

Review

# Recent Advances in Raman Spectral Imaging in Cell Diagnosis and Gene Expression Prediction

Tomonobu M. Watanabe <sup>1,2</sup>, Kensuke Sasaki <sup>2</sup> and Hideaki Fujita <sup>1,\*</sup> 

<sup>1</sup> Department of Stem Cell Biology, Research Institute for Radiation Biology and Medicine, Hiroshima University, 1-2-3 Minami-ku, Hiroshima 734-8553, Japan

<sup>2</sup> Laboratory for Comprehensive Bioimaging, RIKEN Center for Biosystems Dynamics Research (BDR), 2-2-3 Minatogima-minamimachi, Kobe 650-0047, Japan

\* Correspondence: fujitah@hiroshima-u.ac.jp; Tel.: +81-82-257-5808

**Abstract:** Normal and tumor regions within cancer tissue can be distinguished using various methods, such as histological analysis, tumor marker testing, X-ray imaging, or magnetic resonance imaging. Recently, new discrimination methods utilizing the Raman spectra of tissues have been developed and put into practical use. Because Raman spectral microscopy is a non-destructive and non-labeling method, it is potentially compatible for use in the operating room. In this review, we focus on the basics of Raman spectroscopy and Raman imaging in live cells and cell type discrimination, as these form the bases for current Raman scattering-based cancer diagnosis. We also review recent attempts to estimate the gene expression profile from the Raman spectrum of living cells using simple machine learning. Considering recent advances in machine learning techniques, we speculate that cancer type discrimination using Raman spectroscopy will be possible in the near future.

**Keywords:** spectroscopy; machine learning; non-linear optics



**Citation:** Watanabe, T.M.; Sasaki, K.; Fujita, H. Recent Advances in Raman Spectral Imaging in Cell Diagnosis and Gene Expression Prediction. *Genes* **2022**, *13*, 2127. <https://doi.org/10.3390/genes13112127>

Academic Editors: Keiko Kawauchi and Miwako Kato Homma

Received: 19 October 2022

Accepted: 15 November 2022

Published: 16 November 2022

**Publisher's Note:** MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Copyright:** © 2022 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

## 1. Raman Spectroscopy of Mammalian Cells

Biological substances are composed of various biomolecules, each of which has a unique molecular structure with various vibrational modes. Light irradiated on these substances is scattered after the energy is transferred; corresponding to these vibrational modes between the substances and the light is the Raman scattering phenomenon. The Raman scattered light emitted by Raman active substances is the light that has been modulated by all the vibration modes of the constituent molecules. Hence, the obtained spectrum of light contains information on all the molecular constituents present and their physical state, making it possible to determine the molecular composition of the subjects. However, in the case of biological samples, such as cells, is more complex. The Raman spectrum emitted by a cell comprises the vibrational modes of a wide variety of molecular species, including proteins [1–4], lipids [5], and nucleotides [6,7], which are superimposed in too complex a manner to be resolved [8,9]. However, Raman scattering spectra can be acquired non-invasively and can be used as a quantifiable parameter to describe the function and state of cells [10]. To achieve this, the technological development of signal- and information-processing technologies [11], including chemometrics and machine learning [12], is indispensable. Machine learning methods turn the complexity of Raman scattering spectra of cells from a weakness to an advantage [8].

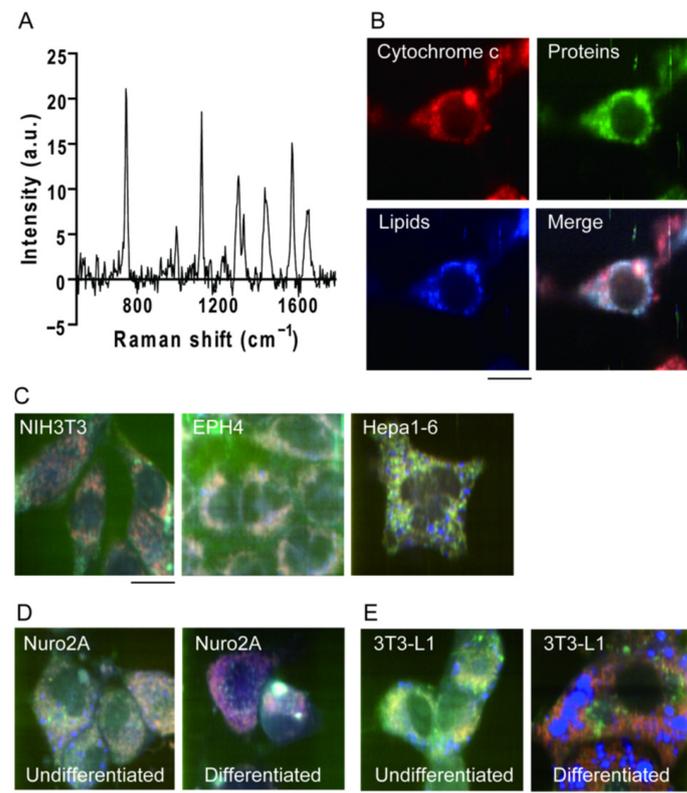
## 2. Raman Observation of Living Cells

Raman scattered light is weak relative to irradiated light. Therefore, in practice, strong light irradiation is required to obtain a good signal; however, this causes photodamage to the cell. When cells are exposed to strong light, reactive oxygen species are internally produced, which can destroy DNA, mitochondria, and other organelles, thereby leading to cell death [13]. Additionally, metabolic activity may change as a result of resistance to light

irradiation, which is included as false signals in the acquired Raman scattering spectrum. To correctly measure the Raman spectra of cells, it is essential to collect weak signals while keeping the cells alive by minimizing light-induced damage. Since the 1980s, with the increase in sensitivity of multichannel spectrometers and detectors and the significant improvement of filter technology, it is now possible to analyze spectra even with weak Raman scattering light [14]. Raman scattering spectroscopy of cells is currently widely used in the fields of cell biology and medicine [15,16].

Katsumasa Fujita, and colleagues constructed line-scanning Raman spectroscopic microscope specialized for observing living cells [17,18]. This line scanning configuration dramatically increased the scanning speed to  $\sim 100\times$  compared to point scanning confocal Raman microscope [19]. Because cells contain numerous molecules that emit autofluorescence, the detected spectrum also includes the autofluorescence of the cells. If the cells are attached to a substrate, such as a glass slide, the signal derived from the glass will also be included. Additionally, microstructures that cause scattering in the cell and multiple scattering that occurs between the cell and the substrate are also included in the detected spectrum. Methods to remove these background signals are a subject of research and development [20]. The objective is to estimate and remove the background signal via polynomial approximation and extract a clear narrow peak or extract only the cell-derived signal by subtracting the signal of the substrate.

