Integration of near-infrared spectroscopy and aquaphotomics for discrimination of cultured cancerous cells using phenol red

Raypah, E. Muna ; Muncan, Jelena ; Sudik, Suhainah ; Omar, Fairuz Ahmad ; Mail, Hafiz Mohd ; Tsenkova, Roumiana ; Seeni, Azman

(Citation)
Chemometrics and Intelligent Laboratory Systems,227:104611

(Issue Date)
2022-06-30

(Resource Type)
journal article

(Version)
Accepted Manuscript

Integration of Near-Infrared Spectroscopy and Aquaphotomics for Discrimination of Cultured Cancerous Cells using Phenol Red

Muna E. Raypah [1], Jelena Muncan [2*], Suhainah Sudik [1], Ahmad Fairuz Omar [1**], Mohd Hafiz Mail [3], Roumiana Tsenkova[2], and Azman Seeni [4]

[1] School of Physics, Universiti Sains Malaysia, 11800, Pulau Penang, Malaysia
[2] Aquaphotomics Research Department, Faculty of Agriculture, Kobe University, Kobe, Japan
[3] Malaysian Institute of Pharmaceuticals and Nutraceuticals, National Institute of Biotechnology Malaysia, Ministry of Energy, Science, Technology, Environment and Climate Change, 11700 Penang, Malaysia
[4] Advanced Medical and Dental Institute, Universiti Sains Malaysia, Bertam, 13200, Pulau Penang, Malaysia
*Corresponding authors e-mails: fairuz_omar@usm.my and jmuncan@people.kobe-u.ac.jp

Abstract

Presently, mammalian cell lines are the most utilized hosts for the production of biopharmaceuticals. Optical spectroscopy is extensively used for the physical or physiological measurements of cellular features to study the diagnosis, prognosis, and treatment of different cancers. Near-infrared (NIR) spectroscopy and aquaphotomics are ubiquitous techniques increasingly used for nondestructive assessment. The key objective of this study is to evaluate the feasibility of absorbance spectra in the NIR region (700-1100 nm) coupled with aquaphotomics analysis to classify the different cancerous cell lines. Human cervix adenocarcinoma cells (HeLa) and human prostate carcinoma (DU145) were the cancerous cell lines, while normal mouse skin fibroblast (L929 cell line) was used as a reference for the assessment. The NIR absorbance spectra for 50,000 to 275,000 cells of L929, DU145, and HeLa cell lines in culture media prepared with and without phenol red (PhR) were subjected to principal component analysis (PCA) and soft independent modeling of class analogies (SIMCA) models. The PCA of the spectral absorbance data revealed a clear discrimination between the three cell lines, with subgroups detected based on the presence or absence of PhR in the culture medium. The SIMCA method showed a high accuracy of classification, where larger class distances were obtained among the non-stained cells.

Keywords: Cancerous Cells; Culture Medium; Phenol Red; Near-infrared Spectroscopy; PCA; SIMCA; Water.

1. Introduction

The prevalence of cancer is persistently rising, resulting in an increased annual global mortality rate. With over 17 million new cases annually and approximately 10 million deaths, cancer persists as one of the foremost health complications, globally [1]. Cancer is a complex disease, typified by reprogrammed signaling, cellular physiological alterations, and uninhibited cell growth. A characteristic trait of cancer cells is their metabolic reprogramming, which facilitates fast cellular reproduction, migration, and alteration of their microenvironment, ultimately leading to metastasis [2]. Surgical biopsy is the gold standard method for cancer diagnosis and has improved in speed and accuracy. However, the surgical biopsy is painful for patients and not suitable for early cancer screening. Liquid biopsy is a non-invasive cancer diagnosis method that detects highly sensitive biomarkers and could be used in early cancer detection. This technique is less efficient due to the cell loss or other cell contamination and potential uncertainties in the subsequent immunoassay. X-ray computed tomography (CT) technology is widely used in recognizing cancer, but it is not suitable for regular screening due to its toxicity. Magnetic resonance imaging (MRI) is a non-toxic technique, whereas it is expensive and time-consuming. Fluorescence imaging is a molecular imaging technique that is widely used in biological sample detection. Other methods such as single-photon emission CT, positron emission tomography, ultrasound, and endoscopy are also commonly used for the diagnosis of cancer. However, all imaging techniques are having considerably low sensitivity for the detection of a small number of cancer cells [3].

In the past 20 years, optical spectroscopy techniques have widely emerged as a non-invasive substitute for different conventional measurement techniques, and as a tool for medical diagnosis. The optical techniques are economic, portable with no side effects, and offer good spatio-temporal

resolution, and real-time functional information. The bulk changes in cell architecture as the tumor grows can be employed for the diagnosis of cancerous cells under light microscopy, thereby enabling a definitive diagnosis of the disease. These changes can also be utilized to determine the stage of cancer growth. Recently, optical spectroscopy has shown its potential non-destructive capability of identifying distinct spectral attributes that discriminate between tumor and normal cell. These marker bands provide a basis for the identification and therapeutic screening of several cancers [4]. The optical spectroscopy technique has been extensively used for detecting various types of cancer [5-7]. It is strongly believed that vibrational spectroscopy offers excellent potential to study the chemical structural characteristics of biological samples. Raman, Fourier transform infrared (FTIR), and NIR spectroscopic techniques collect highly specific information making them suitable for cell analysis. Raman and FTIR spectroscopy techniques are considered potential cancer diagnostic tools. These techniques have been well studied for discriminating normal and cancer states in various biological analytes including cells [8]. Fluorescence imaging and Raman spectroscopy have been applied for the accurate diagnosis of cancer [9]. In addition, the surface-enhanced Raman scattering (SERS) technique is generally used to enhance the conventional Raman scattering signal [10]. In recent years, SERS has been widely used in the diagnosis of various types of cancer [11-16]. Visible (Vis) and near-infrared (NIR) spectroscopy and imaging have been widely utilized for detection, diagnosis, and screening of cancerous cells. NIR spectroscopy has evolved considerably into a versatile technique in medical diagnosis, particularly for the diagnosis of carcinoma [17]. NIR spectroscopy plays a vital role in cancer detection, early prediction of responders during therapeutic interventions, and optimizing the treatment approach [18]. The supposed “optical window” occurs in the NIR part of the spectrum at the wavelength range of 650-1100 nm, which is suitable for most of the non-invasive measurements of biological

systems [19]. This overtone region is also referred to as the “NIR window” or “therapeutic window” for the lower and measurable light absorbance and the richness of information. In addition, the short-wave NIR region (700-1100 nm) is appropriate for in-line and in-situ field measurements using fiber-optic probes and relatively low-cost silicon detectors [20, 21].

Vibrational spectroscopy gives valuable information on the chemical composition based on functional groups detection and spectral analysis of the obtained fingerprints [22] that can be used to characterize normal and cancerous cells. Several studies revealed NIR region as a spectroscopic biomarker that has promising applications and a future design of specialized spectroscopic instrumentation. The NIR spectral changes of mammary gland tissues in the carcinogenic processes of rats were examined. It was found that intensities of DNA bands (1471 and 1911 nm) and water bands (967, 1154, 1402, and 1888 nm) were relatively increased while those of the lipid bands (1209, 1721, and 1764 nm) were reduced in the cancerous site [23]. The content of water in cancerous and normal prostate was analyzed using NIR spectroscopy (400-2400 nm). The water absorption peaks at 1444 and 1944 nm observed in prostate tissues which are related to OH overtone vibrational bands. It was shown that cancerous tissues contained less water than normal tissues [24]. The application of NIR spectroscopy for the detection of human primary pancreatic and colorectal cancers showed that major spectral differences were in the CH-stretching first (6000-5400 cm^{-1}) and second overtone (9000-7900 cm^{-1}) regions [25]. The most significant differences in NIR spectra from malignant and benign colorectal tissues were observed at the CH-stretching second overtone region and water’s first overtone of combination bands (1100-1330 nm) with bands associated with glycoproteins, glycolipids, and carbohydrates as at OH and NH first overtone and CH first overtone combinations (1400-1600 nm and 1300-1420 nm) [26]. Normal and diseased breast tissues were examined by Raman spectroscopy. The spectra of

diseased breast tissue (benign and malignant) showed markedly diminished to absent contributions from lipids (at ~1082, 1302, 1444, and 1652 cm^{-1}) and reduced contributions from carotenoids (at ~1004, 1156, and 152 cm^{-1}) with an absence of the peak at ~1358 cm^{-1} [27]. The FTIR spectroscopy system analyzed prostate cancer cell lines. The ratio of peak intensities at 1030 and 1080 cm^{-1} used as a diagnostic marker to distinguish between the transformed normal cell and cell lines derived from various metastatic sites. This peak corresponds to the glycogen/phosphate ratio and is indicative of the metabolic turnover of the cell [28]. The application of laser trapping Raman spectroscopy to analyze leukemia cells was investigated. Raman markers associated with DNA (1093 and 785 cm^{-1}) and protein (1447 and 1126 cm^{-1}) vibrational modes have been found to exhibit excellent discriminating power for cancer cell identification [29]. Likewise, the Raman spectra at region 700-1750 cm^{-1} of malignant and bladder cancer cells denoted a superior concentration of proteins and nucleic acids in the bladder cells as in cancer cells. Proteins and nucleic acids were more abundant in MGH-U1 than PC-3 cells, while lipids and carbohydrates were more abundant in PC-3 cells [30]. Normal and four different types of human lung cultured cancer cells were identified by Raman spectroscopy. Strong bands at 748, 1129, and 1586 cm^{-1} were assigned to cytochrome c (cyt-c). The strong appearance of these bands suggested that the cancer cells were rich in cyt-c relative to phenylalanine, and possibly rich in mitochondrial cyt-c or mitochondria compared to normal cells [31]. Raman spectroscopic analysis was used to differentiate between normal breast (MCF-10A) and breast cancer (MCF-7 and MDA-MB-436) cell lines. The spectra of the cell lines showed spectral information about proteins, lipids, and nucleic acids. It was shown that the cancer cells presented high lipidic and proteinic information in 3050-2800 cm^{-1} and 1800-500 cm^{-1} regions [32]. Raman spectroscopy was used to evaluate the biomolecular cascade events related to the conversion of a normal cell into an invasive breast

cancer cell. It was observed that the lipid levels were increased for the invasive cells compared with normal ones in regions 2800-3000 cm^{-1} and 700-1800 cm^{-1} [33]. The FTIR spectra to study the development of skin cancer showed that the absorption band at approximately 3062 cm^{-1} was increased, indicating that most of the proteins had the configuration of amide B and the β-sheet protein structure predominated [34]. FTIR-ATR and FT-Raman absorption spectra were used to study the changes in oral cancer. The peak at 1238 cm^{-1} was correlated with nucleic acids symmetrical stretching, the phosphate wavenumbers were lower compared with that of normal tissue, and the shifting of 1030 cm^{-1} was ascribed to -CH_2OH vibrations [35]. FTIR spectroscopy technique was applied to explore the serum characteristics in breast cancer. It was found that the range of 3090-3700 cm^{-1} is the criterion for differentiating breast cancer serum samples from the healthy ones, which can be attributed to protein modifications [36]. With SERS, the normal and cancerous liver tissue were analyzed in the fingerprint region (500-1800 cm^{-1}). The relative intensities of the characteristic vibration peaks at 838 cm^{-1} (amine groups), 1448 cm^{-1} (collagen), and 1585 cm^{-1} (protein and hemoglobin) are significantly changed in the cancerous tissues [37]. FTIR and Raman spectra were obtained from oral cancer cells to discriminate between normal, pre-cancerous, and cancerous conditions. Compared to normal patients, significant differences were observed at 1550, 1580, 1640, 2370, 2330, 2950-3000 and 3650-3750 cm^{-1} for FTIR and 520, 640, 785, 827, 850, 935, 1003, 1175, 1311 and 1606 cm^{-1} for Raman vibrations. The increase in DNA, protein, and lipid contents with malignancy was more evident [38].

The combination of analytical tools has been proven to be of immense value in a variety of bio-scientific applications. Spectroscopy can be coupled with chemometrics to offer a rapid and versatile method for the characterization of raw materials to advance cell culture performance [39]. Presently, artificial intelligence practices are widely applied in the field of computational biology

and bioinformatics [40, 41]. The absorption of molecules in the NIR region results mostly from the absorption of overtones and the merging of stretching-bending bands of atomic groups comprising hydrogen-bearing compounds and complexes that include CH, OH, and NH [42]. The overlap of signals detected in the NIR range increases the broadness of the peaks, resulting in composite spectra and hindering the designation of distinct features to specific compounds. The spectroscopic techniques coupled with multivariate data analysis in mammalian cell culture have been employed to derive data on several bioprocess variables, that include culture medium composition (nutrients and metabolites), viable cell concentration, and number of living cells [43-46].

Biomarkers have shown to be more favorable indicators, and the disparity in their contents is directly associated with specific diseases including cancer. Biomarkers can be specific cells, genes, gene products, hormones, or other molecules that are detectable in tissues or body fluids. However, in the case of early diagnosis, the concentrations of biomarkers are typically very low and difficult to determine using conventional techniques. Therefore, searching for more distinct biomarkers or improving the techniques of measurement remains the focus of research in the field of cancer detection and diagnosis. Given that a disease impairs the body functions at various levels of system organization, the collective impact of these induced variations in water can be measured by means of spectroscopic techniques. Aquaphotomics presents the spectral pattern of water molecular structure as a “molecular mirror” and a novel integrative biomarker [47]. This approach offers a new framework for elucidating variations in water molecular systems obtainable as a water spectral pattern [48, 49]. The rationale is to identify and expand the database of water absorption bands and decipher the distinctive water absorption patterns that can be utilized as biomarkers [50]. Information on water provides a potential diagnostic tool to improve the determination and

imaging of tumor cells based on variation in vibrational overtones of H_2O molecules in the NIR region. The molecular structure of H_2O is proposed as a novel fingerprint marker to signify changes in the NIR region for early detection of various types of tumor cells. So far, no studies can be found in a literature on NIR spectroscopy at short NIR region (700-1100 nm) and aquaphotomics of cultured cancerous cells.

In cell culture, spectroscopic technologies combined with multivariate data analysis can be used to obtain information regarding bioprocess variables in NIR spectra [51]. Generally, media formulations for mammalian cell culture procedures frequently apply phenol red (phenolsulphonephthalein) as a pH indicator, to visually and qualitatively monitor metabolism [52, 53]. Phenol red (PhR) has been utilized in different colorimetric applications, such as the pH assessment of freshwater [54], the diagnosis of filarial infection [55], the measurement of H_2O_2 generated by cells in culture [56] and CO_2 pressure in carbonated liquids [57], and aiding biopsy procedures to verify Helicobacter pylori-infected areas for patients diagnosed early with gastric cancer [58]. PhR also has several other biochemical applications [59-63]. The absorption spectra associated with PhR are differentiated by two absorption bands with two maxima at 438 nm and 559 nm at pH 7.4 [63].

To date, information on cancer cells using optical spectroscopy techniques is not extensively available, especially for qualitative evaluation of cells. This study aims to explore the interaction between light in the short NIR region (700-1100 nm) and cultured cancerous cells and identify the relationship between the absorption spectra and the characteristics of the cancerous cells. The focus of this research is to examine the accuracy of NIR spectroscopy to detect the concentration of different types of cultured cancerous cells and the potential to develop a spectral signature as a cancer biomarker using aquaphotomics. So far, the spectroscopic analysis of cancer cells in the

NIR region using phenol red dye has not been reported. Direct comparisons of the growth of three different cell lines were performed in the culture medium in the presence and devoid of phenol red. The cancerous cells utilized were cervical (HeLa) and prostate (DU145), while the mouse skin fibroblast (L929) cell line was used as a reference normal cell. The absorbance spectra for 50,000 to 275,000 cells of L929, DU145, and HeLa cell lines were measured. Principal component analysis (PCA) and soft independent modeling of class analogies (SIMCA) models were used for exploration and classification analyses. Multivariate analysis of NIR absorption spectra for mammalian cell cultures using phenol red in culture medium in addition to the spectral analysis of non-stained cell culture could be an effective and more viable substitute for biomedia characterization. Besides, aquaphotomics improves the understanding of cancer cell proliferation from the perspective of water molecular structure.

2. Materials and Methods

2.1 Cell Lines

The normal and carcinoma cell lines utilized in this study were procured from American Type Culture Collection (ATCC), Manassas, VA, USA. The cell lines include mouse skin fibroblast (L929), human cervix adenocarcinoma (HeLa), and human prostate carcinoma (DU145). Growth assays were prepared for all cell types in a serum-supplemented culture medium (Minimum Essential Medium 'MEM' and Dulbecco's Modified Eagle Medium 'DMEM'). The L929 cell was cultivated in the MEM, while HeLa and DU145 cells were cultivated in the DMEM supplemented with fetal bovine serum 'FBS' (10%), sodium pyruvate (1 mM), and penicillin (100 units/mL)/streptomycin (100 µg/mL). All growth assays were performed in MEM and DMEM with and without phenol red.

The growth assays were prepared from stock cultures harvested using trypan blue dye treatment. The preparation procedures of the subcultures of the cells are reported in our prior publications [64, 65]. The cell cultures were maintained at 37 ℃ in an incubator with 5% of CO_2 environment. The cultured cells were harvested from T25 tissue culture flasks and suspended in the culture medium. The cells were subsequently stained with trypan blue dye at a 1:1 ratio. Afterwards, cell viability test was carried out in parallel using the trypan blue dye exclusion technique by quantifying the number of viable cells in a hemocytometer. The trypan blue infiltrates the membrane of non-viable cells, which are then stained blue and can thus be differentiated from viable cells [66]. The cell concentration and total number of cells were calculated using Eqs.1 and 2 and presented in Tables 1 and 2.

$$C\ (cells/mL) = (viable\ counted\ cells/Q) \times Df \times Hf \qquad (1)$$

$$Total\ cell\ number\ (cells) = C \times V \qquad (2)$$

where C, Q, Df, Hf and V denote cell concentration, quadrant count (the number of squares is 4 in the haemocytometer and each square has a dimension of 1x1 mm with a 0.1 mm depth), dilution factor (which is typically 2), haemocytometer factor with value of 10^4 (also referred to a correction factor), and the sum of original volume of cell suspension, respectively. The cell lines were seeded in a 6-well culture plate with total cell numbers ranging from 50,000 to 275,000 cells per well. The total volume of culture medium in each well was 2 mL and the number of total cell seeding was derived using Eq. 3 below:

$$M_1V_1 = M_2V_2 \qquad (3)$$

where M_1, V_1, M_2 and V_2 represent the concentration of cell suspension, initial volume of cell suspension, concentration of seeded cells, and volume of seeded cells, respectively. The cells in the media were appended to the well and then incubated for an entire day. The total number of

cells per well, M_2, V_1, and V_2 are presented in Tables 3 and 4. The microscopic images (100x total magnification) of the cells were obtained using an inverted microscope (Olympus modelCKX41SF) in order to decipher the cell morphology and distribution.

Table 1. Number of viable cells, concentration of cell suspension, volume of cell suspension, and total number of cells for cell lines in culture medium with phenol red.

Cell line	Number of viable cells (cells)	Concentration of cell suspension (cells/mL)	Volume of cell suspension (mL)	Total cell number (cells)
L929	581	2 905 000	5	14 525 000
HeLa	436	2 180 000	5	10 900 000
DU145	325	1 625 000	5	8 125 000

Table 2. Number of viable cells, concentration of cell suspension, volume of cell suspension, and total number of cells for cell lines in culture medium without phenol red.

Cell line	Number of viable cells (cells)	Concentration of cell suspension, (cells/mL)	Volume of cell suspension (mL)	Total cell number (cells)
L929	528	2 640 000	6	15 840 000
HeLa	673	2 019 000	6	20 190 000
DU145	211	1 055 000	5	5 275 000

Table 3. Initial volume of cell suspension, concentration of cell seeding, volume of cell seeding, and total number of cells per well for L929, HeLa, and DU145 cells cultured in medium with phenol red.

Initial volume of cell suspension (mL), V_1			Concentration of cell seeding (cells/mL), M_2	Volume of cell seeding (mL), V_2	Total number of cells per well (cells)
L929	HeLa	DU145			
60.2	80.3	92.3	25 000	2	50 000
90.4	120.4	138.5	37 500	2	75 000
120.5	160.6	184.6	50 000	2	100 000
150.6	200.7	230.8	62 500	2	125 000
180.7	240.8	276.9	75 000	2	150 000
210.8	281.0	323.1	87 500	2	175 000
241.0	321.1	369.2	100 000	2	200 000
271.1	361.2	415.4	112 500	2	225 000
301.2	401.4	461.5	125 000	2	250 000
331.3	441.5	507.7	137 500	2	275 000

Table 4. Initial volume of cell suspension, concentration of cell seeding, volume of cell seeding, and total number of cells per well for L929, HeLa, and DU145 cells cultured in medium without phenol red.

Initial volume of cell suspension (mL), V_1			Concentration of cell seeding (cells/mL), M_2	Volume of cell seeding (mL), V_2	Total number of cells per well (cells)
L929	HeLa	DU145			
56.8	44.6	142.2	25 000	2	50 000
85.2	66.9	213.3	37 500	2	75 000
113.6	89.2	284.4	50 000	2	100 000
142.0	111.4	355.5	62 500	2	125 000
170.5	133.7	426.5	75 000	2	150 000
198.9	156.0	497.6	87 500	2	175 000
227.3	178.3	568.7	100 000	2	200 000
255.7	200.6	639.8	112 500	2	225 000
284.1	222.9	710.9	125 000	2	250 000
312.5	245.2	782.0	137 500	2	275 000

2.2 Spectra Acquisition

Vis-NIR spectroscopic technique was used to obtain absorbance spectra of each cell line cultured in a medium with and without PhR. The spectroscopy instrument used in the experiments is from Ocean Optics Inc. (Dunedin, Florida, USA). The spectrometer features and spectroscopic system are similar to those reported in our preceding studies [64, 65]. The measurement of the spectra of the cultured cell lines was achieved using a custom-prepared chamber. Three wells (34.8 mm diameter and 17.65 mm depth) were used from the 6-well cell culture plate. Both optical fiber cables (retrieving and illumination fibers) were connected with a collimating lens to ensure parallel transfer of light rays towards the well and from the sample to the spectrometer. The spot area of the light measured on the bottom of the well was 38.48 mm^2. The illumination light source was connected to a tungsten halogen light source, HL-2000, with a range of emission wavelengths between 360 and 2400 nm and a color temperature of 2800 K. The absorbance measurement was

calibrated using an empty well. The resultant light from the interaction with the cells in the well was transmitted through the well and collected using a retrieving fiber. The fiber was connected to a QE65000 spectrometer with a spectral sensitivity between 350.64 and 1131.24 nm. However, the entire analysis only utilized the wavelength range between 400 and 1100 nm. The acquisition parameters for the spectrometer including integration time, scan to average, and boxcar width were 17 ms, 8, and 3, respectively. Triplicate spectra were acquired for each batch to guarantee the reliability of the measurement and to obtain the average spectrum for multivariate analyses. Spectra Suite Software (Ocean Optics) was used to acquire and analyze the spectra.

2.3 Data Analysis

Principal Component Analysis (PCA) [67] was used as an exploratory analysis of the Vis-NIR spectral data. It is a method for reduction of data dimensionality by orthogonal matrix decomposition. The results of PCA are visualized using scores and loadings plots, where scores are projections of original spectra in the pattern spaces defined by principal components, while loadings show the weight coefficients of original variables. The principal components are orthogonal to each other and define spaces that capture maximum variation in the data. The scores plots allow easy observation of existing data patterns while loadings help analyze the spectral features and how they relate to the observed patterns. Soft independent modeling of class analogies (SIMCA) [68] is a supervised pattern recognition technique used for classification purposes. It is a PCA-based method, that employs PCA models for each of the previously assigned classes in data. Classification of spectra is based on a comparison of Mahalanobis distance, which is the distance between the spectrum and the centroid of each class. The spectra are classified as belonging to the class if the Mahalanobis distance is less than three standard deviations from the

class centroid [69]. Before SIMCA analysis, the spectral data were smoothed using Savitzky Golay 2nd order polynomial filter (21 points) [70], corrected for baseline effects using standard normal variate (SNV) transformation [71], and mean-centered [72]. Both PCA and SIMCA analyses were performed using commercially available multivariate analysis software, Pirouette (version 4.5, Infometrix, Bothell, WA).

3. Results and Discussion

CO_2 and lactate are generated via cellular respiration, which can increase the acidity level of the culture medium in due course, leading to a change in microenvironment of the cells, which is attributable to the small volume of the growth medium [73]. In addition, a rise in number of cells and cellular respiration increase the acidity levels of the cell cultures, leading to yellowing in colour of media formulations containing PhR. The cancer cells require energy metabolism for the stimulation of mitosis (i.e. cell division) and tumour growth, thus high rates of glycolysis are manifested, resulting in the generation of high amounts of lactate. This subsequently leads to increase level of acidity in the microenvironment [74-76] and theoretically explains the relatively faster change in the PhR colour for the cancer cells medium compared to the normal cell line. The comparatively higher intensity in the spectra of cancer cells could be due to variations in scattering, size and thickness of the cancer specimens, which influences the spectral absorbance due to the different depths of photon penetration [77]. It is evident that the absorbance of the samples in the NIR region is neither affected nor predisposed to the colour of the medium.

NIR spectroscopy has the benefit of ease and expediency when applied to aqueous samples [78]. The absorbance spectra in the NIR region (700-1050 nm) of the three cell lines with and without PhR are displayed in Figures 1 to 3. Each spectrum in Figures 1-3 was smoothed using the

Savitzky-Golay algorithm with 2nd order polynomial and a window length of 15 points in order to reduce the noise presents in the raw data. The absorbance values on the y-axis represent the total light attenuated through the sample as well as light scatterings. Each spectrum is a result of nine readings of measurements from three wells of the 6-well culture plates. The measurement of the

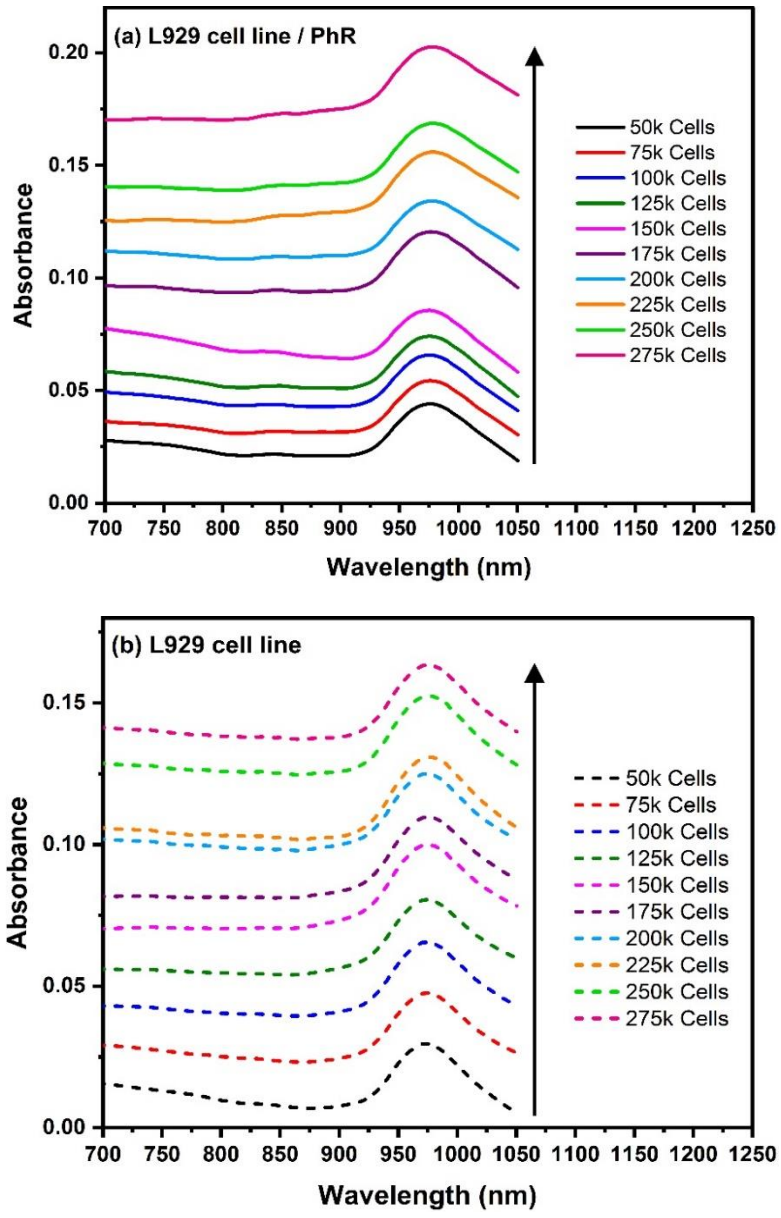


Figure 1. NIR absorbance spectra of L929 cells cultured in medium in: (a) the presence of phenol red and (b) the absence of phenol red.

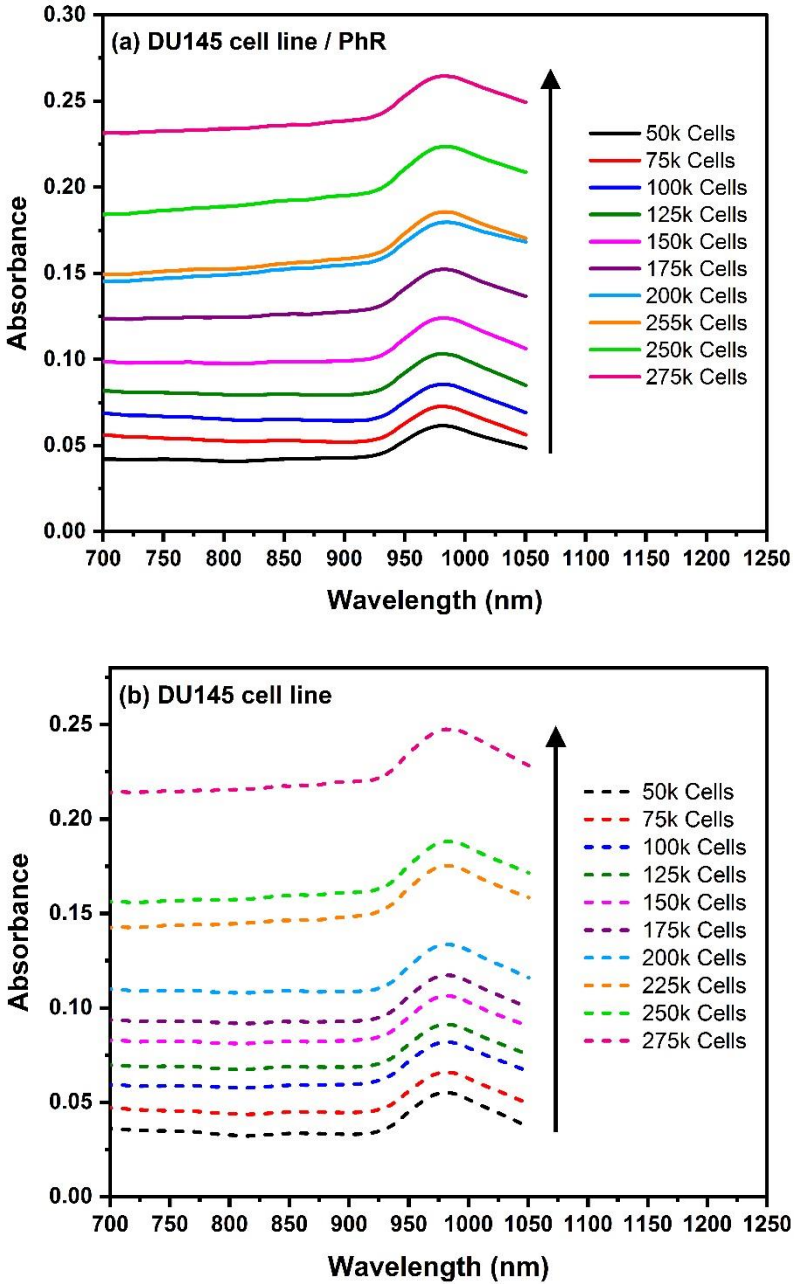


Figure 2. NIR absorbance spectra of DU145 cells cultured in medium in: (a) the presence of phenol red and (b) the absence of phenol red.

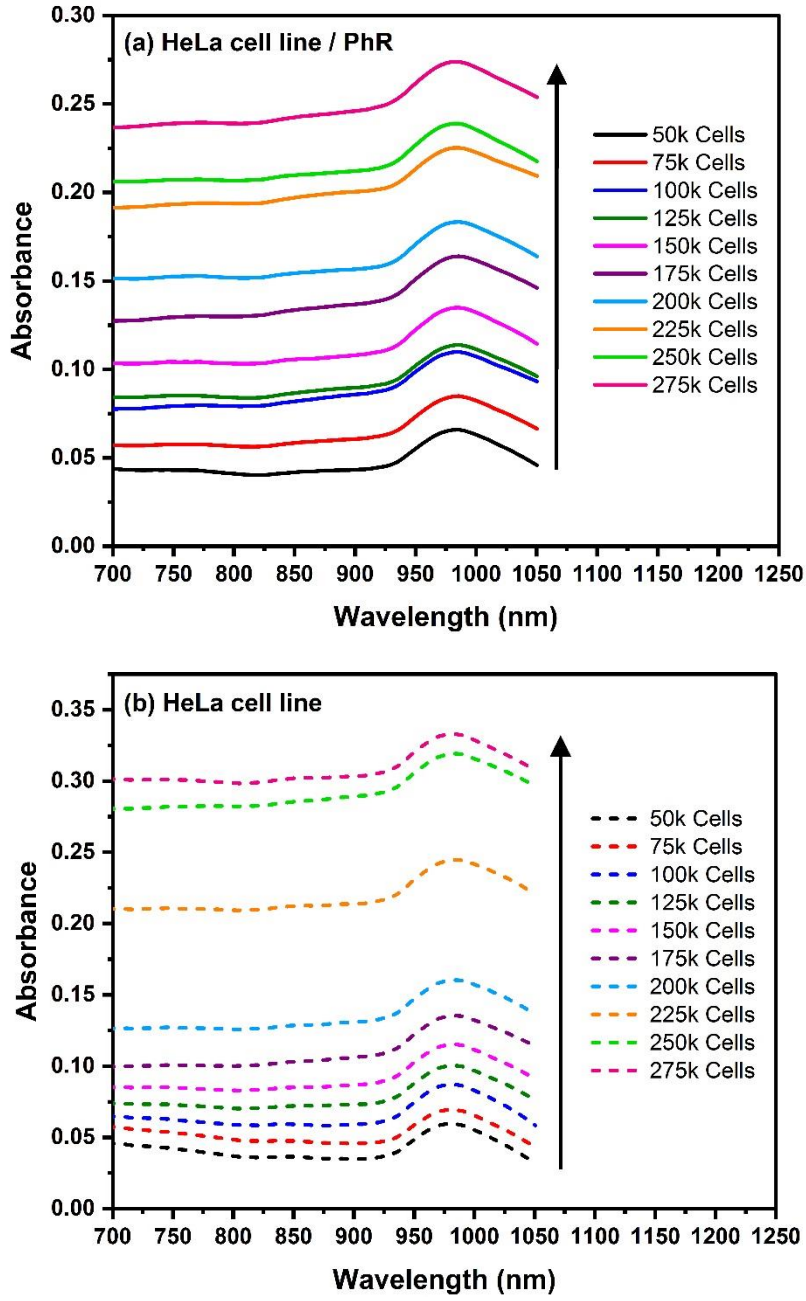


Figure 3. NIR absorbance spectra of HeLa cells cultured in medium in: (a) the presence of phenol red and (b) the absence of phenol red.

spectrum from each well was repeated three times to ensure the repeatability of the measurement and eliminate uncertainties due to sampling. Each spectrum in Figures 1 to 3 is the average of the spectra obtained from three wells of the 6-well plate. Light attenuation apparently increases with

the number of cells in the well. Furthermore, given that the culture medium is composed of water, the distinctive absorbance peaks observed at NIR wavelength (≈ 965-985 nm) denotes the water absorbance overtones [24]. The absorbance of cancerous cells slightly exceeds that of the normal cells for each increment in the number of cells. It is imperative to mention that the absorbance values generated by the spectrometer represent the loss of light due to both absorbance and scattering phenomena. Therefore, the observed absorbance can occur when light is scattered by a specific number of cells.

