Research Article

Chemometrics-Enabled Raman Spectrometric Qualitative Determination and Assessment of Biochemical Alterations during Early Prostate Cancer Proliferation in Model Tissue

John I. Githaiga (iD),[1] **Hudson K. Angeyo** (iD),[1] **Kenneth A. Kaduki** (iD),[1] **and Wallace D. Bulimo** (iD)[2]

[1]*Department of Physics, University of Nairobi, 30197–00100 Nairobi, Kenya*
[2]*Department of Biochemistry, University of Nairobi, 30197–00100 Nairobi, Kenya*

Correspondence should be addressed to John I. Githaiga; gjirungu@uonbi.ac.ke

The use of Raman spectroscopy combined with multivariate chemometrics for disease diagnosis has attracted great attention from researchers in recent years. This is because it is a noninvasive and nondestructive detection approach with enhanced sensitivity. However, a major challenge when analyzing spectra from biological samples has been the detection of subtle biochemical alterations buried in background and fluorescence noise. This work reports a qualitative chemometrics-assisted investigation of subtle biochemical alterations associated with prostate malignancy in model biological tissue (metastatic androgen insensitive (PC3) and immortalized normal (PNT1a) prostate cell lines). Raman spectra were acquired from PC3 and PNT1a cells at various stages of growth, and their biochemical alterations were determined from difference spectra between the two cell lines (for prominent alterations) and principal component analysis (PCA) (for subtle alterations). The Raman difference spectra were computed by subtracting the normalized mean spectral intensities of PNT1a cells from the normalized mean spectral intensities of PC3 cells. These difference spectra revealed prominent biochemical alterations associated with the malignant PC3 cells at $566 \pm 0.70 \, \text{cm}^{-1}$, $630 \, \text{cm}^{-1}$, $1370 \pm 0.86 \, \text{cm}^{-1}$, and $1618 \pm 1.73 \, \text{cm}^{-1}$ bands. The band intensity ratios at $566 \pm 0.70 \, \text{cm}^{-1}$ and $630 \, \text{cm}^{-1}$ suggested that prostate malignancy can be associated with an increase in relative amounts of nucleic acids and lipids, respectively, whereas those at $1370 \pm 0.86 \, \text{cm}^{-1}$ and $1618 \pm 1.73 \, \text{cm}^{-1}$ suggested that prostate malignancy can be associated with a decrease in relative amounts of saccharides and tryptophan, respectively. In the analysis using PCA, intermediate-order and high-order principal components (PCs) were used to extract the subtle biochemical fingerprints associated with the cell lines. This revealed subtle biochemical differences at $1076 \, \text{cm}^{-1}$, $(1232, 1234 \, \text{cm}^{-1})$, $(1276, 1278 \, \text{cm}^{-1})$, $(1330, 1333 \, \text{cm}^{-1})$, $(1434, 1442 \, \text{cm}^{-1})$, and $(1471, 1479 \, \text{cm}^{-1})$. The band intensity ratios at $1076 \, \text{cm}^{-1}$ and $1232 \, \text{cm}^{-1}$ suggested that prostate malignancy can be associated with an increase in subtle amounts of nucleic acids and amide III components, respectively. The method reported here has demonstrated that subtle biochemical alterations can be extracted from Raman spectra of normal and malignant cell lines. The identified subtle bands could play an important role in quantitative monitoring of early biomarker alterations associated with prostate cancer proliferation.

1. Introduction

Cancer is a potentially fatal disease dominated by the uncontrolled growth and metastasis of abnormal cells [1]. By 2018, prostate cancer was estimated to be the second most frequent cancer and the fifth leading cause of cancer death in men [2]. The conventional prostate-specific antigen (PSA) procedure for prostate cancer diagnosis has limitations and risks of early detection and treatment of indolent cancers [1], and Gleason grading may be highly subjective rendering to inter-and intraobserver variations in pathological reporting due to resultant false-positive and/or false-negative

outcomes [3]. This highlights the need to develop highly sensitive, less invasive, real time, and relatively low-cost methods for screening patients at the early stages of cancer development.

In the search for biomolecular differences between normal and diseased biological samples, vibrational spectroscopy methods, e.g., Fourier transform infrared (FTIR) absorption spectroscopy and Raman spectroscopy, have increasingly generated interest in the biomedical sciences. These methods have been applied successfully in disease diagnostics due to their capability of performing quantitative analysis on the different chemical compositions and molecular structures of healthy and pathological tissues [4, 5]. For instance, FTIR spectroscopy has been previously employed to spectrally differentiate between normal and neoplastic human skin samples suffering from epithelioma and basalioma cancers [4] and in characterizing the damage and regeneration caused by chemical agents in liver tissues [5, 6]. In the last two decades, Raman microspectroscopy has become a popular noninvasive spectroscopic technique for in vivo, ex vivo, and in situ human cancer studies [3, 7–10], thanks to the numerous advances in instrumentation, including the use of various types of powerful laser sources for excitation, novel optics arrangements, and sensitive detectors. Besides being a rapid and objective technique, it requires minimal sample preparation, is less invasive, is reagent free, has high spatial and depth resolution (down to $1 \mu m$), and is minimally influenced by water bands [10, 11]. Its potential is enhanced when combined with chemometric algorithms for extracting the most significant chemically relevant information, while the less informative data and insignificant information (noise) are discarded.

Recent works show that Raman microspectroscopy has been extensively used in detecting biochemical alterations in prostatic malignancies. For instance, a 785 nm excitation laser-based spectroscopic study on benign prostate tissue specimens, employing Raman spectroscopy and total reflectance Fourier transform infrared spectroscopy (ATR-FTIR) in liaison with principal component analysis-linear discriminant analysis (PCA-LDA) and variable selection techniques (genetic, successive projection algorithms), was reported by Theophilou et al. [12]. In this study, deoxyribonucleic acid (DNA) and ribonucleic acid (RNA) bands in the 1000–1400 cm^{-1} region were observed to be the most significant for segregation between the prostate tissues archived over a period of three decades. In [13], the applicability of 1064 nm wavelength in prostate cancer diagnosis was evaluated where spectral markers associated with proteins, lipids, and nucleic acids differentiated malignant and benign samples, mainly in the 1000–1450 cm^{-1} region. A support vector machine (SVM) classification model predicted Gleason scores of samples with an overall accuracy of 95%. Elsewhere, a dual excitation wavelength system (based on −785 nm and 671 nm lasers) was employed in assessing human prostate tissues in fingerprinting (500–1800 cm^{-1}) and high wavenumber (2800–3550 cm^{-1}) regions [14]. In this study, application of SVM on spectral datasets showed the 500–1800 cm^{-1} and 2800–3050 cm^{-1} spectral regions contained complementary molecular information for prostate cancer detection. The peak intensities of DNA/RNA, lipids, carbohydrates, and proteins were found to be higher in cancerous tissues when compared to peak intensities in normal tissues.

To understand early progression of prostate cancer, the potential of using monoclonal proliferated immortalized prostate cell lines in prostate cancer studies has been investigated using Raman microspectroscopy and chemometrics. For instance, various chemometric methods (PCA, classical least-squares curve fitting (CLSCF), and self-modelling curve resolution (SMCR)) were evaluated on their capability to detect biochemical differences between two human prostatic cells (PC3 and PNT2) [15]. PCA was found to be more suitable for the imaging process, where the second principal component (PC2) successfully revealed tumor cells exhibiting a modified distribution of cytoplasmic lipid fraction. On the contrary, SMCR revealed concentration levels of proteins and lipids, where a higher concentration increase in lipids (97%) was observed in tumor cells. In another study [16], biochemical differences between immortalized normal prostate cells (PNT2) and metastatic androgen-independent prostate cancer cells (DU145) were investigated. The features of glycogen (485 cm^{-1}), phenylalanine (621 cm^{-1}, 1003 cm^{-1}, and 1031 cm^{-1}), DNA (723 cm^{-1}), and nucleic acids (1184 cm^{-1}) were found to be influential for classification assignment into the cancer class (DU145), whereas the L-arginine (497 cm^{-1}) component was found to be influential for classification assignment into the normal class (PNT2). In this study, a PCA-LDA model distinguished the two cells with specificity and sensitivity values of 88% and 95%, respectively.

Often, however, a major challenge would be the detection of subtle biomolecular differences that have eluded detection, perhaps due to their extremely small concentrations against the pronounced background and fluorescence noise. A possible solution which is yet to be explored for detection of subtle alterations during prostate cancer proliferation is the combination of Raman spectroscopy and principal component analysis (PCA). PCA is a widely used dimensional reduction multivariate analysis technique for removing redundancy in the original datasets. PCA aims at reducing the dimensionality while, at the same time, accounting for as much of the variation in the original dataset as possible [17]. With PCA, the data are transformed into a new set of coordinates or variables (principal components) that are a linear combination of the original variables, whereas the observations (scores) in the new principal component space are uncorrelated. In many spectral analyses, however, only the first few principal components (PCs) that account for large proportions of variance (i.e., most significant information) are retained, while the remaining higher order PCs (typically assumed as noise) are discarded [17, 18]. Limiting spectra data analysis to only a reduced set of the first few principal components is disadvantageous in two ways: firstly, there is no guarantee that the difference between the sample groups of interest will be in the direction of the first few or high-variance principal components [18], and secondly, the low-variance signals (weak Raman bands) in the presence of much higher

variance signals can vanish into the noise of higher principal components [19]. The loss of low-variance signals can exclude most of the subtle information (e.g., subtle biochemical alterations) in the original data concerning between-group variation.

In work by Jolliffe [18], it was shown it is not obvious that the difference between the sample groups will always be in the direction of the first few or high-variance principal components. This is because the first few principal components are useful for identifying differences between sample groups in the following situations: (i) where the between-group variation is much larger than within-group variation and (ii) where within and between-group variation have the same dominant directions [18]. For instance, work by Theophilou et al. [12] showed that the first six principal component scores that accounted for >90% variance could not demonstrate any substantial separation of prostate chippings according to their year of collection although the principal components scores were statistically significant ($p < 0.05$). In another study, the third principal component (PC3) that accounted for 10.2% of total cumulative variance was found to better discriminate two prostatic cells (PNT2 and DU145 cells) than using the first principal component (PC1) that accounted for 62% of total cumulative variance [16]. Because the best subset of principal components is not necessarily limited to those with the largest variances, Pelletier [19] suggested that analysis of the excluded higher principal components could be a potential method of detecting analyte information and other trends in the dataset. This is in agreement with observations made by Jolliffe [18] that the last few PCs are not simply unstructured left-overs after removing the important PCs but can be rather useful in other ways which include detecting unobservable near-constant linear relationships between variables, in regression, and selection of subset of variables, in outlier detection.