Raman scattering in the mouse fibroblast cell line NIH3T3, show a complex spectrum (Figure 1A) with some peaks where the resulting vibrational mode is clear. For example, the peak at  $1340\text{ cm}^{-1}$  was mainly attributed to the  $\text{CH}_2$  stretching mode in the main lipid chain (Figure 1B, blue). The peaks at  $753\text{ cm}^{-1}$ ,  $1131\text{ cm}^{-1}$ , and  $1585\text{ cm}^{-1}$  are attributed to the protein cytochrome c (Figure 1B, red), and the peak at  $1686\text{ cm}^{-1}$  is attributed to peptide bonds in various proteins (Figure 1B, green). By assigning the signal intensities of three of these characteristic peak values ( $753\text{ cm}^{-1}$ ,  $1686\text{ cm}^{-1}$ , and  $1340\text{ cm}^{-1}$ ) to the three blue-red-green colors, a pseudo-color image reflecting the Raman scattering spectrum of the cell line can be created (Figure 1B). Different cell types and conditions contain different ratios of constituent molecules. Therefore, the Raman scattering spectra should exhibit different shapes depending on the cell type. Raman spectra from three cell lines of mouse origin: fibroblasts (NIH3T3), mammary epithelial cells (EpH4), and hepatoma cells (Hepa1-6) were collected [9]. By comparing the pseudo-color images created based on the Raman scattering spectra emitted from each cell, it was confirmed that the color tones were different, that is, the Raman scattering spectra were different (Figure 1C). In addition to differences in cell line types, the changes in the color of the pseudo-color images were observed before and after cell differentiation, both in the case of neuronal differentiation (mouse neuroblasts (Neuro 2a)) (Figure 1D) and adipocyte differentiation (mouse fibroblasts (3T3-L1)) (Figure 1E). This confirms that Raman scattering spectra reflect the cell type and state [9]. Other groups successfully discriminated cell lineages in co-culture of MDCK and CHO cells using partial least-squares discriminant analysis, showing the applicability of Raman spectroscopy in non-label cell type discrimination [21].



**Figure 1.** Raman spectra from living cells. (A) Typical Raman spectrum from a living NIH3T3 cell. (B) Pseudo-color Raman image of NIH3T3 cells. Cells were colored by peak values at  $753\text{ cm}^{-1}$  (cytochrome c; red),  $1686\text{ cm}^{-1}$  (proteins, green) and  $1340\text{ cm}^{-1}$  (lipids; blue). Scale bar,  $10\text{ }\mu\text{m}$ . (C) Raman image of NIH3T3 cells (left), EPH4 cells (middle), and Hepa1-6 cells (right). Scale bar,  $10\text{ }\mu\text{m}$ . (D,E) Raman image of Neuro 2A (D) and 3T3-L1 (E) cells before (left) and after (right) differentiation. Scale bar,  $10\text{ }\mu\text{m}$ . Reproduction from reference [9].

The signal intensity of selective peaks in Raman scattering spectra can be used to evaluate the function and state of cells, although the application is limited because of the large photodamage. For example, cytochrome c, a heme protein bound to the inner mitochondrial membrane, is one of the few molecules that show multiple clear peaks in the complex Raman scattering spectrum. Cytochrome c shows different structures between oxidized and reduced forms, but the Raman scattering intensity of reduced cytochrome c is higher than that of oxidized cytochrome c. Based on this property, malarial infection of red blood cells can be detected using the Raman scattering spectrum [22]. Additionally, when cells undergo apoptosis, cytochrome c is released from the mitochondria into the cytoplasm, which can be observed using Raman scattering spectroscopy [18]. Because Raman spectroscopy signals are far weaker than fluorescence signals, their temporal resolution is much lower than that of fluorescence observation. However, the risk of artifacts due to fluorescent dye staining or other effects was eliminated. By irradiating cells with light and analyzing the scattered light, cell types and conditions can be identified and separated using Raman scattering spectroscopy without the need for staining or manipulation [23].

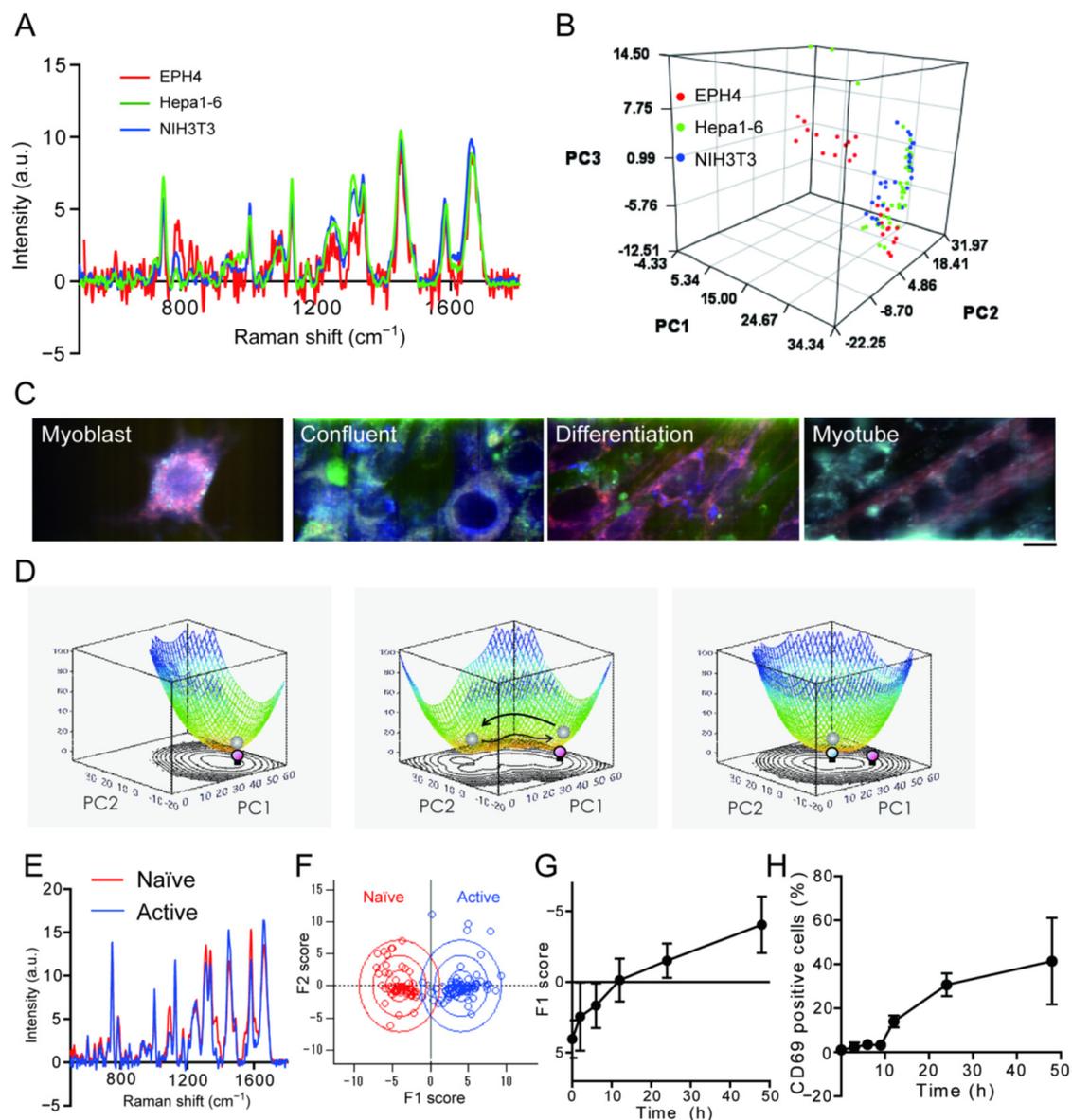
### 3. Use of Machine Learning on Raman Spectrum Analysis