The morphology of cells is used as a basis for differentiating the normal and cancerous cells. Cancer cells are characterized by morphological variations as compared to normal cells. Morphologically, the cancerous cell is differentiated by several features that include a large nucleus, asymmetrical size and shape, prominent nucleoli, scarce cytoplasm and intense or pale colors [79]. The morphology of the cells was assessed prior to the spectroscopic measurements to guarantee the integral state of the cells without any biological impairment. The cells were microscopically examined to check for abnormal components and aberrations. The color of the cells was then inspected using PhR indicator. Microbial contamination is indicated by color alteration of PhR from plain red to cloudy pink. The morphologies of L929, HeLa, and DU145 cells prepared in culture media with PhR are displayed in Figure 4. The images were obtained using an inverted microscope with 100x total magnification and 0.25 numerical aperture of the 10x objective lens. The obviously higher scattering of cancerous cells compared to normal cells is attributed to their larger nucleus, variable sizes, and diverse shapes. The dimension and size of the nuclei as well as the change of the refractive index of cells have a specific effect on the light scattering of the entire cell [80]. In addition, the mitochondria and other small organelles in the cell also contribute to the light scattering [80]. The size of HeLa cells is larger than those of DU145

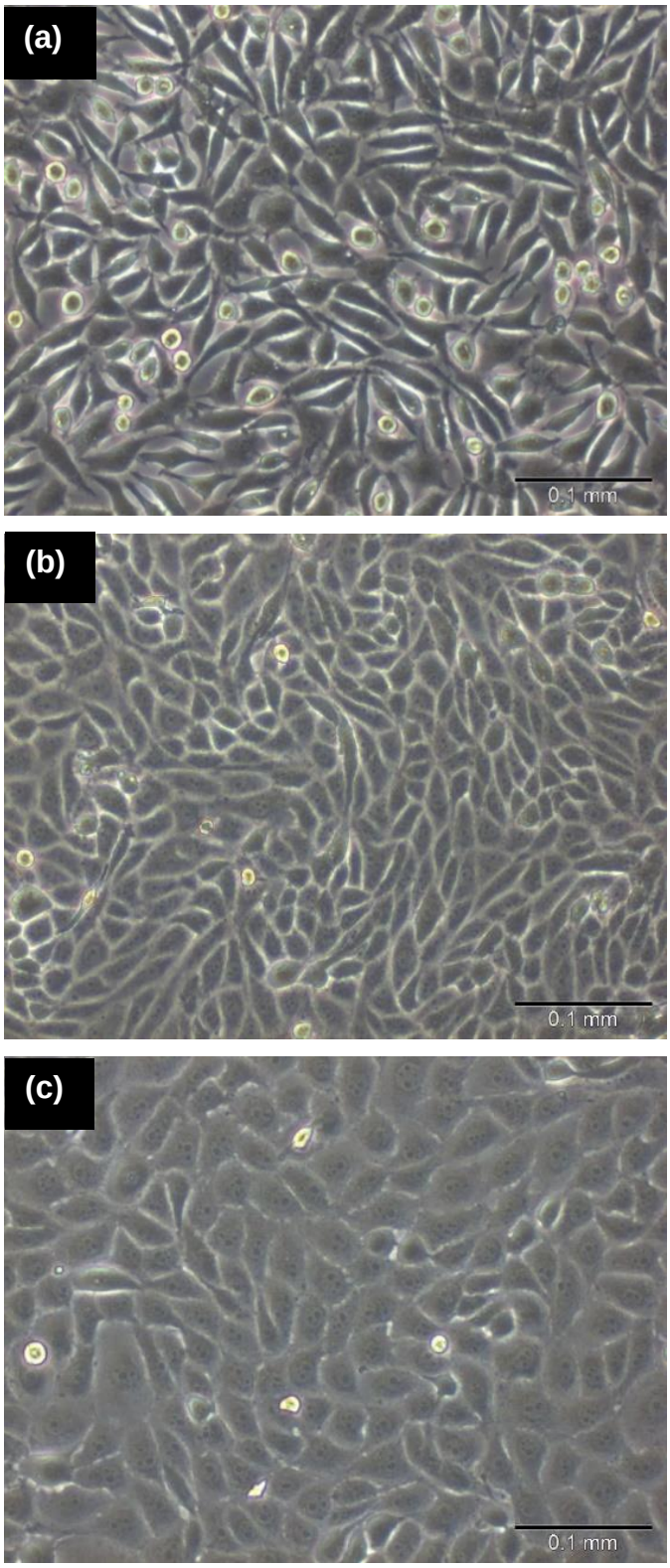


Figure 4. Morphology of (a) L929, (b) DU145, and (c) HeLa cell lines with 100x magnification of an inverted microscope.

and L929 cells, which plausibly explains their higher light scattering [64]. Additionally, increasing the number of cells enhances the probability of light scattering back into the detector, which results in an increase in the measured absorbance of the samples [81]. As observed in the images, the size of HeLa and DU145 cancerous cells is larger than those of L929 cells. The L929 cells are round and oval shaped with a standard size, while HeLa and DU145 have diverse sizes and shapes that are easily discernible at a low number of cells. It was shown that the proliferation or growth of the cancerous cells was slower than normal cells with DU145 cells being the slowest. In addition, it was observed that cell morphology changes as the number of cells increases.

The spectral changes are reflected by shift and broadening/narrowing of distinct water absorption peak that emerge in the NIR region at 975 nm. As observed in Figures 1-3, the characteristic peak of water (≈ 965-985 nm) for cancerous cells is broader compared to normal cells. This peak is attributed to the blend of the second harmonic of O-H symmetric stretch vibration and the intrinsic anti-symmetric stretch vibration from hydrogen bound O-H. In addition, the attributes of this peak are sensitive indicators of the microenvironment of H_2O molecules [82]. The specificity and precision of the shape and position of NIR peak are highly responsive to the molecular character of H_2O [83]. It is well-known that the fraction of adsorbed or bound H_2O measured using optical techniques can relay the molecular vibrational states of H_2O that are related to macro molecular complexes in tissues [84], and may provide further understanding as regards tissue pathophysiology. As a result of the increased volume of H_2O that bind to macromolecules such as proteins, the characteristic absorption peak of H_2O at 970 nm is subjected to both shifting and broadening. The spectral characteristics of H_2O absorption are dependent on the binding state of H_2O. As the volume of hydrogen-bound H_2O molecules decreases, the intensity of water absorption peak increases, the bandwidth becomes narrower, and the peak is shifted to higher

energy (blue-shift). These changes are due to imbalance in the symmetry between hydrogen-bound and free H_2O molecules. Hydrogen bonding between H_2O and macromolecules such as proteins initiates additional spectral broadening and red shifting of the 970 nm water peak [83]. Also, the symmetry of the band increases as the number of cells increases, with a shift in the maximum absorption (Figures 1 to 3), which is possibly due to the increased number of free O-H. This explains the eventual increase in absorbance [85] as the number of cells increases.

The spectra in the range of 400 to 1100 nm were investigated using PCA analysis, with the edges of the spectra trimmed to prevent noise. The samples were designated with "Y" and "N", for instance, L929 Y-1 and L929 N-1 denote the L929 cells with and without PhR for 50,000 cells, while the samples labelled as L929 Y-1 to L929 Y-10 signify the L929 cells for 50,000 to 275,000 cells. Results of PCA analysis of the spectra for L929, HeLa, and DU145 cell lines are shown in Figure 5. The PCA scores plots reveal the hidden structures in spectral data. The first five principal components (PC1 to PC5) explained over 99.99% of the variance in data and their loadings are illustrated in Figure 6. The first principal component (PC1) indicating the highest variance (88.69%) is related to the overall number of cells, where the scores move in the direction of the PC1 axis toward the positive part as the number of cells increases (Figure 5(a)). In the same space, three distinct groups are observable along the PC2 axis. In addition, Figure 5(a) shows the PC1 versus PC2 scores plot which reveals separation along the PC2 axis between the cells stained with PhR (located in the negative part of the PC2 axis) and non-stained cells (located in the positive part of PC2 axis). In the cluster of scores corresponding to PhR stained cells, another separation along the PC2 axis can be observed between cancer and normal cells. It can be seen that the scores of cancer cells are located close to zero, while scores of normal cells are further below the negative part of PC2. This kind of sub-separation is lacking in the scores of non-stained cells. Figure 5(b)

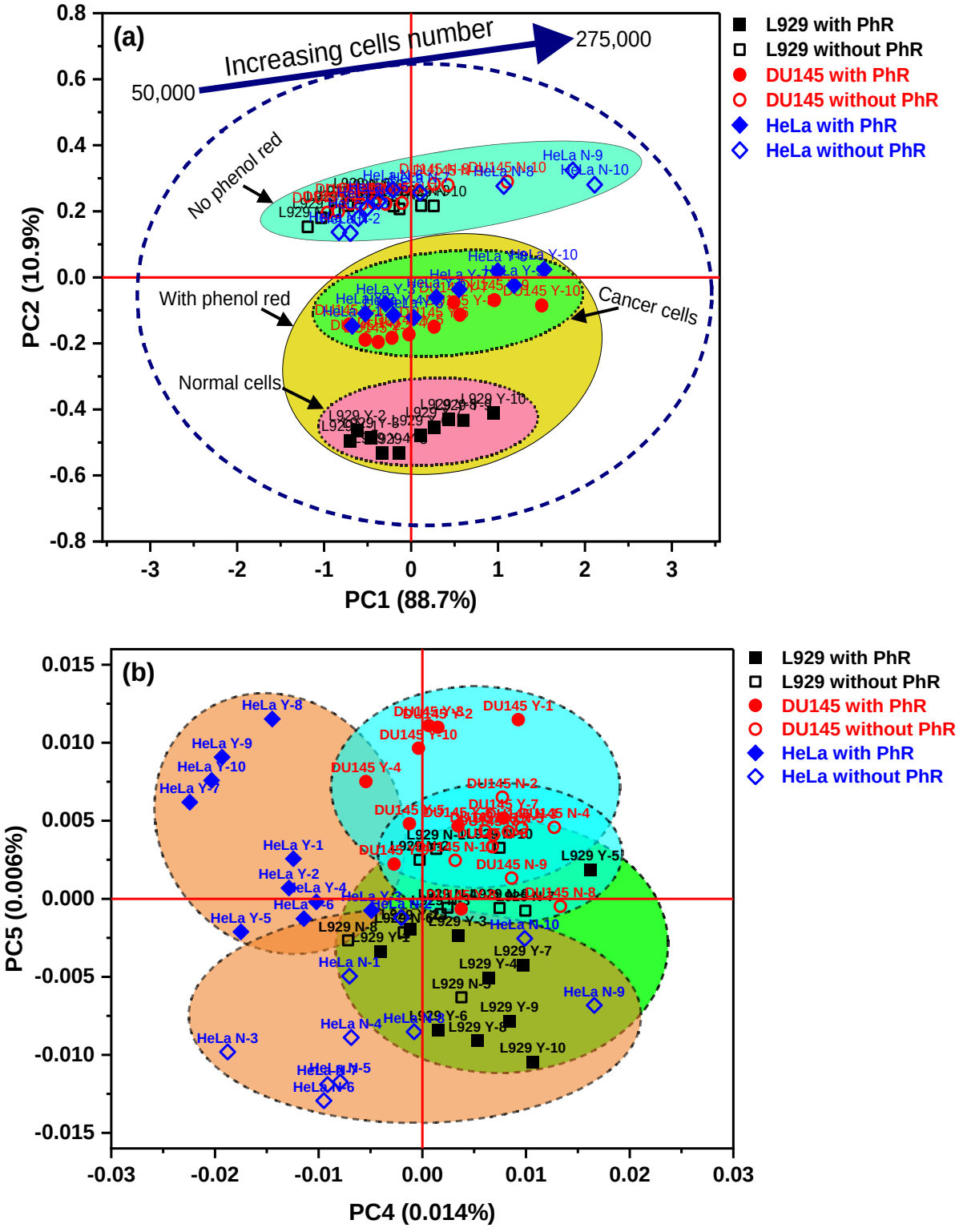


Figure 5. Scores plots of PCA analysis of Vis-NIR raw spectra (400-1100 nm) for L929, DU145, and HeLa cell lines: (a) PC1 versus PC2 and (b) PC4 versus PC5.

presents PC4 versus PC5 scores plot which further reveals the separation of scores between PhR stained and non-stained cancer cells along the PC5 axis. PhR stained HeLa cells scores are located in the positive part of PC5 and negative part of PC4, while scores of HeLa cells without PhR are located entirely in the negative part of PC5. Similarly, both scores of PhR stained DU145 and PhR non-stained DU145 located in the positive part of PC4 and PC5 axes, show some separation along the PC5 axis. The scores of the control cells L929 form one group located mostly in the positive part of the PC4 axis, without distinction between the scores belonging to stained and non-stained classes.

The loadings of the first two principal components are shown in Figures 6(a) and (b). The predominant peak for both loadings is located at 558 nm, which denotes the basic form (deprotonated base form) of phenol red [86, 87]. The peak is positive in PC1, indicating that the absorbance at this wavelength increases concomitantly with an increase in the number of cells, which conforms to the observed trend of scores towards the positive part of the PC1 axis. Besides the colour peak, it is discernible in the NIR part of the spectra that the shape of the PC1 is similar to the baseline of the spectra, with the graph being dominated by an offset and slope. This is consistent with the preliminary findings presented in Figures 1 to 3, where the number of cells is directly associated with light scattering and its effects on the baseline. It can be deduced that the interpretation of this factor is dependent on the number, shape, and size of the cells. Nonetheless, regardless of the principal features in the PC1 being derived from the physical attributes of the cells, numerous small spectral features can also be observed. These spectral features can be attributed to the absorbance bands of H_2O rather than noise.

As aforementioned, the cells and cell medium are enriched with H_2O, which is a strong absorber of NIR light. Specifically, the broad bands observed at approximately 740 nm, 850 nm,

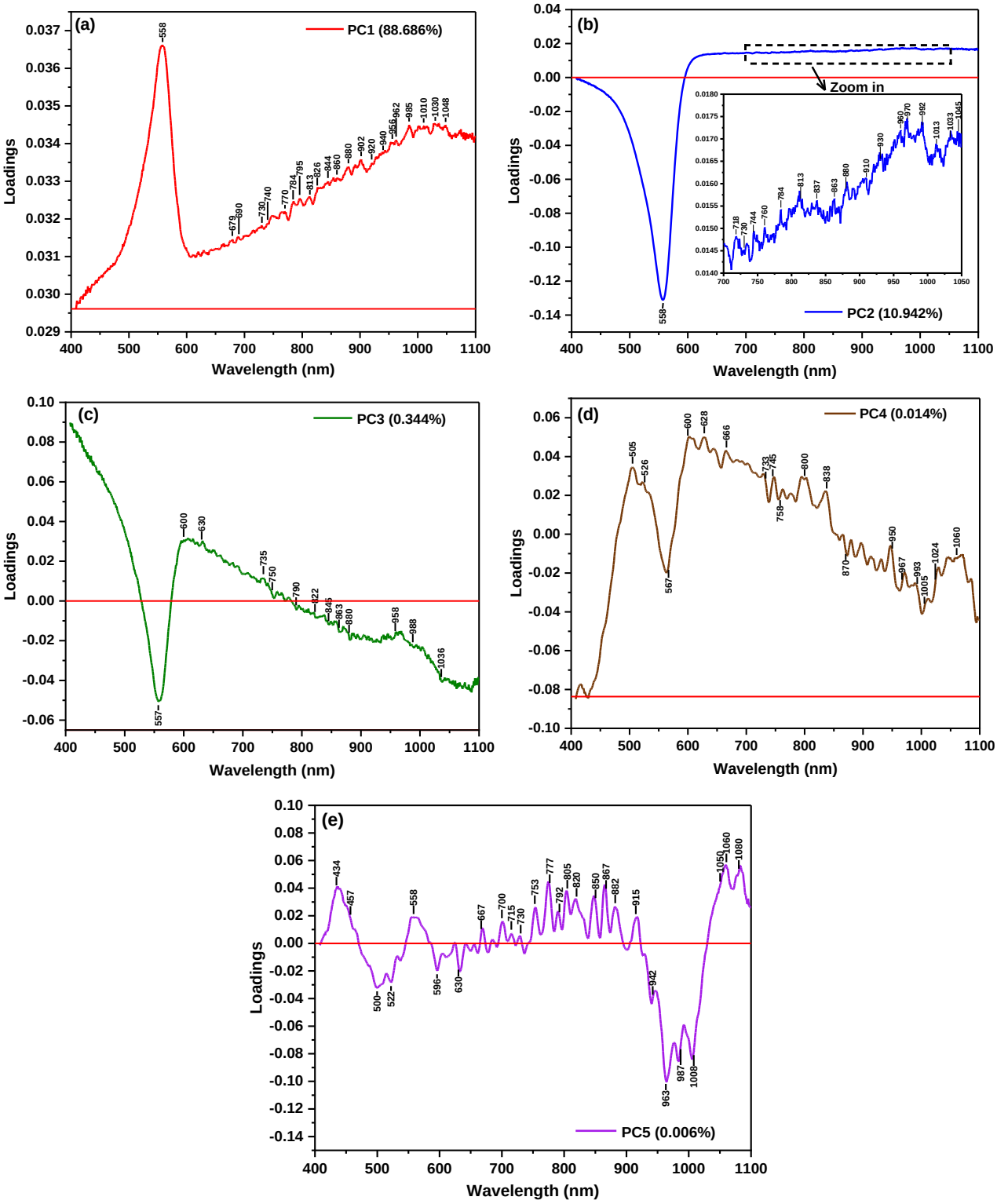


Figure 6. (a)-(e) PCA loadings plots of the calculated PC1 to PC5 at region 400-1100 nm for L929, DU145, and HeLa cell lines. Inset related to PC2 is zoom in of loadings in the NIR region (700 to 1050 nm).

and 970 nm can be assigned to the third overtone of OH stretching vibration $(3\nu_1+\nu_3)$, the second overtone of the combination of stretching and bending vibration $(2\nu_1+\nu_2+\nu_3)$, and the second overtone of OH stretching vibration $(2\nu_1+\nu_3)$ [88-91]. The slight variations in absorbances of the peaks (Figure 6) seem to be small compared to the peak intensity in the visible region, although this is simply as a result of the disparity in the source of the peaks, i.e., electronic and vibrational transitions. The emergence of numerous peaks confirms the complex nature of the molecular structure of H_2O in the cells and cell cultures which becomes altered in the course of cell proliferation. Likewise, PC2 is dominated by the peak denoting PhR at 558 nm, which is negative, while the NIR part shows only baseline offset resulting from light scattering. The features of this component can elucidate the characteristics of the three distinct groups observed at Figure 6(a). It is evident that the scores positioned in the positive part of PC2 axis are related to non-stained cell cultures (without PhR), which for all three cells lines, form a compact, well-defined cluster. The scores matching the PhR stained cell cultures are present in the negative part of the PC2 axis (as indicated by the negative peak at 558 nm in PC2). Remarkably, two subgroups can be observed within relatively large cluster, corresponding to the cell types, where the scores of the normal cell line are located in the lowest portion of the negative part of PC2 axis. An overview of the scores plots in Figure 5(a) and PC2 in Figure 6(b) shows that this component differentiated PhR stained and non-stained cultures. This differentiation is based on the peak denoting PhR in the visible region, while in the NIR part, the difference is indicated in both the baseline offset between PhR stained and non-stained groups (with higher offset for the non-stained cultures) and in the magnified part of Figure 6(b) in the several absorbance bands of water, particularly the area of the 2^{nd} overtone of water ($\sim$ 970 nm). It can be inferred that both water content and molecular structure of the cell cultures (media and cells included) are affected by staining. PhR staining of the cells

changes the molecular structure of water differently for the cancer cells and normal cells. Thus, it is imperative to further elucidate the relationship between pH and the molecular structure of water in cancer and normal cells. The PC4 and PC5 loadings are presented in Figure 6(d) and (e), respectively. The loadings show an increased variety of spectral features, with lots of distinct peaks, particularly the PC5 in the NIR region. The region of 800-910 nm is complex to interpret because of the several overlapping bands denoting diverse functional groups. A number of the bands observed in the loadings are recognized water absorbance bands and consistent with the ones aforementioned in the analysis.

PCA determined the natural clusters in the spectra based on the cell types and PhR staining. For additional examination of these differences, SIMCA discrimination analysis was performed to differentiate the cell lines according to PhR treatment and cell type. The raw spectra of the cell lines were pre-processed in order to improve the discrimination analysis using SIMCA. The pre-processing methods were Savitzky-Golay with 2^{nd} order polynomial filter and 21 data points to remove noise, standard normal variate (SNV) transformation to remove baseline effects and mean-centering. Figure 7(a) depicts the pre-processed spectra of the three cell lines, coloured based on the status of PhR treatment regardless of cell type. In this figure, the spectra represent the relation between wavelength and absorbance at the number of cells from 50,000 to 275,000 for each cell line cultured in the medium with and without PhR. As can be seen in Figure 7(a), there is a distinctive difference between the spectra of the cell lines with and without PhR at the absorbance band located at 556 nm (corresponding to the absorption of the basic form of PhR) and in the spectral region located around 970 nm (corresponding to 2^{nd} overtone of water) where several distinctive bands could be observed. In Figure 7(a), the band of H_2O has equivalent implications for the differentiation between the normal and cancerous cells in the presence and absence of PhR

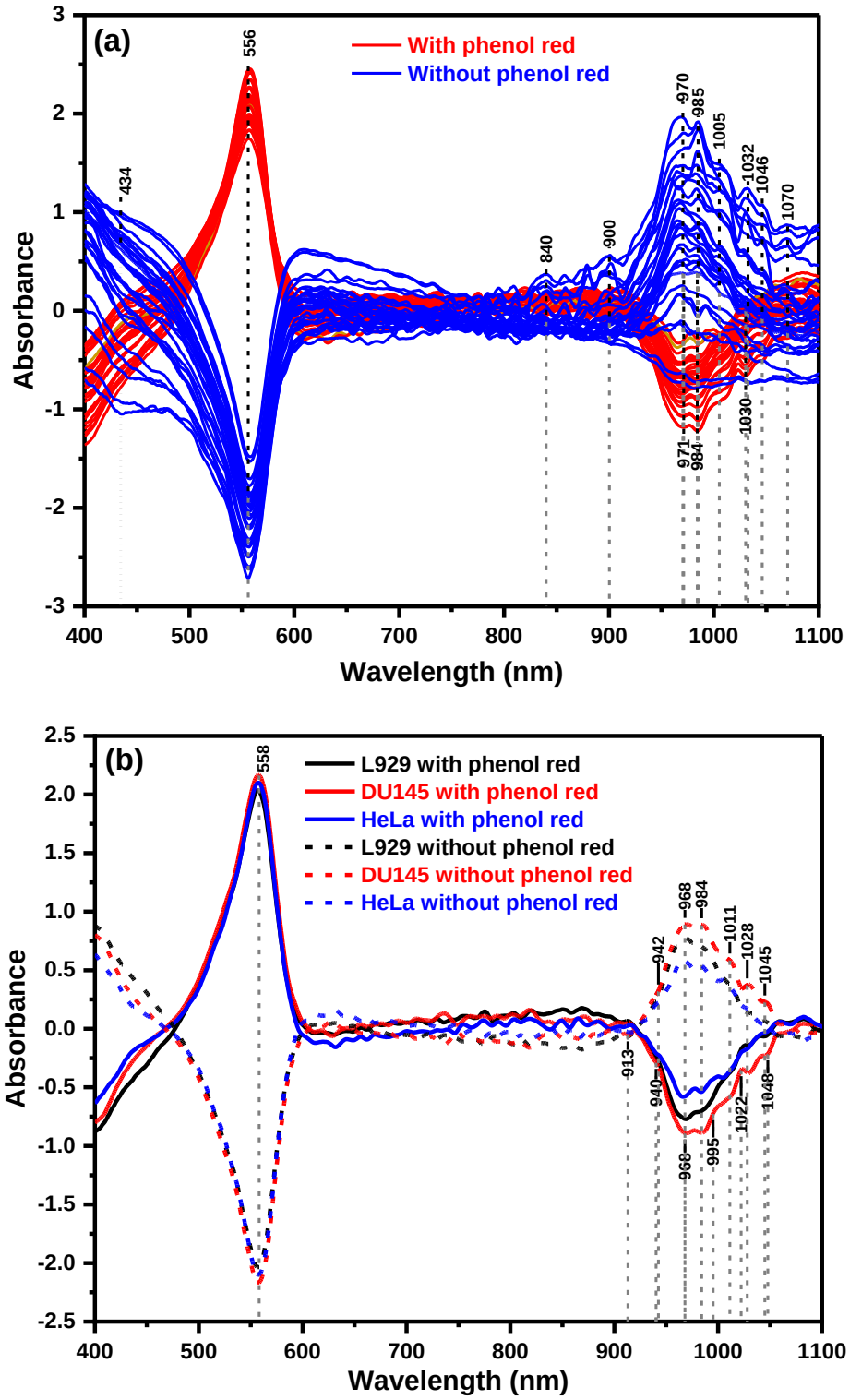


Figure 7. (a) Pre-processed spectra and (b) average of the pre-processed spectra for L929, DU145, and HeLa cell lines cultured in medium with and without phenol red at the number of cells ranging from 50,000 to 275,000.

compared to the peak of PhR at 556 nm. Figure 7(b) displays the averaged spectra for each cell line presented in Figure 7(a) in case of the presence and absence of PhR which means each cell line has two averaged spectra one of them in the presence of phenol red (solid line) and the other one in the absence of phenol red (dashed line). Evidently, there is a common absorbance band in the region of the 2nd overtone of water, irrespectively of the presence or absence of PhR, suggesting that the main spectral features characteristic for each cell line are inherent in this spectral region (NIR region) and do not depend on PhR. It is interesting that the variations of the spectra decrease with adding PhR, which is consistent with the fact that PhR is a perturber for the water molecular system which unifies the samples along its influence leaving only the differences depending on the type of the samples.

The PhR treatment appears to lead to inhibited dissimilarities between the spectra of individual samples and thus the cell types. The presence of more inter-class spectral variations is observable in the samples prepared in the absence of PhR, particularly in the region around 985 nm. The variations also exist in the visible region, around 434 nm and 558 nm, which can be credited to the acid (yellow) and basic (red) forms of PhR [86], respectively. Thus, it can be deduced that PhR influences the molecular structure of H_2O in the samples, either through interfacing with the medium, cells, or both. The averaging of the spectra shown in Figure 7(b) for each cell type, clearly indicates that the average differences in cell type cultures is more evident in the NIR region, and same absorbance bands are affected by the PhR regardless of the cell type, but with dissimilar spectral intensities. These bands are positioned at 940-942 nm, 968 nm, 984 nm, 995 nm, 1011 nm, 1022 nm, 1028 nm, and 1045-1048 nm (Table 5). The band positions, being in the second overtone of H_2O, when recalculated to the 1st overtone, generate the following bands: 1410-1413 nm, 1452 nm, 1476 nm, 1492.5 nm, 1516.5 nm, 1533 nm, 1542 nm, and 1567.5-1572

nm. The first five bands are assigned to the well-defined H_2O matrix coordinates of aquaphotomics. C5, C8, C9, C10, C11, and C12 are assigned to free molecules S_0, H_2O hydration shell, H_2O molecules with two hydrogen bonds, H_2O molecules with three hydrogen bonds S_3, H_2O molecules with four hydrogen bonds S_4, and strongly bound water [92], respectively.

Table 5 shows that the assignment of bands in the region of 995-1030 nm is a difficult procedure because of the complex features of the NIR spectra of both proteins and H_2O. Nonetheless, since the bands in this case are associated with the differences between PhR stained and non-stained cultures, it is rational to assume the assignment of these bands to H_2O involved with proton hydration (also referred to as aqueous protons, which are different H_2O clusters with dissimilar number of H_2O molecules related to hydration). The fact that diverse spectra also exhibit distinct bands of protonated and deprotonated forms of PhR in the Vis region confirms that bands discernible in NIR region explain the same phenomenon i.e., also act as an optical pH indicator of the changes in hydrogen ions (H^+) or hydronium ions (H_3O^+).

The results of SIMCA discriminating analysis (5% significance) performed on the pre-processed spectra in Figure 7(a) are presented in Figure 8. Cooman's plot shown in Figure 8(a) reveals the results of SIMCA discrimination of the spectra of L929, DU145, and HeLa with and without PhR staining. Ellipses drawn on this graph are for presentation purposes emphasizing the complete separation between different cell types in the multivariate space. The yellow ellipse includes the scores of the three cell lines with the different numbers of cells (from 50,000 to 275,000) and cultured in the medium including phenol red dye. This ellipse showed that the use of PhR during culture decreases the interclass distance. The achieved accuracy of classification with SIMCA was 100%, with a confidence interval of 0.95. Table 6 exhibits the number of components (PCs), explained variance (%), and the actual class membership and predicted

Table 5. Tentative assignments of the absorbance bands found to be important for discrimination between the cell cultures with and without phenol red in 2nd overtone of water, and the corresponding bands in the 1st overtone region, where first five of them belong to well-established aquaphotomics 12 Water Matrix Coordinates 'WAMACs' (C1-C12) [92], together with the alternative possible assignments.

Wavelength (nm)	Tentative band assignment	Ref.
940-942	Free water molecules, S_0 (1410-1413 nm, C5 WAMACs)	[92]
	Water vapour (940 nm)	[93-95]
	Water vapour (942 nm)	[96]
968	Hydration shell OH-$(H_2O)_{4,5}$ (1452 nm, C8 WAMACs)	[92]
	Hydroxide absorption peak (967 nm)	[97]
	O–H stretch, 2nd overtone of ROH and H_2O (970 nm)	[42]
	O–H stretch, 2nd overtone of H_2O (960–970 nm)	[89]
984	Water molecules with 3 hydrogen bonds, S_3 (1476 nm, C10 WAMACs)	[92]
	N–H stretch, 2nd overtone of CONHRn (secondary amides) (981 nm)	[89]
995	Water molecules with 4 hydrogen bonds (1492.5 nm, C11 WAMACs)	[92]
	N-H stretch, 2nd overtone (aromatic amines) (995 nm)	[89]
	N-H stretch, 2nd overtone (amines, general) (1000 nm)	[89]
1011	Strongly bound water (1516.5 nm, C12 WAMACs)	[92]
	OH from secondary (1004 nm) and tertiary alcohols (1006 nm)	[89]
	N-H stretch, 2nd overtone (amines, general) (1015 nm)	[89]
	2nd overtone Superoxide Tetrahydrate $O_{2-}.(H_2O)_4$ (1010 nm)	[98]
1022	2× N–H stretch + 2× amide I (protein) (1020 nm)	
	N–H stretch, 2nd overtone (protein) (1007 nm)	[89]
	N–H stretch, combination of RNH2 (primary amine) (1020 nm)	
	Aqueous proton $[H^+.(H_2O)_6]$ - H_2O in $H_5O_2^+$ symmetric stretch, 2nd overt (1018 nm)	[99]
	2nd overtone intramolecular hydrogen bond stretches in (OH-$(H_2O)_4$) (1020 nm)	[100]
1028	N-H stretch, 2nd overtone of RNH2 (1030 nm)	[89]
	Aqueous proton $[H^+\cdot(H_2O)_3]$ - H_3O^+ symmetric stretch, 3rd overt (1033 nm)	[99]
1045-48	N-H stretch, 2nd overtone of amides (primary, bonded; two bonds) (~1050 nm)	
	N-H stretch, combination of CONH (primary amide) (1047 nm)	[89]
	N-H stretch, 3rd overtone of amides (secondary, bonded; one band) (1000-1067 nm)	
	$H_{11}O_5^+$ H-bonded OH stretch, 2nd overt (1041 nm)	[101]
	Aqueous proton $[H^+\cdot(H_2O)_6]$ - H_2O in $H_5O_2^+$ symmetric stretch, 2nd overt (1055 nm)	[99]

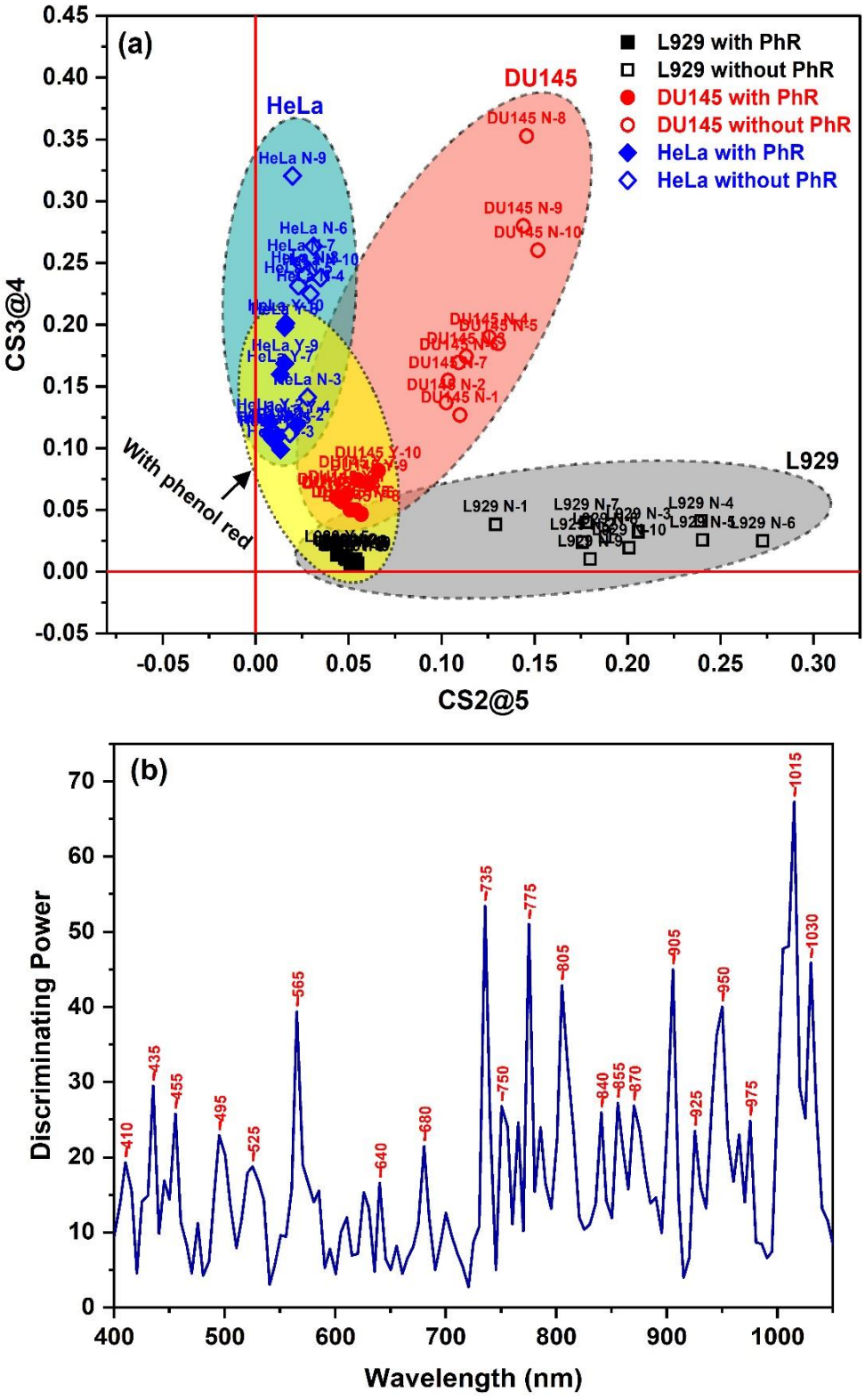


Figure 8. SIMCA classification analysis: (a) Cooman's plot and (b) discriminating power between L929, DU145, and HeLa cell lines.

Table 6. SIMCA classification results for L929, DU145, and HeLa cell lines.

PCs	Explained variance (%)			Actual Class			Correct classification (%)	Misclassifications (%)
				L929	DU145	HeLa		
5	99.93	SIMCA Class	L929	**20**	0	0	100	0
5	99.89		DU145	0	**20**	0	100	0
4	99.86		HeLa	0	0	**20**	100	0

membership using the SIMCA model which clearly reveals no misclassifications. Mahalanobis distances (interclass distance) of 4.07, 6.93, and 5.32 were obtained between HeLa and DU145, HeLa and L929, and DU145 and L929, respectively. The interclass distances were > 3.0, indicating that the separation is reliable. As observed in Cooman's plot, the PhR stained cells are positioned closer together (smaller Mahalanobis distance) than the non- stained cells, which further confirms the importance of preserving the cells in their initial state for effective discrimination. Figure 8(b) displays the discriminating power of the SIMCA model that indicates the absorbance bands with the highest influence for successful discrimination between cell types. The discrimination power shows that the absorbance bands with the highest contribution in the model are located in the NIR region, and they include 735 nm, 775 nm, 805 nm, 905 nm, 950 nm, 1015 nm, and 1030 nm. In the visible region, the bands at 435 nm and 565 nm are assigned to the highest power. The assignments of the absorbance bands are presented in Table 7. These bands may be attributed to hydrogen-bonded H_2O and H_2O interfacing with carbohydrates or proteins, or these compounds directly. Six of these bands can be related to hydrogen-bonded H_2O, which further confirms the significance of using the molecular structure of H_2O to distinguish between the three cell types. Nevertheless, as reported in prior research, the extracellular metabolites are main contributors to

Table 7. Tentative assignments of the absorbance bands found in discriminating power of SIMCA model developed for discrimination between different cell types. The assignments were provided for the bands found directly in 2nd and 3rd overtones of water, or recalculated to 1st overtone WAMACS bands, together with the alternative possible assignments.

Wavelength (nm)	Tentative band assignment	Ref.
735	Hydrogen-bonded -OH, 3rd overt	[102]
	Water molecules with 3 hydrogen bonds, S_3 (1470 nm, C10 WAMACS)	[92]
775	Hydrogen bonded water (770 nm)	[103]
	N–H stretch, 3rd overtone of RNH_2 (779 nm)	[42]
	N–H stretch, 3rd overtone of amides (secondary, bonded; one band) (750-800 nm)	[89]
805	Maximum absorption peak for ice (800 nm)	[96]
905	C-H stretch, 3rd overtone of protein (910 nm)	[104, 105]
	3rd overtone of a C-H stretching mode and the 2nd overtone of an N-H stretching mode of protein (906 nm)	[106]
	Hydration shell OH-$(H_2O)_{1,2,4}$ (1357.5 nm, C2 WAMACs)	[92]
950	O-H stretch (954 nm), 2nd overtone of ROH (alcohols) (940-970 nm)	[107]
	Water interacting with sugar (950 nm)	[108]
	Hydration water (1425 nm, C6 WAMACS)	[92]
	2nd overtone of O-H stretching of water interacting with protein (950-960 nm)	[106]
1015	H_2O deionized; 2nd overt (1010.1 nm)	[89]
	Hydrated proton $[H+(H_2O)_6]$ - H_2O symmetric stretch, 2nd overt (1018.1 nm)	[99]
	2X N-H stretch + 2X amide I (protein) (1020 nm)	
	N-H stretch, 2nd overtone (protein) (1007 nm)	[89]
	N-H stretch, combination of RNH2 (primary amine) (1020 nm)	
	Protein (1020 nm)	[109]
	Strongly bound water (1522.5 nm, C12 WAMACs)	[92]
1030	N-H stretch, 2nd overtone of RNH_2	[89]
	3rd overtone of a C-H stretching mode and the 2nd overtone of an N-H stretching mode of protein (1032 nm)	[106]

the features of the NIR spectral signal [110], which implies that the differentiation of cell types is not based on the physical attributes of cells (whose effect is drastically reduced by pre-processing)

but according to the metabolites which are specific for each cell type because of their metabolic variations. The metabolites in turn affect the hydrogen bond complex of extra- and intracellular water leading to the characteristic spectral signature in the NIR region. Employment of vibrational spectroscopy for a variety of cancer types has been reported marker bands related to water, proteins, lipids, amino acids, carbohydrates, and their combination that are useful for cancer diagnosis. The main absorbers of NIR light in blood-perfused tissues are oxyhemoglobin, deoxyhemoglobin, and water [111], while in the cell culture, the major absorbers are various water molecular species and other components are presented in considerably less quantities. On the other hand, there are numerous reports on the diagnosis of other diseases that showed the diagnostic value of water absorbance bands [25, 38, 112-121]. In some reports, there is a possibility that the discovered bands are also assigned as the water vibrational bands. For example, the bands 1471 nm and 1911 nm that are attributed to DNA of rat mammary gland tumor [23], whereas there is a very likely possibility these bands originate from water [92, 119].