Currently, there is lack of research interrogating diagnostically useful weak band biomolecular alterations during prostate cancer progression based on the utility of low-variance principal components. Therefore, the aim of this study was to investigate the weak detectable biomolecular alterations associated with the metastatic androgen insensitive (PC3) and immortalized normal (PNT1a) prostate cell lines through utility of intermediate and high-order PCs. The weak band alterations in cells were studied in a span of three stages of cell proliferation, i.e., stage 1 (48 hours), stage 2 (72 hours), and stage 3 (96 hours). PCA was used for multivariate analysis. The chosen principal components must be statistically distinct in order to make biological sense; otherwise, they would be only representing random directions [20]. Therefore, the utility of PCs was determined by a combination of two statistical criteria, i.e., the t-test and effect size (Cohen's d and Pearson correlation coefficients (r)).

In brief, the breakdown of this paper is as follows: in Section 2, the adopted samples preparation, Raman spectral data collection, and spectral analysis procedures have been described. Section 3 presents the results and their discussion. Finally, Section 4 highlights the significant findings of the study.

2. Materials and Methods

2.1. Cell Lines Samples Preparation.

The metastatic androgen insensitive (PC3) and immortalized normal (PNT1a) human prostate cell lines at passages 3 and 4, respectively, were prepared at Kenya Medical Research Institute (KEMRI), Nairobi. Both cell lines were cultured in Dulbecco's Modified Eagle Medium (DMEM (1x)) (Gibco® by Life Technologies™, USA), supplemented with 10%, by volume, fetal bovine serum (FBS) (ATCC® 30-2020™, USA), 1% Fungizone (amphotericin B, 250 μg/ml) (Gibco® by Life Technologies™, UK), 1% L-glutamine (200 Mm) (100x) (Gibco® by Life Technologies™, USA), 1% penicillin-streptomycin, and 0.1% Gentamycin solution 50 mg/ml (all from Sigma® Life Science, USA), at 37°C in a humidified atmosphere containing 5% CO^2. Each cell line was simultaneously proliferated in three batches at seeding densities of 1.5×10^6, 1.2×10^6, and 1.0×10^6 viable cells/ml, whose cells were harvested after 48 hours (stage 1), 72 hours (stage 2), and 96 hours (stage 3), respectively. Furthermore, sterilized calcium fluoride substrates (Crystran, UK) were introduced into T-75 plug seal capped cell culture flasks (Corning® Flask) to allow cell attachments. No coating was done on substrates to enhance cell proliferation.

Cell harvesting was done by first removing the substrates (containing attached monolayer cells). The substrates were briefly washed in Hanks' balanced salt solution, fixated in acetone ≥99.5% A.C.S. reagent (Sigma® Life Science, USA) for approximately 8 minutes, rinsed with double distilled Millipore water, and then allowed to dry overnight in the biosafety chamber. The attached monolayer cells were examined using the Raman optical microscope system. The remaining cells in T-75 flasks were detached by the usual trypsinization procedure and centrifuged twice at 1200 revolutions per minute (r. p. m.) for 5 minutes. The resultant cells were twice washed in a 1.5 ml phosphate-buffered saline solution (Sigma-Aldrich®, USA) and centrifuged at 1200 r. p. m. for 5 min after each wash. After removing remaining supernatants, cells were vortexed again, then suspended in 1.5 ml phosphate-buffered saline solution, and stored at −80°C.

2.2. Raman Spectroscopy

2.2.1. Raman Spectroscopy Instrument.

Raman spectra were recorded using the STR series Raman spectrometer set up (Airix Corporation, Japan) that comprised two sources of excitation: a diode-pumped solid-state laser gem diode green 532 nm (class IIIb, 50 mW, Laser quantum Corporation, UK) and a single-mode diode near-infrared 785 nm (class IIIb, 100 mW) lasers. The scattering efficiency is higher at shorter wavelengths (e.g., 532 nm), thereby resulting in shorter integration times [21], but so does autofluorescence and sample photodegradation [22]. Therefore, spectra were collected using 785 nm laser excitation. Importantly, 785 NIR nm excitation source has good penetration depths into the samples (e.g., biological tissues), leading to excitation of large volumes of sample [23]. Raman signals were collected with a 50x infinity corrected microscope objective (Olympus

BX51, Olympus Corporation, Tokyo) that was focused after cutting out more than 99.9% of the elastic scattering light through 785 nm MaxLine® laser clean-up filter (LL01-785, Semrock Corporation) onto the entrance slit of a spectrograph (Princeton Instruments (Acton SpectraPro2300), entrance slit 10 um~3.0 mm, focal length 300~500 mm, grating 600 lines/mm). A charge-coupled device (CCD) camera (Ludl Electronics Products, Ltd) was employed for signal detection. The experimental parameters/control functions (including cosmic-ray removal and file conversions) were set and managed by the Window-based STR Raman version 1.9.3 software (Airix Corporation, Japan).

2.2.2. Raman Spectral Data Collection.

Thawed suspended supernatants were centrifuged at 1200 revolutions per minute (r. p. m.) for 5 minutes. To avoid large background signals that could emanate from the sample substrate, the calcium fluoride substrates (Crystran, UK) were specifically employed for this study [21]. For Raman measurements, a $15\,\mu l$ aliquot was deposited onto freshly sterile prepared calcium fluoride substrates. The cells were allowed to dry at room temperature (approximately $18°C$) in the biosafety chamber for the next 6 hours and immediately taken for spectroscopic measurements. The spectra were collected in the range of $490–2132\,cm^{-1}$ region, with a resolution of $1.0\,cm^{-1}$. The collected spectra for each cell were as follows: 48 hours spectral measurements (PC3 = 79 spectra and PNT1a = 75 spectra), 72 hours spectral measurements (PC3 = 84 spectra and PNT1a = 76 spectra), and 96 hours spectral measurements (PC3 = 97 spectra and PNT1a = 101 spectra). The exposure time was 90 seconds. The laser power irradiation over the samples ranged between 19.80 (±0.2) mW and 21.10 (±0.03) mW. The room temperature was constantly maintained at $22°C$ to ensure wavelength stability. Measurements were taken in the dark to avoid interference from ambient light sources.

2.2.3. Spectral Analysis.

Spectral data analysis was limited to the $500–1800\,cm^{-1}$ region. This region is advantageous in revealing the major relevant biochemical fingerprint alterations in biological samples and circumventing baseline correction issues that would possibly emerge from the large background contributions in the lower wavenumber regions [21]. This was performed in two separate processes as follows:

(i) To examine prominent Raman band alterations in both cells, spectra data were coadded, averaged, baseline-corrected (linearly), and minimally denoised by Daubechies wavelet transformations, followed by spectral smoothing based on Savitzky–Golay (S-G) polynomial filter smoothing algorithm. For plotting, ORIGIN (Origin pro 9.1) software was used. The resultant 6 averaged spectra (3 averaged spectra for PC3 cells and 3 averaged spectra for PNT1a cells) were normalized to their maximum intensity. To further examine the band intensity differences in the spectral profiles, the

difference spectra were computed by subtracting the normalized mean spectrum intensities of PNT1a samples from the normalized mean spectrum intensities of PC3 samples. The two-sample t-test was used to determine their statistical significance.

(ii) To determine subtle biochemical alterations, individual raw spectra were baseline corrected (linearly) and normalized to their maximum intensity. The spectral denoising preprocessing procedure was avoided to preserve all pertinent spectral features in datasets. The PC3 and PNT1a spectra were then combined into single matrices, i.e., $w \times n$, where w = wavenumbers and n = number of spectra, based on the stage of cell proliferation (48 hours, 72 hours, and 96 hours). The resultant stage 1 ($X_{781 \times 154}$), stage 2 ($X_{781 \times 160}$), and stage 3 ($X_{781 \times 198}$) matrices were subjected to singular value decomposition principal component analysis (SVD-PCA) [17]. With this procedure, three matrices U, S, and V were obtained, where matrix U represented the scores, S the size of the scores, and V transpose (V^T) the loadings. This was done in MATLAB 2018a scripting environment. The primary estimation of the number of probable significant PCs to be retained was done by two methods as follows: firstly, by examining the cumulative fraction (or percentage) of the data and/or eigenvalues explained as a function of the number of principal components (scree plots), and secondly, examining the canonical loadings plots of the principal components based on the canonical parameters.

A methodology which categorizes PCs as either low, intermediate, or high-order PC based on the size of variance of PCs and cumulative percentage [24] was adopted, as follows:

(i) Low-order PCs: <99% of the cumulative variance and >1.0 (average) eigenvalue.

(ii) Intermediate-order PCs: between 99% and 99.5% of the cumulative variance.

(iii) High-order PCs: >99.5% of the cumulative variance and <1.0 (average) eigenvalue.

The t-test and effect size statistical criteria [25] were included in determining the utility (relevance) of PCs for chemometric analysis. This was done in Microsoft Excel (2010). The two-sample t-tests were employed to compare the PC scores between the PC3 and PNT1a cells. The determined t-test values were utilized in calculating Cohen's d values for each PC score as follows [26]:

$$d = \frac{(\overline{x})_t - (\overline{x})_c}{s_P},$$

$$s_P = \sqrt{\frac{(n_t - 1)s_t^2 + (n_c - 1)s_c^2}{n_t + n_c}}, \tag{1}$$

where d = Cohen's d effect size, $(\overline{x})$ = mean of diseased or normal conditions, s = standard deviation, n = number of

samples, t = diseased condition, and c = control condition. It should be noted Cohen's d effect sizes were classified as small ($d = 0.2$), medium ($d = 0.5$), and large ($d \geq 0.8$) [25]. The resultant p values were adjusted using a Holm–Bonferroni method to maintain a family-wise alpha of 0.05 for tests on all variables [27]. Pearson correlation coefficients (r) were computed to assess the degree of linear relationship between principal components, where effect sizes were classified as small ($r = 0.2$), medium ($r = 0.5$), and large ($r = 0.8$) [25].

The low-variance signals (weak band Raman band alterations) were extracted from the intermediate- and high-order PC loadings with the help of *peak finding* function in Omnic® software from Thermo Scientific. Two parameters, i.e., threshold and sensitivity levels, were adjusted, where threshold value specified the Y value above which peaks could be found, and sensitivity value took into account the relative size of adjacent spectral features. The sensitivity level was set at a low value (30%) to eliminate noise and other unimportant features above the threshold value.

3. Results and Discussion

3.1. Determination of Prominent Biochemical Alterations in PC3 and PNT1a Proliferation. Figures 1(a) and 1(b) show the highlighted optical photomicrographs of unstained PC3 and PNT1a monolayer grown cells on calcium fluoride (CaF_2) substrates, respectively. As expected, the PC3 and PNT1a are typically adherent cells that grow attached to a substrate in discrete patches, usually with more regular dimensions (ideally polygonal shapes). Here, it is noted the attached monolayer cells have epithelial morphological features with the nucleus (i), cytoplasm (ii), and cell wall (iii) locations discernible in both images.