In Raman scattering spectra emitted by cells, both sharp peaks and peaks that are partially overlapped or broad to be identified are often observed. For example, although collagen and glycogen do not show significant specific peaks, they do show characteristic spectral shapes, which can be analyzed using principal component analysis and reverse convolution. Cells produce collagen and glycogen as resistance responses to irradiation, and this feature can be used to evaluate the radioactivity tolerance of cells using Raman

scattering spectra [24]. Discrimination of cell type/function/state using principal component analysis and reverse convolution has been widely applied, starting from the imaging diagnosis of cancer cells in the past [25] to the identification of peripheral nerve tissue, including myelinated and unmyelinated nerves, to avoid nerve damage after surgery [26]. There are already examples of its use in medical practice, such as image-guided surgery, to specifically identify cancer cells during cancer tissue removal surgery [27].

In the pseudo-color image obtained using the peak values of the Raman scattering spectra, it was clear that there were differences among the three cell lines (Figure 1C). However, the averaged Raman spectrum obtained from the cell only shows a seemingly minor difference (Figure 2A). This difference is due to the fact that in the pseudo-color image, the contrast ratio was adjusted to emphasize the difference. However, differences were also observed in the small and broad peaks other than the characteristic peaks used in the pseudo-color image. In decomposition analyses, such as pseudo-color imaging and detection of characteristic peaks, only a part of the information in the Raman spectrum is used. Thus, the comprehensiveness of the Raman scattering spectrum cannot be effectively exploited. Therefore, multivariate analysis such as principal component analysis is often used to measure Raman scattering spectra of cells, which can be analyzed while retaining as much information as possible. Principal component analysis is among the simplest machine learning methods available, wherein data described by a large number of variables are reduced to a small number of independent variables for analysis. When the Raman scattering spectra obtained from the three cell lines were projected as points on a three-dimensional coordinate axis using principal component analysis, each spectrum formed a cluster with some distribution (Figure 2B). This result shows that the differences in spectra depending on the cell type and the heterogeneity of cell states can be visualized using principal component analysis. Thus, by using principal component analysis, differences can be evaluated in more detail by performing an analysis that captures the entire picture of the spectral shape, rather than selectively analyzing the signal intensity of specific peaks.

Visualization of cellular heterogeneity can aid in the physical understanding of live phenomena. Attempt were made to visualize the process of state transitions during cell differentiation using Raman scattering spectra [28]. Raman spectra of C2C12 mouse myoblasts (a model cell line for studying muscle differentiation) over time were measured and a principal component analysis during cell differentiation were performed. Pseudo-color images of the Raman spectral peaks showed that the Raman spectra of the cells changed over time before differentiation (myoblast), immediately after induction of differentiation (confluent), 3 days after induction of differentiation (differentiation), and upon completion of differentiation into myotubes (myotubes) (Figure 2C). This three-dimensional distribution represents the existence probability distribution of the cell states. The potential energy that represents the cell state can be visualized by plotting the logarithm of the reciprocal of the probability distribution. C2C12 cells immediately after differentiation were narrowly distributed in two-dimensional space (Figure 2D, left). Three days after induction of differentiation, when C2C12 cells showed an elongated morphology, their distribution on the same coordinate widened (Figure 2D, middle), that is, the potential became shallower and wider. After completion of differentiation, the state of the cells fell into a trough that differed from the trough of the potential before differentiation (Figure 2D, right). These observations indicate that C2C12 cells have a stable cell state before and after the completion of differentiation, whereas their internal state is fluctuating during the transition process of cell differentiation. Similar results have been observed in mouse embryonic stem cells [2], and which indicates that these results are not unique to C2C12 cells. Cell differentiation monitoring by Raman spectroscopy was reported in various models such as cardiomyocyte differentiation [29], neuronal differentiation [30], or differentiation of organoids [31].



**Figure 2.** Detection of cell-state change during differentiation using Raman spectra. (A) Averaged Raman spectra of EPH4 cells (red), Hepa1-6 cells (green) and NIH3T3 cells (blue). (B) Principal component analysis score plot of Raman spectra analysis of 3 cell-lines. (C) Raman image of C2C12 cells during differentiation. Scale bar, 10  $\mu\text{m}$ . (D) Change in the potential landscapes during C2C12 differentiation estimated by the inverse of logarithm of cell population. The counter plots corresponding to the potential landscapes are also shown. (E) Averaged Raman spectra of naïve (red) and active (blue) T cells. (F) Results of discriminant analysis of the PCA scores. (G) Time course change of F1 score during T cell activation. (H) Activation time course of T cell activation assessed by CD69. Reproduction from reference [28].

