows: PC1 = 39.2%, PC2 = 8%, PC3 = 6.3%, PC4 = 3.1%, PC5 = 2.6% and PC6 = 1.3%. This in turn means that the first three principal components capture 53.5% and the first six capture 60.6% of the total variance of the sample. Cumulatively, 782 principal components are required to capture total variance of the sample. Three rotations of PCA space using the first three principal components are displayed in Figures 5.95(a), 5.95(b) and 5.95(c). Furthermore, three rotations of LDA space, using 60 principal components, capturing 86.7% of the variance in the sample are displayed in Figures 6.13(d), 6.13(e) and 6.13(f)



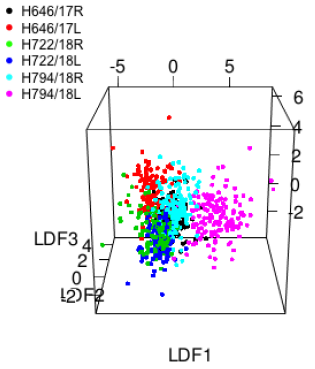
(a) PCA rotation 1



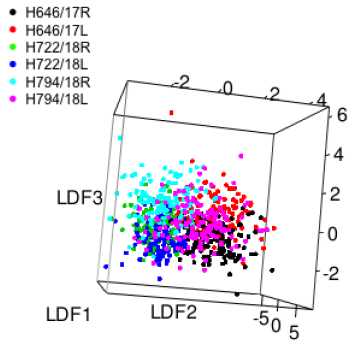
(b) PCA rotation 2



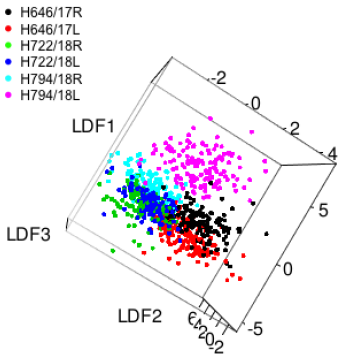
(c) PCA rotation 3



(d) LDA rotation 1



(e) LDA rotation 2



(f) LDA rotation 3

Figure 6.13: Comparison of patient primary cell cultures in the high wavenumber region. PCA space using the first three principal components is used to compare all patient primary cell cultures at different angles in Figs. (a), (b) & (c). LDA analysis using 60 PCs (86.7% population variance capture) has been plotted using LDF1, LDF2 and LDF3, is shown (d), (e) & (f).

The 3D PCA comparison of H646/17R, H646/17L, H722/18R, H722/18L, H794/18R, H794/18L is shown for the high wavenumber region in Figures 6.13(a), 6.13(b) and 6.13(c). Although not clear, it can be seen in all three rotations of PCA space that there is separation happening between different patient primary cell cultures. To find the biomolecular variances occurring in PCA analysis, the corresponding PCA loadings 1,2 and 3 are investigated, in Figures 6.14, 6.15 and 6.16.

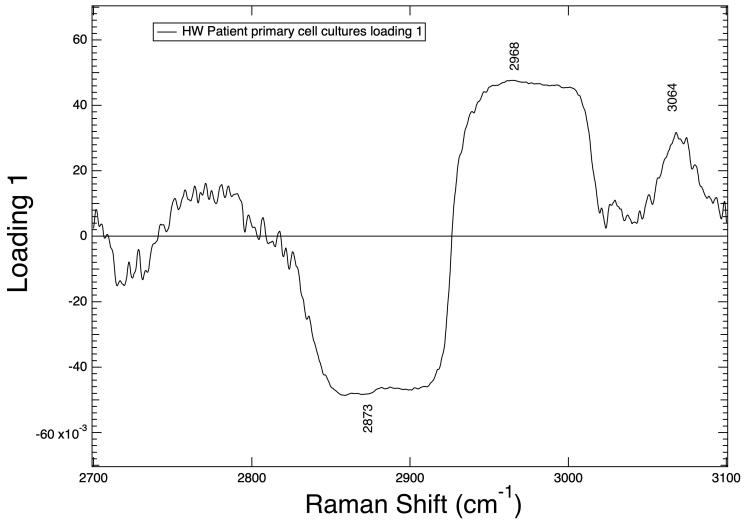


Figure 6.14: The associated loading for PC1, capturing 39.2% of the population variance between all patient primary cell cultures in the fingerprint region, is displayed and annotated. Loading 1 is capturing variance at 2873 cm^{-1} (lipids and proteins), 2967 cm^{-1} (lipids) and 3064 cm^{-1} (lipids). It is noted that the 2873 cm^{-1} and 2967 cm^{-1} peak locations, although not well defined, are placed for reference.

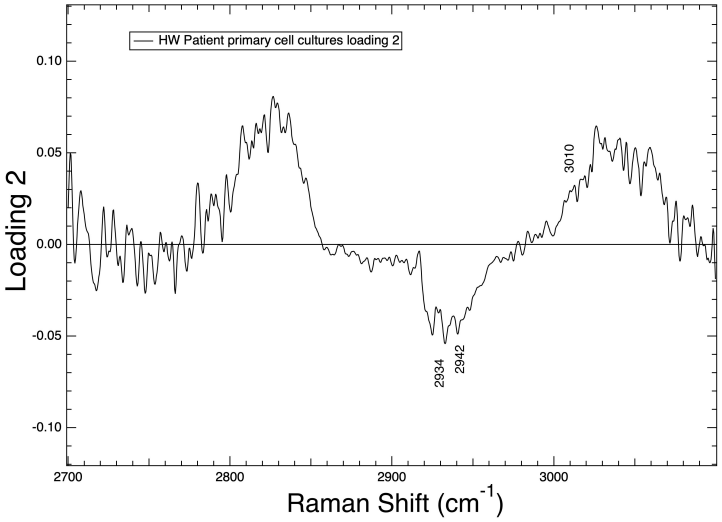


Figure 6.15: The associated loading for PC2, capturing 8% of the population variance between all patient primary cell cultures in the fingerprint region, is displayed and annotated. Loading 2 is capturing variance at 2934 cm^{-1} (protein and lipids), 2942 cm^{-1} (protein and lipids) and 3010 cm^{-1} (lipids).

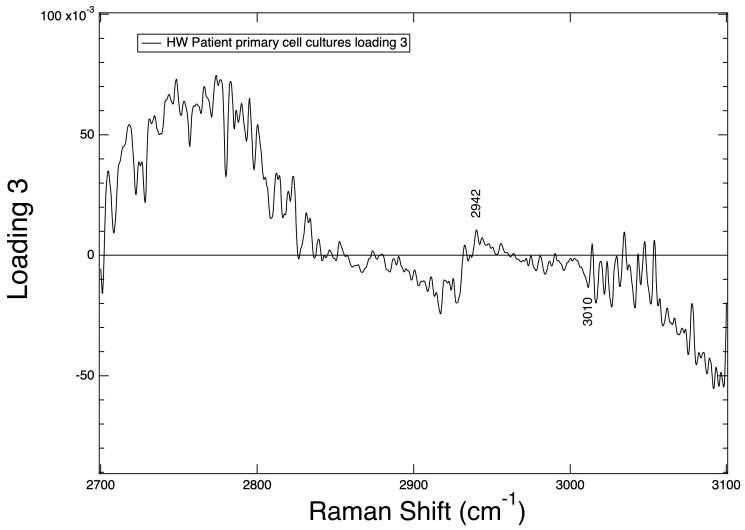


Figure 6.16: The loading representative of PC3 between all primary patient cultures in the high wavenumber region is displayed and annotated. In loading 3, capturing 6.3% of the sample variance, it can be seen that the profile is representative of background variance. This is demonstrated by the large variance observed in the region 2700-3100 cm^{-1}. Although not being the main sources of variance in Loading 3, variance is captured at 2942 cm^{-1} (lipids and proteins), and 3010 cm^{-1} (lipids).

It can be observed in the high wavenumber loadings (Figures 6.14, 6.15 and 6.15) that there is a significant contribution to the variance from the background

profile. This variance is particularly dominating in loading 3. It is for this reason that only the most well defined peak positions are considered. In loading 1 the only well defined peak was found at 3064 cm^{-1} (lipids). In loading 2, three peaks were identified as contributing to the variance, 2934 cm^{-1} (protein and lipids), 2942 cm^{-1} (protein and lipids) and 3010 cm^{-1} (lipids). In loading 3, two peaks were identified as contributing to variance, 2942 cm^{-1} (lipids and proteins) and 3010 cm^{-1} (lipids).

The 3D LDA comparison of H646/17R, H646/17L, H722/18R, H722/18L, H794/18R, H794/18L is shown for the high wavenumber region in Figures 6.13(a), 6.13(b) and 6.13(c). It can be seen in all three rotations of LDA space that there is separation happening between different patient primary cell cultures. It is also observed that H794/18L is showing clear separation along LDF1. It can also be seen in Figure 6.13(c) that separation is occurring between individual patient sets at an angle along LDF1 and LDF3. The use of PCA-LDA as a prediction model for primary patient cell cultures in the high wavenumber region is shown in Table 6.4.

Table 6.4: PCA-LDA leave-one-out cross-validation is performed on primary patient cultures in the high wavenumber region. This used a total of 60 principal components accounting for 86.7% total population variance. The corresponding confusion matrix results are shown in the table below:

Cell line	**H646/17R**	**H646/17L**	**H722/18R**	**H722/18L**	**H794/18R**	**H794/18L**
H646/17R	117	7	5	2	2	3
H646/17L	12	112	6	2	6	0
H722/18R	3	5	95	18	2	0
H722/18L	0	2	15	94	10	0
H794/18R	1	4	9	7	113	0
H794/18L	4	0	1	2	6	127

The PCA-LDA cross-validation model was able to correctly classify 74.1% of the Raman spectra. The associated 25.9% error in Table 6.4 is attributed to the misclassification between patients' normal and malignant cells. It is observed in Table 6.4 that there is a much higher predictive classification accuracy for the primary patient cell culture H794/18. Furthermore, it is noted that the largest misclassification between primary patient cell cultures took place between H794/18R and H722/18R. This, as stated previously when describing the PCA-LDA leave-one-out analysis of the fingerprint region, is expected due to both cell cultures

representing normal cells. This PCA-LDA comparison of the high wavenumber region confirms that Raman spectroscopy can classify the patient primary cell cultures with reasonable accuracy.

It can be seen, using PCA-LDA leave-one-out analysis, that Raman spectroscopy can classify the patient primary cultures in both the fingerprint and high wavenumber region with reasonable accuracy. Furthermore, through comparison of Tables 6.3 and 6.4, that Raman spectroscopy of the high wavenumber region (74% correct classification) was able to marginally outperform the fingerprint region (72.4%) in terms of classification. A common trend in the cross-validation was that in both regions the highest percentage of misclassification was attributed to overlap between the normal cells and cancer cells of individual patients. Additionally, it was found between patients that the highest rate of misclassification occurred in the normal cells of patients. This would signify that there is less variance between the normal cell cultures. There is currently no primary prostate patient cell culture involving Raman spectroscopy to compare this result to, as far as the author is aware. Using hyper-spectral Oliveira *et al.* (2019) analysed patient derived human thyroid cells with 97% classification accuracy [286]. However, the results from Oliveira *et al.* (2019) cannot be used for comparison in this thesis as they are measuring a different biological system using a different Raman technique.

The PCA-LDA cross-validation results from the patient primary cells are considerably lower in predictive accuracy than the standard cell lines. The PCA-LDA model in Chapter 5 section 5.7 was able to correctly classify 93.7% of the standard cell line Raman spectra in the fingerprint region and 96.1% of spectra in the high wavenumber region. It should be stated that this drop in accuracy of the PCA-LDA model is not wholly dependent on the number of normal primary cell cultures that were included in it. It was observed in both results of the PCA-LDA model, in Tables 6.4, the majority of misclassifications were between normal and malignant cell cultures of individual patients. It could also be said that the higher accuracy of the PCA-LDA in the standard cell lines is due to the larger spectrum of disease considered. However, in Tables 5.4 and 5.5, it can be seen

that the misclassifications in the standard cell lines are between the metastatic cell lines and benign hyperplastic cell line. It is hypothesised that, due to the fact that standard cell lines only have one undifferentiated cell type, that it will increase the likelihood of separation in PCA and LDA between different prostate standard cell lines. In the patient primary cell cultures there are stem cells, transit amplifying cells, committed basal cells and luminal cells. The much wider complete cellular heterogeneity, unlike standard cell lines, is thought to be the reason lower PCA-LDA cross-validation prediction values were achieved. It is suggested that the much higher PCA-LDA predictive values of the standard cell lines have significant implications for studies that are based on PCA and LDA analysis. It would infer that studies that have used prostate standard cell lines to measure between disease states may have experienced an artificially higher degree of separation due to the lack of heterogeneity in the cell lines. In future studies, this thesis recommends, where possible, that Raman studies using PCA and LDA analysis on Raman spectra use patient derived primary cell cultures to capture the true heterogeneity between the normal and disease state.

6.4 Comparison of patient primary cell cultures to previously identified biomarkers relating to disease state

In Chapter 5, Subsection 5.7.2, a set of PIR markers that can identify differences between PNT2-C2 (normal), BPH-1 (benign), P4E6 (localised malignancy) and DU145/LNCaP/PC3 (metastatic malignancy). This was carried out after completing detailed Raman characterisation of each standard cell line that was used in this thesis. In this Section the PIR biomarkers that were identified as being able to discriminate between the standard cell lines are applied to the patient primary cell cultures.

It is highlighted that the PIR marker R783/699 can not be mapped to the primary patient cultures due to the associated 699 cm^{-1} band not being present

in the fittings of the patient primary cell cultures.

The first three PIR markers identified for separating disease state in the standard cell lines are analysed using the peak intensity of the 1125 cm^{-1} peak. The 1125 cm^{-1} band is representative of lipids, proteins and carbohydrates, specifically, a protein C-N stretch, C-C stretch in lipids and C-C stretch in carbohydrates. The mapping of the PIR markers, R1001/1125, R715/1125 and R1448/1125, is mapped to the patient primary cell cultures below.

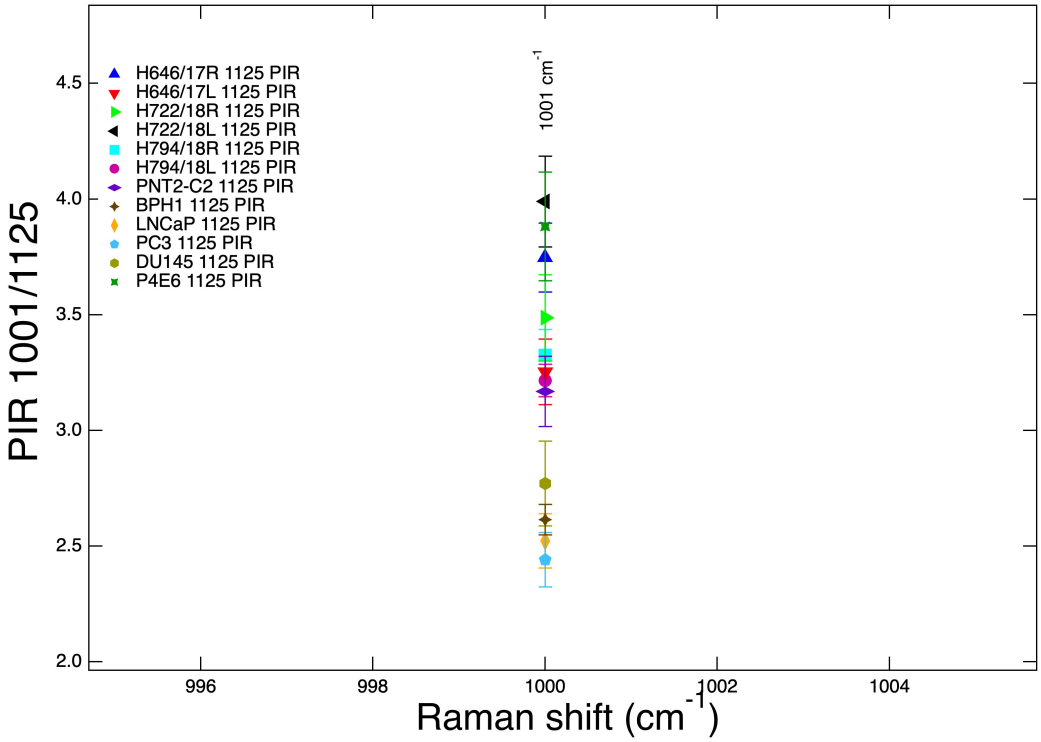


Figure 6.17: It was observed in the standard cell lines, Chapter 5, Section 8, Figure 5.103(a), that the R1001/1125 ratio could be used to separate the normal cell line PNT2-C2 from the diseased states outside of statistical uncertainty. However, when this ratio has been applied to the primary prostate cell cultures, the malignant patient primary cell culture H646/17L is overlapping with PNT2-C2. The corresponding normal patient primary cell culture H646/17R (normal) is overlapping with P4E6 (localised cancer). It is also observed that the H794/18 can not be statistically separated into its sub cell cultures using this ratio. Further observations show that the H722/18R (normal) is grouping with PNT2-C2 and H722/18L is grouping with P4E6. These observations suggest that the diagnostic marker established using the cell lines fails to correctly classify the patient primary cell cultures.

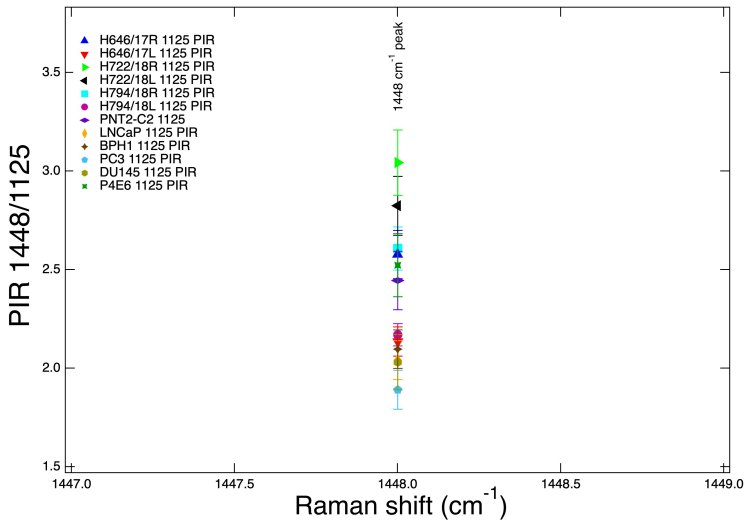


Figure 6.18: It was observed in the standard cell lines, Chapter 5, Section 8, Figure 5.103(b), that the R1448/1125 ratio showed PNT2-C2 and P4E6 clustering away from the other diseased states, outside of statistical uncertainty. When patient cell lines are mapped to the R1448/1125 ratio both H646/18L and H794/18L overlap the with metastatic disease states. However, it is also observed that H722L is overlapping with PNT2-C2.

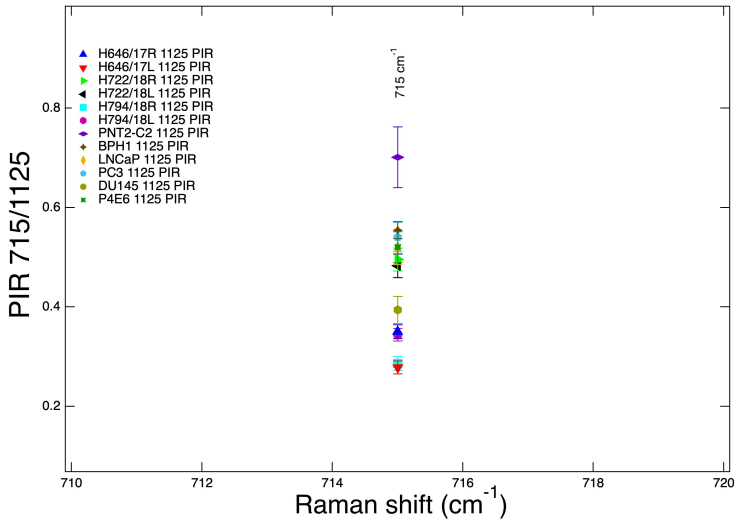


Figure 6.19: It was observed in the standard cell lines, Chapter 5, Section 8, Figure 5.103(c), that the R715/1125 ratio was able to separate PNT2-C2 from the rest of the diseased states outside of statistical uncertainty. This is still the case when the patient primary cell cultures are also considered. However, it is observed that H794/18R (normal) and H646/17L (cancer) are overlapping. It is also seen that H794/18L (cancer) and H646/17R (normal) are overlapping. This demonstrates the variability in the biomolecular peak intensity profile of the patient primary cell cultures.

The mapping of the PIR markers, R1001/1125, R715/1125 and R1448/1125, in Figures 6.17, 6.18 and 6.19, demonstrates that the PIR biomarkers established using the 1125 cm^{-1} peak intensity, are not suitable for the classification of patient primary cell cultures. Figure 5.103(a) shows that the peak intensity ratio R1001/1125 classifies H646/17L (cancer) as being normal due to it overlapping with PNT2-C2. Furthermore, Figure 5.103(a), also shows that the H646/17R (normal) is overlapping with localised cancer, P4E6. Figure 6.18 demonstrates that when cell lines are mapped to the R1448/1125 ratio, both H646/18L (cancer) and H794/18L (cancer) overlap the with metastatic disease states. However, it is also observed that H722/18L (cancer) sample is overlapping with PNT2-C2. In Figure 6.19 further evidence in the wider variance of patient primary cell cultures, it is observed that H794/18R (normal) and H646/17L (cancer) are overlapping. It is also seen that H794/18L (cancer) and H646/17R (normal) are overlapping.

The PIR markers identified for separating disease state in the standard cell lines are analysed using the peak intensity of the 1254 cm^{-1} band (amide III : CH_2 wagging) are now considered against the primary patient cell cultures, in Figures 6.20, 6.21 and 6.22.

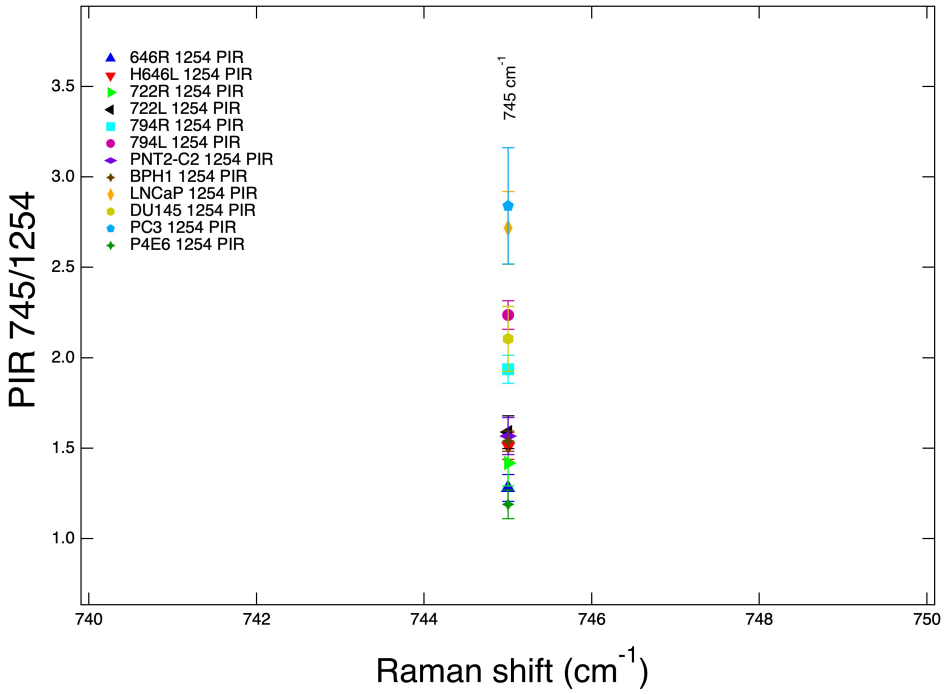


Figure 6.20: It was observed in the standard cell lines, Chapter 5, Section 8, Figure 5.103(d), that the peak intensity ratio R745/1254 was able to discriminate between malignant and non-malignant states. When the R745/1254 peak intensity ratio is applied to patient primary cultures the H722/18R (normal) and H722/L (cancer) patient primary cell cultures both overlap with the PNT2-C2 and BPH-1, previously non-malignant clustering. The patient primary cell culture H646/17R (normal) and H646/17L (cancer) also overlap with PNT2-C2 and BPH-1. In contrast, H794/18R (normal) and H794L (cancer), although separating outside of statistical uncertainty, overlap with the standard cell line DU145 (metastatic). Once again, this highlights the higher variation observed within primary patient cell cultures.

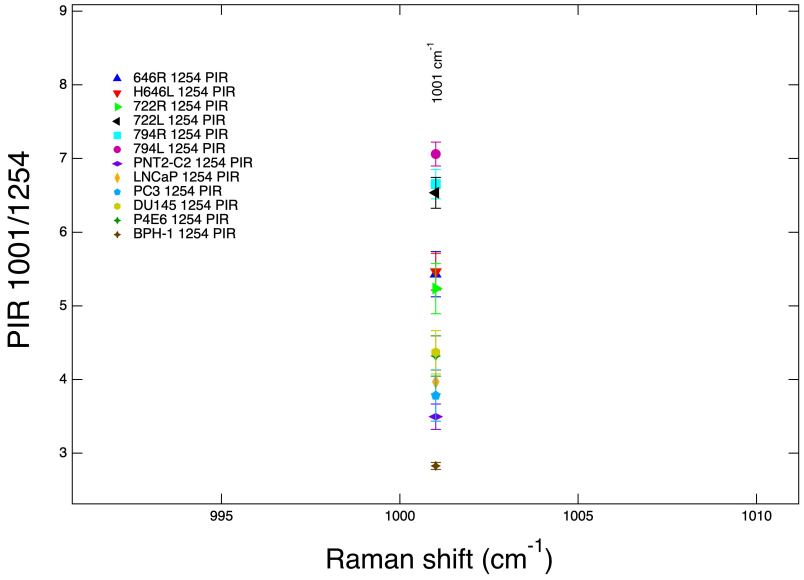


Figure 6.21: It was observed in the standard cell lines, Chapter 5, Section 8, Figure 5.103(e), that BPH-1 was separating from the rest of standard cell lines, using the PIR ratio 1001/1254. When the patient primary cell cultures are considered using the PIR ratio 1001/1254 ratio, BPH-1 is still separated from the rest of the standard cell lines and primary cell cultures.

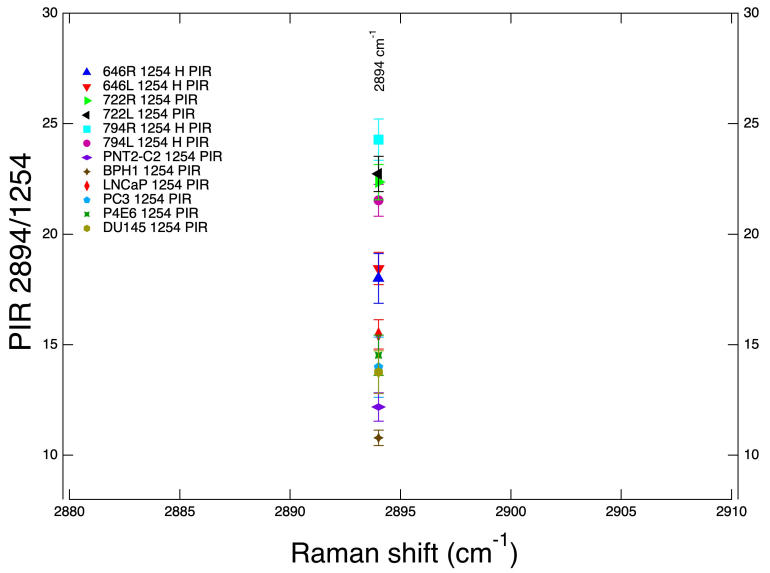


Figure 6.22: It was observed in the standard cell lines, Chapter 5, Section 8, Figure 5.103(f), that BPH-1 was separating from the rest of standard cell lines, using the PIR ratio 2894/1254. When the PIR ratio 2894/1254 is applied to the patient primary cell cultures the BPH-1 ratio is still separating from all other samples. It is also noted that the patient primary cell cultures have separated from the standard cell lines.

The first ratio considered, using the 1254 cm^{-1} peak intensity ratio, was R745/1254, in Figure 6.20. The R745/1254 ratio, previously shown to discriminate between malignant and non-malignant states, could not discriminate between the normal and cancer patient primary cell cultures. It was observed that the normal and malignant cells from H722/18 and H646/17, patient primary cell cultures, both overlap with the PNT2-C2 and BPH-1, previously observed as non-malignant clustering. In contrast, H794/18R and H794L, although separating outside of statistical uncertainty, overlap with the standard cell line DU145. The second ratio examined was R1001/1254, in Figure 6.21. The R1001/1254 ratio had been previously determined to be able to discriminate BPH-1 from other standard cell lines. It was observed, in Figure 6.21, that BPH-1 could still be discriminated against outside of statistical uncertainty using the primary patient cell cultures. Lastly, the ratio 2894/1254, in Figure 6.22, which was also used to discriminate BPH-1 from other cell lines previously, was still able to do so after the inclusion of primaries. This still indicates that a higher relative intensity of the 1254 peak cm^{-1} (amide III) against the 1001 cm^{-1} (proteins) and 2894 (CH stretch of lipids and proteins)) may be an indicator of benign hyperplasia. When the PIR ratio 2894/1254 is applied to the patient primary cell cultures, the BPH-1 ratio is still separating from all other samples. It is also noted that the patient primary cell cultures have separated from the standard cell lines. This could be an early indicator of fundamental differences in the 1254 cm^{-1} (amide III) band and the 2894 cm^{-1} band (lipids and proteins) between standard cell lines and primary cell cultures.

Two further discriminatory PIR biomarkers using the high wavenumber 2873 cm^{-1} peak intensity ratio (CH asymmetric/symmetric stretch in lipids and proteins) were used to discriminate between the standard cell lines. This discriminatory marker is applied to the patient primary cell cultures, in Figures 6.23 and 6.24.

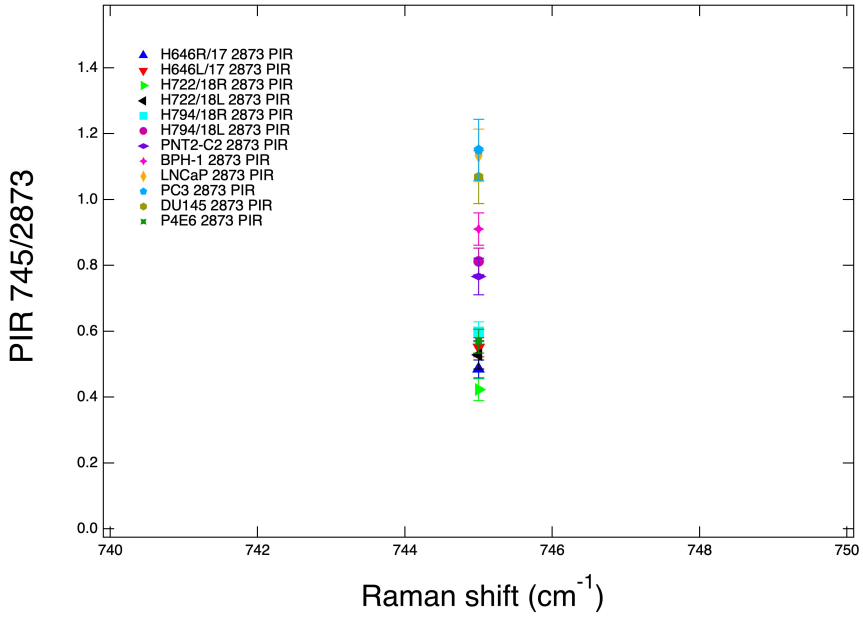


Figure 6.23: It was observed in the standard cell lines, Chapter 5, Section 8, Figure 5.104(b), that the ratio R745/2873 could be used to discriminate the standard cell lines into normal, benign hyperplastic, localised and metastatic states. However, when the ratio R745/2873 was applied to primary patient cell cultures it was found that the patient primary cancer cell culture H794/18L overlapped with PNT2-C2. There was still separation outside of statistical uncertainty between the metastatic standard cell lines and BPH-1 to the rest of the sample. Lastly, it is observed that H794R is overlapping with P4E6.

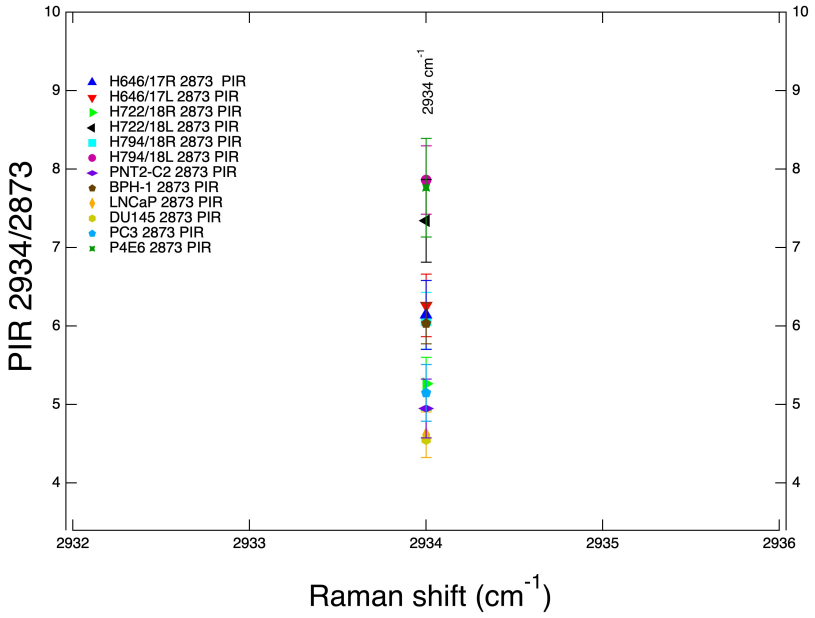


Figure 6.24: It was observed in the standard cell lines, Chapter 5, Section 8, Figure 5.104(c), that the ratio 2934/2873 was able to discriminate BPH-1 and P4E6 from the rest of the standard cell lines. The ratio 2934/2873, when considered for standard cell lines and patient primary cell cultures shows that P4E6, H794/18L (cancer) and H722/18L (cancer) have clustered together and can be discriminated from the rest of the samples. BPH-1 (benign hyperplastic) can no longer be discriminated against, due to H646/17L (cancer) and H646/17R (normal) both overlapping it.

The first ratio considered using the 2873 cm^{-1} peak intensity ratio was R745/2873, in Figure 6.23. The R745/2873 ratio was previously shown to discriminate standard cell lines into normal, benign hyperplastic, localised and metastatic states. However, when the ratio R745/2873 was applied to primary patient cell cultures it was found that the patient primary cancer cell culture H794/18L overlapped with PNT2-C2. There was still separation outside of statistical uncertainty between the metastatic standard cell lines and BPH-1 to the rest of the sample. Lastly, it is observed that H794R is overlapping with P4E6. Whilst this ratio shows that the variance of the patient primary cell cultures is much higher, it was able to discriminate the metastatic and benign states. This would prove useful for further exploration into the biomolecular differences between standard cell lines and primary patient cell cultures. The second ratio R2934/2873 showed that P4E6, H794/18L and H722/18L have clustered together and can

be discriminated from the rest of the samples. However the H646/17L (cancer) sample has clustered with BPH-1. Once again, this highlights the variance that occurs between the patient primary cell cultures.

Lastly, the peak intensity ratio R3015/2851, shown in Figure 6.25, demonstrated that in this thesis it was not possible to discriminate between disease states using standard cell lines, using this ratio. It was observed that the standard cell line DU145 could be discriminated individually. This ratio is investigated for primary patient cell cultures alongside the standard cell lines in Figure 6.25.

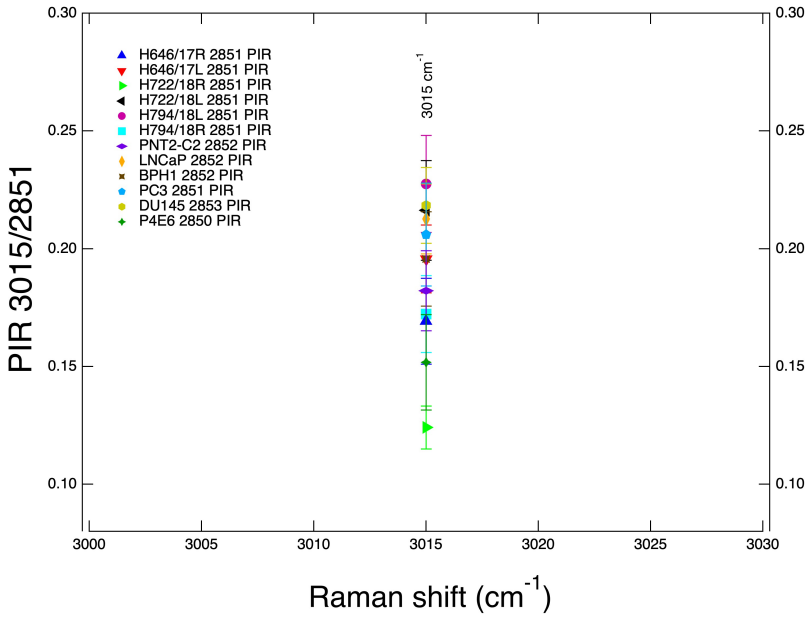


Figure 6.25: It was observed in the standard cell lines, Chapter 5, Section 8, Figure 5.104(d), that it was not possible to discriminate between standard cell line types. It can be seen that there is separation occurring in H722/18 and H794/18. However, there is no separation in the patient primary cell culture H646/17.

The peak intensity ratio marker, R3015/2851, shown in Figure 6.25, was not able to discriminate the standard cell lines based on disease state. Although discrimination was achieve using H722/18 and H794/18 the primary patient cell culture H646/17 did not separate.

Ultimately, the analysis in this section shows that there are multiple inconsistencies of establishing diagnostic markers using standard cell lines, as they do not account for the wider variability and heterogeneity of the patient cell

cultures. This has been demonstrated in the Figures above by the repeated misclassification of the patient primary cell cultures. It is noted that the standard cell line BPH-1 could still be discriminated against using the biomarkers established through analysis of the standard cell lines. However, there was no inclusion of a benign prostate hyperplasia primary culture involved in this study. To measure the performance of the discriminatory markers, derived using standard cell lines for BPH-1, an inclusion of a benign prostate primary cell culture would have to be included.

6.5 Conclusion

The main objective of this chapter was to understand if it was possible to implement a predictive model for the analysis of patient primary cell cultures. The primary patient cell cultures are much more representative of the complex heterogeneity, in the patient environment. This is due to primary patient cell cultures containing contain stem cells, transit amplifying cells, committed basal cells and luminal cells combining to give complete cellular heterogeneity unlike standard cell lines that have only one undifferentiated cell type. To the best of the author's knowledge this the first single cell study on patient primary cell cultures.

To understand the complex cellular heterogeneity of the derived primary patient samples, convergence testing was carried out. Whilst there are variations between different primary cultures, the Raman spectra for both the fingerprint and the high wavenumber region converged at around 50-80 spectra. This was noted as being of a similar range to the standard cell lines.

As the primary patient cell data was converged it was possible to use multivariate analysis to investigate the normal and cancerous cells of the primary patient cultures. It was seen in PCA analysis of the fingerprint and high wavenumber region that separation was occurring between primary patient cultures. It was also seen that separation was occurring between patient's derived normal and cancerous cells. A PCA-LDA prediction model using leave-one-out analysis correctly classified 72.4% percent of the primary patient cells in the fingerprint

region. PCA-LDA analysis was also carried out for the high wavenumber region for patient cells, correctly classifying 74% percent of cells. It is noted that some of the misclassifications occurred between the normal and cancer subsets from different patients. This highlights the challenges of classifying normal vs cancer cells in different patients. However, given that only three patients are involved in this comparison it is not possible to say whether there are any trends that can be observed in inter-patient separation. It is proposed that further work could address this with the inclusion of many more patients to observe if specific patients cluster together or separate. Despite the inter-patient variation it has been shown that Raman spectroscopy can be used to classify individual patient derived primary cell cultures to reasonable predictive accuracy.

The PCA-LDA cross-validation results from the patient primary cells are considerably lower in predictive accuracy than the standard cell lines observed in the previous chapter. The PCA-LDA model in Chapter 5 section 5.7 was able to correctly classify 93.7% of the standard cell line Raman spectra in the fingerprint region and 96.1% of spectra in the high wavenumber region. In the comparison of patient primary cell cultures, in Chapter 6, it was observed in the confusion matrix generated from the PCA-LDA cross-validation that the majority of misclassifications were between the normal and malignant cell cultures of individual patients. This is primarily due to the more complex heterogeneity of patient primary cell cultures. It is suggested that because standard cell lines only have one undifferentiated cell type they experience a much higher degree of separation. However, the high predictive classification rates, due to lack of heterogeneity, are not indicative of a true comparison between the normal and diseased state.

To evaluate the use of discriminant markers of disease state determined by the standard cell lines in Chapter 5, the same PIR markers were applied to the primary cell cultures. To evaluate the performance of this the standard cell lines and patient cell lines were cross compared. It was determined that specific relationships determined using standard cell lines to measure a spectrum of diseased states do not account for the variability of individual patients primary cell cultures.

On the basis of the results found in this chapter it can be shown that prostate standard cell lines are not an accurate model for patient derived primary cell cultures.

CHAPTER 7

CONCLUSION

This thesis has shown that Raman spectroscopy can be used to successfully characterise and discriminate both prostate standard cell lines and prostate primary cell cultures. Raman spectroscopy has been applied to characterise and investigate a standard cell line in its live and air-dried state, to quantify the biomolecular changes that take place in the air-drying process. Furthermore, using Raman spectroscopy, this thesis analyses multiple standard cell lines that are representative of a full spectrum of prostate disease: normal, benign hyperplasia, localised cancer and metastatic cancer. This was done to build a comprehensive model of prostate disease using standard cell lines. Also, using Raman spectroscopy, this thesis examined patient derived primary cell cultures that have a much more complex heterogeneity than standard cell lines. Lastly, using different methods of analysis, the effectiveness of the standard cell lines to characterise disease states was compared against the primary cell cultures.