The biochemical assignments of peaks were done in accordance with the Raman spectroscopy of tissues, body fluids, or biomolecules, as highlighted in the literature. This was done in consideration of position and possible wavelength differences of each particular Raman band. Theoretically, spectral resolution in a dispersive Raman spectrometer is determined by many factors which include spectrometer focal length, diffraction grating, laser wavelength, and the detector [28]. However, the effect of each factor on spectral resolution is considered under the assumption that all other factors remain unchanged, which may include change in slit width, grating line spacing, and effects of small grating movements. However, CCD detector (dispersive Raman) is usually sensitive to the changes in room temperature, owing to the thermal expansion of various elements including the diffraction grating, the detector, and all the other optics. In view of possible detection conditions and experimental errors, the Raman peaks with wavelength difference within $10\,cm^{-1}$ were considered as the same peak. This choice of wavelength difference was further motivated based on the work of McCreely [28] who suggested that most analytical Raman applications involve liquids and solids in which Raman bandwidths are significantly greater than those in the gas phase; hence, the narrowest linewidths encountered in most liquid and solid samples lie in the range of 3 to $10\,cm^{-1}$.

Figure 2(a) shows an example of as-collected averaged, minimally denoised, and baseline-corrected raw spectra in PC3 and PNT1a samples. Examination of both spectra shows the primary Raman bands featuring around $519\,cm^{-1}$, $615\,cm^{-1}$, $638\,cm^{-1}$, (700, $702\,cm^{-1}$), (852, $857\,cm^{-1}$), $889\,cm^{-1}$, (921, $927\,cm^{-1}$), $1001\,cm^{-1}$, $1084\,cm^{-1}$, $1126\,cm^{-1}$, (1256, $1261\,cm^{-1}$), (1445, $1450\,cm^{-1}$), (1552, $1554\,cm^{-1}$), $1602\,cm^{-1}$, and $1657\,cm^{-1}$. The respective Raman band assignments are provided in Table 1. Figure 2(b) shows the stacked mean Raman spectra that explain biochemical alterations occurring during all stages of cell proliferation, for both cell lines. It should be noted the spectra have been linearly offset for comparison purposes. The spectra demonstrate a similar spectral pattern although there are minor Raman shifts. The figure reveals the mean positions (and the standard errors) of major Raman band contributions for the two cell lines, whose biochemical assignments are well explained in literature [9, 13, 16, 29, 30]. They include $519 \pm 1.30\,cm^{-1}$ (phosphatidylinositol), $719 \pm 3.78\,cm^{-1}$ (phospholipids and nucleic acids), $857 \pm 1.84\,cm^{-1}$ (proline and tyrosine proteins), $1003 \pm 1.49\,cm^{-1}$ (phenylalanine), $1083 \pm 1.94\,cm^{-1}$ (C-N stretch of proteins and lipids and nucleic acids), $1253 \pm 1.87\,cm^{-1}$ (nucleic acids, lipids, and proteins), $1449 \pm 1.35\,cm^{-1}$ (C-H vibration of proteins and lipids and nucleic acids), and $1659 \pm 1.33\,cm^{-1}$ (amide I), with the strongest Raman bands occurring at $519 \pm 1.30\,cm^{-1}$, $1253 \pm 1.87\,cm^{-1}$, $1449 \pm 1.35\,cm^{-1}$, and $1659 \pm 1.33\,cm^{-1}$.

Figures 3(a)–3(c) show the difference spectra between the normalized mean spectra of PC3 and PNT1a cells at 48, 72, and 96 hours, respectively. The positive bands explain alterations present in greater concentration in PC3 cells, while negative bands explain alterations abundant in PNT1a cells. Generally, the difference spectra show PC3 cells, exhibiting prominent biochemical alterations at $566 \pm 0.70\,cm^{-1}$ (cytosine and guanine), $630\,cm^{-1}$ (glycerol), $972 \pm 1.17\,cm^{-1}$ (cytosine and proteins), $1186\,cm^{-1}$ (guanine, cytosine, adenine, and antisymmetric phosphate vibrations), $1520 \pm 1.41\,cm^{-1}$ (cytosine, C-C stretch, and C=C stretch mode (β–carotene accumulation)), and $1743\,cm^{-1}$ (ester groups) [9, 16, 29–31]. Similarly, PNT1a samples have prominent biochemical alterations at $550 \pm 0.23\,cm^{-1}$ (cytosine, guanine, tryptophan, and glycerol), $719 \pm 1.31\,cm^{-1}$ (phospholipids and nucleic acids), $852 \pm 0.47\,cm^{-1}$ (proline, tyrosine, and polysaccharides), $948 \pm 1.88\,cm^{-1}$ (valine and proline), $1250 \pm 2.86\,cm^{-1}$ (amide III, lipids, adenine, and cytosine), $1332 \pm 1.64\,cm^{-1}$ (nucleic acids and CH_3CH_2 wagging of collagen), $1450 \pm 2.20\,cm^{-1}$ (lipids and proteins), and $1660\,cm^{-1}$ (amide I and lipids) [13, 30, 31]. If we consider stage 2 and 3 spectral datasets, it is observed PNT1a cells spectra have prominent band alterations at $623\,cm^{-1}$ (phenylalanine and adenine), $664\,cm^{-1}$ (guanine, thymine, and collagen), $898 \pm 0.20\,cm^{-1}$ (proline and saccharides), $1066\,cm^{-1}$ (proline of collagen), $1152 \pm 1.44\,cm^{-1}$ (proteins and carotenoids), $1370 \pm 0.86\,cm^{-1}$ (saccharides), $1573\,cm^{-1}$ (guanine, adenine, and tryptophan proteins), $1618 \pm 1.73\,cm^{-1}$ (tryptophan), and $1675\,cm^{-1}$ (amide I (β-sheet)) [30, 31]. Furthermore, the spectral marker at $576\,cm^{-1}$ (phosphatidylinositol) was only detected during
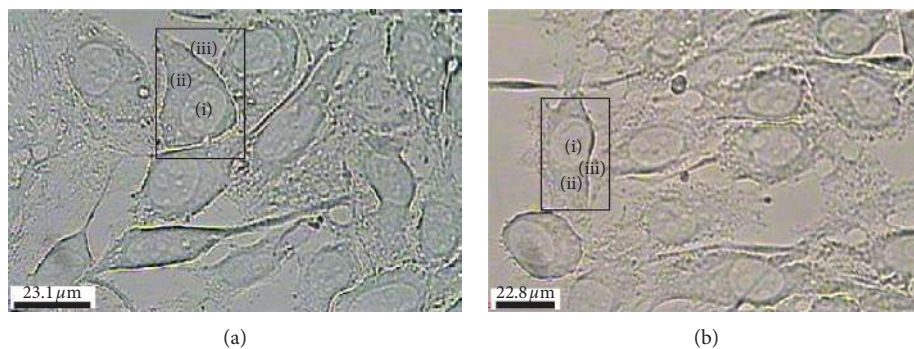
FIGURE 1: Photomicrographs of (a) PC3 and (b) PNT1a monolayer grown cells on calcium fluoride (CaF_2) substrates at 50x magnification, with (i) nucleus, (ii) cytoplasm, and (iii) cell wall constituents clearly visible in both cells.
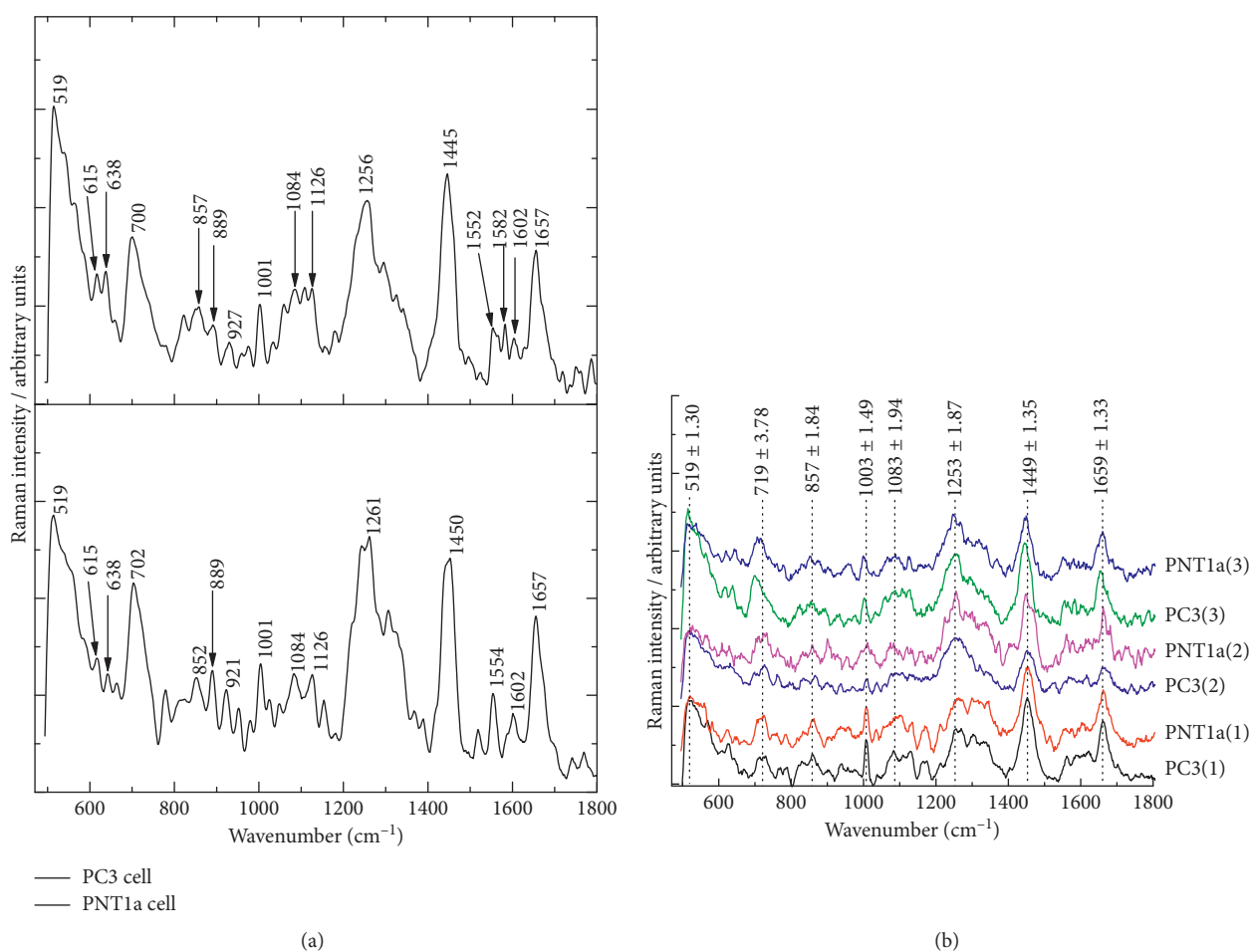

FIGURE 2: (a) Examples of as-collected raw spectra in PC3 and PNT1a cells; (b) the stacked mean spectra of samples according to the stages of cell proliferation. The spectra shown in part (b) have been linearly offset for comparison. The numbers (1, 2, 3) enclosed in brackets identify stage 1 (48 hours), stage 2 (72 hours), and stage 3 (96 hours) spectral measurements.

the first two stages of PNT1a cell proliferation and the late proliferation cycle (stage 3) of malignant (PC3) cells. This suggests presence of enhanced lipid alterations during advanced malignancy. However, to evaluate utility of these bands in prostate cancer diagnosis, the statistical p values were calculated using Student's t-test for each band. The two-sample t-test on averaged intensities around these bands showed they were statistically significant ($p < 0.05$), with exception of 576 cm^{-1} band, meaning the biochemical changes at 576 cm^{-1} band could not be used to discriminate between control (PNT1a) and prostate cancer (PC3) cell groups.