In our findings, seven absorbance bands were found to be most influential for discrimination between the L929, DU145, and HeLa cell lines. Five of these bands can be attributed to water, specifically to hydrogen-bonded water with different numbers of hydrogen bonds or strength, and to the water interacting with either sugar or proteins. A study performed by Ali et al. [24] suggested water bands can be used for the diagnosis of human prostate cancer. It was reported that the 1st overtone of water located around 1450 nm, and other water overtone modes provided key spectroscopic fingerprints for cancer in prostate tissue, citing as the most probable reason the higher-order phases of water and their interaction with cellular or other macromolecules in prostate tissues. Gosh et al. [38] used FTIR spectroscopy of exfoliated cells that revealed five distinctive spectral features attributable to hydrogen-bonded water (3 peaks at 3500-3700, 3700-3800, and

3800-4000 cm^{-1}) and non-hydrogen bonded OH stretch of water (2 peaks at 3300-3400 and 3400-3500 cm^{-1}) that showed changes in intensity in patients with oral cancer compared to healthy volunteers. Their findings are in a good agreement with current work, where within 3rd and 2nd overtone of water, water absorbance bands attributed to different phases of water and water interacting with other macromolecules were found.

It can be deduced that the water molecular structure plays a significant role in differentiating between cancer and healthy cells which is revealed from the current study where the cells of cancers from different organs showed spectral differences. To the authors' knowledge, there is a study using NIRS that showed spectral changes between different cancers (colorectal and pancreatic cancer) [25]. On the other hand, there are studies on aquaphotomics showing an excellent possibility of differentiating the cell types in the case of bacteria [122-124], different types of animal tissues [125], and even normal and abnormal cell development [90]. This study reported for the first time the differences between normal fibroblast cells and two types of cancer – prostate and cervical cells, which demonstrated the discrimination between the cell types is mostly based on the water molecular structure in the NIR short region. This is a very important aspect of cancer research, that can help increase the specificity of cancer diagnosis by enabling differentiation among tumors [25]. Nonetheless, additional research is needed to precisely assign bands to functional groups and relate them to the cell type metabolism and accurately deduce the basis for cell differentiation. An appealing feature of the current findings is that the observed bands are crucial for the discrimination of cells, and relevant for a variety of diagnostic purposes in other biological systems, such as the level of tissues. This offers a basis for advanced studies that explore the application of aquaphotomics NIR spectroscopy to investigate early detection of cancer using cells, tissues, or body fluids.

4. Conclusion

This work presents the application of NIR absorbance spectroscopy and aquaphotomics to classify cultured cancerous cells including HeLa and DU145 cells. The results show that the combined application of absorption spectra and principal component analysis (PCA) improves the selectivity for the precise classification of cancer cells. The PCA achieved a clear differentiation between the three types of cell lines, with subgroups detected based on the presence or absence of the phenol red (PhR) in the culture. SIMCA method achieved a high accuracy of classification, where larger class distances were obtained between the non-stained cells compared to stained cells. Aquaphotomics enhanced the assignment of cancer cell proliferation from the perspective of water molecular structure. The absorbance bands that successfully aided the cancer cell discrimination include a range of H_2O molecular species-protons (vapor) and strongly bonded H_2O, as well as CH and NH bands of proteins. Future studies can further explore the roles of these bands in differentiating between normal and cancer cell types.

This study is a novel approach to the advancement of an online noninvasive spectroscopic system capable of monitoring ideal growth conditions and maximizing cellular productivity. In addition, the integrative spectroscopic biomarker results in a future design of specialized spectroscopic instrumentation for cancer diagnosis. Furthermore, this work unlocks possibilities for at-line sensor development for discriminating mammalian cells using PhR in a culture medium. It also offers certain insights into the influence of PhR on cell dynamics and expands the application of the NIR spectral region to distinguishing between cell types.

Data Availability

All data used to support the findings of this study are available from the corresponding author (A.F.O.) upon reasonable request.

Acknowledgment

The authors gratefully acknowledge the financial support of this research by the Malaysian Ministry of Higher Education–Fundamental Research Grant Scheme (Grant No. 203.PFIZIK.6711897).

References

1. Butler, L., et al., *Lipids and cancer: Emerging roles in pathogenesis, diagnosis and therapeutic intervention.* Advanced drug delivery reviews, 2020.

2. Sorvina, A., et al., *Lipid profiles of prostate cancer cells.* Oncotarget, 2018. **9**(85): p. 35541.

3. Frangioni, J.V., *New technologies for human cancer imaging.* Journal of clinical oncology, 2008. **26**(24): p. 4012.

4. Kondepati, V.R., H.M. Heise, and J. Backhaus, *Recent applications of near-infrared spectroscopy in cancer diagnosis and therapy.* Analytical and bioanalytical chemistry, 2008. **390**(1): p. 125-139.

5. El-Tawil, S.G., et al., *Comparative study between Pap smear cytology and FTIR spectroscopy: a new tool for screening for cervical cancer.* Pathology, 2008. **40**(6): p. 600-603.

6. Wilson, R.H., et al., *Tissue classification using optical spectroscopy accurately differentiates cancer and chronic pancreatitis.* Pancreas, 2017. **46**(2): p. 244.

7. Yang, P.-W., et al., *Visible-absorption spectroscopy as a biomarker to predict treatment response and prognosis of surgically resected esophageal cancer.* Scientific reports, 2016. **6**(1): p. 1-12.

8. Dai, W.Y., S. Lee, and Y.C. Hsu, *Discrimination between oral cancer and healthy cells based on the adenine signature detected by using Raman spectroscopy.* Journal of Raman Spectroscopy, 2018. **49**(2): p. 336-342.

9. Shang, L.-W., et al., *Fluorescence imaging and Raman spectroscopy applied for the accurate diagnosis of breast cancer with deep learning algorithms.* Biomedical Optics Express, 2020. **11**(7): p. 3673-3683.

10. Noothalapati, H., K. Iwasaki, and T. Yamamoto, *Non-invasive diagnosis of colorectal cancer by Raman spectroscopy: Recent developments in liquid biopsy and endoscopy approaches.* Spectrochimica Acta Part A: Molecular and Biomolecular Spectroscopy, 2021. **258**: p. 119818.

11. Del Mistro, G., et al., *Surface-enhanced Raman spectroscopy of urine for prostate cancer detection: a preliminary study.* Analytical and Bioanalytical Chemistry, 2015. **407**(12): p. 3271-3275.

12. Vendrell, M., et al., *Surface-enhanced Raman scattering in cancer detection and imaging.* Trends in biotechnology, 2013. **31**(4): p. 249-257.

13. Wang, G., et al., *Detection of the potential pancreatic cancer marker MUC4 in serum using surface-enhanced Raman scattering.* Analytical chemistry, 2011. **83**(7): p. 2554-2561.

14. Lee, S., et al., *Rapid and sensitive phenotypic marker detection on breast cancer cells using surface-enhanced Raman scattering (SERS) imaging.* Biosensors and Bioelectronics, 2014. **51**: p. 238-243.

15. Lin, D., et al., *Label-free blood plasma test based on surface-enhanced Raman scattering for tumor stages detection in nasopharyngeal cancer.* Scientific reports, 2014. **4**(1): p. 1-8.

16. Shin, H., et al., *Early-stage lung cancer diagnosis by deep learning-based spectroscopic analysis of circulating exosomes.* ACS nano, 2020. **14**(5): p. 5435-5444.

17. Ozaki, Y., et al., *Near-Infrared Spectroscopy: Theory, Spectral Analysis, Instrumentation, and Applications*. 2020: Springer Nature.

18. Jue, T. and K. Masuda, *Application of near infrared spectroscopy in biomedicine*. 2013: Springer.

19. Sakudo, A., *Near-infrared spectroscopy for medical applications: Current status and future perspectives.* Clinica Chimica Acta, 2016. **455**: p. 181-188.

20. Tsenkova, R., et al., *Near-infrared spectroscopy for biomonitoring: influence of somatic cell count on cow's milk composition analysis.* International dairy journal, 2001. **11**(10): p. 779-783.

21. Omar, A.F., H. Atan, and M.Z. MatJafri, *NIR spectroscopic properties of aqueous acids solutions.* Molecules, 2012. **17**(6): p. 7440-7450.

22. Balan, V., et al., *Vibrational spectroscopy fingerprinting in medicine: from molecular to clinical practice.* Materials, 2019. **12**(18): p. 2884.

23. Hirosawa, N., et al., *In vivo investigation of progressive alterations in rat mammary gland tumors by near-infrared spectroscopy.* Analytical biochemistry, 2002. **305**(2): p. 156-165.

24. Ali, J., et al., *Near infrared spectroscopy and imaging to probe differences in water content in normal and cancer human prostate tissues.* Technology in cancer research & treatment, 2004. **3**(5): p. 491-497.

25. Kondepati, V.R., et al., *CH-overtone regions as diagnostic markers for near-infrared spectroscopic diagnosis of primary cancers in human pancreas and colorectal tissue.* Analytical and bioanalytical chemistry, 2007. **387**(5): p. 1633-1641.

26. Ehlen, L., et al., *Synergy of fluorescence and near-infrared spectroscopy in detection of colorectal cancer.* Journal of Surgical Research, 2019. **242**: p. 349-356.

27. Redd, D.C., et al., *Raman spectroscopic characterization of human breast tissues: implications for breast cancer diagnosis.* Applied spectroscopy, 1993. **47**(6): p. 787-791.

28. Gazi, E., 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, 2003. **201**(1): p. 99-108.

29. Chan, J.W., et al., *Nondestructive identification of individual leukemia cells by laser trapping Raman spectroscopy.* Analytical chemistry, 2008. **80**(6): p. 2180-2187.

30. Harvey, T.J., et al., *Spectral discrimination of live prostate and bladder cancer cell lines using Raman optical tweezers.* Journal of biomedical optics, 2008. **13**(6): p. 064004.

31. Oshima, Y., et al., *Discrimination analysis of human lung cancer cells associated with histological type and malignancy using Raman spectroscopy.* Journal of biomedical optics, 2010. **15**(1): p. 017009.

32. Talari, A., et al., *Raman spectroscopic analysis differentiates between breast cancer cell lines.* Journal of Raman Spectroscopy, 2015. **46**(5): p. 421-427.

33. Chaturvedi, D., et al., *Different phases of breast cancer cells: Raman study of immortalized, transformed, and invasive cells.* Biosensors, 2016. **6**(4): p. 57.

34. Kyriakidou, M., et al., *FT-IR spectroscopy study in early diagnosis of skin cancer.* in vivo, 2017. **31**(6): p. 1131-1137.

35. Naurecka, M.L., et al., *FTIR-ATR and FT-Raman spectroscopy for biochemical changes in oral tissue.* American Journal of Analytical Chemistry, 2017. **8**(03).

36. Elmi, F., et al., *Application of FT-IR spectroscopy on breast cancer serum analysis.* Spectrochimica Acta Part A: Molecular and Biomolecular Spectroscopy, 2017. **187**: p. 87-91.

37. Zhang, K., et al., *Diagnosis of liver cancer based on tissue slice surface enhanced Raman spectroscopy and multivariate analysis.* Vibrational Spectroscopy, 2018. **98**: p. 82-87.

38. Ghosh, A., et al., *Chemometric analysis of integrated FTIR and Raman spectra obtained by non-invasive exfoliative cytology for the screening of oral cancer.* Analyst, 2019. **144**(4): p. 1309-1325.

39. Trunfio, N., et al., *Characterization of mammalian cell culture raw materials by combining spectroscopy and chemometrics.* Biotechnology progress, 2017. **33**(4): p. 1127-1138.

40. Bao, W., B. Yang, and B. Chen, *2-hydr_ensemble: lysine 2-hydroxyisobutyrylation identification with ensemble method.* Chemometrics and Intelligent Laboratory Systems, 2021. **215**: p. 104351.

41. Yang, B., W. Bao, and J. Wang, *Active disease-related compound identification based on capsule network.* Briefings in bioinformatics, 2022. **23**(1): p. bbab462.

42. Osborne, B.G., T. Fearn, and P.H. Hindle, *Practical NIR spectroscopy with applications in food and beverage analysis*. 1993: Longman scientific and technical.

43. Bhatia, H., et al., *In-line monitoring of amino acids in mammalian cell cultures using raman spectroscopy and multivariate chemometrics models.* Engineering in life sciences, 2018. **18**(1): p. 55-61.

44. Clavaud, M., et al., *Chemometrics and in-line near infrared spectroscopic monitoring of a biopharmaceutical Chinese hamster ovary cell culture: prediction of multiple cultivation variables.* Talanta, 2013. **111**: p. 28-38.

45. Sellick, C.A., et al., *Rapid monitoring of recombinant antibody production by mammalian cell cultures using Fourier transform infrared spectroscopy and chemometrics.* Biotechnology and bioengineering, 2010. **106**(3): p. 432-442.

46. Kozma, B., A. Salgó, and S. Gergely, *Comparison of multivariate data analysis techniques to improve glucose concentration prediction in mammalian cell cultivations by Raman spectroscopy.* Journal of pharmaceutical and biomedical analysis, 2018. **158**: p. 269-279.

47. Cui, X., et al., *Water as a probe for serum–based diagnosis by temperature–dependent near–infrared spectroscopy.* Talanta, 2019. **204**: p. 359-366.

48. Tsenkova, R., Z. Kovacs, and Y. Kubota, *Aquaphotomics: near infrared spectroscopy and water states in biological systems.* Membrane Hydration, 2015: p. 189-211.

49. Tsenkova, R., et al., *Essentials of aquaphotomics and its chemometrics approaches.* Frontiers in chemistry, 2018. **6**: p. 363.

50. Manley, M., *Near-infrared spectroscopy and hyperspectral imaging: non-destructive analysis of biological materials.* Chemical Society Reviews, 2014. **43**(24): p. 8200-8214.

51. De Beer, T., et al., *Near infrared and Raman spectroscopy for the in-process monitoring of pharmaceutical production processes.* International journal of pharmaceutics, 2011. **417**(1-2): p. 32-47.

52. Sandell, L. and D. Sakai, *Mammalian cell culture.* Current Protocols Essential Laboratory Techniques, 2011. **5**(1): p. 4.3. 1-4.3. 32.

53. Leme, J., et al., *A multivariate calibration procedure for UV/VIS spectrometric monitoring of BHK-21 cell metabolism and growth.* Biotechnology progress, 2014. **30**(1): p. 241-248.

54. Lai, C.Z., et al., *Spectrophotometric measurement of freshwater pH with purified meta-cresol purple and phenol red.* Limnology and Oceanography: Methods, 2016. **14**(12): p. 864-873.

55. Poole, C.B., et al., *Colorimetric tests for diagnosis of filarial infection and vector surveillance using non-instrumented nucleic acid loop-mediated isothermal amplification (NINA-LAMP).* PloS one, 2017. **12**(2): p. e0169011.

56. Pick, E. and Y. Keisari, *A simple colorimetric method for the measurement of hydrogen peroxide produced by cells in culture.* Journal of immunological methods, 1980. **38**(1-2): p. 161-170.

57. Mills, A. and G.A. Skinner, *A novel 'fizziness' indicator.* Analyst, 2011. **136**(5): p. 894-896.

58. Iseki, K., et al., *Helicobacter pylori infection in patients with early gastric cancer by the endoscopic phenol red test.* Gut, 1998. **42**(1): p. 20-23.

59. Takahashi, M.B., et al., *Artificial neural network associated to UV/Vis spectroscopy for monitoring bioreactions in biopharmaceutical processes.* Bioprocess and biosystems engineering, 2015. **38**(6): p. 1045-1054.

60. Baylor, S. and S. Hollingworth, *Absorbance signals from resting frog skeletal muscle fibers injected with the pH indicator dye, phenol red.* The Journal of general physiology, 1990. **96**(3): p. 449-471.

61. Bolsover, S., J. Brown, and T. Goldsmith, *Intracellular pH of Limulus ventral photoreceptor cells: measurement with phenol red.* Society of General Physiologists Series, 1986. **40**: p. 285-310.

62. Nishida, K., et al., *Effect of albumin on the absorption of phenol red, bromphenol blue and bromosulphonphthalein as model drugs from the liver surface membrane in rats.* Biological and Pharmaceutical Bulletin, 1995. **18**(11): p. 1548-1550.

63. Sochacka, J., *Application of phenol red as a marker ligand for bilirubin binding site at subdomain IIA on human serum albumin.* Journal of Photochemistry and Photobiology B: Biology, 2015. **151**: p. 89-99.

64. Abd Ghani, K., et al., *VIS–NIR spectral signature and quantitative analysis of HeLa and DU145 cell line.* Spectrochimica Acta Part A: Molecular and Biomolecular Spectroscopy, 2019. **222**: p. 117241.

65. Amran, E.N., et al., *Potential colorimetric detection of cancer cells using Phenol Red.* Photodiagnosis and photodynamic therapy, 2019. **27**: p. 380-384.

66. Arnold, S.A., et al., *In-situ near infrared spectroscopy to monitor key analytes in mammalian cell cultivation.* Biotechnology and bioengineering, 2003. **84**(1): p. 13-19.

67. Wold, S., K. Esbensen, and P. Geladi, *Principal component analysis.* Chemometrics and intelligent laboratory systems, 1987. **2**(1-3): p. 37-52.

68. Wold, S. and M. Sjöström, *SIMCA: a method for analyzing chemical data in terms of similarity and analogy*. 1977, ACS Publications.

69. Tsenkova, R.N., et al., *Prion protein fate governed by metal binding.* Biochemical and biophysical research communications, 2004. **325**(3): p. 1005-1012.

70. Savitzky, A. and M.J. Golay, *Smoothing and differentiation of data by simplified least squares procedures.* Analytical chemistry, 1964. **36**(8): p. 1627-1639.

71. Barnes, R., M.S. Dhanoa, and S.J. Lister, *Standard normal variate transformation and de-trending of near-infrared diffuse reflectance spectra.* Applied spectroscopy, 1989. **43**(5): p. 772-777.

72. Seasholtz, M.B. and B.R. Kowalski, *The effect of mean centering on prediction in multivariate calibration.* Journal of Chemometrics, 1992. **6**(2): p. 103-111.

73. Magnusson, E.B., et al., *Real-time optical pH measurement in a standard microfluidic cell culture system.* Biomedical optics express, 2013. **4**(9): p. 1749-1758.

74. Hanahan, D. and R.A. Weinberg, *Hallmarks of cancer: the next generation.* cell, 2011. **144**(5): p. 646-674.

75. Gatenby, R.A. and R.J. Gillies, *Why do cancers have high aerobic glycolysis?* Nature reviews cancer, 2004. **4**(11): p. 891-899.

76. Granja, S., et al. *Value of pH regulators in the diagnosis, prognosis and treatment of cancer*. in *Seminars in cancer biology*. 2017. Elsevier.

77. W Huck, C., Y. Ozaki, and V. A Huck-Pezzei, *Critical review upon the role and potential of fluorescence and near-infrared imaging and absorption spectroscopy in cancer related cells, serum, saliva, urine and tissue analysis.* Current Medicinal Chemistry, 2016. **23**(27): p. 3052-3077.

78. Liu, K.-Z., et al., *Quantitative determination of serum LDL cholesterol by near-infrared spectroscopy.* Vibrational spectroscopy, 2005. **38**(1-2): p. 203-208.

79. Baba, A.I. and C. Câtoi, *Tumor cell morphology*, in *Comparative Oncology*. 2007, The Publishing House of the Romanian Academy.

80. Lin, X., et al., *Light scattering from normal and cervical cancer cells.* Applied optics, 2017. **56**(12): p. 3608-3614.

81. Stevenson, K., et al., *General calibration of microbial growth in microplate readers.* Scientific reports, 2016. **6**(1): p. 1-7.

82. Chung, S., et al., *Non-invasive tissue temperature measurements based on quantitative diffuse optical spectroscopy (DOS) of water.* Physics in Medicine & Biology, 2010. **55**(13): p. 3753.

83. Chung, S., et al., *In vivo water state measurements in breast cancer using broadband diffuse optical spectroscopy.* Physics in Medicine & Biology, 2008. **53**(23): p. 6713.

84. Brubach, J.-B., et al., *Signatures of the hydrogen bonding in the infrared bands of water.* The Journal of chemical physics, 2005. **122**(18): p. 184509.

85. Afrin, T., et al., *Water structure modification by sugars and its consequence on micellization behavior of cetyltrimethylammonium bromide in aqueous solution.* Journal of Solution Chemistry, 2013. **42**(7): p. 1488-1499.

86. Guo, Y., et al., *A pH-responsive colorimetric strategy for DNA detection by acetylcholinesterase catalyzed hydrolysis and cascade amplification.* Biosensors and Bioelectronics, 2017. **94**: p. 651-656.

87. Yang, Y., et al., *A pH-responsive bioassay for paper-based diagnosis of exosomes via mussel-inspired surface chemistry.* Talanta, 2019. **192**: p. 325-330.

88. Bonner, O.D. and G.B. Woolsey, *Effect of solutes and temperature on the structure of water.* The Journal of Physical Chemistry, 1968. **72**(3): p. 899-905.

89. Workman, J. and J. Workman, *Handbook of Organic Compounds: Methods and Interpretations*. Vol. 1. 2001: Academic Press.

90. Kovacs, Z., et al., *Water Spectral Patterns Reveals Similarities and Differences in Rice Germination and Induced Degenerated Callus Development.* Plants, 2021. **10**(9): p. 1832.

91. Okubo, N. and Y. Kurata, *Nondestructive classification analysis of green coffee beans by using near-infrared spectroscopy.* Foods, 2019. **8**(2): p. 82.

92. Tsenkova, R., *Aquaphotomics: dynamic spectroscopy of aqueous and biological systems describes peculiarities of water.* Journal of Near Infrared Spectroscopy, 2009. **17**(6): p. 303-313.

93. Cattaneo, H., T. Laurila, and R. Hernberg, *VCSEL based detection of water vapor near 940 nm.* Spectrochimica Acta Part A: Molecular and Biomolecular Spectroscopy, 2004. **60**(14): p. 3269-3275.

94. Cattaneo, H. and R. Hernberg, *Detection of high-temperature water vapor at 940 nm with vertical-cavity surface-emitting lasers.* Applied optics, 2005. **44**(31): p. 6593-6598.

95. Sierk, B., et al., *Field measurements of water vapor continuum absorption in the visible and near-infrared.* Journal of Geophysical Research: Atmospheres, 2004. **109**(D8).

96. Workman, J. and L. Weyer, *Practical Guide and Spectral Atlas for Interpretive Near-Infrared*. 2012: CRC.

97. Heiman, A. and S. Licht, *Fundamental baseline variations in aqueous near-infrared analysis.* Analytica chimica acta, 1999. **394**(2-3): p. 135-147.

98. Weber, J.M., et al., *Isolating the spectroscopic signature of a hydration shell with the use of clusters: Superoxide tetrahydrate.* Science, 2000. **287**(5462): p. 2461-2463.

99. Headrick, J.M., et al., *Spectral signatures of hydrated proton vibrations in water clusters.* Science, 2005. **308**(5729): p. 1765-1769.

100. Xantheas, S.S., *Ab initio studies of cyclic water clusters (H2O) n, n= 1–6. III. Comparison of density functional with MP2 results.* The Journal of chemical physics, 1995. **102**(11): p. 4505-4517.

101. Mizuse, K. and A. Fujii, *Tuning of the internal energy and isomer distribution in small protonated water clusters H+ (H2O) 4–8: An application of the inert gas messenger technique.* The Journal of Physical Chemistry A, 2012. **116**(20): p. 4868-4877.

102. Davis, J.G., et al., *Water structural transformation at molecular hydrophobic interfaces.* Nature, 2012. **491**(7425): p. 582-585.

103. Golic, M., K. Walsh, and P. Lawson, *Short-wavelength near-infrared spectra of sucrose, glucose, and fructose with respect to sugar concentration and temperature.* Applied spectroscopy, 2003. **57**(2): p. 139-145.

104. Osborne, B. and T. Fearn, *Near Infrared Spectroscopy in Food Analysis; Longman Scientific & Technical: Essex*. 1986, UK.

105. Ciurczak, E.W. and B. Igne, *Pharmaceutical and medical applications of near-infrared spectroscopy*. 2014: CRC Press.

106. Šašić, S. and Y. Ozaki, *Short-wave near-infrared spectroscopy of biological fluids. 1. Quantitative analysis of fat, protein, and lactose in raw milk by partial least-squares regression and band assignment.* Analytical chemistry, 2001. **73**(1): p. 64-71.

107. Workman, J., *Handbook of organic compounds: NIR, IR, Raman and UV-Vis spectra featuring polymers and surfactants (a 3-volume set). 3. IR and Raman spectra*. 2001: Academic Press.

108. Kawano, S., H. Watanabe, and M. Iwamoto, *Determination of sugar content in intact peaches by near infrared spectroscopy with fiber optics in interactance mode.* Journal of the Japanese Society for Horticultural Science, 1992. **61**(2): p. 445-451.

109. Woo, Y.-A., et al., *Development of a new measurement unit (MilkSpec-1) for rapid determination of fat, lactose, and protein in raw milk using near-infrared transmittance spectroscopy.* Applied spectroscopy, 2002. **56**(5): p. 599-604.

110. Nakakimura, Y., et al., *Extracellular metabolites play a dominant role in near-infrared spectroscopic quantification of bacteria at food-safety level concentrations.* Analytical Methods, 2012. **4**(5): p. 1389-1394.

111. Fantini, S. and M.A. Franceschini, *Frequency-domain techniques for tissue spectroscopy and imaging.* Handbook of optical biomedical diagnostics, 2002. **1**.

112. Cerussi, A.E., et al., *In vivo absorption, scattering, and physiologic properties of 58 malignant breast tumors determined by broadband diffuse optical spectroscopy.* Journal of biomedical optics, 2006. **11**(4): p. 044005.

113. Beć, K.B., J. Grabska, and C.W. Huck, *Near-infrared spectroscopy in bio-applications.* Molecules, 2020. **25**(12): p. 2948.

114. Heffer, E.L., et al., *Near-infrared imaging of the human breast: complementing hemoglobin concentration maps with oxygenation images.* Journal of Biomedical Optics, 2004. **9**(6): p. 1152-1160.

115. Hollis, V.S., T. Binzoni, and D.T. Delpy. *Noninvasive monitoring of brain tissue temperature by near-infrared spectroscopy*. in *Optical Tomography and Spectroscopy of Tissue IV*. 2001. SPIE.

116. Hornung, R., et al., *Quantitative near-infrared spectroscopy of cervical dysplasia in vivo.* Human Reproduction, 1999. **14**(11): p. 2908-2916.

117. Li, Y., et al., *Early Diagnosis of Type 2 Diabetes Based on Near-Infrared Spectroscopy Combined With Machine Learning and Aquaphotomics.* Frontiers in Chemistry, 2020: p. 1133.

118. McIntosh, L.M., et al., *Towards non-invasive screening of skin lesions by near-infrared spectroscopy.* Journal of Investigative Dermatology, 2001. **116**(1): p. 175-181.

119. Muncan, J. and R. Tsenkova, *Aquaphotomics—From innovative knowledge to integrative platform in science and technology.* Molecules, 2019. **24**(15): p. 2742.

120. Simick, M.K., et al., *Non-ionizing near-infrared radiation transillumination spectroscopy for breast tissue density and assessment of breast cancer risk.* Journal of biomedical optics, 2004. **9**(4): p. 794-803.

121. Tromberg, B.J., et al., *Imaging in breast cancer: diffuse optics in breast cancer: detecting tumors in pre-menopausal women and monitoring neoadjuvant chemotherapy.* Breast Cancer Research, 2005. **7**(6): p. 1-7.

122. Kovacs, Z., et al., *Rapid bacteria selection using Aquaphotomics and near infrared spectroscopy.* 2019.

123. Slavchev, A., et al., *Monitoring of water spectral patterns of lactobacilli development as a tool for rapid selection of probiotic candidates.* Journal of Near Infrared Spectroscopy, 2017. **25**(6): p. 423-431.

124. Slavchev, A., et al., *Monitoring of water spectral pattern reveals differences in probiotics growth when used for rapid bacteria selection.* PLoS One, 2015. **10**(7): p. e0130698.

125. Sakudo, A., et al., *Ex vivo tissue discrimination by visible and near-infrared spectra with chemometrics.* Journal of veterinary medical science, 2006. **68**(12): p. 1375-1378.

We would like to express our sincere gratitude to the editor of Chemometrics and Intelligent Laboratory Systems Journal for allowing us to improve our manuscript for potential publication in the journal and to the reviewers who have critically evaluated our work and given constructive comments to enhance the quality of the manuscript. We have tried our best to address all the technical issues raised by the reviewers and we are ready to make any further modifications to the manuscript if it is required. All the corrections are written in red font in the revised manuscript. Below is a point-by-point response to the comments:

Reviewer #1:

Comment 1: Authors should give a detailed description of figure 5, figure 7 and figure 8.

Response: More detailed descriptions for Figures 5, 7, and 8 are provided in the revised manuscript as highlighted on PAGEs 21, 23, 26, 28, 29, and 32.

Comment 2: It was suggested that authors should share the data of the data and the code of this work.???

Response: We have added the data availability statement on PAGE 36 in the revised manuscript. It is stated that all the data used in this work is available upon request from the corresponding author Ahmad Fairuz Omar. Regarding the codes, we do not have any codes to share because the analysis was performed using commercially available software for multivariate analysis. However, as we mentioned in the statement on PAGE 36, all raw and derived data is available upon request.

Comment 3: Some references, including 10.1093/bib/bbab462, and 10.1093/bib/bbab462, should be discussed in this work.

Response: The suggested two references by the reviewer are added in the Introduction Section and highlighted in the revised manuscript on PAGE 6.

Comment 4: The language of this work should be carefully polished. Some grammar and spelling errors should be avoided in this manuscript.

Response: The language in the revised manuscript is now carefully polished, and grammar and spelling errors are corrected.

Reviewer #2:

Comment 1: Introduction part needs to be improved. Existing methods should be discussed more.

Response: The Introduction Section is now improved and a discussion on existing methods is added in the revised manuscript as highlighted on PAGEs 2 and 3.

Comment 2: For classification analysis PCA is very old model. There are a number of classification models which gives very good accuracy than PCA. Why those models are not considered?

Response: The principal component analysis was not used in this work for classification purposes. It was used as an unsupervised for exploratory analysis in order to observe and understand the variations in the spectra as a first step before any classification algorithms were tested. Using exploratory step of the analysis, we were able to observe the patterns in the terms of separation of scores among different cell types and cells cultured with and without phenol red staining. We also observed a strong trend depending on the number of cells.

In this work, we focused more on exploratory and qualitative study with the objective of understanding better which spectral region should be used in our further studies, and if we should use phenol red or not. Principal components revealed particular regions of the spectrum that represent important physical and chemical properties or entities within the cells that contribute to these observations. PCA also provided us with information on whether we need to apply pre-processing or not. Despite PCA being old chemometrics and multivariate data analysis method, for all these purposes that we mentioned, we found PCA to be the ideal choice in our study case. The help of PCA with visualizing high-order multivariate data was in our opinion of immense informational value for this work.

To conclude, PCA was not used for classification, just as an exploratory method of analysis to help us visualize the data, decide on pre-processing of data, and understand the causes of major variations in the spectra, all of which were achieved.

Comment 3: Today Deep learning models produce high accuracy?
Why deep learning or machine learning models not considered?

Response: This is a good suggestion, and we do consider using deep learning models, however not for this study. This work was about making the most important decisions on how we want to proceed, mainly whether we should use the entire Vis-NIR region and if phenol red should be used or not. For this purpose, we acquired a certain number of spectra, which we considered large enough to provide us with this information but insufficient to be used for any deep learning model reliably. The number of variables in this study is 1044 and the number of spectra (before averaging the triplicates) is 180. According to the commentary and guidelines by Foster et al. "the rule of thumb" is to have more than 10 "events" per each attribute to result in a classifier with reasonable predictive value [1]. The "black box" character of the machine learning methods makes them prone to false discovery and finding spurious associations, resulting in an impressive performance but useless when applied to new data [2]. So, it was very important to us to understand our data first and make correct choice about the future sample preparation when we want to obtain large number of spectra. Choice of wavelength region and most influential variables (absorbance bands related to the chemical information that we were able to relate to the features of normal and cancer cells) we now consider to be good candidates for attributes in deep learning models.

We do plan to use machine learning/deep learning models in a future research because based on this study we concluded that we can focus on the NIR region, we even have some absorbance bands (7 bands provided in Table 7) identified as of interest that can be used as attributes without need to use phenol red. Due to resource constraints, the current study was based on a small sample dataset, but we think we can now proceed to obtain additional samples for further research. Although there are strategies for the data augmentation approach that may help in the case of small datasets [3], we consider the best direction is to obtain more real samples and follow the conclusions that are currently provided.

Comment 4: Results should be compared with more latest models.

Response: As we explained in one of the previous responses, PCA was used as an exploratory data analysis tool and the PCA-based method Soft Modelling of Class Analogies (SIMCA) for classification. Since one more reviewer also asked to provide information about the accuracy of classification, we revised the previous text in the original manuscript on PAGEs 29 and 32 to describe the results of classification.

The achieved accuracy of classification with SIMCA was 100%, with a confidence interval of 0.95. We also provided in the revised manuscript Table 6 to show the actual class membership and predicted membership using the SIMCA model which clearly shows there were no misclassifications.

We explained earlier that our study at this stage was exploratory and qualitative to mainly provide the basis for decisions about future research direction. Especially due to the constraints of the size of our datasets we considered the comparative study performance with different classifiers would be weak at this stage. We already performed similar work with the comparison of different regression models for quantification of the number of cells of the three cell lines, which is currently under submission.

Reviewer #3:

Comment 1: Introduction part of the study is insufficient. Comparison with other studies regarding the characteristics of cancer cells in fingerprinting region is missing. Several recent studies are available where cancer cells are well defined under the said range of wavelength. Authors may include and discuss the following references:

Alsamad Fatima et al., Analyst, 2020, 145, 2945, DOI: 10.1039/C9AN02155H

Ghosh et al., Chemometrics and Intelligent Laboratory Systems, Volume 224, 2022, 104548.

Noothalapati et al., Spectrochimica Acta Part A: Molecular and Biomolecular Spectroscopy, Volume 258, 5 September 2021, 119818

Zhang et al., Vibrational Spectroscopy, Volume 98, September 2018, Pages 82-87.

Ghosh, et al., Analyst, Volume 144, Issue 4, 2019, 1309-1325

Wang, S. et al.,. Sci Rep 11, 23842 (2021). https://doi.org/10.1038/s41598-021-02687-0

Response: The introduction Section is now improved and studies regarding the characteristics of cancer cells in fingerprinting regions are included in the revised manuscript as highlighted on PAGEs 4-6. The relevant references suggested by the reviewer are cited in the revised manuscript with additions of few other works we found them important.

Comment 2: The novelty of present study in comparison to the available literature is not certain. A paragraph including the novelty may helpful. It is also not clear about the rationale of investigating cancer cells with and without phenol red indicator. As authors have already mentioned that the microscopy study of cell clearly indicate the morphological alterations, what further information phenol red is providing?

Response:

- A paragraph is added to the Introduction Section in the revised manuscript as highlighted on PAGE 8 to show the novelty of the current study.
- The rationale for investigating cancer cells cultured with and without phenol red is the following:

  As we showed in this paper, our results indicate that culturing different types of cells in medium with phenol red, leads to diminished interclass distance, and decreased the potential for discrimination between the cell types. On the other hand, this has a benefit of easier quantification by creating general regression model that can be used for quantification of cells despite of their type. The quantification results are the topic of another paper that currently under submission. In addition, in specific cases, some studies preferred culturing the cells in a non-phenol red medium and others with the presence of phenol red. The aim of investigating the cells cultured in a medium with and without phenol red is to show the implication of the phenol red in the culture medium on cells in the NIR region. We investigated the NIR absorbance spectra of the cells which are less affected or predisposed to the colour of the medium due to phenol red (in the Vis region). Further, we showed that both water content and molecular structure of the cell culture (media and cells included) are affected by staining. The averaging of the spectra shown in Figure 7(b) for each cell type, clearly indicates that the average differences in cell type cultures are more evident in the NIR region, and the same absorbance bands are affected by the phenol red regardless of the cell type, but with dissimilar spectral intensities. What is more, seven absorbance bands provided in Table 7 revealed that these bands are identified as of interest that can be used as attributes in deep learning machine models and these bands were obtained irrespective of the presence of phenol red in the cell culture medium.

  In summary, owing to our findings, for the problems of discrimination between cells it is better to not use phenol red, while if the problem is

quantification, it is better to use phenol red because it allows creation of one general quantitative model that can be used regardless of cell type.

Comment 3: Methodology part need a through revision:

a) Spectral acquisition part is not clearly mentioned - Laser wavelength, power, spot size and duration of exposure etc.
b) Cells are investigated in which form - as cellular smear or in flask/petridish?
c) Is any pre-processing of spectra performed for removing spectral inconsistency?
d) Why selecting soft independent modelling of class analogies (SIMCA) and not any deep learning based method?

Response:

a) The spectral acquisition part is now modified to include the information requested by the reviewer. The additional information is highlighted in the revised manuscript on PAGEs 12-13.

b) The three cell lines were grown in T25 tissue culture flasks and kept in an incubator at 37 ℃ in a humidified atmosphere with 5% CO_2. The cells were investigated in a 6-well cell culture plate as mentioned in the Spectral Acquisition and Results & Discussion Sections. To record the spectra, we used three wells of the 6-well cell culture plate at each number of the cells for each cell line.

c) Each spectrum in Figures 1-3 was smoothed using the Savitzky-Golay algorithm with 2^{nd} order polynomial and a window length of 15 points using the OriginPro 2021 Software to reduce the noise in the raw data as highlighted in the Results & Discussion Section in the revised manuscript on PAGE 14 and 15. In multivariate analysis, the PCA analysis was performed without any pre-processing methods. On the other hand, the spectra were pre-processed before performing SIMCA as mentioned in the Data Analysis Section as highlighted on PAGEs 14.

d) This comment is also raised by Reviewer 2 in Comment 3 as “why deep learning/machine learning models not considered”. So, to answer this

question, we would first like to point out that the current research was exploratory in nature, aiming to provide us with answers to which wavelength region (visible or near-infrared) will offer more information about the types and number of cells and in particular does staining with phenol red is necessary or not. To answer these questions, we first performed PCA which is one of the most used exploratory chemometrics methods to observe hidden patterns in the data and understand the chemical information from the data. The PCA method provided the answers that it is possible to see the trend in the increase of the number of cells, the differences between cell types, and also differences between the stained and non-stained cells. SIMCA, as it is a supervised classification method and based on the PCA, where separate PCA models are created for each of the class (cell type) was a natural, most logical choice as we already realized the separation between cell types in PCA. With SIMCA, we had 100% accuracy in the classification and the newly Table 6 in the revised manuscript on PAGE 32 provides information about the models for each class (number of components and explained variance in %), SIMCA-predicted class membership, and % of correct and incorrect classification.