Discriminant analysis is one of the simplest supervised machine-learning methods [32]. Discriminant analysis is a method that sets up criteria and axes to separate groups, and calculates the similarity to each group as a score for data that has not been classified. T cells, which are at the center of the biological immune response, are activated by recognizing specific antigen-presenting cells. Stimulation with anti-CD3 and anti-CD28 antibodies mimics the activation of T cells *in vitro*. First, Raman scattering spectra are collected on T cells before stimulation and 48 h after stimulation; we assume that T cells were fully activated (Figure 2E), and these were used as teacher data for discriminant analysis. From

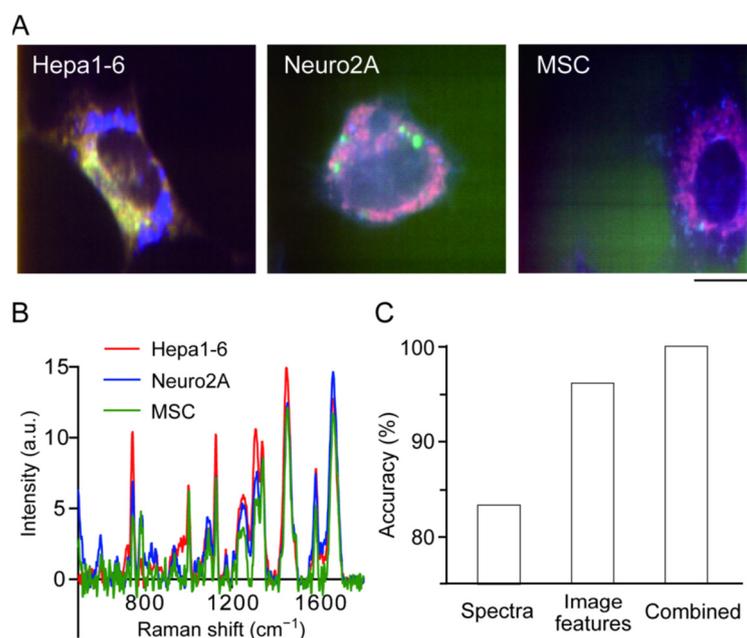
the Raman scattering spectra after an arbitrary elapsed time following stimulation, a score was calculated based on the teacher data (Figure 2F). The scores calculated by measuring the Raman scattering spectra of T cells after 2, 6, 12, 24, and 48 h increased with increasing time after stimulation (Figure 2G). Immune cell activation was assessed by flow cytometry. Upon activation, immune cells expressed specific proteins on the cell surface, which resembled the activation time course obtained by Raman scattering spectra (Figure 2H). Using the discriminant analysis method, the transition process of the cell state can be evaluated without staining using the Raman scattering spectrum. In this manner, the complex Raman scattering spectra of a cell can be effectively used as a “cell fingerprint” using machine learning in combination. Because of the high-content data obtained by Raman spectroscopy, it is compatible to other machine learning methods such as random-forest classification, support vector machine or k-means classification and being used in cancer diagnosis or evaluation of cancer treatment [33–35]. Use of Raman spectra in cancer discrimination is recently becoming popular in combination with machine-learning, including artificial intelligence [35–37], which is expected to further develop in the near future.

#### 4. Utilization of Image Obtained by Raman Spectral Peaks

Morphology of cultured cells are often different between cell types, and discrimination of cell types or state can be possible using bright field or fluorescent image of cells with the use of machine learning [38,39], and more recently by aid of deep learning [40,41]. In Raman spectroscopy of cells, it is necessary to efficiently collect weak Raman scattered light emitted by the cells to minimize light-induced damage [42,43]. Therefore, most examples of Raman scattering spectroscopy of cells are measurements conducted using microscopes equipped with high numerical aperture objectives. Confocal optics is often applied in more practical cases [44–46]. In confocal optics, laser light is focused on the cell, Raman light is generated from the molecules in the focused spot, and the light is focused again on the incident pinhole of the spectrometer to be detected as a spectrum. By scanning the position of the focusing spot in the focal plane, a hyperspectral image can be acquired wherein each pixel contains spectral information. Thus, it is possible to construct a pseudo-color image. Therefore, using information from hyper “Raman” spectral images, we compared the accuracy of discrimination of cell types for Raman spectra only, images composed of specific peaks alone, and both [47].

Raman spectra from three mouse-derived hepatocellular carcinoma cell lines (Hepa1-6), neuroblastoma (Neuro2A), and mesenchymal stem cells (MSC) were collected, which clearly show differences in Raman scattering spectra and have distinctive shapes [47]. The differences between the three cell lines were clearly visible when the signal intensities of the cytochrome c-derived peak ( $753\text{ cm}^{-1}$ ), protein-derived peak ( $1687\text{ cm}^{-1}$ ), and lipid-derived peak ( $1340\text{ cm}^{-1}$ ) were used to create a pseudo-color image (Figure 3A). However, when the entire Raman spectrum averaged over the cell region was compared, it was very similar, indicating that there were only a few spectral peaks that contributed to the discrimination of the three cell lines (Figure 3B). Three combinations were compared in this study: (1) image information alone, (2) Raman spectrum information alone, or (3) both, and set up a spatial axis where the difference between the three cell lines could be best expressed by principal component analysis, and investigated the discrimination accuracy by performing discriminant analysis on that principal component space. The accuracy of discriminant analysis of the three cell lines using the average spectrum alone was only 83% (Figure 3C). For analysis using image information, image features of the 300 spatial patterns were obtained using the 11 mathematical algorithms described by Orlov et al. [48]. When discriminant analysis was performed using these features, the accuracy was as high as 95% (Figure 3C). This is a natural result as three species with clear differences in shape were used in the experiment; however, 5% of the patterns remained misidentified. By using both, that is, all the information in the hyperspectral images, the discrimination accuracy reached 100% (Figure 3C). This indicates that the spectral information and spatial

distribution information contribute to discrimination in both common and complementary ways.



**Figure 3.** Cell type discrimination using image feature of Raman image. (A) Raman image of Hepa1-6 cells (left), Neuro2A cells (middle) and mesenchymal stem cells (MSC) (right). Scale bar, 10  $\mu\text{m}$ . (B) Averaged Raman spectra of Hepa1-6 cells (red), Neuro2A cells (blue) and MSC (green). (C) Accuracy of cell type discrimination using Raman spectra (left), image features (middle) and both (right). Reproduction from reference [47].