TABLE 1: Raman band assignments of PC3 and PNT1a prostatic cells.

| Raman shift (cm^{-1})	Functional groups and molecular vibration assignments	References
519	Phosphatidylinositol	[9]
615	Cholesterol esters	[9]
638	C-C stretch, C-C twisting of tyrosine proteins	[9, 16]
700, 702	Cholesterol esters	[9]
852, 857	Ring-breathing modes (tyrosine proteins)	[16]
889	Structural protein modes of tumors	[9]
921	C-C stretch (proline ring, glucose, and lactic acid)	[9]
927	Proteins	[9]
1001	Proteins (phenylalanine)	[9, 13]
1084	Phosphodiester groups in nucleic acids	[9]
1126	Protein, phospholipids	[13]
1256, 1261	C-N stretch (proteins), CH$_2$ twisting and wagging (lipids and proteins), C-H bend (phospholipids)	[29]
1445	CH$_2$ bending (proteins)	[16]
1450	DNA/RNA, proteins, lipids	[13, 16]
1552, 1554	$_v$(C=C) stretching modes (tryptophan), amide II	[9]
1582	C=C bending mode of phenylalanine (proteins)	[16]
1602	C=C bending modes (phenylalanine and tyrosine proteins)	[16]
1657	C=O stretch (proteins and lipids), C=C stretch (lipids), nucleic acids	[9, 29]

(a)

(b)

(c)
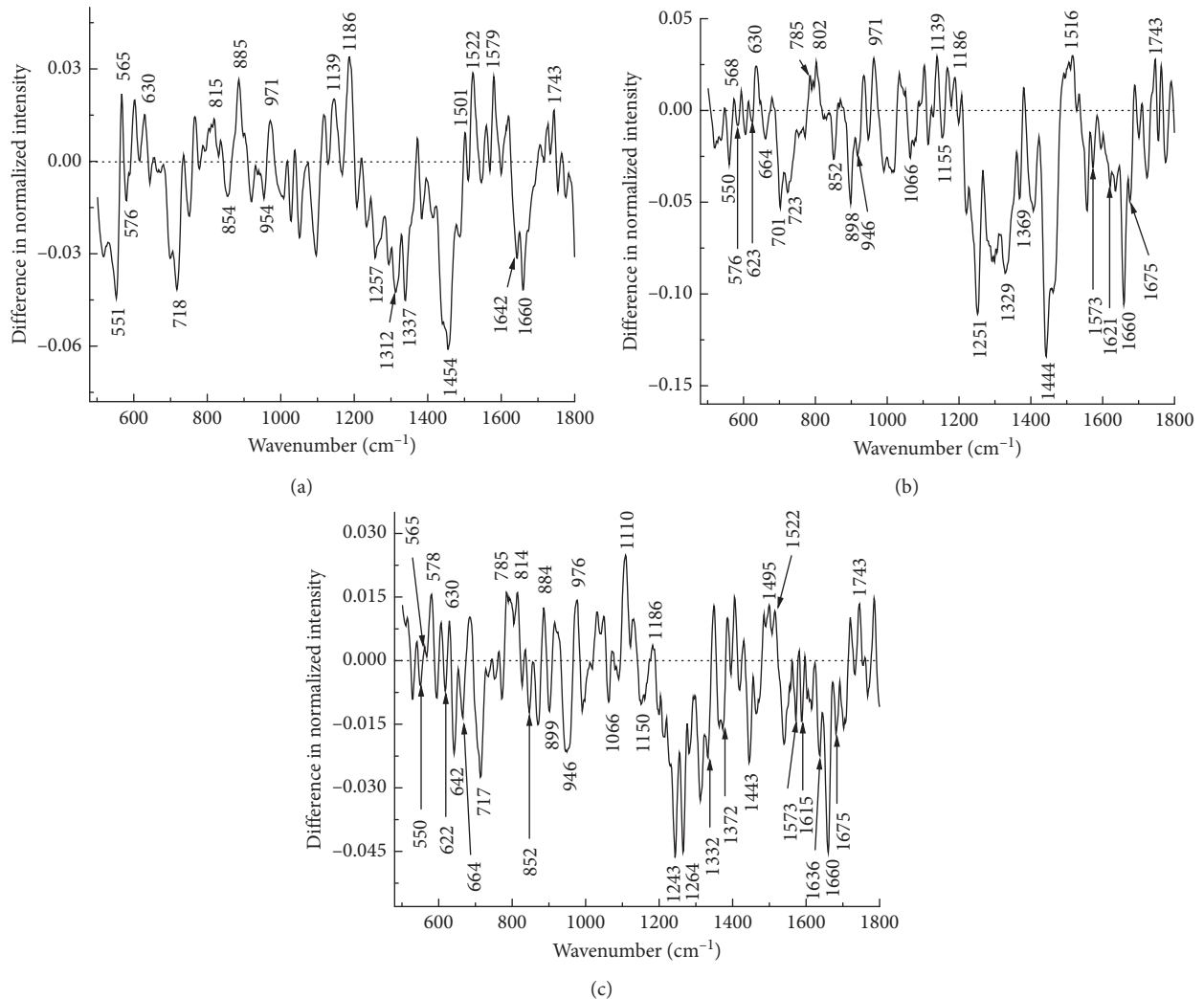
FIGURE 3: Difference spectra between the normalized mean spectra of PC3 and PNT1a cells for (a) stage 1 (48 hours), (b) stage 2 (72 hours), and (c) stage 3 (96 hours) spectral measurements. It is observed the PC3 samples have prominent spectral markers at 566 ± 0.70 cm^{-1}, 630 cm^{-1}, 972 ± 1.17 cm^{-1}, 1186 cm^{-1}, 1520 ± 1.41 cm^{-1}, and 1743 cm^{-1}. The PNT1a samples have prominent spectral markers at 550 ± 0.23 cm^{-1}, 719 ± 1.31 cm^{-1}, 852 ± 0.47 cm^{-1}, 948 ± 1.88 cm^{-1} 1250 ± 2.86 cm^{-1}, 1332 ± 1.64 cm^{-1}, 1450 ± 2.20 cm^{-1}, and 1660 cm^{-1}.

The peak intensity ratios of Raman spectra measurements have been previously utilized in classifying diseased and healthy samples [32]. A similar analysis was performed on the observed prominent difference bands in Figures 3(a)–3(c), where ratio values (i.e., I_C/I_N) were calculated by dividing the normalized intensities of PC3 spectra (I_C) by normalized intensities of PNT1a spectra (I_N). Table 2 highlights the bands ($566 \pm 0.70\,\mathrm{cm}^{-1}$, $630\,\mathrm{cm}^{-1}$, $1370 \pm 0.86\,\mathrm{cm}^{-1}$, and $1618 \pm 1.73\,\mathrm{cm}^{-1}$) whose band intensity ratios were found to increase or decrease with the stages of cell proliferation.

The band intensity ratio values at $566 \pm 0.70\,\mathrm{cm}^{-1}$ and $630\,\mathrm{cm}^{-1}$ were found to increase with the stages of the cell proliferation, while the band intensity ratio values at $1370 \pm 0.86\,\mathrm{cm}^{-1}$ and $1618 \pm 1.73\,\mathrm{cm}^{-1}$ were found to decrease with the stages of cell proliferation. The increasing peak ratios at $566 \pm 0.70\,\mathrm{cm}^{-1}$ band suggest biochemical changes due to nucleic acids bases (cytosine and guanines) increased with malignancy. This correlates with other closely related prostatic studies [33–35] which have suggested that spectral intensities of DNA-related bands in prostate samples increase with malignancy, a factor attributed to enlarged nuclei in malignant cells than for normal cells and therefore greater abundance of DNA content in malignant samples [34]. For instance, qualitative findings by Crow et al. [33] showed nucleic acid contents in malignant prostate biopsies were higher compared to benign prostatic hyperplasia (BPH) ones. Elsewhere, a study that investigated biochemical alterations for the different tissue pathologies within the bladder and prostate gland region observed the DNA contents increased with malignancy [35], while work by Taleb et al. [34] observed DNA contents were higher in malignant (LNCaP) cells when compared with DNA contents in normal (PNT1a) cells.

The increasing band intensity ratio values at $630\,\mathrm{cm}^{-1}$ band suggest biochemical changes due to lipids increased with malignancy. A closely related study by Matias et al. [36] observed that adipocyte levels increased with the severity of malignancy, a factor attributed to the production of lipids via de novo lipogenesis [37]. The $1370\,\mathrm{cm}^{-1}$ band is a pronounced saccharide band in biological samples, while $1618\,\mathrm{cm}^{-1}$ band is assigned to υ (C=C) vibrations due to tryptophan protein [30]. The decreasing band intensity ratio values observed at $1370 \pm 0.86\,\mathrm{cm}^{-1}$ and $1618 \pm 1.73\,\mathrm{cm}^{-1}$ bands suggest that biochemical changes due to saccharides and tryptophan decreased with malignancy. The observed decrease in saccharide levels with malignancy fit in with the findings of a previous qualitative cell line-based study where glycogen content was found lower in prostatic adenocarcinoma pathologies, compared to benign pathologies [33], a factor that can be attributed to enhanced glucose uptake by cells during onset of tumor development for conversion to lactate molecules necessary for energy production during cell proliferation [38]. The decrement of tryptophan levels with malignancy is an indication of prostate malignancy undergoing increased tryptophan degradation.