We would also like to point out that multivariate analysis methods such as PCA, SIMCA, PLS-DA, and PCA-LDA are still very much in use for classification between cancer and healthy cells/tissues and provide a deeper understanding of the differences at the level of biomolecules. Some of the examples that the Reviewer pointed out from the recent literature include such studies [4, 5] among many others [6].

Comment 4: Sensitivity, specificity, accuracy of the studies need to include.

Response: The accuracy is now included in the paper, but since SIMCA results are about the classification between three cell types, and we did not want to focus here on diagnostics, i.e., to discriminate between healthy cells and cancer cells, the sensitivity and specificity are not calculated since they require the test to have only two classes (healthy and cancer). Technically, it can be said, based on classification accuracy of 100% as presented in Table 6 in the revised manuscript on PAGE 32, the sensitivity and specificity are also in this case 100%, but we did not want to make such claims, since there was only one type of non-cancerous cells and two

types of cancer cells from different tissues. Therefore, it would not be appropriate to assign general classes healthy and cancer and present our results as if they are an outcome of a diagnostic test.

Comment 5: A through revision of language and grammatical error is required.

Response: A revision of language is made, and grammatical errors are corrected in the revised manuscript.

Comment 6: Comparison of functional groups between cancer and normal fibroblast cells is not reflected in discussion part.

Response: Several studies showed the functional groups that reflected the difference between normal and cancerous tissues/cells together with characteristics of cancer cells in fingerprinting region (Comment 1) have been added in the Introduction Section as highlighted in the revised manuscript on PAGEs 4-6. Furthermore, we have provided a comparison of functional groups focusing on water bands in the Results & Discussion Section as highlighted in the revised manuscript on PAGEs 34-35. This comparison and discussion with studies supporting our findings can explain the importance of our findings for the increased specificity of cancer diagnosis by enabling differentiation among tumours using the water bands.

References

1. Foster, K.R., R. Koprowski, and J.D. Skufca, *Machine learning, medical diagnosis, and biomedical engineering research-commentary.* Biomedical engineering online, 2014. **13**(1): p. 1-9.

2. Broadhurst, D.I. and D.B. Kell, *Statistical strategies for avoiding false discoveries in metabolomics and related experiments.* Metabolomics, 2006. **2**(4): p. 171-196.

3. Wu, M., et al., *Deep learning data augmentation for Raman spectroscopy cancer tissue classification.* Scientific reports, 2021. **11**(1): p. 1-13.

4. Zhang, K., et al., *Diagnosis of liver cancer based on tissue slice surface enhanced Raman spectroscopy and multivariate analysis.* Vibrational Spectroscopy, 2018. **98**: p. 82-87.

5. Ghosh, A., et al., *Chemometric analysis of integrated FTIR and Raman spectra obtained by non-invasive exfoliative cytology for the screening of oral cancer.* Analyst, 2019. **144**(4): p. 1309-1325.

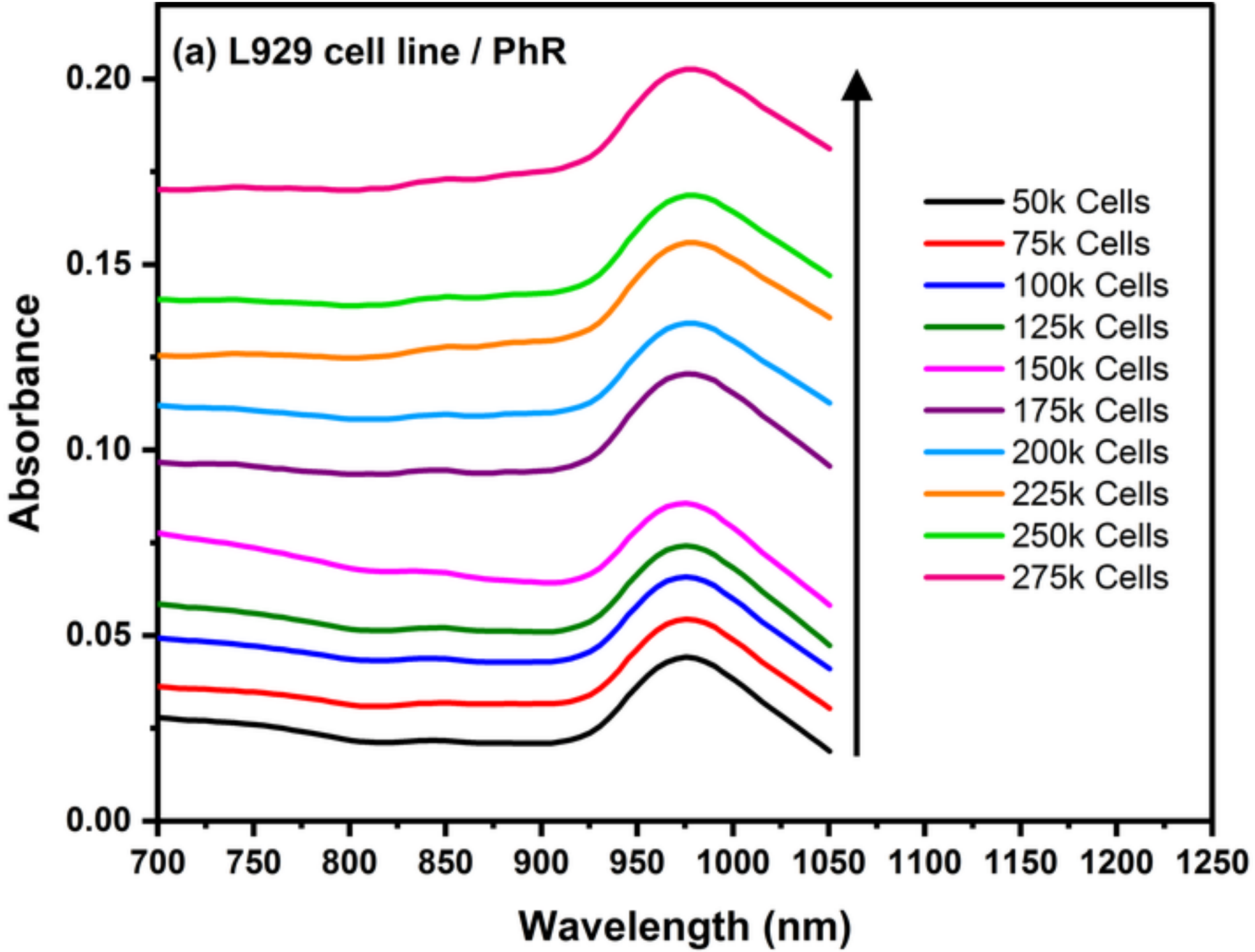
6. Simick, M.K., et al., *Non-ionizing near-infrared radiation transillumination spectroscopy for breast tissue density and assessment of breast cancer risk.* Journal of biomedical optics, 2004. **9**(4): p. 794-803.

Highlights

- HeLa, DU145, and L929 cells were cultured.
- Absorbance spectra were recorded for 50k-275k cells of L929, DU145, and HeLa cultured in the presence and absence of phenol red.
- PCA achieved clear discrimination between three cell lines with subgroups detected based on phenol red.
- SIMCA showed a high accuracy of classification, where larger class distances were obtained between the non-stained cells.