In Chapter 4, Raman spectroscopy was used to characterise and discriminate between the normal prostate standard cell line PNT2-C2, in its live and dried state. There were multiple biomolecular differences that were found in this comparison. The PCA analysis and corresponding loading results, in the fingerprint region, indicated that there was a large amount of variance attributed to changes in the

DNA/RNA, proteins and lipids between the live and dried state. Additionally, PCA analysis of the high wavenumber region detailed further biomolecular differences, between the proteins and lipids bands, between the live and dried cells. The peaks corresponding to high variance, in both the fingerprint and high wavenumber PCA loadings, were used to inform the PIR analysis, to detail relative biomolecular differences between PNT2-C2 in its live and dried state. It was found that the 745 cm^{-1} peak intensity ratio was identified as a robust biomarker to discriminate between the live and dried cells. The 745 cm^{-1} band is attributed to the thymine ring breathing mode (DNA) [105, 128, 224, 248]. It was found that there was a much lower relative intensity of the 745 cm^{-1} thymine peak in the PNT2-C2 dried sample. This was attributed to a change in DNA conformation state. In the high wavenumber region the relative peak intensity of the 3015 cm^{-1} band, representative of unsaturated lipids [64, 128, 251, 252], was also a robust biomarker to differentiate between the live and dried sample. The 3015 cm^{-1} peak intensity ratio showed that there is a decrease in the relative intensity of unsaturated fatty acids in PNT2-C2 dried. The findings from this chapter enhance the understanding of the implications of using dried prostate standard cell lines for research. Although this study was based only on a small sample, one specific prostate standard cell line, it suggests that Raman studies using dried prostate standard cell lines can generate results that may result in false positives or negatives. This is based on the evidence in chapter 4 that PNT2-C2 dried cells were not representative of PNT2-C2 live cells.

In Chapter 5, Raman spectroscopy was used to fully characterise live prostate standard cell lines that covered a wide spectrum of disease. The prostate standard cell lines included in this study were PNT2-C2 (normal), BPH-1 (benign prostate hyperplasia), P4E6 (androgen-sensitive localised) LNCaP (androgen-sensitive metastatic), DU145 (androgen-independent metastatic) and PC3 (androgen-independant metastatic). Each standard prostate cell line was fully characterised in the fingerprint and high wavenumber regions. It was shown that Raman spectroscopy could differentiate between the normal PNT2-C2 cell line and each diseased state. This was done using a combination of the mean profile subtraction, PCA analysis and

PIR analysis. The corresponding biomarkers were identified between PNT2-C2 and each disease state and highlighted differences in relative intensity ratios DNA/RNA, nucleic acids, lipids and proteins. PCA and LDA analysis was also used to compare all of the prostate standard cell lines against each other, demonstrating a high degree of separation in both the fingerprint and high wavenumber regions. PCA-LDA analysis was used to correctly classify 94.8% of the Raman spectra from all prostate standard cell lines in the fingerprint region and 96.9% of spectra in the high wavenumber region. A comprehensive model of prostate disease was established using PCA and PIR analysis to discriminate between different disease states. It was found using the peak intensity ratio of the 745 cm^{-1} peak (DNA) against the 1254 cm^{-1} peak (amide III) that standard cell lines could be separated into non-malignant, localised and metastatic subsets. Furthermore, the peak intensity ratio of the 745 cm^{-1} peak (DNA) against the 2873 cm^{-1} peak (lipid and proteins), demonstrated that the prostate cell lines could be separated into normal, benign hyperplasia, localised cancer and metastatic. However, no PIRs could be determined to separate the metastatic prostate cell lines based on androgen dependency. The work in this chapter adds to the existing body of knowledge of Raman spectroscopy of prostate standard cell lines by building a complete model of prostate disease, using live cell lines. The most important limitation to the research in this chapter lies in the fact that prostate standard cell lines consist of only one undifferentiated cell type and do not account for the heterogeneity present in an *in vivo* environment.

In chapter 6, Raman spectroscopy was used for the analysis of patient primary cell cultures. These are much more complex cell cultures and unlike the standard cell lines used in chapter 5, are duly representative of the variability of disease in each patient. A PCA-LDA model was used as a predictive model for the analysis of patient primary cell cultures. The PCA-LDA model correctly classified 72.4% percent of the primary patient cells in the fingerprint region and 74% percent of cells in the high wavenumber region. These results show as a major finding that Raman spectroscopy is able to classify patient primary cell cultures to a reasonable level of predictive accuracy. The second major finding

of this chapter was that when the discriminatory model, based on PIR markers from the standard cell lines in chapter 5, was applied to the patient primary cell cultures, it could not discriminate between the normal and cancer cells of each patient, resulting in numerous misclassifications. This result demonstrated that, due to the higher heterogeneity of patient cell cultures, the standard cell lines are not duly representative of a disease state in patients. One limiting factor of this study is that all the cancer cells are attributed to a Gleason score of 7(3+4) and thus does not encompass the complete spectrum of prostate disease. This work contributes to existing knowledge in the field of Raman spectroscopy of prostate cancer, as it is thought, based on extensive literature search, that this is the first time a PCA-LDA model has been used to classify prostate primary cell cultures. Furthermore, through investigation of the literature, it is believed to be the first time that standard prostate cell lines have been compared to primary patient cell cultures using Raman spectroscopy.

The research in this thesis has resulted in raising many further questions in need of investigation. Future work could look to fully characterise the patient primary cell cultures using Raman spectroscopy. This would include PCA/LDA and PIR analysis of the normal and malignant cells in each individual patient. It was further observed that there is an additional challenge that exists in classifying normal vs cancer cells due to the variation between patients, highlighted by separation of patients in chapter 6, using PCA-LDA analysis. It is proposed that to better understand the separation of patients that the number of patients involved in future studies be greatly increased. This would allow for not only a better understanding of the increased heterogeneity of the patient primary cell cultures and individual patients but also the biomolecular differences that exist between the normal and disease states. These future results could be then be further compared to the standard prostate cell lines to see if there are any trends emerging in the biomolecular differences. Further work also needs to be done to establish the biomolecular differences that occur in different Gleason scores in patient primary cell cultures. This would potentially be able to model comprehensively the complex range of disease states that afflict patients with

prostate cancer.

APPENDIX

A.1 List of Abbreviations

ABM Anti-biotic anti-mycotic

AVG Average

BPH Benign Prostate Hyperplasia

CaF2 Calcium Fluoride

CZ Central Zone

FCS Feotal calf syrome

HBSS Hanks' Balanced Salt Solution

LDA Linear Discriminant Analysis

PBS phosphate buffered saline

PCA Principal Component Analysis

PCa Prostate Cancer

PCs Principal components

PSa Prostate-specific antigen

RP Radical prostatectomy

RMPI Roswell Park Memorial Institute

R10 RMPI +

SE Standard Error

TUFA/TFA Total unsaturated fatty acids/ total fatty acids

2SD twice the standard deviation

N/A not available

APPENDIX B

Raman Peak Positions

Table B.1: This table contains the peak positions found for PNT2-C2 live and PNT2-C2 dried as well as the corresponding average.

PNT2-C2 Live	**PNT2-C2 Dried**	**AVG**
617.9	618.5	618
640.1	640.3	640
670.6	667.0	669
697.5	N/A	N/A
713.9	715.8	715
724.3	723.2	724
745.0	745.4	745
755.9	754.3	755
781.4	780.8	781
807.3	808.6	808
824.8	824.5	825
850.1	850.2	850
874.6	875.6	875
891.8	885.5	889
934.4	932.7	934
969.5	965.4	967
1000.7	1001.2	1001
1029.2	1029.8	1030
1061.8	1060.0	1061
1088.7	1090.4	1090
1125.2	1125.5	1125
1154.5	1154.6	1155
1171.2	1171.3	1171
1205.1	1205.6	1205
1230.1	1238.8	1234
1252.9	1259.4	1256
1306.7	1306.2	1306
1336.9	1335.2	1336
1397.8	NA	N/A
1448.2	1447.7	1448
1482.9	1483.4	1483
1551.9	1551.7	1552
1576.2	1574.1	1575
1602.4	1602.7	1603
1615.1	1613.7	1614
1656.4	1656.1	1656
1681.7	1681.2	1681
2850.3	2852.4	2851
2871.2	2872.0	2872
2893.8	2895.2	2895
2932.0	2930.0	2931
2940.4	2940.3	2940
2965.5	2962.7	2964
3013.8	3014.8	3014
3063.5	3064.3	3064

Table B.2: This table contains the peak positions found for PNT2-C2 live, BPH-1, P4E6, LNCaP, DU145, PC3 as well as the corresponding average.

PNT2-C2	**BPH-1**	**P4E6**	**LNCaP**	**DU145**	**PC3**	**AVG**
617.913	619.388	616.819	619.576	618.993	618.95	619
640.14	641.497	639.765	641.632	641.412	640.57	641
670.582	672.965	671.272	673.15	670.845	677.891	673
697.513	699.424	696.436	698.629	N/A	698.128	698
713.86	714.067	713.2	715.097	716.14	714.102	714
724.301	723.527	724.461	723.355	N/A	723.358	724
744.954	745.54	743.909	745.899	745.766	746.049	745
755.907	756.44	755.994	755.294	756.737	755.716	756
781.409	782.932	781.325	783.401	783.78	783.087	783
807.311	808.831	806.215	809.626	806.563	809.033	808
824.768	826.736	824.511	824.956	821.356	824.824	825
850.148	851.498	849.908	850.558	849.428	851.003	850
874.588	875.505	875.869	875.721	N/A	876.434	876
891.842	891.542	892.644	892.669	N/A	892.576	892
934.448	933.883	934.814	935.991	932.953	933.886	934
969.459	976.052	N/A	975.786	972.826	973.608	974
1000.72	1001.32	999.927	1001.34	1000.69	1000.71	1001
1029.24	1029.43	1028.57	1029.81	1029.69	1029.19	1029
1061.82	1070.5	1062.04	1059.84	1063.52	1059.3	1063
1088.7	1090.99	1090.3	1087.17	1088.3	1087.9	1089
1125.2	1126.01	1124.49	1125.38	1124.67	1125	1125
1154.54	1155.46	1154.17	1154.54	1154.6	1153.64	1154
1171.23	1172.3	1171.04	1171.04	1170.03	1170.91	1171
1205.14	1205.83	1203.81	1205.76	1204.91	1205.1	1205
1230.08	1228.22	1232.01	1234.21	1230.79	1231.55	1231
1252.9	1249.13	1252.9	1254.47	1254.35	1250.22	1252
1306.65	1307.2	1306.48	1306.83	1308.86	1307.97	1307
1336.88	1337.31	1336.38	1337.49	1337.85	1337.33	1337
1397.76	N/A	1400.18	1396.48	N/A	1396.38	1398
1448.17	1449.6	1448.77	1448.07	1449.93	1448.73	1449
1482.86	1484.29	1482.29	1483.76	1484.94	1484.07	1484
1551.93	1554.97	1550.48	1553.65	1554.92	1553.24	1553
1576.19	1577.3	1575.42	1578.23	1578.91	1577.91	1577
1602.41	1603.29	1601.95	1602.99	1604.57	1602.41	1603
1615.1	1615.8	1615.28	1616.12	N/A	1615.87	1616
1656.43	1657.33	1654.48	1657.33	1658.49	1657.23	1657
1681.65	1682.84	1679.32	1683.57	1685.53	1683.64	1683
2850.28	2851.72	2850.01	2852.62	2853.13	2851.36	2852
2871.18	2872.33	2871.88	2873.04	2873.39	2872.04	2872
2893.75	2894.3	2890.41	2896.23	2895.26	2894.81	2894
2932.04	2933.35	2932.01	2933.86	2933.39	2932.95	2933
2940.44	2941.67	2941.34	2941.89	2942.2	2941.34	2941
2965.52	2968.24	2967	2966.58	2963.52	2967.19	2966
3013.76	3015.69	3009.86	3016.27	3016.76	3015.39	3015
3063.54	3064.47	3063.17	3065.35	3064.76	3064.46	3064

Table B.3: This table contains the peak positions found for H646/17L, H646/17R, H722/18L, H722/18R, H794/18L, H794/18R as well as the corresponding average.

H64617/L	**H64617/R**	**H722/18L**	**H722/18R**	**H794/18L**	**H794/18R**	**AVG**
617.272	617.999	619.719	619.339	618.505	618.549	619
639.409	640.262	642.026	641.268	640.816	640.722	641
665.212	666.111	669.773	674.86	666.347	667.579	668
N/A	695.233	N/A	698.021	N/A	N/A	697
717.303	717.695	718.522	716.839	718.767	717.703	718
743.067	743.702	745.278	745.217	744.316	744.25	744
754.154	754.812	754.587	754.814	755.574	755.671	755
782.156	781.814	782.946	782.59	783.369	783.341	783
807.855	807.239	808.862	809.162	806.254	N/A	808
821.212	821.789	824.836	825.602	820.136	820.835	822
848.254	849.164	851.799	850.859	849.286	848.755	850
N/A	N/A	880.083	878.467	N/A	N/A	879
N/A	N/A	895.936	895.505	N/A	N/A	896
933.598	934.253	935.97	935.661	933.351	935.549	935
998.792	999.438	1001.31	1001.04	999.916	999.8	1000
1028.1	1028.52	1029.85	1029.57	1029.18	1028.86	1029
1063.37	1062.99	N/A	1059.7	1063.55	1063.55	1063
1087.22	1087.71	1089.1	1087.6	1088.06	1086.8	1088
1123.11	1123.46	1125.6	1125.06	1124.06	1123.8	1124
1152.94	1154.02	1154.07	1154.06	1154.57	1154.69	1154
1168.83	1170.22	1172.17	1171.29	1169.5	1170.08	1170
1202.11	1203.06	1205.71	1205.41	1203.12	1202.88	1204
1229.26	1230.26	1237.04	1234.28	1231.75	1230.72	1232
1248.9	1248.68	1255.19	1253.77	1250.49	1249.76	1251
1307.47	1307.79	1307.75	1305.54	1308.1	1305.49	1307
1336.46	1336.82	1337.84	1337.17	1337.92	1336.44	1337
1449.01	1449	1449.77	1448.16	1449.05	1448.58	1449
1482.38	1483.25	1484.02	1484.29	1484.05	1485.43	1484
1553.3	1552.99	1550.87	1550.4	1554.56	1554.08	1553
1576.95	1577.21	1577.04	1577.41	1577.67	1578.71	1577
1602.46	1603.23	1602.94	1603.14	1603.65	1603.94	1603
1613.34	1614.3	1616.03	1616.65	1616.53	N/A	1615
1653.97	1653.77	1655.55	1654.32	1654.38	1653.83	1654
1678.71	1677.08	1681.79	1679.42	1680.2	1679.5	1679
2851.62	2852.04	2852.83	2852.32	2852.92	2852.81	2852
2872.32	2872.7	2873.73	2872.76	2873.51	2872.87	2873
2893.77	2893.91	2894.17	2895.53	2893.62	2895.71	2894
2933.23	2933.39	2934.58	2934.12	2934.59	2934.48	2934
2941.15	2941.73	2943.09	2942.52	2942.45	2942.44	2942
2967.53	2966.83	2969.26	2967.57	2969.8	2968.76	2968
3008.85	3010.14	3009.33	3010.71	3008.96	3009.52	3010
3063.44	3063.88	3064.98	3065.15	3064.54	3064.82	3064

APPENDIX C

Peak Intensity Ratio Calculation Tables

Table C.1: Peak ratio analysis (PIR) of the comparison between PNT2-C2 live and PNT2-C2 dried. The letter D specifies PIR separation between live and dried samples. The letter C specifies that the PIR marker is overlapping between samples. The letter X specifies that the PIR comparison could not take place as the band was not present in PNT2-C2 dried.

Panel	PIR Matches	% separationD
618	640(C), 669(D), 698(X), 715(C), 724(D), 745(D), 755(C), 781(C), 808(D), 825(D), 850(D), 875(D), 889(C), 934(C), 967(C), 1001(C),1030(C), 1061(C), 1090(D), 1125(D), 1155(D), 1171(C), 1205(D), 1234(D), 1256(D), 1306(C), 1336(C), 1398(X), 1448(C), 1483(D), 1552(C), 1575(C), 1603(D), 1614(C), 1656(C), 1681(D), 2851(D), 2872(D), 2895(D), 2931(D), 2940(D), 2964(D), 3014(D), 3064(D)	57%
640	618(C), 669(D), 698(X), 715(D), 724(D), 745(D), 755(C), 781(D), 808(D), 825(D), 850(D), 875(D), 889(C), 934(D), 967(C), 1001(C),1030(C), 1061(C), 1090(D), 1125(D), 1155(D), 1171(C), 1205(C), 1234(D), 1256(D), 1306(C), 1336(C), 1398(X), 1448(D), 1483(D), 1552(C), 1575(C), 1603(D), 1614(C), 1656(D), 1681(D), 2851(C), 2872(D), 2895(D), 2931(C), 2940(D), 2964(D), 3014(D), 3064(D)	62%
669	618(D), 640(D), 698(X), 715(C), 724(D), 745(D), 755(C), 781(C), 808(D), 825(D), 850(D), 875(D), 889(C), 934(C), 967(D), 1001(D),1030(C), 1061(D), 1090(D), 1125(C), 1155(C), 1171(C), 1205(D), 1234(D), 1256(D), 1306(D), 1336(D), 1398(X), 1448(D), 1483(D), 1552(D), 1575(C), 1603(D), 1614(C), 1656(D), 1681(D), 2851(D), 2872(D), 2895(D), 2931(D), 2940(D), 2964(D), 3014(C), 3064(D)	71%
715	618(C), 640(D), 669(C), 698(X), 724(D), 745(D), 755(C), 781(C), 808(D), 825(D), 850(D), 875(D), 889(C), 934(C), 967(D), 1001(C),1030(C), 1061(C), 1090(C), 1125(C), 1155(C), 1171(C), 1205(D), 1234(D), 1256(D), 1306(D), 1336(C), 1398(X), 1448(D), 1483(D), 1552(C), 1575(C), 1603(D), 1614(C), 1656(C), 1681(D), 2851(D), 2872(D), 2895(D), 2931(D), 2940(D), 2964(D), 3014(C), 3064(D)	55%
724	618(D), 640(D), 669(D), 698(X), 715(D), 745(D), 755(D), 781(D), 808(D), 825(D), 850(D), 875(D), 889(D), 934(D), 967(D), 1001(D),1030(D), 1061(D), 1090(D), 1125(D), 1155(D), 1171(D), 1205(D), 1234(D), 1256(C), 1306(D), 1336(D), 1398(X), 1448(D), 1483(D), 1552(D), 1575(D), 1603(D), 1614(D), 1656(D), 1681(D), 2851(D), 2872(D), 2895(D), 2931(D), 2940(D), 2964(D), 3014(D), 3064(D)	98%
745	618(D), 640(D), 669(D), 698(X), 715(D), 724(C), 755(D), 781(D), 808(D), 825(D), 850(D), 875(D), 889(D), 934(D), 967(D), 1001(D),1030(D), 1061(D), 1090(D), 1125(D), 1155(D), 1171(D), 1205(D), 1234(D), 1256(D), 1306(D), 1336(D), 1398(X), 1448(D), 1483(D), 1552(D), 1575(D), 1603(D), 1614(D), 1656(D), 1681(D), 2851(D), 2872(D), 2895(D), 2931(D), 2940(D), 2964(D), 3014(D), 3064(D)	98%
755	618(D), 640(D), 669(D), 698(X), 715(D), 724(D), 745(D), 781(D), 808(D), 825(D), 850(D), 875(D), 889(D), 934(D), 967(D), 1001(D),1030(D), 1061(C), 1090(D), 1125(D), 1155(D), 1171(D), 1205(D), 1234(C), 1256(D), 1306(D), 1336(D), 1398(X), 1448(D), 1483(C), 1552(D), 1575(D), 1603(C), 1614(D), 1656(D), 1681(C), 2851(D), 2872(C), 2895(D), 2931(D), 2940(D), 2964(D), 3014(D), 3064(D)	86%
781	618(C), 640(C), 669(C), 698(X), 715(C), 724(C), 745(D), 755(C), 808(D), 825(D), 850(D), 875(D), 889(D), 934(C), 967(C), 1001(C),1030(C), 1061(C), 1090(C), 1125(D), 1155(D), 1171(C), 1205(D), 1234(D), 1256(D), 1306(C), 1336(C), 1398(X), 1448(D), 1483(D), 1552(C), 1575(C), 1603(D), 1614(C), 1656(C), 1681(D), 2851(D), 2872(D), 2895(D), 2931(D), 2940(D), 2964(D), 3014(C), 3064(D)	52%
808	618(D), 640(D), 669(D), 698(X), 715(D), 724(D), 745(D), 755(D), 781 (D), 825(C), 850(D), 875(D), 889(D), 934(D), 967(C), 1001(D),1030(D), 1061(C), 1090(D), 1125(D), 1155(D), 1171(D), 1205(D), 1234(D), 1256(D), 1306(D), 1336(D), 1398(X), 1448(C), 1483(C), 1552(D), 1575(D), 1603(D), 1614(D), 1656(D), 1681(D), 2851(C), 2872(D), 2895(C), 2931(C), 2940(D), 2964(C), 3014(D), 3064(D)	79%
825	618(D), 640(D), 669(D), 698(X), 715(D), 724(D), 745(D), 755(D), 781(D), 808(C), 850(C), 875(D), 889(D), 934(D), 967(C), 1001(D),1030(D), 1061(C), 1090(D), 1125(D), 1155(D), 1171(D), 1205(D), 1234(D), 1256(D), 1306(D), 1336(D), 1398(X), 1448(C), 1483(C), 1552(D), 1575(D), 1603(D), 1614(D), 1656(C), 1681(D), 2851(C), 2872(D), 2895(C), 2931(C), 2940(D), 2964(C), 3014(D), 3064(C)	71%
850	618(D), 640(D), 669(D), 698(X), 715(D), 724(D), 745(D), 755(D), 781(D), 808(C), 825(C), 875(D), 889(D), 934(D), 967(C), 1001(D),1030(D), 1061(C), 1090(D), 1125(D), 1155(D), 1171(D), 1205(C), 1234(D), 1256(D), 1306(D), 1336(D), 1398(X), 1448(C), 1483(C), 1552(C), 1575(D), 1603(D), 1614(D), 1656(C), 1681(D), 2851(C), 2872(D), 2895(C), 2931(C), 2940(D), 2964(C), 3014(D), 3064(C)	67%
875	618(D), 640(D), 669(D), 698(X), 715(D), 724(C), 745(D), 755(D), 781(D), 808(D), 825(D), 850(D), 889(D), 934(D), 967(D), 1001(D),1030(D), 1061(D), 1090(D), 1125(D), 1155(C), 1171(D), 1205(D), 1234(D), 1256(D), 1306(C), 1336(D), 1398(X), 1448(D), 1483(D), 1552(D), 1575(D), 1603(D), 1614(C), 1656(D), 1681(C), 2851(D), 2872(D), 2895(D), 2931(D), 2940(D), 2964(D), 3014(D), 3064(D)	88%
889	618(C), 640(C), 669(D), 698(X), 715(C), 724(D), 745(D), 755(C), 781(C), 808(D), 825(D), 850(D), 875(D), 934(C), 967(C), 1001(C),1030(C), 1061(C), 1090(C), 1125(D), 1155(D), 1171(C), 1205(D), 1234(D), 1256(D), 1306(C), 1336(C), 1398(X), 1448(D), 1483(D), 1552(C), 1575(C), 1603(D), 1614(C), 1656(C), 1681(C), 2851(D), 2872(D), 2895(D), 2931(D), 2940(D), 2964(D), 3014(D), 3064(C)	52%
934	618(C), 640(C), 669(C), 698(X), 715(C), 724(D), 745(D), 755(C), 781(C), 808(D), 825(D), 850(D), 875(D), 889(C), 967(C), 1001(C),1030(C), 1061(C), 1090(C), 1125(C), 1155(C), 1171(C), 1205(C), 1234(D), 1256(D), 1306(C), 1336(C), 1398(X), 1448(D), 1483(D), 1552(C), 1575(C), 1603(D), 1614(C), 1656(C), 1681(D), 2851(D), 2872(D), 2895(D), 2931(D), 2940(D), 2964(D), 3014(C), 3064(D)	45%
967	618(C), 640(C), 669(D), 698(X), 715(D), 724(D), 745(D), 755(D), 781(D), 808(C), 825(C), 850(C), 875(D), 889(D), 934(D), 1001(C),1030(C), 1061(C), 1090(D), 1125(D), 1155(D), 1171(D), 1205(C), 1234(D), 1256(D), 1306(C), 1336(C), 1398(X), 1448(C), 1483(C), 1552(C), 1575(C), 1603(D), 1614(D), 1656(C), 1681(D), 2851(C), 2872(C), 2895(C), 2931(C), 2940(D), 2964(C), 3014(D), 3064(C)	48%
1001	618(C), 640(C), 669(D), 698(X), 715(C), 724(D), 745(D), 755(C), 781(C), 808(D), 825(D), 850(D), 875(D), 889(C), 934(C), 967(C), 1030(C), 1061(C), 1090(C), 1125(C), 1155(D), 1171(D), 1205(C), 1234(D), 1256(D), 1306(C), 1336(C), 1398(X), 1448(D), 1483(D), 1552(C), 1575(C), 1603(D), 1614(C), 1656(C), 1681(D), 2851(D), 2872(D), 2895(D), 2931(D), 2940(D), 2964(D), 3014(D), 3064(D)	55%
1030	618(C), 640(C), 669(C), 698(X), 715(C), 724(D), 745(D), 755(C), 781(C), 808(D), 825(D), 850(D), 875(C), 889(C), 934(C), 967(D), 1001(C), 1061(C), 1090(C), 1125(C), 1155(D), 1171(C), 1205(D), 1234(D), 1256(D), 1306(C), 1336(C), 1398(X), 1448(C), 1483(D), 1552(C), 1575(C), 1603(D), 1614(C), 1656(C), 1681(D), 2851(D), 2872(D), 2895(D), 2931(D), 2940(D), 2964(D), 3014(C), 3064(D)	48%
1061	618(C), 640(C), 669(D), 698(X), 715(C), 724(D), 745(D), 755(C), 781(C), 808(C), 825(C), 850(C), 875(C), 889(D), 934(C), 967(C), 1001(C),1030(D), 1090(C), 1125(D), 1155(D), 1171(C), 1205(C), 1234(D), 1256(D), 1306(C), 1336(C), 1398(X), 1448(C), 1483(C), 1552(C), 1575(C), 1603(D), 1614(D), 1656(C), 1681(D), 2851(C), 2872(D), 2895(C), 2931(C), 2940(D), 2964(C), 3014(C), 3064(C)	33%
1090	618(C), 640(C), 669(C), 698(X), 715(C), 724(D), 745(D), 755(C), 781(C), 808(D), 825(D), 850(D), 875(D), 889(C), 934(C), 967(D), 1001(C),1030(C), 1061(C), 1125(D), 1155(D), 1171(D), 1205(C), 1234(D), 1256(D), 1306(C), 1336(C), 1398(X), 1448(D), 1483(D), 1552(C), 1575(C), 1603(D), 1614(C), 1656(C), 1681(D), 2851(D), 2872(D), 2895(D), 2931(D), 2940(D), 2964(D), 3014(C), 3064(D)	55%
1125	618(D), 640(D), 669(C), 698(X), 715(C), 724(D), 745(D), 755(D), 781(D), 808(D), 825(D), 850(D), 875(D), 889(D), 934(C), 967(D), 1001(C),1030(C), 1061(D), 1090(D), 1155(C), 1171(D), 1205(D), 1234(D), 1256(D), 1306(D), 1336(D), 1398(X), 1448(D), 1483(D), 1552(D), 1575(D), 1603(D), 1614(C), 1656(D), 1681(D), 2851(D), 2872(D), 2895(D), 2931(D), 2940(D), 2964(D), 3014(C), 3064(D)	83%
1155	618(D), 640(D), 669(C), 698(X), 715(D), 724(C), 745(D), 755(D), 781(D), 808(D), 825(D), 850(D), 875(D), 889(D), 934(C), 967(D), 1001(D),1030(D), 1061(D), 1090(D), 1125(C), 1171(D), 1205(D), 1234(D), 1256(D), 1306(D), 1336(D), 1398(X), 1448(D), 1483(D), 1552(D), 1575(D), 1603(D), 1614(C), 1656(D), 1681(D), 2851(D), 2872(D), 2895(D), 2931(D), 2940(D), 2964(D), 3014(C), 3064(D)	86%
1171	618(C), 640(D), 669(C), 698(X), 715(C), 724(D), 745(D), 755(C), 781(C), 808(D), 825(D), 850(D), 875(D), 889(C), 934(C), 967(C), 1001(C),1030(C), 1061(C), 1090(C), 1125(D), 1155(D), 1205(D), 1234(D), 1256(D), 1306(C), 1336(C), 1398(X), 1448(C), 1483(D), 1552(C), 1575(C), 1603(D), 1614(C), 1656(C), 1681(D), 2851(D), 2872(D), 2895(D), 2931(D), 2940(D), 2964(D), 3014(C), 3064(D)	52%
1205	618(C), 640(C), 669(D), 698(X), 715(D), 724(D), 745(D), 755(C), 781(D), 808(D), 825(C), 850(C), 875(D), 889(C), 934(C), 967(C), 1001(D),1030(C), 1061(C), 1090(D), 1125(D), 1155(D), 1171(D), 1234(D), 1256(D), 1306(C), 1336(C), 1398(X), 1448(D), 1483(C), 1552(C), 1575(D), 1603(D), 1614(D), 1656(C), 1681(D), 2851(C), 2872(D), 2895(C), 2931(C), 2940(D), 2964(D), 3014(D), 3064(C)	55%
1234	618(D), 640(D), 669(D), 698(X), 715(D), 724(D), 745(C), 755(C), 781(C), 808(D), 825(D), 850(D), 875(D), 889(D), 934(D), 967(D), 1001(D),1030(D), 1061(D), 1090(D), 1125(D), 1155(D), 1171(D), 1205(D), 1256(D), 1306(D), 1336(D), 1398(X), 1448(D), 1483(D), 1552(D), 1575(D), 1603(C), 1614(D), 1656(C), 1681(C), 2851(D), 2872(C), 2895(D), 2931(D), 2940(D), 2964DC), 3014(D), 3064(D)	83%
1256	618(D), 640(D), 669(D), 698(X), 715(D), 724(C), 745(D), 755(D), 781(D), 808(D), 825(D), 850(D), 875(C), 889(D), 934(D), 967(D), 1001(D),1030(D), 1061(D), 1090(D), 1125(D), 1155(D), 1171(D), 1205(D), 1234(D), 1306(D), 1336(D), 1398(X), 1448(D), 1483(D), 1552(D), 1575(D), 1603(D), 1614(D), 1656(D), 1681(D), 2851(D), 2872(D), 2895(D), 2931(D), 2940(D), 2964(D), 3014(D), 3064(D)	95%
1306	618(C), 640(C), 669(D), 698(X), 715(C), 724(D), 745(D), 755(C), 781(C), 808(D), 825(D), 850(C), 875(D), 889(C), 934(C), 967(D), 1001(C),1030(C), 1061(C), 1090(C), 1125(D), 1155(D), 1171(C), 1205(C), 1234(D), 1256(D), 1336(C), 1398(X), 1448(D), 1483(D), 1552(C), 1575(C), 1603(D), 1614(C), 1656(C), 1681(D), 2851(C), 2872(D), 2895(D), 2931(C), 2940(D), 2964(D), 3014(C), 3064(C)	48%
1336	618(C), 640(C), 669(C), 698(X), 715(C), 724(D), 745(D), 755(C), 781(C), 808(D), 825(D), 850(D), 875(D), 889(C), 934(C), 967(C), 1001(C),1030(C), 1061(C), 1090(C), 1125(D), 1155(D), 1171(C), 1205(C), 1234(D), 1256(D), 1306(C), 1398(X), 1448(D), 1483(D), 1552(C), 1575(C), 1603(D), 1614(C), 1656(C), 1681(D), 2851(D), 2872(D), 2895(D), 2931(D), 2940(D), 2964(D), 3014(C), 3064(D)	50%
1448	618(D), 640(D), 669(D), 698(X), 715(D), 724(D), 745(D), 755(D), 781(D), 808(C), 825(C), 850(C), 875(D), 889(D), 934(D), 967(C), 1001(D),1030(D), 1061(C), 1090(D), 1125(D), 1155(D), 1171(D), 1205(D), 1234(D), 1256(D), 1306(D), 1336(D), 1398(X), 1483(C), 1552(D), 1575(D), 1603(D), 1614(D), 1656(D), 1681(D), 2851(C), 2872(D), 2895(C), 2931(C), 2940(D), 2964(C), 3014(D), 3064(C)	74%
1483	618(D), 640(D), 669(D), 698(X), 715(D), 724(D), 745(D), 755(D), 781(D), 808(C), 825(C), 850(C), 875(D), 889(D), 934(D), 967(C), 1001(D),1030(D), 1061(C), 1090(D), 1125(D), 1155(D), 1171(D), 1205(C), 1234(D), 1256(D), 1306(D), 1336(D), 1398(X), 1448(C), 1552(D), 1575(D), 1603(D), 1614(D), 1656(C), 1681(D), 2851(C), 2872(C), 2895(C), 2931(C), 2940(D), 2964(C), 3014(D), 3064(C)	67%
1552	618(C), 640(C), 669(D), 698(X), 715(C), 724(D), 745(D), 755(C), 781(C), 808(D), 825(D), 850(D), 875(D), 889(D), 934(C), 967(C), 1001(C),1030(C), 1061(C), 1090(C), 1125(D), 1155(D), 1171(C), 1205(C), 1234(D), 1256(D), 1306(C), 1336(C), 1398(X), 1448(D), 1483(D), 1575(C), 1603(D), 1614(C), 1656(C), 1681(D), 2851(C), 2872(D), 2895(D), 2931(C), 2940(D), 2964(D), 3014(D), 3064(D)	50%
1575	618(C), 640(C), 669(C), 698(X), 715(C), 724(D), 745(D), 755(C), 781(C), 808(D), 825(D), 850(D), 875(D), 889(C), 934(C), 967(C), 1001(C),1030(C), 1061(C), 1090(C), 1125(D), 1155(D), 1171(C), 1205(D), 1234(D), 1256(D), 1306(C), 1336(C), 1398(X), 1448(D), 1483(D), 1552(C), 1603(D), 1614(C), 1656(C), 1681(D), 2851(D), 2872(D), 2895(D), 2931(D), 2940(D), 2964D(D), 3014(D), 3064(C)	52%
1603	618(D), 640(D), 669(D), 698(X), 715(D), 724(D), 745(D), 755(D), 781(D), 808(D), 825(D), 850(D), 875(D), 889(D), 934(D), 967(C), 1001(D),1030(D), 1061(D), 1090(D), 1125(D), 1155(D), 1171(D), 1205(D), 1234(C), 1256(D), 1306(D), 1336(D), 1398(X), 1448(D), 1483(D), 1552(D), 1575(D), 1614(D), 1656(D), 1681(D), 2851(D), 2872(D), 2895(D), 2931(D), 2940(C), 2964(D), 3014(D), 3064(D)	88%
1614	618(C), 640(C), 669(C), 698(X), 715(C), 724(C), 745(D), 755(C), 781(C), 808(D), 825(D), 850(D), 875(D), 889(C), 934(C), 967(D), 1001(D),1030(C), 1061(C), 1090(C), 1125(C), 1155(C), 1171(C), 1205(D), 1234(D), 1256(D), 1306(C), 1336(C), 1398(X), 1448(D), 1483(D), 1552(C), 1575(C), 1603(D), 1656(C), 1681(D), 2851(D), 2872(D), 2895(D), 2931(D), 2940(D), 2964(D), 3014(C), 3064(D)	50%
1656	618(C), 640(C), 669(D), 698(X), 715(C), 724(D), 745(D), 755(C), 781(C), 808(D), 825(C), 850(C), 875(D), 889(C), 934(C), 967(C), 1001(C),1030(C), 1061(C), 1090(C), 1125(D), 1155(D), 1171(C), 1205(C), 1234(D), 1256(D), 1306(C), 1336(C), 1398(X), 1448(D), 1483(D), 1552(C), 1575(C), 1603(D), 1614(C), 1681(D), 2851(C), 2872(D), 2895(C), 2931(C), 2940(D), 2964(D), 3014(C), 3064(C)	38%
1681	618(D), 640(D), 669(D), 698(X), 715(D), 724(D), 745(D), 755(D), 781(D), 808(D), 825(D), 850(D), 875(D), 889(D), 934(D), 967(D), 1001(D),1030(D), 1061(D), 1090(D), 1125(D), 1155(D), 1171(D), 1205(D), 1234(C), 1256(D), 1306(D), 1336(D), 1398(X), 1448(D), 1483(D), 1552(D), 1575(D), 1603(C), 1614(D), 1656(D), 2851(D), 2872(C), 2895(D), 2931(D), 2940(C), 2964(D), 3014(D), 3064(D)	90%
2851	618(D), 640(C), 669(DD), 698(X), 715(D), 724(D), 745(D), 755(D), 781(D), 808(C), 825(C), 850(C), 875(D), 889(D), 934(D), 967(C), 1001(D),1030(D), 1061(C), 1090(D), 1125(D), 1155(D), 1171(D), 1205(C), 1234(D), 1256(D), 1306(C), 1336(D), 1398(X), 1448(C), 1483(C), 1552(C), 1575(D), 1603(D), 1614(D), 1656(C), 1681(D), 2872(D), 2895(C), 2931(C), 2940(D), 2964(C), 3014(D), 3064(C)	60%
2872	618(D), 640(D), 669(D), 698(X), 715(D), 724(D), 745(D), 755(D), 781(D), 808(C), 825(D), 850(D), 875(D), 889(D), 934(D), 967(D), 1001(C),1030(D), 1061(D), 1090(D), 1125(D), 1155(D), 1171(D), 1205(D), 1234(D), 1256(D), 1306(D), 1336(D), 1398(X), 1448(C), 1483(D), 1552(C), 1575(D), 1603(D), 1614(D), 1656(C), 1681(C), 2851(D), 2895(D), 2931(D), 2940(C), 2964(D), 3014(D), 3064(D)	83%
2895	618(D), 640(D), 669(D), 698(X), 715(D), 724(D), 745(D), 755(D), 781(D), 808(C), 825(C), 850(C), 875(D), 889(D), 934(D), 967(C), 1001(D),1030(D), 1061(C), 1090(D), 1125(D), 1155(D), 1171(D), 1205(D), 1234(D), 1256(D), 1306(D), 1336(D), 1398(X), 1448(C), 1483(C), 1552(D), 1575(D), 1603(D), 1614(D), 1656(C), 1681(D), 2851(D), 2872(D), 2931(C), 2940(D), 2964(C), 3014(D), 3064(C)	74%
2931	618(D), 640(C), 669(D), 698(X), 715(D), 724(D), 745(D), 755(C), 781(D), 808(C), 825(C), 850(C), 875(D), 889(D), 934(D), 967(C), 1001(C),1030(D), 1061(C), 1090(D), 1125(D), 1155(D), 1171(D), 1205(C), 1234(D), 1256(D), 1306(C), 1336(C), 1398(X), 1448(C), 1483(C), 1552(C), 1575(C), 1603(D), 1614(D), 1656(C), 1681(D), 2851(C), 2872(D), 2895(C), 2940(D), 2964(C), 3014(D), 3064(C)	57%
2940	618(D), 640(D), 669(D), 698(X), 715(D), 724(D), 745(D), 755(D), 781(D), 808(D), 825(D), 850(D), 875(D), 889(D), 934(D), 967(D), 1001(D),1030(D), 1061(D), 1090(D), 1125(D), 1155(D), 1171(D), 1205(D), 1234(D), 1256(D), 1306(D), 1336(D), 1398(X), 1448(D), 1483(D), 1552(D), 1575(D), 1603(C), 1614(D), 1656(D), 1681(C), 2851(D), 2872(C), 2895(D), 2931(D), 2964(D), 3014(D), 3064(D)	90%
2964	618(D), 640(D), 669(D), 698(X), 715(D), 724(D), 745(D), 755(D), 781(D), 808(C), 825(C), 850(C), 875(D), 889(D), 934(D), 967(C), 1001(D), 1030(D), 1061(C), 1090(D), 1125(D), 1155(D), 1171(D), 1205(D), 1234(D), 1256(D), 1306(D), 1336(D), 1398(X), 1448(C), 1483(C), 1552(D), 1575(D), 1603(D), 1614(D), 1656(D), 1681(D), 2851(C), 2872(D), 2895(D), 2931(D), 2940(D), 3014(D), 3064(C)	79%
3014	618(D), 640(D), 669(C), 698(X), 715(C), 724(D), 745(D), 755(D), 781(D), 808(D), 825(D), 850(D), 875(D), 889(C), 934(C), 967(D), 1001(D),1030(C), 1061(C), 1090(C), 1125(C), 1155(C), 1171(C), 1205(D), 1234(D), 1256(D), 1306(D), 1336(D), 1398(X), 1448(D), 1483(D), 1552(C), 1575(D), 1603(D), 1614(C), 1656(D), 1681(D), 2851(D), 2872(D), 2895(D), 2931(D), 2940(D), 2964(D), 3064(D)	67%
3064	618(D), 640(D), 669(D), 698(X), 715(D), 724(D), 745(D), 755(D), 781(D), 808(C), 825(C), 850(C), 875(C), 889(D), 934(C), 967(C), 1001(D),1030(D), 1061(C), 1090(D), 1125(D), 1155(D), 1171(D), 1205(C), 1234(D), 1256(D), 1306(D), 1336(D), 1398(X), 1448(C), 1483(C), 1552(D), 1575(D), 1603(D), 1614(D), 1656(C), 1681(D), 2851(C), 2872(D), 2895(C), 2931(C), 2940(D), 2964(C), 3014(C)	62%

Table C.2: Peak ratio analysis (PIR) of the comparison between PNT2-C2 live and BPH-1. The letter D specifies PIR separation between PNT2-C2 and BPH-1. The letter C specifies that the PIR marker is overlapping between samples. The letter X specifies that the PIR comparison could not take place as the band was not present in BPH-1.