3.2. Determination of Subtle Biochemical Alterations during PC3 and PNT1a Proliferation. Figure 4 shows the

eigenvalues plotted as a function of the number of principal components in our raw datasets. As highlighted in Section 2.2.3, the raw datasets were baseline corrected by the linear method and normalized to their maximum intensity, but without denoising to preserve all pertinent spectral features in the datasets. Table 3 shows categorization of principal components as either low, intermediate, or high-order principal components based on the size of variances and cumulative percentages of the total variation. It was noted that the first principal components (PC1) could be categorized as a low-order principal component. For stage 1 datasets of both cells, principal components 2 to 16 were categorized as intermediate-order PCs. For stages 2 and 3 datasets, intermediate-order principal components were 2 to 22 and 2, respectively. The remaining principal components were categorized as high-order principal components according to the criteria given in Table 3. The first two principal components (PC1 and PC2) were found to account for the largest variance of the data. Overall, the total cumulative variances accounted for by the first two principal components (PC1 and PC2) in stage 1, stage 2 and, stage 3 datasets were 99.13%, 99.02%, and 99.58%, respectively.

To determine which PCs scores were potentially significant for sample discrimination, the PCs were subjected to the two-sample t-test and effect size (Cohen's d and Pearson's correlation coefficient r) statistical criteria, as described in Section 2.2.3. Table 4 shows the p values, Cohen's d, and Pearson correlation coefficients (r) computed for the first 10 principal components. Significant differences ($p < 0.05$) between the principal component scores were observed in PCs 1 to 5 for the stage 1 dataset, PCs 1 to 3 and PC 8 for the stage 2 dataset, and PCs 1 to 6 and PC 8 for the stage 3 dataset. Generally, the first principal component (PC1) had the largest effect sizes (Cohen's d) for all the datasets. The value of d was 1.22 for stage 1, 1.32 for stage 2, and 2.27 for stage 3 datasets. The respective explained total variances were 98.85%, 98.88%, and 99.19%. The utility of observed PCs was further evaluated by plotting their canonical variable distributions (Figures 5(a), 5(c), and 5(e)). Figure 5(a) shows that PCs 2, 3, 4, and 5 had the largest standardized loadings for the stage 1 dataset. For the stage 2 dataset (Figure 5(c)), the largest standardized loading values were for PCs 2, 3, 8, and 9, while PCs 2, 4, 5, and 6 had the largest standardized loadings for the stage 3 dataset (Figure 5(e)). These are the principal components that were used for cell discrimination.

Cell discrimination was done by examining the scatter plots of PC scores and their respective loadings vectors. Each loading vector was related to the original spectrum by PC scores, which refer to the weight of that particular biochemical component in each spectrum [16]. The score plots highlighted the natural groupings of each of the principal components for the cell types. It was found that PC 2 (0.29%), PC 2 (0.42%), and PC 4 (0.39%) had best grouping of both cell types for stages 1, 2, and 3 spectral data, respectively. Examination of loading vectors showed the groupings/discriminations were attributed to the prominent biochemical differences observed in Figure 3. It was also observed that some level of clustering was present due to

TABLE 2: Comparison of peak intensity ratios between malignant (PC3) and normal cells (PNT1a) at $566 \pm 0.70\,\text{cm}^{-1}$, $630\,\text{cm}^{-1}$, $1370 \pm 0.86\,\text{cm}^{-1}$, and $1618 \pm 1.73\,\text{cm}^{-1}$ spectral regions.

| Raman shift (cm^{-1}) | Raman intensities (arbitrary units) | | | | | | Band ratios | | |
| | Stage 1 | | Stage 2 | | Stage 3 | | | | |
	PNT1a	PC3	PNT1a	PC3	PNT1a	PC3	PC3(1)/PNT1a(1)	PC3(2)/PNT1a(2)	PC3(3)/PNT1a(3)
565	0.90	0.87	0.75	0.85	0.75	0.87	0.967	1.133	1.16
568	0.81	0.87	0.77	0.86	0.77	0.89	1.074	1.117	1.156
630	0.69	0.79	0.73	0.86	0.62	0.85	1.145	1.178	1.371
1368	0.51	0.47	0.72	0.66	0.68	0.58	0.922	0.917	0.853
1372	0.40	0.44	0.71	0.65	0.60	0.53	1.1	0.915	0.883
1617	0.32	0.35	0.46	0.43	0.58	0.53	1.094	0.935	0.914
1621	0.30	0.32	0.59	0.54	0.49	0.44	1.067	0.915	0.898

(a)

(b)

(c)
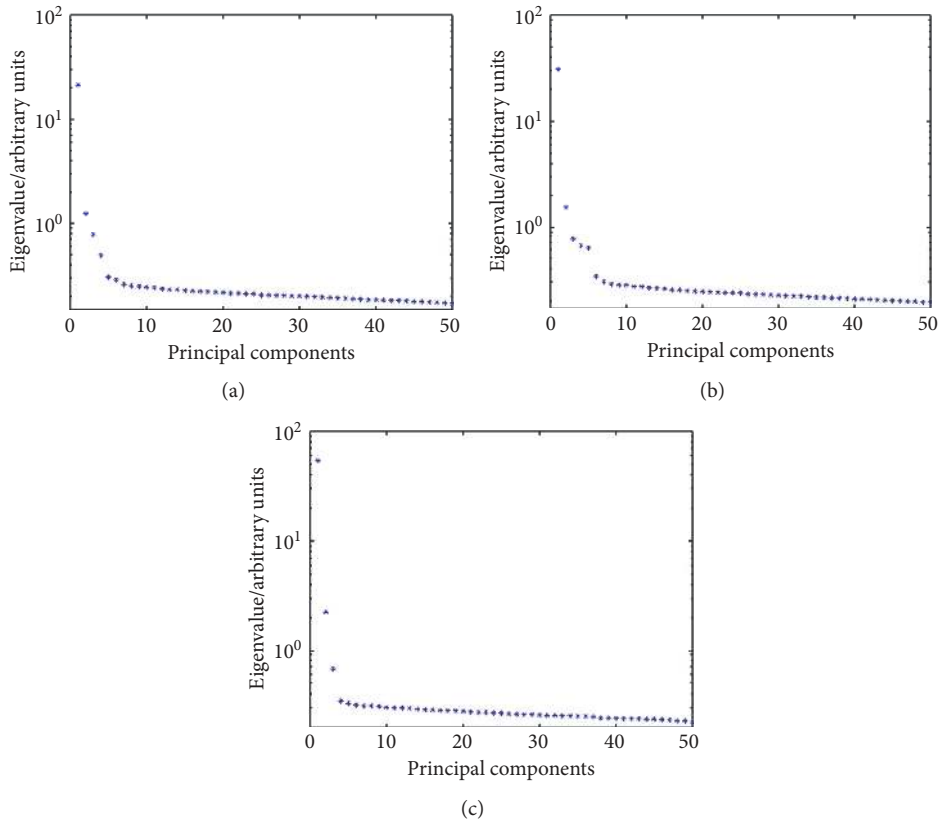
FIGURE 4: Scree plots showing eigenvalues explained as a function of the number of principal components for (a) stage 1, (b) stage 2, and (c) stage 3 datasets. For the sake of clarity, only the first 50 principal components are shown.

TABLE 3: Categorization of principal components (PCs) based on their cumulative percentage of total variation and the variance sizes: low-, intermediate-, and high-order PCs.

Spectral dataset	Low-order PCs (<99% of the cumulative variance and >1.0 average eigenvalue)	Intermediate-order PCs (between 99% and 99.5% of the cumulative variance)	High-order PCs (>99.5% of the cumulative variance and <1.0 average eigenvalue)
Stage 1	1	2–16	17–154
Stage 2	1	2–22	23–160
Stage 3	1	2	3–198

TABLE 4: Student's t-test (p values) and effect sizes (Cohen's d and Pearson's correlation coefficients (r)) showing the relationship between the principal component scores of normal (PNT1a) and malignant (PC3) cells.