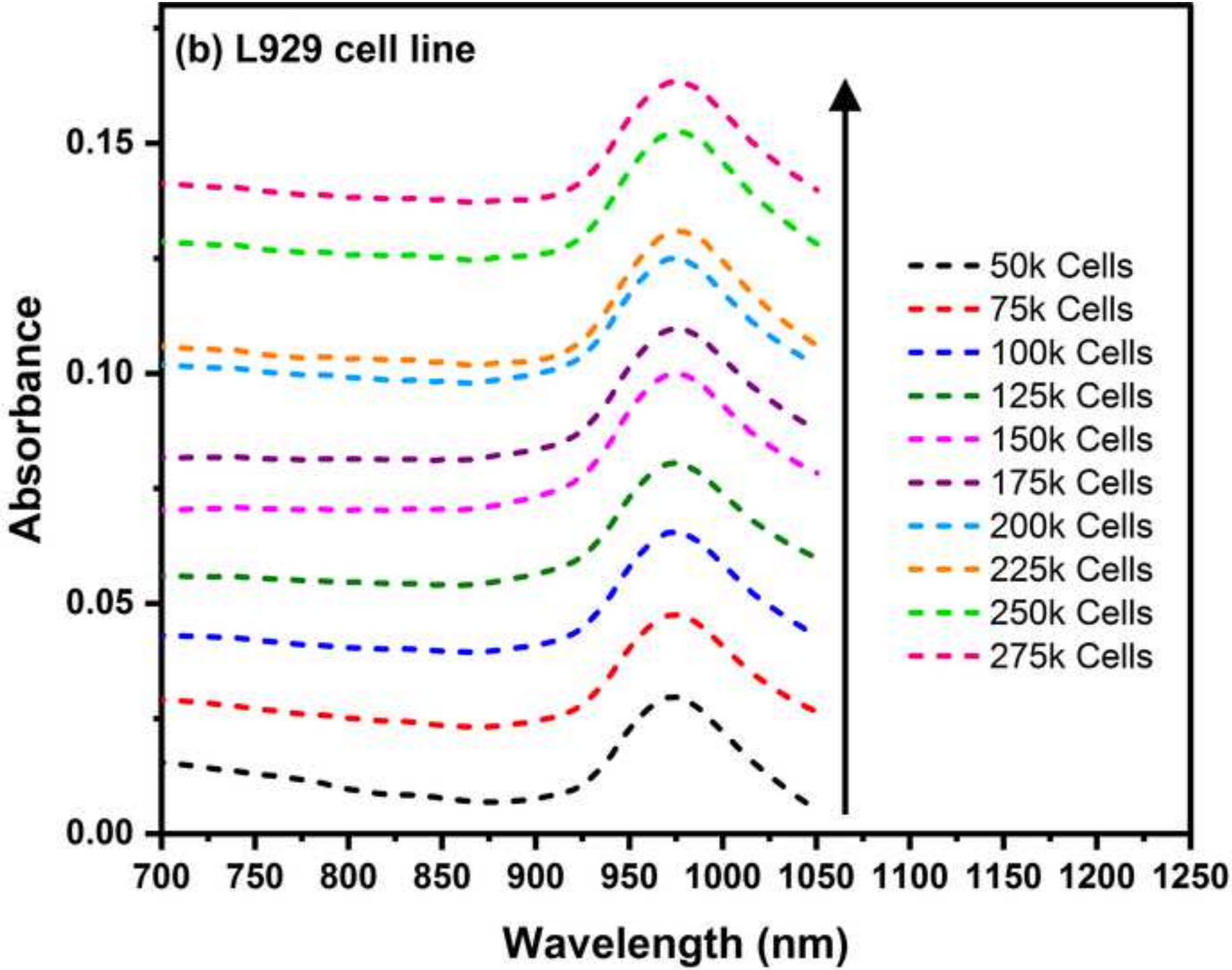
Figure

Click here to access/download;Figure;Figure 1(a).jpg



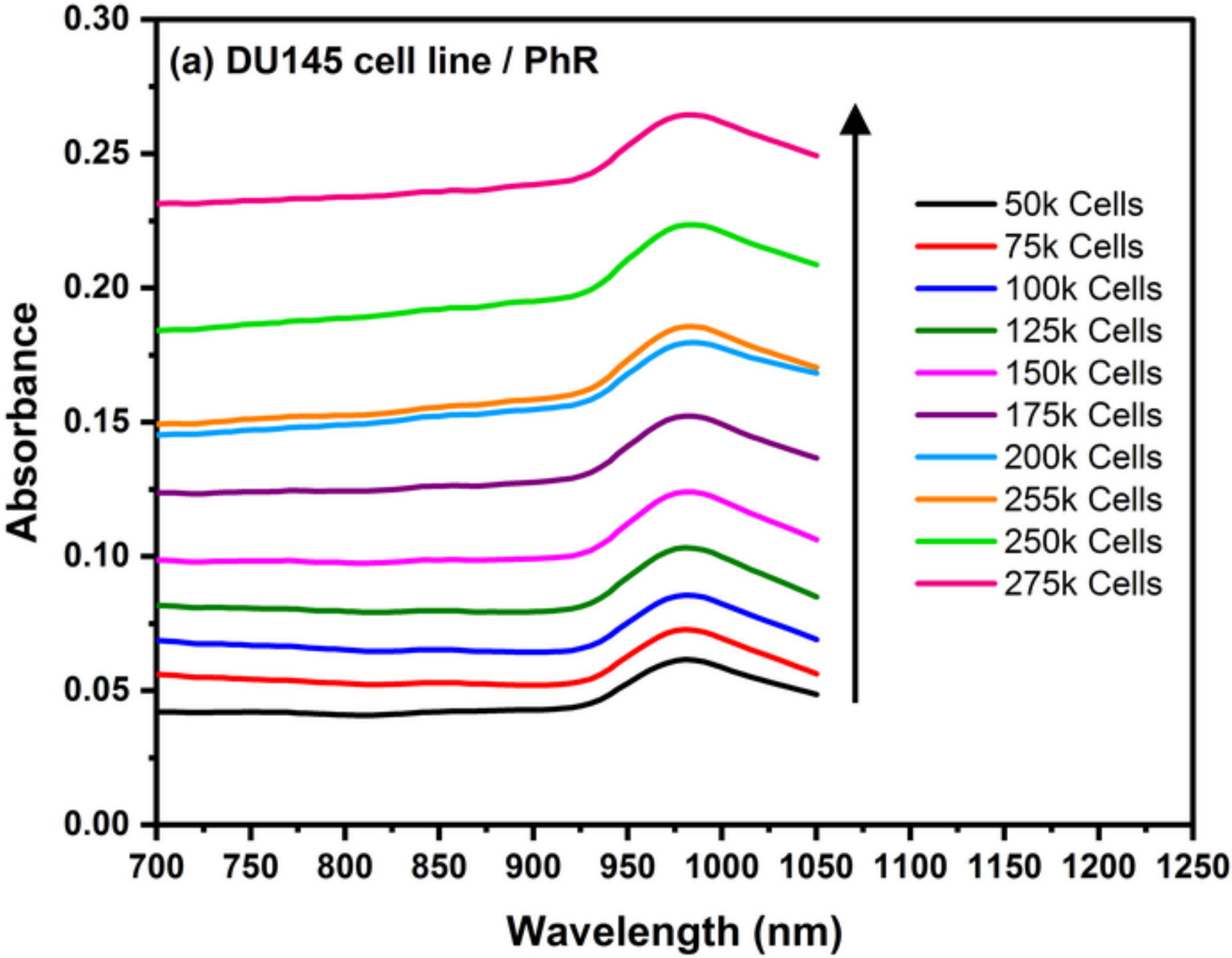
Figure

Click here to access/download;Figure;Figure 1(b).jpg



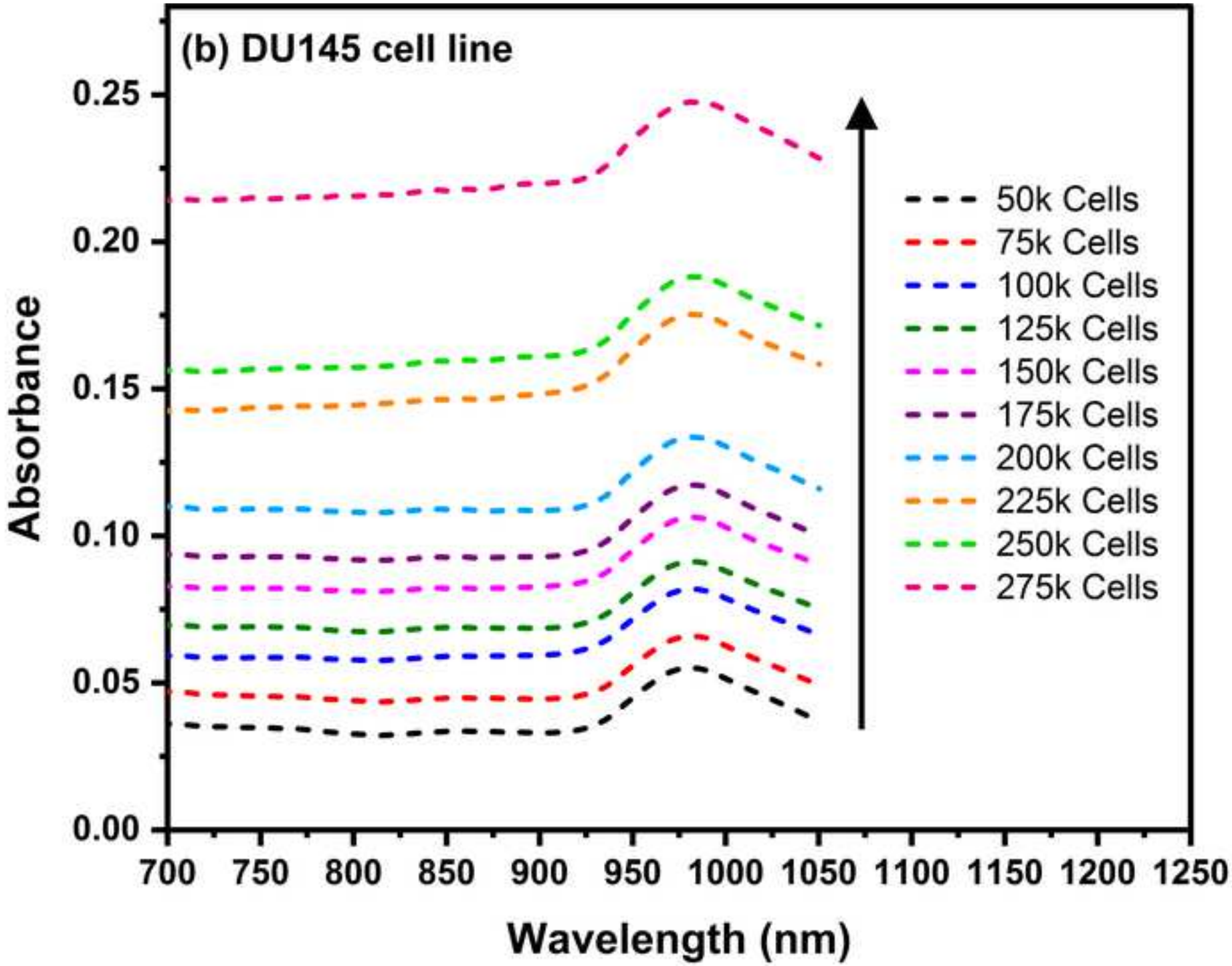
Figure

Click here to access/download;Figure;Figure 2(a).jpg



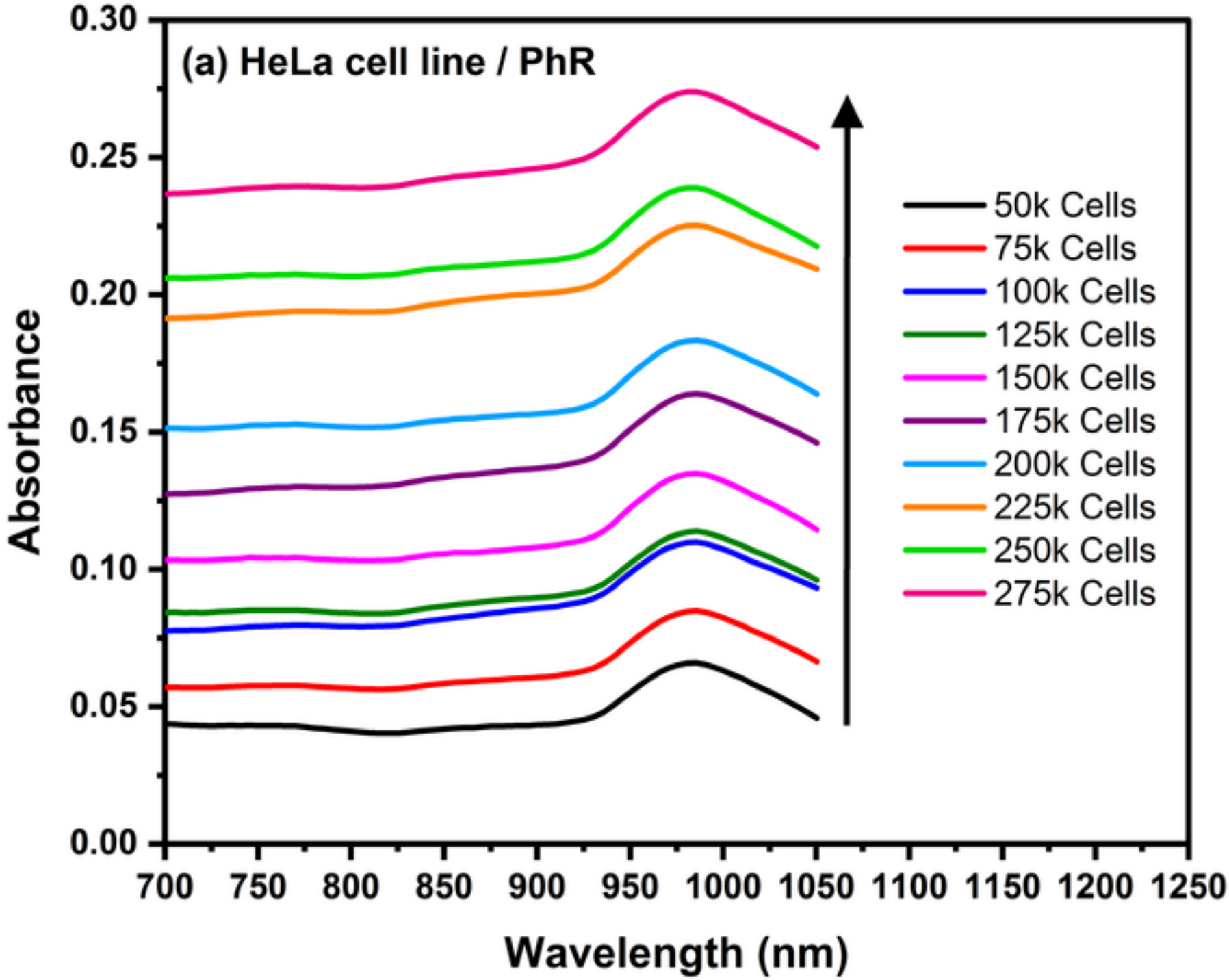
Figure

Click here to access/download;Figure;Figure 2(b).jpg



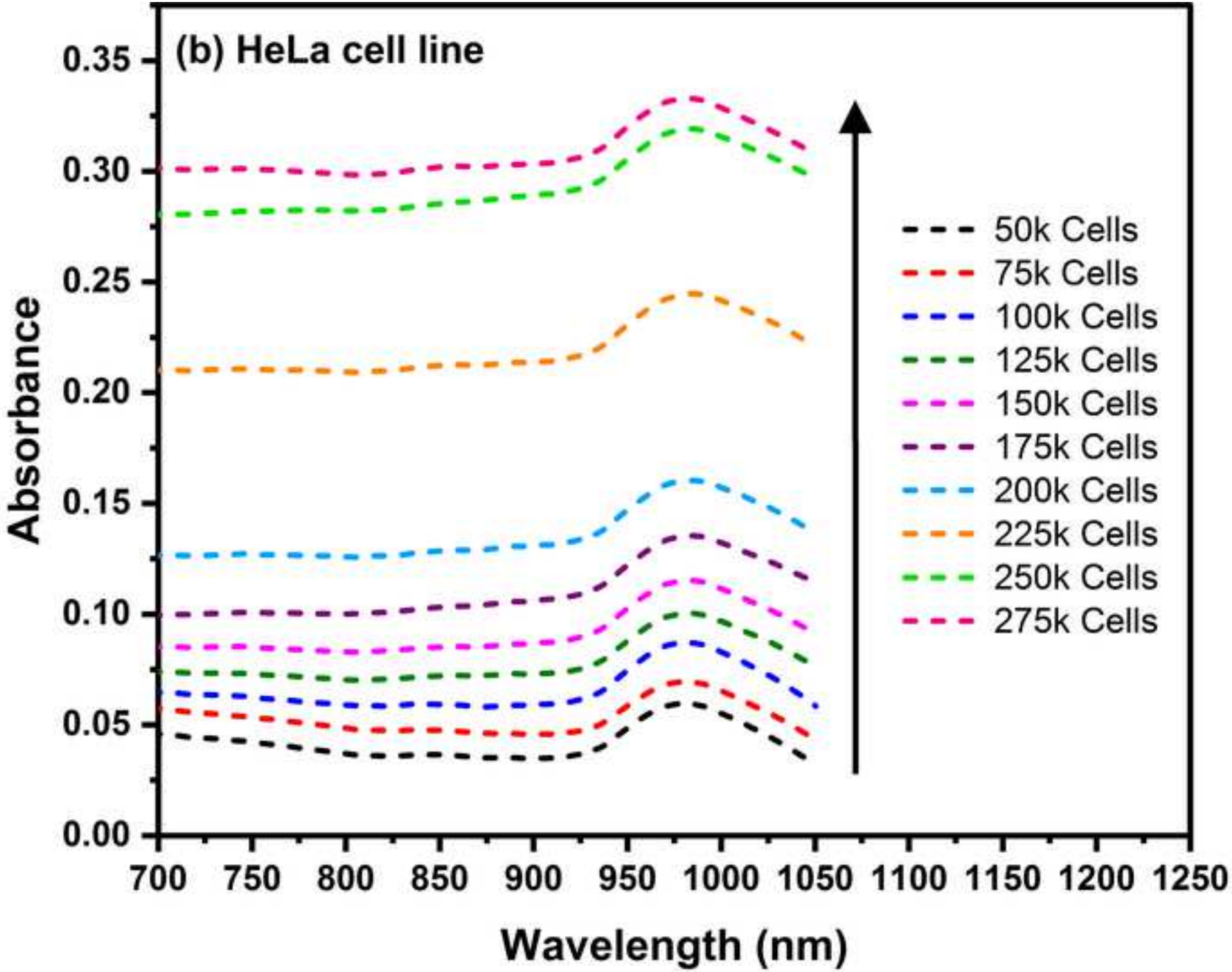
Figure

Click here to access/download;Figure;Figure 3(a).jpg



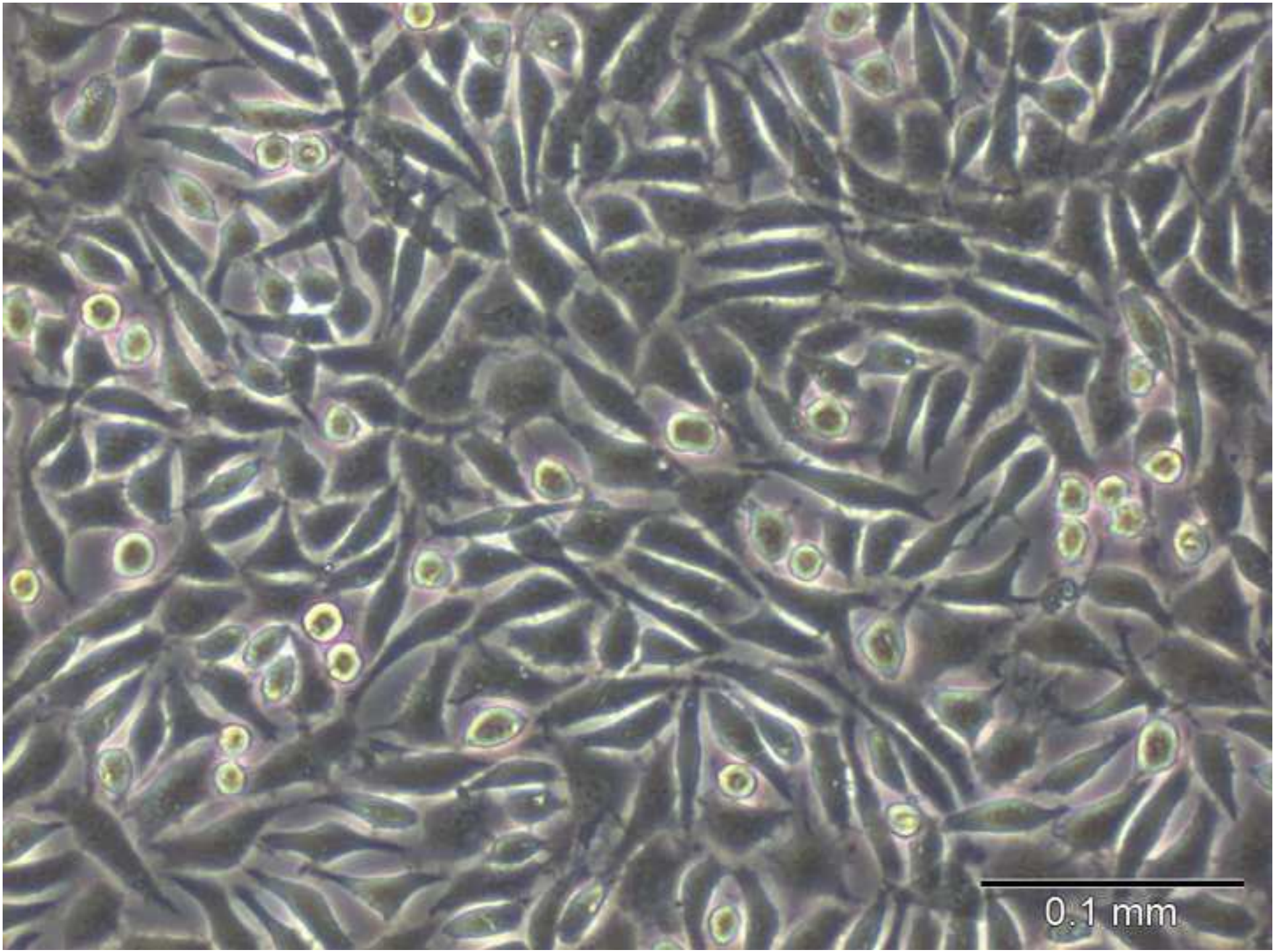
Figure

Click here to access/download;Figure;Figure 3(b).jpg



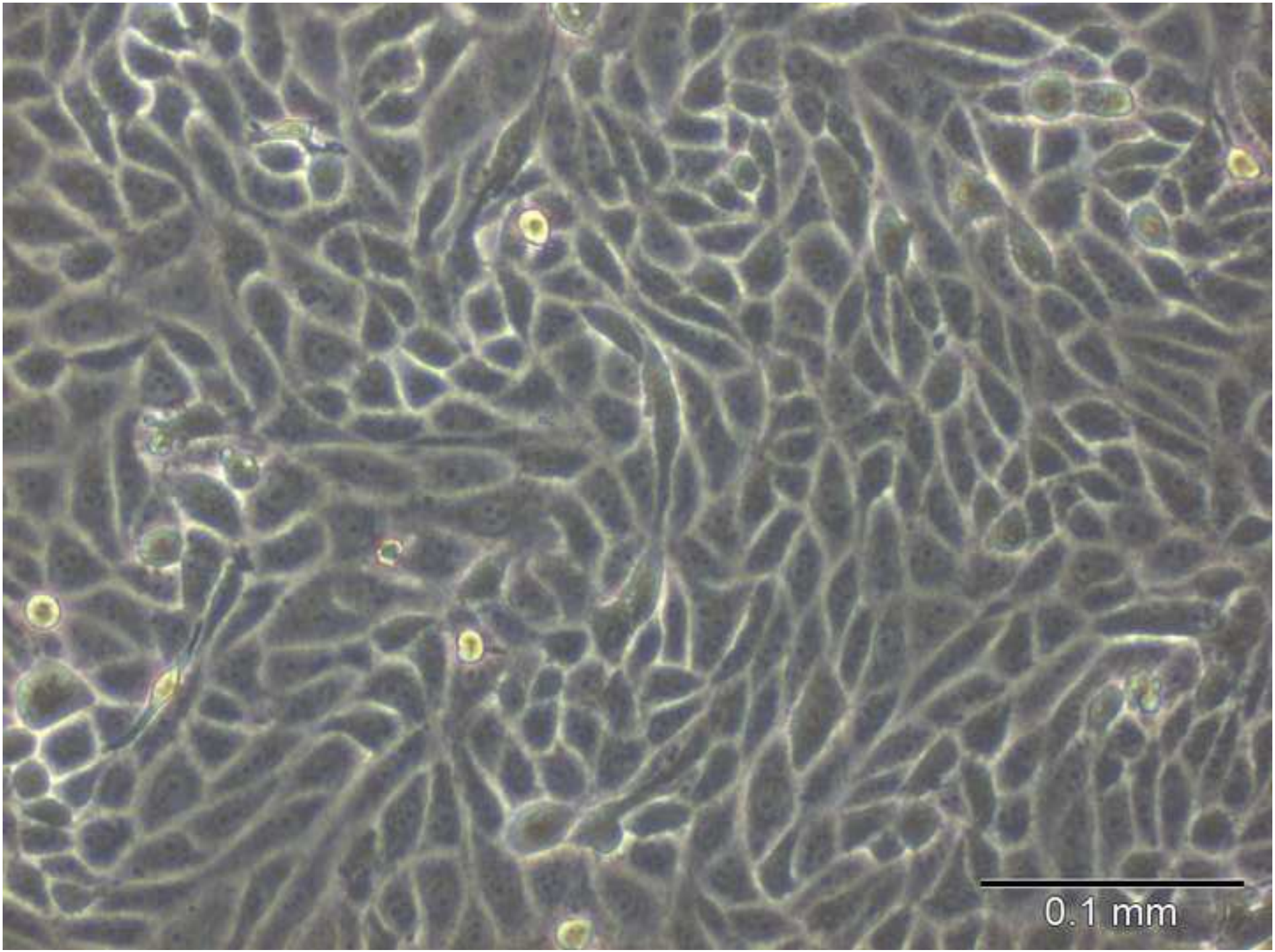
Figure

Click here to access/download;Figure;Figure 4(a).jpeg



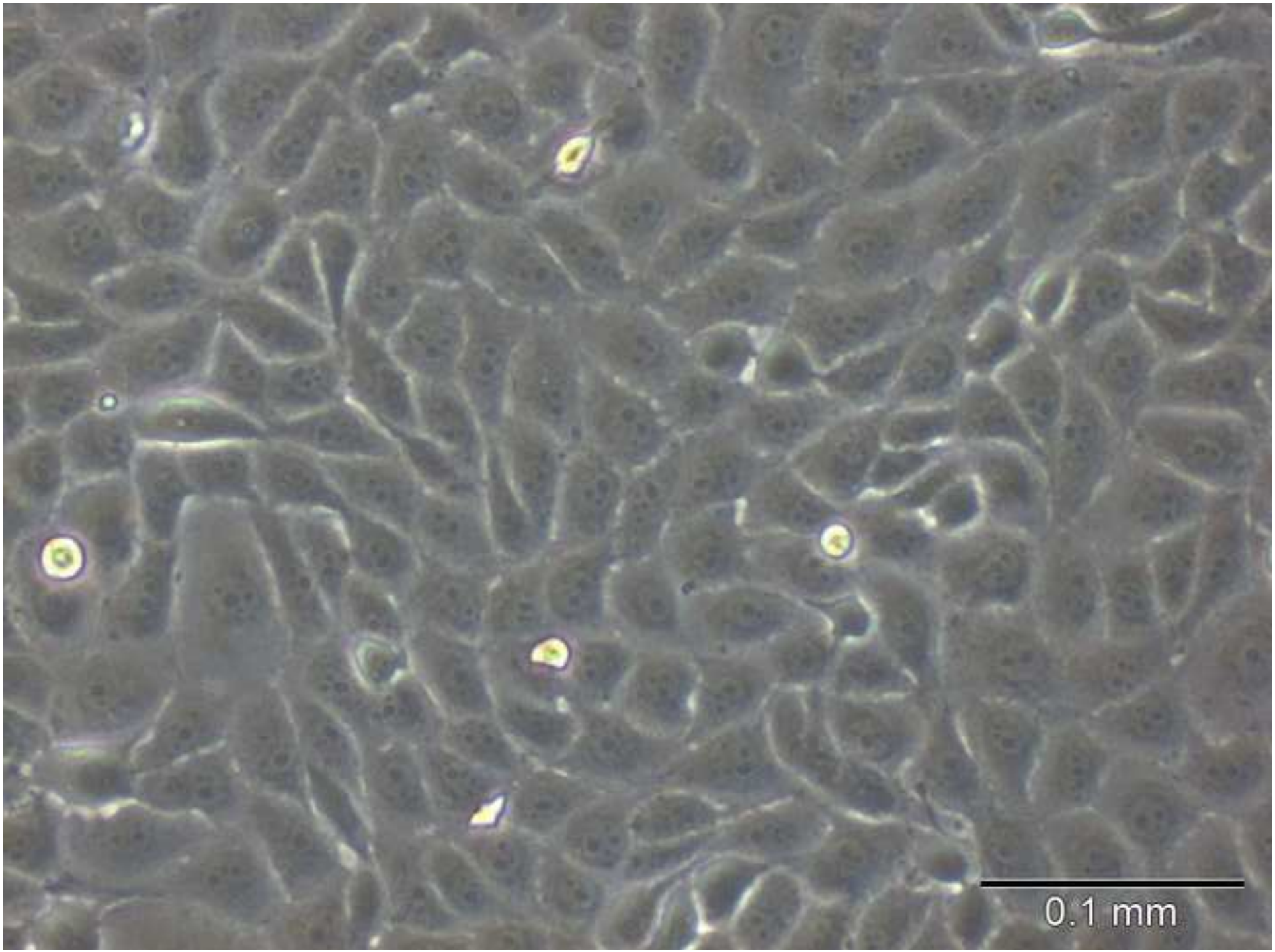
Figure

Click here to access/download;Figure;Figure 4(b).jpeg



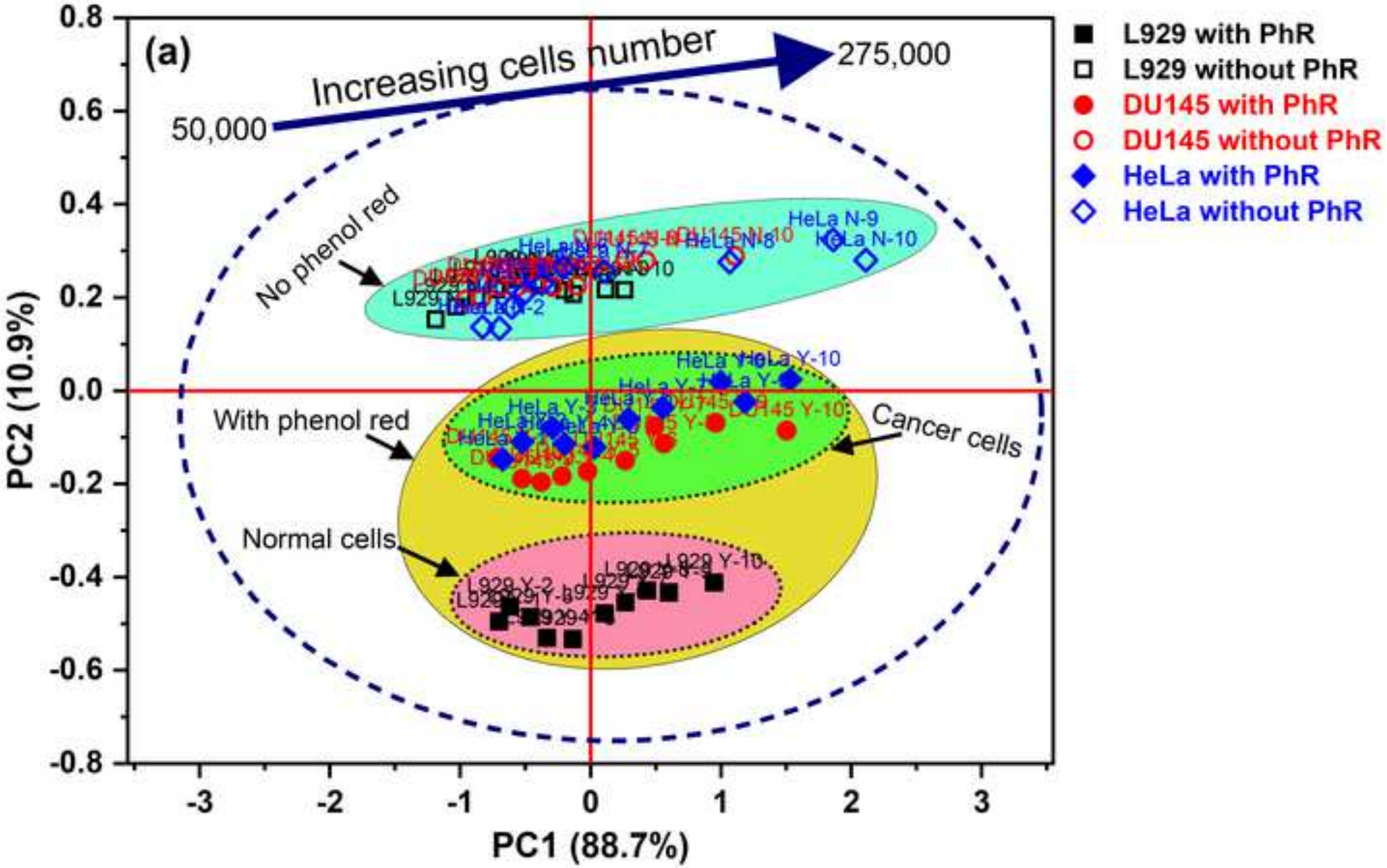
Figure

Click here to access/download;Figure;Figure 4(c).jpeg



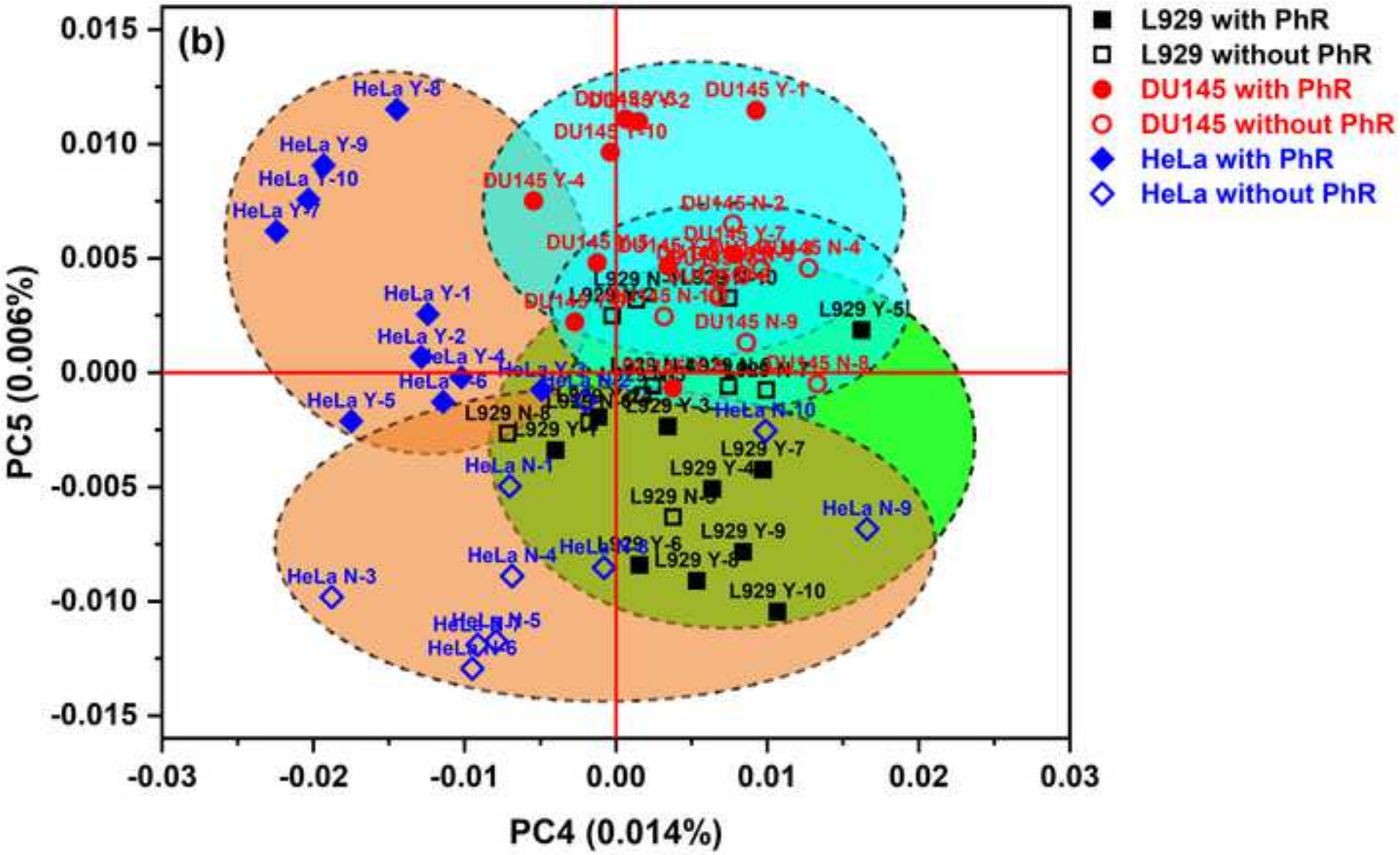
Figure

Click here to access/download;Figure;Figure 5(a).jpg



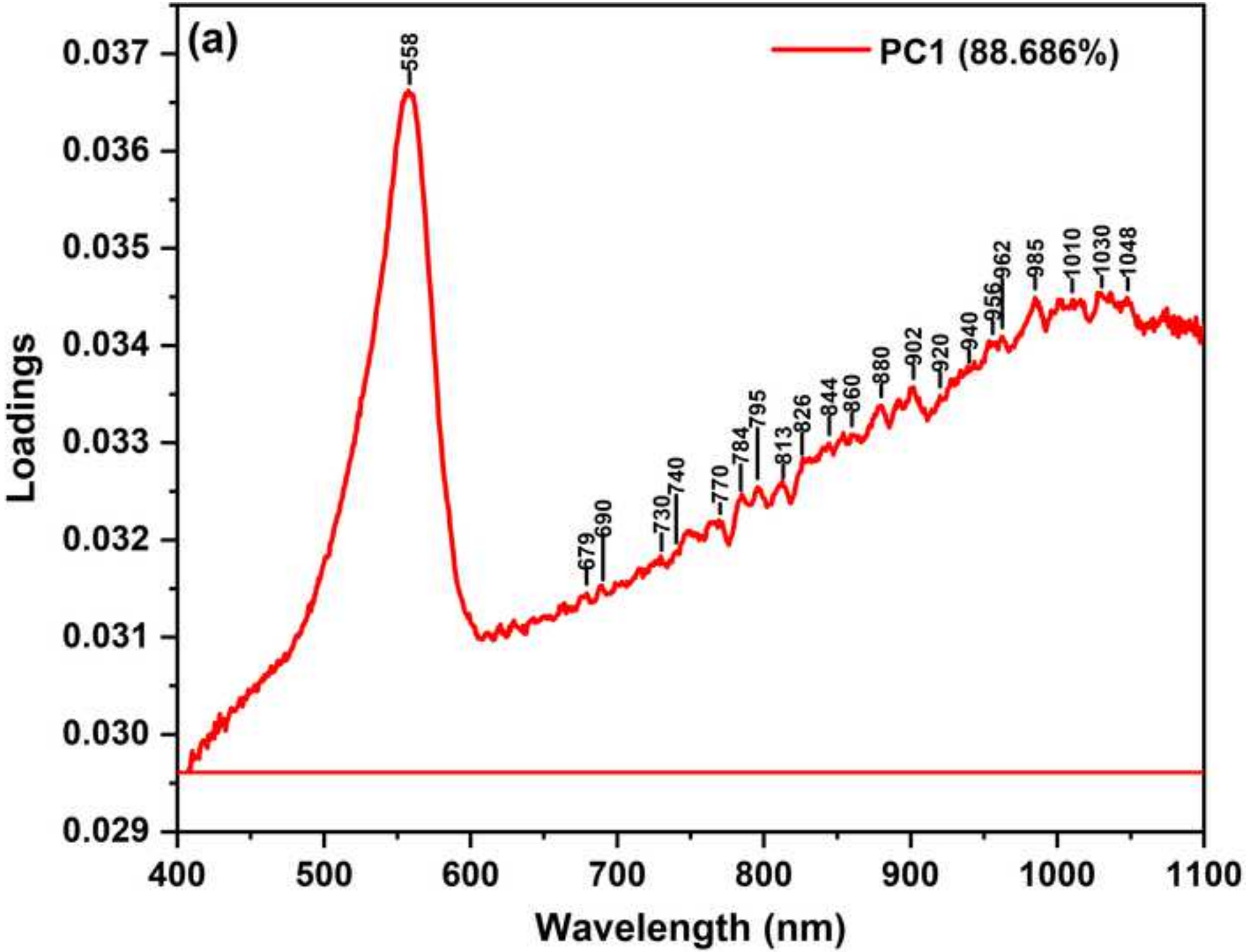
Figure

Click here to access/download;Figure;Figure 5(b).jpg



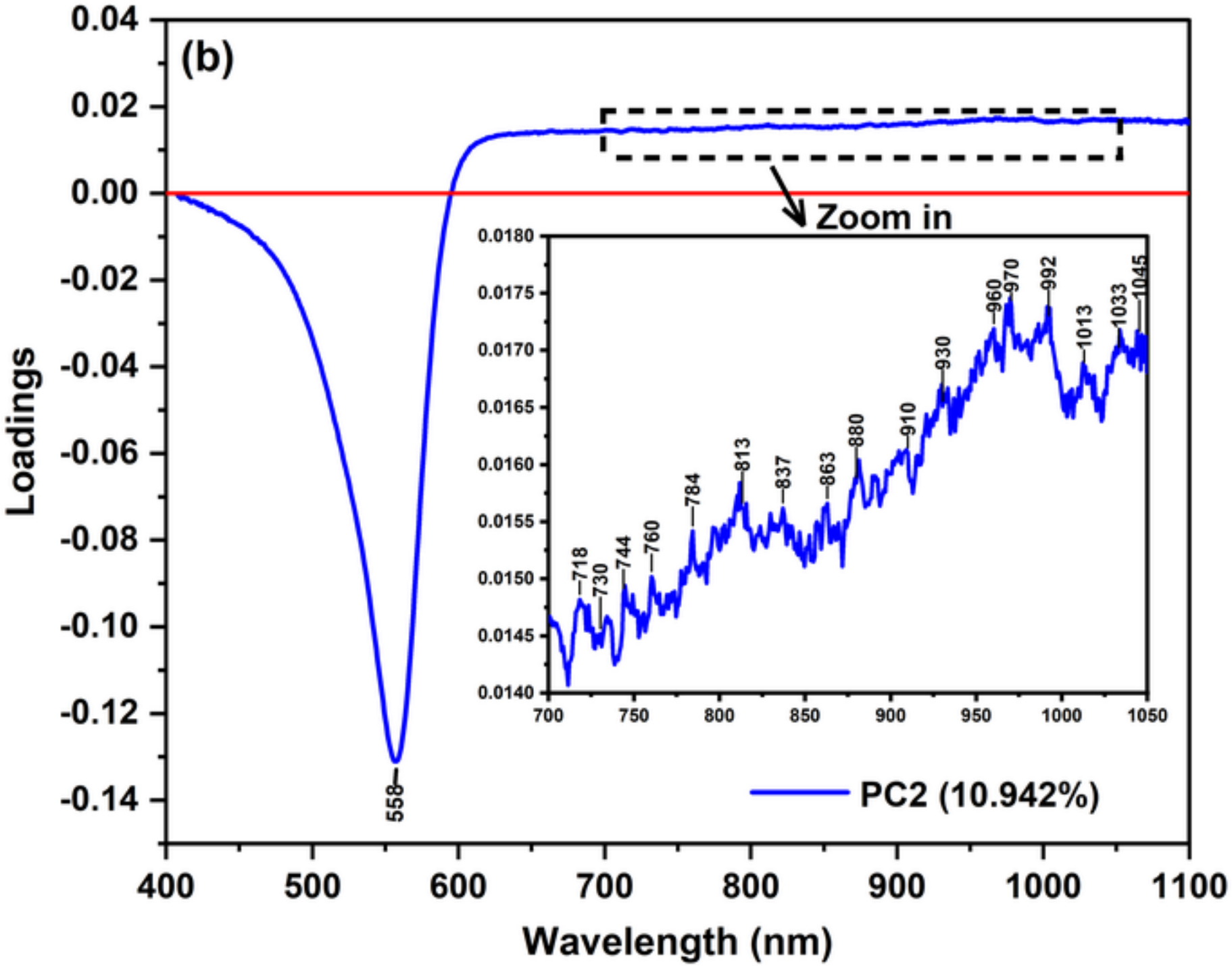
Figure

Click here to access/download;Figure;Figure 6(a).jpg



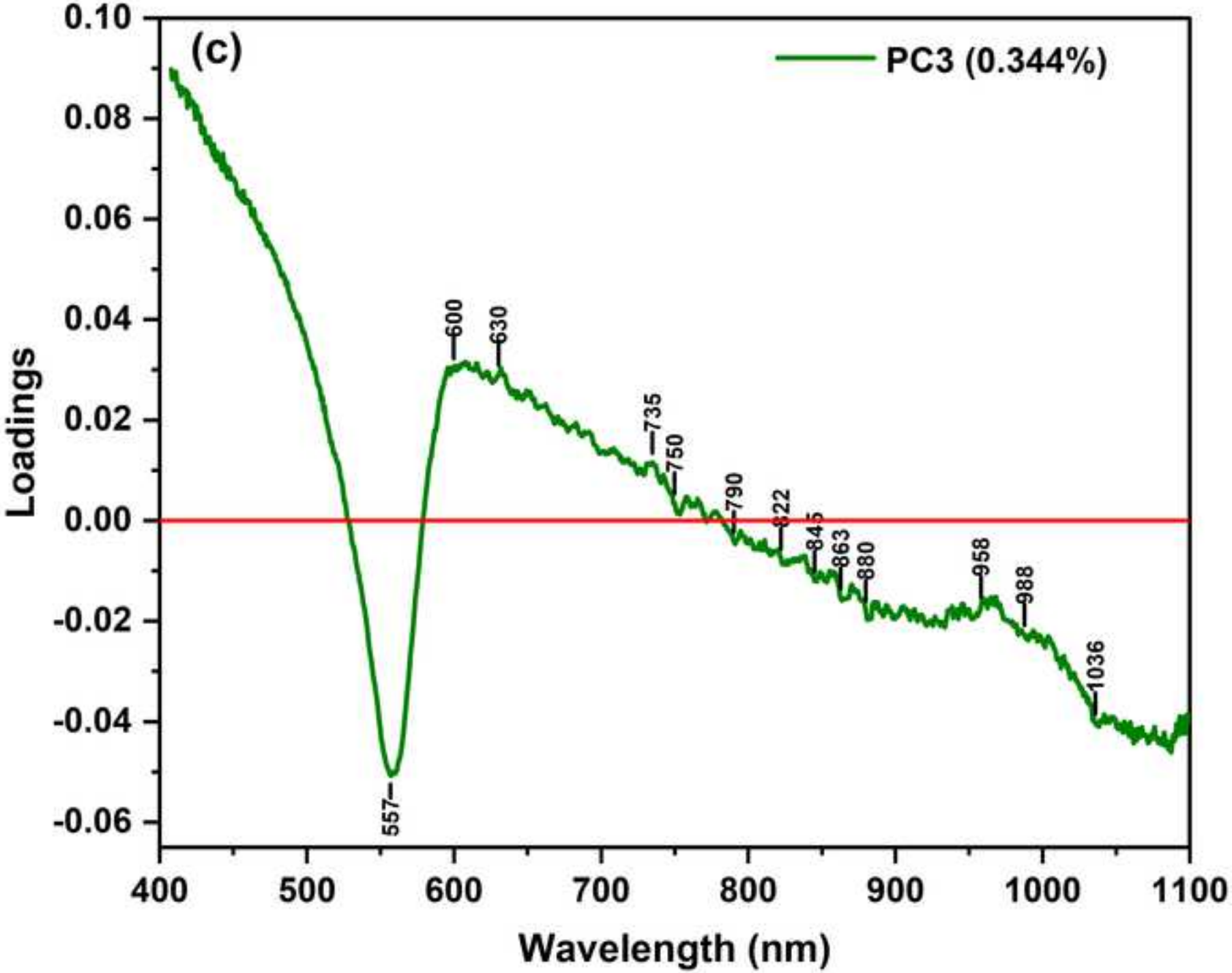
Figure

Click here to access/download;Figure;Figure 6(b).jpg



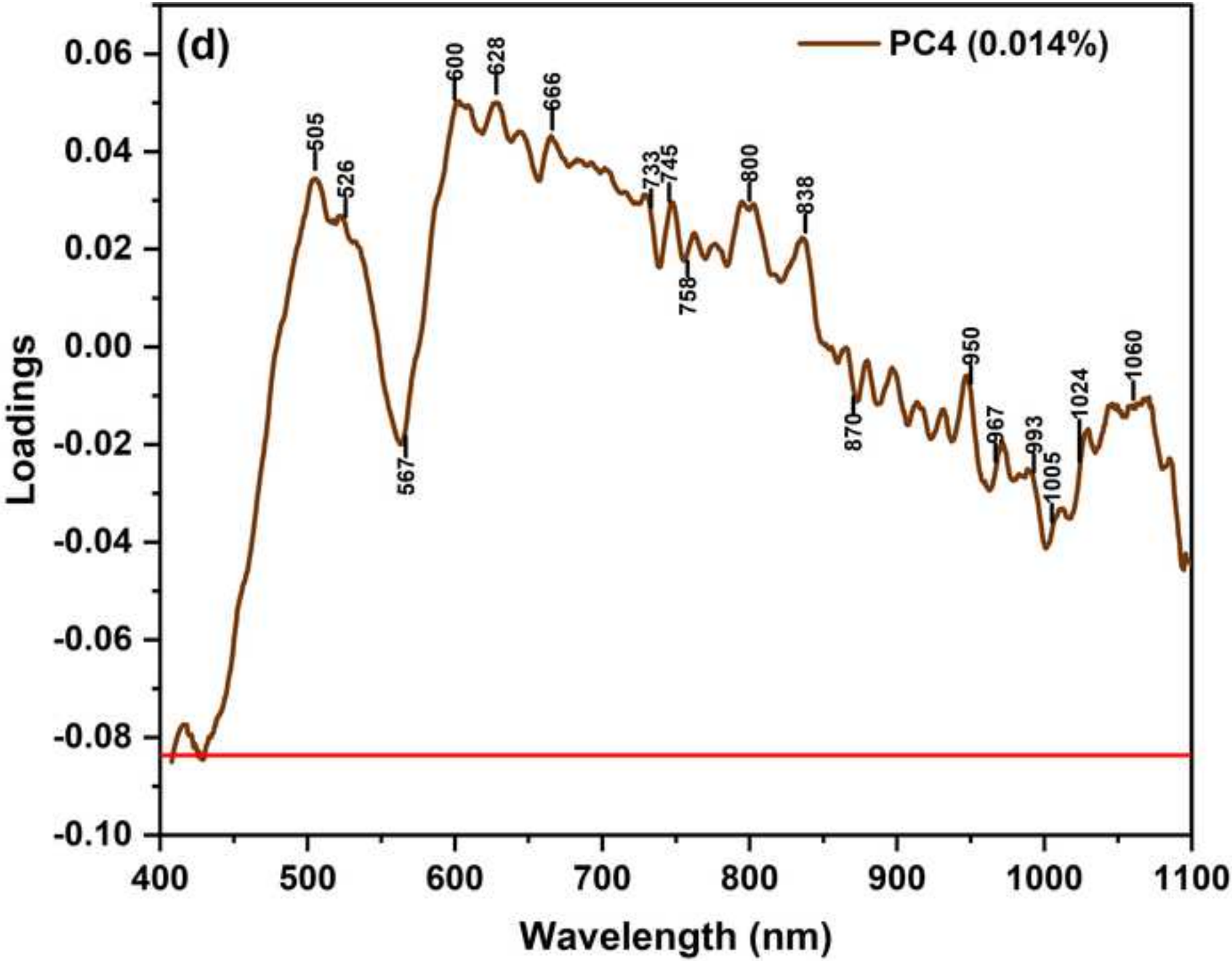
Figure

Click here to access/download;Figure;Figure 6(c).jpg



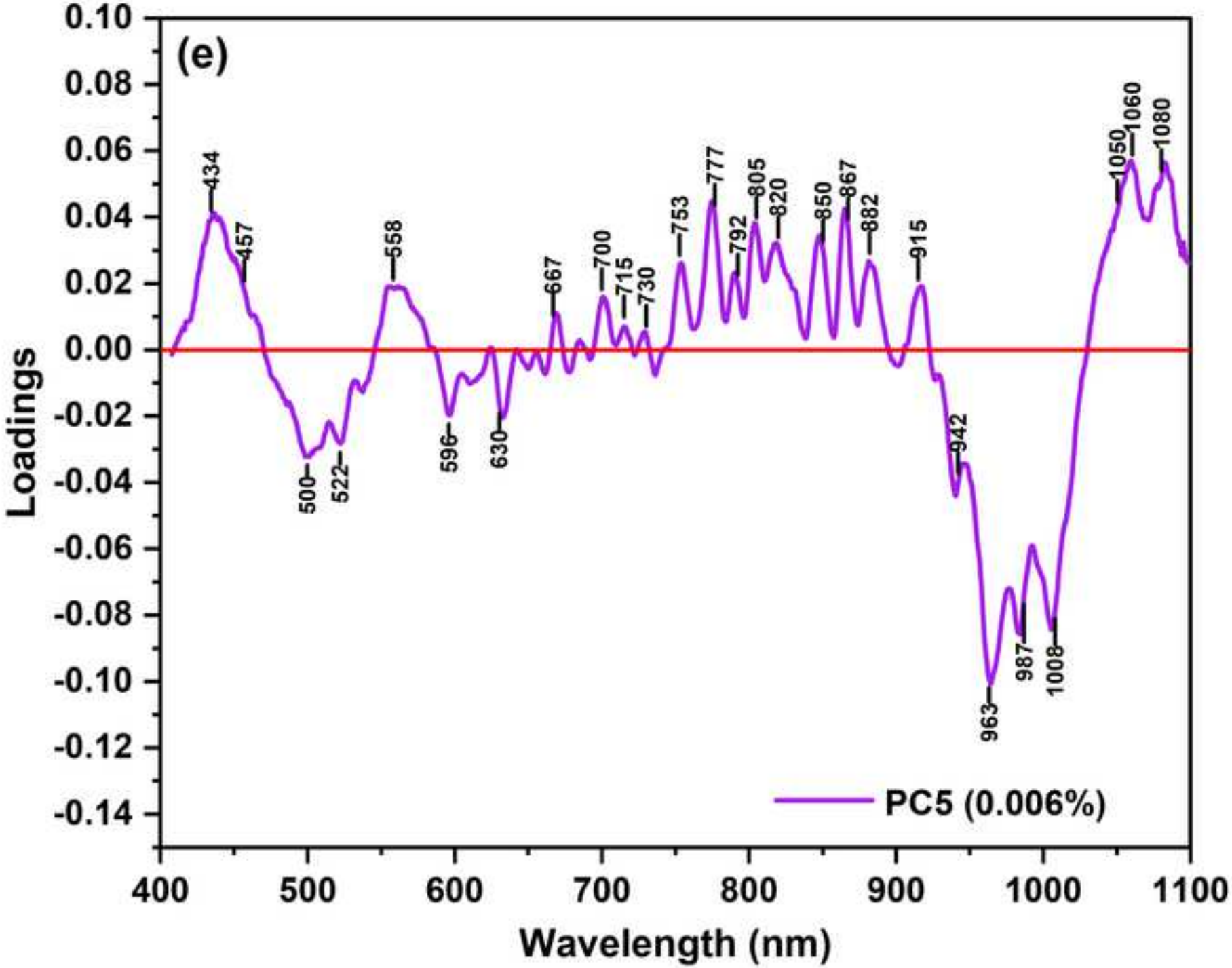
Figure

Click here to access/download;Figure;Figure 6(d).jpg



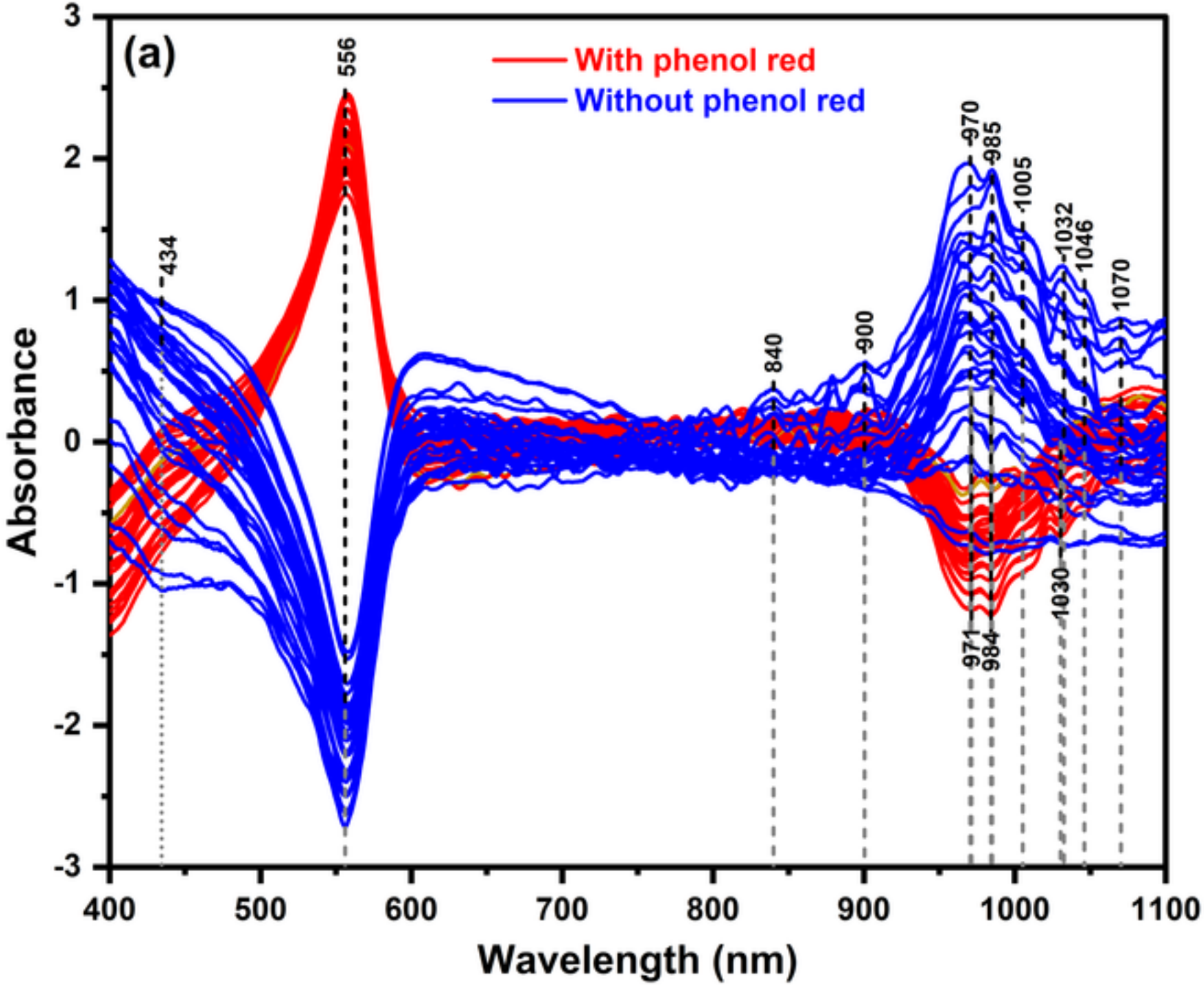
Figure

Click here to access/download;Figure;Figure 6(e).jpg



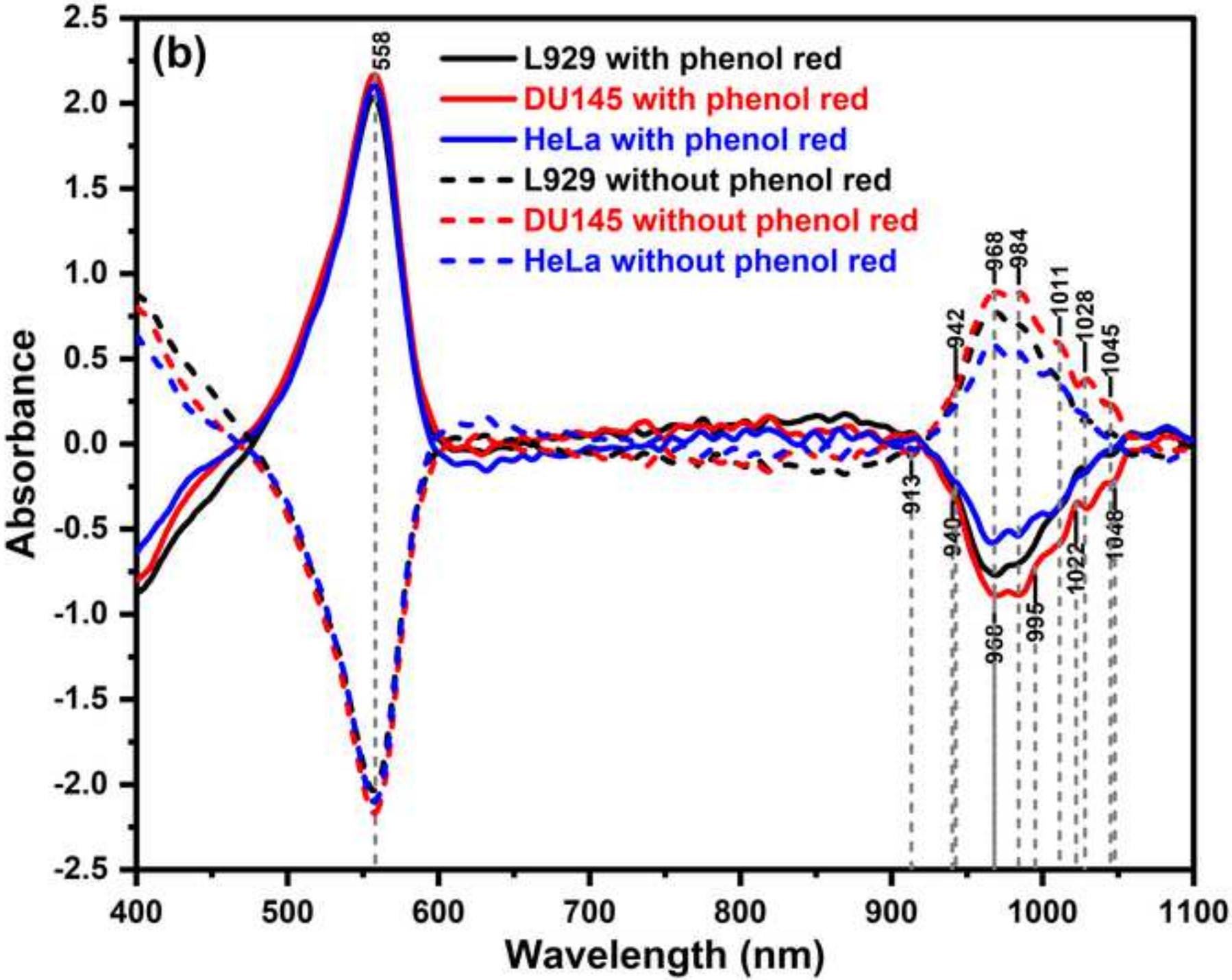
Figure

Click here to access/download;Figure;Figure 7(a).jpg



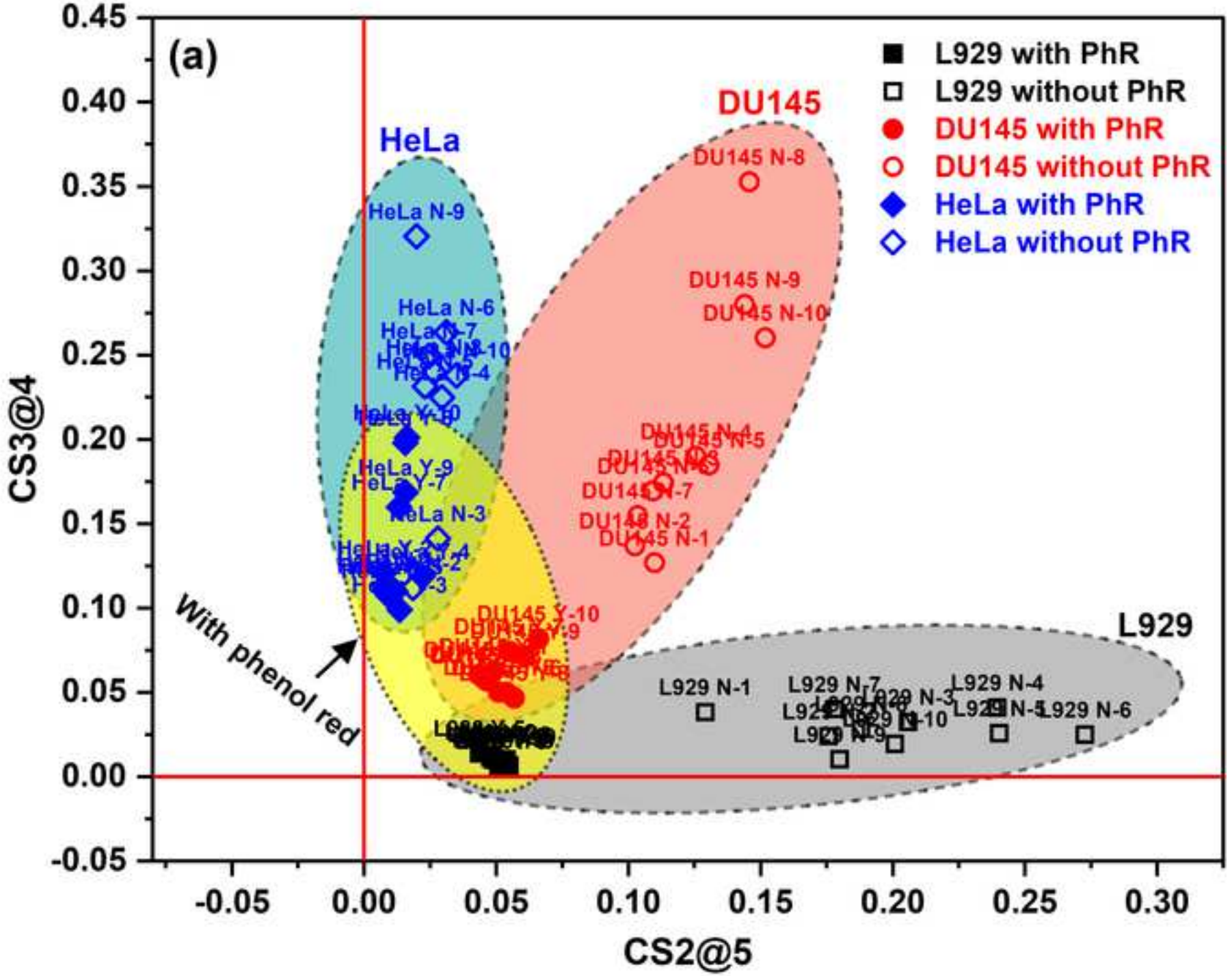
Figure

Click here to access/download;Figure;Figure 7(b).jpg



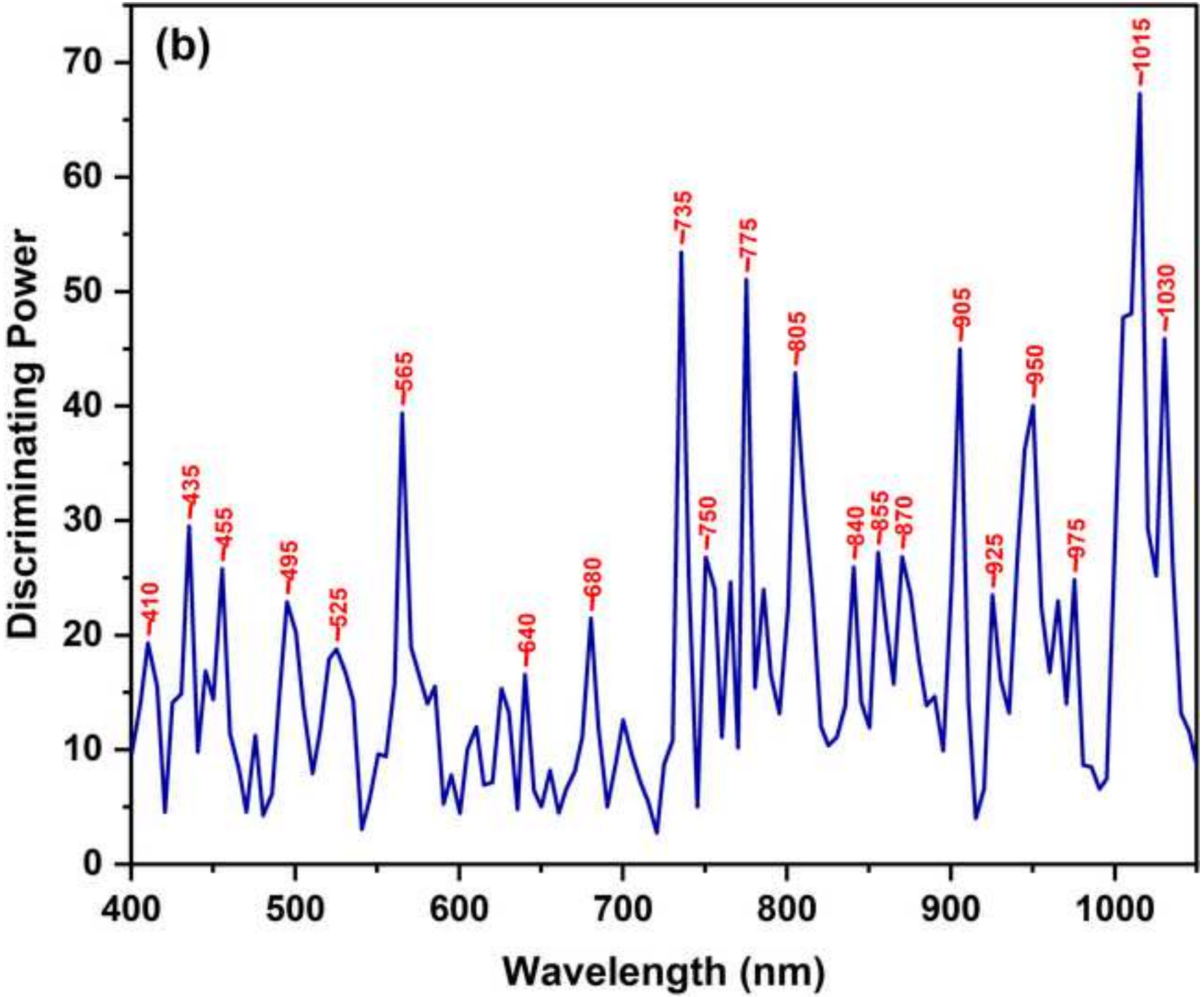
Figure

Click here to access/download;Figure;Figure 8(a).jpg



Figure

Click here to access/download;Figure;Figure 8(b).jpg



Cell line	Number of viable cells (cells)	Concentration of cell suspension (cells/mL)	Volume of cell suspension (mL)	Total cell number (cells)
L929	581	2 905 000	5	14 525 000
HeLa	436	2 180 000	5	10 900 000
DU145	325	1 625 000	5	8 125 000