Panel	PIR Matches	% matchD
619	641(C), 673(D), 690(D), 715(C), 723(D), 745(D), 755(C), 783(D), 810(D), 825(C), 851(D), 876(D), 893(D), 938(C), 976(D), 1001(C), 1030(C), 1056(D), 1097(C), 1125(D), 1156(D), 1171(D), 1206(D), 1234(D), 1254(D), 1307(C), 1337(D), 1396(X), 1448(C), 1484(D), 1554(D), 1578(D), 1603(D), 1616(C), 1657(C), 1684(D), 2852(C), 2873(C), 2898(D), 2934(D), 2942(C), 2967(D), 3016(C), 3065(D)	63%
641	619(C), 673(D), 690(D), 715(C), 723(D), 745(D), 755(C), 783(D), 810(D), 825(C), 851(D), 876(D), 893(D), 938(C), 976(D), 1001(C), 1030(C), 1056(D), 1097(C), 1125(D), 1156(D), 1171(D), 1206(D), 1234(D), 1254(D), 1307(C), 1337(D), 1396(X), 1448(C), 1484(D), 1554(D), 1578(D), 1603(D), 1616(C), 1657(C), 1684(D), 2852(C), 2873(C), 2898(D), 2934(D), 2942(C), 2967(D), 3016(C), 3065(D)	58%
673	619(D), 641(D), 690(D), 715(D), 723(D), 745(D), 755(D), 783(D), 810(D), 825(D), 851(D), 876(D), 893(D), 938(C), 976(D), 1001(D), 1030(D), 1056(C), 1097(D), 1125(D), 1156(D), 1171(D), 1206(D), 1234(D), 1254(D), 1307(D), 1337(D), 1396(X), 1448(D), 1484(C), 1554(D), 1578(C), 1603(D), 1616(D), 1657(D), 1684(D), 2852(D), 2873(D), 2898(D), 2934(D), 2942(D), 2967(D), 3016(D), 3065(D)	91%
689	619(D), 641(D), 673(D), 715(D), 723(D), 745(D), 755(D), 783(D), 810(D), 825(C), 851(C), 876(D), 893(D), 938(D), 976(D), 1001(D), 1030(D), 1056(C), 1097(D), 1125(D), 1156(D), 1171(D), 1206(D), 1234(C), 1254(D), 1307(D), 1337(D), 1396(X), 1448(D), 1484(C), 1554(C), 1578(D), 1603(D), 1616(D), 1657(D), 1684(D), 2852(D), 2873(D), 2898(D), 2934(D), 2942(D), 2967(D), 3016(D), 3065(D)	91%
715	619(D), 641(D), 673(D), 689(C), 723(D), 745(D), 755(D), 783(D), 810(C), 825(D), 851(C), 876(D), 893(D), 938(D), 976(D), 1001(D), 1030(D), 1056(D), 1097(D), 1125(D), 1156(D), 1171(D), 1206(D), 1234(C), 1254(D), 1307(D), 1337(D), 1396(X), 1448(D), 1484(C), 1554(C), 1578(D), 1603(C), 1616(D), 1657(D), 1684(C), 2852(D), 2873(D), 2898(D), 2934(D), 2942(D), 2967(D), 3016(D), 3065(D)	91%
723	619(D), 641(D), 673(D), 689(D), 715(D), 745(C), 755(D), 783(D), 810(D), 825(D), 851(D), 876(D), 893(D), 938(D), 976(D), 1001(D), 1030(D), 1056(D), 1097(D), 1125(C), 1156(D), 1171(D), 1206(C), 1234(D), 1254(C), 1307(D), 1337(D), 1396(X), 1448(D), 1484(D), 1554(D), 1578(C), 1603(C), 1616(D), 1657(D), 1684(C), 2852(D), 2873(D), 2898(D), 2934(D), 2942(D), 2967(D), 3016(D), 3065(D)	77%
745	619(D), 641(D), 673(D), 689(D), 715(D), 723(D), 755(D), 783(D), 810(D), 825(D), 851(D), 876(D), 893(D), 938(C), 976(D), 1001(D), 1030(D), 1056(D), 1097(D), 1125(C), 1156(D), 1171(D), 1206(C), 1234(D), 1254(C), 1307(C), 1337(D), 1396(X), 1448(D), 1484(D), 1554(D), 1578(C), 1603(D), 1616(D), 1657(D), 1684(D), 2852(D), 2873(D), 2898(D), 2934(D), 2942(C), 2967(D), 3016(C), 3065(C)	63%
755	619(D), 641(D), 673(D), 689(D), 715(D), 723(D), 745(D), 783(D), 810(D), 825(D), 851(D), 876(D), 893(D), 938(D), 976(D), 1001(D), 1030(D), 1056(D), 1097(D), 1125(D), 1156(D), 1171(D), 1206(D), 1234(D), 1254(D), 1307(D), 1337(D), 1396(X), 1448(D), 1484(D), 1554(D), 1578(D), 1603(D), 1616(D), 1657(D), 1684(D), 2852(D), 2873(D), 2898(D), 2934(D), 2942(D), 2967(D), 3016(D), 3065(C)	68%
783	619(D), 641(D), 673(D), 689(D), 715(D), 723(D), 745(D), 755(D), 810(D), 825(D), 851(D), 876(D), 893(D), 938(D), 976(D), 1001(D), 1030(D), 1056(D), 1097(D), 1125(D), 1156(D), 1171(D), 1206(D), 1234(D), 1254(D), 1307(D), 1337(D), 1396(X), 1448(D), 1484(D), 1554(D), 1578(D), 1603(D), 1616(D), 1657(D), 1684(D), 2852(D), 2873(D), 2898(D), 2934(D), 2942(D), 2967(D), 3016(D), 3065(C)	74%
810	619(D), 641(D), 673(D), 689(D), 715(D), 723(D), 745(D), 755(D), 783(D), 825(D), 851(D), 876(D), 893(D), 938(D), 976(D), 1001(D), 1030(D), 1056(D), 1097(D), 1125(D), 1156(D), 1171(D), 1206(D), 1234(C), 1254(D), 1307(D), 1337(D), 1396(X), 1448(D), 1484(C), 1554(C), 1578(D), 1603(D), 1616(D), 1657(D), 1684(D), 2852(D), 2873(D), 2898(D), 2934(D), 2942(D), 2967(D), 3016(D), 3065(D)	91%
825	619(D), 641(C), 673(D), 689(D), 715(C), 723(D), 745(C), 754(C), 783(D), 810(D), 851(C), 876(D), 893(D), 938(C), 976(D), 1001(D), 1030(C), 1056(D), 1097(C), 1125(D), 1156(D), 1171(D), 1206(D), 1234(D), 1254(D), 1307(D), 1337(D), 1396(X), 1448(D), 1484(D), 1554(D), 1578(D), 1603(D), 1616(C), 1657(C), 1684(C), 2852(D), 2873(C), 2898(D), 2934(D), 2942(C), 2967(D), 3016(C), 3065(D)	70%
851	619(D), 641(D), 673(D), 689(C), 715(D), 723(D), 745(D), 755(D), 783(D), 810(D), 825(D), 876(D), 893(D), 938(C), 976(D), 1001(D), 1030(D), 1056(D), 1097(D), 1125(D), 1156(D), 1171(D), 1206(D), 1234(D), 1254(D), 1307(D), 1337(D), 1396(X), 1448(D), 1484(D), 1554(D), 1578(D), 1603(D), 1616(C), 1657(D), 1684(D), 2852(C), 2873(D), 2898(D), 2934(D), 2942(D), 2967(D), 3016(C), 3065(D)	86%
876	619(D), 641(D), 673(D), 689(D), 715(D), 723(D), 745(D), 755(D), 783(D), 810(D), 825(D), 851(D), 893(D), 938(D), 976(D), 1001(D), 1030(D), 1056(D), 1097(D), 1125(D), 1156(D), 1171(D), 1206(D), 1234(D), 1254(D), 1307(D), 1337(D), 1396(X), 1448(D), 1484(D), 1554(C), 1578(D), 1603(D), 1616(D), 1657(D), 1684(D), 2852(D), 2873(D), 2898(D), 2934(D), 2942(D), 2967(D), 3016(D), 3065(D)	95%
893	619(D), 641(D), 673(D), 689(D), 715(D), 723(D), 745(D), 755(D), 783(D), 810(D), 825(D), 851(D), 876(D), 938(D), 976(D), 1001(D), 1030(D), 1056(D), 1097(D), 1125(D), 1156(D), 1171(D), 1206(D), 1234(D), 1254(D), 1307(D), 1337(D), 1396(X), 1448(D), 1484(C), 1554(D), 1578(D), 1603(D), 1616(D), 1657(D), 1684(D), 2852(D), 2873(D), 2898(D), 2934(D), 2942(D), 2967(D), 3016(D), 3065(D)	95%
938	619(C), 641(C), 673(D), 689(D), 715(C), 723(D), 745(D), 755(C), 783(D), 810(D), 825(C), 851(D), 876(D), 893(D), 976(D), 1001(D), 1030(D), 1056(D), 1097(C), 1125(D), 1156(D), 1171(D), 1206(D), 1234(D), 1254(D), 1307(D), 1337(D), 1396(X), 1448(D), 1484(D), 1554(D), 1578(D), 1603(D), 1616(D), 1657(D), 1684(D), 2852(D), 2873(D), 2898(D), 2934(D), 2942(D), 2967(D), 3016(D), 3065(D)	70%
976	619(D), 641(D), 673(D), 689(D), 715(D), 723(C), 745(C), 755(D), 783(D), 810(D), 825(D), 851(D), 876(D), 893(D), 938(D), 1001(D), 1030(D), 1056(D), 1097(C), 1125(D), 1156(D), 1171(C), 1206(C), 1234(D), 1254(D), 1307(C), 1337(D), 1396(X), 1448(D), 1484(C), 1554(D), 1578(C), 1603(D), 1616(D), 1657(C), 1684(C), 2852(D), 2873(D), 2898(C), 2934(C), 2942(D), 2967(C), 3016(C), 3065(C)	49%
1001	619(C), 641(C), 673(D), 689(D), 715(C), 723(D), 745(D), 755(C), 783(D), 810(D), 825(C), 851(D), 876(D), 893(D), 938(D), 976(D), 1030(C), 1056(D), 1097(C), 1125(D), 1156(D), 1171(C), 1206(D), 1234(D), 1254(D), 1307(C), 1337(D), 1396(X), 1448(C), 1484(D), 1554(D), 1578(D), 1603(D), 1616(C), 1657(C), 1684(D), 2852(C), 2873(C), 2898(D), 2934(D), 2942(C), 2967(D), 3016(C), 3065(D)	56%
1030	619(C), 641(C), 673(D), 689(D), 715(C), 723(D), 745(D), 755(C), 783(D), 810(D), 825(D), 851(D), 876(D), 893(D), 938(D), 976(D), 1001(C), 1056(D), 1097(C), 1125(D), 1156(D), 1171(D), 1206(D), 1234(D), 1254(D), 1307(D), 1337(D), 1396(X), 1448(D), 1484(D), 1554(D), 1578(D), 1603(D), 1616(C), 1657(D), 1684(C), 2852(D), 2873(C), 2898(D), 2934(D), 2942(D), 2967(D), 3016(D), 3065(D)	70%
1056	619(D), 641(D), 673(D), 689(D), 715(D), 723(D), 745(D), 755(D), 783(D), 810(D), 825(D), 851(D), 876(D), 893(D), 938(D), 976(D), 1001(D), 1030(D), 1097(D), 1125(D), 1156(D), 1171(D), 1206(D), 1234(D), 1254(D), 1307(D), 1337(D), 1396(X), 1448(D), 1484(D), 1554(D), 1578(D), 1603(D), 1616(D), 1657(C), 1684(D), 2852(D), 2873(D), 2898(D), 2934(D), 2942(D), 2967(D), 3016(D), 3065(D)	91%
1097	619(C), 641(C), 673(D), 689(D), 715(C), 723(D), 745(D), 755(C), 783(D), 810(D), 825(D), 851(D), 876(D), 893(D), 938(C), 976(C), 1001(C), 1030(C), 1056(D), 1125(D), 1156(D), 1171(C), 1206(D), 1234(D), 1254(D), 1307(D), 1337(D), 1396(X), 1448(C), 1484(D), 1554(D), 1578(C), 1603(D), 1616(C), 1657(C), 1684(C), 2852(C), 2873(C), 2898(D), 2934(D), 2942(C), 2967(C), 3016(C), 3065(C)	53%
1125	619(D), 641(D), 673(D), 689(D), 715(D), 723(D), 745(D), 755(D), 783(D), 810(D), 825(D), 851(D), 876(D), 893(D), 938(D), 976(D), 1001(D), 1030(D), 1056(D), 1097(D), 1156(D), 1171(D), 1206(D), 1234(D), 1254(D), 1307(D), 1337(D), 1396(X), 1448(D), 1484(D), 1554(D), 1578(D), 1603(D), 1616(D), 1657(D), 1684(D), 2852(D), 2873(D), 2898(D), 2934(D), 2942(D), 2967(D), 3016(D), 3065(D)	77%
1156	619(D), 641(D), 673(D), 689(D), 715(D), 723(D), 745(D), 755(D), 783(D), 810(D), 825(D), 851(D), 876(D), 893(D), 938(D), 976(D), 1001(D), 1030(D), 1056(D), 1097(D), 1125(D), 1171(D), 1206(D), 1234(D), 1254(D), 1307(D), 1337(D), 1396(X), 1448(D), 1484(D), 1554(D), 1578(D), 1603(D), 1616(D), 1657(D), 1684(D), 2852(D), 2873(D), 2898(D), 2934(D), 2942(D), 2967(D), 3016(D), 3065(D)	100%
1171	619(D), 641(D), 673(D), 689(D), 715(D), 723(D), 745(D), 755(D), 783(D), 810(D), 825(D), 851(D), 876(D), 893(D), 938(D), 976(D), 1001(D), 1030(D), 1056(D), 1097(D), 1125(D), 1156(D), 1206(D), 1234(D), 1254(D), 1307(D), 1337(D), 1396(X), 1448(D), 1484(D), 1554(D), 1578(D), 1603(D), 1616(D), 1657(D), 1684(D), 2852(D), 2873(D), 2898(D), 2934(D), 2942(D), 2967(D), 3016(D), 3065(D)	100%
1206	619(D), 641(D), 673(D), 689(D), 715(D), 723(D), 745(C), 755(D), 783(D), 810(D), 825(D), 851(D), 876(D), 893(D), 938(D), 976(C), 1001(D), 1030(D), 1056(D), 1097(C), 1125(D), 1156(D), 1171(D), 1234(D), 1254(D), 1307(C), 1337(D), 1396(X), 1448(D), 1484(D), 1554(D), 1578(C), 1603(D), 1616(D), 1657(D), 1684(D), 2852(D), 2873(D), 2898(D), 2934(C), 2942(D), 2967(C), 3016(C), 3065(D)	81%
1234	619(D), 641(D), 673(D), 689(D), 715(D), 723(D), 745(D), 755(D), 783(D), 810(C), 825(D), 851(D), 876(D), 893(D), 938(D), 976(D), 1001(D), 1030(D), 1056(D), 1097(D), 1125(D), 1156(D), 1171(D), 1206(D), 1254(C), 1307(D), 1337(D), 1396(X), 1448(D), 1484(D), 1554(C), 1578(D), 1603(D), 1616(D), 1657(D), 1684(D), 2852(D), 2873(D), 2898(D), 2934(D), 2942(D), 2967(D), 3016(D), 3065(D)	95%
1254	619(D), 641(D), 673(D), 689(D), 715(D), 723(D), 745(C), 755(D), 783(D), 810(D), 825(D), 851(D), 876(D), 893(D), 938(D), 976(D), 1001(D), 1030(D), 1056(D), 1097(D), 1125(C), 1156(D), 1171(D), 1206(D), 1234(D), 1307(D), 1337(C), 1396(X), 1448(D), 1484(D), 1554(D), 1578(D), 1603(D), 1616(D), 1657(D), 1684(D), 2852(D), 2873(D), 2898(D), 2934(C), 2942(D), 2967(D), 3016(D), 3065(D)	85%
1307	619(C), 641(C), 673(D), 689(D), 715(C), 723(D), 745(C), 755(C), 783(D), 810(D), 825(D), 851(D), 876(D), 893(D), 938(D), 976(C), 1001(C), 1030(C), 1056(D), 1097(C), 1125(C), 1156(D), 1171(D), 1206(D), 1234(D), 1254(C), 1337(D), 1396(X), 1448(C), 1484(D), 1554(D), 1578(D), 1603(D), 1616(C), 1657(C), 1684(C), 2852(C), 2873(C), 2898(D), 2934(D), 2942(C), 2967(C), 3016(C), 3065(C)	40%
1337	619(D), 641(D), 673(D), 689(D), 715(D), 723(C), 745(C), 755(D), 783(D), 810(D), 825(D), 851(D), 876(D), 893(D), 938(D), 976(C), 1001(D), 1030(D), 1056(D), 1097(D), 1125(C), 1156(D), 1171(D), 1206(C), 1234(D), 1254(C), 1307(D), 1396(X), 1448(D), 1484(D), 1554(D), 1578(D), 1603(C), 1616(D), 1657(D), 1684(D), 2852(D), 2873(D), 2898(D), 2934(D), 2942(D), 2967(D), 3016(D), 3065(D)	79%
1448	619(C), 641(C), 673(D), 689(D), 715(C), 723(D), 745(D), 755(C), 783(D), 810(D), 825(D), 851(D), 876(D), 893(D), 938(C), 976(C), 1001(C), 1030(C), 1056(D), 1097(C), 1125(D), 1156(D), 1171(C), 1206(D), 1234(D), 1254(D), 1307(C), 1337(D), 1396(X), 1484(D), 1554(D), 1578(D), 1603(D), 1616(C), 1657(C), 1684(C), 2852(C), 2873(C), 2898(C), 2934(D), 2942(C), 2967(C), 3016(C), 3065(C)	53%
1484	619(D), 641(D), 673(D), 689(D), 715(D), 723(C), 745(C), 755(D), 783(D), 810(D), 825(D), 851(D), 876(D), 893(D), 938(D), 976(C), 1001(D), 1030(D), 1056(D), 1097(D), 1125(D), 1156(D), 1171(D), 1206(D), 1234(D), 1254(D), 1307(D), 1337(D), 1396(X), 1448(D), 1554(D), 1578(D), 1603(D), 1616(D), 1657(D), 1684(D), 2852(D), 2873(D), 2898(D), 2934(D), 2942(D), 2967(D), 3016(D), 3065(D)	86%
1554	619(D), 641(D), 673(D), 689(D), 715(D), 723(D), 745(D), 755(C), 783(D), 810(D), 825(D), 851(D), 876(D), 893(D), 938(D), 976(D), 1001(D), 1030(D), 1056(D), 1097(D), 1125(D), 1156(D), 1171(D), 1206(D), 1234(C), 1254(D), 1307(D), 1337(D), 1396(X), 1448(D), 1484(C), 1578(D), 1603(D), 1616(D), 1657(D), 1684(D), 2852(D), 2873(D), 2898(D), 2934(C), 2942(D), 2967(D), 3016(D), 3065(D)	93%
1578	619(D), 641(D), 673(D), 689(D), 715(D), 723(C), 745(C), 755(D), 783(D), 810(D), 825(D), 851(D), 876(D), 893(D), 938(D), 976(C), 1001(D), 1030(D), 1056(D), 1097(C), 1125(D), 1156(D), 1171(D), 1206(C), 1234(D), 1254(C), 1307(D), 1337(C), 1396(X), 1448(D), 1484(D), 1554(C), 1603(D), 1616(D), 1657(D), 1684(C), 2852(D), 2873(C), 2898(D), 2934(C), 2942(D), 2967(C), 3016(C), 3065(C)	69%
1603	619(D), 641(D), 673(D), 689(D), 715(D), 723(D), 745(D), 755(D), 783(D), 810(D), 825(D), 851(D), 876(D), 893(D), 938(D), 976(D), 1001(D), 1030(D), 1056(D), 1097(D), 1125(D), 1156(D), 1171(D), 1206(D), 1234(D), 1254(D), 1307(D), 1337(D), 1396(X), 1448(D), 1484(D), 1554(D), 1578(C), 1616(D), 1657(D), 1684(D), 2852(D), 2873(D), 2898(D), 2934(D), 2942(D), 2967(D), 3016(D), 3065(C)	28%
1657	619(C), 641(C), 673(D), 689(D), 715(C), 723(D), 745(C), 755(C), 783(D), 810(D), 825(D), 851(D), 876(D), 893(D), 938(D), 976(C), 1001(C), 1030(C), 1056(D), 1097(C), 1125(D), 1156(D), 1171(C), 1206(D), 1234(D), 1254(D), 1307(C), 1337(D), 1396(X), 1448(C), 1484(D), 1554(D), 1578(C), 1603(D), 1616(D), 1684(C), 2852(C), 2873(C), 2898(D), 2934(D), 2942(C), 2967(C), 3016(C), 3065(C)	53%
1684	619(D), 641(D), 673(D), 689(D), 715(D), 723(D), 745(D), 755(C), 783(D), 810(D), 825(D), 851(D), 876(D), 893(D), 938(D), 976(D), 1001(D), 1030(D), 1056(D), 1097(D), 1125(D), 1156(D), 1171(D), 1206(D), 1234(D), 1254(D), 1307(D), 1337(D), 1396(X), 1448(D), 1484(D), 1554(D), 1578(D), 1603(D), 1616(D), 1657(D), 2852(D), 2873(D), 2898(D), 2934(D), 2942(D), 2967(D), 3016(D), 3065(C)	51%
2852	619(C), 641(C), 673(D), 689(D), 715(C), 723(D), 745(C), 755(C), 783(D), 810(D), 825(D), 851(D), 876(D), 893(D), 938(D), 976(C), 1001(C), 1030(C), 1056(D), 1097(C), 1125(D), 1156(D), 1171(C), 1206(D), 1234(D), 1254(D), 1307(C), 1337(D), 1396(X), 1448(C), 1484(D), 1554(D), 1578(C), 1603(D), 1616(C), 1657(C), 1684(C), 2873(C), 2898(D), 2934(D), 2942(C), 2967(C), 3016(C), 3065(C)	55%
2873	619(C), 641(C), 673(D), 689(D), 715(C), 723(D), 745(D), 755(C), 783(D), 810(D), 825(D), 851(D), 876(D), 893(D), 938(D), 976(C), 1001(C), 1030(C), 1056(D), 1097(C), 1125(D), 1156(D), 1171(C), 1206(D), 1234(D), 1254(D), 1307(C), 1337(D), 1396(X), 1448(C), 1484(D), 1554(D), 1578(C), 1603(D), 1616(C), 1657(C), 1684(C), 2852(C), 2898(D), 2934(D), 2942(C), 2967(C), 3016(C), 3065(C)	49%
2898	619(D), 641(D), 673(D), 689(D), 715(D), 723(D), 745(C), 755(D), 783(D), 810(D), 825(D), 851(D), 876(D), 893(D), 938(C), 976(C), 1001(D), 1030(D), 1056(D), 1097(C), 1125(D), 1156(D), 1171(D), 1206(C), 1234(D), 1254(D), 1307(D), 1337(D), 1396(X), 1448(C), 1484(D), 1554(D), 1578(C), 1603(D), 1616(D), 1657(C), 1684(D), 2852(D), 2873(C), 2934(D), 2942(D), 2967(D), 3016(D), 3065(C)	65%
2934	619(D), 641(D), 673(D), 689(D), 715(D), 723(D), 745(D), 755(D), 783(D), 810(D), 825(D), 851(D), 876(D), 893(D), 938(D), 976(C), 1001(D), 1030(D), 1056(D), 1097(D), 1125(D), 1156(D), 1171(D), 1206(D), 1234(D), 1254(D), 1307(D), 1337(D), 1396(X), 1448(D), 1484(D), 1554(D), 1578(D), 1603(D), 1616(D), 1657(D), 1684(D), 2852(D), 2873(D), 2898(D), 2942(D), 2967(D), 3016(D), 3065(C)	70%
2942	619(D), 641(D), 673(D), 689(D), 715(D), 723(D), 745(D), 755(C), 783(D), 810(D), 825(D), 851(D), 876(D), 893(D), 938(D), 976(D), 1001(D), 1030(D), 1056(D), 1097(D), 1125(D), 1156(D), 1171(D), 1206(D), 1234(D), 1254(D), 1307(D), 1337(D), 1396(X), 1448(D), 1484(D), 1554(D), 1578(C), 1603(D), 1616(D), 1657(D), 1684(D), 2852(D), 2873(C), 2898(D), 2934(D), 2967(D), 3016(D), 3065(C)	70%
2967	619(D), 641(D), 673(D), 689(D), 715(D), 723(D), 745(D), 755(C), 783(D), 810(D), 825(D), 851(D), 876(D), 893(D), 938(D), 976(D), 1001(D), 1030(D), 1056(D), 1097(D), 1125(D), 1156(D), 1171(D), 1206(D), 1234(D), 1254(D), 1307(D), 1337(D), 1396(X), 1448(D), 1484(D), 1554(D), 1578(C), 1603(D), 1616(D), 1657(C), 1684(C), 2852(C), 2873(C), 2898(D), 2934(D), 2942(C), 2967(D), 3016(C), 3065(C)	65%
3016	619(C), 641(C), 673(D), 689(D), 715(C), 723(D), 745(D), 755(C), 783(D), 810(D), 825(C), 851(D), 876(D), 893(D), 938(D), 976(D), 1001(D), 1030(C), 1056(D), 1097(C), 1125(D), 1156(D), 1171(C), 1206(D), 1234(D), 1254(D), 1307(C), 1337(D), 1396(X), 1448(C), 1484(D), 1554(D), 1578(C), 1603(D), 1616(C), 1657(C), 1684(C), 2852(C), 2873(C), 2898(D), 2934(D), 2942(C), 2967(C), 3065(C)	56%
3065	619(D), 641(D), 673(D), 689(D), 715(D), 723(D), 745(D), 755(D), 783(D), 810(D), 825(D), 851(D), 876(D), 893(D), 938(C), 976(D), 1001(D), 1030(D), 1056(D), 1097(D), 1125(D), 1156(D), 1171(C), 1206(D), 1234(D), 1254(D), 1307(D), 1337(D), 1396(X), 1448(C), 1484(D), 1554(D), 1578(C), 1603(D), 1616(D), 1657(C), 1684(D), 2852(D), 2873(C), 2898(D), 2934(D), 2942(D), 2967(D), 3016(D)	70%

Table C.3: Peak ratio analysis (PIR) of the comparison between PNT2-C2 live and LNCaP. The letter D specifies PIR separation between PNT2-C2 and LNCaP. The letter C specifies that the PIR marker is overlapping between samples.

Panel	PIRMatches	% matchD
619	641(D), 673(D), 689(D), 715(D), 723(C), 745(D), 755(D), 783(D), 810(D), 825(D), 851(D), 874(C), 893(C), 936(D), 976(D), 1001(D), 1030(D), 1060(D), 1087(D), 1125(D), 1155(C), 1171(D), 1206(D), 1254(D), 1297(D), 1337(D), 1396(D), 1448(D), 1484(C), 1554(D), 1578(D), 1603(C), 1616(D), 1657(D), 1684(C), 2852(D), 2873(D), 2896(D), 2934(D), 2942(D), 2967(D), 3016(D), 3065(D)	84%
641	619(D), 673(C), 689(D), 715(C), 723(D), 745(D), 755(C), 783(D), 810(D), 825(D), 851(C), 874(C), 893(C), 936(D), 976(D), 1001(D), 1030(D), 1060(C), 1087(C), 1125(D), 1155(C), 1171(D), 1206(D), 1254(C), 1297(D), 1337(C), 1396(D), 1448(C), 1484(D), 1554(D), 1578(C), 1603(C), 1616(C), 1657(D), 1684(D), 2852(D), 2873(D), 2896(D), 2934(C), 2942(C), 2967(D), 3016(D), 3065(C)	59%
673	619(D), 641(C), 689(C), 715(C), 723(D), 745(D), 755(C), 783(D), 810(C), 825(C), 851(C), 874(C), 893(C), 936(C), 976(D), 1001(C), 1030(C), 1060(C), 1087(C), 1125(C), 1155(C), 1171(D), 1206(C), 1254(C), 1297(C), 1337(C), 1396(D), 1448(C), 1484(D), 1554(D), 1578(D), 1603(C), 1616(C), 1657(C), 1684(C), 2852(D), 2873(C), 2896(C), 2934(C), 2942(C), 2967(C), 3016(D), 3065(C)	27%
689	641(D), 673(C), 715(C), 723(D), 745(D), 755(C), 783(D), 810(C), 825(C), 851(D), 874(C), 893(C), 936(C), 976(D), 1001(D), 1030(D), 1060(C), 1087(D), 1125(D), 1155(C), 1171(D), 1206(C), 1254(D), 1297(C), 1337(C), 1396(D), 1448(C), 1484(D), 1554(D), 1578(D), 1603(C), 1616(C), 1657(C), 1684(C), 2852(C), 2873(C), 2896(C), 2934(C), 2942(C), 2967(C), 3016(D), 3065(C)	45%
715	619(D), 641(C), 673(C), 689(C), 723(C), 745(D), 755(D), 783(D), 810(D), 825(D), 851(C), 874(C), 893(C), 936(C), 976(D), 1001(D), 1030(C), 1060(C), 1087(C), 1125(D), 1155(C), 1171(C), 1206(D), 1254(C), 1297(C), 1337(C), 1396(D), 1448(C), 1484(C), 1554(D), 1578(C), 1603(C), 1616(C), 1657(D), 1684(C), 2852(C), 2873(D), 2896(D), 2934(C), 2942(C), 2967(C), 3016(D), 3065(C)	34%
723	619(C), 641(D), 673(D), 689(D), 715(D), 745(D), 755(D), 783(C), 810(D), 825(D), 851(D), 874(C), 893(C), 936(C), 976(D), 1001(D), 1030(D), 1060(C), 1087(C), 1125(D), 1155(C), 1171(D), 1206(D), 1254(D), 1297(D), 1337(D), 1396(D), 1448(C), 1484(C), 1554(D), 1578(D), 1603(C), 1616(C), 1657(D), 1684(C), 2852(C), 2873(D), 2896(D), 2934(C), 2942(D), 2967(D), 3016(D), 3065(D)	86%
745	619(D), 641(D), 673(D), 689(D), 715(D), 723(D), 755(C), 783(D), 810(D), 825(D), 851(D), 874(C), 893(C), 936(D), 976(C), 1001(D), 1030(D), 1060(C), 1087(D), 1125(D), 1155(D), 1171(D), 1206(D), 1254(D), 1297(D), 1337(D), 1396(C), 1448(D), 1484(D), 1554(D), 1578(D), 1603(C), 1616(C), 1657(D), 1684(D), 2852(C), 2873(D), 2896(D), 2934(D), 2942(D), 2967(D), 3016(C), 3065(D)	87%
755	619(D), 641(C), 673(C), 689(C), 715(C), 723(D), 745(D), 783(D), 810(D), 825(C), 851(D), 874(C), 893(C), 936(C), 976(D), 1001(D), 1030(D), 1060(C), 1087(C), 1125(C), 1155(C), 1171(D), 1206(D), 1254(C), 1297(C), 1337(C), 1396(D), 1448(C), 1484(D), 1554(D), 1578(D), 1603(C), 1616(C), 1657(C), 1684(C), 2852(D), 2873(D), 2896(D), 2934(D), 2942(C), 2967(C), 3016(D), 3065(D)	59%
781	619(C), 641(D), 673(D), 689(D), 715(D), 723(D), 745(D), 755(D), 783(D), 810(D), 825(C), 851(D), 874(C), 893(C), 936(D), 976(D), 1001(D), 1030(D), 1060(C), 1087(D), 1125(C), 1155(C), 1171(C), 1206(D), 1254(D), 1297(D), 1337(D), 1396(C), 1448(D), 1484(C), 1554(D), 1578(D), 1603(D), 1616(D), 1657(D), 1684(C), 2852(D), 2873(D), 2896(D), 2934(D), 2942(D), 2967(D), 3016(D), 3065(D)	50%
803	619(C), 641(C), 673(C), 689(C), 715(C), 723(D), 745(D), 755(D), 783(D), 810(D), 825(D), 851(C), 874(C), 893(C), 936(C), 976(D), 1001(D), 1030(D), 1060(C), 1087(C), 1125(D), 1155(C), 1171(D), 1206(D), 1254(D), 1297(C), 1337(C), 1396(D), 1448(D), 1484(C), 1554(D), 1578(D), 1603(C), 1616(C), 1657(D), 1684(C), 2852(D), 2873(D), 2896(D), 2934(C), 2942(C), 2967(D), 3016(D), 3065(D)	59%
816	619(D), 641(C), 673(C), 689(C), 715(C), 723(D), 745(D), 755(C), 783(D), 810(D), 825(C), 851(D), 874(D), 893(C), 936(D), 976(D), 1001(D), 1030(D), 1060(C), 1087(C), 1125(C), 1155(C), 1171(D), 1206(D), 1254(C), 1297(C), 1337(C), 1396(D), 1448(C), 1484(D), 1554(D), 1578(D), 1603(C), 1616(C), 1657(C), 1684(C), 2852(D), 2873(C), 2896(D), 2934(D), 2942(D), 2967(C), 3016(D), 3065(D)	82%
976	619(D), 641(D), 673(D), 689(D), 715(D), 723(D), 745(D), 755(D), 783(D), 810(D), 825(D), 851(D), 874(D), 893(C), 936(D), 976(C), 1001(D), 1030(D), 1060(D), 1087(D), 1125(D), 1155(D), 1171(D), 1206(D), 1254(D), 1297(D), 1337(D), 1396(D), 1448(D), 1484(D), 1554(D), 1578(D), 1603(D), 1616(D), 1657(D), 1684(D), 2852(D), 2873(D), 2896(D), 2934(D), 2942(D), 2967(D), 3016(C), 3065(D)	98%
1001	619(D), 641(D), 673(C), 689(C), 715(C), 723(D), 745(D), 755(D), 783(D), 810(D), 825(D), 851(D), 876(C), 893(D), 936(C), 976(D), 1001(C), 1030(C), 1060(C), 1087(C), 1125(D), 1155(D), 1171(D), 1206(C), 1254(D), 1297(C), 1337(C), 1396(D), 1448(C), 1484(D), 1554(D), 1578(C), 1603(C), 1616(C), 1657(C), 1684(C), 2852(D), 2873(C), 2896(D), 2934(C), 2942(C), 2967(C), 3016(D), 3065(C)	48%
1030	619(D), 641(D), 673(C), 689(C), 715(C), 723(D), 745(D), 755(D), 783(D), 810(D), 825(D), 851(C), 876(D), 893(C), 936(D), 976(D), 1001(C), 1030(C), 1060(C), 1087(D), 1125(D), 1155(D), 1171(D), 1206(C), 1254(D), 1297(D), 1337(C), 1396(D), 1448(D), 1484(D), 1554(D), 1578(C), 1603(C), 1616(C), 1657(C), 1684(C), 2852(D), 2873(C), 2896(D), 2934(C), 2942(C), 2967(C), 3016(D), 3065(C)	50%
1060	619(D), 641(C), 673(C), 689(C), 715(C), 723(D), 745(D), 755(C), 783(D), 810(C), 825(C), 851(C), 876(C), 893(C), 936(C), 976(D), 1001(C), 1030(C), 1060(C), 1087(C), 1125(C), 1155(C), 1171(C), 1206(C), 1234(C), 1254(C), 1297(C), 1337(C), 1396(C), 1448(C), 1484(D), 1554(D), 1578(C), 1603(C), 1616(C), 1657(C), 1684(C), 2852(C), 2873(C), 2896(C), 2934(C), 2942(C), 2967(C), 3016(D), 3065(C)	18%
1087	619(D), 641(C), 673(C), 689(D), 715(C), 723(D), 745(D), 755(D), 783(D), 810(D), 825(D), 851(C), 876(C), 893(C), 936(C), 976(D), 1001(C), 1030(C), 1060(C), 1125(D), 1155(C), 1171(D), 1206(D), 1234(C), 1254(C), 1297(C), 1337(C), 1396(D), 1448(D), 1484(D), 1554(D), 1578(C), 1603(C), 1616(C), 1657(C), 1684(C), 2852(D), 2873(C), 2896(D), 2934(D), 2942(C), 2967(C), 3016(D), 3065(D)	48%
1125	619(D), 641(D), 673(D), 689(D), 715(D), 723(D), 745(D), 755(C), 783(D), 810(C), 825(C), 851(D), 876(D), 893(D), 936(D), 976(D), 1001(D), 1030(D), 1060(D), 1087(D), 1155(D), 1171(C), 1206(D), 1234(C), 1254(D), 1297(D), 1337(D), 1396(C), 1448(D), 1484(D), 1554(D), 1578(D), 1603(D), 1616(D), 1657(D), 1684(D), 2852(D), 2873(D), 2896(D), 2934(D), 2942(D), 2967(D), 3016(D), 3065(D)	80%
1155	619(D), 641(C), 673(C), 689(C), 715(C), 723(D), 745(D), 755(D), 783(D), 810(D), 825(D), 851(D), 876(C), 893(C), 936(C), 976(D), 1001(D), 1030(D), 1060(C), 1087(C), 1125(D), 1155(C), 1171(D), 1206(D), 1234(D), 1254(C), 1297(D), 1337(D), 1396(D), 1448(D), 1484(C), 1554(D), 1578(C), 1603(C), 1616(C), 1657(D), 1684(D), 2852(D), 2873(D), 2896(D), 2934(C), 2942(C), 2967(D), 3016(D), 3065(D)	64%
1171	619(D), 641(D), 673(D), 689(D), 715(D), 723(D), 745(D), 755(D), 783(D), 810(D), 825(C), 851(D), 876(D), 893(D), 936(D), 976(D), 1001(D), 1030(D), 1060(C), 1087(D), 1125(C), 1155(D), 1206(D), 1234(D), 1254(D), 1297(D), 1337(D), 1396(D), 1448(D), 1484(D), 1554(D), 1578(D), 1603(C), 1616(C), 1657(D), 1684(D), 2852(C), 2873(D), 2896(D), 2934(D), 2942(D), 2967(D), 3016(D), 3065(D)	82%
1206	619(D), 641(D), 673(C), 689(C), 715(C), 723(D), 745(D), 755(C), 783(D), 810(C), 825(C), 851(D), 876(D), 893(D), 936(C), 976(D), 1001(C), 1030(C), 1060(C), 1087(D), 1125(C), 1155(C), 1171(D), 1206(C), 1234(C), 1254(C), 1297(C), 1337(C), 1396(D), 1448(C), 1484(C), 1554(D), 1578(C), 1603(C), 1616(C), 1657(C), 1684(C), 2852(D), 2873(C), 2896(C), 2934(C), 2942(C), 2967(C), 3016(D), 3065(C)	52% 11
1254	619(D), 641(C), 673(C), 689(C), 715(C), 723(D), 745(D), 755(C), 783(D), 810(D), 825(D), 851(C), 876(C), 893(C), 936(C), 976(D), 1001(C), 1030(C), 1060(C), 1087(C), 1125(D), 1155(C), 1171(D), 1206(D), 1234(C), 1254(C), 1297(C), 1337(C), 1396(D), 1448(D), 1484(C), 1554(D), 1578(C), 1603(C), 1616(C), 1657(C), 1684(C), 2852(D), 2873(C), 2896(C), 2934(C), 2942(C), 2967(C), 3016(D), 3065(C)	55%
1307	619(D), 641(D), 673(C), 689(C), 715(C), 723(D), 745(D), 755(C), 783(D), 810(D), 825(D), 851(D), 876(D), 893(D), 936(C), 976(D), 1001(C), 1030(C), 1060(C), 1087(C), 1125(D), 1155(D), 1171(D), 1206(D), 1234(C), 1254(C), 1297(C), 1337(C), 1396(D), 1448(C), 1484(D), 1554(D), 1578(C), 1603(C), 1616(C), 1657(C), 1684(C), 2852(D), 2873(C), 2896(C), 2934(C), 2942(C), 2967(C), 3016(D), 3065(C)	45% 30
1337	619(D), 641(C), 673(C), 689(C), 715(C), 723(D), 745(D), 755(C), 783(D), 810(D), 825(D), 851(C), 876(C), 893(C), 936(C), 976(D), 1001(C), 1030(C), 1060(C), 1087(C), 1125(D), 1155(D), 1171(D), 1206(C), 1234(C), 1254(C), 1297(C), 1337(C), 1396(D), 1448(C), 1484(D), 1554(D), 1578(C), 1603(C), 1616(C), 1657(C), 1684(C), 2852(D), 2873(C), 2896(C), 2934(C), 2942(C), 2967(C), 3016(D), 3065(C)	39%
1396	619(D), 641(D), 673(D), 689(D), 715(D), 723(D), 745(D), 755(D), 783(D), 810(D), 825(D), 851(D), 876(D), 893(D), 936(D), 976(C), 1001(C), 1030(D), 1060(D), 1087(D), 1125(D), 1155(D), 1171(D), 1206(D), 1234(D), 1254(D), 1297(D), 1337(D), 1396(C), 1448(D), 1484(D), 1554(D), 1578(D), 1603(D), 1616(D), 1657(D), 1684(D), 2852(D), 2873(D), 2896(D), 2934(D), 2942(D), 2967(D), 3016(C), 3065(D)	84%
1448	619(D), 641(D), 673(C), 689(D), 715(D), 723(D), 745(D), 755(C), 783(D), 810(D), 825(D), 851(D), 876(D), 893(D), 936(D), 976(C), 1001(C), 1030(C), 1060(C), 1087(D), 1125(D), 1155(D), 1171(D), 1206(D), 1234(D), 1254(D), 1297(D), 1337(C), 1396(C), 1448(D), 1484(D), 1554(D), 1578(C), 1603(C), 1616(C), 1657(C), 1684(C), 2852(D), 2873(C), 2896(C), 2934(C), 2942(C), 2967(C), 3016(C), 3065(D)	45%
1484	619(C), 641(C), 673(C), 689(D), 715(C), 723(D), 745(D), 755(D), 783(C), 810(D), 825(D), 851(D), 876(C), 893(C), 936(D), 976(D), 1001(D), 1030(D), 1060(D), 1087(D), 1125(D), 1155(D), 1171(C), 1206(D), 1234(D), 1254(D), 1297(C), 1337(D), 1396(D), 1448(C), 1484(D), 1554(D), 1578(C), 1603(C), 1616(C), 1657(D), 1684(D), 2852(C), 2873(C), 2896(D), 2934(D), 2942(C), 2967(D), 3016(D), 3065(D)	70%
1554	619(D), 641(D), 673(D), 689(D), 715(D), 723(C), 745(D), 755(D), 783(D), 810(D), 825(D), 851(D), 876(D), 893(D), 936(D), 976(D), 1001(D), 1030(D), 1060(D), 1087(D), 1125(D), 1155(D), 1171(D), 1206(D), 1234(D), 1254(D), 1297(D), 1337(D), 1396(D), 1448(D), 1484(D), 1554(D), 1578(D), 1603(C), 1616(C), 1657(D), 1684(D), 2852(D), 2873(D), 2896(D), 2934(D), 2942(D), 2967(D), 3016(D), 3065(D)	98%
1578	619(D), 641(C), 673(C), 689(C), 715(C), 723(D), 745(D), 755(D), 783(D), 810(D), 825(D), 851(C), 876(C), 893(C), 936(C), 976(D), 1001(D), 1030(D), 1060(C), 1087(C), 1125(D), 1155(D), 1171(D), 1206(C), 1234(C), 1254(C), 1297(C), 1337(C), 1396(D), 1448(D), 1484(D), 1554(D), 1578(C), 1603(C), 1616(C), 1657(C), 1684(C), 2852(D), 2873(C), 2896(D), 2934(C), 2942(C), 2967(C), 3016(D), 3065(C)	43%
1603	619(D), 641(C), 673(C), 689(C), 715(C), 723(D), 745(D), 755(C), 783(D), 810(D), 825(D), 851(C), 876(C), 893(C), 936(C), 976(D), 1001(C), 1030(C), 1060(C), 1087(C), 1125(D), 1155(C), 1171(D), 1206(C), 1234(C), 1254(C), 1297(C), 1337(C), 1396(D), 1448(C), 1484(C), 1554(D), 1578(C), 1603(C), 1616(C), 1657(C), 1684(C), 2852(D), 2873(C), 2896(C), 2934(C), 2942(C), 2967(C), 3016(D), 3065(C)	30%
1615	619(D), 641(D), 673(C), 689(C), 715(C), 723(D), 745(D), 755(C), 783(D), 810(D), 825(C), 851(C), 876(C), 893(C), 936(C), 976(D), 1001(C), 1030(C), 1060(C), 1087(C), 1125(C), 1155(C), 1171(C), 1206(C), 1234(C), 1254(C), 1297(C), 1337(C), 1396(D), 1448(C), 1484(D), 1554(D), 1578(C), 1603(C), 1616(C), 1657(C), 1684(C), 2852(D), 2873(C), 2896(C), 2934(C), 2942(C), 2967(C), 3016(D), 3065(C)	25%
1657	619(D), 641(D), 673(C), 689(C), 715(C), 723(D), 745(D), 755(C), 783(D), 810(D), 825(D), 851(C), 876(D), 893(D), 936(C), 976(D), 1001(C), 1030(C), 1060(C), 1087(C), 1125(D), 1155(D), 1171(D), 1206(D), 1234(C), 1254(D), 1297(D), 1337(C), 1396(D), 1448(C), 1484(D), 1554(D), 1578(C), 1603(C), 1616(C), 1657(C), 1684(C), 2852(D), 2873(C), 2896(C), 2934(C), 2942(C), 2967(C), 3016(D), 3065(C)	43%
1684	619(C), 641(C), 673(C), 689(C), 715(C), 723(C), 745(D), 755(C), 783(D), 810(C), 825(D), 851(C), 876(C), 893(C), 936(C), 976(D), 1001(C), 1030(C), 1060(C), 1087(C), 1125(D), 1155(C), 1171(D), 1206(D), 1234(C), 1254(D), 1297(C), 1337(C), 1396(D), 1448(C), 1484(C), 1554(D), 1578(C), 1603(C), 1616(C), 1657(C), 1684(C), 2852(D), 2873(C), 2896(C), 2934(C), 2942(C), 2967(C), 3016(D), 3065(C)	27%
1852	619(D), 641(D), 673(D), 689(C), 715(C), 723(D), 745(D), 755(C), 783(D), 810(D), 825(D), 851(D), 876(D), 893(D), 936(C), 976(D), 1001(D), 1030(D), 1060(C), 1087(D), 1125(C), 1155(C), 1171(C), 1206(D), 1234(C), 1254(D), 1297(D), 1337(D), 1396(D), 1448(C), 1484(D), 1554(D), 1578(C), 1603(C), 1616(C), 1657(D), 1684(D), 2852(C), 2873(C), 2896(C), 2934(C), 2942(C), 2967(C), 3016(D), 3065(C)	15%
2873	619(D), 641(D), 673(D), 689(C), 715(C), 723(D), 745(D), 755(C), 783(D), 810(D), 825(D), 851(D), 876(D), 893(D), 936(C), 976(D), 1001(D), 1030(D), 1060(C), 1087(D), 1125(C), 1155(C), 1171(D), 1206(D), 1234(C), 1254(D), 1297(D), 1337(D), 1396(D), 1448(C), 1484(D), 1554(D), 1578(C), 1603(C), 1616(C), 1657(D), 1684(D), 2852(C), 2873(C), 2896(D), 2934(D), 2942(D), 2967(D), 3016(C), 3065(C)	45%
2896	619(D), 641(D), 673(C), 689(C), 715(C), 723(D), 745(D), 755(C), 783(D), 810(C), 825(C), 851(D), 876(D), 893(D), 936(C), 976(D), 1001(C), 1030(C), 1060(C), 1087(D), 1125(D), 1155(D), 1171(D), 1206(D), 1234(C), 1254(D), 1297(D), 1337(D), 1396(D), 1448(C), 1484(D), 1554(D), 1578(C), 1603(C), 1616(C), 1657(C), 1684(D), 2852(C), 2873(C), 2896(C), 2934(D), 2942(C), 2967(C), 3016(C), 3065(C)	52%
2934	619(D), 641(C), 673(C), 689(C), 715(C), 723(D), 745(D), 755(D), 783(D), 810(D), 825(D), 851(C), 876(C), 893(C), 936(C), 976(D), 1001(C), 1030(C), 1060(C), 1087(C), 1125(D), 1155(C), 1171(D), 1206(C), 1234(C), 1254(C), 1297(C), 1337(C), 1396(D), 1448(C), 1484(C), 1554(D), 1578(C), 1603(C), 1616(C), 1657(C), 1684(C), 2852(D), 2873(C), 2896(D), 2934(C), 2942(C), 2967(C), 3016(D), 3065(C)	34%
2942	619(D), 641(D), 673(C), 689(C), 715(C), 723(D), 745(D), 755(C), 783(D), 810(C), 825(D), 851(C), 876(C), 893(C), 936(C), 976(D), 1001(C), 1030(C), 1060(C), 1087(C), 1125(D), 1155(C), 1171(D), 1206(D), 1234(D), 1254(C), 1297(C), 1337(C), 1396(D), 1448(C), 1484(C), 1554(D), 1578(C), 1603(C), 1616(C), 1657(C), 1684(C), 2852(D), 2873(C), 2896(C), 2934(C), 2942(C), 2967(C), 3016(D), 3065(C)	48%
2967	619(D), 641(D), 673(C), 689(D), 715(C), 723(D), 745(D), 755(C), 783(D), 810(C), 825(D), 851(D), 876(D), 893(D), 936(C), 976(D), 1001(C), 1030(C), 1060(C), 1087(D), 1125(D), 1155(C), 1171(D), 1206(D), 1234(C), 1254(C), 1297(C), 1337(C), 1396(D), 1448(C), 1484(D), 1554(D), 1578(C), 1603(C), 1616(C), 1657(C), 1684(C), 2852(D), 2873(C), 2896(C), 2934(C), 2942(C), 2967(C), 3016(D), 3065(C)	91%
3016	619(D), 641(D), 673(D), 689(D), 715(D), 723(D), 745(C), 755(D), 783(D), 810(C), 825(D), 851(D), 876(D), 893(D), 936(C), 976(D), 1001(C), 1030(C), 1060(D), 1087(D), 1125(D), 1155(D), 1171(D), 1206(C), 1234(C), 1254(D), 1297(D), 1337(C), 1396(C), 1448(C), 1484(D), 1554(D), 1578(C), 1603(C), 1616(D), 1657(D), 1684(D), 2852(C), 2873(C), 2896(D), 2934(D), 2942(D), 2967(D), 3016(D), 3065(D)	55%
3065	619(D), 641(D), 673(C), 689(C), 715(C), 723(D), 745(D), 755(D), 783(D), 810(D), 825(D), 851(D), 876(D), 893(D), 936(C), 976(D), 1001(C), 1030(C), 1060(D), 1087(D), 1125(D), 1155(D), 1171(D), 1206(C), 1234(C), 1254(D), 1297(D), 1337(C), 1396(C), 1448(C), 1484(D), 1554(D), 1578(C), 1603(C), 1616(D), 1657(C), 1684(C), 2852(D), 2873(C), 2896(D), 2934(C), 2942(C), 2967(C), 3016(D)	