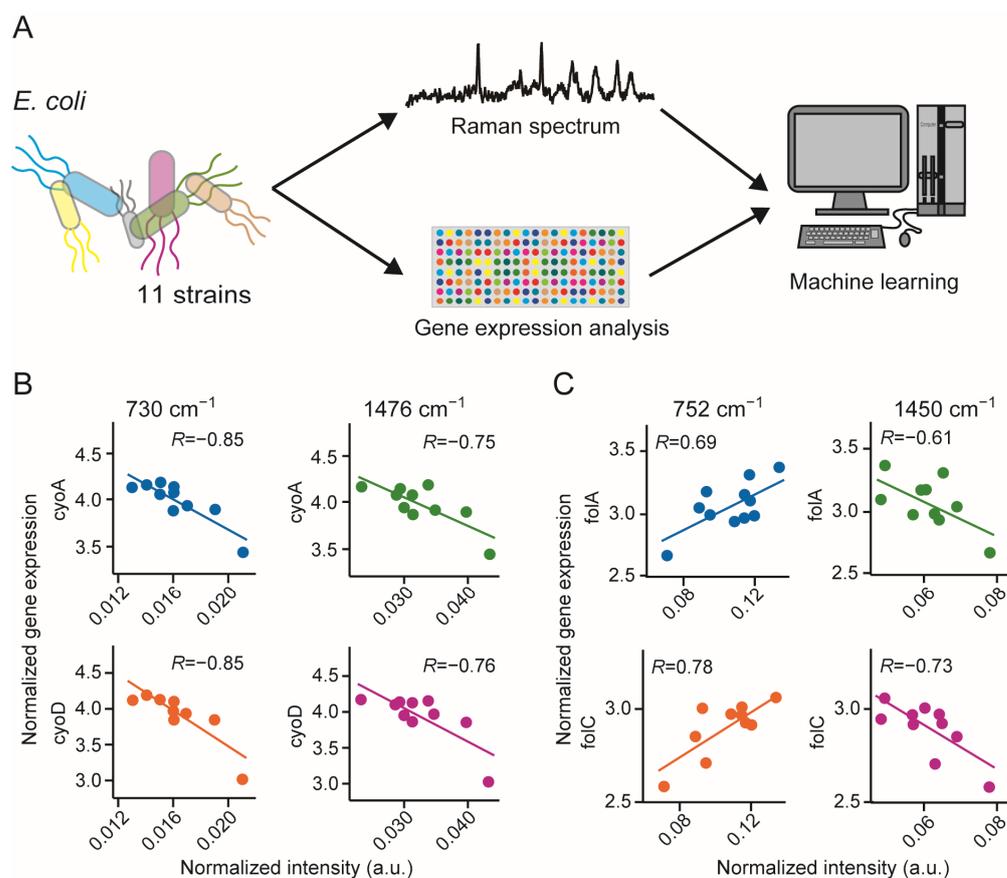
Thus, it was confirmed that the discrimination accuracy was indeed improved by using both spectral and spatial distribution information; however, the amount of computation increased dramatically. Therefore, the next step was to investigate the effective use of the image features for discrimination. Using a color image created with three spectral peaks, discriminant analysis were performed using only the ten image features with the highest discrimination contribution out of 300 features and found that the discrimination accuracy was 85%; however, using 50 features, the discrimination accuracy reached 97%. Even for images consisting of a single specific peak, the discrimination accuracy reached 90% using the peak derived from cytochrome c ( $753\text{ cm}^{-1}$ ); however, the information from the other peaks further improved the discrimination accuracy. Thus, by appropriately selecting the spectral peaks and image features to be used, the number of calculations can be effectively reduced. Recently, by using stimulated Raman scattering (SRS) microscopy which enables fast acquisition of specific Raman peaks, SRS-image based sorting was reported opening the door to non-label cell sorting [49].

### 5. Relationship between Raman Spectra and Gene Expression

The results of the abovementioned study imply that the spectral shape of the Raman spectrum from the cell is patterned according to the state and function of the cell, although the spectral shape becomes so complex that it is difficult to decompose and analyze. Because Raman scattered light comprises all the molecular vibration modes of the observed object, the spectral shape is determined by the type and content ratio of the molecular vibration modes of the observed object. The number of molecular vibration modes in a cell is large but not infinite. This is because the types of molecules that constitute a cell are determined by the types of compounds inside the cell, such as nucleic acids, proteins, and lipids, and are finite. The content ratio of intracellular compounds depends on the type and function of cells. As the type and function of a cell are defined by its gene expression pattern, it can be expected that the Raman spectrum of a cell is correlated with its gene expression pattern.

A cell is an assembly of a vast number of elements of various types, such as proteins, genes, and metabolic molecules, interacting in a complex manner; however, the number of functions and states that a cell can adopt is limited. For example, a mammalian cell has more than 20,000 genes, but because cell functions are a result of the complex interaction of these genes, the cell functions can be aggregated and classified into only ~300 different types [50]. Both Raman spectra and gene expression patterns are complex; however, they are interrelated through a finite number of cellular functions.

Although the above hypothesis makes sense, it has not been proven based on actual mammalian cells. However, attempts were made in simpler organism with less genes; *Escherichia coli* (*E. coli*) which only have 4400 genes, to confirm the feasibility of the above assumption [16]. Unlike mammalian cells, the Raman scattering spectra of *E. coli* can be obtained while the cells float in the culture medium. Raman spectra from floating cells are less likely to be contaminated by signals from the glass substrate, allowing for a more quantitative analysis. Gene expression patterns of 11 drug-resistant *E. coli* strains were collected and created paired data corresponding to the Raman spectra acquired simultaneously (Figure 4A). Using these data pairs, correlation spectra were created by calculating the correlation coefficient between each spectral peak and the gene expression. Correlation between the increase or decrease of a certain peak value in the Raman scattering spectrum and the increase or decrease of a certain gene's expression were reported [51]. For example, the expression of *cyoA* and *cyoD*, genes related to protein synthesis, was negatively correlated with the peak intensities at  $730\text{ cm}^{-1}$  and  $1476\text{ cm}^{-1}$  (Figure 4B). Alternatively, the expression of genes related to folate synthesis, *floA* and *floC*, was positively correlated with the peak intensity at  $752\text{ cm}^{-1}$  and negatively correlated with the peak intensity at  $1450\text{ cm}^{-1}$  (Figure 4C). The peak at  $1450\text{ cm}^{-1}$  is associated with CH<sub>2</sub> deformation. Because there is a correlation between each spectral peak and each gene expression, it is possible, by regression, to estimate the gene expression profile from the acquired Raman spectra. Here, we estimated the gene expression profiles of unknown *E. coli* strains using the partial least squares regression method and found that gene expression could be estimated with an accuracy of 85–90% for all cell lines. However, Raman scattering spectrum analysis is not a suitable substitute for gene expression analysis. This was because the two were mapped through patterning. For example, it is not possible to predict gene expression that is not associated with patterning. There is another important implication from the results of this experiment. The Raman scattering spectrum of a cell depends strongly on its metabolic network. Using this method, it is possible to analyze metabolism and gene expression as a cross-network. In the near future, we expect that Raman scattering spectroscopy of cells will be used in combination with omics technologies.



**Figure 4.** Gene expression prediction using Raman spectra. (A) Schematic illustration of the concept of gene expression prediction. (B) Relationship between Raman peak intensities at 730  $\text{cm}^{-1}$  (left) or 1476  $\text{cm}^{-1}$  and quantile-normalized gene expression of *cyoA* (top) or *cyoD* (bottom). (C) Relationship between Raman peak intensities at 752  $\text{cm}^{-1}$  (left) or 1450  $\text{cm}^{-1}$  and gene expression of *folA* (top) or *folC* (bottom). Reproduction from reference [51].