Cell line	Number of viable cells (cells)	Concentration of cell suspension, (cells/mL)	Volume of cell suspension (mL)	Total cell number (cells)
L929	528	2 640 000	6	15 840 000
HeLa	673	2 019 000	6	20 190 000
DU145	211	1 055 000	5	5 275 000

Initial volume of cell suspension (mL), V_1			Concentration of cell seeding (cells/mL), M_2	Volume of cell seeding (mL), V_2	Total number of cells per well (cells)
L929	HeLa	DU145			
60.2	80.3	92.3	25 000	2	50 000
90.4	120.4	138.5	37 500	2	75 000
120.5	160.6	184.6	50 000	2	100 000
150.6	200.7	230.8	62 500	2	125 000
180.7	240.8	276.9	75 000	2	150 000
210.8	281.0	323.1	87 500	2	175 000
241.0	321.1	369.2	100 000	2	200 000
271.1	361.2	415.4	112 500	2	225 000
301.2	401.4	461.5	125 000	2	250 000
331.3	441.5	507.7	137 500	2	275 000

Initial volume of cell suspension (mL), V_1			Concentration of cell seeding (cells/mL), M_2	Volume of cell seeding (mL), V_2	Total number of cells per well (cells)
L929	HeLa	DU145			
56.8	44.6	142.2	25 000	2	50 000
85.2	66.9	213.3	37 500	2	75 000
113.6	89.2	284.4	50 000	2	100 000
142.0	111.4	355.5	62 500	2	125 000
170.5	133.7	426.5	75 000	2	150 000
198.9	156.0	497.6	87 500	2	175 000
227.3	178.3	568.7	100 000	2	200 000
255.7	200.6	639.8	112 500	2	225 000
284.1	222.9	710.9	125 000	2	250 000
312.5	245.2	782.0	137 500	2	275 000

Wavelength (nm)	Tentative band assignment	Ref.
940-942	Free water molecules, S_0 (1410-1413 nm, C5 WAMACs)	[87]
	Water vapour (940 nm)	[88-90]
	Water vapour (942 nm)	[91]
968	Hydration shell OH-$(H_2O)_{4,5}$ (1452 nm, C8 WAMACs)	[87]
	Hydroxide absorption peak (967 nm)	[92]
	O–H stretch, 2nd overtone of ROH and H_2O (970 nm)	[11]
	O–H stretch, 2nd overtone of H_2O (960–970 nm)	[84]
984	Water molecules with 3 hydrogen bonds, S_3 (1476 nm, C10 WAMACs)	[87]
	N–H stretch, 2nd overtone of CONHRn (secondary amides) (981 nm)	[84]
995	Water molecules with 4 hydrogen bonds (1492.5 nm, C11 WAMACs)	[87]
	N-H stretch, 2nd overtone (aromatic amines) (995 nm)	[84]
	N-H stretch, 2nd overtone (amines, general) (1000 nm)	[84]
1011	Strongly bound water (1516.5 nm, C12 WAMACs)	[87]
	OH from secondary (1004 nm) and tertiary alcohols (1006 nm)	[84]
	N-H stretch, 2nd overtone (amines, general) (1015 nm)	[84]
	2nd overtone Superoxide Tetrahydrate $O_{2^-}.(H_2O)_4$ (1010 nm)	[93]
1022	2× N–H stretch + 2× amide I (protein) (1020 nm)	
	N–H stretch, 2nd overtone (protein) (1007 nm)	[84]
	N–H stretch, combination of RNH2 (primary amine) (1020 nm)	
	Aqueous proton $[H^+.(H_2O)_6]$ - H_2O in $H_5O_2^+$ symmetric stretch, 2nd overt (1018 nm)	[94]
	2nd overtone intramolecular hydrogen bond stretches in (OH-$(H_2O)_4$) (1020 nm)	[95]
1028	N-H stretch, 2nd overtone of RNH2 (1030 nm)	[84]
	Aqueous proton $[H^+\cdot(H_2O)_3]$ - H_3O^+ symmetric stretch, 3rd overt (1033 nm)	[94]
1045-48	N-H stretch, 2nd overtone of amides (primary, bonded; two bonds) (~1050 nm)	
	N-H stretch, combination of CONH (primary amide) (1047 nm)	[84]
	N-H stretch, 3rd overtone of amides (secondary, bonded; one band) (1000-1067 nm)	
	$H_{11}O_5^+$ H-bonded OH stretch, 2nd overt (1041 nm)	[96]
	Aqueous proton $[H^+\cdot(H_2O)_6]$ - H_2O in $H_5O_2^+$ symmetric stretch, 2nd overt (1055 nm)	[94]

PCs	Explained Variance (%)			Actual Class			Correct classification (%)	Misclassifications (%)
				L929	DU145	HeLa		
5	99.93	SIMCA Class	L929	**20**	0	0	100	0
5	99.89		DU145	0	**20**	0	100	0
4	99.86		HeLa	0	0	**20**	100	0

Wavelength (nm)	Tentative band assignment	Ref.
735	Hydrogen-bonded -OH, 3rd overt	[97]
	Water molecules with 3 hydrogen bonds, S_3 (1470 nm, C10 WAMACS)	[87]
775	Hydrogen bonded water (770 nm)	[98]
	N–H stretch, 3rd overtone of RNH_2 (779 nm)	[11]
	N–H stretch, 3rd overtone of amides (secondary, bonded; one band) (750-800 nm)	[84]
805	Maximum absorption peak for ice (800 nm)	[91]
905	C-H stretch, 3rd overtone of protein (910 nm)	[99, 100]
	3rd overtone of a C-H stretching mode and the 2nd overtone of an N-H stretching mode of protein (906 nm)	[101]
	Hydration shell OH-$(H_2O)_{1,2,4}$ (1357.5 nm, C2 WAMACs)	[87]
950	O-H stretch (954 nm), 2nd overtone of ROH (alcohols) (940-970 nm)	[102]
	Water interacting with sugar (950 nm)	[103]
	Hydration water (1425 nm, C6 WAMACS)	[87]
	2nd overtone of O-H stretching of water interacting with protein (950-960 nm)	[101]
1015	H_2O deionized; 2nd overt (1010.1 nm)	[84]
	Hydrated proton $[H^+(H_2O)_6]$ - H_2O symmetric stretch, 2nd overt (1018.1 nm)	[94]
	2X N-H stretch + 2X amide I (protein) (1020 nm)	
	N-H stretch, 2nd overtone (protein) (1007 nm)	[84]
	N-H stretch, combination of RNH2 (primary amine) (1020 nm)	
	Protein (1020 nm)	[104]
	Strongly bound water (1522.5 nm, C12 WAMACs)	[87]
1030	N-H stretch, 2nd overtone of RNH_2	[84]
	3rd overtone of a C-H stretching mode and the 2nd overtone of an N-H stretching mode of protein (1032 nm)	[101]

Declaration of Interest Statement

The authors declare that they have no conflict of interest.

Corresponding Author

Ahmad Fairuz Omar

Declaration of Interest Statement

The authors declare that they have no conflict of interest.

Corresponding Author

Ahmad Fairuz Omar

Integration of Near-Infrared Spectroscopy and Aquaphotomics for Discrimination of Cultured Cancerous Cells using Phenol Red

Muna E. Raypah [1], Jelena Muncan [2*], Suhainah Sudik [1], Ahmad Fairuz Omar [1**], Mohd Hafiz Mail [3], Roumiana Tsenkova[2], and Azman Seeni [4]

[1] School of Physics, Universiti Sains Malaysia, 11800, Pulau Penang, Malaysia
[2] Aquaphotomics Research Department, Faculty of Agriculture, Kobe University, Kobe, Japan
[3] Malaysian Institute of Pharmaceuticals and Nutraceuticals, National Institute of Biotechnology Malaysia, Ministry of Energy, Science, Technology, Environment and Climate Change, 11700 Penang, Malaysia
[4] Advanced Medical and Dental Institute, Universiti Sains Malaysia, Bertam, 13200, Pulau Penang, Malaysia
*Corresponding authors e-mails: fairuz_omar@usm.my and jmuncan@people.kobe-u.ac.jp

Abstract

Presently, mammalian cell lines are the most utilized hosts for the production of biopharmaceuticals. Optical spectroscopy is extensively used for the physical or physiological measurements of cellular features to study the diagnosis, prognosis, and treatment of different cancers. Near-infrared (NIR) spectroscopy and aquaphotomics are ubiquitous techniques increasingly used for nondestructive assessment. The key objective of this study is to evaluate the feasibility of absorbance spectra in the NIR region (700-1100 nm) coupled with aquaphotomics analysis to classify the different cancerous cell lines. Human cervix adenocarcinoma cells (HeLa) and human prostate carcinoma (DU145) were the cancerous cell lines, while normal mouse skin fibroblast (L929 cell line) was used as a reference for the assessment. The NIR absorbance spectra for 50,000 to 275,000 cells of L929, DU145, and HeLa cell lines in culture media prepared with and without phenol red (PhR) were subjected to principal component analysis (PCA) and soft independent modeling of class analogies (SIMCA) models. The PCA of the spectral absorbance data revealed a clear discrimination between the three cell lines, with subgroups detected based on the presence or absence of PhR in the culture medium. The SIMCA method showed a high accuracy of classification, where larger class distances were obtained among the non-stained cells.

Keywords: Cancerous Cells; Culture Medium; Phenol Red; Near-infrared Spectroscopy; PCA; SIMCA; Water.

1. Introduction

The prevalence of cancer is persistently rising, resulting in an increased annual global mortality rate. With over 17 million new cases annually and approximately 10 million deaths, cancer persists as one of the foremost health complications, globally [1]. Cancer is a complex disease, typified by reprogrammed signaling, cellular physiological alterations, and uninhibited cell growth. A characteristic trait of cancer cells is their metabolic reprogramming, which facilitates fast cellular reproduction, migration, and alteration of their microenvironment, ultimately leading to metastasis [2]. Surgical biopsy is the gold standard method for cancer diagnosis and has improved in speed and accuracy. However, the surgical biopsy is painful for patients and not suitable for early cancer screening. Liquid biopsy is a non-invasive cancer diagnosis method that detects highly sensitive biomarkers and could be used in early cancer detection. This technique is less efficient due to the cell loss or other cell contamination and potential uncertainties in the subsequent immunoassay. X-ray computed tomography (CT) technology is widely used in recognizing cancer, but it is not suitable for regular screening due to its toxicity. Magnetic resonance imaging (MRI) is a non-toxic technique, whereas it is expensive and time-consuming. Fluorescence imaging is a molecular imaging technique that is widely used in biological sample detection. Other methods such as single-photon emission CT, positron emission tomography, ultrasound, and endoscopy are also commonly used for the diagnosis of cancer. However, all imaging techniques are having considerably low sensitivity for the detection of a small number of cancer cells [3].

In the past 20 years, optical spectroscopy techniques have widely emerged as a non-invasive substitute for different conventional measurement techniques, and as a tool for medical diagnosis. The optical techniques are economic, portable with no side effects, and offer good spatio-temporal

resolution, and real-time functional information. The bulk changes in cell architecture as the tumor grows can be employed for the diagnosis of cancerous cells under light microscopy, thereby enabling a definitive diagnosis of the disease. These changes can also be utilized to determine the stage of cancer growth. Recently, optical spectroscopy has shown its potential non-destructive capability of identifying distinct spectral attributes that discriminate between tumor and normal cell. These marker bands provide a basis for the identification and therapeutic screening of several cancers [4]. The optical spectroscopy technique has been extensively used for detecting various types of cancer [5-7]. It is strongly believed that vibrational spectroscopy offers excellent potential to study the chemical structural characteristics of biological samples. Raman, Fourier transform infrared (FTIR), and NIR spectroscopic techniques collect highly specific information making them suitable for cell analysis. Raman and FTIR spectroscopy techniques are considered potential cancer diagnostic tools. These techniques have been well studied for discriminating normal and cancer states in various biological analytes including cells [8]. Fluorescence imaging and Raman spectroscopy have been applied for the accurate diagnosis of cancer [9]. In addition, the surface-enhanced Raman scattering (SERS) technique is generally used to enhance the conventional Raman scattering signal [10]. In recent years, SERS has been widely used in the diagnosis of various types of cancer [11-16]. Visible (Vis) and near-infrared (NIR) spectroscopy and imaging have been widely utilized for detection, diagnosis, and screening of cancerous cells. NIR spectroscopy has evolved considerably into a versatile technique in medical diagnosis, particularly for the diagnosis of carcinoma [17]. NIR spectroscopy plays a vital role in cancer detection, early prediction of responders during therapeutic interventions, and optimizing the treatment approach [18]. The supposed “optical window” occurs in the NIR part of the spectrum at the wavelength range of 650-1100 nm, which is suitable for most of the non-invasive measurements of biological

systems [19]. This overtone region is also referred to as the "NIR window" or "therapeutic window" for the lower and measurable light absorbance and the richness of information. In addition, the short-wave NIR region (700-1100 nm) is appropriate for in-line and in-situ field measurements using fiber-optic probes and relatively low-cost silicon detectors [20, 21].

Vibrational spectroscopy gives valuable information on the chemical composition based on functional groups detection and spectral analysis of the obtained fingerprints [22] that can be used to characterize normal and cancerous cells. Several studies revealed NIR region as a spectroscopic biomarker that has promising applications and a future design of specialized spectroscopic instrumentation. The NIR spectral changes of mammary gland tissues in the carcinogenic processes of rats were examined. It was found that intensities of DNA bands (1471 and 1911 nm) and water bands (967, 1154, 1402, and 1888 nm) were relatively increased while those of the lipid bands (1209, 1721, and 1764 nm) were reduced in the cancerous site [23]. The content of water in cancerous and normal prostate was analyzed using NIR spectroscopy (400-2400 nm). The water absorption peaks at 1444 and 1944 nm observed in prostate tissues which are related to OH overtone vibrational bands. It was shown that cancerous tissues contained less water than normal tissues [24]. The application of NIR spectroscopy for the detection of human primary pancreatic and colorectal cancers showed that major spectral differences were in the CH-stretching first (6000-5400 cm^{-1}) and second overtone (9000-7900 cm^{-1}) regions [25]. The most significant differences in NIR spectra from malignant and benign colorectal tissues were observed at the CH-stretching second overtone region and water's first overtone of combination bands (1100-1330 nm) with bands associated with glycoproteins, glycolipids, and carbohydrates as at OH and NH first overtone and CH first overtone combinations (1400-1600 nm and 1300-1420 nm) [26]. Normal and diseased breast tissues were examined by Raman spectroscopy. The spectra of

diseased breast tissue (benign and malignant) showed markedly diminished to absent contributions from lipids (at ~1082, 1302, 1444, and 1652 cm^{-1}) and reduced contributions from carotenoids (at ~1004, 1156, and 152 cm^{-1}) with an absence of the peak at ~1358 cm^{-1} [27]. The FTIR spectroscopy system analyzed prostate cancer cell lines. The ratio of peak intensities at 1030 and 1080 cm^{-1} used as a diagnostic marker to distinguish between the transformed normal cell and cell lines derived from various metastatic sites. This peak corresponds to the glycogen/phosphate ratio and is indicative of the metabolic turnover of the cell [28]. The application of laser trapping Raman spectroscopy to analyze leukemia cells was investigated. Raman markers associated with DNA (1093 and 785 cm^{-1}) and protein (1447 and 1126 cm^{-1}) vibrational modes have been found to exhibit excellent discriminating power for cancer cell identification [29]. Likewise, the Raman spectra at region 700-1750 cm^{-1} of malignant and bladder cancer cells denoted a superior concentration of proteins and nucleic acids in the bladder cells as in cancer cells. Proteins and nucleic acids were more abundant in MGH-U1 than PC-3 cells, while lipids and carbohydrates were more abundant in PC-3 cells [30]. Normal and four different types of human lung cultured cancer cells were identified by Raman spectroscopy. Strong bands at 748, 1129, and 1586 cm^{-1} were assigned to cytochrome c (cyt-c). The strong appearance of these bands suggested that the cancer cells were rich in cyt-c relative to phenylalanine, and possibly rich in mitochondrial cyt-c or mitochondria compared to normal cells [31]. Raman spectroscopic analysis was used to differentiate between normal breast (MCF-10A) and breast cancer (MCF-7 and MDA-MB-436) cell lines. The spectra of the cell lines showed spectral information about proteins, lipids, and nucleic acids. It was shown that the cancer cells presented high lipidic and proteinic information in 3050-2800 cm^{-1} and 1800-500 cm^{-1} regions [32]. Raman spectroscopy was used to evaluate the biomolecular cascade events related to the conversion of a normal cell into an invasive breast

cancer cell. It was observed that the lipid levels were increased for the invasive cells compared with normal ones in regions 2800-3000 cm^{-1} and 700-1800 cm^{-1} [33]. The FTIR spectra to study the development of skin cancer showed that the absorption band at approximately 3062 cm^{-1} was increased, indicating that most of the proteins had the configuration of amide B and the β-sheet protein structure predominated [34]. FTIR-ATR and FT-Raman absorption spectra were used to study the changes in oral cancer. The peak at 1238 cm^{-1} was correlated with nucleic acids symmetrical stretching, the phosphate wavenumbers were lower compared with that of normal tissue, and the shifting of 1030 cm^{-1} was ascribed to -CH_2OH vibrations [35]. FTIR spectroscopy technique was applied to explore the serum characteristics in breast cancer. It was found that the range of 3090- 3700 cm^{-1} is the criterion for differentiating breast cancer serum samples from the healthy ones, which can be attributed to protein modifications [36]. With SERS, the normal and cancerous liver tissue were analyzed in the fingerprint region (500-1800 cm^{-1}). The relative intensities of the characteristic vibration peaks at 838 cm^{-1} (amine groups), 1448 cm^{-1} (collagen), and 1585 cm^{-1} (protein and hemoglobin) are significantly changed in the cancerous tissues [37]. FTIR and Raman spectra were obtained from oral cancer cells to discriminate between normal, pre-cancerous, and cancerous conditions. Compared to normal patients, significant differences were observed at 1550, 1580, 1640, 2370, 2330, 2950-3000 and 3650-3750 cm^{-1} for FTIR and 520, 640, 785, 827, 850, 935, 1003, 1175, 1311 and 1606 cm^{-1} for Raman vibrations. The increase in DNA, protein, and lipid contents with malignancy was more evident [38].

The combination of analytical tools has been proven to be of immense value in a variety of bio-scientific applications. Spectroscopy can be coupled with chemometrics to offer a rapid and versatile method for the characterization of raw materials to advance cell culture performance [39]. Presently, artificial intelligence practices are widely applied in the field of computational biology

and bioinformatics [40, 41]. The absorption of molecules in the NIR region results mostly from the absorption of overtones and the merging of stretching-bending bands of atomic groups comprising hydrogen-bearing compounds and complexes that include CH, OH, and NH [42]. The overlap of signals detected in the NIR range increases the broadness of the peaks, resulting in composite spectra and hindering the designation of distinct features to specific compounds. The spectroscopic techniques coupled with multivariate data analysis in mammalian cell culture have been employed to derive data on several bioprocess variables, that include culture medium composition (nutrients and metabolites), viable cell concentration, and number of living cells [43-46].

Biomarkers have shown to be more favorable indicators, and the disparity in their contents is directly associated with specific diseases including cancer. Biomarkers can be specific cells, genes, gene products, hormones, or other molecules that are detectable in tissues or body fluids. However, in the case of early diagnosis, the concentrations of biomarkers are typically very low and difficult to determine using conventional techniques. Therefore, searching for more distinct biomarkers or improving the techniques of measurement remains the focus of research in the field of cancer detection and diagnosis. Given that a disease impairs the body functions at various levels of system organization, the collective impact of these induced variations in water can be measured by means of spectroscopic techniques. Aquaphotomics presents the spectral pattern of water molecular structure as a "molecular mirror" and a novel integrative biomarker [47]. This approach offers a new framework for elucidating variations in water molecular systems obtainable as a water spectral pattern [48, 49]. The rationale is to identify and expand the database of water absorption bands and decipher the distinctive water absorption patterns that can be utilized as biomarkers [50]. Information on water provides a potential diagnostic tool to improve the determination and

imaging of tumor cells based on variation in vibrational overtones of H_2O molecules in the NIR region. The molecular structure of H_2O is proposed as a novel fingerprint marker to signify changes in the NIR region for early detection of various types of tumor cells. So far, no studies can be found in a literature on NIR spectroscopy at short NIR region (700-1100 nm) and aquaphotomics of cultured cancerous cells.

In cell culture, spectroscopic technologies combined with multivariate data analysis can be used to obtain information regarding bioprocess variables in NIR spectra [51]. Generally, media formulations for mammalian cell culture procedures frequently apply phenol red (phenolsulphonephthalein) as a pH indicator, to visually and qualitatively monitor metabolism [52, 53]. Phenol red (PhR) has been utilized in different colorimetric applications, such as the pH assessment of freshwater [54], the diagnosis of filarial infection [55], the measurement of H_2O_2 generated by cells in culture [56] and CO_2 pressure in carbonated liquids [57], and aiding biopsy procedures to verify Helicobacter pylori-infected areas for patients diagnosed early with gastric cancer [58]. PhR also has several other biochemical applications [59-63]. The absorption spectra associated with PhR are differentiated by two absorption bands with two maxima at 438 nm and 559 nm at pH 7.4 [63].

To date, information on cancer cells using optical spectroscopy techniques is not extensively available, especially for qualitative evaluation of cells. This study aims to explore the interaction between light in the short NIR region (700-1100 nm) and cultured cancerous cells and identify the relationship between the absorption spectra and the characteristics of the cancerous cells. The focus of this research is to examine the accuracy of NIR spectroscopy to detect the concentration of different types of cultured cancerous cells and the potential to develop a spectral signature as a cancer biomarker using aquaphotomics. So far, the spectroscopic analysis of cancer cells in the

NIR region using phenol red dye has not been reported. Direct comparisons of the growth of three different cell lines were performed in the culture medium in the presence and devoid of phenol red. The cancerous cells utilized were cervical (HeLa) and prostate (DU145), while the mouse skin fibroblast (L929) cell line was used as a reference normal cell. The absorbance spectra for 50,000 to 275,000 cells of L929, DU145, and HeLa cell lines were measured. Principal component analysis (PCA) and soft independent modeling of class analogies (SIMCA) models were used for exploration and classification analyses. Multivariate analysis of NIR absorption spectra for mammalian cell cultures using phenol red in culture medium in addition to the spectral analysis of non-stained cell culture could be an effective and more viable substitute for biomedia characterization. Besides, aquaphotomics improves the understanding of cancer cell proliferation from the perspective of water molecular structure.

2. Materials and Methods

2.1 Cell Lines

The normal and carcinoma cell lines utilized in this study were procured from American Type Culture Collection (ATCC), Manassas, VA, USA. The cell lines include mouse skin fibroblast (L929), human cervix adenocarcinoma (HeLa), and human prostate carcinoma (DU145). Growth assays were prepared for all cell types in a serum-supplemented culture medium (Minimum Essential Medium 'MEM' and Dulbecco's Modified Eagle Medium 'DMEM'). The L929 cell was cultivated in the MEM, while HeLa and DU145 cells were cultivated in the DMEM supplemented with fetal bovine serum 'FBS' (10%), sodium pyruvate (1 mM), and penicillin (100 units/mL)/streptomycin (100 µg/mL). All growth assays were performed in MEM and DMEM with and without phenol red.

The growth assays were prepared from stock cultures harvested using trypan blue dye treatment. The preparation procedures of the subcultures of the cells are reported in our prior publications [64, 65]. The cell cultures were maintained at 37 ℃ in an incubator with 5% of CO_2 environment. The cultured cells were harvested from T25 tissue culture flasks and suspended in the culture medium. The cells were subsequently stained with trypan blue dye at a 1:1 ratio. Afterwards, cell viability test was carried out in parallel using the trypan blue dye exclusion technique by quantifying the number of viable cells in a hemocytometer. The trypan blue infiltrates the membrane of non-viable cells, which are then stained blue and can thus be differentiated from viable cells [66]. The cell concentration and total number of cells were calculated using Eqs.1 and 2 and presented in Tables 1 and 2.

$$C\ (cells/mL) = (viable\ counted\ cells/Q) \times Df \times Hf \qquad (1)$$

$$Total\ cell\ number\ (cells) = C \times V \qquad (2)$$

where C, Q, Df, Hf and V denote cell concentration, quadrant count (the number of squares is 4 in the haemocytometer and each square has a dimension of 1x1 mm with a 0.1 mm depth), dilution factor (which is typically 2), haemocytometer factor with value of 10^4 (also referred to a correction factor), and the sum of original volume of cell suspension, respectively. The cell lines were seeded in a 6-well culture plate with total cell numbers ranging from 50,000 to 275,000 cells per well. The total volume of culture medium in each well was 2 mL and the number of total cell seeding was derived using Eq. 3 below:

$$M_1V_1 = M_2V_2 \qquad (3)$$

where M_1, V_1, M_2 and V_2 represent the concentration of cell suspension, initial volume of cell suspension, concentration of seeded cells, and volume of seeded cells, respectively. The cells in the media were appended to the well and then incubated for an entire day. The total number of

cells per well, M_2, V_1, and V_2 are presented in Tables 3 and 4. The microscopic images (100x total magnification) of the cells were obtained using an inverted microscope (Olympus modelCKX41SF) in order to decipher the cell morphology and distribution.

Table 1. Number of viable cells, concentration of cell suspension, volume of cell suspension, and total number of cells for cell lines in culture medium with phenol red.

Cell line	Number of viable cells (cells)	Concentration of cell suspension (cells/mL)	Volume of cell suspension (mL)	Total cell number (cells)
L929	581	2 905 000	5	14 525 000
HeLa	436	2 180 000	5	10 900 000
DU145	325	1 625 000	5	8 125 000

Table 2. Number of viable cells, concentration of cell suspension, volume of cell suspension, and total number of cells for cell lines in culture medium without phenol red.

Cell line	Number of viable cells (cells)	Concentration of cell suspension, (cells/mL)	Volume of cell suspension (mL)	Total cell number (cells)
L929	528	2 640 000	6	15 840 000
HeLa	673	2 019 000	6	20 190 000
DU145	211	1 055 000	5	5 275 000

Table 3. Initial volume of cell suspension, concentration of cell seeding, volume of cell seeding, and total number of cells per well for L929, HeLa, and DU145 cells cultured in medium with phenol red.

Initial volume of cell suspension (mL), V_1			Concentration of cell seeding (cells/mL), M_2	Volume of cell seeding (mL), V_2	Total number of cells per well (cells)
L929	HeLa	DU145			
60.2	80.3	92.3	25 000	2	50 000
90.4	120.4	138.5	37 500	2	75 000
120.5	160.6	184.6	50 000	2	100 000
150.6	200.7	230.8	62 500	2	125 000
180.7	240.8	276.9	75 000	2	150 000
210.8	281.0	323.1	87 500	2	175 000
241.0	321.1	369.2	100 000	2	200 000
271.1	361.2	415.4	112 500	2	225 000
301.2	401.4	461.5	125 000	2	250 000
331.3	441.5	507.7	137 500	2	275 000

Table 4. Initial volume of cell suspension, concentration of cell seeding, volume of cell seeding, and total number of cells per well for L929, HeLa, and DU145 cells cultured in medium without phenol red.

Initial volume of cell suspension (mL), V_1			Concentration of cell seeding (cells/mL), M_2	Volume of cell seeding (mL), V_2	Total number of cells per well (cells)
L929	HeLa	DU145			
56.8	44.6	142.2	25 000	2	50 000
85.2	66.9	213.3	37 500	2	75 000
113.6	89.2	284.4	50 000	2	100 000
142.0	111.4	355.5	62 500	2	125 000
170.5	133.7	426.5	75 000	2	150 000
198.9	156.0	497.6	87 500	2	175 000
227.3	178.3	568.7	100 000	2	200 000
255.7	200.6	639.8	112 500	2	225 000
284.1	222.9	710.9	125 000	2	250 000
312.5	245.2	782.0	137 500	2	275 000

2.2 Spectra Acquisition

Vis-NIR spectroscopic technique was used to obtain absorbance spectra of each cell line cultured in a medium with and without PhR. The spectroscopy instrument used in the experiments is from Ocean Optics Inc. (Dunedin, Florida, USA). The spectrometer features and spectroscopic system are similar to those reported in our preceding studies [64, 65]. The measurement of the spectra of the cultured cell lines was achieved using a custom-prepared chamber. Three wells (34.8 mm diameter and 17.65 mm depth) were used from the 6-well cell culture plate. Both optical fiber cables (retrieving and illumination fibers) were connected with a collimating lens to ensure parallel transfer of light rays towards the well and from the sample to the spectrometer. The spot area of the light measured on the bottom of the well was 38.48 mm^2. The illumination light source was connected to a tungsten halogen light source, HL-2000, with a range of emission wavelengths between 360 and 2400 nm and a color temperature of 2800 K. The absorbance measurement was

calibrated using an empty well. The resultant light from the interaction with the cells in the well was transmitted through the well and collected using a retrieving fiber. The fiber was connected to a QE65000 spectrometer with a spectral sensitivity between 350.64 and 1131.24 nm. However, the entire analysis only utilized the wavelength range between 400 and 1100 nm. The acquisition parameters for the spectrometer including integration time, scan to average, and boxcar width were 17 ms, 8, and 3, respectively. Triplicate spectra were acquired for each batch to guarantee the reliability of the measurement and to obtain the average spectrum for multivariate analyses. Spectra Suite Software (Ocean Optics) was used to acquire and analyze the spectra.

2.3 Data Analysis

Principal Component Analysis (PCA) [67] was used as an exploratory analysis of the Vis-NIR spectral data. It is a method for reduction of data dimensionality by orthogonal matrix decomposition. The results of PCA are visualized using scores and loadings plots, where scores are projections of original spectra in the pattern spaces defined by principal components, while loadings show the weight coefficients of original variables. The principal components are orthogonal to each other and define spaces that capture maximum variation in the data. The scores plots allow easy observation of existing data patterns while loadings help analyze the spectral features and how they relate to the observed patterns. Soft independent modeling of class analogies (SIMCA) [68] is a supervised pattern recognition technique used for classification purposes. It is a PCA-based method, that employs PCA models for each of the previously assigned classes in data. Classification of spectra is based on a comparison of Mahalanobis distance, which is the distance between the spectrum and the centroid of each class. The spectra are classified as belonging to the class if the Mahalanobis distance is less than three standard deviations from the

class centroid [69]. Before SIMCA analysis, the spectral data were smoothed using Savitzky Golay 2nd order polynomial filter (21 points) [70], corrected for baseline effects using standard normal variate (SNV) transformation [71], and mean-centered [72]. Both PCA and SIMCA analyses were performed using commercially available multivariate analysis software, Pirouette (version 4.5, Infometrix, Bothell, WA).

3. Results and Discussion

CO_2 and lactate are generated via cellular respiration, which can increase the acidity level of the culture medium in due course, leading to a change in microenvironment of the cells, which is attributable to the small volume of the growth medium [73]. In addition, a rise in number of cells and cellular respiration increase the acidity levels of the cell cultures, leading to yellowing in colour of media formulations containing PhR. The cancer cells require energy metabolism for the stimulation of mitosis (i.e. cell division) and tumour growth, thus high rates of glycolysis are manifested, resulting in the generation of high amounts of lactate. This subsequently leads to increase level of acidity in the microenvironment [74-76] and theoretically explains the relatively faster change in the PhR colour for the cancer cells medium compared to the normal cell line. The comparatively higher intensity in the spectra of cancer cells could be due to variations in scattering, size and thickness of the cancer specimens, which influences the spectral absorbance due to the different depths of photon penetration [77]. It is evident that the absorbance of the samples in the NIR region is neither affected nor predisposed to the colour of the medium.

NIR spectroscopy has the benefit of ease and expediency when applied to aqueous samples [78]. The absorbance spectra in the NIR region (700-1050 nm) of the three cell lines with and without PhR are displayed in Figures 1 to 3. Each spectrum in Figures 1-3 was smoothed using the

Savitzky-Golay algorithm with 2nd order polynomial and a window length of 15 points in order to reduce the noise presents in the raw data. The absorbance values on the y-axis represent the total light attenuated through the sample as well as light scatterings. Each spectrum is a result of nine readings of measurements from three wells of the 6-well culture plates. The measurement of the

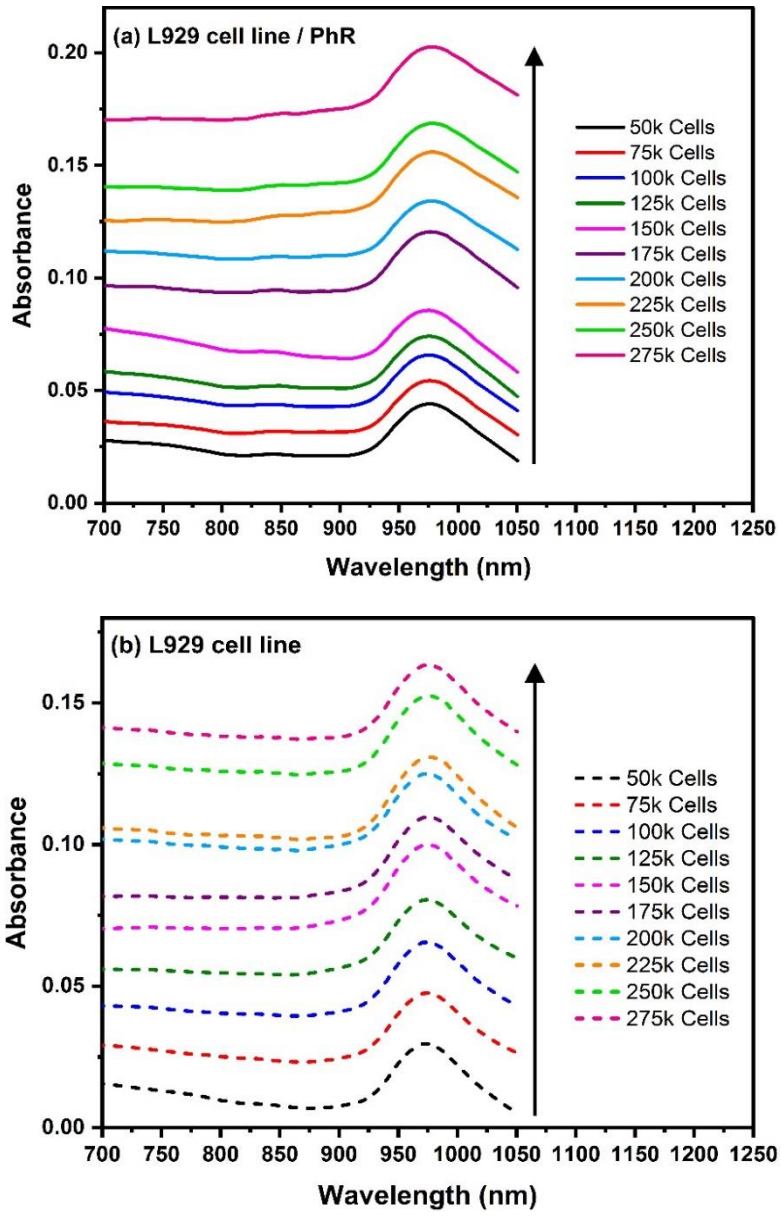


Figure 1. NIR absorbance spectra of L929 cells cultured in medium in: (a) the presence of phenol red and (b) the absence of phenol red.