Table C.4: Peak ratio analysis (PIR) of the comparison between PNT2-C2 live and DU145. The letter D specifies PIR separation between PNT2-C2 and DU145. The letter C specifies that the PIR marker is overlapping between samples. The letter X specifies that the PIR comparison could not take place as the band was not present in DU145.

Panel	PIR Matches	% matchID
619	641(D), 673(D), 698(X, 715(D), 723(X, 745(C), 755(D), 788(D), 810(D), 825(D), 851(D), 876(X, 893(X), 936(D), 976(C), 1001(D),1030(D), 1060(D), 1087(D), 1125(C), 1155(D), 1171(D), 1206(D), 1234(D), 1307(D), 1337(D), 1396(X), 1448(D), 1484(D), 1554(D), 1578(D), 1603(D), 1616(X), 1657(D), 1684(D), 2852(D), 2873(D), 2896(D), 2934(D), 2942(D), 2967(D), 3016(C), 3065(D)	90%
641	619(D), 673(D), 698(X, 715(D), 723(X, 745(D), 755(D), 788(D), 810(D), 825(D), 851(D), 876(X, 893(X), 936(D), 976(D), 1001(D),1030(D), 1060(C), 1087(D), 1125(D), 1155(D), 1171(D), 1206(D), 1234(D), 1307(D), 1337(D), 1396(X), 1448(D), 1484(D), 1554(C), 1578(D), 1603(C), 1616(X), 1657(D), 1684(D), 2852(D), 2873(D), 2896(D), 2934(D), 2942(D), 2967(D), 3016(D), 3065(D)	95%
673	619(D), 641(D), 698(X, 715(C), 723(X, 745(D), 755(D), 788(D), 810(C), 825(D), 851(D), 876(X, 893(X), 936(C), 976(D), 1001(D),1030(D), 1060(D), 1087(D), 1125(D), 1155(D), 1171(C), 1206(C), 1234(D), 1307(D), 1337(D), 1396(X), 1448(D), 1484(C), 1554(D), 1578(D), 1603(D), 1616(X), 1657(D), 1684(C), 2852(D), 2873(C), 2896(D), 2934(C), 2942(D), 2967(D), 3016(D), 3065(D)	76%
715	619(D), 641(D), 673(D), 698(X, 723(X, 745(D), 755(D), 788(D), 810(C), 825(C), 851(D), 876(X, 893(X), 936(C), 976(D), 1001(D),1030(D), 1060(D), 1087(D), 1125(D), 1155(D), 1171(D), 1206(C), 1234(D), 1307(D), 1337(D), 1396(X), 1448(D), 1484(C), 1554(C), 1578(D), 1603(D), 1616(X), 1657(D), 1684(D), 2852(D), 2873(C), 2896(D), 2934(C), 2942(D), 2967(D), 3016(D), 3065(D)	82%
745	619(C), 641(D), 673(D), 698(X, 715(D), 723(X, 755(C), 788(D), 810(D), 825(D), 851(D), 876(X, 893(X), 936(D), 976(C), 1001(C),1030(C), 1060(D), 1087(D), 1125(C), 1155(C), 1171(D), 1206(D), 1234(D), 1307(D), 1337(D), 1396(X), 1448(C), 1484(D), 1554(D), 1578(D), 1603(D), 1616(X), 1657(C), 1684(D), 2852(C), 2873(D), 2896(D), 2934(D), 2942(D), 2967(D), 3016(C), 3065(C)	68%
755	619(D), 641(D), 673(D), 698(X, 715(D), 723(X), 745(C), 788(D), 810(D), 825(D), 851(D), 876(X, 893(X), 936(D), 976(C), 1001(C),1030(C), 1060(D), 1087(D), 1125(C), 1155(C), 1171(D), 1206(D), 1234(D), 1307(D), 1337(D), 1396(X), 1448(C), 1484(D), 1554(D), 1578(D), 1603(D), 1616(X), 1657(C), 1684(D), 2852(C), 2873(D), 2896(D), 2934(D), 2942(D), 2967(D), 3016(D), 3065(C)	71%
783	619(D), 641(D), 673(D), 698(X, 715(D), 723(X), 745(D), 755(D), 810(D), 825(D), 851(C), 876(X, 893(X), 936(D), 976(D), 1001(D),1030(D), 1060(D), 1087(C), 1125(D), 1155(D), 1171(D), 1206(D), 1234(D), 1307(C), 1337(D), 1396(X), 1448(C), 1484(D), 1554(D), 1578(C), 1603(D), 1616(X), 1657(C), 1684(D), 2852(C), 2873(D), 2896(C), 2934(D), 2942(C), 2967(C), 3016(D), 3065(C)	71%
810	619(D), 641(D), 673(C), 698(X), 715(C), 723(X), 745(D), 755(D), 788(D), 825(D), 851(D), 876(X), 893(X), 936(C), 976(D), 1001(D),1030(D), 1060(D), 1087(D), 1125(D), 1155(D), 1171(D), 1206(D), 1234(D), 1307(D), 1337(D), 1396(X), 1448(D), 1484(D), 1554(D), 1578(D), 1603(D), 1616(X), 1657(D), 1684(D), 2852(D), 2873(D), 2896(D), 2934(C), 2942(D), 2967(D), 3016(D), 3065(D)	84%
825	619(D), 641(D), 673(D), 679(D, 698(X), 715(C), 723(X), 745(D), 755(D), 788(D), 810(C), 851(D), 876(X), 893(X), 936(D), 976(D), 1001(D),1030(D), 1060(D), 1087(D), 1125(D), 1155(D), 1171(D), 1206(D), 1234(D), 1307(D), 1337(D), 1396(X), 1448(D), 1484(D), 1554(D), 1578(D), 1603(D), 1616(X), 1657(D), 1684(D), 2852(D), 2873(D), 2896(D), 2934(D), 2942(D), 2967(D), 3016(D), 3065(D)	95%
851	619(D), 641(D), 673(D), 698(X), 715(D), 723(X), 745(D), 755(D), 788(D), 810(C), 825(D), 876(X), 893(X), 936(C), 976(D), 1001(D),1030(D), 1060(D), 1087(D), 1125(D), 1155(D), 1171(D), 1206(C), 1234(C), 1307(C), 1337(D), 1396(X), 1448(D), 1484(D), 1554(D), 1578(D), 1603(D), 1616(X), 1657(D), 1684(C), 2852(D), 2873(D), 2896(D), 2934(D), 2942(D), 2967(C), 3016(D), 3065(D)	71%
868	619(D), 641(D), 673(D), 698(X), 715(C), 723(X), 745(D), 755(D), 788(D), 810(C), 825(D), 851(C), 876(X), 893(X), 936(D), 976(D), 1001(D),1030(D), 1060(D), 1087(D), 1125(D), 1155(D), 1171(C), 1206(C), 1234(C), 1307(C), 1337(D), 1396(X), 1448(D), 1484(C), 1554(C), 1578(D), 1603(D), 1616(X), 1657(D), 1684(C), 2852(D), 2873(D), 2896(D), 2934(C), 2942(D), 2967(C), 3016(D), 3065(D)	55%
876	619(C), 641(D), 673(D), 698(X), 715(D), 723(X), 745(C), 755(C), 788(D), 810(D), 825(D), 851(D), 893(X), 936(D), 976(C), 1001(C),1030(C), 1060(D), 1087(D), 1125(C), 1155(C), 1171(D), 1206(D), 1234(D), 1307(D), 1337(D), 1396(X), 1448(C), 1484(D), 1554(D), 1578(D), 1603(D), 1616(X), 1657(C), 1684(D), 2852(C), 2873(D), 2896(D), 2934(D), 2942(C), 2967(D), 3016(C), 3065(C)	55%
1001	619(D), 641(D), 673(D), 698(X), 715(D), 723(X), 745(C), 755(C), 788(D), 810(D), 825(D), 851(D), 876(X), 893(X), 936(D), 976(C), 1030(C), 1060(D), 1087(D), 1125(C), 1155(C), 1171(D), 1206(D), 1234(D), 1307(D), 1337(D), 1396(X), 1448(C), 1484(D), 1554(D), 1578(D), 1603(D), 1616(X), 1657(C), 1684(D), 2852(C), 2873(D), 2896(D), 2934(D), 2942(C), 2967(D), 3016(D), 3065(C)	71%
1030	619(D), 641(D), 673(D), 698(X), 715(D), 723(X), 745(C), 755(C), 788(D), 810(D), 825(D), 851(D), 876(X), 893(X), 936(D), 976(D), 1001(C), 1060(D), 1087(D), 1125(C), 1155(C), 1171(D), 1206(D), 1234(D), 1307(D), 1337(D), 1396(X), 1448(C), 1484(D), 1554(D), 1578(D), 1603(D), 1616(X), 1657(C), 1684(D), 2852(C), 2873(D), 2896(D), 2934(D), 2942(C), 2967(D), 3016(D), 3065(C)	74%
1060	619(D), 641(C), 673(D), 698(X), 715(D), 723(X), 745(D), 755(D), 788(D), 810(D), 825(D), 851(D), 876(X), 893(X), 936(D), 976(D), 1001(D), 1030(D), 1087(D), 1125(D), 1155(D), 1171(D), 1206(D), 1234(D), 1307(D), 1337(D), 1396(X), 1448(D), 1484(D), 1554(D), 1578(D), 1603(C), 1616(X), 1657(D), 1684(D), 2852(D), 2873(D), 2896(D), 2934(D), 2942(D), 2967(D), 3016(D), 3065(D)	95%
1087	619(D), 641(D), 673(D), 698(X), 715(D), 723(X), 745(D), 755(D), 788(D), 810(C), 851(D), 876(X), 893(X), 936(D), 976(C), 1001(C),1030(D), 1060(D), 1125(D), 1155(D), 1171(D), 1206(D), 1234(D), 1307(C), 1337(D), 1396(X), 1448(C), 1484(D), 1554(D), 1578(C), 1603(D), 1616(X), 1657(C), 1684(D), 2852(C), 2873(D), 2896(C), 2934(D), 2942(C), 2967(C), 3016(D), 3065(C)	66%
1125	619(C), 641(D), 673(D), 698(X), 715(D), 723(X), 745(C), 755(C), 788(D), 810(D), 825(D), 851(D), 876(X), 893(X), 936(D), 976(C), 1001(C),1030(D), 1060(D), 1087(D), 1155(C), 1171(D), 1206(D), 1234(D), 1307(D), 1337(D), 1396(X), 1448(C), 1484(D), 1554(D), 1578(D), 1603(D), 1616(X), 1657(C), 1684(D), 2852(C), 2873(D), 2896(D), 2934(D), 2942(D), 2967(D), 3016(C), 3065(D)	79%
1155	619(D), 641(D), 673(D), 698(X), 715(D), 723(X), 745(C), 755(C), 788(D), 810(D), 825(D), 851(D), 876(X), 893(X), 936(D), 976(C), 1001(C),1030(D), 1060(D), 1087(D), 1125(C), 1171(D), 1206(D), 1234(D), 1307(D), 1337(D), 1396(X), 1448(D), 1484(C), 1554(C), 1578(D), 1603(D), 1616(X), 1657(D), 1684(C), 2852(D), 2873(D), 2896(D), 2934(D), 2942(D), 2967(D), 3016(C), 3065(C)	71%
1171	619(D), 641(D), 673(D), 698(X), 715(D), 723(X), 745(D), 755(C), 788(D), 810(D), 825(D), 851(D), 876(X), 893(X), 936(D), 976(C), 1001(D),1030(D), 1060(D), 1087(D), 1125(D), 1155(D), 1206(C), 1234(C), 1307(C), 1337(C), 1396(X), 1448(D), 1484(C), 1554(C), 1578(D), 1603(D), 1616(X), 1657(D), 1684(C), 2852(D), 2873(C), 2896(D), 2934(C), 2942(D), 2967(D), 3016(D), 3065(D)	71%
1206	619(D), 641(D), 673(C), 698(X), 715(D), 723(X), 745(D), 755(D), 788(D), 810(D), 825(D), 851(C), 876(X), 893(X), 936(D), 976(D), 1001(D),1030(D), 1060(D), 1087(D), 1125(D), 1155(D), 1171(C), 1234(C), 1307(C), 1337(C), 1396(X), 1448(D), 1484(C), 1554(C), 1578(C), 1603(D), 1616(X), 1657(D), 1684(C), 2852(D), 2873(C), 2896(C), 2934(C), 2942(D), 2967(D), 3016(D), 3065(D)	66%
1234	619(D), 641(D), 673(D), 698(X), 715(D), 723(X), 745(D), 755(C), 788(D), 810(D), 825(D), 851(D), 876(X), 893(X), 936(D), 976(D), 1001(D),1030(D), 1060(D), 1087(D), 1125(D), 1155(D), 1171(D), 1206(C), 1234(C), 1307(C), 1337(C), 1396(X), 1448(D), 1484(C), 1554(C), 1578(D), 1603(D), 1616(X), 1657(D), 1684(C), 2852(D), 2873(C), 2896(D), 2934(C), 2942(D), 2967(D), 3016(D), 3065(D)	71%
1307	619(D), 641(D), 673(D), 698(X), 715(D), 723(X), 745(D), 755(D), 788(C), 810(D), 825(D), 851(C), 876(X), 893(X), 936(C), 976(D), 1001(D),1030(D), 1060(D), 1087(C), 1125(D), 1155(D), 1171(D), 1206(D), 1234(C), 1307(C), 1337(C), 1396(X), 1448(D), 1484(D), 1554(C), 1578(C), 1603(D), 1616(X), 1657(C), 1684(C), 2852(D), 2873(C), 2896(C), 2934(C), 2942(C), 2967(C), 3016(D), 3065(D)	58%
1337	619(D), 641(D), 673(D), 698(X), 715(D), 723(X), 745(D), 755(D), 788(D), 810(D), 825(D), 851(C), 876(X), 893(X), 936(C), 976(D), 1001(D),1030(D), 1060(D), 1087(D), 1125(D), 1155(D), 1171(D), 1206(C), 1234(C), 1307(C), 1337(C), 1396(X), 1448(D), 1484(D), 1554(C), 1578(C), 1603(D), 1616(X), 1657(D), 1684(C), 2852(D), 2873(C), 2896(C), 2934(C), 2942(C), 2967(C), 3016(D), 3065(D)	66%
1448	619(D), 641(D), 673(D), 679(D, 698(X), 715(D), 723(X), 744(C), 755(C), 788(C), 810(D), 825(D), 851(C), 876(X), 893(X), 936(D), 976(C), 1001(C),1030(C), 1060(D), 1087(D), 1125(D), 1155(C), 1171(D), 1206(D), 1234(D), 1254(D), 1307(C), 1337(D), 1396(X), 1448(D), 1484(D), 1554(D), 1578(C), 1603(D), 1616(X), 1657(D), 1684(D), 2852(C), 2873(D), 2896(C), 2934(D), 2942(C), 2967(C), 3016(D), 3065(D)	55%
1484	619(D), 641(D), 673(D), 679(D, 698(X), 715(C), 723(X), 744(C), 755(C), 788(C), 810(D), 825(D), 851(D), 876(X), 893(X), 936(D), 976(C), 1001(C),1030(C), 1060(D), 1087(D), 1125(D), 1155(C), 1171(D), 1206(C), 1234(C), 1254(D), 1307(D), 1337(D), 1396(X), 1448(D), 1484(C), 1554(C), 1578(D), 1603(D), 1616(X), 1657(D), 1684(C), 2852(D), 2873(D), 2896(C), 2934(C), 2942(D), 2967(C), 3016(D), 3065(D)	71%
1554	619(D), 641(D), 673(D), 679(D, 698(X), 715(C), 723(X), 744(C), 755(C), 788(C), 810(D), 825(D), 851(D), 876(X), 893(X), 936(D), 976(C), 1001(C),1030(C), 1060(D), 1087(D), 1125(D), 1155(C), 1171(D), 1206(C), 1234(C), 1254(D), 1307(D), 1337(D), 1396(X), 1448(D), 1484(C), 1554(C), 1578(C), 1603(D), 1616(X), 1657(D), 1684(C), 2852(D), 2873(D), 2896(C), 2934(C), 2942(D), 2967(C), 3016(D), 3065(D)	71%
1578	619(D), 641(D), 673(D), 698(X), 715(D), 723(X), 744(C), 755(C), 788(C), 810(D), 825(D), 851(D), 876(X), 893(X), 936(D), 976(C), 1001(D),1030(D), 1060(D), 1087(D), 1125(C), 1155(D), 1171(D), 1206(D), 1234(D), 1254(C), 1307(C), 1337(D), 1396(X), 1448(D), 1484(D), 1554(D), 1578(C), 1603(D), 1616(X), 1657(D), 1684(D), 2852(C), 2873(D), 2896(C), 2934(D), 2942(C), 2967(D), 3016(D), 3065(D)	61%
1603	619(D), 641(C), 673(D), 698(X), 715(D), 723(X), 745(D), 755(D), 788(D), 810(D), 825(D), 851(D), 876(X), 893(X), 936(D), 976(C), 1001(D),1030(D), 1060(D), 1087(D), 1125(C), 1155(D), 1171(D), 1206(D), 1234(D), 1254(D), 1307(D), 1337(D), 1396(X), 1448(D), 1484(D), 1554(D), 1578(D), 1616(X), 1657(D), 1684(D), 2852(D), 2873(D), 2896(D), 2934(D), 2942(D), 2967(D), 3016(C), 3065(D)	87%
1657	619(D), 641(D), 673(D), 698(X), 715(D), 723(X), 745(C), 755(C), 788(C), 810(D), 825(D), 851(D), 876(X), 893(X), 936(D), 976(C), 1001(C),1030(C), 1060(D), 1087(C), 1125(D), 1155(D), 1171(C), 1206(D), 1234(D), 1254(D), 1307(D), 1337(D), 1396(X), 1448(C), 1484(D), 1554(D), 1578(C), 1603(D), 1616(X), 1684(D), 2852(C), 2873(D), 2896(C), 2934(D), 2942(C), 2967(C), 3016(D), 3065(C)	61%
1684	619(D), 641(D), 673(C), 698(X), 715(D), 723(X), 745(C), 755(C), 788(C), 810(D), 825(D), 851(D), 876(X), 893(X), 936(D), 976(D), 1001(D),1030(D), 1060(D), 1087(D), 1125(D), 1155(D), 1171(C), 1206(C), 1234(C), 1254(C), 1307(C), 1337(C), 1396(X), 1448(C), 1484(C), 1554(C), 1578(D), 1603(D), 1616(X), 1657(D), 1684(C), 2852(D), 2873(D), 2896(C), 2934(C), 2942(D), 2967(C), 3016(D), 3065(D)	68%
2852	619(D), 641(D), 673(D), 698(X), 715(D), 723(X), 745(C), 755(C), 788(C), 810(D), 825(D), 851(C), 876(X), 893(X), 936(D), 976(C), 1001(C),1030(C), 1060(D), 1087(C), 1125(C), 1155(C), 1171(D), 1206(D), 1234(D), 1254(D), 1307(C), 1337(C), 1396(X), 1448(C), 1484(D), 1554(D), 1578(C), 1603(D), 1616(X), 1657(C), 1684(D), 2873(D), 2896(C), 2934(D), 2942(C), 2967(C), 3016(D), 3065(C)	50%
2873	619(D), 641(D), 673(C), 698(X), 715(D), 723(X), 745(C), 755(C), 788(C), 810(D), 825(D), 851(D), 876(X), 893(X), 936(D), 976(D), 1001(D),1030(D), 1060(D), 1087(D), 1125(C), 1155(C), 1171(D), 1206(D), 1234(D), 1254(C), 1307(C), 1337(C), 1396(X), 1448(D), 1484(C), 1554(C), 1578(C), 1603(D), 1616(X), 1657(D), 1684(C), 2852(D), 2896(D), 2934(C), 2942(D), 2967(D), 3016(D), 3065(D)	68%
2896	619(D), 641(D), 673(D), 698(X), 715(D), 723(X), 745(C), 755(C), 788(C), 810(D), 825(D), 851(C), 876(X), 893(X), 936(D), 976(D), 1001(D),1030(D), 1060(D), 1087(C), 1125(C), 1155(D), 1171(D), 1206(D), 1234(D), 1254(D), 1307(C), 1337(C), 1396(X), 1448(C), 1484(D), 1554(D), 1578(C), 1603(D), 1616(X), 1657(C), 1684(D), 2852(C), 2873(D), 2934(D), 2942(C), 2967(C), 3016(D), 3065(D)	66%
2934	619(D), 641(D), 673(D), 679(D, 698(X), 715(C), 723(X), 745(D), 755(D), 788(D), 810(C), 825(D), 851(D), 876(X), 893(X), 936(D), 976(D), 1001(D),1030(D), 1060(D), 1087(D), 1125(D), 1155(D), 1171(D), 1206(D), 1234(D), 1254(C), 1307(D), 1337(D), 1396(X), 1448(D), 1484(C), 1554(C), 1578(D), 1603(D), 1616(X), 1657(D), 1684(C), 2852(D), 2873(D), 2896(D), 2942(D), 2967(D), 3016(D), 3065(D)	66%
2942	619(D), 641(D), 673(D), 679(D, 698(X), 715(D), 723(X), 745(D), 755(C), 788(C), 810(D), 825(D), 851(C), 876(X), 893(X), 936(D), 976(C), 1001(C),1030(C), 1060(D), 1087(C), 1125(D), 1155(D), 1171(D), 1206(D), 1234(D), 1254(D), 1307(D), 1337(D), 1396(X), 1448(C), 1484(D), 1554(D), 1578(C), 1603(D), 1616(X), 1657(C), 1684(D), 2852(D), 2873(C), 2896(D), 2934(D), 2967(C), 3016(D), 3065(C)	63%
2967	619(D), 641(D), 673(D), 679(D, 698(X), 715(D), 723(X), 745(D), 755(D), 788(C), 810(D), 825(D), 851(C), 876(X), 893(X), 936(D), 976(C), 1001(C),1030(C), 1060(D), 1087(C), 1125(D), 1155(D), 1171(D), 1206(D), 1234(D), 1254(D), 1307(D), 1337(D), 1396(X), 1448(C), 1484(D), 1554(D), 1578(C), 1603(D), 1616(X), 1657(C), 1684(D), 2852(C), 2873(D), 2896(C), 2934(D), 2942(C), 2967(C), 3016(D), 3065(D)	66%
3016	619(C), 641(D), 673(D), 679(D, 698(X), 715(D), 723(X), 745(C), 755(C), 788(D), 810(D), 825(D), 851(D), 876(X), 893(X), 936(D), 976(C), 1001(C),1030(C), 1060(D), 1087(D), 1125(C), 1155(C), 1171(D), 1206(D), 1234(D), 1254(D), 1307(D), 1337(D), 1396(X), 1448(C), 1484(C), 1554(D), 1578(C), 1603(D), 1616(X), 1657(D), 1684(D), 2852(C), 2873(D), 2896(D), 2934(D), 2942(D), 2967(D), 3016(C), 3065(D)	82%
3065	619(C), 641(D), 673(D), 679(D, 698(X), 715(D), 723(X), 745(C), 755(C), 788(D), 810(D), 825(D), 851(D), 876(X), 893(X), 936(D), 976(C), 1001(C),1030(C), 1060(D), 1087(D), 1125(C), 1155(C), 1171(C), 1206(D), 1234(D), 1254(D), 1307(D), 1337(D), 1396(X), 1448(C), 1484(C), 1554(D), 1578(C), 1603(D), 1616(X), 1657(C), 1684(D), 2852(C), 2873(D), 2896(D), 2934(D), 2942(D), 2967(D), 3016(C), 3065(C)	66%

Table C.5: Peak ratio analysis (PIR) of the comparison between PNT2-C2 live and PC3. The letter D specifies PIR separation between PNT2-C2 and DU145. The letter C specifies that the PIR marker is overlapping between samples.