## 6. Conclusions

Here, we discuss the applicability of Raman spectroscopy in cell state estimation, which is the basis of diagnostic use of Raman spectroscopy. Although its nondestructive and ready-to-measure nature is favorable for its application, there remains a major barrier for its practical and general use in life science and medicine [52]. As all molecules have the ability to emit Raman scattering light, the shape of the acquired Raman spectrum depends on the characteristics of the optical components used. It is also affected by the density of the cells and the substrate to which the cells are attached [53]. Therefore, Raman scattering measurement is principally less reproducible than other microscopy methods, and device calibration is required for every measurement. In combination with machine learning, large information hidden in the complicated Raman spectra and Raman images can be utilized. Because machine learning, particularly deep learning, requires a large amount of training data, it is necessary to perform reproducible and high-quality Raman scattering measurements on unknown cells in a high-throughput manner. Furthermore, as with biological and biochemical methods, manual observation is required, which will also cause unfavorable deviations in data quality. We believe that the key to the useful integration of Raman scattering spectroscopy and machine learning is to automate/data-transform the procedures related to microscopic observations, including the establishment of correction methods for modifications due to optical properties and the pretreatment of samples. To overcome these issues, Raman scattering spectroscopy is an alternative to the current method of pathology diagnosis. Recently, several groups reported the prediction of gene expression using Raman spectra in mammalian cells [54,55], which can be further

expanded in the future. By prediction of gene expression, it may be possible to determine the type of disease, such as cancer, and select effective medication to treat it by measuring the Raman spectrum and predicting gene expression profiles of the tissue in situ in a clinical environment.

**Author Contributions:** Conceptualization, T.M.W. and H.F.; visualization, K.S.; writing, K.S. and H.F.; funding, T.M.W.; All authors have read and agreed to the published version of the manuscript.

**Funding:** This work was supported by Secom Science and Technology Foundation (Specific Field Research grant FY 2017), Japan Agency for Medical Research and Development, grant Number 17bm0804008 and MEXT Grant-in-Aid for Scientific Research on Innovative Areas “Singularity Biology”, grant Number JP18H0540.

**Conflicts of Interest:** The authors declare no conflict of interest.

## References

1. Rygula, A.; Majzner, K.; Marzec, K.M.; Kaczor, A.; Pilarczyk, M.; Baranska, M. Raman spectroscopy of proteins: A review. *J. Raman Spectrosc.* **2013**, *44*, 1061–1076. [[CrossRef](#)]
2. Benevides, J.M.; Overman, S.A.; Thomas, G.J., Jr. Raman Spectroscopy of Proteins. *Curr. Protoc. Protein. Sci.* **2003**, *33*, 17.8.1–17.8.35. [[CrossRef](#)] [[PubMed](#)]
3. Nemecek, D.; Stepanek, J.; Thomas, G.J., Jr. Raman Spectroscopy of Proteins and Nucleoproteins. *Curr. Protoc. Protein. Sci.* **2013**, *71*, 17–18. [[CrossRef](#)]
4. Thomas, G.J. Raman spectroscopy of protein and nucleic acid assemblies. *Annu. Rev. Biophys. Biomol. Struct.* **1999**, *28*, 1–27. [[CrossRef](#)]
5. Czamara, K.; Majzner, K.; Pacia, M.Z.; Kochan, K.; Kaczor, A.; Baranska, M. Raman spectroscopy of lipids: A review. *J. Raman Spectrosc.* **2015**, *46*, 4–20. [[CrossRef](#)]
6. Puppels, G.; Garritsen, H.; Segers-Nolten, G.; De Mul, F.; Greve, J. Raman microspectroscopic approach to the study of human granulocytes. *Biophys. J.* **1991**, *60*, 1046–1056. [[CrossRef](#)]
7. De Gelder, J.; De Gussem, K.; Vandenebeele, P.; Moens, L. Reference database of Raman spectra of biological molecules. *J. Raman Spectrosc.* **2007**, *38*, 1133–1147. [[CrossRef](#)]
8. Guo, S.; Popp, J.; Bocklitz, T. Chemometric analysis in Raman spectroscopy from experimental design to machine learning-based modeling. *Nat. Protoc.* **2021**, *16*, 5426–5459. [[CrossRef](#)]
9. Ichimura, T.; Chiu, L.-D.; Fujita, K.; Kawata, S.; Watanabe, T.M.; Yanagida, T.; Fujita, H. Visualizing Cell State Transition Using Raman Spectroscopy. *PLoS ONE* **2014**, *9*, e84478. [[CrossRef](#)]
10. Petry, R.; Schmitt, M.; Popp, J. Raman Spectroscopy-A Prospective Tool in the Life Sciences. *ChemPhysChem* **2002**, *4*, 14–30. [[CrossRef](#)]
11. Senger, R.S.; Scherr, D. Resolving complex phenotypes with Raman spectroscopy and chemometrics. *Curr. Opin. Biotechnol.* **2020**, *66*, 277–282. [[CrossRef](#)]
12. Ralbovsky, N.M.; Lednev, I.K. Raman spectroscopy and chemometrics: A potential universal method for diagnosing cancer. *Spectrochim. Acta Part A Mol. Biomol. Spectrosc.* **2019**, *219*, 463–487. [[CrossRef](#)] [[PubMed](#)]
13. Talone, B.; Bazzarelli, M.; Schirato, A.; Vicario, F.D.; Viola, D.; Jacchetti, E.; Bregonzio, M.; Raimondi, M.T.; Cerullo, G.; Polli, D. Phototoxicity Induced in Living Hela Cells by Focused Femtosecond Laser Pulses: A Data-Driven Approach. *Biomed. Opt. Express* **2021**, *12*, 7886. [[CrossRef](#)] [[PubMed](#)]
14. Puppels, G.J.; De Mul, F.F.M.; Otto, C.; Greve, J.; Robert-Nicoud, M.; Arndt-Jovin, D.J.; Jovin, T.M. Studying single living cells and chromosomes by confocal Raman microspectroscopy. *Nature* **1990**, *347*, 301–303. [[CrossRef](#)] [[PubMed](#)]
15. Ember, K.J.I.; Hoeve, M.A.; McAughtrie, S.L.; Bergholt, M.S.; Dwyer, B.J.; Stevens, M.M.; Faulds, K.; Forbes, S.J.; Campbell, C.J. Raman spectroscopy and regenerative medicine: A review. *Npj Regen. Med.* **2017**, *2*, 12. [[CrossRef](#)]
16. Smith, R.; Wright, K.L.; Ashton, L. Raman spectroscopy: An evolving technique for live cell studies. *Analyst* **2016**, *141*, 3590–3600. [[CrossRef](#)]
17. Hamada, K.; Fujita, K.; Smith, N.I.; Kobayashi, M.; Inouye, Y.; Kawata, S. Raman microscopy for dynamic molecular imaging of living cells. *J. Biomed. Opt.* **2008**, *13*, 044027. [[CrossRef](#)]
18. Okada, M.; Smith, N.I.; Palonpon, A.F.; Endo, H.; Kawata, S.; Sodeoka, M.; Fujita, K. Label-free Raman observation of cytochrome c dynamics during apoptosis. *Proc. Natl. Acad. Sci. USA* **2011**, *109*, 28–32. [[CrossRef](#)]
19. Klein, K.; Gigler, A.M.; Aschenbrenner, T.; Monetti, R.; Bunk, W.; Jamitzky, F.; Morfill, G.; Stark, R.; Schlegel, J. Label-Free Live-Cell Imaging with Confocal Raman Microscopy. *Biophys. J.* **2012**, *102*, 360–368. [[CrossRef](#)]
20. Gautam, R.; Vanga, S.; Ariese, F.; Umaphathy, S. Review of multidimensional data processing approaches for Raman and infrared spectroscopy. *EPJ Tech. Instrum.* **2015**, *2*, 8. [[CrossRef](#)]
21. Ilin, Y.; Kraft, M.L. Identifying the lineages of individual cells in cocultures by multivariate analysis of Raman spectra. *Anal.* **2014**, *139*, 2177–2185. [[CrossRef](#)] [[PubMed](#)]