| PC	Stage 1			Stage 2			Stage 3		
	p value	Cohen's d	r	p value	Cohen's d	r	p value	Cohen's d	r
1	0.003	1.22	0.52	0.04	1.32	0.55	0.44	2.27	0.75
2	4.04×10^{-19}	0.99	0.44	2.11×10^{-39}	1.18	0.51	3.56×10^{-39}	2.36	0.76
3	0.0095	0.39	0.19	9.98×10^{-10}	0.99	0.44	0.029	0.18	0.09
4	7.13×10^{-18}	1.17	0.50	0.42	0.33	0.02	1.51×10^{-18}	1.38	0.57
5	3.24×10^{-10}	1.08	0.48	0.20	0.13	0.06	0.0088	0.34	0.17
6	0.18	0.15	0.07	0.32	0.08	0.04	0.0082	0.19	0.09
7	0.21	0.13	0.06	0.36	0.06	0.03	0.413	0.03	0.02
8	0.09	0.21	0.11	0.03	0.29	0.14	0.02	0.03	0.01
9	0.20	0.14	0.07	0.23	0.41	0.20	0.46	0.02	0.01
10	0.11	0.19	0.09	0.33	0.07	0.04	0.44	0.02	0.01
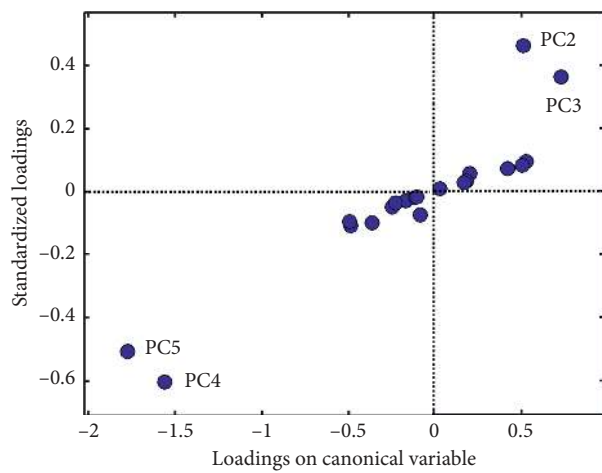
(a)
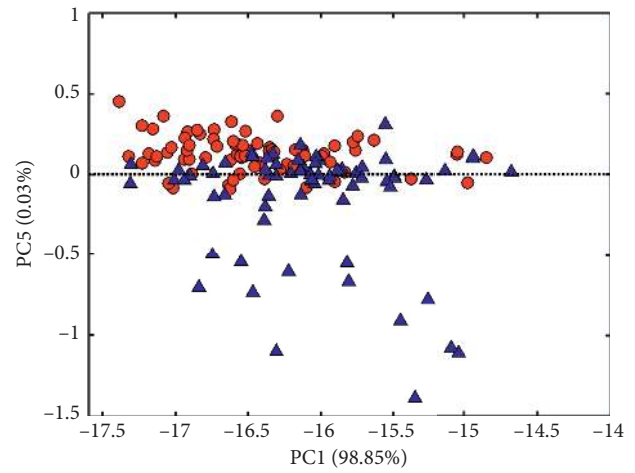
(b)
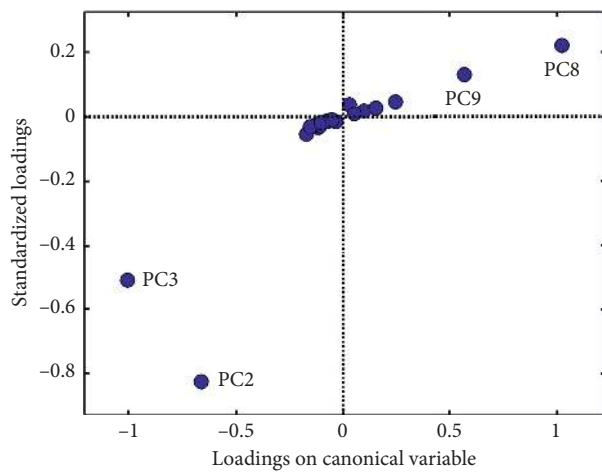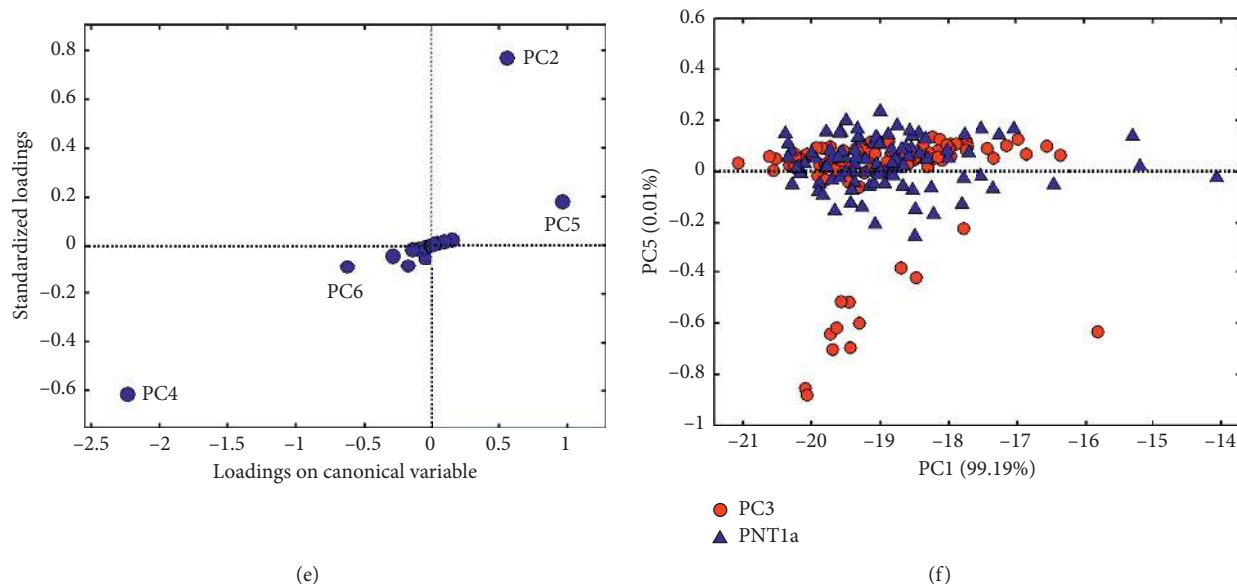

(c)
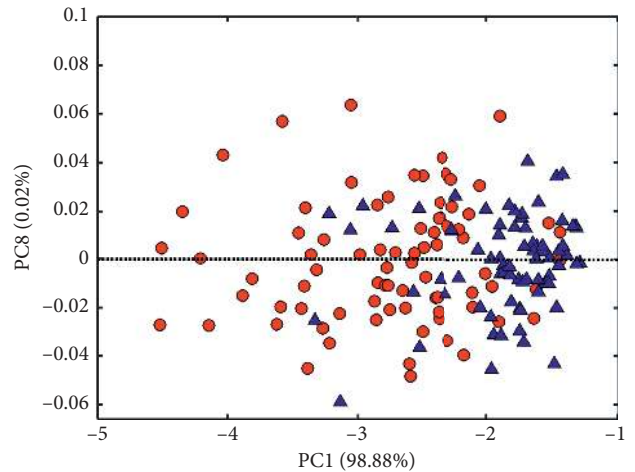
(d)

FIGURE 5: Continued.

(e)

(f)

FIGURE 5: Canonical variable distribution of principal components for (a) stage 1, (c) stage 2, and (e) stage 3 measurements and the score plots of the first principal component (PC 1) against the intermediate- and high-order PCs for (b) stage 1, (d) stage 2, and (f) stage 3 measurements. The clustered principal components (centered on origin) had poor discrimination of scores and were therefore discarded from the analysis.

PC5 (0.03%), PC8 (0.02%), and PC5 (0.01%) in stages 1, 2, and 3 spectral data, respectively (Figures 5(b), 5(d), and 5(f)). These discriminations were attributed to the subtle biochemical differences between the malignant and normal cells.

Since the aim of this study was to investigate subtle biochemical alterations (weak variance signals) associated with prostate cancer progression, the following principal components were selected for further analysis: PC5 (0.03%) for stage 1, PC8 (0.02%) for stage 2, and PC5 (0.01%) for stage 3 datasets. The principal component loadings that explained scores discrimination due to PC5 (0.03%) for stage 1, PC8 (0.02%) for stage 2, and PC5 (0.01%) for stage 3 datasets are shown in Figures 6(a)–6(c), respectively.

As explained in Section 2.2.3, Cohen's d effect sizes were classified as small ($d = 0.2$), medium ($d = 0.5$), and large ($d \geq 0.8$) [25]. Table 4 shows PC5 (0.03%) for stage 1 dataset had a large effect size, whereas PC8 (0.02%) for stage 2 and PC5 (0.01%) for stage 3 datasets had medium effect sizes. Examination of both the scatter plots (Figures 5(b), 5(d), and 5(f)) and the loading spectrum (Figures 6(a)–6(c)) showed the positive bands in the loading spectrum were present in cells to which higher scores were attributed, and vice versa. The loadings profiles were noted to contain useful peaks amid noisy features. A possible method of extracting useful peaks would have been to denoise or smooth the loading profiles. However, this would have introduced spectral artifacts, peak shifts, and loss of useful peaks. Therefore, the peaks were extracted in their raw form by adjusting threshold and sensitivity levels/parameters of *peak finding* function in Omnic® software from Thermo Scientific. It should be noted the sensitivity levels were kept at low values (30%) to eliminate noisy peaks, and only the largest loading

vectors that explained scores distribution in Figures 5(b), 5(d), and 5(f) were extracted and used for further analysis. Detailed information concerning the loading vectors (which in this case are weak Raman bands) that explained natural groupings of PNT1a and PC3 scores using intermediate- and high-order PC5 (0.03%) for stage 1, PC8 (0.02%) for stage 2, and PC5 (0.01%) for stage 3 datasets are shown in Table 5.

It was observed (Table 5) that the loading vectors (Raman bands) at 771 cm^{-1}, 871 cm^{-1}, 1122 cm^{-1}, 1176 cm^{-1}, and 1614 cm^{-1} (for stage 1), 1068 cm^{-1}, 1191 cm^{-1}, 1333 cm^{-1}, 1471 cm^{-1}, 1524 cm^{-1}, and 1586 cm^{-1} (for stage 2), and 1076, 1089 cm^{-1}, 1278 cm^{-1}, 1596 cm^{-1}, 1631 cm^{-1}, and 1712 cm^{-1} (for stage 3) had the most influence for the assignment of scores into the malignant class (PC3 cells). Likewise, the loading vectors at (1014, 1092, 1232, 1468, and 1776 cm^{-1}), (704, 1217, 1271, 1274, and 1531 cm^{-1}), and (641, 697, 1319, 1377, and 1577 cm^{-1}) had the most influence for the assignment of scores into the normal class (PNT1a) for stages 1, 2, and 3 measurements, respectively. Note that the loading vectors at (1232, 1330, and 1442 cm^{-1}), (1471 cm^{-1}), and (1076 and 1278 cm^{-1}) were common for the two cell lines in stages 1, 2, and 3 datasets, respectively. The two-sample t-test on averaged normalized intensities on these bands indicated they were statistically significant ($p < 0.05$).

To better understand these differences, band intensity ratios (I_C/I_N) were determined at all observed loading vectors (weak Raman bands), by dividing the normalized intensities of diseased (PC3 cells) Raman spectra (I_C) by the respective normalized intensities of control (PNT1a cells) Raman spectra (I_N). The band intensity ratio values at 1076 cm^{-1} and 1232 cm^{-1} bands were found to increase with the stage of cell proliferation (Table 6). The 1076 cm^{-1} band
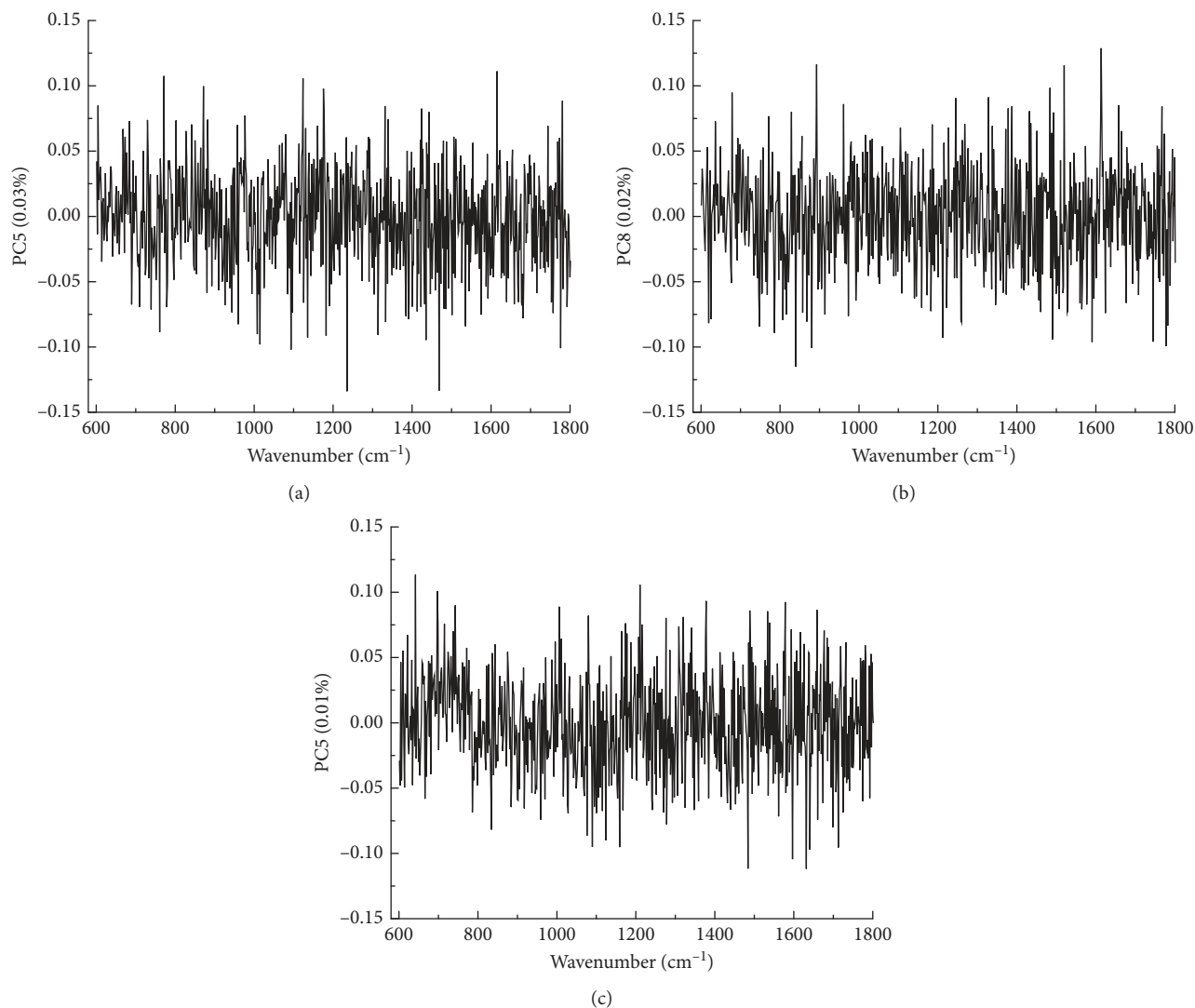
(a)