Panel	PIR Matches	% matchD
619	641(D), 673(D), 689(D), 715(D), 723(D), 745(D), 755(D), 783(D), 810(C), 825(D), 851(D), 876(C), 893(D), 936(D), 976(D), 1001(D),1030(D), 1060(D), 1087(D), 1125(D), 1155(D), 1171(D), 1206(D), 1234(D), 1254(D), 1307(D), 1337(D), 1396(D), 1448(D), 1484(D), 1554(D), 1578(D), 1603(D), 1616(C), 1657(D), 1684(D), 2852(D),2873(D), 2896(D), 2934(D), 2942(D), 2967(D), 3016(D), 3065(D)	35%
641	619(D), 673(C), 689(D), 715(C), 723(D), 745(D), 755(C), 783(C), 810(C), 825(D), 851(D), 876(C), 893(C), 936(C), 976(C), 1001(C),1030(D), 1060(C), 1087(C), 1125(D), 1155(C), 1171(D), 1206(D), 1234(C), 1254(D), 1307(D), 1337(C), 1396(D), 1448(C), 1484(D), 1554(C), 1578(C), 1603(C), 1616(C), 1657(C), 1684(D), 2852(C), 2873(D), 2896(D), 2934(D), 2942(C), 2967(D), 3016(D), 3065(D)	48%
673	619(D), 641(D), 689(D), 715(C), 723(C), 745(D), 755(D), 783(D), 810(C), 825(D), 851(D), 876(D), 893(D), 936(C), 976(C), 1001(C),1030(C), 1060(D), 1087(D), 1125(D), 1155(C), 1171(D), 1206(C), 1234(D), 1254(C), 1307(C), 1337(C), 1396(D), 1448(C), 1484(D), 1554(C), 1578(D), 1603(C), 1616(D), 1657(C), 1684(D), 2852(C), 2873(D), 2896(D), 2934(C), 2942(D), 2967(C), 3016(C), 3065(C)	45%
689	619(D), 641(D), 673(D), 715(C), 723(C), 745(D), 755(C), 783(D), 810(C), 825(C), 851(D), 876(D), 893(C), 936(D), 976(C), 1001(D),1030(C), 1060(D), 1087(D), 1125(C), 1155(C), 1171(C), 1206(D), 1234(D), 1254(C), 1307(D), 1337(C), 1396(C), 1448(D), 1484(C), 1554(C), 1578(C), 1603(C), 1616(C), 1657(C), 1684(D), 2852(D), 2873(C), 2896(D), 2934(C), 2942(D), 2967(D), 3016(C), 3065(C)	55%
715	619(D), 641(C), 673(C), 689(C), 723(D), 745(D), 755(D), 783(C), 810(C), 825(D), 851(C), 876(D), 893(D), 936(C), 976(C), 1001(D),1030(C), 1060(D), 1087(D), 1125(C), 1155(C), 1171(D), 1206(C), 1234(D), 1254(C), 1307(C), 1337(C), 1396(D), 1448(C), 1484(D), 1554(C), 1578(C), 1603(C), 1616(D), 1657(C), 1684(D), 2852(C), 2873(C), 2896(C), 2934(D), 2942(C), 2967(D), 3016(D), 3065(C)	32%
723	619(D), 641(D), 673(C), 689(C), 715(D), 745(D), 755(C), 783(C), 810(C), 825(C), 851(D), 876(D), 893(C), 936(D), 976(C), 1001(D),1030(C), 1060(D), 1087(D), 1125(C), 1155(C), 1171(C), 1206(D), 1234(D), 1254(D), 1307(C), 1337(C), 1396(C), 1448(C), 1484(D), 1554(C), 1578(C), 1603(C), 1616(D), 1657(C), 1684(D), 2852(C), 2873(C), 2896(C), 2934(D), 2942(C), 2967(C), 3016(C), 3065(C)	34%
745	619(D), 641(D), 673(D), 689(D), 715(D), 723(D), 755(D), 783(D), 810(C), 825(D), 851(D), 876(D), 893(D), 936(D), 976(C), 1001(D),1030(D), 1060(D), 1087(D), 1125(D), 1155(D), 1171(D), 1206(D), 1234(D), 1254(D), 1307(D), 1337(D), 1396(D), 1448(D), 1484(D), 1554(D), 1578(D), 1603(D), 1616(D), 1657(D), 1684(D), 2852(D), 2873(D), 2896(D), 2934(D), 2942(D), 2967(D), 3016(D), 3065(D)	90%
755	619(D), 641(C), 673(C), 689(D), 715(C), 723(D), 745(D), 783(C), 810(C), 825(C), 851(D), 876(C), 893(C), 936(C), 976(C), 1001(D),1030(C), 1060(C), 1087(D), 1125(D), 1155(C), 1171(D), 1206(C), 1234(D), 1254(D), 1307(C), 1337(C), 1396(D), 1448(C), 1484(D), 1554(C), 1578(D), 1603(C), 1616(C), 1657(C), 1684(D), 2852(C), 2873(C), 2896(C), 2934(D), 2942(C), 2967(C), 3016(D), 3065(D)	89%
783	619(D), 641(C), 673(C), 689(D), 715(C), 723(C), 745(D), 755(C), 810(C), 825(C), 851(C), 876(C), 893(C), 936(C), 976(C), 1001(D),1030(C), 1060(C), 1087(C), 1125(D), 1155(C), 1171(D), 1206(C), 1234(D), 1254(D), 1307(C), 1337(C), 1396(D), 1448(C), 1484(D), 1554(C), 1578(D), 1603(C), 1616(C), 1657(C), 1684(D), 2852(C), 2873(C), 2896(C), 2934(D), 2942(C), 2967(C), 3016(D), 3065(D)	89%
810	619(D), 641(C), 673(C), 689(D), 715(C), 723(C), 745(D), 755(C), 783(C), 825(C), 851(C), 876(C), 893(C), 936(C), 976(C), 1001(D),1030(C), 1060(C), 1087(C), 1125(D), 1155(C), 1171(D), 1206(C), 1234(D), 1254(D), 1307(C), 1337(C), 1396(D), 1448(C), 1484(D), 1554(C), 1578(D), 1603(C), 1616(C), 1657(C), 1684(D), 2852(C), 2873(C), 2896(C), 2934(D), 2942(C), 2967(C), 3016(D), 3065(D)	89%
825	619(D), 641(C), 673(C), 689(D), 715(C), 723(C), 745(D), 755(C), 783(C), 810(C), 851(C), 876(C), 893(C), 936(C), 976(C), 1001(D),1030(C), 1060(C), 1087(C), 1125(D), 1155(C), 1171(D), 1206(C), 1234(D), 1254(D), 1307(C), 1337(C), 1396(D), 1448(C), 1484(D), 1554(C), 1578(D), 1603(C), 1616(C), 1657(C), 1684(D), 2852(C), 2873(C), 2896(C), 2934(D), 2942(C), 2967(C), 3016(D), 3065(D)	89%
851	619(D), 641(C), 673(D), 689(D), 715(D), 723(D), 745(D), 755(D), 783(C), 810(C), 825(D), 876(D), 893(D), 936(C), 976(C), 1001(D),1030(D), 1060(D), 1087(D), 1125(D), 1155(D), 1171(D), 1206(C), 1234(D), 1254(D), 1307(D), 1337(D), 1396(D), 1448(D), 1484(D), 1554(D), 1578(D), 1603(D), 1616(D), 1657(D), 1684(D), 2852(D), 2873(D), 2896(D), 2934(D), 2942(D), 2967(D), 3016(D), 3065(D)	89%
876	619(C), 641(D), 673(D), 689(D), 715(D), 723(D), 745(D), 755(C), 783(D), 810(C), 825(C), 851(D), 893(D), 936(D), 976(D), 1001(D),1030(D), 1060(C), 1087(D), 1125(D), 1155(D), 1171(D), 1206(C), 1234(C), 1254(D), 1307(D), 1337(D), 1396(D), 1448(D), 1484(D), 1554(D), 1578(D), 1603(D), 1616(D), 1657(D), 1684(D), 2852(D), 2873(D), 2896(D), 2934(D), 2942(D), 2967(D), 3016(D), 3065(D)	89%
893	619(D), 641(D), 673(C), 689(D), 715(C), 723(C), 745(D), 755(C), 783(D), 810(C), 825(C), 851(D), 876(C), 936(C), 976(C), 1001(D),1030(C), 1060(C), 1087(D), 1125(D), 1155(C), 1171(D), 1206(C), 1234(D), 1254(D), 1307(C), 1337(C), 1396(D), 1448(C), 1484(D), 1554(C), 1578(D), 1603(C), 1616(C), 1657(C), 1684(D), 2852(C), 2873(C), 2896(C), 2934(D), 2942(C), 2967(C), 3016(D), 3065(D)	55%
936	619(D), 641(C), 673(C), 689(D), 715(C), 723(C), 745(D), 755(C), 783(C), 810(C), 825(C), 851(C), 876(C), 893(C), 976(C), 1001(D),1030(C), 1060(C), 1087(C), 1125(D), 1155(C), 1171(D), 1206(C), 1234(D), 1254(D), 1307(C), 1337(C), 1396(D), 1448(C), 1484(D), 1554(C), 1578(D), 1603(C), 1616(C), 1657(C), 1684(D), 2852(C), 2873(C), 2896(C), 2934(D), 2942(C), 2967(C), 3016(D), 3065(D)	16%
976	619(D), 641(C), 673(C), 689(D), 715(C), 723(C), 745(D), 755(C), 783(C), 810(C), 825(C), 851(C), 876(C), 893(C), 936(C), 1001(D),1030(C), 1060(C), 1087(C), 1125(D), 1155(C), 1171(D), 1206(C), 1234(D), 1254(D), 1307(C), 1337(C), 1396(D), 1448(C), 1484(C), 1554(C), 1578(D), 1603(C), 1616(C), 1657(C), 1684(C), 2852(C), 2873(C), 2896(C), 2934(C), 2942(C), 2967(C), 3016(C), 3065(C)	16%
1001	619(D), 641(C), 673(C), 689(D), 715(C), 723(C), 745(D), 755(D), 783(C), 810(C), 825(D), 851(D), 876(D), 893(C), 936(C), 976(C), 1030(C), 1060(C), 1087(C), 1125(D), 1155(C), 1171(D), 1206(C), 1234(D), 1254(C), 1307(C), 1337(C), 1396(D), 1448(C), 1484(D), 1554(C), 1578(D), 1603(C), 1616(C), 1657(C), 1684(D), 2852(C), 2873(C), 2896(C), 2934(D), 2942(C), 2967(C), 3016(C), 3065(D)	36%
1030	619(D), 641(D), 673(C), 689(D), 715(C), 723(D), 745(D), 755(D), 783(D), 810(C), 825(D), 851(D), 876(D), 893(D), 936(D), 976(C), 1001(C), 1060(D), 1087(D), 1125(D), 1155(C), 1171(D), 1206(C), 1234(D), 1254(C), 1307(C), 1337(C), 1396(D), 1448(C), 1484(D), 1554(C), 1578(D), 1603(C), 1616(D), 1657(C), 1684(D), 2852(C), 2873(C), 2896(C), 2934(C), 2942(C), 2967(C), 3016(C), 3065(C)	45%
1060	619(D), 641(C), 673(D), 689(D), 715(C), 723(D), 745(D), 755(C), 783(D), 810(C), 825(D), 851(C), 876(D), 893(D), 936(C), 976(C), 1001(C), 1030(D), 1060(C), 1125(D), 1155(C), 1171(C), 1206(D), 1234(D), 1254(C), 1307(D), 1337(C), 1396(D), 1448(C), 1484(C), 1554(C), 1578(D), 1603(C), 1616(C), 1657(D), 1684(C), 2852(C), 2873(D), 2896(D), 2934(D), 2942(C), 2967(D), 3016(D), 3065(D)	57%
1087	619(D), 641(C), 673(D), 689(D), 715(C), 723(D), 745(D), 755(D), 783(C), 810(C), 825(D), 851(C), 876(D), 893(D), 936(C), 976(C), 1001(C), 1030(D), 1060(C), 1125(D), 1155(C), 1171(D), 1206(C), 1234(D), 1254(C), 1307(D), 1337(D), 1396(D), 1448(C), 1484(C), 1554(C), 1578(D), 1603(C), 1616(C), 1657(C), 1684(C), 2852(C), 2873(D), 2896(D), 2934(D), 2942(D), 2967(D), 3016(D), 3065(D)	52%
1125	619(D), 641(D), 673(D), 689(D), 715(D), 723(C), 745(D), 755(D), 783(D), 810(C), 825(D), 851(D), 876(D), 893(D), 936(D), 976(D), 1001(D),1030(D), 1060(D), 1087(D), 1155(D), 1171(D), 1206(C), 1234(D), 1254(D), 1307(D), 1337(D), 1396(C), 1448(D), 1484(C), 1554(D), 1578(C), 1603(D), 1616(D), 1657(D), 1684(D), 2852(D), 2873(D), 2896(D), 2934(D), 2942(D), 2967(D), 3016(D), 3065(D)	82%
1155	619(D), 641(C), 673(C), 689(C), 715(C), 723(C), 745(D), 755(C), 783(C), 810(C), 825(C), 851(D), 876(D), 893(C), 936(C), 976(C), 1001(C),1030(C), 1060(C), 1087(C), 1125(C), 1171(D), 1206(C), 1234(C), 1254(C), 1307(C), 1337(C), 1396(D), 1448(C), 1484(C), 1554(C), 1578(C), 1603(C), 1616(C), 1657(C), 1684(D), 2852(C), 2873(C), 2896(C), 2934(C), 2942(C), 2967(C), 3016(C), 3065(C)	25%
1171	619(D), 641(C), 673(C), 689(D), 715(C), 723(D), 745(D), 755(D), 783(C), 810(C), 825(D), 851(C), 876(C), 893(C), 936(C), 976(C), 1001(C),1030(C), 1060(C), 1087(C), 1125(C), 1155(C), 1206(C), 1234(D), 1254(D), 1307(C), 1337(C), 1396(D), 1448(C), 1484(C), 1554(C), 1578(C), 1603(C), 1616(C), 1657(C), 1684(C), 2852(C), 2873(C), 2896(C), 2934(C), 2942(C), 2967(C), 3016(C), 3065(C)	86%
1206	619(D), 641(D), 673(D), 689(D), 715(D), 723(D), 745(D), 755(C), 783(D), 810(C), 825(D), 851(D), 876(D), 893(D), 936(D), 976(D), 1001(D),1030(D), 1060(D), 1087(D), 1125(C), 1155(C), 1171(D), 1234(D), 1254(D), 1307(D), 1337(D), 1396(D), 1448(D), 1484(D), 1554(D), 1578(D), 1603(D), 1616(D), 1657(D), 1684(D), 2852(D), 2873(D), 2896(D), 2934(D), 2942(D), 2967(D), 3016(D), 3065(D)	32%
1234	619(D), 641(D), 673(D), 689(D), 715(D), 723(D), 745(D), 755(C), 783(D), 810(C), 825(D), 851(D), 876(D), 893(D), 936(D), 976(D), 1001(D),1030(D), 1060(D), 1087(D), 1125(D), 1155(C), 1171(D), 1206(C), 1254(D), 1307(D), 1337(D), 1396(D), 1448(D), 1484(D), 1554(D), 1578(D), 1603(D), 1616(D), 1657(D), 1684(D), 2852(D), 2873(D), 2896(D), 2934(D), 2942(D), 2967(D), 3016(D), 3065(D)	84%
1254	619(D), 641(C), 673(C), 689(D), 715(C), 723(C), 745(D), 755(D), 783(C), 810(C), 825(D), 851(C), 876(D), 893(D), 936(C), 976(C), 1001(C),1030(C), 1060(C), 1087(C), 1125(D), 1155(C), 1171(D), 1206(C), 1234(D), 1254(C), 1307(D), 1337(C), 1396(D), 1448(C), 1484(C), 1554(C), 1578(D), 1603(C), 1616(C), 1657(C), 1684(C), 2852(C), 2873(C), 2896(C), 2934(C), 2942(C), 2967(C), 3016(C), 3065(C)	41%
1307	619(D), 641(D), 673(C), 689(D), 715(C), 723(D), 745(D), 755(C), 783(D), 810(C), 825(C), 851(D), 876(D), 893(C), 936(C), 976(C), 1001(C),1030(C), 1060(C), 1087(D), 1125(D), 1155(C), 1171(D), 1206(C), 1234(D), 1254(C), 1307(D), 1337(D), 1396(D), 1448(C), 1484(D), 1554(C), 1578(D), 1603(C), 1616(D), 1657(C), 1684(D), 2852(C), 2873(C), 2896(C), 2934(C), 2942(D), 2967(C), 3016(C), 3065(C)	36%
1337	619(D), 641(C), 673(C), 689(D), 715(C), 723(D), 745(D), 755(D), 783(C), 810(C), 825(D), 851(D), 876(D), 893(C), 936(C), 976(C), 1001(C),1030(C), 1060(C), 1087(C), 1125(D), 1155(C), 1171(D), 1206(C), 1234(D), 1254(C), 1307(C), 1337(C), 1396(D), 1448(C), 1484(D), 1554(C), 1578(D), 1603(C), 1616(C), 1657(C), 1684(D), 2852(C), 2873(C), 2896(C), 2934(C), 2942(D), 2967(C), 3016(C), 3065(C)	34%
1396	619(D), 641(D), 673(D), 689(C), 715(D), 723(D), 745(D), 755(C), 783(D), 810(C), 825(D), 851(D), 876(D), 893(D), 936(D), 976(C), 1001(D),1030(D), 1060(D), 1087(D), 1125(D), 1155(D), 1171(D), 1206(D), 1234(D), 1254(D), 1307(D), 1337(D), 1396(C), 1448(D), 1484(D), 1554(D), 1578(D), 1603(D), 1616(D), 1657(D), 1684(D), 2852(D), 2873(D), 2896(D), 2934(D), 2942(D), 2967(D), 3016(D), 3065(D)	91%
1448	619(D), 641(C), 673(C), 689(D), 715(C), 723(C), 745(D), 755(D), 783(D), 810(C), 825(D), 851(D), 876(D), 893(C), 936(C), 976(C), 1001(C),1030(C), 1060(C), 1087(D), 1125(D), 1155(C), 1171(D), 1206(C), 1234(D), 1254(D), 1307(C), 1337(C), 1396(C), 1484(D), 1554(C), 1578(C), 1603(C), 1616(D), 1657(C), 1684(D), 2852(C), 2873(C), 2896(C), 2934(D), 2942(C), 2967(C), 3016(C), 3065(C)	39%
1484	619(D), 641(D), 673(D), 689(C), 715(D), 723(C), 745(D), 755(C), 783(D), 810(C), 825(C), 851(D), 876(D), 893(D), 936(D), 976(C), 1001(D),1030(C), 1060(D), 1087(D), 1125(C), 1155(C), 1171(C), 1206(C), 1234(D), 1254(D), 1307(D), 1337(D), 1396(C), 1448(D), 1554(C), 1578(C), 1603(D), 1616(D), 1657(D), 1684(D), 2852(C), 2873(C), 2896(C), 2934(C), 2942(C), 2967(C), 3016(C), 3065(C)	43%
1554	619(D), 641(C), 673(C), 689(D), 715(C), 723(C), 745(D), 755(D), 783(C), 810(C), 825(C), 851(D), 876(D), 893(C), 936(C), 976(C), 1001(C),1030(C), 1060(C), 1087(C), 1125(C), 1155(C), 1171(D), 1206(C), 1234(D), 1254(C), 1307(C), 1337(C), 1396(D), 1448(D), 1484(D), 1578(D), 1603(C), 1616(C), 1657(C), 1684(C), 2852(C), 2873(C), 2896(C), 2934(C), 2942(C), 2967(C), 3016(C), 3065(C)	27%
1578	619(D), 641(D), 673(D), 689(C), 715(D), 723(C), 745(D), 755(C), 783(D), 810(C), 825(C), 851(D), 876(D), 893(D), 936(D), 976(D), 1001(D),1030(D), 1060(D), 1087(D), 1125(C), 1155(D), 1171(C), 1206(D), 1234(D), 1254(D), 1307(D), 1337(D), 1396(C), 1448(D), 1484(D), 1554(C), 1603(D), 1616(D), 1657(D), 1684(D), 2852(C), 2873(D), 2896(D), 2934(C), 2942(D), 2967(D), 3016(D), 3065(D)	84%
1603	619(D), 641(C), 673(C), 689(D), 715(C), 723(D), 745(D), 755(D), 783(C), 810(C), 825(D), 851(D), 876(D), 893(C), 936(C), 976(C), 1001(C),1030(C), 1060(C), 1087(C), 1125(D), 1155(C), 1171(D), 1206(C), 1234(D), 1254(D), 1307(C), 1337(C), 1396(D), 1448(C), 1484(D), 1554(C), 1578(D), 1616(D), 1657(D), 1684(D), 2852(C), 2873(C), 2896(C), 2934(C), 2942(C), 2967(C), 3016(C), 3065(C)	66%
1657	619(D), 641(C), 673(C), 689(D), 715(C), 723(C), 745(D), 755(D), 783(C), 810(C), 825(D), 851(D), 876(D), 893(C), 936(C), 976(C), 1001(C),1030(C), 1060(C), 1087(C), 1125(D), 1155(C), 1171(D), 1206(C), 1234(D), 1254(D), 1307(C), 1337(C), 1396(D), 1448(C), 1484(D), 1554(C), 1578(D), 1603(C), 1616(D), 1684(D), 2852(C), 2873(C), 2896(C), 2934(C), 2942(C), 2967(C), 3016(C), 3065(C)	57%
1684	619(D), 641(C), 673(D), 689(D), 715(C), 723(D), 745(D), 755(C), 783(C), 810(C), 825(D), 851(C), 876(D), 893(D), 936(C), 976(C), 1001(D),1030(D), 1060(D), 1087(D), 1125(D), 1155(C), 1171(D), 1206(C), 1234(D), 1254(D), 1307(D), 1337(D), 1396(D), 1448(C), 1484(C), 1554(C), 1578(D), 1603(C), 1616(D), 1657(C), 2852(C), 2873(D), 2896(D), 2934(D), 2942(D), 2967(D), 3016(D), 3065(D)	39%
2852	619(D), 641(D), 673(C), 689(D), 715(C), 723(C), 745(D), 755(C), 783(D), 810(C), 825(C), 851(D), 876(D), 893(C), 936(C), 976(C), 1001(C),1030(C), 1060(C), 1087(C), 1125(D), 1155(C), 1171(D), 1206(C), 1234(D), 1254(C), 1307(C), 1337(C), 1396(D), 1448(C), 1484(C), 1554(C), 1578(D), 1603(C), 1616(C), 1657(C), 1684(D), 2873(C), 2896(C), 2934(C), 2942(C), 2967(C), 3016(C), 3065(C)	39%
2873	619(D), 641(D), 673(D), 689(D), 715(D), 723(D), 745(D), 755(C), 783(D), 810(C), 825(D), 851(D), 876(D), 893(D), 936(D), 976(C), 1001(D),1030(D), 1060(D), 1087(D), 1125(C), 1155(C), 1171(D), 1206(C), 1234(D), 1254(D), 1307(D), 1337(D), 1396(D), 1448(D), 1484(C), 1554(C), 1578(D), 1603(D), 1616(D), 1657(D), 1684(D), 2852(C), 2896(C), 2934(D), 2942(D), 2967(D), 3016(D), 3065(D)	39%
2896	619(D), 641(D), 673(C), 689(D), 715(C), 723(C), 745(D), 755(C), 783(D), 810(C), 825(C), 851(D), 876(D), 893(C), 936(C), 976(C), 1001(C),1030(C), 1060(C), 1087(D), 1125(C), 1155(C), 1171(D), 1206(C), 1234(D), 1254(C), 1307(C), 1337(C), 1396(D), 1448(C), 1484(C), 1554(C), 1578(D), 1603(C), 1616(D), 1657(C), 1684(D), 2852(C), 2873(C), 2934(C), 2942(C), 2967(C), 3016(C), 3065(C)	39%
2934	619(D), 641(D), 673(C), 689(D), 715(C), 723(C), 745(D), 755(C), 783(D), 810(C), 825(C), 851(D), 876(D), 893(C), 936(D), 976(C), 1001(D),1030(C), 1060(C), 1087(D), 1125(C), 1155(C), 1171(D), 1206(C), 1234(D), 1254(D), 1307(D), 1337(C), 1396(D), 1448(D), 1484(C), 1554(C), 1578(D), 1603(C), 1616(D), 1657(C), 1684(D), 2852(C), 2873(C), 2896(C), 2942(D), 2967(D), 3016(D), 3065(D)	39%
2942	619(D), 641(D), 673(D), 689(D), 715(C), 723(D), 745(D), 755(C), 783(D), 810(C), 825(C), 851(D), 876(D), 893(C), 936(D), 976(C), 1001(D),1030(C), 1060(C), 1087(D), 1125(C), 1155(C), 1171(D), 1206(C), 1234(D), 1254(D), 1307(D), 1337(C), 1396(D), 1448(D), 1484(C), 1554(C), 1578(D), 1603(C), 1616(D), 1657(C), 1684(D), 2852(C), 2873(C), 2896(C), 2934(D), 2967(D), 3016(D), 3065(D)	57%
2967	619(D), 641(D), 673(C), 689(D), 715(C), 723(C), 745(D), 755(C), 783(D), 810(C), 825(C), 851(D), 876(D), 893(C), 936(C), 976(C), 1001(C),1030(C), 1060(C), 1087(D), 1125(C), 1155(C), 1171(D), 1206(C), 1234(D), 1254(D), 1307(D), 1337(C), 1396(D), 1448(D), 1484(C), 1554(C), 1578(D), 1603(C), 1616(D), 1657(C), 1684(D), 2852(C), 2873(C), 2896(C), 2934(D), 2942(D), 3016(D), 3065(D)	39%
3016	619(D), 641(D), 673(C), 689(C), 715(C), 723(C), 745(D), 755(C), 783(D), 810(C), 825(C), 851(D), 876(D), 893(C), 936(C), 976(C), 1001(D),1030(C), 1060(D), 1087(D), 1125(C), 1155(C), 1171(D), 1206(C), 1234(D), 1254(D), 1307(D), 1337(C), 1396(D), 1448(D), 1484(C), 1554(C), 1578(D), 1603(C), 1616(D), 1657(C), 1684(D), 2852(C), 2873(C), 2896(C), 2934(C), 2942(D), 2967(C), 3065(C)	39%
3065	619(D), 641(D), 673(C), 689(D), 715(C), 723(C), 745(D), 755(C), 783(D), 810(C), 825(C), 851(D), 876(D), 893(C), 936(C), 976(C), 1001(D),1030(C), 1060(D), 1087(D), 1125(C), 1155(C), 1171(D), 1206(C), 1234(D), 1254(D), 1307(D), 1337(C), 1396(D), 1448(D), 1484(C), 1554(C), 1578(D), 1603(C), 1616(D), 1657(C), 1684(D), 2852(C), 2873(C), 2896(C), 2934(C), 2942(D), 2967(C), 3016(C)	47%

Table C.6: Peak ratio analysis (PIR) of the comparison between PNT2-C2 live and P4E6. The letter D specifies PIR separation between PNT2-C2 and P4E6. The letter C specifies that the PIR marker is overlapping between samples. The letter X specifies that the PIR comparison could not take place as the band was not present in P4E6.

Panel	PIR Matches	% matchD
619	641(D), 673(D), 698(D), 715(D), 723(D), 745(D), 755(D), 783(D), 810(C), 825(D), 851(D), 876(D), 893(C), 936(D), 976(X), 1001(D), 1030(D), 1060(D), 1087(D), 1125(D), 1155(D), 1171(D), 1206(D), 1234(D), 1307(D), 1396(D), 1448(D), 1484(D), 1554(D), 1578(D), 1603(D), 1616(C), 1657(D), 1684(C), 2852(D), 2873(C), 2898(D), 2934(D), 2942(D), 2967(D), 3016(D), 3065(D)	88%
641	619(D), 673(D), 698(D), 715(D), 723(D), 745(D), 755(C), 783(C), 810(D), 825(C), 851(D), 876(D), 893(C), 936(C), 976(X), 1001(C), 1030(D), 1060(C), 1087(C), 1125(D), 1155(D), 1171(C), 1206(C), 1234(C), 1254(D), 1307(D), 1396(D), 1448(D), 1484(C), 1554(C), 1578(C), 1603(D), 1616(C), 1657(D), 1684(C), 2852(D), 2873(C), 2898(D), 2934(D), 2942(C), 2967(D), 3016(D), 3065(D)	80%
673	619(C), 641(D), 698(C), 715(D), 723(D), 745(D), 755(D), 783(D), 810(C), 825(C), 851(D), 876(D), 893(D), 936(D), 976(X), 1001(D), 1030(D), 1060(D), 1087(D), 1125(D), 1155(D), 1171(D), 1206(D), 1234(D), 1254(D), 1307(D), 1396(D), 1448(D), 1484(D), 1554(D), 1578(D), 1603(C), 1616(C), 1657(D), 1684(C), 2852(D), 2873(D), 2898(D), 2934(D), 2942(D), 2967(D), 3016(D), 3065(D)	79%
698	619(D), 641(D), 673(D), 715(D), 723(D), 745(D), 755(D), 783(D), 810(D), 825(D), 851(D), 876(D), 893(D), 936(D), 976(X), 1001(D), 1030(D), 1060(D), 1087(D), 1125(D), 1155(D), 1171(D), 1206(D), 1234(D), 1254(D), 1307(D), 1396(D), 1448(D), 1484(D), 1554(D), 1578(D), 1603(D), 1616(C), 1657(D), 1684(C), 2852(D), 2873(D), 2898(D), 2934(D), 2942(D), 2967(D), 3016(D), 3065(D)	88%
715	619(D), 641(D), 673(D), 698(D), 723(D), 745(D), 755(D), 783(D), 810(D), 825(D), 851(D), 876(C), 893(D), 936(D), 976(X), 1001(D), 1030(D), 1060(D), 1087(D), 1125(D), 1155(D), 1171(D), 1206(D), 1234(D), 1254(D), 1307(D), 1396(D), 1448(D), 1484(D), 1554(D), 1578(D), 1603(D), 1616(D), 1657(D), 1684(D), 2852(C), 2873(D), 2898(D), 2934(D), 2942(D), 2967(D), 3016(D), 3065(D)	91%
723	619(C), 641(C), 673(C), 698(D), 715(D), 745(D), 755(C), 783(C), 810(C), 825(C), 851(C), 876(D), 893(C), 936(C), 976(X), 1001(C), 1030(C), 1060(C), 1087(C), 1125(D), 1155(C), 1171(C), 1206(C), 1234(C), 1254(D), 1307(D), 1396(C), 1448(C), 1484(C), 1554(C), 1578(C), 1603(D), 1616(C), 1657(C), 1684(C), 2852(D), 2873(C), 2898(C), 2934(D), 2942(C), 2967(C), 3016(D), 3065(D)	40%
745	619(D), 641(C), 673(C), 698(D), 715(D), 723(D), 755(C), 783(C), 810(C), 825(C), 851(C), 876(D), 893(C), 936(C), 976(X), 1001(C), 1030(C), 1060(C), 1087(C), 1125(D), 1155(C), 1171(C), 1206(C), 1234(C), 1254(D), 1307(D), 1396(C), 1448(C), 1484(C), 1554(C), 1578(C), 1603(D), 1616(C), 1657(C), 1684(C), 2852(D), 2873(C), 2898(C), 2934(D), 2942(C), 2967(C), 3016(D), 3065(D)	63%
755	619(D), 641(C), 673(C), 698(D), 715(D), 723(D), 745(C), 783(C), 810(C), 825(C), 851(C), 876(D), 893(C), 936(C), 976(X), 1001(C), 1030(C), 1060(C), 1087(C), 1125(D), 1155(C), 1171(C), 1206(C), 1234(C), 1254(D), 1307(D), 1396(C), 1448(C), 1484(C), 1554(C), 1578(C), 1603(D), 1616(C), 1657(C), 1684(C), 2852(D), 2873(C), 2898(C), 2934(D), 2942(C), 2967(C), 3016(D), 3065(D)	63%
783	619(D), 641(C), 673(D), 698(D), 715(D), 723(C), 745(D), 755(C), 810(D), 825(D), 851(D), 876(D), 893(D), 936(C), 976(X), 1001(C), 1030(D), 1060(C), 1087(C), 1125(D), 1155(D), 1171(C), 1206(D), 1234(D), 1254(D), 1307(D), 1396(D), 1448(D), 1484(C), 1554(C), 1578(C), 1603(D), 1616(D), 1657(D), 1684(C), 2852(D), 2873(C), 2898(C), 2934(D), 2942(C), 2967(D), 3016(D), 3065(D)	63%
810	619(C), 641(D), 673(C), 698(C), 715(D), 723(C), 745(D), 755(D), 783(D), 825(D), 851(D), 876(D), 893(C), 936(D), 976(X), 1001(D), 1030(D), 1060(D), 1087(D), 1125(D), 1155(C), 1171(D), 1206(D), 1234(D), 1254(D), 1307(D), 1396(D), 1448(D), 1484(C), 1554(D), 1578(D), 1603(D), 1616(C), 1657(D), 1684(C), 2852(D), 2873(D), 2898(D), 2934(C), 2942(D), 2967(C), 3016(D), 3065(D)	74%
825	619(D), 641(D), 673(D), 698(D), 715(D), 723(C), 745(D), 755(D), 783(C), 810(C), 851(C), 876(D), 893(C), 936(C), 976(X), 1001(C), 1030(C), 1060(C), 1087(D), 1125(D), 1155(C), 1171(D), 1206(D), 1234(D), 1254(D), 1307(D), 1396(D), 1448(D), 1484(C), 1554(D), 1578(D), 1603(D), 1616(D), 1657(D), 1684(C), 2852(D), 2873(D), 2898(D), 2934(D), 2942(C), 2967(C), 3016(D), 3065(D)	67%
851	619(D), 641(D), 673(D), 698(D), 715(D), 723(C), 745(D), 755(D), 783(D), 810(D), 825(C), 876(D), 893(D), 936(C), 976(X), 1001(D), 1030(D), 1060(C), 1087(D), 1125(D), 1155(C), 1171(D), 1206(C), 1234(D), 1254(D), 1307(D), 1396(D), 1448(D), 1484(C), 1554(D), 1578(D), 1603(D), 1616(D), 1657(D), 1684(D), 2852(D), 2873(D), 2898(C), 2934(D), 2942(C), 2967(D), 3016(D), 3065(D)	72%
876	619(D), 641(D), 673(D), 698(D), 715(D), 723(C), 745(C), 755(D), 783(D), 810(D), 825(D), 851(D), 893(D), 936(D), 976(X), 1001(D), 1030(D), 1060(D), 1087(D), 1125(D), 1155(D), 1171(D), 1206(D), 1234(D), 1254(D), 1307(D), 1396(C), 1448(D), 1484(D), 1554(D), 1578(D), 1603(D), 1616(D), 1657(D), 1684(D), 2852(C), 2873(D), 2898(C), 2934(D), 2942(D), 2967(C), 3016(D), 3065(D)	88%
893	619(C), 641(D), 673(C), 698(D), 715(D), 723(C), 745(D), 755(D), 783(D), 810(D), 825(D), 851(D), 876(D), 936(C), 976(X), 1001(D), 1030(D), 1060(D), 1087(D), 1125(D), 1155(C), 1171(D), 1206(D), 1234(D), 1254(D), 1307(D), 1396(D), 1448(D), 1484(C), 1554(D), 1578(D), 1603(D), 1616(D), 1657(D), 1684(D), 2852(D), 2873(D), 2898(C), 2934(D), 2942(D), 2967(C), 3016(D), 3065(C)	63%
936	619(D), 641(C), 673(D), 698(D), 715(D), 723(C), 745(D), 755(C), 783(C), 810(D), 825(D), 851(D), 876(D), 893(C), 976(X), 1001(C), 1030(D), 1060(C), 1087(C), 1125(D), 1155(C), 1171(C), 1206(C), 1234(C), 1254(D), 1307(D), 1396(C), 1448(C), 1484(C), 1554(C), 1578(C), 1603(D), 1616(C), 1657(D), 1684(C), 2852(D), 2873(C), 2898(C), 2934(D), 2942(C), 2967(C), 3016(D), 3065(D)	35%
976	619(D), 641(C), 673(D), 698(D), 715(D), 723(C), 745(D), 755(C), 783(C), 810(D), 825(C), 851(C), 876(D), 893(D), 936(C), 1001(C), 1030(C), 1060(C), 1087(C), 1125(D), 1155(C), 1171(C), 1206(C), 1234(C), 1254(D), 1307(D), 1396(C), 1448(C), 1484(C), 1554(C), 1578(C), 1603(D), 1616(D), 1657(D), 1684(D), 2852(D), 2873(C), 2898(C), 2934(D), 2942(C), 2967(C), 3016(D), 3065(D)	81%
1001	619(D), 641(C), 673(D), 698(D), 715(D), 723(C), 745(D), 755(C), 783(C), 810(D), 825(C), 851(C), 876(D), 893(D), 936(C), 976(X), 1030(C), 1060(C), 1087(C), 1125(D), 1155(C), 1171(D), 1206(C), 1234(D), 1254(D), 1307(D), 1396(D), 1448(D), 1484(C), 1554(D), 1578(C), 1603(D), 1616(D), 1657(D), 1684(D), 2852(D), 2873(D), 2898(C), 2934(D), 2942(C), 2967(D), 3016(D), 3065(D)	60%
1030	619(D), 641(D), 673(D), 698(D), 715(D), 723(C), 745(D), 755(C), 783(C), 810(D), 825(C), 851(C), 876(D), 893(D), 936(C), 976(X), 1001(C), 1060(C), 1087(D), 1125(D), 1155(C), 1171(D), 1206(C), 1234(D), 1254(D), 1307(D), 1396(D), 1448(D), 1484(C), 1554(D), 1578(C), 1603(D), 1616(D), 1657(D), 1684(D), 2852(D), 2873(D), 2898(C), 2934(D), 2942(C), 2967(C), 3016(D), 3065(D)	58%
1060	619(D), 641(C), 673(D), 698(D), 715(D), 723(C), 745(D), 755(C), 783(C), 810(C), 825(C), 851(C), 876(D), 893(D), 936(C), 976(X), 1001(C), 1030(D), 1060(C), 1087(C), 1125(D), 1155(C), 1171(C), 1206(C), 1234(C), 1254(C), 1307(C), 1337(D), 1396(D), 1448(C), 1484(C), 1554(C), 1578(C), 1603(D), 1616(D), 1657(C), 1684(D), 2852(D), 2873(C), 2898(C), 2934(D), 2942(C), 2967(D), 3016(D), 3065(D)	42%
1087	619(D), 641(C), 673(D), 698(D), 715(D), 723(C), 745(D), 755(C), 783(C), 810(C), 825(D), 851(D), 876(D), 893(D), 936(C), 976(X), 1001(C), 1030(D), 1060(C), 1125(C), 1155(D), 1171(D), 1206(D), 1234(D), 1254(C), 1307(C), 1337(D), 1396(D), 1448(C), 1484(C), 1554(C), 1578(C), 1603(D), 1616(D), 1657(C), 1684(D), 2852(D), 2873(C), 2898(C), 2934(D), 2942(C), 2967(D), 3016(D), 3065(D)	58%
1125	619(D), 641(D), 673(D), 698(D), 715(D), 723(D), 745(D), 755(D), 783(C), 810(C), 825(D), 851(D), 876(D), 893(D), 936(D), 976(X), 1001(D), 1030(D), 1060(D), 1087(C), 1155(D), 1171(D), 1206(D), 1234(C), 1254(C), 1307(C), 1337(D), 1396(D), 1448(C), 1484(C), 1554(C), 1578(C), 1603(D), 1616(D), 1657(C), 1684(D), 2852(D), 2873(C), 2898(C), 2934(D), 2942(C), 2967(D), 3016(D), 3065(D)	74%
1155	619(D), 641(D), 673(D), 698(D), 715(D), 723(D), 745(D), 755(D), 783(C), 810(C), 825(D), 851(D), 876(D), 893(D), 936(C), 976(X), 1001(D), 1030(D), 1060(C), 1087(D), 1125(D), 1171(D), 1206(C), 1234(D), 1254(D), 1307(C), 1337(D), 1396(D), 1448(C), 1484(C), 1554(D), 1578(D), 1603(D), 1616(D), 1657(D), 1684(D), 2852(D), 2873(D), 2898(C), 2934(D), 2942(C), 2967(C), 3016(D), 3065(D)	60%
1171	619(D), 641(C), 673(D), 698(D), 715(D), 723(C), 745(D), 755(C), 783(C), 810(C), 825(C), 851(C), 876(D), 893(C), 936(C), 976(X), 1001(C), 1030(C), 1060(C), 1087(D), 1125(D), 1155(C), 1206(C), 1234(C), 1254(D), 1307(D), 1337(D), 1396(C), 1448(C), 1484(C), 1554(C), 1578(C), 1603(D), 1616(C), 1657(D), 1684(C), 2852(D), 2873(C), 2898(C), 2934(D), 2942(C), 2967(C), 3016(D), 3065(D)	67%
1206	619(D), 641(C), 673(D), 698(D), 715(D), 723(C), 745(D), 755(C), 783(C), 810(C), 825(C), 851(C), 876(D), 893(D), 936(C), 976(X), 1001(C), 1030(C), 1060(C), 1087(C), 1125(D), 1155(C), 1171(C), 1234(C), 1254(D), 1307(D), 1337(D), 1396(C), 1448(C), 1484(C), 1554(C), 1578(C), 1603(D), 1616(C), 1657(D), 1684(C), 2852(D), 2873(C), 2898(C), 2934(D), 2942(C), 2967(C), 3016(D), 3065(D)	67%
1234	619(D), 641(D), 673(D), 698(D), 715(D), 723(C), 745(D), 755(C), 783(C), 810(D), 825(D), 851(D), 876(D), 893(D), 936(C), 976(X), 1001(D), 1030(D), 1060(D), 1087(D), 1125(D), 1155(D), 1171(D), 1206(D), 1254(D), 1307(D), 1337(D), 1396(D), 1448(C), 1484(C), 1554(D), 1578(D), 1603(D), 1616(D), 1657(D), 1684(D), 2852(D), 2873(D), 2898(C), 2934(D), 2942(C), 2967(D), 3016(D), 3065(D)	81%
1254	619(D), 641(D), 673(D), 698(D), 715(D), 723(D), 745(D), 755(D), 783(D), 810(D), 825(D), 851(D), 876(D), 893(D), 936(C), 976(X), 1001(D), 1030(D), 1060(C), 1087(C), 1125(C), 1155(D), 1171(D), 1206(D), 1234(C), 1307(D), 1337(D), 1396(D), 1448(C), 1484(C), 1554(C), 1578(C), 1603(D), 1616(D), 1657(C), 1684(D), 2852(D), 2873(D), 2898(C), 2934(D), 2942(D), 2967(C), 3016(D), 3065(D)	72%
1307	619(C), 641(C), 673(D), 698(D), 715(D), 723(C), 745(D), 755(C), 783(C), 810(C), 825(C), 851(C), 876(D), 893(C), 936(C), 976(X), 1001(C), 1030(C), 1060(C), 1087(C), 1125(D), 1155(C), 1171(C), 1206(C), 1234(C), 1254(D), 1307(D), 1337(D), 1396(D), 1448(C), 1484(C), 1554(C), 1578(C), 1603(D), 1616(D), 1657(D), 1684(D), 2852(C), 2873(D), 2898(C), 2934(D), 2942(C), 2967(C), 3016(D), 3065(D)	49%
1337	619(C), 641(D), 673(C), 698(D), 715(D), 723(D), 745(D), 755(D), 783(D), 810(C), 825(C), 851(D), 876(D), 893(D), 936(C), 976(X), 1001(D), 1030(D), 1060(D), 1087(D), 1125(D), 1155(D), 1171(D), 1206(D), 1234(D), 1254(D), 1307(D), 1337(D), 1396(D), 1448(D), 1484(D), 1554(D), 1578(D), 1603(D), 1616(D), 1657(D), 1684(D), 2852(D), 2873(D), 2898(D), 2934(D), 2942(C), 2967(D), 3016(D), 3065(C)	79%
1396	619(D), 641(D), 673(D), 698(D), 715(C), 723(D), 745(C), 755(D), 783(D), 810(D), 825(D), 851(D), 876(D), 893(D), 936(D), 976(X), 1001(D), 1030(D), 1060(D), 1087(D), 1125(C), 1155(D), 1171(D), 1206(D), 1234(C), 1254(C), 1307(D), 1337(D), 1396(D), 1448(C), 1484(D), 1554(D), 1578(D), 1603(D), 1616(D), 1657(D), 1684(D), 2852(D), 2873(D), 2898(D), 2934(D), 2942(C), 2967(D), 3016(D), 3065(D)	74%
1448	619(D), 641(D), 673(D), 698(D), 715(D), 723(D), 745(D), 755(D), 783(D), 810(D), 825(D), 851(D), 876(D), 893(D), 936(C), 976(X), 1001(D), 1030(D), 1060(D), 1087(C), 1125(C), 1155(D), 1171(C), 1206(D), 1234(C), 1254(C), 1307(D), 1337(D), 1396(C), 1448(D), 1484(D), 1554(D), 1578(D), 1603(D), 1616(D), 1657(C), 1684(D), 2852(D), 2873(C), 2898(C), 2934(D), 2942(D), 2967(D), 3016(D), 3065(D)	72%
1484	619(D), 641(C), 673(D), 698(D), 715(D), 723(C), 745(D), 755(C), 783(C), 810(C), 825(C), 851(C), 876(D), 893(D), 936(C), 976(X), 1001(C), 1030(C), 1060(C), 1087(D), 1125(D), 1155(C), 1171(C), 1206(D), 1234(D), 1254(C), 1307(C), 1337(D), 1396(D), 1448(D), 1484(D), 1554(C), 1578(C), 1603(D), 1616(D), 1657(C), 1684(C), 2852(D), 2873(C), 2898(C), 2934(D), 2942(C), 2967(C), 3016(D), 3065(D)	37%
1554	619(D), 641(C), 673(D), 698(D), 715(C), 723(D), 745(D), 755(C), 783(C), 810(D), 825(C), 851(C), 876(D), 893(D), 936(C), 976(X), 1001(D), 1030(D), 1060(C), 1087(C), 1125(C), 1155(C), 1171(C), 1206(D), 1234(D), 1254(D), 1307(D), 1337(D), 1396(D), 1448(C), 1484(C), 1554(D), 1578(C), 1603(D), 1616(D), 1657(D), 1684(D), 2852(C), 2873(C), 2898(C), 2934(D), 2942(D), 2967(D), 3016(D), 3065(D)	53%
1578	619(D), 641(C), 673(D), 698(D), 715(D), 723(C), 745(D), 755(C), 783(C), 810(C), 825(D), 851(D), 876(D), 893(D), 936(C), 976(X), 1001(C), 1030(D), 1060(D), 1087(D), 1125(D), 1155(D), 1171(D), 1206(D), 1234(D), 1254(D), 1307(D), 1337(D), 1396(D), 1448(D), 1484(D), 1554(D), 1578(D), 1603(C), 1616(C), 1657(D), 1684(C), 2852(D), 2873(D), 2898(D), 2934(D), 2942(D), 2967(D), 3016(D), 3065(D)	56%
1603	619(D), 641(D), 673(D), 698(C), 715(D), 723(D), 745(D), 755(D), 783(D), 810(D), 825(D), 851(D), 876(D), 893(D), 936(D), 976(X), 1001(D), 1030(D), 1060(D), 1087(D), 1125(D), 1155(D), 1171(D), 1206(D), 1234(D), 1254(D), 1307(D), 1337(D), 1396(D), 1448(D), 1484(D), 1554(D), 1578(D), 1603(D), 1616(C), 1657(D), 1684(D), 2852(D), 2873(D), 2898(D), 2934(D), 2942(D), 2967(D), 3016(D), 3065(D)	84%
1616	619(C), 641(D), 673(C), 698(C), 715(D), 723(C), 745(D), 755(C), 783(C), 810(D), 825(C), 851(C), 876(D), 893(C), 936(C), 976(X), 1001(C), 1030(C), 1060(C), 1087(C), 1125(D), 1155(C), 1171(C), 1206(C), 1234(C), 1254(C), 1307(C), 1337(D), 1396(C), 1448(C), 1484(C), 1554(C), 1578(C), 1603(D), 1616(C), 1657(C), 1684(C), 2852(D), 2873(C), 2898(C), 2934(D), 2942(C), 2967(C), 3016(D), 3065(D)	56%
1657	619(D), 641(D), 673(D), 698(D), 715(D), 723(D), 745(D), 755(D), 783(D), 810(D), 825(D), 851(D), 876(D), 893(D), 936(D), 976(X), 1001(D), 1030(D), 1060(D), 1087(D), 1125(D), 1155(D), 1171(D), 1206(D), 1234(D), 1254(D), 1307(D), 1337(D), 1396(D), 1448(D), 1484(D), 1554(D), 1578(D), 1603(D), 1616(D), 1657(C), 1684(D), 2852(D), 2873(D), 2898(D), 2934(D), 2942(D), 2967(D), 3016(D), 3065(D)	65%
1684	619(C), 641(C), 673(C), 698(C), 715(C), 723(C), 745(C), 755(C), 783(C), 810(C), 825(C), 851(C), 876(D), 893(C), 936(C), 976(X), 1001(C), 1030(C), 1060(C), 1087(C), 1125(C), 1155(C), 1171(C), 1206(C), 1234(C), 1254(C), 1307(C), 1337(C), 1396(C), 1448(C), 1484(C), 1554(C), 1578(C), 1603(C), 1616(C), 1657(C), 1684(C), 2852(C), 2873(C), 2898(C), 2934(C), 2942(C), 2967(C), 3016(C), 3065(C)	84%
2852	619(D), 641(C), 673(D), 698(D), 715(D), 723(C), 745(D), 755(C), 783(C), 810(D), 825(D), 851(D), 876(D), 893(D), 936(C), 976(X), 1001(D), 1030(D), 1060(C), 1087(C), 1125(C), 1155(D), 1171(C), 1206(D), 1234(C), 1254(C), 1307(D), 1337(D), 1396(C), 1448(C), 1484(C), 1554(C), 1578(C), 1603(D), 1616(D), 1657(C), 1684(D), 2852(D), 2873(C), 2898(C), 2934(D), 2942(C), 2967(C), 3016(D), 3065(D)	65%
2873	619(D), 641(C), 673(D), 698(D), 715(D), 723(C), 745(D), 755(C), 783(C), 810(D), 825(D), 851(D), 876(D), 893(D), 936(C), 976(X), 1001(D), 1030(D), 1060(C), 1087(C), 1125(C), 1155(D), 1171(C), 1206(D), 1234(C), 1254(C), 1307(D), 1337(D), 1396(C), 1448(C), 1484(C), 1554(C), 1578(C), 1603(D), 1616(D), 1657(C), 1684(D), 2852(D), 2873(C), 2898(C), 2934(D), 2942(C), 2967(C), 3016(D), 3065(D)	63%
2898	619(D), 641(C), 673(D), 698(D), 715(D), 723(C), 745(D), 755(C), 783(C), 810(D), 825(D), 851(D), 876(D), 893(D), 936(C), 976(X), 1001(C), 1030(D), 1060(C), 1087(C), 1125(D), 1155(C), 1171(C), 1206(C), 1234(D), 1254(D), 1307(D), 1337(D), 1396(D), 1448(D), 1484(C), 1554(D), 1578(C), 1603(D), 1616(D), 1657(C), 1684(C), 2852(D), 2873(C), 2898(C), 2934(D), 2942(C), 2967(C), 3016(D), 3065(D)	67%
2934	619(D), 641(C), 673(D), 698(D), 715(D), 723(C), 745(D), 755(C), 783(C), 810(D), 825(D), 851(D), 876(D), 893(D), 936(C), 976(X), 1001(C), 1030(D), 1060(C), 1087(C), 1125(D), 1155(C), 1171(C), 1206(C), 1234(D), 1254(D), 1307(D), 1337(D), 1396(D), 1448(D), 1484(C), 1554(D), 1578(C), 1603(D), 1616(D), 1657(D), 1684(C), 2852(D), 2873(D), 2898(C), 2934(C), 2942(C), 2967(C), 3016(D), 3065(D)	58%
2942	619(D), 641(D), 673(D), 698(D), 715(D), 723(C), 745(D), 755(C), 783(C), 810(D), 825(D), 851(D), 876(D), 893(D), 936(C), 976(X), 1001(C), 1030(D), 1060(C), 1087(C), 1125(D), 1155(C), 1171(C), 1206(C), 1234(D), 1254(D), 1307(D), 1337(D), 1396(D), 1448(D), 1484(C), 1554(D), 1578(C), 1603(D), 1616(D), 1657(D), 1684(C), 2852(D), 2873(D), 2898(C), 2934(D), 2942(C), 2967(C), 3016(D), 3065(D)	74%
2967	619(D), 641(D), 673(D), 698(D), 715(D), 723(C), 745(D), 755(C), 783(C), 810(D), 825(D), 851(D), 876(D), 893(D), 936(C), 976(X), 1001(D), 1030(D), 1060(D), 1087(D), 1125(D), 1155(D), 1171(D), 1206(D), 1234(D), 1254(D), 1307(C), 1337(C), 1396(D), 1448(D), 1484(C), 1554(D), 1578(C), 1603(D), 1616(D), 1657(D), 1684(D), 2852(D), 2873(D), 2898(D), 2934(D), 2942(D), 2967(C), 3016(D), 3065(D)	93%
3065	619(D), 641(D), 673(C), 698(C), 715(D), 723(C), 745(D), 755(D), 783(D), 810(D), 825(D), 851(D), 876(D), 893(C), 936(D), 976(X), 1001(D), 1030(D), 1060(D), 1087(D), 1125(D), 1155(D), 1171(D), 1206(D), 1234(D), 1254(D), 1307(D), 1337(C), 1396(D), 1448(D), 1484(D), 1554(D), 1578(D), 1603(C), 1616(C), 1657(D), 1684(C), 2852(D), 2873(D), 2898(C), 2934(C), 2942(D), 2967(D), 3016(D)	77%

Table C.7: Peak ratio analysis (PIR) of the comparison between H646/17R and H646/17L. The letter D specifies PIR separation between H646/17R and H646/17L. The letter C specifies that the PIR marker is overlapping between samples. The letter X specifies that the PIR comparison could not take place as the band was not present in one of the samples.