22. Hobro, A.J.; Konishi, A.; Coban, C.; Smith, N.I. Raman spectroscopic analysis of malaria disease progression via blood and plasma samples. *Analyst* **2013**, *138*, 3927–3933. [[CrossRef](#)] [[PubMed](#)]
23. Manoharan, R.; Wang, Y.; Feld, M.S. Histochemical Analysis of Biological Tissues Using Raman Spectroscopy. *Spectrochim. Acta Part A Mol. Biomol. Spectrosc.* **1996**, *52*, 215–249. [[CrossRef](#)]
24. Paidi, S.K.; Diaz, P.M.; Dadgar, S.; Jenkins, S.V.; Quick, C.M.; Griffin, R.J.; Dings, R.P.; Rajaram, N.; Barman, I. Label-Free Raman Spectroscopy Reveals Signatures of Radiation Resistance in the Tumor Microenvironment. *Cancer Res.* **2019**, *79*, 2054–2064. [[CrossRef](#)] [[PubMed](#)]
25. Frank, C.J.; Redd, D.C.B.; Gansler, T.S.; McCreery, R.L. Characterization of human breast biopsy specimens with near-IR Raman spectroscopy. *Anal. Chem.* **1994**, *66*, 319–326. [[CrossRef](#)] [[PubMed](#)]
26. Minamikawa, T.; Harada, Y.; Takamatsu, T. Ex vivo peripheral nerve detection of rats by spontaneous Raman spectroscopy. *Sci. Rep.* **2015**, *5*, 17165. [[CrossRef](#)] [[PubMed](#)]
27. Santos, I.P.; Barroso, E.M.; Bakker Schut, T.C.; Caspers, P.J.; van Lanschot, C.G.F.; Choi, D.-H.; van der Kamp, M.F.; Smits, R.W.H.; van Doorn, R.; Verdijk, R.M.; et al. Raman spectroscopy for cancer detection and cancer surgery guidance: Translation to the clinics. *Analyst* **2017**, *142*, 3025–3047. [[CrossRef](#)]
28. Ichimura, T.; Chiu, L.-D.; Fujita, K.; Machiyama, H.; Kawata, S.; Watanabe, T.M.; Fujita, H. Visualizing the appearance and disappearance of the attractor of differentiation using Raman spectral imaging. *Sci. Rep.* **2015**, *5*, 11358. [[CrossRef](#)]
29. Pascut, F.C.; Kalra, S.; George, V.; Welch, N.; Denning, C.; Notingher, I. Non-invasive label-free monitoring the cardiac differentiation of human embryonic stem cells in-vitro by Raman spectroscopy. *Biochim. Et Biophys. Acta (BBA) Gen. Subj.* **2013**, *1830*, 3517–3524. [[CrossRef](#)]
30. Hsu, C.-C.; Xu, J.; Brinkhof, B.; Wang, H.; Cui, Z.; Huang, W.E.; Ye, H. A single-cell Raman-based platform to identify developmental stages of human pluripotent stem cell-derived neurons. *Proc. Natl. Acad. Sci. USA* **2020**, *117*, 18412–18423. [[CrossRef](#)]
31. Tubbesing, K.; Moskwa, N.; Khoo, T.C.; Nelson, D.A.; Sharikova, A.; Feng, Y.; Larsen, M.; Khmaladze, A. Raman microspectroscopy fingerprinting of organoid differentiation state. *Cell. Mol. Biol. Lett.* **2022**, *27*, 53. [[CrossRef](#)] [[PubMed](#)]
32. Ichimura, T.; Chiu, L.-D.; Fujita, K.; Machiyama, H.; Yamaguchi, T.; Watanabe, T.M.; Fujita, H. Non-label immune cell state prediction using Raman spectroscopy. *Sci. Rep.* **2016**, *6*, 37562. [[CrossRef](#)] [[PubMed](#)]
33. Milligan, K.; Van Nest, S.J.; Deng, X.; Ali-Adeeb, R.; Shreeves, P.; Punch, S.; Costie, N.; Pavey, N.; Crook, J.M.; Berman, D.M.; et al. Raman spectroscopy and supervised learning as a potential tool to identify high-dose-rate -brachytherapy induced biochemical profiles of prostate cancer. *J. Biophotonics* **2022**, *15*, e202200121. [[CrossRef](#)]
34. Huang, W.; Shang, Q.; Xiao, X.; Zhang, H.; Gu, Y.; Yang, L.; Shi, G.; Yang, Y.; Hu, Y.; Yuan, Y.; et al. Raman spectroscopy and machine learning for the classification of esophageal squamous carcinoma. *Spectrochim. Acta Part A Mol. Biomol. Spectrosc.* **2022**, *281*, 121654. [[CrossRef](#)] [[PubMed](#)]
35. Kothari, R.; Jones, V.; Mena, D.; Bermúdez Reyes, V.; Shon, Y.; Smith, J.P.; Schmolze, D.; Cha, P.D.; Lai, L.; Storrie-Lombardi, M.C.; et al. Raman spectroscopy and artificial intelligence to predict the Bayesian probability of breast cancer. *Sci. Rep.* **2021**, *11*, 6482. [[CrossRef](#)]
36. He, Q.; Yang, W.; Luo, W.; Wilhelm, S.; Weng, B. Label-Free Differentiation of Cancer and Non-Cancer Cells Based on Machine-Learning-Algorithm-Assisted Fast Raman Imaging. *Biosensors* **2022**, *12*, 250. [[CrossRef](#)]
37. Kumar, Y.; Gupta, S.; Singla, R.; Hu, Y.-C. A Systematic Review of Artificial Intelligence Techniques in Cancer Prediction and Diagnosis. *Arch. Comput. Methods Eng.* **2022**, *29*, 2043–2070. [[CrossRef](#)]
38. Raza, S.H.; Sharma, Y.; Chaudry, Q.; Young, A.N.; Wang, M.D.; Raza, H.; Parry, R.M.; Moffitt, R.A. Automated classification of renal cell carcinoma subtypes using scale invariant feature transform. *Annu. Int. Conf. IEEE Eng. Med. Biol. Soc.* **2009**, *2009*, 6687–6690. [[CrossRef](#)]
39. Tokunaga, K.; Saitoh, N.; Goldberg, I.G.; Sakamoto, C.; Yasuda, Y.; Yoshida, Y.; Yamanaka, S.; Nakao, M. Computational image analysis of colony and nuclear morphology to evaluate human induced pluripotent stem cells. *Sci. Rep.* **2014**, *4*, 6996. [[CrossRef](#)]
40. Cruz-Roa, A.; Gilmore, H.; Basavanthally, A.; Feldman, M.; Ganesan, S.; Shih, N.N.; Tomaszewski, J.; González, F.A.; Madabhushi, A. Accurate and reproducible invasive breast cancer detection in whole-slide images: A Deep Learning approach for quantifying tumor extent. *Sci. Rep.* **2017**, *7*, 46450. [[CrossRef](#)]
41. Piotrowski, T.; Rippel, O.; Elanzew, A.; Nießing, B.; Stucken, S.; Jung, S.; König, N.; Haupt, S.; Stappert, L.; Brüstle, O.; et al. Deep-learning-based multi-class segmentation for automated, non-invasive routine assessment of human pluripotent stem cell culture status. *Comput. Biol. Med.* **2020**, *129*, 104172. [[CrossRef](#)] [[PubMed](#)]
42. Li, G.-Y.; Fan, B.; Ma, T.-H. Visible light may directly induce nuclear DNA damage triggering the death pathway in RGC-5 cells. *Mol. Vis.* **2011**, *17*, 3279–3289. [[PubMed](#)]
43. Hockberger, P.E.; Skimina, T.A.; Centonze, V.E.; Lavin, C.; Chu, S.; Dadras, S.; Reddy, J.K.; White, J.G. Activation of Flavin-Containing Oxidases Underlies Light-Induced Production of H<sub>2</sub>O<sub>2</sub> in Mammalian Cells. *Proc. Natl. Acad. Sci. USA* **1999**, *96*, 6255–6260. [[CrossRef](#)] [[PubMed](#)]
44. Bauer, N.J.; Wicksted, J.P.; Jongasma, F.H.; March, W.F.; Hendrikse, F.; Motamedi, M. Noninvasive assessment of the hydration gradient across the cornea using confocal Raman spectroscopy. *Investig. Ophthalmol. Vis. Sci.* **1998**, *39*, 831–835.
45. Feofanov, A.V.; Grichine, A.I.; Shitova, L.A.; Karmakova, T.A.; Yakubovskaya, R.I.; Egret-Charlier, M.; Vigny, P. Confocal Raman Microspectroscopy and Imaging Study of Theraphthal in Living Cancer Cells. *Biophys. J.* **2000**, *78*, 499–512. [[CrossRef](#)]