(b)

(c)

FIGURE 6: Principal component loadings that explain natural grouping of PNT1a and PC3 scores due to (a) PC5 (0.03%) for stage 1, (b) PC8 (0.02%) for stage 2, and (c) PC5 (0.01%) for stage 3 datasets. The loadings profiles are noted to contain useful peaks amid noisy features. The useful peaks (weak Raman bands) were extracted from the raw loading profiles by adjustment of threshold and sensitivity levels/parameters in peak finding function in Omnic® software from Thermo Scientific. Sensitivity levels were set at a low value (30%) so that noise and other unimportant features above threshold value were eliminated.

is associated with symmetric phosphate stretching modes which normally originate from the phosphodiester groups in nucleic acids and is thought to suggest an increase in the nucleic acids in malignant tissues [30]. The alterations around $1232 \, cm^{-1}$ band indicate antisymmetric phosphate stretching vibrations due to nucleic acids and amide III components [39]. Therefore, the band intensity ratio values of normalized intensities of PC3 cells (I_C) to normalized intensities of PNT1a cells (I_N) at $1076 \, cm^{-1}$ and $1232 \, cm^{-1}$ bands suggested prostate malignancy could be associated with an increase in relative amounts of nucleic acids and amide III components. The rest of weak Raman bands (Table 5) can be mainly attributed to subtle nucleic acids and protein alterations, and their specific assignments are explained elsewhere [9, 30, 31].

The use of high-order principal components is still underexplored as a chemometric tool for the analysis of prostatic tissues. In fact, Raman spectroscopy-based studies for the diagnosis of prostate cancer favor the use of the first few principal components (PCs) that account for large proportions of variance [12, 15, 16, 36, 40, 41]. A previous review on applications of Raman spectroscopy for prostate cancer [3] showed the majority of prostate-based Raman studies reporting bands around $1000 \, cm^{-1}$ (phenylalanine ring breathing), 1200 to $1350 \, cm^{-1}$ (overlapping of the amide III, CH_2 "wagging," and CN "stretch" peaks present mainly in lipids and nucleic acids), $1450 \, cm^{-1}$ (CH_2 deformation vibration present in nucleic acids, proteins, and lipids), and in the region of 1600 to $1700 \, cm^{-1}$ (C O stretch and amide I, present in nucleic acids and proteins).

TABLE 5: The Raman bands (loading vectors) that explain natural groupings of PNT1a and PC3 scores using PC 5 (0.03%) for stage 1, PC 8 (0.02%) for stage 2, and PC 5 (0.01%) for stage 3 datasets.

Prostatic cells					
PC3			PNT1a		
Stage 1	Stage 2	Stage 3	Stage 1	Stage 2	Stage 3
604	508	574	1007	517	641 *
668	564	786	1014 *	704 **	697 *
730	869	833	1092 *	716	1076[c]
771 *	913	884	1232 *[a]	752	1209
841	1068 *	900	1330[a]	772	1278[c]
871 *	1077	969	1391	1217 *	1319 *
976	1191 *	1027	1442[a]	1271 *	1377 *
1122 *	1333 *	1076 *[c]	1468 *	1274 *	1577 *
1176 *	1471 *[b]	1089 *	1681	1364	
1232[a]	1524 *	1094	1776 *	1471[b]	
1330[a]	1586 *	1100		1514	
1339		1123		1531 *	
1442[a]		1158		1557	
1614 *		1240			
		1269			
		1278 *[c]			
		1439			
		1482			
		1561			
		1596 *			
		1631 *			
		1641			
		1699			
		1712 *			
		1724			

*Loading vectors (weak Raman bands) that had the most influence in the assignment of scores for each of the two cell lines. Superscripts a, b, and c identify the loading vectors (weak Raman bands) that were common for both cell lines. The values are in units of cm^{-1}.

To the best of authors' knowledge, spectroscopic prostate studies based on the utility of intermediate- and high-order principal components which account for negligible variance (assumed noise) has not been previously reported. Our findings have shown it is possible to discriminate prostatic cells using weak band alterations determined with aid of intermediate- and high-order principal components. Interestingly, earlier studies on prostate cancer cells by Corsetti et al. [16] and Crow et al. [40] reported prominent bands around (880, 1184, 1588, and 1614 cm^{-1}) and (1094, 1096, 1125, and 1576 cm^{-1}), respectively. Similarly, the 1328 cm^{-1} and 643 cm^{-1} bands were observed as prominent bands in works by Matias et al. [36] and Patel and Martin-Hirsch [41], respectively. In our study, intermediate- and high-order principal components demonstrated the weak bands detected at 871 cm^{-1}, 1094 cm^{-1}, 1122 cm^{-1}, 1123 cm^{-1}, 1191 cm^{-1}, 1330 cm^{-1}, 1333 cm^{-1}, 1586 cm^{-1}, and 1614 cm^{-1} (Table 5) were relevant for grouping of normal and diseased prostatic cells. However, these bands were subtle in intensity and could not be detected in spectral profiles, as shown in Figures 2(b) and 3(a)–3(c). This apparent discrepancy in Raman band intensities between their results and our findings is most likely due to the different experimental conditions employed. In this study, the observed subtle bands can be explained in consideration of different experimental conditions that could have limited detection of optimum Raman signals, e.g., scattering efficiency of laser wavelength (and hence the laser power irradiation over the sample), and the large background due to scattering from the sample itself, substrate, and the microscope objective [21]. Generally, the weak Raman signals that can be potentially useful for disease diagnostics may be swamped in large background effects. Consequently, useful Raman bands may not be optimally detected in all related experiments. It should be noted the works by Corsetti et al. [16], Crow et al. [40], Matias et al., [36] and Martin-Hirsch et al. [41] were based on 785 nm, 830 nm, 830 nm, and 785 nm excitation lasers with values of laser power at 135 mW, 300 mW, 80 mW, and 100 mW, respectively. It is therefore not entirely clear if the laser power played a major role in detection of optimum Raman signals in their experiments.

Our results showed the weak bands in Table 5 were not spectrally visible in Figures 2(b) and 3(a)–3(c) although they were statistically significant ($p < 0.05$). However, as seen in Figures 5(b), 5(d), and 5(f), the score plots demonstrated a reasonable level of scores discrimination, significantly strengthening the view that there were inherent subtle molecular differences between the two cell lines. These current results are encouraging and should motivate utility and a deeper investigation of alternative multivariate analysis methods of detecting weak band Raman alterations (e.g., higher principal components), particularly when performing Raman experiments under limited experimental conditions.

TABLE 6: Comparison of band intensity ratios between malignant (PC3) and normal cells (PNT1a) based on the weak Raman bands at $1076 \, \text{cm}^{-1}$ and $1232 \, \text{cm}^{-1}$.

Raman shift (cm^{-1})	Raman intensities (arbitrary units)								
	Stage 1		Stage 2		Stage 3		Band ratios		
	PNT1a	PC3	PNT1a	PC3	PNT1a	PC3	PC3(1)/PNT1a(1)	PC3(2)/PNT1a(2)	PC3(3)/PNT1a(3)
1076	0.67	0.65	0.64	0.64	0.74	0.77	0.970	1.0	1.04
1232	0.60	0.55	0.74	0.68	0.79	0.78	0.917	0.919	0.987

The numbers 1, 2, and 3 enclosed in brackets identify the stage of cell proliferation.

4. Conclusion

In this study, the utility of Raman spectroscopy in interrogating subtle molecular alterations in cultured PC3 and PNT1a cells in vitro has been demonstrated. The findings from this study suggest that Raman spectroscopy combined with principal component analysis (PCA) has potential applications as a research tool in prostate cancer diagnostics. Analyses using Raman difference spectra, PCA, and Raman band intensity ratios were used to identify prominent and subtle biochemical alterations associated with prostate malignancy. The $566 \pm 0.70 \, \text{cm}^{-1}$, $630 \, \text{cm}^{-1}$, $1370 \pm 0.86 \, \text{cm}^{-1}$, and $1618 \pm 1.73 \, \text{cm}^{-1}$ bands were identified from the difference spectra and were associated with prominent biochemical differences between diseased (PC3) and healthy (PNT1a) cells. The band intensity ratios at $566 \pm 0.70 \, \text{cm}^{-1}$ and $630 \, \text{cm}^{-1}$ were found to increase with the stage of cell proliferation. This suggested that prostate malignancy can be linked to an increase in relative amounts of nucleic acids and lipids, respectively. On the contrary, decreasing band intensity ratios at $1370 \pm 0.86 \, \text{cm}^{-1}$ and $1618 \pm 1.73 \, \text{cm}^{-1}$ suggested that prostate malignancy can be associated with decrease in relative amounts of saccharides and tryptophan, respectively, as the cells proliferate. PCA was used to detect subtle biochemical alterations through analysis of intermediate-order and high-order principal components (PCs). Subtle bands around $1076 \, \text{cm}^{-1}$, (1232, $1234 \, \text{cm}^{-1}$), (1276, $1278 \, \text{cm}^{-1}$), (1330, $1333 \, \text{cm}^{-1}$), (1434, $1442 \, \text{cm}^{-1}$), and (1471, $1479 \, \text{cm}^{-1}$) were observed during proliferation of both cell lines. The band intensity ratios at $1076 \, \text{cm}^{-1}$ and $1232 \, \text{cm}^{-1}$ were found to increase with the stage of cell proliferation, suggesting that prostate malignancy can be associated with an increase in relative amounts of nucleic acids and amide III components, respectively. The results of this study show that subtle biochemical alterations can be detected and extracted from the Raman spectra of prostate cell lines. There is a need for further exploration and quantification of such biochemical alterations, especially at the subcellular level.