Panel	PIR Matches	% matchD
619	641(D), 688(D), 718(D), 745(C), 755(D), 783(D), 808(C), 822(D), 856(D), 935(D), 1000(D), 1028(D), 1063(D), 1088(D), 1124(D), 1154(D), 1170(D), 1204(D), 1232(C), 1251(D), 1307(D), 1337(D), 1449(D), 1484(D), 1553(D), 1577(D), 1603(D), 1615(D), 1659(D), 1684(C), 2852(D), 2873(D), 2894(D), 2934(C), 2942(D), 2968(D), 3010(C), 3064(D)	84%
641	619(D), 688(D), 718(D), 745(D), 755(C), 783(C), 808(D), 822(C), 856(D), 935(D), 1000(C), 1028(C), 1063(D), 1088(C), 1124(D), 1154(D), 1170(D), 1204(C), 1232(D), 1251(C), 1307(C), 1337(C), 1449(C), 1484(D), 1553(D), 1577(C), 1603(C), 1615(C), 1659(C), 1684(D), 2852(C), 2873(C), 2894(C), 2934(C), 2942(C), 2968(C), 3010(D), 3064(C)	39%
668	619(D), 641(D), 718(C), 745(D), 755(C), 783(D), 808(D), 822(C), 856(D), 935(C), 1000(C), 1028(C), 1063(X), 1088(C), 1124(C), 1154(D), 1170(D), 1204(D), 1232(D), 1251(C), 1307(C), 1337(C), 1449(C), 1484(C), 1553(D), 1577(C), 1603(C), 1615(C), 1659(C), 1684(D), 2852(D), 2873(C), 2894(C), 2934(C), 2942(C), 2968(C), 3010(D), 3064(D)	39%
718	619(D), 641(D), 668(C), 745(C), 755(C), 783(D), 808(D), 822(C), 856(D), 935(C), 1000(C), 1028(C), 1063(C), 1088(D), 1124(D), 1154(D), 1170(D), 1204(D), 1232(D), 1251(C), 1307(C), 1337(C), 1449(C), 1484(C), 1553(D), 1577(D), 1603(C), 1615(C), 1659(C), 1684(D), 2852(C), 2873(D), 2894(D), 2934(D), 2942(D), 2968(C), 3010(D), 3064(D)	55%
745	619(C), 641(D), 668(D), 718(D), 755(D), 783(D), 808(C), 822(D), 856(D), 935(D), 1000(D), 1028(D), 1063(D), 1088(D), 1124(C), 1154(D), 1170(C), 1204(C), 1232(C), 1251(D), 1307(D), 1337(D), 1449(D), 1484(D), 1553(C), 1577(D), 1603(D), 1615(C), 1659(D), 1684(C), 2852(D), 2873(D), 2894(D), 2934(C), 2942(D), 2968(D), 3010(C), 3064(D)	71%
755	619(D), 641(C), 668(D), 718(C), 745(D), 783(C), 808(D), 822(D), 856(D), 935(D), 1000(C), 1028(C), 1063(D), 1088(D), 1124(C), 1154(D), 1170(D), 1204(C), 1232(D), 1251(D), 1307(D), 1337(C), 1449(C), 1484(C), 1553(D), 1577(C), 1603(C), 1615(C), 1659(C), 1684(D), 2852(C), 2873(C), 2894(C), 2934(C), 2942(C), 2968(C), 3010(D), 3064(C)	37%
783	619(D), 641(C), 668(C), 718(D), 745(D), 755(C), 808(D), 822(C), 856(C), 935(D), 1000(C), 1028(C), 1063(C), 1088(D), 1124(D), 1154(D), 1170(D), 1204(C), 1232(D), 1251(C), 1307(C), 1337(C), 1449(C), 1484(C), 1553(D), 1577(C), 1603(C), 1615(C), 1659(C), 1684(D), 2852(C), 2873(C), 2894(C), 2934(C), 2942(C), 2968(C), 3010(D), 3064(C)	37%
808	619(C), 641(D), 668(D), 718(D), 745(C), 755(D), 783(D), 822(C), 856(C), 935(D), 1000(D), 1028(D), 1063(D), 1088(D), 1124(C), 1154(D), 1170(C), 1204(C), 1232(C), 1251(D), 1307(D), 1337(D), 1449(D), 1484(D), 1553(C), 1577(C), 1603(D), 1615(C), 1659(C), 1684(C), 2852(C), 2873(C), 2894(D), 2934(C), 2942(D), 2968(D), 3010(C), 3064(D)	55%
822	619(D), 641(C), 668(C), 718(D), 745(D), 755(C), 783(C), 808(D), 856(C), 935(C), 1000(C), 1028(C), 1063(C), 1088(C), 1124(D), 1154(D), 1170(D), 1204(C), 1232(D), 1251(C), 1307(C), 1337(C), 1449(C), 1484(C), 1553(D), 1577(C), 1603(C), 1615(C), 1659(C), 1684(D), 2852(C), 2873(C), 2894(C), 2934(C), 2942(C), 2968(C), 3010(D), 3064(C)	29%
856	619(D), 641(C), 668(D), 718(D), 745(D), 755(C), 783(C), 808(D), 822(C), 935(D), 1000(D), 1028(C), 1063(D), 1088(D), 1124(D), 1154(D), 1170(C), 1204(C), 1232(D), 1251(C), 1307(C), 1337(C), 1449(C), 1484(C), 1553(D), 1577(C), 1603(C), 1615(C), 1659(C), 1684(C), 2852(C), 2873(C), 2894(C), 2934(C), 2942(C), 2968(C), 3010(C), 3064(D)	29%
935	619(D), 641(D), 668(C), 718(C), 745(D), 755(C), 783(D), 808(D), 822(C), 856(D), 1000(D), 1028(C), 1063(D), 1088(D), 1124(D), 1154(D), 1170(D), 1204(D), 1232(D), 1251(C), 1307(C), 1337(C), 1449(C), 1484(C), 1553(D), 1577(D), 1603(C), 1615(C), 1659(C), 1684(D), 2852(C), 2873(C), 2894(C), 2934(C), 2942(D), 2968(D), 3010(D), 3064(D)	48%
1000	619(D), 641(C), 668(C), 718(C), 745(D), 755(C), 783(C), 808(D), 822(C), 856(C), 935(D), 1028(C), 1063(D), 1088(D), 1124(D), 1154(D), 1170(D), 1204(C), 1232(D), 1251(C), 1307(C), 1337(C), 1449(C), 1484(D), 1553(C), 1577(C), 1603(C), 1615(C), 1659(C), 1684(D), 2852(C), 2873(C), 2894(C), 2934(C), 2942(C), 2968(C), 3010(D), 3064(C)	32%
1029	619(D), 641(C), 668(C), 718(C), 745(D), 755(C), 783(D), 808(C), 822(C), 856(D), 935(C), 1000(C), 1063(C), 1088(C), 1124(D), 1154(D), 1170(D), 1204(D), 1232(D), 1251(C), 1307(C), 1337(C), 1449(C), 1484(C), 1553(D), 1577(C), 1603(C), 1615(C), 1659(C), 1684(D), 2852(D), 2873(C), 2894(C), 2934(C), 2942(C), 2968(D), 3010(D), 3064(C)	42%
1063	619(D), 641(D), 668(C), 718(C), 745(D), 755(D), 783(D), 808(D), 822(C), 856(D), 935(C), 1000(D), 1028(C), 1088(D), 1124(D), 1154(D), 1170(D), 1204(D), 1232(D), 1251(C), 1307(D), 1337(C), 1449(C), 1484(C), 1553(D), 1577(D), 1603(C), 1615(D), 1659(C), 1684(D), 2852(C), 2873(D), 2894(D), 2934(D), 2942(D), 2968(D), 3010(D), 3064(D)	68%
1088	619(D), 641(C), 668(C), 718(D), 744(D), 755(C), 783(C), 808(C), 822(C), 856(C), 935(D), 1000(D), 1028(C), 1063(D), 1124(D), 1154(D), 1170(D), 1204(C), 1232(C), 1251(D), 1307(D), 1337(C), 1449(C), 1484(D), 1553(C), 1577(D), 1603(C), 1615(C), 1659(C), 1684(D), 2852(C), 2873(C), 2894(C), 2934(C), 2942(C), 2968(C), 3010(C), 3064(C)	29%
1124	619(C), 641(D), 668(D), 718(D), 745(D), 755(D), 783(D), 808(C), 822(D), 856(D), 935(D), 1000(D), 1029(D), 1063(D), 1088(D), 1154(C), 1170(D), 1204(C), 1232(C), 1251(D), 1307(D), 1337(D), 1449(D), 1484(D), 1553(C), 1577(D), 1603(D), 1615(C), 1659(D), 1684(C), 2852(C), 2873(D), 2894(D), 2934(C), 2942(C), 2968(D), 3010(C), 3064(D)	71%
1154	619(D), 641(D), 668(D), 718(D), 745(D), 755(D), 783(D), 808(D), 822(D), 856(D), 935(D), 1000(D), 1029(D), 1063(D), 1088(D), 1124(D), 1170(D), 1204(D), 1232(D), 1251(D), 1307(D), 1337(D), 1449(D), 1484(D), 1553(C), 1577(C), 1603(D), 1615(D), 1659(D), 1684(C), 2852(D), 2873(C), 2894(C), 2934(C), 2942(C), 2968(D), 3010(C), 3064(D)	100%
1170	619(C), 641(D), 668(D), 718(D), 745(C), 755(D), 783(D), 808(C), 822(D), 856(C), 935(D), 1000(D), 1029(D), 1063(D), 1088(C), 1124(C), 1154(C), 1204(C), 1232(D), 1251(C), 1307(C), 1337(C), 1449(D), 1484(D), 1553(C), 1577(C), 1603(D), 1615(C), 1659(C), 1684(C), 2852(C), 2873(C), 2894(C), 2934(C), 2942(C), 2968(D), 3010(C), 3064(D)	45%
1204	619(D), 641(C), 668(D), 718(D), 745(D), 755(C), 783(C), 808(C), 822(C), 856(C), 935(D), 1000(C), 1029(D), 1063(D), 1088(C), 1124(C), 1154(D), 1170(C), 1232(D), 1251(C), 1307(C), 1337(C), 1449(D), 1484(D), 1553(C), 1577(C), 1603(C), 1615(C), 1659(C), 1684(C), 2852(C), 2873(C), 2894(C), 2934(C), 2942(C), 2968(C), 3010(C), 3064(C)	29%
1232	619(C), 641(D), 668(D), 718(D), 745(C), 755(D), 783(D), 808(C), 822(D), 856(D), 935(D), 1000(D), 1029(D), 1063(D), 1088(D), 1124(C), 1154(D), 1170(C), 1204(D), 1251(D), 1307(D), 1337(D), 1449(D), 1484(D), 1553(D), 1577(D), 1603(D), 1615(D), 1659(D), 1684(C), 2852(D), 2873(D), 2894(D), 2934(D), 2942(D), 2968(D), 3010(C), 3064(D)	82%
1251	619(D), 641(C), 668(C), 718(C), 745(D), 755(C), 783(C), 808(D), 822(C), 856(C), 935(C), 1000(C), 1029(C), 1063(C), 1088(C), 1124(D), 1154(D), 1170(C), 1204(C), 1232(D), 1307(C), 1337(C), 1449(C), 1484(C), 1553(C), 1577(C), 1603(C), 1615(C), 1659(C), 1684(D), 2852(C), 2873(C), 2894(C), 2934(C), 2942(C), 2968(C), 3010(D), 3064(C)	21%
1307	619(D), 641(C), 668(C), 718(C), 745(D), 755(C), 783(C), 808(C), 822(C), 856(C), 935(C), 1000(C), 1029(C), 1063(D), 1088(C), 1124(D), 1154(D), 1170(C), 1204(C), 1232(D), 1251(C), 1337(C), 1449(C), 1484(C), 1553(C), 1577(C), 1603(C), 1615(C), 1659(C), 1684(D), 2852(C), 2873(C), 2894(C), 2934(C), 2942(C), 2968(C), 3010(C), 3064(C)	16%
1337	619(D), 641(C), 668(C), 718(C), 745(D), 755(C), 783(C), 808(D), 822(C), 856(C), 935(C), 1000(C), 1029(C), 1063(C), 1088(C), 1124(C), 1154(D), 1170(D), 1204(C), 1232(D), 1251(C), 1307(C), 1449(C), 1484(C), 1553(C), 1577(C), 1603(C), 1615(C), 1659(C), 1684(C), 2852(C), 2873(C), 2894(C), 2934(C), 2942(C), 2968(C), 3010(C), 3064(C)	16%
1449	619(D), 641(C), 668(C), 718(C), 745(D), 755(C), 783(C), 808(D), 822(C), 856(D), 935(C), 1000(C), 1029(C), 1063(C), 1088(C), 1124(C), 1154(D), 1170(D), 1204(D), 1232(D), 1251(D), 1307(C), 1337(C), 1484(C), 1553(D), 1577(C), 1603(C), 1615(C), 1659(C), 1684(D), 2852(C), 2873(C), 2894(C), 2934(C), 2942(C), 2968(C), 3010(C), 3064(C)	32%
1484	619(D), 641(D), 668(C), 718(C), 745(D), 755(C), 783(C), 808(D), 822(D), 856(D), 935(D), 1000(D), 1029(C), 1063(C), 1088(D), 1124(D), 1154(D), 1170(D), 1204(D), 1232(D), 1251(C), 1307(C), 1337(C), 1449(C), 1553(D), 1577(D), 1603(C), 1615(C), 1658(C), 1684(D), 2852(C), 2873(D), 2894(D), 2934(D), 2942(D), 2968(C), 3010(D), 3064(D)	61%
1553	619(C), 641(D), 668(D), 718(D), 745(C), 755(D), 783(C), 808(C), 822(C), 856(C), 935(D), 1000(C), 1029(D), 1063(D), 1088(C), 1124(C), 1154(D), 1170(C), 1204(C), 1232(D), 1251(C), 1307(C), 1337(C), 1449(C), 1484(D), 1553(C), 1577(C), 1603(D), 1615(C), 1659(C), 1684(C), 2852(C), 2873(C), 2894(C), 2934(C), 2942(C), 2968(C), 3010(C), 3064(C)	24%
1577	619(D), 641(D), 668(C), 718(D), 745(D), 755(C), 783(C), 808(C), 822(C), 856(C), 935(D), 1000(C), 1029(C), 1063(D), 1088(C), 1124(D), 1154(D), 1170(C), 1204(C), 1232(D), 1251(C), 1307(C), 1337(C), 1449(C), 1484(D), 1553(C), 1577(D), 1603(C), 1615(C), 1658(C), 1684(C), 2852(C), 2873(C), 2894(C), 2934(C), 2942(C), 2968(C), 3010(C), 3064(C)	24%
1603	619(D), 641(C), 668(C), 718(C), 745(D), 755(C), 783(C), 808(D), 822(C), 856(C), 935(C), 1000(C), 1029(C), 1063(D), 1088(C), 1124(D), 1154(D), 1170(D), 1204(C), 1232(D), 1251(C), 1307(C), 1337(C), 1449(C), 1484(C), 1553(D), 1577(C), 1615(C), 1658(C), 1684(D), 2852(C), 2873(C), 2894(C), 2934(C), 2942(C), 2968(C), 3010(D), 3064(C)	29%
1615	619(D), 641(C), 668(C), 718(C), 745(D), 755(C), 783(C), 808(D), 822(C), 856(C), 935(C), 1000(C), 1029(C), 1063(C), 1088(C), 1124(C), 1154(D), 1170(D), 1204(C), 1232(D), 1251(C), 1307(C), 1337(C), 1449(D), 1484(C), 1553(D), 1577(C), 1603(C), 1659(C), 1684(D), 2852(C), 2873(C), 2894(C), 2934(C), 2942(C), 2968(D), 3010(C), 3064(C)	11%
1659	619(D), 641(C), 668(C), 718(C), 745(D), 755(C), 783(C), 808(C), 822(C), 856(C), 935(C), 1000(C), 1029(C), 1063(D), 1088(C), 1124(C), 1154(D), 1170(C), 1204(C), 1232(D), 1251(C), 1307(C), 1337(C), 1449(C), 1484(C), 1553(C), 1577(C), 1603(C), 1615(C), 1658(C), 1684(D), 2852(C), 2873(C), 2894(C), 2934(C), 2942(C), 2968(C), 3010(C), 3064(C)	37%
1684	619(C), 641(D), 668(D), 718(D), 745(C), 755(D), 783(D), 808(C), 822(D), 856(C), 935(C), 1000(D), 1029(D), 1063(D), 1088(C), 1124(C), 1154(D), 1170(D), 1204(C), 1232(C), 1251(D), 1307(C), 1337(C), 1449(D), 1484(D), 1553(D), 1577(C), 1603(D), 1615(C), 1659(C), 1684(C), 2852(C), 2873(C), 2894(D), 2934(C), 2942(C), 2968(D), 3010(C), 3064(D)	50%
2852	619(D), 641(D), 668(C), 718(C), 745(D), 755(D), 783(D), 808(C), 822(C), 856(C), 935(C), 1000(C), 1029(C), 1063(D), 1088(C), 1124(C), 1154(D), 1170(D), 1204(C), 1232(D), 1251(C), 1307(C), 1337(C), 1449(C), 1484(D), 1553(C), 1577(C), 1603(C), 1615(C), 1659(C), 1684(D), 2852(C), 2873(C), 2894(C), 2934(C), 2942(C), 2968(C), 3010(C), 3064(C)	11%
2873	619(D), 641(C), 668(C), 718(C), 745(D), 755(C), 783(C), 808(C), 822(C), 856(C), 935(C), 1000(C), 1029(C), 1063(D), 1088(C), 1124(D), 1154(D), 1170(C), 1204(C), 1232(D), 1251(C), 1307(C), 1337(C), 1449(C), 1484(D), 1553(C), 1577(C), 1603(C), 1615(C), 1659(C), 1684(D), 2852(C), 2894(C), 2934(C), 2942(C), 2968(C), 3010(C), 3064(C)	26%
2894	619(D), 641(C), 668(C), 718(D), 745(D), 755(C), 783(C), 808(D), 822(C), 856(C), 935(D), 1000(C), 1029(C), 1063(D), 1088(D), 1124(D), 1154(D), 1170(D), 1204(C), 1232(D), 1251(C), 1307(C), 1337(C), 1449(C), 1484(D), 1553(C), 1577(C), 1603(C), 1615(C), 1659(C), 1684(D), 2852(C), 2873(C), 2894(C), 2934(C), 2942(C), 2968(C), 3010(C), 3064(C)	34%
2934	619(C), 641(C), 668(C), 718(C), 745(C), 755(C), 783(C), 808(D), 822(C), 856(C), 935(D), 1000(C), 1029(C), 1063(D), 1088(C), 1124(C), 1154(D), 1170(D), 1204(C), 1232(D), 1251(C), 1307(C), 1337(C), 1449(C), 1484(D), 1553(C), 1577(C), 1603(C), 1615(C), 1659(C), 1684(D), 2852(C), 2873(C), 2894(C), 2934(C), 2942(C), 2968(C), 3010(C), 3064(C)	16%
2942	619(D), 641(C), 668(C), 718(C), 745(D), 755(C), 783(C), 808(D), 822(C), 856(C), 935(D), 1000(C), 1029(C), 1063(C), 1088(C), 1124(C), 1154(D), 1170(C), 1204(C), 1232(D), 1251(C), 1307(C), 1337(C), 1449(C), 1484(D), 1553(C), 1577(C), 1603(C), 1615(C), 1659(C), 1684(D), 2852(C), 2873(C), 2894(C), 2934(C), 2942(C), 3010(D), 3064(C)	26%
2968	619(D), 641(C), 668(C), 718(C), 745(D), 755(C), 783(C), 808(D), 822(C), 856(D), 935(D), 1000(D), 1029(C), 1063(C), 1088(C), 1124(D), 1154(D), 1170(D), 1204(C), 1232(D), 1251(C), 1307(C), 1337(C), 1449(C), 1484(C), 1553(D), 1577(C), 1603(C), 1615(C), 1659(C), 1684(D), 2852(C), 2873(C), 2894(C), 2934(C), 2942(C), 3010(D), 3064(C)	34%
3010	619(C), 641(D), 668(D), 718(D), 745(C), 755(D), 783(D), 808(D), 822(D), 856(C), 935(D), 1000(D), 1029(D), 1063(D), 1088(C), 1124(C), 1154(C), 1170(D), 1204(C), 1232(C), 1251(D), 1307(C), 1337(C), 1449(D), 1484(D), 1553(C), 1577(C), 1603(C), 1615(C), 1659(D), 1684(C), 2852(C), 2873(C), 2894(C), 2934(C), 2942(D), 2968(D), 3064(D)	50%
3064	619(D), 641(C), 668(C), 718(D), 745(D), 755(C), 783(C), 808(C), 822(C), 856(C), 935(D), 1000(C), 1029(C), 1063(D), 1088(C), 1124(D), 1154(D), 1170(D), 1204(C), 1232(D), 1251(C), 1307(C), 1337(C), 1449(C), 1484(D), 1553(C), 1577(C), 1603(C), 1615(C), 1659(C), 1684(D), 2852(C), 2873(C), 2894(C), 2934(C), 2942(C), 2968(C), 3010(D)	32%

Table C.8: Peak ratio analysis (PIR) of the comparison between H722/18R and H722/18L. The letter D specifies PIR separation between H722/18R and H722/18L. The letter C specifies that the PIR marker is overlapping between samples. The letter X specifies that the PIR comparison could not take place as the band was not present in one of the samples.

Panel	PIR Matches	% matchD
619	641(D), 668(D), 697(X), 718(D) / 745(C), 755(C), 788(D), 808(D)	55%(C)
641	619(D), 668(D), 697(X), 718(D) / 745(D), 755(D), 783(D), 808(D)	79%(C)
668	619(D), 641(D), 697(X), 718(D) / 745(D), 755(D), 783(D), 808(D)	100%(C)
718	619(D), 641(D), 668(D), 697(X) / 745(D), 755(D), 780(D), 808(D)	76%(C)
745	619(C), 641(D), 668(D), 697(X), 718(D) / 755(C), 783(D), 808(D)	39%(C)
755	619(C), 641(D), 668(D), 697(X), 718(D) / 745(C), 780(D), 808(D)	61%(C)
783	619(D), 641(D), 668(D), 697(X), 718(D) / 745(D), 755(D), 808(D), 822(D)	95%(C)
808	619(D), 641(D), 668(C), 697(X), 718(D) / 745(D), 755(D), 783(D), 822(D)	100%(C)
822	619(D), 641(C), 668(D), 697(X), 718(D) / 745(C), 755(C), 783(D), 808(D)	47%(C)
850	619(D), 641(C), 668(D), 697(X), 718(D) / 745(D), 755(D), 783(D), 808(D), 822(D)	79%(C)
879	619(D), 641(D), 668(D), 697(X), 718(D) / 745(D), 755(D), 783(C), 808(D), 822(D), 850(D)	95%(C)
896	619(C), 641(D), 668(D), 697(X), 718(D) / 744(C), 755(C), 783(D), 808(D), 822(C), 850(C), 879(D)	55%(C)
935	619(C), 641(D), 668(D), 697(X), 718(D) / 745(C), 755(C), 783(D), 808(D), 822(D), 850(D), 879(D), 896(D)	74%(C)
1000	619(D), 641(D), 668(D), 697(X), 718(D) / 745(C), 755(C), 783(D), 808(D), 822(C), 850(C), 879(D), 896(D), 935(D)	58%(C)
1029	619(D), 641(C), 668(D), 697(X), 718(D) / 745(D), 755(D), 783(D), 808(D), 822(C), 850(C), 879(D), 896(D), 935(D), 1000(C)	66%(C)
1088	619(D), 641(D), 668(D), 697(X), 718(D) / 745(D), 755(D), 783(D), 808(D), 822(D), 850(D), 879(D), 896(D), 935(D), 1000(D), 1029(D)	82%(C)
1124	619(C), 641(D), 668(D), 697(X), 718(C) / 744(C), 755(C), 783(D), 808(D), 822(D), 850(D), 879(D), 896(C), 935(C), 1000(D), 1029(D), 1063(X)	66%(C)
1154	619(C), 641(D), 668(D), 697(X), 718(C) / 745(C), 755(C), 783(D), 808(D), 822(D), 850(D), 879(D), 896(D), 935(C), 1000(D), 1029(D), 1063(X), 1088(D)	55%(C)
1170	619(C), 641(D), 668(D), 697(X), 718(D) / 745(C), 755(C), 783(D), 808(D), 822(C), 850(C), 879(D), 896(D), 935(C), 1000(D), 1029(D), 1063(X), 1088(D), 1124(C)	65%(C)
1206	619(D), 641(D), 668(D), 697(X), 718(D) / 745(C), 755(C), 783(D), 808(D), 822(D), 850(D), 879(D), 896(D), 935(D), 1000(D), 1029(D), 1063(X), 1088(D), 1124(D), 1154(D)	95%(C)
1247	619(D), 641(D), 668(D), 697(X), 718(D) / 745(C), 755(C), 783(D), 808(D), 822(C), 850(C), 879(D), 896(D), 935(D), 1000(D), 1029(D), 1063(X), 1088(D), 1124(D), 1154(D), 1170(D)	96%(C)
1307	619(C), 641(D), 668(D), 697(X), 718(D) / 745(C), 755(C), 783(D), 808(D), 822(C), 850(C), 879(D), 896(D), 935(C), 1000(D), 1029(D), 1063(X), 1088(D), 1124(D), 1154(D), 1170(D), 1204(D)	71%(C)
1337	619(C), 641(C), 668(D), 697(X), 718(D) / 745(C), 755(C), 783(D), 808(D), 822(C), 850(C), 879(D), 896(D), 935(D), 1000(C), 1029(D), 1063(X), 1088(D), 1124(D), 1154(D), 1170(D), 1204(D), 1232(D)	42%(C)
1449	619(D), 641(D), 668(D), 697(X), 718(C) / 745(C), 755(D), 783(D), 808(D), 822(D), 850(D), 879(D), 896(D), 935(D), 1000(D), 1029(D), 1063(X), 1088(D), 1124(C), 1154(C), 1170(C), 1204(D), 1232(D), 1251(D)	97%(C)
1484	619(D), 641(D), 668(D), 697(X), 718(D) / 744(C), 756(D), 783(D), 808(D), 822(D), 850(D), 879(C), 896(D), 935(D), 1000(D), 1029(D), 1063(X), 1088(D), 1124(D), 1154(D), 1170(D), 1204(D), 1232(C), 1251(D), 1307(D)	97%(C)
1553	619(C), 641(D), 668(D), 697(X), 718(D) / 745(C), 755(C), 783(D), 808(D), 822(C), 850(D), 879(D), 896(D), 935(C), 1000(D), 1029(D), 1063(X), 1088(D), 1124(C), 1154(C), 1170(C), 1204(C), 1232(D), 1251(D), 1307(D), 1337(C)	58%(C)
1577	619(D), 641(C), 668(D), 697(X), 718(D), 745(D), 755(D), 783(D), 808(D), 822(C), 850(C), 879(D), 896(D), 935(D), 1000(D), 1029(D), 1063(X), 1088(D), 1124(D), 1154(D), 1170(D), 1204(D), 1232(D), 1251(D), 1307(D), 1337(C), 1449(D)	82%(C)
1603	619(D), 641(C), 668(D), 697(X), 718(D), 745(D), 755(D), 783(D), 808(D), 822(C), 850(C), 879(D), 896(D), 935(D), 1000(D), 1029(D), 1063(X), 1088(D), 1124(D), 1154(D), 1170(D), 1204(D), 1232(D), 1251(D), 1307(D), 1337(C), 1449(D), 1484(D), 1553(D)	79%(C)
1615	619(D), 641(C), 668(D), 697(X), 718(D), 745(D), 755(D), 783(D), 808(D), 822(C), 850(C), 879(D), 896(D), 935(D), 1000(D), 1029(D), 1063(X), 1088(D), 1124(D), 1154(D), 1170(D), 1204(D), 1232(D), 1251(D), 1307(D), 1337(C), 1449(D), 1484(D), 1553(D), 1577(C)	79%(C)
1659	619(C), 641(D), 668(D), 697(X), 718(D), 745(C), 755(C), 783(D), 808(D), 822(C), 850(C), 879(D), 896(D), 935(C), 1000(D), 1029(D), 1063(X), 1088(D), 1124(C), 1154(C), 1170(C), 1204(C), 1232(D), 1251(D), 1307(D), 1337(C), 1449(D), 1484(D), 1553(C), 1577(D), 1603(D)	58%(C)
1684	619(C), 641(D), 668(D), 697(X), 718(C), 745(C), 755(C), 783(D), 808(D), 822(C), 850(D), 879(D), 896(D), 935(C), 1000(D), 1029(D), 1063(X), 1088(D), 1124(C), 1154(C), 1170(D), 1204(C), 1232(D), 1251(D), 1307(D), 1337(C), 1449(D), 1484(D), 1553(C), 1577(D), 1603(D), 1615(D)	60%(C)
2852	619(D), 641(D), 668(D), 697(X), 718(D), 745(D), 755(D), 783(D), 808(D), 822(D), 850(D), 879(D), 896(D), 935(D), 1000(D), 1029(D), 1063(X), 1088(D), 1124(D), 1154(D), 1170(D), 1204(D), 1232(D), 1251(D), 1307(D), 1337(D), 1449(D), 1484(D), 1553(D), 1577(D), 1603(D), 1615(D), 1659(D)	100%(D)
2873	619(D), 641(D), 668(D), 697(X), 718(D), 745(D), 755(D), 783(D), 808(D), 822(D), 850(D), 879(D), 896(D), 935(D), 1000(D), 1029(D), 1063(X), 1088(D), 1124(D), 1154(D), 1170(D), 1204(D), 1232(D), 1251(D), 1307(D), 1337(D), 1449(C), 1484(D), 1553(D), 1577(D), 1603(D), 1615(D), 1659(D), 1684(D)	97%(C)
2894	619(D), 641(D), 668(D), 697(X), 718(C), 745(C), 755(D), 783(D), 808(D), 822(D), 850(D), 879(D), 896(D), 935(C), 1000(D), 1029(D), 1063(X), 1088(D), 1124(C), 1154(C), 1170(D), 1204(D), 1232(D), 1251(C), 1307(D), 1337(D), 1449(C), 1484(D), 1553(D), 1577(D), 1603(D), 1615(D), 1659(D), 1684(C), 2852(D)	74%(C)
2934	619(C), 641(C), 668(D), 697(X), 718(D), 745(C), 755(C), 783(D), 808(D), 822(C), 850(C), 879(D), 896(D), 935(C), 1000(C), 1029(D), 1063(X), 1088(D), 1124(D), 1154(C), 1170(C), 1204(C), 1232(D), 1251(D), 1307(D), 1337(D), 1449(C), 1484(D), 1553(C), 1577(C), 1603(C), 1615(C), 1659(C), 1684(D), 2852(D)	50%(C)
2942	619(D), 641(D), 668(D), 697(X), 718(C), 745(C), 755(D), 783(D), 808(D), 822(D), 850(D), 879(D), 896(D), 935(C), 1000(D), 1029(D), 1063(X), 1088(C), 1124(C), 1154(C), 1170(D), 1204(D), 1232(D), 1251(C), 1307(D), 1337(D), 1449(C), 1484(D), 1553(D), 1577(D), 1603(D), 1615(D), 1659(D), 1684(C), 2852(D)	68%(C)
2968	619(C), 641(D), 668(D), 697(X), 718(C), 745(C), 755(C), 783(D), 808(D), 822(C), 850(C), 879(D), 896(D), 935(C), 1000(D), 1029(S), 1063(X), 1088(D), 1124(C), 1154(C), 1170(C), 1204(D), 1232(D), 1251(D), 1307(C), 1337(C), 1449(D), 1484(D), 1553(D), 1577(D), 1603(D), 1615(D), 1659(C), 1684(C), 2852(D)	58%(C)
3010	619(D), 641(C), 668(D), 697(X), 718(D), 745(C), 755(C), 783(D), 808(D), 822(C), 850(D), 879(D), 896(D), 935(D), 1000(C), 1029(C), 1063(X), 1088(D), 1124(C), 1154(C), 1170(C), 1204(C), 1232(D), 1251(D), 1307(D), 1337(C), 1449(D), 1484(D), 1553(C), 1577(C), 1603(C), 1615(C), 1659(D), 1684(C), 2852(D)	55%(C)
3064	619(C), 641(D), 668(D), 697(X), 718(D), 745(C), 755(C), 783(D), 808(D), 822(C), 850(D), 879(D), 896(C), 935(D), 1000(C), 1029(D), 1063(X), 1088(D), 1124(C), 1154(C), 1170(C), 1204(C), 1232(D), 1251(D), 1307(D), 1337(C), 1449(D), 1484(D), 1553(C), 1577(D), 1603(D), 1615(D), 1659(C), 1684(C), 2852(D)	58%(C)

Table C.9: Peak ratio analysis (PIR) of the comparison between H794/18R and H794/18L. The letter D specifies PIR separation between H794/18R and H794/18L. The letter C specifies that the PIR marker is overlapping between samples. The letter X specifies that the PIR comparison could not take place as the band was not present in one of the samples.