46. Huleihel, M.; Shufan, E.; Zeiri, L.; Salman, A. Detection of Vero Cells Infected with Herpes Simplex Types 1 and 2 and Varicella Zoster Viruses Using Raman Spectroscopy and Advanced Statistical Methods. *PLoS ONE* **2016**, *11*, e0153599. [[CrossRef](#)]
47. Germond, A.; Ichimura, T.; Chiu, L.-D.; Fujita, K.; Watanabe, T.M.; Fujita, H. Cell type discrimination based on image features of molecular component distribution. *Sci. Rep.* **2018**, *8*, 11726. [[CrossRef](#)]
48. Orlov, N.; Shamir, L.; Macura, T.; Johnston, J.; Eckley, D.M.; Goldberg, I. WND-CHARM: Multi-purpose image classification using compound image transforms. *Pattern Recognit. Lett.* **2008**, *29*, 1684–1693. [[CrossRef](#)]
49. Nitta, N.; Iino, T.; Isozaki, A.; Yamagishi, M.; Kitahama, Y.; Sakuma, S.; Suzuki, Y.; Tezuka, H.; Oikawa, M.; Arai, F.; et al. Raman image-activated cell sorting. *Nat. Commun.* **2020**, *11*, 3452. [[CrossRef](#)]
50. Valentine, J.W.; Collins, A.G.; Meyer, C.P. Morphological complexity increase in metazoans. *Paleobiology* **1994**, *20*, 131–142. [[CrossRef](#)]
51. Germond, A.; Ichimura, T.; Horinouchi, T.; Fujita, H.; Furusawa, C.; Watanabe, T.M. Raman spectral signature reflects transcriptomic features of antibiotic resistance in *Escherichia coli*. *Commun. Biol.* **2018**, *1*, 85. [[CrossRef](#)] [[PubMed](#)]
52. Choo-Smith, L.P.; Edwards, H.G.M.; Endtz, H.P.; Kros, J.M.; Heule, F.; Barr, H.; Robinson, J.S.; Bruining, H.A.; Puppels, G.J. Medical Applications of Raman Spectroscopy: From Proof of Principle to Clinical Implementation. *Biopolymers* **2002**, *67*, 1–9. [[CrossRef](#)] [[PubMed](#)]
53. Bergholt, M.S.; Serio, A.; Albro, M.B. Raman Spectroscopy: Guiding Light for the Extracellular Matrix. *Front. Bioeng. Biotechnol.* **2019**, *7*, 303. [[CrossRef](#)] [[PubMed](#)]
54. Kobayashi-Kirschvink, K.J.; Nakaoka, H.; Oda, A.; Kamei, K.-I.F.; Noshio, K.; Fukushima, H.; Kanesaki, Y.; Yajima, S.; Masaki, H.; Ohta, K.; et al. Linear Regression Links Transcriptomic Data and Cellular Raman Spectra. *Cell Syst.* **2018**, *7*, 104–117.e4. [[CrossRef](#)]
55. Le Reste, P.; Pilalis, E.; Aubry, M.; McMahon, M.; Cano, L.; Etcheverry, A.; Chatziioannou, A.; Chevet, E.; Fautrel, A. Integration of Raman spectra with transcriptome data in glioblastoma multiforme defines tumour subtypes and predicts patient outcome. *J. Cell. Mol. Med.* **2021**, *25*, 10846–10856. [[CrossRef](#)]