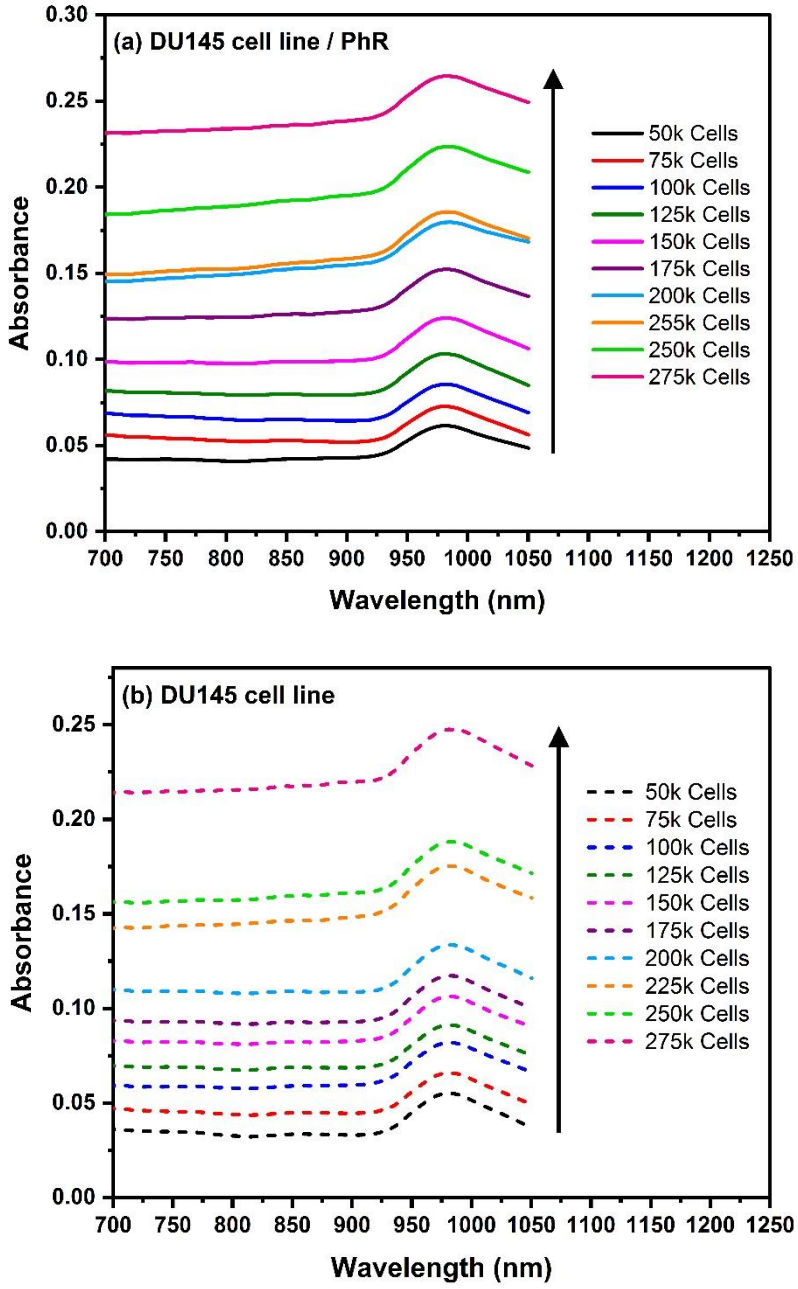


Figure 2. NIR absorbance spectra of DU145 cells cultured in medium in: (a) the presence of phenol red and (b) the absence of phenol red.

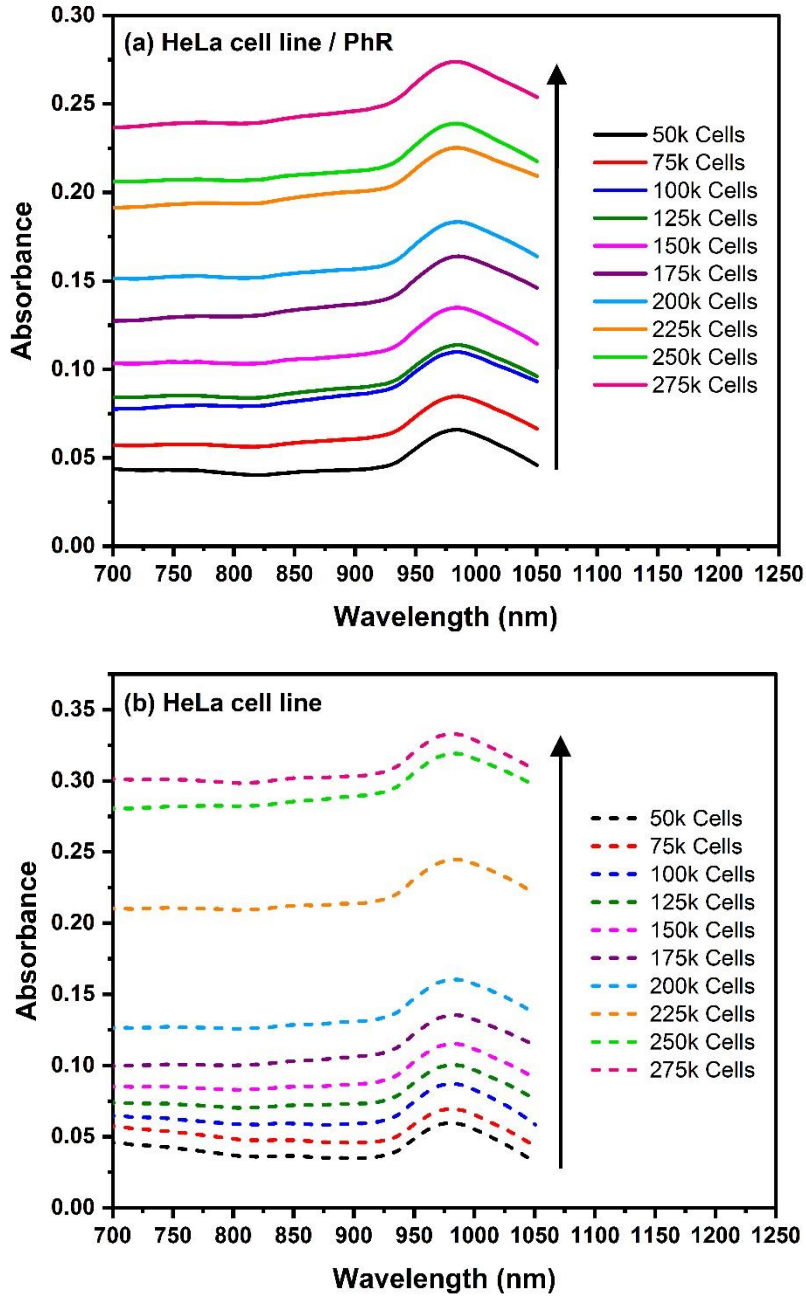


Figure 3. NIR absorbance spectra of HeLa cells cultured in medium in: (a) the presence of phenol red and (b) the absence of phenol red.

spectrum from each well was repeated three times to ensure the repeatability of the measurement and eliminate uncertainties due to sampling. Each spectrum in Figures 1 to 3 is the average of the spectra obtained from three wells of the 6-well plate. Light attenuation apparently increases with

the number of cells in the well. Furthermore, given that the culture medium is composed of water, the distinctive absorbance peaks observed at NIR wavelength (≈ 965-985 nm) denotes the water absorbance overtones [24]. The absorbance of cancerous cells slightly exceeds that of the normal cells for each increment in the number of cells. It is imperative to mention that the absorbance values generated by the spectrometer represent the loss of light due to both absorbance and scattering phenomena. Therefore, the observed absorbance can occur when light is scattered by a specific number of cells.

The morphology of cells is used as a basis for differentiating the normal and cancerous cells. Cancer cells are characterized by morphological variations as compared to normal cells. Morphologically, the cancerous cell is differentiated by several features that include a large nucleus, asymmetrical size and shape, prominent nucleoli, scarce cytoplasm and intense or pale colors [79]. The morphology of the cells was assessed prior to the spectroscopic measurements to guarantee the integral state of the cells without any biological impairment. The cells were microscopically examined to check for abnormal components and aberrations. The color of the cells was then inspected using PhR indicator. Microbial contamination is indicated by color alteration of PhR from plain red to cloudy pink. The morphologies of L929, HeLa, and DU145 cells prepared in culture media with PhR are displayed in Figure 4. The images were obtained using an inverted microscope with 100x total magnification and 0.25 numerical aperture of the 10x objective lens. The obviously higher scattering of cancerous cells compared to normal cells is attributed to their larger nucleus, variable sizes, and diverse shapes. The dimension and size of the nuclei as well as the change of the refractive index of cells have a specific effect on the light scattering of the entire cell [80]. In addition, the mitochondria and other small organelles in the cell also contribute to the light scattering [80]. The size of HeLa cells is larger than those of DU145

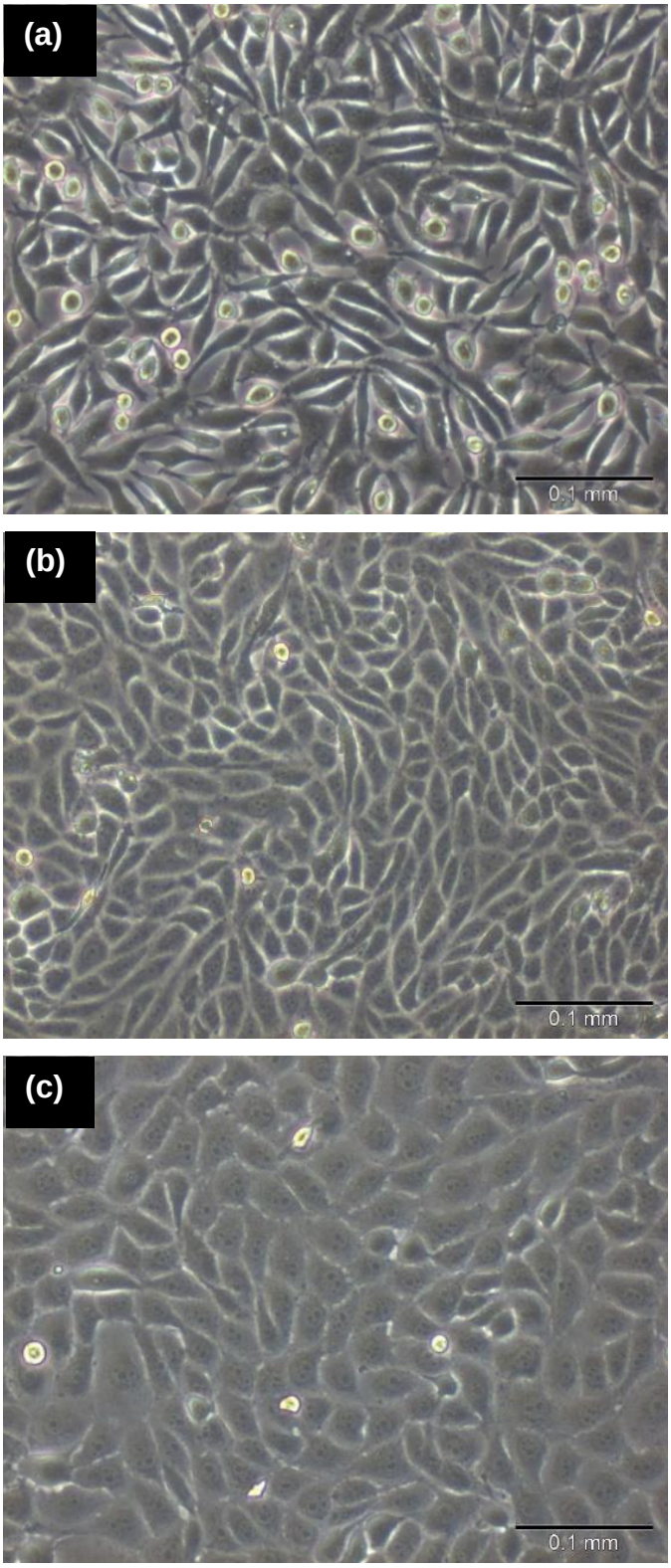


Figure 4. Morphology of (a) L929, (b) DU145, and (c) HeLa cell lines with 100x magnification of an inverted microscope.

and L929 cells, which plausibly explains their higher light scattering [64]. Additionally, increasing the number of cells enhances the probability of light scattering back into the detector, which results in an increase in the measured absorbance of the samples [81]. As observed in the images, the size of HeLa and DU145 cancerous cells is larger than those of L929 cells. The L929 cells are round and oval shaped with a standard size, while HeLa and DU145 have diverse sizes and shapes that are easily discernible at a low number of cells. It was shown that the proliferation or growth of the cancerous cells was slower than normal cells with DU145 cells being the slowest. In addition, it was observed that cell morphology changes as the number of cells increases.

The spectral changes are reflected by shift and broadening/narrowing of distinct water absorption peak that emerge in the NIR region at 975 nm. As observed in Figures 1-3, the characteristic peak of water (≈ 965-985 nm) for cancerous cells is broader compared to normal cells. This peak is attributed to the blend of the second harmonic of O-H symmetric stretch vibration and the intrinsic anti-symmetric stretch vibration from hydrogen bound O-H. In addition, the attributes of this peak are sensitive indicators of the microenvironment of H_2O molecules [82]. The specificity and precision of the shape and position of NIR peak are highly responsive to the molecular character of H_2O [83]. It is well-known that the fraction of adsorbed or bound H_2O measured using optical techniques can relay the molecular vibrational states of H_2O that are related to macro molecular complexes in tissues [84], and may provide further understanding as regards tissue pathophysiology. As a result of the increased volume of H_2O that bind to macromolecules such as proteins, the characteristic absorption peak of H_2O at 970 nm is subjected to both shifting and broadening. The spectral characteristics of H_2O absorption are dependent on the binding state of H_2O. As the volume of hydrogen-bound H_2O molecules decreases, the intensity of water absorption peak increases, the bandwidth becomes narrower, and the peak is shifted to higher

energy (blue-shift). These changes are due to imbalance in the symmetry between hydrogen-bound and free H_2O molecules. Hydrogen bonding between H_2O and macromolecules such as proteins initiates additional spectral broadening and red shifting of the 970 nm water peak [83]. Also, the symmetry of the band increases as the number of cells increases, with a shift in the maximum absorption (Figures 1 to 3), which is possibly due to the increased number of free O-H. This explains the eventual increase in absorbance [85] as the number of cells increases.

The spectra in the range of 400 to 1100 nm were investigated using PCA analysis, with the edges of the spectra trimmed to prevent noise. The samples were designated with "Y" and "N", for instance, L929 Y-1 and L929 N-1 denote the L929 cells with and without PhR for 50,000 cells, while the samples labelled as L929 Y-1 to L929 Y-10 signify the L929 cells for 50,000 to 275,000 cells. Results of PCA analysis of the spectra for L929, HeLa, and DU145 cell lines are shown in Figure 5. The PCA scores plots reveal the hidden structures in spectral data. The first five principal components (PC1 to PC5) explained over 99.99% of the variance in data and their loadings are illustrated in Figure 6. The first principal component (PC1) indicating the highest variance (88.69%) is related to the overall number of cells, where the scores move in the direction of the PC1 axis toward the positive part as the number of cells increases (Figure 5(a)). In the same space, three distinct groups are observable along the PC2 axis. In addition, Figure 5(a) shows the PC1 versus PC2 scores plot which reveals separation along the PC2 axis between the cells stained with PhR (located in the negative part of the PC2 axis) and non-stained cells (located in the positive part of PC2 axis). In the cluster of scores corresponding to PhR stained cells, another separation along the PC2 axis can be observed between cancer and normal cells. It can be seen that the scores of cancer cells are located close to zero, while scores of normal cells are further below the negative part of PC2. This kind of sub-separation is lacking in the scores of non-stained cells. Figure 5(b)

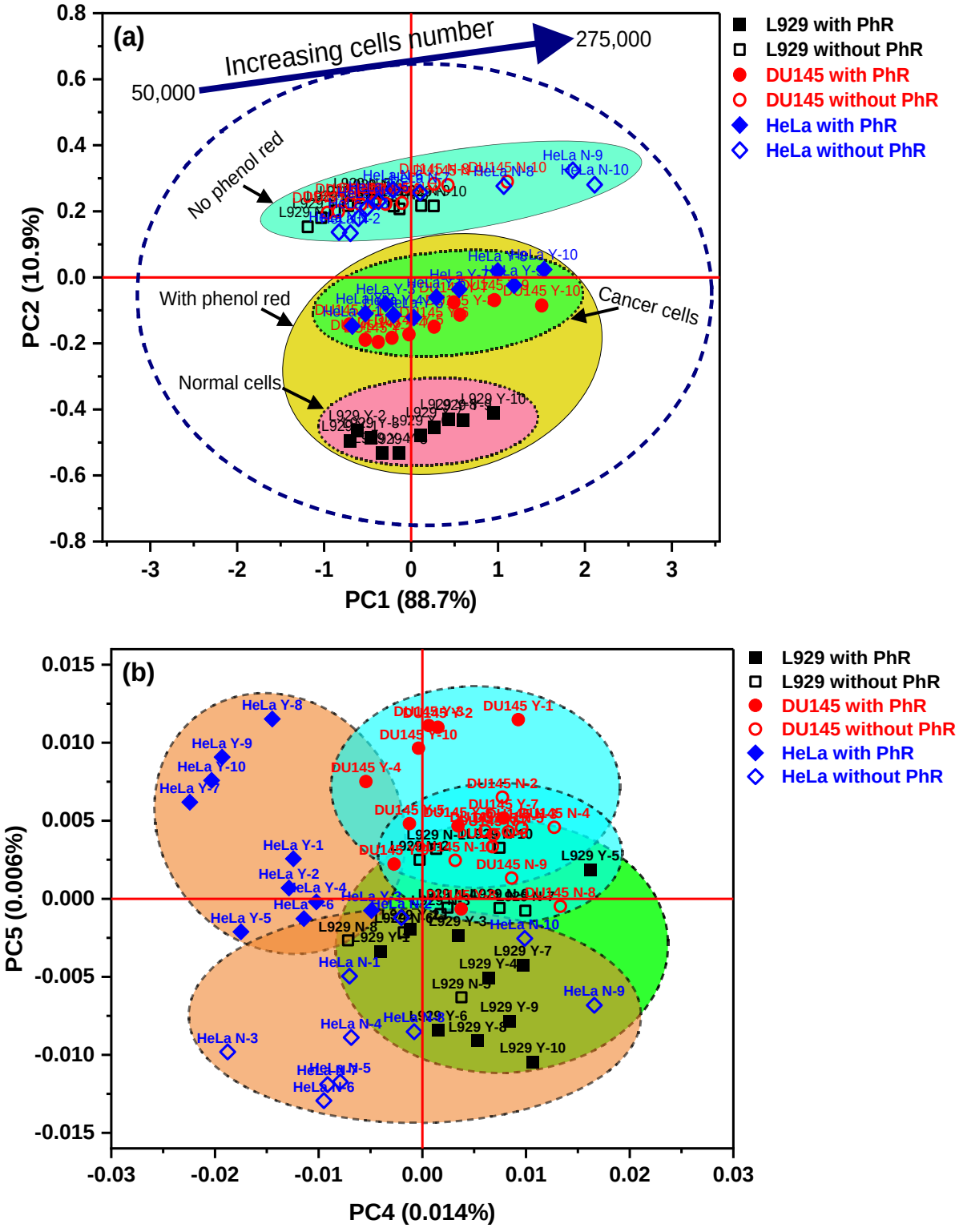


Figure 5. Scores plots of PCA analysis of Vis-NIR raw spectra (400-1100 nm) for L929, DU145, and HeLa cell lines: (a) PC1 versus PC2 and (b) PC4 versus PC5.

presents PC4 versus PC5 scores plot which further reveals the separation of scores between PhR stained and non-stained cancer cells along the PC5 axis. PhR stained HeLa cells scores are located in the positive part of PC5 and negative part of PC4, while scores of HeLa cells without PhR are located entirely in the negative part of PC5. Similarly, both scores of PhR stained DU145 and PhR non-stained DU145 located in the positive part of PC4 and PC5 axes, show some separation along the PC5 axis. The scores of the control cells L929 form one group located mostly in the positive part of the PC4 axis, without distinction between the scores belonging to stained and non-stained classes.

The loadings of the first two principal components are shown in Figures 6(a) and (b). The predominant peak for both loadings is located at 558 nm, which denotes the basic form (deprotonated base form) of phenol red [86, 87]. The peak is positive in PC1, indicating that the absorbance at this wavelength increases concomitantly with an increase in the number of cells, which conforms to the observed trend of scores towards the positive part of the PC1 axis. Besides the colour peak, it is discernible in the NIR part of the spectra that the shape of the PC1 is similar to the baseline of the spectra, with the graph being dominated by an offset and slope. This is consistent with the preliminary findings presented in Figures 1 to 3, where the number of cells is directly associated with light scattering and its effects on the baseline. It can be deduced that the interpretation of this factor is dependent on the number, shape, and size of the cells. Nonetheless, regardless of the principal features in the PC1 being derived from the physical attributes of the cells, numerous small spectral features can also be observed. These spectral features can be attributed to the absorbance bands of H_2O rather than noise.

As aforementioned, the cells and cell medium are enriched with H_2O, which is a strong absorber of NIR light. Specifically, the broad bands observed at approximately 740 nm, 850 nm,

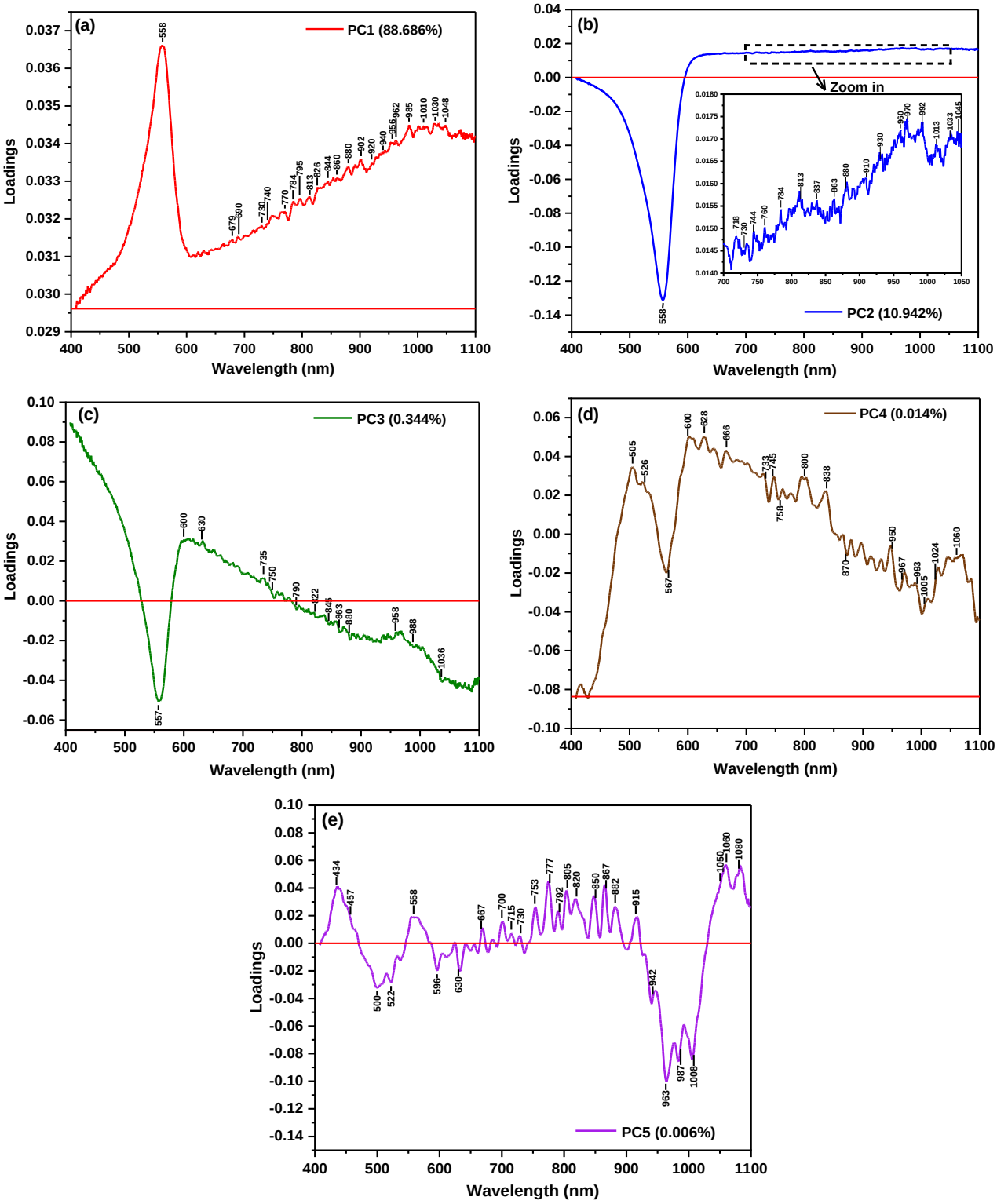


Figure 6. (a)-(e) PCA loadings plots of the calculated PC1 to PC5 at region 400-1100 nm for L929, DU145, and HeLa cell lines. Inset related to PC2 is zoom in of loadings in the NIR region (700 to 1050 nm).

and 970 nm can be assigned to the third overtone of OH stretching vibration $(3\nu_1+\nu_3)$, the second overtone of the combination of stretching and bending vibration $(2\nu_1+\nu_2+\nu_3)$, and the second overtone of OH stretching vibration $(2\nu_1+\nu_3)$ [88-91]. The slight variations in absorbances of the peaks (Figure 6) seem to be small compared to the peak intensity in the visible region, although this is simply as a result of the disparity in the source of the peaks, i.e., electronic and vibrational transitions. The emergence of numerous peaks confirms the complex nature of the molecular structure of H_2O in the cells and cell cultures which becomes altered in the course of cell proliferation. Likewise, PC2 is dominated by the peak denoting PhR at 558 nm, which is negative, while the NIR part shows only baseline offset resulting from light scattering. The features of this component can elucidate the characteristics of the three distinct groups observed at Figure 6(a). It is evident that the scores positioned in the positive part of PC2 axis are related to non-stained cell cultures (without PhR), which for all three cells lines, form a compact, well-defined cluster. The scores matching the PhR stained cell cultures are present in the negative part of the PC2 axis (as indicated by the negative peak at 558 nm in PC2). Remarkably, two subgroups can be observed within relatively large cluster, corresponding to the cell types, where the scores of the normal cell line are located in the lowest portion of the negative part of PC2 axis. An overview of the scores plots in Figure 5(a) and PC2 in Figure 6(b) shows that this component differentiated PhR stained and non-stained cultures. This differentiation is based on the peak denoting PhR in the visible region, while in the NIR part, the difference is indicated in both the baseline offset between PhR stained and non-stained groups (with higher offset for the non-stained cultures) and in the magnified part of Figure 6(b) in the several absorbance bands of water, particularly the area of the 2^{nd} overtone of water ($\sim$ 970 nm). It can be inferred that both water content and molecular structure of the cell cultures (media and cells included) are affected by staining. PhR staining of the cells

changes the molecular structure of water differently for the cancer cells and normal cells. Thus, it is imperative to further elucidate the relationship between pH and the molecular structure of water in cancer and normal cells. The PC4 and PC5 loadings are presented in Figure 6(d) and (e), respectively. The loadings show an increased variety of spectral features, with lots of distinct peaks, particularly the PC5 in the NIR region. The region of 800-910 nm is complex to interpret because of the several overlapping bands denoting diverse functional groups. A number of the bands observed in the loadings are recognized water absorbance bands and consistent with the ones aforementioned in the analysis.

PCA determined the natural clusters in the spectra based on the cell types and PhR staining. For additional examination of these differences, SIMCA discrimination analysis was performed to differentiate the cell lines according to PhR treatment and cell type. The raw spectra of the cell lines were pre-processed in order to improve the discrimination analysis using SIMCA. The pre-processing methods were Savitzky-Golay with 2^{nd} order polynomial filter and 21 data points to remove noise, standard normal variate (SNV) transformation to remove baseline effects and mean-centering. Figure 7(a) depicts the pre-processed spectra of the three cell lines, coloured based on the status of PhR treatment regardless of cell type. In this figure, the spectra represent the relation between wavelength and absorbance at the number of cells from 50,000 to 275,000 for each cell line cultured in the medium with and without PhR. As can be seen in Figure 7(a), there is a distinctive difference between the spectra of the cell lines with and without PhR at the absorbance band located at 556 nm (corresponding to the absorption of the basic form of PhR) and in the spectral region located around 970 nm (corresponding to 2^{nd} overtone of water) where several distinctive bands could be observed. In Figure 7(a), the band of H_2O has equivalent implications for the differentiation between the normal and cancerous cells in the presence and absence of PhR

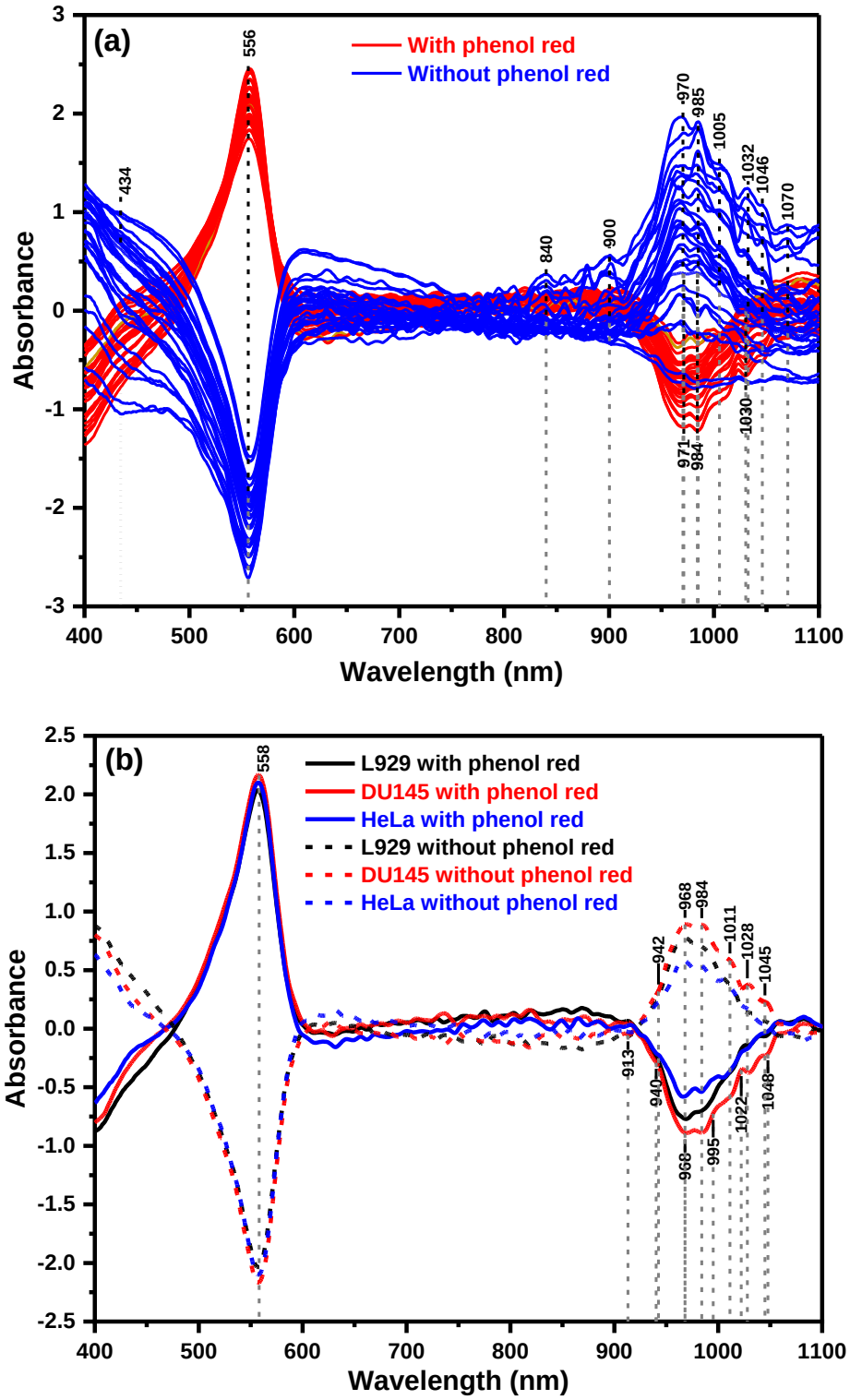


Figure 7. (a) Pre-processed spectra and (b) average of the pre-processed spectra for L929, DU145, and HeLa cell lines cultured in medium with and without phenol red at the number of cells ranging from 50,000 to 275,000.

compared to the peak of PhR at 556 nm. Figure 7(b) displays the averaged spectra for each cell line presented in Figure 7(a) in case of the presence and absence of PhR which means each cell line has two averaged spectra one of them in the presence of phenol red (solid line) and the other one in the absence of phenol red (dashed line). Evidently, there is a common absorbance band in the region of the 2nd overtone of water, irrespectively of the presence or absence of PhR, suggesting that the main spectral features characteristic for each cell line are inherent in this spectral region (NIR region) and do not depend on PhR. It is interesting that the variations of the spectra decrease with adding PhR, which is consistent with the fact that PhR is a perturber for the water molecular system which unifies the samples along its influence leaving only the differences depending on the type of the samples.

The PhR treatment appears to lead to inhibited dissimilarities between the spectra of individual samples and thus the cell types. The presence of more inter-class spectral variations is observable in the samples prepared in the absence of PhR, particularly in the region around 985 nm. The variations also exist in the visible region, around 434 nm and 558 nm, which can be credited to the acid (yellow) and basic (red) forms of PhR [86], respectively. Thus, it can be deduced that PhR influences the molecular structure of H_2O in the samples, either through interfacing with the medium, cells, or both. The averaging of the spectra shown in Figure 7(b) for each cell type, clearly indicates that the average differences in cell type cultures is more evident in the NIR region, and same absorbance bands are affected by the PhR regardless of the cell type, but with dissimilar spectral intensities. These bands are positioned at 940-942 nm, 968 nm, 984 nm, 995 nm, 1011 nm, 1022 nm, 1028 nm, and 1045-1048 nm (Table 5). The band positions, being in the second overtone of H_2O, when recalculated to the 1st overtone, generate the following bands: 1410-1413 nm, 1452 nm, 1476 nm, 1492.5 nm, 1516.5 nm, 1533 nm, 1542 nm, and 1567.5-1572

nm. The first five bands are assigned to the well-defined H_2O matrix coordinates of aquaphotomics. C5, C8, C9, C10, C11, and C12 are assigned to free molecules S_0, H_2O hydration shell, H_2O molecules with two hydrogen bonds, H_2O molecules with three hydrogen bonds S_3, H_2O molecules with four hydrogen bonds S_4, and strongly bound water [92], respectively.

Table 5 shows that the assignment of bands in the region of 995-1030 nm is a difficult procedure because of the complex features of the NIR spectra of both proteins and H_2O. Nonetheless, since the bands in this case are associated with the differences between PhR stained and non-stained cultures, it is rational to assume the assignment of these bands to H_2O involved with proton hydration (also referred to as aqueous protons, which are different H_2O clusters with dissimilar number of H_2O molecules related to hydration). The fact that diverse spectra also exhibit distinct bands of protonated and deprotonated forms of PhR in the Vis region confirms that bands discernible in NIR region explain the same phenomenon i.e., also act as an optical pH indicator of the changes in hydrogen ions (H^+) or hydronium ions (H_3O^+).

The results of SIMCA discriminating analysis (5% significance) performed on the pre-processed spectra in Figure 7(a) are presented in Figure 8. Cooman's plot shown in Figure 8(a) reveals the results of SIMCA discrimination of the spectra of L929, DU145, and HeLa with and without PhR staining. Ellipses drawn on this graph are for presentation purposes emphasizing the complete separation between different cell types in the multivariate space. The yellow ellipse includes the scores of the three cell lines with the different numbers of cells (from 50,000 to 275,000) and cultured in the medium including phenol red dye. This ellipse showed that the use of PhR during culture decreases the interclass distance. The achieved accuracy of classification with SIMCA was 100%, with a confidence interval of 0.95. Table 6 exhibits the number of components (PCs), explained variance (%), and the actual class membership and predicted

Table 5. Tentative assignments of the absorbance bands found to be important for discrimination between the cell cultures with and without phenol red in 2nd overtone of water, and the corresponding bands in the 1st overtone region, where first five of them belong to well-established aquaphotomics 12 Water Matrix Coordinates 'WAMACs' (C1-C12) [92], together with the alternative possible assignments.