Panel	PIR Matches	% matchD
619	641(D), 668(D), 718(D), 745(D), 755(D), 783(D), 808(X), 822(C), 850(C), 935(C), 1000(C), 1029(C), 1063(D), 1088(D), 1124(C), 1154(D), 1176(C), 1204(C), 1232(D), 1251(C), 1307(C), 1337(C), 1449(D), 1484(D), 1553(D), 1577(D), 1603(C), 1659(C), 1684(C), 2852(D), 2873(D), 2894(C), 2934(D), 2942(C), 2968(C), 3010(C), 3064(C)	47%
641	619(D), 668(D), 718(C), 745(D), 755(C), 783(D), 808(X), 822(D), 850(D), 935(D), 1000(D), 1029(D), 1063(D), 1088(D), 1124(D), 1154(C), 1176(D), 1204(D), 1232(D), 1251(D), 1307(D), 1337(D), 1449(D), 1484(D), 1553(D), 1577(D), 1603(D), 1659(D), 1684(D), 2852(D), 2873(D), 2894(D), 2934(D), 2942(D), 2968(D), 3010(D), 3064(D)	92%
668	619(D), 641(D), 718(D), 745(D), 755(D), 783(C), 808(X), 822(D), 850(D), 935(D), 1000(D), 1029(D), 1063(D), 1088(D), 1124(D), 1154(C), 1176(D), 1204(D), 1232(D), 1251(D), 1307(D), 1337(D), 1449(D), 1484(C), 1553(D), 1577(D), 1603(D), 1659(D), 1684(D), 2852(D), 2873(D), 2894(D), 2934(D), 2942(D), 2968(D), 3010(D), 3064(D)	89%
718	619(D), 641(C), 668(D), 745(D), 755(C), 783(D), 808(X), 822(D), 850(D), 935(D), 1000(D), 1029(D), 1063(D), 1088(D), 1124(D), 1154(C), 1176(D), 1204(D), 1232(D), 1251(D), 1307(D), 1337(D), 1449(D), 1484(D), 1553(D), 1577(D), 1603(D), 1659(D), 1684(D), 2852(D), 2873(D), 2894(D), 2934(D), 2942(D), 2968(D), 3010(D), 3064(D)	92%
745	619(D), 641(D), 668(D), 718(D), 755(D), 783(D), 808(X), 822(C), 850(C), 935(C), 1000(D), 1029(D), 1063(D), 1088(D), 1124(C), 1154(D), 1176(D), 1204(C), 1232(D), 1251(D), 1307(D), 1337(C), 1449(D), 1484(D), 1553(C), 1577(C), 1603(D), 1659(D), 1684(D), 2852(D), 2873(D), 2894(D), 2934(C), 2942(D), 2968(C), 3010(D), 3064(D)	68%
755	619(D), 641(C), 668(D), 718(C), 745(D), 783(D), 808(X), 822(D), 850(D), 935(D), 1000(D), 1029(D), 1063(D), 1088(D), 1124(D), 1154(C), 1176(D), 1204(D), 1232(D), 1251(D), 1307(D), 1337(D), 1449(D), 1484(D), 1553(D), 1577(D), 1603(D), 1659(D), 1684(D), 2852(D), 2873(D), 2894(D), 2934(D), 2942(D), 2968(D), 3010(D), 3064(D)	87%
783	619(D), 641(D), 668(C), 718(D), 745(D), 755(D), 808(X), 822(D), 850(D), 935(D), 1000(D), 1029(D), 1063(C), 1088(D), 1124(D), 1154(D), 1176(D), 1204(D), 1232(C), 1251(D), 1307(D), 1337(D), 1449(D), 1484(D), 1553(D), 1577(D), 1603(D), 1659(D), 1684(D), 2852(D), 2873(D), 2894(D), 2934(D), 2942(D), 2968(D), 3010(D), 3064(D)	87%
822	619(C), 641(D), 668(D), 718(D), 745(C), 755(D), 783(D), 808(X), 850(C), 935(C), 1000(C), 1029(C), 1063(D), 1088(C), 1124(C), 1154(D), 1176(C), 1204(C), 1232(D), 1251(C), 1307(C), 1337(C), 1449(D), 1484(D), 1553(C), 1577(C), 1603(D), 1659(C), 1684(C), 2852(D), 2873(D), 2894(D), 2934(C), 2942(D), 2968(C), 3010(D), 3064(C)	44%
850	619(C), 641(D), 668(D), 718(D), 745(C), 755(D), 783(D), 808(X), 822(C), 935(C), 1000(C), 1029(D), 1063(D), 1088(C), 1124(C), 1154(D), 1176(C), 1204(C), 1232(D), 1251(D), 1307(C), 1337(C), 1449(D), 1484(D), 1553(C), 1577(C), 1603(D), 1659(D), 1684(D), 2852(D), 2873(D), 2894(D), 2934(C), 2942(D), 2968(C), 3010(D), 3064(C)	56%
935	619(C), 641(D), 668(D), 718(D), 745(C), 755(D), 783(D), 808(X), 822(C), 850(C), 1000(C), 1029(D), 1063(D), 1088(C), 1124(C), 1154(D), 1176(C), 1204(C), 1232(D), 1251(D), 1307(D), 1337(C), 1449(D), 1484(D), 1553(C), 1577(C), 1603(D), 1659(D), 1684(D), 2852(D), 2873(D), 2894(D), 2934(C), 2942(D), 2968(D), 3010(D), 3064(C)	61%
1000	619(C), 641(D), 668(D), 718(D), 745(D), 755(D), 783(D), 808(X), 822(C), 850(C), 935(C), 1029(C), 1063(D), 1088(D), 1124(C), 1154(D), 1176(C), 1204(C), 1232(D), 1251(D), 1307(D), 1337(C), 1449(D), 1484(D), 1553(D), 1577(D), 1603(C), 1659(D), 1684(D), 2852(D), 2873(D), 2894(D), 2934(C), 2942(D), 2968(D), 3010(D), 3064(C)	64%
1029	619(C), 641(D), 668(D), 718(D), 745(D), 755(D), 783(D), 808(X), 822(C), 850(D), 935(D), 1000(C), 1063(D), 1088(D), 1124(D), 1154(D), 1176(C), 1204(C), 1232(D), 1251(C), 1307(C), 1337(D), 1449(C), 1484(D), 1553(D), 1577(D), 1603(C), 1659(C), 1684(C), 2852(D), 2873(D), 2894(D), 2934(C), 2942(C), 2968(C), 3010(C), 3064(C)	56%
1063	619(D), 641(D), 668(C), 718(D), 745(D), 755(D), 783(C), 808(X), 822(D), 850(D), 935(D), 1000(D), 1029(D), 1088(D), 1124(D), 1154(D), 1176(D), 1204(D), 1232(C), 1251(D), 1307(D), 1337(D), 1449(D), 1484(C), 1553(D), 1577(D), 1603(D), 1659(D), 1684(D), 2852(D), 2873(D), 2894(D), 2934(D), 2942(D), 2968(D), 3010(D), 3064(D)	92%
1088	619(D), 641(D), 668(D), 718(D), 745(C), 755(D), 783(D), 808(X), 822(X), 850(X), 935(X), 1000(C), 1029(D), 1063(D), 1124(C), 1154(D), 1176(D), 1204(C), 1232(D), 1251(D), 1307(D), 1337(D), 1449(D), 1484(D), 1553(C), 1577(C), 1603(D), 1659(D), 1684(C), 2852(D), 2873(D), 2894(D), 2934(C), 2942(D), 2968(D), 3010(D), 3064(D)	67%
1124	619(C), 641(D), 668(D), 718(D), 745(C), 755(D), 783(D), 808(X), 822(C), 850(C), 935(C), 1000(C), 1029(D), 1063(D), 1088(C), 1154(D), 1176(C), 1204(C), 1232(D), 1251(D), 1307(C), 1337(C), 1449(D), 1484(D), 1553(C), 1577(C), 1603(D), 1659(D), 1684(D), 2852(D), 2873(D), 2894(D), 2934(C), 2942(D), 2968(D), 3010(D), 3064(C)	58%
1154	619(D), 641(C), 668(D), 718(C), 745(D), 755(C), 783(D), 808(X), 822(D), 850(D), 935(D), 1000(D), 1029(D), 1063(D), 1088(D), 1124(D), 1176(D), 1204(D), 1232(D), 1251(D), 1307(D), 1337(D), 1449(D), 1484(D), 1553(D), 1577(D), 1603(C), 1659(C), 1684(D), 2852(D), 2873(D), 2894(D), 2934(C), 2942(D), 2968(D), 3010(D), 3064(D)	92% 3
1170	619(C), 641(D), 668(D), 718(D), 745(D), 755(D), 783(D), 808(X), 822(C), 850(C), 935(C), 1000(C), 1029(C), 1063(D), 1088(D), 1124(C), 1154(D), 1176(D), 1204(D), 1232(D), 1251(C), 1307(C), 1337(C), 1449(D), 1484(D), 1553(D), 1577(D), 1603(C), 1658(C), 1684(C), 2852(D), 2873(D), 2894(D), 2934(C), 2942(D), 2968(C), 3010(C), 3064(C)	50%
1204	619(C), 641(D), 668(D), 718(D), 745(C), 755(D), 783(D), 808(X), 822(C), 850(C), 935(C), 1000(C), 1029(C), 1063(D), 1088(C), 1124(C), 1154(D), 1176(C), 1232(D), 1251(C), 1307(C), 1337(C), 1449(D), 1484(D), 1553(C), 1577(C), 1603(C), 1658(C), 1684(D), 2852(D), 2873(D), 2894(D), 2934(C), 2942(D), 2968(C), 3010(D), 3064(C)	44%
1232	619(D), 641(D), 668(C), 718(D), 745(D), 755(D), 783(C), 808(X), 822(D), 850(D), 935(D), 1000(D), 1029(C), 1063(D), 1088(D), 1124(D), 1154(D), 1176(D), 1204(D), 1251(D), 1307(D), 1337(D), 1449(D), 1484(C), 1553(D), 1577(D), 1603(D), 1659(D), 1684(D), 2852(D), 2873(D), 2894(D), 2934(D), 2942(D), 2968(D), 3010(D), 3064(D)	92%
1251	619(C), 641(D), 668(D), 718(D), 745(D), 755(D), 783(D), 808(X), 822(D), 850(D), 935(D), 1000(C), 1029(C), 1063(D), 1088(D), 1124(D), 1154(D), 1176(C), 1204(C), 1232(D), 1307(C), 1337(D), 1449(D), 1484(D), 1553(D), 1577(D), 1603(C), 1659(C), 1684(C), 2852(D), 2873(D), 2894(D), 2934(D), 2942(C), 2968(C), 3010(C), 3064(D)	67%
1307	619(C), 641(D), 668(D), 718(D), 745(D), 755(D), 783(D), 808(X), 822(C), 850(C), 935(D), 1000(C), 1029(C), 1063(D), 1088(D), 1124(C), 1154(D), 1176(C), 1204(C), 1232(D), 1251(C), 1337(C), 1449(D), 1484(D), 1553(D), 1577(D), 1603(C), 1658(C), 1684(C), 2852(D), 2873(D), 2894(D), 2934(D), 2942(C), 2968(C), 3010(C), 3064(C)	50%
1337	619(C), 641(D), 668(D), 718(D), 745(C), 755(D), 783(D), 808(X), 822(C), 850(C), 935(C), 1000(C), 1029(D), 1063(D), 1088(C), 1124(C), 1154(D), 1176(C), 1204(C), 1232(D), 1251(D), 1307(C), 1449(D), 1484(D), 1553(C), 1577(C), 1603(D), 1659(D), 1684(D), 2852(D), 2873(D), 2894(D), 2934(D), 2942(D), 2968(D), 3010(D), 3064(C)	58%
1449	619(D), 641(D), 668(D), 718(D), 745(D), 755(D), 783(D), 808(X), 822(D), 850(D), 935(D), 1000(D), 1029(C), 1063(D), 1088(D), 1124(D), 1154(D), 1176(D), 1204(D), 1232(D), 1251(D), 1307(D), 1337(D), 1484(C), 1553(D), 1577(D), 1603(D), 1659(D), 1684(C), 2852(D), 2873(C), 2894(C), 2934(D), 2942(C), 2968(C), 3010(C), 3064(D)	56%
1484	619(D), 641(D), 668(C), 718(D), 745(D), 755(D), 783(C), 808(X), 822(D), 850(D), 935(D), 1000(D), 1029(D), 1063(C), 1088(D), 1124(D), 1154(D), 1176(D), 1204(C), 1232(D), 1251(D), 1307(D), 1337(D), 1449(C), 1553(D), 1577(D), 1603(C), 1659(D), 1684(D), 2852(D), 2873(D), 2894(D), 2934(D), 2942(C), 2968(D), 3010(D), 3064(D)	89%
1553	619(D), 641(D), 668(D), 718(D), 745(C), 755(D), 783(D), 808(X), 822(C), 850(C), 935(C), 1000(D), 1029(D), 1063(D), 1088(C), 1124(C), 1154(D), 1176(D), 1204(C), 1232(D), 1251(C), 1307(D), 1337(C), 1449(D), 1484(D), 1577(D), 1603(D), 1659(D), 1684(D), 2852(D), 2873(D), 2894(D), 2934(C), 2942(D), 2968(D), 3010(D), 3064(D)	72%
1577	619(D), 641(D), 668(D), 718(D), 745(C), 755(D), 783(D), 808(X), 822(C), 850(C), 935(C), 1000(D), 1029(D), 1063(D), 1088(C), 1124(C), 1154(D), 1176(D), 1204(C), 1232(D), 1251(D), 1307(D), 1337(C), 1449(D), 1484(D), 1553(C), 1603(D), 1659(D), 1684(D), 2852(D), 2873(D), 2894(D), 2934(C), 2942(D), 2968(D), 3010(D), 3064(D)	72%
1603	619(C), 641(D), 668(D), 718(D), 745(D), 755(D), 783(D), 808(X), 822(C), 850(D), 935(D), 1000(D), 1029(C), 1063(D), 1088(D), 1124(D), 1154(D), 1176(C), 1204(D), 1232(C), 1251(C), 1307(C), 1337(D), 1449(D), 1484(D), 1553(D), 1577(D), 1659(C), 1684(C), 2852(D), 2873(D), 2894(D), 2934(D), 2942(C), 2968(C), 3010(C), 3064(D)	58%
1659	619(C), 641(C), 668(D), 718(D), 745(D), 755(D), 783(D), 808(X), 822(C), 850(D), 935(D), 1000(D), 1029(C), 1063(D), 1088(D), 1124(D), 1154(D), 1176(C), 1204(C), 1232(D), 1251(C), 1307(C), 1337(D), 1449(D), 1484(D), 1553(D), 1577(D), 1603(C), 1658(C), 1684(C), 2852(D), 2873(D), 2894(D), 2934(D), 2942(D), 2968(C), 3010(C), 3064(D)	67%
1684	619(C), 641(D), 668(D), 718(D), 745(D), 755(D), 783(D), 808(X), 822(D), 850(D), 935(D), 1000(D), 1029(C), 1063(D), 1088(D), 1124(D), 1154(D), 1176(D), 1204(D), 1232(D), 1251(D), 1307(C), 1337(C), 1449(D), 1484(D), 1553(D), 1577(D), 1603(C), 1659(C), 1684(C), 2852(D), 2873(C), 2894(D), 2934(D), 2942(C), 2968(C), 3010(C), 3064(D)	69%
2852	619(D), 641(D), 668(D), 718(D), 745(D), 755(D), 783(D), 808(X), 822(D), 850(D), 935(D), 1000(D), 1029(D), 1063(D), 1088(D), 1124(D), 1154(D), 1176(D), 1204(D), 1232(D), 1251(D), 1307(D), 1337(D), 1449(D), 1484(D), 1553(D), 1577(D), 1603(D), 1659(D), 1684(D), 2873(D), 2894(D), 2934(D), 2942(D), 2968(D), 3010(D), 3064(D)	100%
2873	619(D), 641(D), 668(D), 718(D), 745(D), 755(D), 783(D), 808(X), 822(D), 850(D), 935(D), 1000(D), 1029(D), 1063(D), 1088(D), 1124(D), 1154(D), 1176(D), 1204(D), 1232(D), 1251(D), 1307(D), 1337(D), 1449(C), 1484(D), 1553(D), 1577(D), 1603(D), 1659(D), 1684(C), 2852(D), 2894(C), 2934(D), 2942(D), 2968(D), 3010(C), 3064(D)	89%
2894	619(D), 641(D), 668(D), 718(D), 745(D), 755(D), 783(D), 808(X), 822(D), 850(D), 935(D), 1000(D), 1029(D), 1063(D), 1088(D), 1124(D), 1154(D), 1176(D), 1204(D), 1232(D), 1251(D), 1307(D), 1337(D), 1449(C), 1484(C), 1553(D), 1577(D), 1603(D), 1659(D), 1684(C), 2852(D), 2873(C), 2934(D), 2942(C), 2968(D), 3010(C), 3064(D)	83%
2934	619(C), 641(D), 668(D), 718(D), 745(C), 755(D), 783(D), 808(X), 822(D), 850(C), 935(C), 1000(C), 1029(C), 1063(D), 1088(C), 1124(C), 1154(D), 1176(C), 1204(C), 1232(D), 1251(D), 1307(C), 1337(C), 1449(D), 1484(D), 1553(C), 1577(C), 1603(C), 1659(D), 1684(C), 2852(D), 2873(D), 2894(D), 2942(D), 2968(D), 3010(D), 3064(C)	53%
2942	619(D), 641(D), 668(D), 718(D), 745(D), 755(D), 783(D), 808(X), 822(D), 850(D), 935(D), 1000(D), 1029(C), 1063(D), 1088(D), 1124(D), 1154(D), 1176(D), 1204(D), 1232(D), 1251(C), 1307(C), 1337(C), 1449(C), 1484(D), 1553(D), 1577(D), 1603(C), 1659(C), 1684(C), 2852(D), 2873(C), 2894(C), 2934(D), 2968(C), 3010(C), 3064(D)	69%
2968	619(C), 641(D), 668(D), 718(D), 745(D), 755(D), 783(D), 808(X), 822(C), 850(C), 935(D), 1000(C), 1029(C), 1063(D), 1088(D), 1124(D), 1154(D), 1176(C), 1204(C), 1232(D), 1251(C), 1307(C), 1337(D), 1449(D), 1484(D), 1553(D), 1577(D), 1603(C), 1659(C), 1684(C), 2852(D), 2873(D), 2894(D), 2934(D), 2942(D), 3010(D), 3064(D)	69%
3010	619(D), 641(D), 668(D), 718(D), 745(D), 755(C), 783(C), 808(X), 822(C), 850(D), 935(D), 1000(D), 1029(C), 1063(D), 1088(D), 1124(D), 1154(D), 1176(C), 1204(D), 1232(D), 1251(C), 1307(C), 1337(D), 1449(C), 1484(D), 1553(D), 1577(D), 1603(D), 1659(C), 1684(C), 2852(D), 2873(C), 2894(C), 2934(D), 2942(C), 2968(C), 3064(D)	61%
3064	619(C), 641(D), 668(D), 718(D), 745(C), 755(C), 783(D), 808(X), 822(C), 850(C), 935(C), 1000(C), 1029(C), 1063(D), 1088(D), 1124(D), 1154(C), 1176(D), 1204(C), 1232(D), 1251(D), 1307(C), 1337(C), 1449(D), 1484(D), 1553(D), 1577(D), 1603(C), 1659(D), 1684(D), 2852(D), 2873(D), 2894(D), 2934(C), 2942(D), 2968(D), 3010(D)	67%

BIBLIOGRAPHY

[1] Hyuna Sung, Jacques Ferlay, Rebecca L Siegel, Mathieu Laversanne, Isabelle Soerjomataram, Ahmedin Jemal, and Freddie Bray. Global cancer statistics 2020: Globocan estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA: a cancer journal for clinicians*, 71(3):209–249, 2021.

[2] Lindsey A Torre, Freddie Bray, Rebecca L Siegel, Jacques Ferlay, Joannie Lortet-Tieulent, and Ahmedin Jemal. Global cancer statistics, 2012. *CA: a cancer journal for clinicians*, 65(2):87–108, 2015.

[3] Ahmedin Jemal, Freddie Bray, Melissa M Center, Jacques Ferlay, Elizabeth Ward, and David Forman. Global cancer statistics. *CA: a cancer journal for clinicians*, 61(2):69–90, 2011.

[4] Freddie Bray, Jacques Ferlay, Isabelle Soerjomataram, Rebecca L Siegel, Lindsey A Torre, and Ahmedin Jemal. Global cancer statistics 2018: Globocan estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA: a cancer journal for clinicians*, 68(6):394–424, 2018.

[5] Melissa M Center, Ahmedin Jemal, Joannie Lortet-Tieulent, Elizabeth Ward, Jacques Ferlay, Otis Brawley, and Freddie Bray. International variation in prostate cancer incidence and mortality rates. *European urology*, 61(6):1079–1092, 2012.

[6] Jean-Luc Descotes. Diagnosis of prostate cancer. *Asian journal of urology*, 6(2):129–136, 2019.

[7] Claartje Gosselaar, Monique J Roobol, Stijn Roemeling, and Fritz H Schröder. The role of the digital rectal examination in subsequent screening visits in the european randomized study of screening for prostate cancer (erspc), rotterdam. *European urology*, 54(3):581–588, 2008.

[8] Daniel Jones, Charlotte Friend, Andreas Dreher, Victoria Allgar, and Una Macleod. The diagnostic test accuracy of rectal examination for prostate cancer diagnosis in symptomatic patients: a systematic review. *BMC family practice*, 19(1):1–6, 2018.

[9] RH Flocks, Vern C Urich, Chandulal A Patel, and John M Opitz. Studies on the antigenic properties of prostatic tissue i. *The Journal of Urology*, 84(1):134–143, 1960.

[10] Manabu Kuriyama, Ming C Wang, Lawrence D Papsidero, Carl S Killian, Takashi Shimano, Luis Valenzuela, Tsuneo Nishiura, Gerald P Murphy, and T Ming Chu. Quantitation of prostate-specific antigen in serum by a sensitive enzyme immunoassay. *Cancer research*, 40(12):4658–4662, 1980.

[11] Thomas A Stamey, Norman Yang, Alan R Hay, John E McNeal, Fuad S Freiha, and Elise Redwine. Prostate-specific antigen as a serum marker for adenocarcinoma of the prostate. *New England Journal of Medicine*, 317(15):909–916, 1987.

[12] JE Oesterling. Age-specific reference ranges for serum prostate-specific antigen. *The Canadian journal of urology*, 2(Supp1):23–29, 1995.

[13] Javier Hernández and Ian M Thompson. Prostate-specific antigen: a review of the validation of the most commonly used cancer biomarker. *Cancer*, 101(5):894–904, 2004.

[14] Amrith Raj Rao, Hanif G Motiwala, and Omer MA Karim. The discovery of prostate-specific antigen. *BJU international*, 101(1):5–10, 2008.

[15] Seamus J Teahan and Laurence H Klotz. Current role of prostate-specific antigen kinetics in managing patients with prostate cancer. *BJU international*, 97(3):451–455, 2006.

[16] Andrew J Vickers and Simon F Brewster. Psa velocity and doubling time in diagnosis and prognosis of prostate cancer. *British Journal of Medical and Surgical Urology*, 5(4):162–168, 2012.

[17] Sema Nur Ayyıldız and Ali Ayyıldız. Psa, psa derivatives, propsa and prostate health index in the diagnosis of prostate cancer. *Turkish journal of urology*, 40(2):82, 2014.

[18] Bob Djavan, Kathleen Nelson, Amir Kazzazi, Aron Bruhn, Helen Sadri, Alejandro Gomez-Pinillos, and Anna C Ferrari. Immunotherapy in the treatment of advanced prostate cancer. *The Canadian journal of urology*, 18(5):5865, 2011.

[19] Mary F McNaughton-Collins, Michael J Barry, et al. One man at a time,Äîresolving the psa controversy. *N Engl J Med*, 365(21):1951–1953, 2011.

[20] William J Catalona, Jerome P Richie, Frederick R Ahmann, M,ÄôLiss A Hudson, Peter T Scardino, Robert C Flanigan, Jean B DeKernion, Timothy L Ratliff, Louis R Kavoussi, Bruce L Dalkin, et al. Comparison of digital rectal examination and serum prostate specific antigen in the early detection of prostate cancer: results of a multicenter clinical trial of 6,630 men. *The Journal of urology*, 151(5):1283–1290, 1994.

[21] Carsten Stephan, Henning Cammann, Hellmuth-A Meyer, Michael Lein, and Klaus Jung. Psa and new biomarkers within multivariate models to improve early detection of prostate cancer. *Cancer letters*, 249(1):18–29, 2007.

[22] Leonard G Gomella, Xiaolong S Liu, Edouard J Trabulsi, Wm Kevin Kelly, Ronald Myers, Timothy Showalter, Adam Dicker, and Richard Wender. Screening for prostate cancer: the current evidence and guidelines controversy. *Canadian Journal of Urology*, 18(5):5875, 2011.

[23] Peter S Bunting. A guide to the interpretation of serum prostate specific antigen levels. *Clinical biochemistry*, 28(3):221–241, 1995.

[24] Stuart N Lloyd, Gerald N Collins, George B McKelvie, Michael Hehir, and Anthony CN Rogers. Predicted and actual change in serum psa following prostatectomy for bph. *Urology*, 43(4):472–479, 1994.

[25] Durwood E Neal Jr, Timothy D Moon, Sanda Clejan, and Deba Sarma. Prostate specific antigen and prostatitis i. effect of prostatitis on serum psa in the human and nonhuman primate. *The Prostate*, 20(2):105–111, 1992.

[26] William J Glenski, Reza S Malek, JAMES F Myrtle, and Joseph E Oesterling. Sustained, substantially increased concentration of prostate-specific antigen in the absence of prostatic malignant disease: an unusual clinical scenario. In *Mayo Clinic Proceedings*, volume 67, pages 249–252. Elsevier, 1992.

[27] Fritz H Schröder, Jonas Hugosson, Monique J Roobol, Teuvo LJ Tammela, Stefano Ciatto, Vera Nelen, Maciej Kwiatkowski, Marcos Lujan, Hans Lilja, Marco Zappa, et al. Prostate-cancer mortality at 11 years of follow-up. *New England Journal of Medicine*, 366(11):981–990, 2012.

[28] Andrew J Vickers, Angel M Cronin, Gunnar Aus, Carl-Gustav Pihl, Charlotte Becker, Kim Pettersson, Peter T Scardino, Jonas Hugosson, and Hans Lilja. A panel of kallikrein markers can reduce unnecessary biopsy for prostate cancer: data from the european randomized study of prostate cancer screening in göteborg, sweden. *BMC medicine*, 6(1):19, 2008.

[29] Stacy Loeb, H Ballentine Carter, Sonja I Berndt, Winnie Ricker, and Edward M Schaeffer. Complications after prostate biopsy: data from seer-medicare. *The Journal of urology*, 186(5):1830–1834, 2011.

[30] Oluwole Fadare, Sa Wang, and M Rajan Mariappan. Practice patterns of clinicians following isolated diagnoses of atypical small acinar proliferation on prostate biopsy specimens. *Archives of pathology & laboratory medicine*, 128(5):557–560, 2004.

[31] Greta Del Mistro, Silvia Cervo, Elena Mansutti, Riccardo Spizzo, Alfonso Colombatti, Pietro Belmonte, Renzo Zucconelli, Agostino Steffan, Valter

Sergo, and Alois Bonifacio. Surface-enhanced raman spectroscopy of urine for prostate cancer detection: a preliminary study. *Analytical and Bioanalytical Chemistry*, 407(12):3271–3275, 2015.

[32] Veeru Kasivisvanathan, Antti S Rannikko, Marcelo Borghi, Valeria Panebianco, Lance A Mynderse, Markku H Vaarala, Alberto Briganti, Lars Budäus, Giles Hellawell, Richard G Hindley, et al. Mri-targeted or standard biopsy for prostate-cancer diagnosis. *New England Journal of Medicine*, 378(19):1767–1777, 2018.

[33] Sangjun Yoo, Se Young Choi, Dalsan You, and Choung-Soo Kim. New drugs in prostate cancer. *Prostate international*, 4(2):37–42, 2016.

[34] Martin CS Wong, William B Goggins, Harry HX Wang, Franklin DH Fung, Colette Leung, Samuel YS Wong, Chi Fai Ng, and Joseph JY Sung. Global incidence and mortality for prostate cancer: analysis of temporal patterns and trends in 36 countries. *European urology*, 70(5):862–874, 2016.

[35] Ismail Kola and John Landis. Can the pharmaceutical industry reduce attrition rates? *Nature reviews Drug discovery*, 3(8):711–716, 2004.

[36] Alberto Ocana, Atanasio Pandiella, Lillian L Siu, and Ian F Tannock. Preclinical development of molecular-targeted agents for cancer. *Nature reviews Clinical oncology*, 8(4):200–209, 2011.

[37] Bryan T Hennessy, Debra L Smith, Prahlad T Ram, Yiling Lu, and Gordon B Mills. Exploiting the pi3k/akt pathway for cancer drug discovery. *Nature reviews Drug discovery*, 4(12):988–1004, 2005.

[38] Jean-Pierre Gillet, Sudhir Varma, and Michael M Gottesman. The clinical relevance of cancer cell lines. *Journal of the National Cancer Institute*, 105(7):452–458, 2013.

[39] Inge Seim, Penny L Jeffery, Patrick B Thomas, Colleen C Nelson, and Lisa K Chopin. Whole-genome sequence of the metastatic pc3 and lncap human prostate cancer cell lines. *G3: Genes, Genomes, Genetics*, 7(6):1731–1741, 2017.

[40] RE Sobel and MD Sadar. Cell lines used in prostate cancer research: a compendium of old and new lines-part 1. *The Journal of urology*, 173(2):342–359, 2005.

[41] Amos Bairoch. The cellosaurus, a cell-line knowledge resource. *Journal of biomolecular techniques: JBT*, 29(2):25, 2018.

[42] John RW Masters. Human cancer cell lines: fact and fantasy. *Nature reviews Molecular cell biology*, 1(3):233–236, 2000.

[43] Sarah E Burdall, Andrew M Hanby, Mark RJ Lansdown, and Valerie Speirs. Breast cancer cell lines: friend or foe? *Breast cancer research*, 5(2):1–7, 2003.

[44] Norman J Maitland, Catherine A Macintosh, Christian Schmitz, and Shona H Lang. Immortalization of human prostate cells with the human papillomavirus type 16 e6 gene. In *Cancer Cell Culture*, pages 275–285. Springer, 2004.

[45] C Kent Osborne, Kimberly Hobbs, and Jeffrey M Trent. Biological differences among mcf-7 human breast cancer cell lines from different laboratories. *Breast cancer research and treatment*, 9(2):111–121, 1987.

[46] Olivera Markovic and Nenad Markovic. Cell cross-contamination in cell cultures: the silent and neglected danger. *In Vitro Cellular & Developmental Biology-Animal*, 34(1):1–8, 1998.

[47] Roderick AF MacLeod, Wilhelm G Dirks, Yoshinobu Matsuo, Maren Kaufmann, Herbert Milch, and Hans G Drexler. Widespread intraspecies cross-contamination of human tumor cell lines arising at source. *International Journal of Cancer*, 83(4):555–563, 1999.

[48] Norman J Maitland. Getting closer to prostate cancer in patients-what scientists should want from clinicians. *J Cancer Metastasis Treat*, 3:262–70, 2017.

[49] JS Horoszewicz, SS Leong, T Ming Chu, ZL Wajsman, M Friedman, L Papsidero, U Kim, LS Chai, S Kakati, SK Arya, et al. The lncap cell line–a new model for studies on human prostatic carcinoma. *Progress in clinical and biological research*, 37:115–132, 1980.

[50] Jennifer L Wilding and Walter F Bodmer. Cancer cell lines for drug discovery and development. *Cancer research*, 74(9):2377–2384, 2014.

[51] Yunxin Lai, Xinru Wei, Shouheng Lin, Le Qin, Lin Cheng, and Peng Li. Current status and perspectives of patient-derived xenograft models in cancer research. *Journal of hematology & oncology*, 10(1):1–14, 2017.

[52] Lisa Hutchinson and Rebecca Kirk. High drug attrition rates-where are we going wrong? *Nature reviews Clinical oncology*, 8(4):189–190, 2011.

[53] Gerry Northam. How cancer studies wasted cash. *BBC News Channel. on*, 4:7104070, 2007.

[54] Fiona M Frame, Amanda R Noble, Sandra Klein, Hannah F Walker, Rakesh Suman, Richard Kasprowicz, Vin M Mann, Matt S Simms, and Norman J Maitland. Tumor heterogeneity and therapy resistance-implications for future treatments of prostate cancer. *J. Cancer Metastasis*, 3:302, 2017.

[55] Heather Payne and Philip Cornford. Prostate-specific antigen: an evolving role in diagnosis, monitoring, and treatment evaluation in prostate cancer. In *Urologic Oncology: Seminars and Original Investigations*, volume 29, pages 593–601. Elsevier, 2011.

[56] Chandrasekhara Venkata Raman and Kariamanikkam Srinivasa Krishnan. A new type of secondary radiation. *Nature*, 121(3048):501–502, 1928.

[57] Geeta 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(10):1460–1464, 2006.

[58] Dongqi Wang, Peisheng He, Zijian Wang, Guangyu Li, Nehreen Majed, and April Z Gu. Advances in single cell raman spectroscopy technologies for bio-

logical and environmental applications. *Current Opinion in Biotechnology*, 64:218–229, 2020.

[59] PRGDJ Graves and D Gardiner. Practical raman spectroscopy. *Springer*, 1989.

[60] Ewen Smith and Geoffrey Dent. Modern raman spectroscopy: a practical approach. 2005.

[61] Mengqiu Li, Jian Xu, Maria Romero-Gonzalez, Steve A Banwart, and Wei E Huang. Single cell raman spectroscopy for cell sorting and imaging. *Current opinion in biotechnology*, 23(1):56–63, 2012.

[62] Derek Albert Long. Raman spectroscopy. *New York*, pages 1–12, 1977.

[63] 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(12):2166–2170, 2005.

[64] Mariana C Potcoava, Gregory L Futia, Jessica Aughenbaugh, Isabel R Schlaepfer, and Emily A 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(11):111605, 2014.

[65] LaTayia Aaron, Omar E Franco, and Simon W Hayward. Review of prostate anatomy and embryology and the etiology of benign prostatic hyperplasia. *Urologic Clinics*, 43(3):279–288, 2016.

[66] Kulkarni Prakash, Gregorio Pirozzi, Michael Elashoff, William Munger, Iwao Waga, Rajiv Dhir, Yoshiyuki Kakehi, and Robert H Getzenberg. Symptomatic and asymptomatic benign prostatic hyperplasia: molecular differentiation by using microarrays. *Proceedings of the National Academy of Sciences*, 99(11):7598–7603, 2002.

[67] B Raychaudhuri and D Cahill. Pelvic fasciae in urology. *The Annals of The Royal College of Surgeons of England*, 90(8):633–637, 2008.

[68] Elaine Nicpon Marieb and Katja Hoehn. *Human anatomy & physiology*. Pearson education, 2007.

[69] John E McNeal. The zonal anatomy of the prostate. *The prostate*, 2(1):35–49, 1981.

[70] Angelo M De Marzo, Elizabeth A Platz, Siobhan Sutcliffe, Jianfeng Xu, Henrik Grönberg, Charles G Drake, Yasutomo Nakai, William B Isaacs, and William G Nelson. Inflammation in prostate carcinogenesis. *Nature Reviews Cancer*, 7(4):256–269, 2007.

[71] Gervaise H Henry, Alicia Malewska, Diya B Joseph, Venkat S Malladi, Jeon Lee, Jose Torrealba, Ryan J Mauck, Jeffrey C Gahan, Ganesh V Raj, Claus G Roehrborn, et al. A cellular anatomy of the normal adult human prostate and prostatic urethra. *Cell reports*, 25(12):3530–3542, 2018.

[72] Charles Huggins, William W Scott, and J Henry Heinen. Chemical composition of human semen and of the secretions of the prostate and seminal vesicles. *American Journal of Physiology-Legacy Content*, 136(3):467–473, 1942.

[73] Jürgen H Gross. *Mass spectrometry: a textbook*. Springer Science & Business Media, 2006.

[74] Lars Jarup, Nicky Best, Mireille B Toledano, Jon Wakefield, and Paul Elliott. Geographical epidemiology of prostate cancer in great britain. *International journal of cancer*, 97(5):695–699, 2002.

[75] E David Crawford. Epidemiology of prostate cancer. *Urology*, 62(6):3–12, 2003.

[76] Ali Ziada, Mark Rosenblum, and E David Crawford. Benign prostatic hyperplasia: an overview. *Urology*, 53(3):1–6, 1999.

[77] A Thorpe and D Neal. Benign prostatic hyperplasia. *The Lancet*, 361(9366):1359–1367, 2003.

[78] Kevin T McVary. Bph: epidemiology and comorbidities. *The American journal of managed care*, 12(5 Suppl):S122–8, 2006.

[79] Wade Bushman. Etiology, epidemiology, and natural history. *Urologic Clinics*, 36(4):403–415, 2009.

[80] XI Wang, Marianna Kruithof-de Julio, Kyriakos D Economides, David Walker, Hailong Yu, M Vivienne Halili, Ya-Ping Hu, Sandy M Price, Cory Abate-Shen, and Michael M Shen. A luminal epithelial stem cell that is a cell of origin for prostate cancer. *Nature*, 461(7263):495–500, 2009.

[81] Paolo Verze, Tommaso Cai, and Stefano Lorenzetti. The role of the prostate in male fertility, health and disease. *Nature Reviews Urology*, 13(7):379, 2016.

[82] Norman J Maitland, Fiona M Frame, Euan S Polson, John L Lewis, and Anne T Collins. Prostate cancer stem cells: do they have a basal or luminal phenotype? *Hormones and Cancer*, 2(1):47–61, 2011.

[83] Tsai-Ling Lu, Yi-Fen Huang, Li-Ru You, Nai-Chen Chao, Fang-Yi Su, Junn-Liang Chang, and Chun-Ming Chen. Conditionally ablated pten in prostate basal cells promotes basal-to-luminal differentiation and causes invasive prostate cancer in mice. *The American journal of pathology*, 182(3):975–991, 2013.

[84] Zhu A Wang, Roxanne Toivanen, Sarah K Bergren, Pierre Chambon, and Michael M Shen. Luminal cells are favored as the cell of origin for prostate cancer. *Cell reports*, 8(5):1339–1346, 2014.

[85] Dingxiao Zhang, Shuhong Zhao, Xinyun Li, Jason S Kirk, and Dean G Tang. Prostate luminal progenitor cells in development and cancer. *Trends in cancer*, 4(11):769–783, 2018.

[86] Devon A Lawson, Yang Zong, Sanaz Memarzadeh, Li Xin, Jiaoti Huang, and Owen N Witte. Basal epithelial stem cells are efficient targets for prostate cancer initiation. *Proceedings of the National Academy of Sciences*, 107(6):2610–2615, 2010.

[87] Oh-Joon Kwon, Li Zhang, Michael M Ittmann, and Li Xin. Prostatic inflammation enhances basal-to-luminal differentiation and accelerates initiation of prostate cancer with a basal cell origin. *Proceedings of the National Academy of Sciences*, 111(5):E592–E600, 2014.

[88] Matthew Bui and Robert E Reiter. Stem cell genes in androgen-independent prostate cancer. *Cancer and Metastasis Reviews*, 17(4):391–399, 1998.

[89] Jacques Ferlay, Isabelle Soerjomataram, Rajesh Dikshit, Sultan Eser, Colin Mathers, Marise Rebelo, Donald Maxwell Parkin, David Forman, and Freddie Bray. Cancer incidence and mortality worldwide: sources, methods and major patterns in globocan 2012. *International journal of cancer*, 136(5):E359–E386, 2015.

[90] Gati K Panigrahi, Prakash P Praharaj, Hiroki Kittaka, Asit R Mridha, Olen M Black, Rakesh Singh, Roger Mercer, Adrie Van Bokhoven, Kathleen C Torkko, Chapla Agarwal, et al. Exosome proteomic analyses identify inflammatory phenotype and novel biomarkers in african american prostate cancer patients. *Cancer medicine*, 8(3):1110–1123, 2019.

[91] Prashanth Rawla. Epidemiology of prostate cancer. *World journal of oncology*, 10(2):63, 2019.

[92] Subas Neupane, Freddie Bray, and Anssi Auvinen. National economic and development indicators and international variation in prostate cancer incidence and mortality: an ecological analysis. *World journal of urology*, 35(6):851–858, 2017.

[93] David S Parda, Peter J Thraves, Michael R Kuettel, Myeong-Seon Lee, Paul Arnstein, M Edward Kaighn, Johng S Rhim, and Anatoly Dritschilo.

Neoplastic transformation of a human prostate epithelial cell line by the v-ki-ras oncogene. *The Prostate*, 23(2):91–98, 1993.

[94] O Cussenot, Ph Berthon, R Berger, I Mowszowicz, A Faille, F Hojman, P Teillac, A Le Duc, and F Calvo. Immortalization of human adult normal prostatic epithelial cells by liposomes containing large t-sv40 gene. *The Journal of urology*, 146(3):881–886, 1991.

[95] J Miki and JS Rhim. Prostate cell cultures as in vitro models for the study of normal stem cells and cancer stem cells. *Prostate cancer and prostatic diseases*, 11(1):32–39, 2008.

[96] Reza Izadpanah, Deepak Kaushal, Christopher Kriedt, Fern Tsien, Bindiya Patel, Jason Dufour, and Bruce A Bunnell. Long-term in vitro expansion alters the biology of adult mesenchymal stem cells. *Cancer research*, 68(11):4229–4238, 2008.

[97] Svend A Engelholm, Lars L Vindeløv, Mogens Spang-Thomsen, Nils Brünner, Niels Tommerup, Morten H Nielsen, and Heine H Hansen. Genetic instability of cell lines derived from a single human small cell carcinoma of the lung. *European Journal of Cancer and Clinical Oncology*, 21(7):815–824, 1985.

[98] Robert L Margolis. Tetraploidy and tumor development. *Cancer cell*, 8(5):353–354, 2005.

[99] DM Peehl. Primary cell cultures as models of prostate cancer development. *Endocrine-related cancer*, 12(1):19–47, 2005.

[100] Thomas Tyrel Goodman, Chee Ping Ng, and Suzie Hwang Pun. 3-d tissue culture systems for the evaluation and optimization of nanoparticle-based drug carriers. *Bioconjugate chemistry*, 19(10):1951–1959, 2008.

[101] Abdullah Chandra Sekhar Talari, Ahtasham Raza, Shazza Rehman, and Ihtesham U Rehman. Analyzing normal proliferating, hypoxic and necrotic regions of t-47d human breast cancer spheroids using raman spectroscopy. *Applied Spectroscopy Reviews*, 52(10):909–924, 2017.

[102] Mine Altunbek, Demet Çetin, Zekiye Suludere, and Mustafa Çulha. Surface-enhanced raman spectroscopy based 3d spheroid culture for drug discovery studies. *Talanta*, 191:390–399, 2019.

[103] Abhilash K Pandya, Gulay K Serhatkulu, Alex Cao, Rachel E Kast, Houbei Dai, Raja Rabah, Janet Poulik, Sanjeev Banerjee, Ratna Naik, Volkan Adsay, et al. Evaluation of pancreatic cancer with raman spectroscopy in a mouse model. *Pancreas*, 36(2):e1–e8, 2008.

[104] Kuniaki Tanahashi, Atsushi Natsume, Fumiharu Ohka, Hiroyuki Momota, Akira Kato, Kazuya Motomura, Naoki Watabe, Shuichi Muraishi, Hitoshi Nakahara, Yahachi Saito, et al. Assessment of tumor cells in a mouse model of diffuse infiltrative glioma by raman spectroscopy. *BioMed research international*, 2014, 2014.

[105] James W Chan, Douglas S Taylor, Theodore Zwerdling, Stephen M Lane, Ko Ihara, and Thomas Huser. Micro-raman spectroscopy detects individual neoplastic and normal hematopoietic cells. *Biophysical journal*, 90(2):648–656, 2006.

[106] Adolf Smekal. Zur quantentheorie der dispersion. *Naturwissenschaften*, 11(43):873–875, 1923.

[107] Robin R Jones, David C Hooper, Liwu Zhang, Daniel Wolverson, and Ventsislav K Valev. Raman techniques: fundamentals and frontiers. *Nanoscale research letters*, 14(1):1–34, 2019.

[108] Olaf Hollricher. Raman instrumentation for confocal raman microscopy. In *Confocal Raman Microscopy*, pages 43–60. Springer, 2010.

[109] Derek Albert Long and DA Long. The raman effect: a unified treatment of the theory of raman scattering by molecules. 2002.

[110] Sapun H Parekh, Young Jong Lee, Khaled A Aamer, and Marcus T Cicerone. Label-free cellular imaging by broadband coherent anti-stokes raman scattering microscopy. *Biophysical journal*, 99(8):2695–2704, 2010.

[111] Holly J Butler, Lorna Ashton, Benjamin Bird, Gianfelice Cinque, Kelly Curtis, Jennifer Dorney, Karen Esmonde-White, Nigel J Fullwood, Benjamin Gardner, Pierre L Martin-Hirsch, et al. Using raman spectroscopy to characterize biological materials. *Nature protocols*, 11(4):664, 2016.

[112] Eliana Cordero, Ines Latka, Christian Matthaus, Iwan W Schie, and Jurgen Popp. In-vivo raman spectroscopy: from basics to applications. *Journal of biomedical optics*, 23(7):071210, 2018.

[113] Jill A Macoska, Mark A Micale, Wael A Sakr, Philip D Benson, and Sandra R Wolman. Extensive genetic alterations in prostate cancer revealed by dual pcr and fish analysis. *Genes, Chromosomes and Cancer*, 8(2):88–97, 1993.

[114] Yilong Sun, Jian Lin, Aaron E Katz, and Paul B Fisher. Human prostatic carcinoma oncogene pti-1 is expressed in human tumor cell lines and prostate carcinoma patient blood samples. *Cancer research*, 57(1):18–23, 1997.