Data Availability

The data used to support the findings of this study are available from the corresponding author upon request.

Conflicts of Interest

The authors declare that there are no conflicts of interest regarding the publication of this paper.

Acknowledgments

References

[1] American Cancer Society, *Global Cancer Facts and Figures*, American Cancer Society, Atlanta, GA, USA, 3rd edition, 2015.

[2] F. Bray, J. Ferlay, I. Soerjomataram, R. L. Siegel, L. A. Torre, and A. Jemal, "Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries," *CA: A Cancer Journal for Clinicians*, vol. 68, no. 6, pp. 394–424, 2018.

[3] R. E. Kast, S. C. Tucker, K. Killian, M. Trexler, V. Kenneth, and G. W. Auner, "Emerging technology : applications of Raman spectroscopy for prostate cancer," *Cancer Metastasis Review*, vol. 33, no. 673, pp. 1–21, 2014.

[4] V. Crupi, D. De Domenico, S. Interdonato et al., "FT-IR spectroscopy study on cutaneous neoplasie," *Molecular Structure*, vol. 563-564, pp. 115–118, 2001.

[5] V. Crupi, D. Majolino, P. Migliardo, M. R. Mondello, M. P. Germanò, and S. Pergolizzi, "FT-IR molecular evidence of liver damage by chemical agents," *Vibrational Spectroscopy*, vol. 25, no. 2, pp. 213–222, 2001.

[6] V. Crupi, D. Majolino, P. Migliardo, M. R. Mondello, S. Pergolizzi, and V. Venuti, "FT-IR spectroscopy for the detection of liver damage," *Spectroscopy*, vol. 18, no. 1, pp. 67–73, 2004.

[7] C. Kendall, M. Isabelle, F. Bazant-Hegemark et al., "Vibrational spectroscopy: a clinical tool for cancer diagnostics," *The Analyst*, vol. 134, no. 6, pp. 1029–1045, 2009.

[8] L. M. Moreira, L. Silveira Jr., F. V. Santos et al., "Raman spectroscopy: a powerful technique for biochemical analysis and diagnosis," *Spectroscopy*, vol. 22, no. 1, pp. 1–19, 2008.

[9] A. Chandra, S. Talari, Z. Movasaghi, and S. Rehman, "Raman spectroscopy of biological tissues," *Applied Spectroscopy Reviews*, vol. 50, pp. 46–111, 2015.

[10] M. V. P. Chowdary, K. K. Kumar, J. Kurien, S. Mathew, and C. M. Krishna, "Discrimination of normal, benign, and malignant breast tissues by Raman spectroscopy," *Biopolymers*, vol. 83, no. 5, pp. 556–569, 2006.

[11] A. C. S. Talari, C. A. Evans, I. Holen, R. E. Coleman, and I. U. Rehman, "Raman spectroscopic analysis differentiates between breast cancer cell lines," *Journal of Raman Spectroscopy*, vol. 46, no. 5, pp. 421–427, 2015.

[12] G. Theophilou, K. M. G. Lima, M. Briggs, P. L. Martin-hirsch, H. F. Stringfellow, and F. L. Martin, "A biospectroscopic analysis of human prostate tissue obtained from different time periods points to a trans-generational alteration in spectral phenotype," *Scientific Reports*, vol. 5, no. 13465, pp. 1–14, 2015.

[13] F. L. Magalhães, A. M. C. Machado, S. K. Paulino et al., "Raman spectroscopy with a 1064-nm wavelength laser as a potential molecular tool for prostate cancer diagnosis: a pilot study," *Journal of Biomedical Optics*, vol. 23, no. 12, pp. 1–6, 2018.

[14] K. Aubertin, J. Desroches, M. Jermyn et al., "Combining high wavenumber and fingerprint Raman spectroscopy for the detection of prostate cancer during radical prostatectomy," *Biomedical. Optics. Express*, vol. 9, no. 9, pp. 4484–4489, 2018.

[15] P. Musto, A. Calarco, M. Pannico et al., "Hyperspectral Raman imaging of human prostatic cells: an attempt to differentiate normal and malignant cell lines by univariate and multivariate data analysis," *Spectrochimica Acta Part A: Molecular and Biomolecular Spectroscopy*, vol. 173, pp. 476–488, 2017.

[16] S. Corsetti, T. Rabl, D. Mcgloin, and N. Ghulam, "Raman spectroscopy for accurately characterizing biomolecular changes in androgen-independent prostate cancer cells," *Biophotonics*, vol. 11, pp. 1–8, 2018.

[17] W. L. Martinez and A. R. Martinez, *Exploratory Data Analysis with MATLAB®*, CRC Press, Boca Raton, FL, USA, 2005.

[18] I. T. Jolliffe, *Principal Component Analysis*, Springer-Verlag, New York, NY, USA, 2nd edition, 2002.

[19] M. J. Pelletier, "Quantitative analysis using Raman spectrometry," *Applied Spectroscopy*, vol. 57, no. 1, pp. 20A–42A, 2003.

[20] M. Björklund, "Be careful with your principal components," *Evolution*, vol. 73, no. 10, pp. 2151–2158, 2019.

[21] H. Byrne, L. Kerr, and B. M. Hennelly, "Optimal choice of sample substrate and laser wavelength for Raman spectroscopic analysis of biological specimen," *Analytical Methods*, vol. 7, pp. 1–14, 2015.

[22] I. W. Schie and T. Huser, "Methods and applications of Raman microspectroscopy to single-cell analysis," *Applied Spectroscopy*, vol. 67, no. 8, pp. 813–828, 2013.

[23] W. Wang, J. Zhao, M. Short, and H. Zeng, "Real-time in vivo cancer diagnosis using Raman spectroscopy," *Biophotonics*, vol. 19, pp. 1–19, 2014.

[24] A. Phinyomark, B. A. Hettinga, S. Osis, and R. Ferber, "Do intermediate- and higher-order principal components contain useful information to detect subtle changes in lower extremity biomechanics during running?" *Human Movement Science*, vol. 44, pp. 91–101, 2015.

[25] G. M. Sullivan and R. Feinn, "Using effect size-or why the p value is not enough," *Journal of Graduate Medical Education*, vol. 4, no. 3, pp. 279–282, 2012.

[26] W. Thalheimer and S. Cook, "How to calculate effect sizes from published research : a simplified methodology," 2020, http://work-learning.com/effect_sizes.htm.

[27] M. Jafari and N. Ansari-pour, "Why, when and how to adjust your p values ?" *Cell*, vol. 20, no. 4, pp. 604–607, 2019.

[28] R. L. McCreely, *Raman Spectroscopy for Chemical Analysis*, John Wiley & Sons Ltd., Hoboken, NJ, USA, 2001.

[29] L. S. Jr, K. R. M. Leite, F. L. Silveira et al., "Discrimination of prostate carcinoma from benign prostate tissue fragments in vitro by estimating the gross biochemical alterations through Raman spectroscopy," *Lasers in Medical Science*, vol. 29, no. 4, pp. 1469–1477, 2014.

[30] Z. Movasaghi, S. Rehman, and I. U. Rehman, "Raman spectroscopy of biological tissues," *Applied Spectroscopy Reviews*, vol. 42, no. 5, pp. 493–541, 2007.

[31] J. De Gelder, K. De Gussem, P. Vandenabeele, and L. Moens, "Reference database of Raman spectra of biological molecules," *Journal of Raman Spectroscopy*, vol. 38, no. 9, pp. 1133–1147, 2007.

[32] Z. Huang, A. McWilliams, H. Lui, D. I. McLean, S. Lam, and H. Zeng, "Near-infrared Raman spectroscopy for optical diagnosis of lung cancer," *International Journal of Cancer*, vol. 107, no. 6, pp. 1047–1052, 2003.

[33] P. Crow, N. Stone, C. A. Kendall et al., "The use of Raman spectroscopy to identify and grade prostatic adenocarcinoma in vitro," *British Journal of Cancer*, vol. 89, no. 1, pp. 106–108, 2003.

[34] A. Taleb, J. Diamond, J. J. Mcgarvey, J. R. Beattie, C. Toland, and P. W. Hamilton, "Raman microscopy for the chemometric analysis of tumor cells," *The Journal of Physical Chemistry B*, vol. 110, no. 39, pp. 19625–19631, 2006.

[35] N. Stone, M. C. Hart Prieto, P. Crow, J. Uff, A. W. Ritchie, and A. W. Ritchie, "The use of Raman spectroscopy to provide an estimation of the gross biochemistry associated with urological pathologies," *Analytical and Bioanalytical Chemistry*, vol. 387, no. 5, pp. 1657–1668, 2007.

[36] R. Matias, L. Silveira, M. Augusto, and R. S. Silva, "Diagnostic model based on Raman spectra of normal, hyperplasia and prostate adenocarcinoma tissues in vitro," *Spectroscopy*, vol. 25, pp. 89–102, 2011.

[37] J. Long, C. Zhang1, N. Zhu et al., "Lipid metabolism and carcinogenesis, cancer development," *Americal Journal of Cancer Research*, vol. 8, no. 5, pp. 778–791, 2018.

[38] R. J. Klement and U. Kämmerer, "Is there a role for carbohydrate restriction in the treatment and prevention of cancer?" *Nutrition and Metabolism*, vol. 8, no. 1, pp. 257–294, 2013.

[39] I. Rehman, Z. Movasaghi, and S. Rehman, *Vibrational Spectroscopy for Tissue Analysis*, CRC Press, Boca Raton, FL, USA, 2013.

[40] P. Crow, B. Barrass, C. Kendall et al., "The use of Raman spectroscopy to differentiate between different prostatic adenocarcinoma cell lines," *British Journal of Cancer*, vol. 92, no. 12, pp. 2166–2170, 2005.

[41] I. I. Patel and F. L. Martin-Hirsch, "Discrimination of zone-specific spectral signatures in normal human prostate using Raman spectroscopy," *The Analyst*, vol. 135, no. 12, pp. 3060–3069, 2010.