Wavelength (nm)	Tentative band assignment	Ref.
940-942	Free water molecules, S_0 (1410-1413 nm, C5 WAMACs)	[92]
	Water vapour (940 nm)	[93-95]
	Water vapour (942 nm)	[96]
968	Hydration shell OH-$(H_2O)_{4,5}$ (1452 nm, C8 WAMACs)	[92]
	Hydroxide absorption peak (967 nm)	[97]
	O–H stretch, 2nd overtone of ROH and H_2O (970 nm)	[42]
	O–H stretch, 2nd overtone of H_2O (960–970 nm)	[89]
984	Water molecules with 3 hydrogen bonds, S_3 (1476 nm, C10 WAMACs)	[92]
	N–H stretch, 2nd overtone of CONHRn (secondary amides) (981 nm)	[89]
995	Water molecules with 4 hydrogen bonds (1492.5 nm, C11 WAMACs)	[92]
	N-H stretch, 2nd overtone (aromatic amines) (995 nm)	[89]
	N-H stretch, 2nd overtone (amines, general) (1000 nm)	[89]
1011	Strongly bound water (1516.5 nm, C12 WAMACs)	[92]
	OH from secondary (1004 nm) and tertiary alcohols (1006 nm)	[89]
	N-H stretch, 2nd overtone (amines, general) (1015 nm)	[89]
	2nd overtone Superoxide Tetrahydrate $O_{2-}.(H_2O)_4$ (1010 nm)	[98]
1022	2× N–H stretch + 2× amide I (protein) (1020 nm)	
	N–H stretch, 2nd overtone (protein) (1007 nm)	[89]
	N–H stretch, combination of RNH2 (primary amine) (1020 nm)	
	Aqueous proton $[H^+.(H_2O)_6]$ - H_2O in $H_5O_2^+$ symmetric stretch, 2nd overt (1018 nm)	[99]
	2nd overtone intramolecular hydrogen bond stretches in (OH-$(H_2O)_4$) (1020 nm)	[100]
1028	N-H stretch, 2nd overtone of RNH2 (1030 nm)	[89]
	Aqueous proton $[H^+\cdot(H_2O)_3]$ - H_3O^+ symmetric stretch, 3rd overt (1033 nm)	[99]
1045-48	N-H stretch, 2nd overtone of amides (primary, bonded; two bonds) (~1050 nm)	
	N-H stretch, combination of CONH (primary amide) (1047 nm)	[89]
	N-H stretch, 3rd overtone of amides (secondary, bonded; one band) (1000-1067 nm)	
	$H_{11}O_5^+$ H-bonded OH stretch, 2nd overt (1041 nm)	[101]
	Aqueous proton $[H^+\cdot(H_2O)_6]$ - H_2O in $H_5O_2^+$ symmetric stretch, 2nd overt (1055 nm)	[99]

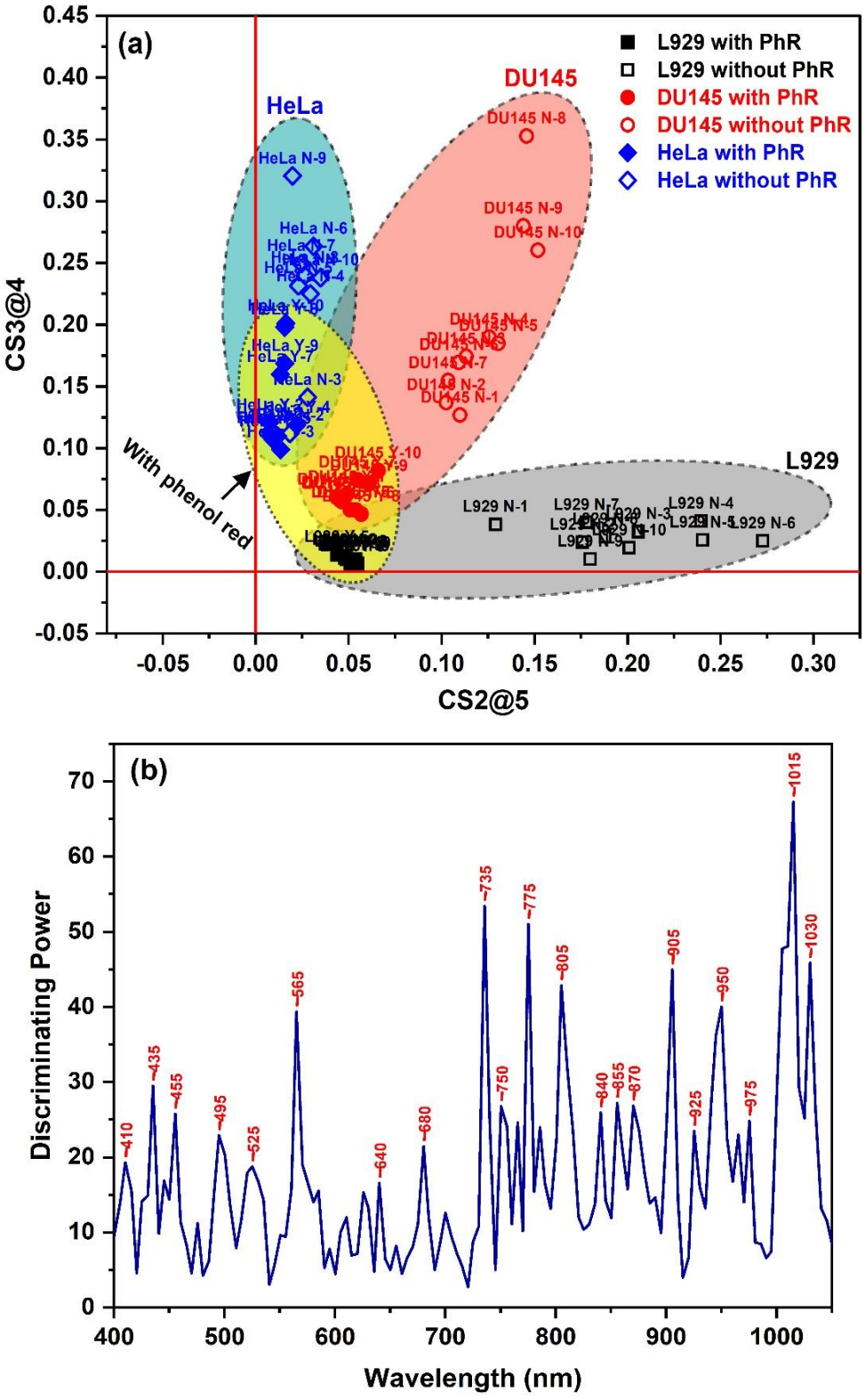


Figure 8. SIMCA classification analysis: (a) Cooman's plot and (b) discriminating power between L929, DU145, and HeLa cell lines.

Table 6. SIMCA classification results for L929, DU145, and HeLa cell lines.

PCs	Explained variance (%)			Actual Class			Correct classification (%)	Misclassifications (%)
				L929	DU145	HeLa		
5	99.93	SIMCA Class	L929	**20**	0	0	100	0
5	99.89		DU145	0	**20**	0	100	0
4	99.86		HeLa	0	0	**20**	100	0

membership using the SIMCA model which clearly reveals no misclassifications. Mahalanobis distances (interclass distance) of 4.07, 6.93, and 5.32 were obtained between HeLa and DU145, HeLa and L929, and DU145 and L929, respectively. The interclass distances were > 3.0, indicating that the separation is reliable. As observed in Cooman's plot, the PhR stained cells are positioned closer together (smaller Mahalanobis distance) than the non- stained cells, which further confirms the importance of preserving the cells in their initial state for effective discrimination. Figure 8(b) displays the discriminating power of the SIMCA model that indicates the absorbance bands with the highest influence for successful discrimination between cell types. The discrimination power shows that the absorbance bands with the highest contribution in the model are located in the NIR region, and they include 735 nm, 775 nm, 805 nm, 905 nm, 950 nm, 1015 nm, and 1030 nm. In the visible region, the bands at 435 nm and 565 nm are assigned to the highest power. The assignments of the absorbance bands are presented in Table 7. These bands may be attributed to hydrogen-bonded H_2O and H_2O interfacing with carbohydrates or proteins, or these compounds directly. Six of these bands can be related to hydrogen-bonded H_2O, which further confirms the significance of using the molecular structure of H_2O to distinguish between the three cell types. Nevertheless, as reported in prior research, the extracellular metabolites are main contributors to

Table 7. Tentative assignments of the absorbance bands found in discriminating power of SIMCA model developed for discrimination between different cell types. The assignments were provided for the bands found directly in 2nd and 3rd overtones of water, or recalculated to 1st overtone WAMACS bands, together with the alternative possible assignments.

Wavelength (nm)	Tentative band assignment	Ref.
735	Hydrogen-bonded -OH, 3rd overt	[102]
	Water molecules with 3 hydrogen bonds, S_3 (1470 nm, C10 WAMACS)	[92]
775	Hydrogen bonded water (770 nm)	[103]
	N–H stretch, 3rd overtone of RNH_2 (779 nm)	[42]
	N–H stretch, 3rd overtone of amides (secondary, bonded; one band) (750-800 nm)	[89]
805	Maximum absorption peak for ice (800 nm)	[96]
905	C-H stretch, 3rd overtone of protein (910 nm)	[104, 105]
	3rd overtone of a C-H stretching mode and the 2nd overtone of an N-H stretching mode of protein (906 nm)	[106]
	Hydration shell OH-$(H_2O)_{1,2,4}$ (1357.5 nm, C2 WAMACs)	[92]
950	O-H stretch (954 nm), 2nd overtone of ROH (alcohols) (940-970 nm)	[107]
	Water interacting with sugar (950 nm)	[108]
	Hydration water (1425 nm, C6 WAMACS)	[92]
	2nd overtone of O-H stretching of water interacting with protein (950-960 nm)	[106]
1015	H_2O deionized; 2nd overt (1010.1 nm)	[89]
	Hydrated proton $[H+(H_2O)_6]$ - H_2O symmetric stretch, 2nd overt (1018.1 nm)	[99]
	2X N-H stretch + 2X amide I (protein) (1020 nm)	
	N-H stretch, 2nd overtone (protein) (1007 nm)	[89]
	N-H stretch, combination of RNH2 (primary amine) (1020 nm)	
	Protein (1020 nm)	[109]
	Strongly bound water (1522.5 nm, C12 WAMACs)	[92]
1030	N-H stretch, 2nd overtone of RNH_2	[89]
	3rd overtone of a C-H stretching mode and the 2nd overtone of an N-H stretching mode of protein (1032 nm)	[106]

the features of the NIR spectral signal [110], which implies that the differentiation of cell types is not based on the physical attributes of cells (whose effect is drastically reduced by pre-processing)

but according to the metabolites which are specific for each cell type because of their metabolic variations. The metabolites in turn affect the hydrogen bond complex of extra- and intracellular water leading to the characteristic spectral signature in the NIR region. Employment of vibrational spectroscopy for a variety of cancer types has been reported marker bands related to water, proteins, lipids, amino acids, carbohydrates, and their combination that are useful for cancer diagnosis. The main absorbers of NIR light in blood-perfused tissues are oxyhemoglobin, deoxyhemoglobin, and water [111], while in the cell culture, the major absorbers are various water molecular species and other components are presented in considerably less quantities. On the other hand, there are numerous reports on the diagnosis of other diseases that showed the diagnostic value of water absorbance bands [25, 38, 112-121]. In some reports, there is a possibility that the discovered bands are also assigned as the water vibrational bands. For example, the bands 1471 nm and 1911 nm that are attributed to DNA of rat mammary gland tumor [23], whereas there is a very likely possibility these bands originate from water [92, 119].

In our findings, seven absorbance bands were found to be most influential for discrimination between the L929, DU145, and HeLa cell lines. Five of these bands can be attributed to water, specifically to hydrogen-bonded water with different numbers of hydrogen bonds or strength, and to the water interacting with either sugar or proteins. A study performed by Ali et al. [24] suggested water bands can be used for the diagnosis of human prostate cancer. It was reported that the 1st overtone of water located around 1450 nm, and other water overtone modes provided key spectroscopic fingerprints for cancer in prostate tissue, citing as the most probable reason the higher-order phases of water and their interaction with cellular or other macromolecules in prostate tissues. Gosh et al. [38] used FTIR spectroscopy of exfoliated cells that revealed five distinctive spectral features attributable to hydrogen-bonded water (3 peaks at 3500-3700, 3700-3800, and

3800-4000 cm^{-1}) and non-hydrogen bonded OH stretch of water (2 peaks at 3300-3400 and 3400-3500 cm^{-1}) that showed changes in intensity in patients with oral cancer compared to healthy volunteers. Their findings are in a good agreement with current work, where within 3rd and 2nd overtone of water, water absorbance bands attributed to different phases of water and water interacting with other macromolecules were found.

It can be deduced that the water molecular structure plays a significant role in differentiating between cancer and healthy cells which is revealed from the current study where the cells of cancers from different organs showed spectral differences. To the authors' knowledge, there is a study using NIRS that showed spectral changes between different cancers (colorectal and pancreatic cancer) [25]. On the other hand, there are studies on aquaphotomics showing an excellent possibility of differentiating the cell types in the case of bacteria [122-124], different types of animal tissues [125], and even normal and abnormal cell development [90]. This study reported for the first time the differences between normal fibroblast cells and two types of cancer – prostate and cervical cells, which demonstrated the discrimination between the cell types is mostly based on the water molecular structure in the NIR short region. This is a very important aspect of cancer research, that can help increase the specificity of cancer diagnosis by enabling differentiation among tumors [25]. Nonetheless, additional research is needed to precisely assign bands to functional groups and relate them to the cell type metabolism and accurately deduce the basis for cell differentiation. An appealing feature of the current findings is that the observed bands are crucial for the discrimination of cells, and relevant for a variety of diagnostic purposes in other biological systems, such as the level of tissues. This offers a basis for advanced studies that explore the application of aquaphotomics NIR spectroscopy to investigate early detection of cancer using cells, tissues, or body fluids.

4. Conclusion

This work presents the application of NIR absorbance spectroscopy and aquaphotomics to classify cultured cancerous cells including HeLa and DU145 cells. The results show that the combined application of absorption spectra and principal component analysis (PCA) improves the selectivity for the precise classification of cancer cells. The PCA achieved a clear differentiation between the three types of cell lines, with subgroups detected based on the presence or absence of the phenol red (PhR) in the culture. SIMCA method achieved a high accuracy of classification, where larger class distances were obtained between the non-stained cells compared to stained cells. Aquaphotomics enhanced the assignment of cancer cell proliferation from the perspective of water molecular structure. The absorbance bands that successfully aided the cancer cell discrimination include a range of H_2O molecular species-protons (vapor) and strongly bonded H_2O, as well as CH and NH bands of proteins. Future studies can further explore the roles of these bands in differentiating between normal and cancer cell types.

This study is a novel approach to the advancement of an online noninvasive spectroscopic system capable of monitoring ideal growth conditions and maximizing cellular productivity. In addition, the integrative spectroscopic biomarker results in a future design of specialized spectroscopic instrumentation for cancer diagnosis. Furthermore, this work unlocks possibilities for at-line sensor development for discriminating mammalian cells using PhR in a culture medium. It also offers certain insights into the influence of PhR on cell dynamics and expands the application of the NIR spectral region to distinguishing between cell types.

Data Availability

All data used to support the findings of this study are available from the corresponding author (A.F.O.) upon reasonable request.

Acknowledgment

The authors gratefully acknowledge the financial support of this research by the Malaysian Ministry of Higher Education–Fundamental Research Grant Scheme (Grant No. 203.PFIZIK.6711897).

References

1. Butler, L., et al., *Lipids and cancer: Emerging roles in pathogenesis, diagnosis and therapeutic intervention.* Advanced drug delivery reviews, 2020.

2. Sorvina, A., et al., *Lipid profiles of prostate cancer cells.* Oncotarget, 2018. **9**(85): p. 35541.

3. Frangioni, J.V., *New technologies for human cancer imaging.* Journal of clinical oncology, 2008. **26**(24): p. 4012.

4. Kondepati, V.R., H.M. Heise, and J. Backhaus, *Recent applications of near-infrared spectroscopy in cancer diagnosis and therapy.* Analytical and bioanalytical chemistry, 2008. **390**(1): p. 125-139.

5. El-Tawil, S.G., et al., *Comparative study between Pap smear cytology and FTIR spectroscopy: a new tool for screening for cervical cancer.* Pathology, 2008. **40**(6): p. 600-603.

6. Wilson, R.H., et al., *Tissue classification using optical spectroscopy accurately differentiates cancer and chronic pancreatitis.* Pancreas, 2017. **46**(2): p. 244.

7. Yang, P.-W., et al., *Visible-absorption spectroscopy as a biomarker to predict treatment response and prognosis of surgically resected esophageal cancer.* Scientific reports, 2016. **6**(1): p. 1-12.

8. Dai, W.Y., S. Lee, and Y.C. Hsu, *Discrimination between oral cancer and healthy cells based on the adenine signature detected by using Raman spectroscopy.* Journal of Raman Spectroscopy, 2018. **49**(2): p. 336-342.

9. Shang, L.-W., et al., *Fluorescence imaging and Raman spectroscopy applied for the accurate diagnosis of breast cancer with deep learning algorithms.* Biomedical Optics Express, 2020. **11**(7): p. 3673-3683.

10. Noothalapati, H., K. Iwasaki, and T. Yamamoto, *Non-invasive diagnosis of colorectal cancer by Raman spectroscopy: Recent developments in liquid biopsy and endoscopy approaches.* Spectrochimica Acta Part A: Molecular and Biomolecular Spectroscopy, 2021. **258**: p. 119818.

11. Del Mistro, G., et al., *Surface-enhanced Raman spectroscopy of urine for prostate cancer detection: a preliminary study.* Analytical and Bioanalytical Chemistry, 2015. **407**(12): p. 3271-3275.

12. Vendrell, M., et al., *Surface-enhanced Raman scattering in cancer detection and imaging.* Trends in biotechnology, 2013. **31**(4): p. 249-257.

13. Wang, G., et al., *Detection of the potential pancreatic cancer marker MUC4 in serum using surface-enhanced Raman scattering.* Analytical chemistry, 2011. **83**(7): p. 2554-2561.

14. Lee, S., et al., *Rapid and sensitive phenotypic marker detection on breast cancer cells using surface-enhanced Raman scattering (SERS) imaging.* Biosensors and Bioelectronics, 2014. **51**: p. 238-243.

15. Lin, D., et al., *Label-free blood plasma test based on surface-enhanced Raman scattering for tumor stages detection in nasopharyngeal cancer.* Scientific reports, 2014. **4**(1): p. 1-8.

16. Shin, H., et al., *Early-stage lung cancer diagnosis by deep learning-based spectroscopic analysis of circulating exosomes.* ACS nano, 2020. **14**(5): p. 5435-5444.

17. Ozaki, Y., et al., *Near-Infrared Spectroscopy: Theory, Spectral Analysis, Instrumentation, and Applications*. 2020: Springer Nature.

18. Jue, T. and K. Masuda, *Application of near infrared spectroscopy in biomedicine*. 2013: Springer.

19. Sakudo, A., *Near-infrared spectroscopy for medical applications: Current status and future perspectives.* Clinica Chimica Acta, 2016. **455**: p. 181-188.

20. Tsenkova, R., et al., *Near-infrared spectroscopy for biomonitoring: influence of somatic cell count on cow's milk composition analysis.* International dairy journal, 2001. **11**(10): p. 779-783.

21. Omar, A.F., H. Atan, and M.Z. MatJafri, *NIR spectroscopic properties of aqueous acids solutions.* Molecules, 2012. **17**(6): p. 7440-7450.

22. Balan, V., et al., *Vibrational spectroscopy fingerprinting in medicine: from molecular to clinical practice.* Materials, 2019. **12**(18): p. 2884.

23. Hirosawa, N., et al., *In vivo investigation of progressive alterations in rat mammary gland tumors by near-infrared spectroscopy.* Analytical biochemistry, 2002. **305**(2): p. 156-165.

24. Ali, J., et al., *Near infrared spectroscopy and imaging to probe differences in water content in normal and cancer human prostate tissues.* Technology in cancer research & treatment, 2004. **3**(5): p. 491-497.

25. Kondepati, V.R., et al., *CH-overtone regions as diagnostic markers for near-infrared spectroscopic diagnosis of primary cancers in human pancreas and colorectal tissue.* Analytical and bioanalytical chemistry, 2007. **387**(5): p. 1633-1641.

26. Ehlen, L., et al., *Synergy of fluorescence and near-infrared spectroscopy in detection of colorectal cancer.* Journal of Surgical Research, 2019. **242**: p. 349-356.

27. Redd, D.C., et al., *Raman spectroscopic characterization of human breast tissues: implications for breast cancer diagnosis.* Applied spectroscopy, 1993. **47**(6): p. 787-791.

28. Gazi, E., 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, 2003. **201**(1): p. 99-108.

29. Chan, J.W., et al., *Nondestructive identification of individual leukemia cells by laser trapping Raman spectroscopy.* Analytical chemistry, 2008. **80**(6): p. 2180-2187.

30. Harvey, T.J., et al., *Spectral discrimination of live prostate and bladder cancer cell lines using Raman optical tweezers.* Journal of biomedical optics, 2008. **13**(6): p. 064004.

31. Oshima, Y., et al., *Discrimination analysis of human lung cancer cells associated with histological type and malignancy using Raman spectroscopy.* Journal of biomedical optics, 2010. **15**(1): p. 017009.

32. Talari, A., et al., *Raman spectroscopic analysis differentiates between breast cancer cell lines.* Journal of Raman Spectroscopy, 2015. **46**(5): p. 421-427.

33. Chaturvedi, D., et al., *Different phases of breast cancer cells: Raman study of immortalized, transformed, and invasive cells.* Biosensors, 2016. **6**(4): p. 57.

34. Kyriakidou, M., et al., *FT-IR spectroscopy study in early diagnosis of skin cancer.* in vivo, 2017. **31**(6): p. 1131-1137.

35. Naurecka, M.L., et al., *FTIR-ATR and FT-Raman spectroscopy for biochemical changes in oral tissue.* American Journal of Analytical Chemistry, 2017. **8**(03).

36. Elmi, F., et al., *Application of FT-IR spectroscopy on breast cancer serum analysis.* Spectrochimica Acta Part A: Molecular and Biomolecular Spectroscopy, 2017. **187**: p. 87-91.

37. Zhang, K., et al., *Diagnosis of liver cancer based on tissue slice surface enhanced Raman spectroscopy and multivariate analysis.* Vibrational Spectroscopy, 2018. **98**: p. 82-87.

38. Ghosh, A., et al., *Chemometric analysis of integrated FTIR and Raman spectra obtained by non-invasive exfoliative cytology for the screening of oral cancer.* Analyst, 2019. **144**(4): p. 1309-1325.

39. Trunfio, N., et al., *Characterization of mammalian cell culture raw materials by combining spectroscopy and chemometrics.* Biotechnology progress, 2017. **33**(4): p. 1127-1138.

40. Bao, W., B. Yang, and B. Chen, *2-hydr_ensemble: lysine 2-hydroxyisobutyrylation identification with ensemble method.* Chemometrics and Intelligent Laboratory Systems, 2021. **215**: p. 104351.

41. Yang, B., W. Bao, and J. Wang, *Active disease-related compound identification based on capsule network.* Briefings in bioinformatics, 2022. **23**(1): p. bbab462.

42. Osborne, B.G., T. Fearn, and P.H. Hindle, *Practical NIR spectroscopy with applications in food and beverage analysis*. 1993: Longman scientific and technical.

43. Bhatia, H., et al., *In- line monitoring of amino acids in mammalian cell cultures using raman spectroscopy and multivariate chemometrics models.* Engineering in life sciences, 2018. **18**(1): p. 55-61.

44. Clavaud, M., et al., *Chemometrics and in-line near infrared spectroscopic monitoring of a biopharmaceutical Chinese hamster ovary cell culture: prediction of multiple cultivation variables.* Talanta, 2013. **111**: p. 28-38.

45. Sellick, C.A., et al., *Rapid monitoring of recombinant antibody production by mammalian cell cultures using Fourier transform infrared spectroscopy and chemometrics.* Biotechnology and bioengineering, 2010. **106**(3): p. 432-442.

46. Kozma, B., A. Salgó, and S. Gergely, *Comparison of multivariate data analysis techniques to improve glucose concentration prediction in mammalian cell cultivations by Raman spectroscopy.* Journal of pharmaceutical and biomedical analysis, 2018. **158**: p. 269-279.

47. Cui, X., et al., *Water as a probe for serum–based diagnosis by temperature–dependent near–infrared spectroscopy.* Talanta, 2019. **204**: p. 359-366.

48. Tsenkova, R., Z. Kovacs, and Y. Kubota, *Aquaphotomics: near infrared spectroscopy and water states in biological systems.* Membrane Hydration, 2015: p. 189-211.

49. Tsenkova, R., et al., *Essentials of aquaphotomics and its chemometrics approaches.* Frontiers in chemistry, 2018. **6**: p. 363.

50. Manley, M., *Near-infrared spectroscopy and hyperspectral imaging: non-destructive analysis of biological materials.* Chemical Society Reviews, 2014. **43**(24): p. 8200-8214.

51. De Beer, T., et al., *Near infrared and Raman spectroscopy for the in-process monitoring of pharmaceutical production processes.* International journal of pharmaceutics, 2011. **417**(1-2): p. 32-47.

52. Sandell, L. and D. Sakai, *Mammalian cell culture.* Current Protocols Essential Laboratory Techniques, 2011. **5**(1): p. 4.3. 1-4.3. 32.

53. Leme, J., et al., *A multivariate calibration procedure for UV/VIS spectrometric monitoring of BHK-21 cell metabolism and growth.* Biotechnology progress, 2014. **30**(1): p. 241-248.

54. Lai, C.Z., et al., *Spectrophotometric measurement of freshwater pH with purified meta- cresol purple and phenol red.* Limnology and Oceanography: Methods, 2016. **14**(12): p. 864-873.

55. Poole, C.B., et al., *Colorimetric tests for diagnosis of filarial infection and vector surveillance using non-instrumented nucleic acid loop-mediated isothermal amplification (NINA-LAMP).* PloS one, 2017. **12**(2): p. e0169011.

56. Pick, E. and Y. Keisari, *A simple colorimetric method for the measurement of hydrogen peroxide produced by cells in culture.* Journal of immunological methods, 1980. **38**(1-2): p. 161-170.

57. Mills, A. and G.A. Skinner, *A novel 'fizziness' indicator.* Analyst, 2011. **136**(5): p. 894-896.

58. Iseki, K., et al., *Helicobacter pylori infection in patients with early gastric cancer by the endoscopic phenol red test.* Gut, 1998. **42**(1): p. 20-23.

59. Takahashi, M.B., et al., *Artificial neural network associated to UV/Vis spectroscopy for monitoring bioreactions in biopharmaceutical processes.* Bioprocess and biosystems engineering, 2015. **38**(6): p. 1045-1054.

60. Baylor, S. and S. Hollingworth, *Absorbance signals from resting frog skeletal muscle fibers injected with the pH indicator dye, phenol red.* The Journal of general physiology, 1990. **96**(3): p. 449-471.

61. Bolsover, S., J. Brown, and T. Goldsmith, *Intracellular pH of Limulus ventral photoreceptor cells: measurement with phenol red.* Society of General Physiologists Series, 1986. **40**: p. 285-310.

62. Nishida, K., et al., *Effect of albumin on the absorption of phenol red, bromphenol blue and bromosulphonphthalein as model drugs from the liver surface membrane in rats.* Biological and Pharmaceutical Bulletin, 1995. **18**(11): p. 1548-1550.

63. Sochacka, J., *Application of phenol red as a marker ligand for bilirubin binding site at subdomain IIA on human serum albumin.* Journal of Photochemistry and Photobiology B: Biology, 2015. **151**: p. 89-99.

64. Abd Ghani, K., et al., *VIS–NIR spectral signature and quantitative analysis of HeLa and DU145 cell line.* Spectrochimica Acta Part A: Molecular and Biomolecular Spectroscopy, 2019. **222**: p. 117241.

65. Amran, E.N., et al., *Potential colorimetric detection of cancer cells using Phenol Red.* Photodiagnosis and photodynamic therapy, 2019. **27**: p. 380-384.

66. Arnold, S.A., et al., *In- situ near infrared spectroscopy to monitor key analytes in mammalian cell cultivation.* Biotechnology and bioengineering, 2003. **84**(1): p. 13-19.

67. Wold, S., K. Esbensen, and P. Geladi, *Principal component analysis.* Chemometrics and intelligent laboratory systems, 1987. **2**(1-3): p. 37-52.

68. Wold, S. and M. Sjöström, *SIMCA: a method for analyzing chemical data in terms of similarity and analogy*. 1977, ACS Publications.

69. Tsenkova, R.N., et al., *Prion protein fate governed by metal binding.* Biochemical and biophysical research communications, 2004. **325**(3): p. 1005-1012.

70. Savitzky, A. and M.J. Golay, *Smoothing and differentiation of data by simplified least squares procedures.* Analytical chemistry, 1964. **36**(8): p. 1627-1639.

71. Barnes, R., M.S. Dhanoa, and S.J. Lister, *Standard normal variate transformation and de-trending of near-infrared diffuse reflectance spectra.* Applied spectroscopy, 1989. **43**(5): p. 772-777.

72. Seasholtz, M.B. and B.R. Kowalski, *The effect of mean centering on prediction in multivariate calibration.* Journal of Chemometrics, 1992. **6**(2): p. 103-111.

73. Magnusson, E.B., et al., *Real-time optical pH measurement in a standard microfluidic cell culture system.* Biomedical optics express, 2013. **4**(9): p. 1749-1758.

74. Hanahan, D. and R.A. Weinberg, *Hallmarks of cancer: the next generation.* cell, 2011. **144**(5): p. 646-674.

75. Gatenby, R.A. and R.J. Gillies, *Why do cancers have high aerobic glycolysis?* Nature reviews cancer, 2004. **4**(11): p. 891-899.

76. Granja, S., et al. *Value of pH regulators in the diagnosis, prognosis and treatment of cancer*. in *Seminars in cancer biology*. 2017. Elsevier.

77. W Huck, C., Y. Ozaki, and V. A Huck-Pezzei, *Critical review upon the role and potential of fluorescence and near-infrared imaging and absorption spectroscopy in cancer related cells, serum, saliva, urine and tissue analysis.* Current Medicinal Chemistry, 2016. **23**(27): p. 3052-3077.

78. Liu, K.-Z., et al., *Quantitative determination of serum LDL cholesterol by near-infrared spectroscopy.* Vibrational spectroscopy, 2005. **38**(1-2): p. 203-208.

79. Baba, A.I. and C. Câtoi, *Tumor cell morphology*, in *Comparative Oncology*. 2007, The Publishing House of the Romanian Academy.

80. Lin, X., et al., *Light scattering from normal and cervical cancer cells.* Applied optics, 2017. **56**(12): p. 3608-3614.

81. Stevenson, K., et al., *General calibration of microbial growth in microplate readers.* Scientific reports, 2016. **6**(1): p. 1-7.

82. Chung, S., et al., *Non-invasive tissue temperature measurements based on quantitative diffuse optical spectroscopy (DOS) of water.* Physics in Medicine & Biology, 2010. **55**(13): p. 3753.

83. Chung, S., et al., *In vivo water state measurements in breast cancer using broadband diffuse optical spectroscopy.* Physics in Medicine & Biology, 2008. **53**(23): p. 6713.

84. Brubach, J.-B., et al., *Signatures of the hydrogen bonding in the infrared bands of water.* The Journal of chemical physics, 2005. **122**(18): p. 184509.

85. Afrin, T., et al., *Water structure modification by sugars and its consequence on micellization behavior of cetyltrimethylammonium bromide in aqueous solution.* Journal of Solution Chemistry, 2013. **42**(7): p. 1488-1499.

86. Guo, Y., et al., *A pH-responsive colorimetric strategy for DNA detection by acetylcholinesterase catalyzed hydrolysis and cascade amplification.* Biosensors and Bioelectronics, 2017. **94**: p. 651-656.

87. Yang, Y., et al., *A pH-responsive bioassay for paper-based diagnosis of exosomes via mussel-inspired surface chemistry.* Talanta, 2019. **192**: p. 325-330.

88. Bonner, O.D. and G.B. Woolsey, *Effect of solutes and temperature on the structure of water.* The Journal of Physical Chemistry, 1968. **72**(3): p. 899-905.

89. Workman, J. and J. Workman, *Handbook of Organic Compounds: Methods and Interpretations*. Vol. 1. 2001: Academic Press.

90. Kovacs, Z., et al., *Water Spectral Patterns Reveals Similarities and Differences in Rice Germination and Induced Degenerated Callus Development.* Plants, 2021. **10**(9): p. 1832.

91. Okubo, N. and Y. Kurata, *Nondestructive classification analysis of green coffee beans by using near-infrared spectroscopy.* Foods, 2019. **8**(2): p. 82.

92. Tsenkova, R., *Aquaphotomics: dynamic spectroscopy of aqueous and biological systems describes peculiarities of water.* Journal of Near Infrared Spectroscopy, 2009. **17**(6): p. 303-313.

93. Cattaneo, H., T. Laurila, and R. Hernberg, *VCSEL based detection of water vapor near 940 nm.* Spectrochimica Acta Part A: Molecular and Biomolecular Spectroscopy, 2004. **60**(14): p. 3269-3275.

94. Cattaneo, H. and R. Hernberg, *Detection of high-temperature water vapor at 940 nm with vertical-cavity surface-emitting lasers.* Applied optics, 2005. **44**(31): p. 6593-6598.

95. Sierk, B., et al., *Field measurements of water vapor continuum absorption in the visible and near-infrared*. Journal of Geophysical Research: Atmospheres, 2004. **109**(D8).

96. Workman, J. and L. Weyer, *Practical Guide and Spectral Atlas for Interpretive Near-Infrared*. 2012: CRC.

97. Heiman, A. and S. Licht, *Fundamental baseline variations in aqueous near-infrared analysis.* Analytica chimica acta, 1999. **394**(2-3): p. 135-147.

98. Weber, J.M., et al., *Isolating the spectroscopic signature of a hydration shell with the use of clusters: Superoxide tetrahydrate.* Science, 2000. **287**(5462): p. 2461-2463.

99. Headrick, J.M., et al., *Spectral signatures of hydrated proton vibrations in water clusters.* Science, 2005. **308**(5729): p. 1765-1769.

100. Xantheas, S.S., *Ab initio studies of cyclic water clusters (H2O) n, n= 1–6. III. Comparison of density functional with MP2 results.* The Journal of chemical physics, 1995. **102**(11): p. 4505-4517.

101. Mizuse, K. and A. Fujii, *Tuning of the internal energy and isomer distribution in small protonated water clusters H+ (H2O) 4–8: An application of the inert gas messenger technique.* The Journal of Physical Chemistry A, 2012. **116**(20): p. 4868-4877.

102. Davis, J.G., et al., *Water structural transformation at molecular hydrophobic interfaces.* Nature, 2012. **491**(7425): p. 582-585.

103. Golic, M., K. Walsh, and P. Lawson, *Short-wavelength near-infrared spectra of sucrose, glucose, and fructose with respect to sugar concentration and temperature.* Applied spectroscopy, 2003. **57**(2): p. 139-145.

104. Osborne, B. and T. Fearn, *Near Infrared Spectroscopy in Food Analysis; Longman Scientific & Technical: Essex*. 1986, UK.

105. Ciurczak, E.W. and B. Igne, *Pharmaceutical and medical applications of near-infrared spectroscopy*. 2014: CRC Press.

106. Šašić, S. and Y. Ozaki, *Short-wave near-infrared spectroscopy of biological fluids. 1. Quantitative analysis of fat, protein, and lactose in raw milk by partial least-squares regression and band assignment.* Analytical chemistry, 2001. **73**(1): p. 64-71.

107. Workman, J., *Handbook of organic compounds: NIR, IR, Raman and UV-Vis spectra featuring polymers and surfactants (a 3-volume set). 3. IR and Raman spectra*. 2001: Academic Press.

108. Kawano, S., H. Watanabe, and M. Iwamoto, *Determination of sugar content in intact peaches by near infrared spectroscopy with fiber optics in interactance mode.* Journal of the Japanese Society for Horticultural Science, 1992. **61**(2): p. 445-451.

109. Woo, Y.-A., et al., *Development of a new measurement unit (MilkSpec-1) for rapid determination of fat, lactose, and protein in raw milk using near-infrared transmittance spectroscopy.* Applied spectroscopy, 2002. **56**(5): p. 599-604.

110. Nakakimura, Y., et al., *Extracellular metabolites play a dominant role in near-infrared spectroscopic quantification of bacteria at food-safety level concentrations.* Analytical Methods, 2012. **4**(5): p. 1389-1394.

111. Fantini, S. and M.A. Franceschini, *Frequency-domain techniques for tissue spectroscopy and imaging.* Handbook of optical biomedical diagnostics, 2002. **1**.

112. Cerussi, A.E., et al., *In vivo absorption, scattering, and physiologic properties of 58 malignant breast tumors determined by broadband diffuse optical spectroscopy.* Journal of biomedical optics, 2006. **11**(4): p. 044005.

113. Beć, K.B., J. Grabska, and C.W. Huck, *Near-infrared spectroscopy in bio-applications.* Molecules, 2020. **25**(12): p. 2948.

114. Heffer, E.L., et al., *Near-infrared imaging of the human breast: complementing hemoglobin concentration maps with oxygenation images.* Journal of Biomedical Optics, 2004. **9**(6): p. 1152-1160.

115. Hollis, V.S., T. Binzoni, and D.T. Delpy. *Noninvasive monitoring of brain tissue temperature by near-infrared spectroscopy*. in *Optical Tomography and Spectroscopy of Tissue IV*. 2001. SPIE.

116. Hornung, R., et al., *Quantitative near-infrared spectroscopy of cervical dysplasia in vivo.* Human Reproduction, 1999. **14**(11): p. 2908-2916.

117. Li, Y., et al., *Early Diagnosis of Type 2 Diabetes Based on Near-Infrared Spectroscopy Combined With Machine Learning and Aquaphotomics.* Frontiers in Chemistry, 2020: p. 1133.

118. McIntosh, L.M., et al., *Towards non-invasive screening of skin lesions by near-infrared spectroscopy.* Journal of Investigative Dermatology, 2001. **116**(1): p. 175-181.

119. Muncan, J. and R. Tsenkova, *Aquaphotomics—From innovative knowledge to integrative platform in science and technology.* Molecules, 2019. **24**(15): p. 2742.

120. Simick, M.K., et al., *Non-ionizing near-infrared radiation transillumination spectroscopy for breast tissue density and assessment of breast cancer risk.* Journal of biomedical optics, 2004. **9**(4): p. 794-803.

121. Tromberg, B.J., et al., *Imaging in breast cancer: diffuse optics in breast cancer: detecting tumors in pre-menopausal women and monitoring neoadjuvant chemotherapy.* Breast Cancer Research, 2005. **7**(6): p. 1-7.

122. Kovacs, Z., et al., *Rapid bacteria selection using Aquaphotomics and near infrared spectroscopy.* 2019.

123. Slavchev, A., et al., *Monitoring of water spectral patterns of lactobacilli development as a tool for rapid selection of probiotic candidates.* Journal of Near Infrared Spectroscopy, 2017. **25**(6): p. 423-431.

124. Slavchev, A., et al., *Monitoring of water spectral pattern reveals differences in probiotics growth when used for rapid bacteria selection.* PLoS One, 2015. **10**(7): p. e0130698.

125. Sakudo, A., et al., *Ex vivo tissue discrimination by visible and near-infrared spectra with chemometrics.* Journal of veterinary medical science, 2006. **68**(12): p. 1375-1378.