[115] RH Barker Jr. Use of pcr in the field. *Parasitology Today*, 10(3):117–119, 1994.

[116] Howard M Shapiro. *Practical flow cytometry*. John Wiley & Sons, 2005.

[117] Robert A Stephenson, Brent C James, Helen Gay, William R Fair, Willet F Whitmore, and Myron R Melamed. Flow cytometry of prostate cancer: relationship of dna content to survival. *Cancer research*, 47(9):2504–2507, 1987.

[118] Michael L Friedlander, David W Hedley, Ian W Taylor, Peter Russell, Alan S Coates, and Martin HN Tattersall. Influence of cellular dna content on survival in advanced ovarian cancer. *Cancer Research*, 44(1):397–400, 1984.

[119] Lisa H Cazares, Bao-Ling Adam, Michael D Ward, Suhail Nasim, Paul F Schellhammer, O John Semmes, and George L Wright. Normal, benign, preneoplastic, and malignant prostate cells have distinct protein expression profiles resolved by surface enhanced laser desorption/ionization mass spectrometry. *Clinical Cancer Research*, 8(8):2541–2552, 2002.

[120] Stephen A Bustin. How to speed up the polymerase chain reaction. *Biomolecular detection and quantification*, 12:10–14, 2017.

[121] J Nicholas Housby. *Mass spectrometry and genomic analysis*, volume 2. Springer Science & Business Media, 2001.

[122] James V Watson. *Introduction to flow cytometry*. Cambridge University Press, 2004.

[123] James W Chan, Douglas S Taylor, Stephen M Lane, Theodore Zwerdling, Joseph Tuscano, and Thomas Huser. Nondestructive identification of individual leukemia cells by laser trapping raman spectroscopy. *Analytical chemistry*, 80(6):2180–2187, 2008.

[124] Ehsan Gazi, John Dwyer, Peter Gardner, A Ghanbari-Siahkali, AP Wade, J Miyan, Nicholas P Lockyer, John C Vickerman, Noel W Clarke, Jonathan H Shanks, et al. Applications of fourier transform infrared microspectroscopy in studies of benign prostate and prostate cancer. a pilot study. *The Journal of Pathology: A Journal of the Pathological Society of Great Britain and Ireland*, 201(1):99–108, 2003.

[125] I Notingher, JR Jones, S Verrier, I Bisson, P Embanga, P Edwards, JM Polak, and LL Hench. Application of ftir and raman spectroscopy to characterisation of bioactive materials and living cells. *Spectroscopy*, 17(2, 3):275–288, 2003.

[126] Christian Matthäus, Benjamin Bird, Miloš Miljković, Tatyana Chernenko, Melissa Romeo, and Max Diem. Infrared and raman microscopy in cell biology. *Methods in cell biology*, 89:275–308, 2008.

[127] Rachael Smith, Karen L Wright, and Lorna Ashton. Raman spectroscopy: an evolving technique for live cell studies. *Analyst*, 141(12):3590–3600, 2016.

[128] Zanyar Movasaghi, Shazza Rehman, and Ihtesham U Rehman. Raman spectroscopy of biological tissues. *Applied Spectroscopy Reviews*, 42(5):493–541, 2007.

[129] Abdullah Chandra Sekhar Talari, Zanyar Movasaghi, Shazza Rehman, and Ihtesham Ur Rehman. Raman spectroscopy of biological tissues. *Applied Spectroscopy Reviews*, 50(1):46–111, 2015.

[130] Masaya Okada, Nicholas Isaac Smith, Almar Flotildes Palonpon, Hiromi Endo, Satoshi Kawata, Mikiko Sodeoka, and Katsumasa Fujita. Label-free raman observation of cytochrome c dynamics during apoptosis. *Proceedings of the National Academy of Sciences*, 109(1):28–32, 2012.

[131] John William Strutt Baron Rayleigh. *On the scattering of light by small particles*. 1871.

[132] Gustav Mie. Pioneering mathematical description of scattering by spheres. *Ann. Phys*, 25(337):0003–3804, 1908.

[133] Andrew T Young. Rayleigh scattering. *Phys. Today*, 35(1):42–48, 1982.

[134] David J Lockwood. Rayleigh and mie scattering. *Encyclopedia of Color Science and Technology, Springer, New York, NY*, pages 1–12, 2016.

[135] Cécile Malgrange, Christian Ricolleau, and Michel Schlenker. Second-rank tensors. In *Symmetry and Physical Properties of Crystals*, pages 205–223. Springer, 2014.

[136] Peter Vandenabeele. *Practical Raman spectroscopy: an introduction*. Wiley Online Library, 2013.

[137] Richard L McCreery. *Raman spectroscopy for chemical analysis*, volume 225. John Wiley & Sons, 2005.

[138] John R Ferraro. *Introductory raman spectroscopy*. Elsevier, 2003.

[139] Dustin W Shipp, Faris Sinjab, and Ioan Notingher. Raman spectroscopy: techniques and applications in the life sciences. *Advances in Optics and Photonics*, 9(2):315–428, 2017.

[140] Claudia Nieva, Monica Marro, Naiara Santana-Codina, Satish Rao, Dmitri Petrov, and Angels Sierra. The lipid phenotype of breast cancer cells charac-

terized by raman microspectroscopy: towards a stratification of malignancy. *PloS one*, 7(10):e46456, 2012.

[141] P Crow, N Stone, CA Kendall, JS Uff, JAM Farmer, H Barr, and MPJ Wright. The use of raman spectroscopy to identify and grade prostatic adenocarcinoma in vitro. *British journal of cancer*, 89(1):106–108, 2003.

[142] Ahcene Taleb, James Diamond, John J McGarvey, J Renwick Beattie, Carla Toland, and Peter W Hamilton. Raman microscopy for the chemometric analysis of tumor cells. *The Journal of Physical Chemistry B*, 110(39):19625–19631, 2006.

[143] Patrizio Candeloro, Elisabetta Grande, Raffaella Raimondo, Daniele Di Mascolo, Francesco Gentile, Maria Laura Coluccio, Gerardo Perozziello, Natalia Malara, Marco Francardi, and Enzo Di Fabrizio. Raman database of amino acids solutions: a critical study of extended multiplicative signal correction. *Analyst*, 138(24):7331–7340, 2013.

[144] Vanessa Russo, Patrizio Candeloro, Natalia Malara, Gerardo Perozziello, Michelangelo Iannone, Miriam Scicchitano, Rocco Mollace, Vincenzo Musolino, Micaela Gliozzi, Cristina Carresi, et al. Key role of cytochrome c for apoptosis detection using raman microimaging in an animal model of brain ischemia with insulin treatment. *Applied spectroscopy*, 73(10):1208–1217, 2019.

[145] Karl Pearson. Principal components analysis. *The London, Edinburgh, and Dublin Philosophical Magazine and Journal of Science*, 6(2):559, 1901.

[146] Harold Hotelling. Analysis of a complex of statistical variables into principal components. *Journal of educational psychology*, 24(6):417, 1933.

[147] Ian T Joliffe and BJT Morgan. Principal component analysis and exploratory factor analysis. *Statistical methods in medical research*, 1(1):69–95, 1992.

[148] Markus Ringnér. What is principal component analysis? *Nature biotechnology*, 26(3):303–304, 2008.

[149] Shruti Sehgal, Harpreet Singh, Mohit Agarwal, V Bhasker, et al. Data analysis using principal component analysis. In *2014 international conference on medical imaging, m-health and emerging communication systems (MedCom)*, pages 45–48. IEEE, 2014.

[150] Ian T Jolliffe and Jorge Cadima. Principal component analysis: a review and recent developments. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, 374(2065):20150202, 2016.

[151] Laurinda FS Siqueira, Raimundo F Araújo Júnior, Aurigena Antunes de Araújo, Camilo LM Morais, and Kássio MG Lima. Lda vs. qda for ft-mir prostate cancer tissue classification. *Chemometrics and Intelligent Laboratory Systems*, 162:123–129, 2017.

[152] James Joyce. Bayes' theorem. 2003.

[153] Richard O Duda, Peter E Hart, and David G Stork. *Pattern classification and scene analysis*, volume 3. Wiley New York, 1973.

[154] Alaa Tharwat. Linear vs. quadratic discriminant analysis classifier: a tutorial. *International Journal of Applied Pattern Recognition*, 3(2):145–180, 2016.

[155] Ronald A Fisher. The use of multiple measurements in taxonomic problems. *Annals of eugenics*, 7(2):179–188, 1936.

[156] James W Chan, Deborah K Lieu, Thomas Huser, and Ronald A Li. Label-free separation of human embryonic stem cells and their cardiac derivatives using raman spectroscopy. *Analytical chemistry*, 81(4):1324–1331, 2009.

[157] Pnt2 cell line human $cb_9$5012613.

[158] Steven L Gonias, Todd A Hembrough, and Mauricio Sankovic. Cytokeratin 8 functions as a major plasminogen receptor in select epithelial and carcinoma cells. *Front Biosci*, 6(3):D1403–D1411, 2001.

[159] PHILIPPE BERTHON, OLIVIER CUSSENOT, LUCY HOPWOOD, ALAIN LEDUC, and NORMAN J MAITLAND. Functional expression of sv40 in normal human prostatic epithelial and fibroblastic cells-differentiation pattern of nontumorigenic cell-lines. *International journal of oncology*, 6(2):333–343, 1995.

[160] SH Lang, M Stower, and NJ Maitland. In vitro modelling of epithelial and stromal interactions in non-malignant and malignant prostates. *British journal of cancer*, 82(4):990–997, 2000.

[161] Yuzhuo Wang, Simon W Hayward, Mei Cao, Kristina A Thayer, and Gerald R Cunha. Cell differentiation lineage in the prostate. *Differentiation*, 68(4-5):270–279, 2001.

[162] Roger S Kirby, John D McConnell, John M Fitzpatrick, Claus G Roehrborn, and Peter Boyle. *Textbook of benign prostatic hyperplasia*. CRC Press, 2004.

[163] Sonny Schelin. *Development of Feedback Microwave Thermotherapy in Symptomatic Benign Prostatic Hyperplasia*. Number 25. Lund University, 2006.

[164] Julie Dutil, Zhihua Chen, Alvaro N Monteiro, Jamie K Teer, and Steven A Eschrich. An interactive resource to probe genetic diversity and estimated ancestry in cancer cell lines. *Cancer research*, 79(7):1263–1273, 2019.

[165] SW Hayward, R Dahiya, GR Cunha, J Bartek, N Deshpande, and P Narayan. Establishment and characterization of an immortalized but non-transformed human prostate epithelial cell line: Bph-1. *In Vitro Cellular & Developmental Biology-Animal*, 31(1):14–24, 1995.

[166] Web. German collection of microorganisms and cell cultures gmbh: details, BPH-1. `https://www.dsmz.de/collection/catalogue/details/culture/ACC-143/`, 2021. [Online; accessed 31-DEC-2021].

[167] Julius S Horoszewicz, Susan S Leong, Elzbieta Kawinski, James P Karr, Hannah Rosenthal, T Ming Chu, Edwin A Mirand, and Gerald P Murphy.

Lncap model of human prostatic carcinoma. *Cancer research*, 43(4):1809–1818, 1983.

[168] Tsukasa Igawa, Fen-Fen Lin, Ming-Shyue Lee, Dev Karan, Surinder K Batra, and Ming-Fong Lin. Establishment and characterization of androgen-independent human prostate cancer lncap cell model. *The Prostate*, 50(4):222–235, 2002.

[169] Saleh Altuwaijri, Hui-Kuan Lin, Kuang-Hsiang Chuang, Wen-Jye Lin, Shuyuan Yeh, LeRoy A Hanchett, Mujib M Rahman, Hong-Yo Kang, Meng-Ying Tsai, Yanqing Zhang, et al. Interruption of nuclear factor κb signaling by the androgen receptor facilitates 12-o-tetradecanoylphorbolacetate-induced apoptosis in androgen-sensitive prostate cancer lncap cells. *Cancer Research*, 63(21):7106–7112, 2003.

[170] Pamela J Russell, Paul Jackson, and Elizabeth A Kingsley. *Prostate cancer methods and protocols*, volume 81. Springer, 2003.

[171] Raymond B Nagle, Frederick R Ahmann, Kathleen M McDaniel, Maria L Paquin, Virginia A Clark, and Abbie Celniker. Cytokeratin characterization of human prostatic carcinoma and its derived cell lines. *Cancer research*, 47(1):281–286, 1987.

[172] Kenneth R. Stone, Don D. Mickey, Heidi Wunderli, George H. Mickey, and David F. Paulson. Isolation of a human prostate carcinoma cell line (du 145). *International Journal of Cancer*, 21(3):274–281, 1978.

[173] Fatouma Alimirah, Jianming Chen, Zane Basrawala, Hong Xin, and Divaker Choubey. Du-145 and pc-3 human prostate cancer cell lines express androgen receptor: Implications for the androgen receptor functions and regulation. *FEBS letters*, 580(9):2294–2300, 2006.

[174] ME Kaighn, K Shankar Narayan, Y Ohnuki, JF Lechner, and LW Jones. Establishment and characterization of a human prostatic carcinoma cell line (pc-3). *Investigative urology*, 17(1):16, 1979.

[175] David F Jarrard, Barry F Blitz, Robert C Smith, Barbara L Patai, and Daniel B Rukstalis. Effect of epidermal growth factor on prostate cancer cell line pc3 growth and invasion. *The Prostate*, 24(1):46–53, 1994.

[176] Marileila Varella-Garcia, Theresa Boomer, and Gary J Miller. Karyotypic similarity identified by multiplex-fish relates four prostate adenocarcinoma cell lines: Pc-3, ppc-1, alva-31, and alva-41. *Genes, Chromosomes and Cancer*, 31(4):303–315, 2001.

[177] Sheng Tai, Yin Sun, Jill M Squires, Hong Zhang, William K Oh, Chao-Zhao Liang, and Jiaoti Huang. Pc3 is a cell line characteristic of prostatic small cell carcinoma. *The Prostate*, 71(15):1668–1679, 2011.

[178] Marianne D Sadar. Androgen-independent induction of prostate-specific antigen gene expression via cross-talk between the androgen receptor and protein kinase a signal transduction pathways. *Journal of Biological Chemistry*, 274(12):7777–7783, 1999.

[179] Edward R Sherwood, Lori A Berg, Nancy J Mitchell, John E McNeal, James M Kozlowski, and Chung Lee. Differential cytokeratin expression in normal, hyperplastic and malignant epithelial cells from human prostate. *The Journal of urology*, 143(1):167–171, 1990.

[180] Norman J Maitland, Catherine A Macintosh, June Hall, Mike Sharrard, Gary Quinn, and Shona Lang. In vitro models to study cellular differentiation and function in human prostate cancers. *Radiation research*, 155(1):133–142, 2001.

[181] Shona H Lang, Joel Smith, Catherine Hyde, Catherine Macintosh, Michael Stower, and Norman J Maitland. Differentiation of prostate epithelial cellcultures by materigel/stromal cell glandular reconstruction. *In Vitro Cellular & Developmental Biology-Animal*, 42(8-9):273–280, 2006.

[182] NJ Maitland, SD Bryce, MJ Stower, and AT Collins. Prostate cancer stem cells: a target for new therapies. In *Cancer Stem Cells*, pages 155–179. Springer, 2007.

[183] RE Sobel and MD Sadar. Cell lines used in prostate cancer research: a compendium of old and new lines-part 2. *The Journal of urology*, 173(2):360–372, 2005.

[184] Christian D Toouli, Lily I Huschtscha, Axel A Neumann, Jane R Noble, Lorel M Colgin, Bharati Hukku, and Roger R Reddel. Comparison of human mammary epithelial cells immortalized by simian virus 40 t-antigen or by the telomerase catalytic subunit. *Oncogene*, 21(1):128–139, 2002.

[185] D. F. GLEASON. Classification of prostatic carcinomas. *Cancer Chemother. Rep.*, 50:125–128, 1966.

[186] Jonathan I Epstein. An update of the gleason grading system. *The Journal of urology*, 183(2):433–440, 2010.

[187] Jonathan I Epstein, William C Allsbrook Jr, Mahul B Amin, Lars L Egevad, ISUP Grading Committee, et al. The 2005 international society of urological pathology (isup) consensus conference on gleason grading of prostatic carcinoma. *The American journal of surgical pathology*, 29(9):1228–1242, 2005.

[188] Jonathan I Epstein, Lars Egevad, Mahul B Amin, Brett Delahunt, John R Srigley, and Peter A Humphrey. The 2014 international society of urological pathology (isup) consensus conference on gleason grading of prostatic carcinoma. *The American journal of surgical pathology*, 40(2):244–252, 2016.

[189] William J Catalona and Willet F Whitmore. New staging systems for prostate cancer, 1989.

[190] Joseph E Oesterling, Vera Jean Suman, Horst Zincke, and David G Bostwick. Psa-detected (clinical stage t1c or b0) prostate cancer: pathologically significant tumors. *Urologic Clinics of North America*, 20(4):687–693, 1993.

[191] Ni Chen and Qiao Zhou. The evolving gleason grading system. *Chinese Journal of Cancer Research*, 28(1):58, 2016.

[192] Brett Delahunt, Rose J Miller, John R Srigley, Andrew J Evans, and Hemamali Samaratunga. Gleason grading: past, present and future. *Histopathology*, 60(1):75–86, 2012.

[193] Peter A Humphrey. Gleason grading and prognostic factors in carcinoma of the prostate. *Modern pathology*, 17(3):292–306, 2004.

[194] Lori E Kamemoto, Anupam K Misra, Shiv K Sharma, Marc T Goodman, Hugh Luk, Ava C Dykes, and Tayro Acosta. Near-infrared micro-raman spectroscopy for in vitro detection of cervical cancer. *Applied spectroscopy*, 64(3):255–261, 2010.

[195] Zhiwei Huang, Annette McWilliams, Stephen Lam, John English, David I McLean, Harvey Lui, and Haishan Zeng. Effect of formalin fixation on the near-infrared raman spectroscopy of normal and cancerous human bronchial tissues. *International journal of oncology*, 23(3):649–655, 2003.

[196] Valeriu Vrabie, Cyril Gobinet, Olivier Piot, Ali Tfayli, Philippe Bernard, Régis Huez, and Michel Manfait. Independent component analysis of raman spectra: Application on paraffin-embedded skin biopsies. *Biomedical Signal Processing and Control*, 2(1):40–50, 2007.

[197] Mythily Srinivasan, Daniel Sedmak, and Scott Jewell. Effect of fixatives and tissue processing on the content and integrity of nucleic acids. *The American journal of pathology*, 161(6):1961–1971, 2002.

[198] Aidan D Meade, Colin Clarke, Florence Draux, Ganesh D Sockalingum, Michel Manfait, Fiona M Lyng, and Hugh J Byrne. Studies of chemical fixation effects in human cell lines using raman microspectroscopy. *Analytical and bioanalytical chemistry*, 396(5):1781–1791, 2010.

[199] Ehsan Gazi, John Dwyer, Nicholas P Lockyer, J Miyan, Peter Gardner, C Hart, M Brown, and Noel W Clarke. Fixation protocols for subcellular imaging by synchrotron-based fourier transform infrared microspectroscopy. *Biopolymers: Original Research on Biomolecules*, 77(1):18–30, 2005.

[200] Anthony SY Leong. Fixation and fixatives. *Laboratory Histopathology: A Complete Reference. London: Churchill Livingstone*, 4:1–26, 1994.

[201] Suneetha Devpura, Jagdish S Thakur, Fazlul H Sarkar, Wael A Sakr, Vaman M Naik, and Ratna Naik. Detection of benign epithelia, prostatic intraepithelial neoplasia, and cancer regions in radical prostatectomy tissues using raman spectroscopy. *Vibrational Spectroscopy*, 53(2):227–232, 2010.

[202] Max Diem, Miloš Miljković, Benjamin Bird, Tatyana Chernenko, Jen Schubert, Ellen Marcsisin, Antonella Mazur, Erin Kingston, Evgenia Zuser, Kostas Papamarkakis, et al. Applications of infrared and raman microspectroscopy of cells and tissue in medical diagnostics: present status and future promises. *Spectroscopy: An International Journal*, 27, 2012.

[203] Saori Fujiwara, Songjoon Baek, Lyuba Varticovski, Sohyoung Kim, and Gordon L Hager. High quality atac-seq data recovered from cryopreserved breast cell lines and tissue. *Scientific reports*, 9(1):1–11, 2019.

[204] Steven Busschots, Sharon O,ÄôToole, John J O,ÄôLeary, and Britta Stordal. Non-invasive and non-destructive measurements of confluence in cultured adherent cell lines. *MethodsX*, 2:8–13, 2015.

[205] Fiona M Frame, Davide Pellacani, Anne T Collins, and Norman J Maitland. Harvesting human prostate tissue material and culturing primary prostate epithelial cells. In *The Nuclear Receptor Superfamily*, pages 181–201. Springer, 2016.

[206] Norman Colthup. *Introduction to infrared and Raman spectroscopy*. Elsevier, 2012.

[207] Ruslan I Dmitriev. *Multi-Parametric Live Cell Microscopy of 3D Tissue Models*. Springer, 2017.

[208] Charles Kershaw. Raman spectroscopy studies of prostate cancer and streptomyces bacteria. Master's thesis, University of York, 2017.

[209] Iwan W Schie and Thomas Huser. Methods and applications of raman microspectroscopy to single-cell analysis. *Applied spectroscopy*, 67(8):813–828, 2013.

[210] I Notingher, S Verrier, H Romanska, AE Bishop, JM Polak, and LL Hench. In situ characterisation of living cells by raman spectroscopy. *Spectroscopy*, 16, 1970.

[211] Warren Strober. Trypan blue exclusion test of cell viability. *Current protocols in immunology*, 21(1):A–3B, 1997.

[212] I Notingher, S Verrier, H Romanska, AE Bishop, JM Polak, and LL Hench. In situ characterisation of living cells by raman spectroscopy. *Spectroscopy*, 16(2):43–51, 2002.

[213] Frank J Anscombe. Rejection of outliers. *Technometrics*, 2(2):123–146, 1960.

[214] Jason W Osborne and Amy Overbay. The power of outliers (and why researchers should always check for them). *Practical Assessment, Research, and Evaluation*, 9(1):6, 2004.

[215] Ute Neugebauer, Joachim H Clement, Thomas Bocklitz, Christoph Krafft, and Jürgen Popp. Identification and differentiation of single cells from peripheral blood by raman spectroscopic imaging. *Journal of biophotonics*, 3(8-9):579–587, 2010.

[216] Chenxu Yu, Erin Gestl, Kristin Eckert, David Allara, and Joseph Irudayaraj. Characterization of human breast epithelial cells by confocal raman microspectroscopy. *Cancer detection and prevention*, 30(6):515–522, 2006.

[217] OP Gnatyuk, GI Dovbeshko, A Yershov, SO Karakhim, Oleksii Ilchenko, and O Yu Posudievsky. 2d-bn nanoparticles as a spectroscopic marker and drug delivery system with protection properties. *RSC advances*, 8(53):30404–30411, 2018.

[218] Maciej Roman, Tomasz P Wrobel, Agnieszka Panek, Czeslawa Paluszkiewicz, and Wojciech M Kwiatek. Lipid droplets in prostate cancer cells and effect of irradiation studied by raman microspectroscopy. *Biochimica et Biophysica Acta (BBA)-Molecular and Cell Biology of Lipids*, 1865(9):158753, 2020.

[219] Yanting Shen, Lixia Yang, Lijia Liang, Zhiyuan Li, Jing Zhang, Wei Shi, Chongyang Liang, Weiqing Xu, and Shuping Xu. Ex situ and in situ surface-enhanced raman spectroscopy for macromolecular profiles of cell nucleus. *Analytical and bioanalytical chemistry*, 411(23):6021–6029, 2019.

[220] Florence Draux, Pierre Jeannesson, Abdelilah Beljebbar, Ali Tfayli, Nicolas Fourre, Michel Manfait, Josep Sulé-Suso, and Ganesh D Sockalingum. Raman spectral imaging of single living cancer cells: a preliminary study. *Analyst*, 134(3):542–548, 2009.

[221] Marc-Olivier Baradez, Daniela Biziato, Enas Hassan, and Damian Marshall. application of raman spectroscopy and univariate modelling as a process analytical technology for cell therapy bioprocessing. *Frontiers in Medicine*, 5:47, 2018.

[222] R Core Team. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria, 2017.

[223] Hadley Wickham. *ggplot2: elegant graphics for data analysis*. springer, 2016.

[224] Stella Corsetti, Thomas Rabl, David McGloin, and Ghulam Nabi. Raman spectroscopy for accurately characterizing biomolecular changes in androgen-independent prostate cancer cells. *Journal of biophotonics*, 11(3):e201700166, 2018.

[225] Melissa M Mariani, Peter Lampen, Juergen Popp, Bayden R Wood, and Volker Deckert. Impact of fixation on in vitro cell culture lines monitored with raman spectroscopy. *Analyst*, 134(6):1154–1161, 2009.

[226] Fiona Lyng, Ehsan Gazi, and Peter Gardner. Preparation of tissues and cells for infrared and raman spectroscopy and imaging. *Biomedical Applications of Synchrotron Infrared Microspectroscopy: A Practical Approach Royal Society of Chemistry, Cambridge*, pages 147–191, 2010.

[227] Kathryn A Wikenheiser, Diane K Vorbroker, Ward R Rice, Jean C Clark, Cindy J Bachurski, Herbert K Oie, and Jeffrey A Whitsett. Production of immortalized distal respiratory epithelial cell lines from surfactant protein c/simian virus 40 large tumor antigen transgenic mice. *Proceedings of the National Academy of Sciences*, 90(23):11029–11033, 1993.

[228] Ioan Notingher. Raman spectroscopy cell-based biosensors. *sensors*, 7(8):1343–1358, 2007.

[229] Karen E Shafer-Peltier, Abigail S Haka, Jason T Motz, Maryann Fitzmaurice, Ramachandra R Dasari, and Michael S Feld. Model-based biological raman spectral imaging. *Journal of Cellular Biochemistry*, 87(S39):125–137, 2002.

[230] Christoph Krafft. Bioanalytical applications of raman spectroscopy. *Analytical and bioanalytical chemistry*, 378(1):60–62, 2004.

[231] Christoph Krafft, Benjamin Dietzek, and Jürgen Popp. Raman and cars microspectroscopy of cells and tissues. *Analyst*, 134(6):1046–1057, 2009.

[232] Christoph Krafft, Thomas Knetschke, Axel Siegner, Richard HW Funk, and Reiner Salzer. Mapping of single cells by near infrared raman microspectroscopy. *Vibrational Spectroscopy*, 32(1):75–83, 2003.

[233] Christian Matthäus, Susie Boydston-White, Miloš Miljković, Melissa Romeo, and Max Diem. Raman and infrared microspectral imaging of mitotic cells. *Applied spectroscopy*, 60(1):1–8, 2006.

[234] Florence Draux, Cyril Gobinet, Josep Sulé-Suso, Aurélie Trussardi, Michel Manfait, Pierre Jeannesson, and Ganesh D Sockalingum. Raman spectral imaging of single cancer cells: probing the impact of sample fixation methods. *Analytical and bioanalytical chemistry*, 397(7):2727–2737, 2010.

[235] David Moss. *Biomedical Applications of Synchrotron Infrared Microspectroscopy: A Practical Approach*. Royal Society of Chemistry, 2011.

[236] Soogeun Kim, Kyung Min Byun, and Soo Yeol Lee. Influence of water content on raman spectroscopy characterization of skin sample. *Biomedical Optics Express*, 8(2):1130–1138, 2017.

[237] Martin G Shim and Brian C Wilson. The effects of ex vivo handling procedures on the near-infrared raman spectra of normal mammalian tissues. *Photochemistry and photobiology*, 63(5):662–671, 1996.

[238] Y Ozaki, A Mizuno, H Sato, K Kawauchi, and S Muraishi. Biomedical application of near-infrared fourier transform raman spectroscopy. part i: the 1064-nm excited raman spectra of blood and met hemoglobin. *Applied spectroscopy*, 46(3):533–536, 1992.

[239] Jeffrey M Vargason, Keith Henderson, and P Shing Ho. A crystallographic map of the transition from b-dna to a-dna. *Proceedings of the National Academy of Sciences*, 98(13):7265–7270, 2001.

[240] Sebastian Dochow, Christoph Krafft, Ute Neugebauer, Thomas Bocklitz, Thomas Henkel, Günter Mayer, Jens Albert, and Jürgen Popp. Tumour cell identification by means of raman spectroscopy in combination with optical traps and microfluidic environments. *Lab on a Chip*, 11(8):1484–1490, 2011.

[241] J Sulé-Suso, D Skingsley, GD Sockalingum, A Kohler, G Kegelaer, M Manfait, and AJ El Haj. Ft-ir microspectroscopy as a tool to assess lung cancer cells response to chemotherapy. *Vibrational Spectroscopy*, 38(1-2):179–184, 2005.

[242] Laura T Kerr, Katarina Domijan, Ivor Cullen, and Bryan M Hennelly. Applications of raman spectroscopy to the urinary bladder for cancer diagnostics. *Photonics & Lasers in Medicine*, 3(3):193–224, 2014.

[243] TJ Harvey, E Gazi, A Henderson, RD Snook, NW Clarke, M Brown, and P Gardner. Factors influencing the discrimination and classification of

prostate cancer cell lines by ftir microspectroscopy. *Analyst*, 134(6):1083–1091, 2009.

[244] Martin J Willemink, Wojciech A Koszek, Cailin Hardell, Jie Wu, Dominik Fleischmann, Hugh Harvey, Les R Folio, Ronald M Summers, Daniel L Rubin, and Matthew P Lungren. Preparing medical imaging data for machine learning. *Radiology*, 295(1):4–15, 2020.

[245] Nicholas Stone, Catherine Kendall, Jenny Smith, Paul Crow, and Hugh Barr. Raman spectroscopy for identification of epithelial cancers. *Faraday discussions*, 126:141–157, 2004.

[246] H Georg Schulze, Stanislav O Konorov, Nicolas J Caron, James M Piret, Michael W Blades, and Robin FB Turner. Assessing differentiation status of human embryonic stem cells noninvasively using raman microspectroscopy. *Analytical chemistry*, 82(12):5020–5027, 2010.

[247] James W Chan, Deborah Motton, John C Rutledge, NL Keim, and Thomas Huser. Raman spectroscopic analysis of biochemical changes in individual triglyceride-rich lipoproteins in the pre-and postprandial state. *Analytical chemistry*, 77(18):5870–5876, 2005.

[248] Tom C Bakker Schut, Peter J Caspers, Gerwin J Puppels, Annieke Nijssen, Freerk Heule, Martino HA Neumann, and Donal P Hayes. Discriminating basal cell carcinoma from its surrounding tissue by raman spectroscopy. *Journal of investigative dermatology*, 119(1):64–69, 2002.

[249] I Notingher, S Verrier, S Haque, JM Polak, and LL Hench. Spectroscopic study of human lung epithelial cells (a549) in culture: living cells versus dead cells. *Biopolymers: Original Research on Biomolecules*, 72(4):230–240, 2003.

[250] Ricardo Matias Lopes, Landulfo Silveira, Marcos Augusto RS Silva, Katia Ramos M Leite, Carlos Augusto G Pasqualucci, and Marcos Tadeu T Pacheco. Diagnostic model based on raman spectra of normal, hyperplasia

and prostate adenocarcinoma tissues in vitro. *Spectroscopy*, 25:89–102, 2011.

[251] Luis Felipe CS Carvalho, Franck Bonnier, Claudio Tellez, Laurita Dos Santos, Kate O'Callaghan, Jeff O'Sullivan, Luis Eduardo S Soares, Stephen Flint, Airton A Martin, Fiona M Lyng, et al. Raman spectroscopic analysis of oral cells in the high wavenumber region. *Experimental and Molecular Pathology*, 103(3):255–262, 2017.

[252] Halina Abramczyk and Anna Imiela. The biochemical, nanomechanical and chemometric signatures of brain cancer. *Spectrochimica Acta Part A: Molecular and Biomolecular Spectroscopy*, 188:8–19, 2018.

[253] Ronald OP Draga, Matthijs CM Grimbergen, Peter LM Vijverberg, Christiaan FP van Swol, Trudy GN Jonges, J Alain Kummer, and JLH Ruud Bosch. In vivo bladder cancer diagnosis by high-volume raman spectroscopy. *Analytical chemistry*, 82(14):5993–5999, 2010.

[254] Lei Wang, Dalin He, Jin Zeng, Zhenfeng Fuan, Qiang Dang, Xinyang Wang, Jun Wang, Liqing Huang, Peilong Cao, Guanjun Zhang, et al. Raman spectroscopy, a potential tool in diagnosis and prognosis of castration-resistant prostate cancer. *Journal of biomedical optics*, 18(8):087001, 2013.

[255] Hemlata Bhatia, Hamidreza Mehdizadeh, Denis Drapeau, and Seongkyu Yoon. In-line monitoring of amino acids in mammalian cell cultures using raman spectroscopy and multivariate chemometrics models. *Engineering in life sciences*, 18(1):55–61, 2018.

[256] György Tóth, Zsolt Lengyel, László Balkay, Morshed A Salah, Lajos Tron, and Csaba Toth. Detection of prostate cancer with 11c-methionine positron emission tomography. *The Journal of urology*, 173(1):66–69, 2005.

[257] Petra Boukamp, Rule T Petrussevska, Dirk Breitkreutz, Jurgen Hornung, Alex Markham, and Norbert E Fusenig. Normal keratinization in a spontaneously immortalized aneuploid human keratinocyte cell line. *Journal of Cell Biology*, 106(3):761–771, 1988.

[258] HW Löms Ziegler-Heitbroc, Eckhard Thiel, Agnes Futterer, Volker Herzog, Antje Wirtz, and Gert Riethmüller. Establishment of a human cell line (mono mac 6) with characteristics of mature monocytes. *International journal of cancer*, 41(3):456–461, 1988.

[259] RA Rocha, James Martin Fox, PG Genever, and Yvette Hancock. Biomolecular phenotyping and heterogeneity assessment of mesenchymal stromal cells using label-free raman spectroscopy. *Scientific reports*, 11(1):1–16, 2021.

[260] J Renwick Beattie, Steven EJ Bell, and Bruce W Moss. A critical evaluation of raman spectroscopy for the analysis of lipids: fatty acid methyl esters. *Lipids*, 39(5):407–419, 2004.

[261] AT Tu. Peptide backbone conformation and microenvironment of protein side chains. *Advances in infrared and Raman spectroscopy*, 13:47–112, 1986.

[262] Hassan Khesbak, Olesya Savchuk, Satoru Tsushima, and Karim Fahmy. The role of water h-bond imbalances in b-dna substate transitions and peptide recognition revealed by time-resolved ftir spectroscopy. *Journal of the american chemical society*, 133(15):5834–5842, 2011.

[263] Tatang Hernawan and Graham Fleet. Chemical and cytological changes during the autolysis of yeasts. *Journal of industrial microbiology*, 14(6):440–450, 1995.

[264] Lisa Vaccari, G Birarda, L Businaro, Sabrina Pacor, and G Grenci. Infrared microspectroscopy of live cells in microfluidic devices (md-irms): toward a powerful label-free cell-based assay. *Analytical chemistry*, 84(11):4768–4775, 2012.

[265] Shan Lu and Daniel E Epner. Molecular mechanisms of cell cycle block by methionine restriction in human prostate cancer cells. *Nutrition and cancer*, 38(1):123–130, 2000.

[266] Nicholas Stone, Pela Stavroulaki, Catherine Kendall, Martin Birchall, and Hugh Barr. Raman spectroscopy for early detection of laryngeal malignancy: preliminary results. *The Laryngoscope*, 110(10):1756–1763, 2000.

[267] Michael S Feld, Ramasamy Manoharan, Juha Salenius, Jacobo Orenstein-Carndona, Tjeerd J Roemer, James F Brennan III, Ramachandra R Dasari, and Yang Wang. Detection and characterization of human tissue lesions with near-infrared raman spectroscopy. In *Advances in fluorescence sensing technology II*, volume 2388, pages 99–104. International Society for Optics and Photonics, 1995.

[268] TC Bakker Schut, H Van Dekken, HW Tilanus, HA Bruining, and GJ Puppels. Nir raman spectroscopy of healthy and diseased oesophagus. *Spectroscopy of Biological Molecules: Modern Trends*, pages 455–456, 1997.

[269] Anita Mahadevan-Jansen, Michele Follen Mitchell, Nirmala Ramanujamf, Anais Malpica, Sharon Thomsen, Urs Utzinger, and Rebecca Richards-Kortumt. Near-infrared raman spectroscopy for in vitro detection of cervical precancers. *Photochemistry and photobiology*, 68(1):123–132, 1998.

[270] Guro F Giskeødegård, Ailin Falkmo Hansen, Helena Bertilsson, Susana Villa Gonzalez, Kåre Andre Kristiansen, Per Bruheim, Svein A Mjøs, Anders Angelsen, Tone Frost Bathen, and May-Britt Tessem. Metabolic markers in blood can separate prostate cancer from benign prostatic hyperplasia. *British journal of cancer*, 113(12):1712–1719, 2015.

[271] Maximo A Benavides, Karen L Hagen, Wenfeng Fang, Pan Du, Simon Lin, Mary P Moyer, Wancai Yang, Kirby I Bland, William E Grizzle, and Maarten C Bosland. Suppression by l-methionine of cell cycle progression in lncap and mcf-7 cells but not benign cells. *Anticancer research*, 30(6):1881–1885, 2010.

[272] Nilgun Gurbuz, Margaret A Park, Paul Dent, Asim B Abdel Mageed, Suresh C Sikka, and Asli Baykal. Cystine dimethyl ester induces apoptosis through regulation of pkc-δ and pkc-ε in prostate cancer cells. *Anti-Cancer Agents in*

Medicinal Chemistry (Formerly Current Medicinal Chemistry-Anti-Cancer Agents), 15(2):217–227, 2015.

[273] KD Kent, WJ Harper, and JA Bomser. Effect of whey protein isolate on intracellular glutathione and oxidant-induced cell death in human prostate epithelial cells. *Toxicology in vitro*, 17(1):27–33, 2003.

[274] RICHA RP RASTOGI. Apoptosis: molecular mechanisms and pathogenicity. 2009.

[275] Alessandro Laviano, Michael M Meguid, Antonia Cascino, Alessio Molfino, and Filippo Rossi Fanelli. Tryptophan in wasting diseases: at the crossing between immune function and behaviour. *Current Opinion in Clinical Nutrition & Metabolic Care*, 12(4):392–397, 2009.

[276] Isabel R Schlaepfer, Carolyn A Hitz, Miguel A Gijón, Bryan C Bergman, Robert H Eckel, and Britta M Jacobsen. Progestin modulates the lipid profile and sensitivity of breast cancer cells to docetaxel. *Molecular and cellular endocrinology*, 363(1-2):111–121, 2012.

[277] Louise A Dawson, Norman J Maitland, Paul Berry, Anthony J Turner, and Badar A Usmani. Expression and localization of endothelin-converting enzyme-1 in human prostate cancer. *Experimental biology and medicine*, 231(6):1106–1110, 2006.

[278] Kanji Endoh, Mayuko Nishi, Hitoshi Ishiguro, Hiroji Uemura, Yohei Miyagi, Ichiro Aoki, Hisashi Hirano, Yoshinobu Kubota, and Akihide Ryo. Identification of phosphorylated proteins involved in the oncogenesis of prostate cancer via pin1-proteomic analysis. *The Prostate*, 72(6):626–637, 2012.

[279] George J Thomas Jr, James M Benevides, Stacy A Overman, Toyotoshi Ueda, Koichi Ushizawa, Masaaki Saitoh, and Masamichi Tsuboi. Polarized raman spectra of oriented fibers of a dna and b dna: anisotropic and isotropic local raman tensors of base and backbone vibrations. *Biophysical journal*, 68(3):1073–1088, 1995.

[280] Satish Rao, Saurabh Raj, Benjamin Cossins, Monica Marro, Victor Guallar, and Dmitri Petrov. Direct observation of single dna structural alterations at low forces with surface-enhanced raman scattering. *Biophysical journal*, 104(1):156–162, 2013.

[281] Bas WD De Jong, Tom C Bakker Schut, Kees Maquelin, Theo van der Kwast, Chris H Bangma, Dirk-Jan Kok, and Gerwin J Puppels. Discrimination between nontumor bladder tissue and tumor by raman spectroscopy. *Analytical chemistry*, 78(22):7761–7769, 2006.

[282] Evelien Rysman, Koen Brusselmans, Katryn Scheys, Leen Timmermans, Rita Derua, Sebastian Munck, Paul P Van Veldhoven, David Waltregny, Veerle W Daniëls, Jelle Machiels, et al. De novo lipogenesis protects cancer cells from free radicals and chemotherapeutics by promoting membrane lipid saturation. *Cancer research*, 70(20):8117–8126, 2010.

[283] Bjørn T Gjertsen and Stein O Døskeland. Protein phosphorylation in apoptosis. *Biochimica et Biophysica Acta (BBA)-Molecular Cell Research*, 1269(2):187–199, 1995.

[284] Martin Hedegaard, Christoph Krafft, Henrik J Ditzel, Lene E Johansen, Søren Hassing, and JuÃàrgen Popp. Discriminating isogenic cancer cells and identifying altered unsaturated fatty acid content as associated with metastasis status, using k-means clustering and partial least squares-discriminant analysis of raman maps. *Analytical chemistry*, 82(7):2797–2802, 2010.

[285] Lucy Faulkner. Raman spectroscopy as a diagnostic method for prostate cancer cell lines on a graphene substrate. Master's thesis, University of York, 2019.

[286] Marcos AS De Oliveira, Michael Campbell, Alaa M Afify, Eric C Huang, and James W Chan. Hyperspectral raman microscopy can accurately differentiate single cells of different human thyroid nodules. *Biomedical optics express*, 10(9):4411–4421, 2019.