High-Throughput Cell Imaging and Classification by Narrowband and Low-Spectral-Resolution Raman Microscopy

Yasuaki Kumamoto,[†,‡] Kentaro Mochizuki,[†] Kosuke Hashimoto,[‡] Yoshinori Harada,[‡] Hideo Tanaka,[‡] Katsumasa Fujita[†,‡,§,‖]*

[†]Department of Applied Physics, Graduate School of Engineering, Osaka University, 2-1 Yamadaoka, Suita, Osaka 5650871, Japan

[‡]Department of Pathology and Cell Regulation, Graduate School of Medical Sciences, Kyoto Prefectural University of Medicine, 465 Kajiicho, Kawaramachi-Hirokoji, Kamigyo, Kyoto, Kyoto 6028566, Japan

[§]Advanced Photonics and Biosensing Open Innovation Laboratory, AIST-Osaka University, 2-1 Yamadaoka, Suita, Osaka 5650871, Japan

[‖]Transdimensional Life Imaging Division, Institute for Open and Transdisciplinary Research Initiatives, Osaka University, 2-1 Yamadaoka, Suita, Osaka 5650871, Japan

ABSTRACT. We investigated the use of narrowband Raman spectra for rapid label-free molecular imaging aimed at cell classification using principal component regression and linear discriminant analysis. In the classification of breast non-tumorigenic epithelial and cancer cell lines, the classification accuracies using a spectral range of 100cm^{-1} were equivalent to or better than that with using the fingerprint and high-wavenumber regions. Narrowing the Raman spectral range for analysis allows reduction of the CCD pixels required for spectrum detection, resulting in the improvement of image acquisition speed with adequate classification accuracy. Our measurements revealed that the wavenumber region at 1397-1501 cm^{-1} can provide molecular information sufficient for cell classification without causing notable errors in the baseline-correction. A spectral resolution of ~9 cm^{-1} was found to be sufficient to provide high accuracy in cell classification, which allowed us to apply pixel binning at the CCD readout for further acceleration of the imaging speed. As a result, the acquisition time for a 1200 x 1500 pixels Raman hyperspectral image at 1397-1501 cm^{-1} was reduced to 21 min. Under this condition, different cell lines were classified at accuracies higher than 90%. The presented approach will improve throughput of cell and tissue analysis and classification using Raman spectroscopy and extend practical uses of Raman imaging in biology and medicine.

Introduction

In cell biology, quantitative molecular analysis and imaging are essential for understanding functions as well as states of cells. Raman scattering microscopy is a powerful tool enabling such analysis and imaging. A Raman scattering spectrum represents the molecular composition of a cell and a tissue.[1] With spectral analysis, the cell function as well as the state of cell and tissue

can be identified accurately.[1-9] When spectral analysis is combined with microscopy, distributions of molecules, cells, and tissues are visualized without any specific labeling of a target.[1,10,11] Raman images can provide information that is not obtained from Raman spectra in analyzing the function, state and type of a cell.[12,13] However, the acquisition of a Raman image takes several tens of minutes or even hours in general.[11,14,15] Despite many technological advances made in the past twenty years[16-21] to tackle this issue, Raman imaging is still much slower than other cell analysis techniques such as fluorescence imaging[22] and flow cytometry.[23]

Here, we demonstrate the narrowband and low-spectral resolution analysis for high-speed Raman imaging. Narrowband and low-spectral resolution Raman hyperspectral imaging allows reduction of the CCD pixels of the detector for capturing the spectrum, resulting to 10- to 20-fold improvement of the image acquisition time. Although the Raman spectrum contains comprehensive information of intracellular molecules, the use of wideband and high-spectral-resolution Raman spectra are not always necessary and practical depending on the application.[24-32] In this research, we investigate the use of narrowband and low-spectral-resolution Raman spectra for classifying breast non-tumorigenic epithelial and cancer cell lines. By using the result of the analysis, we demonstrated high-speed and wide field-of-view Raman imaging of live cells. Our technique can improve throughput of cell and tissue analysis and classification based on Raman imaging and expand the applications of label-free imaging in biology and medicine.

Materials and Methods

We used a human breast cancer cell line (MCF-7) and a human breast non-tumorigenic epithelial cell (MCF-10A). The cells were cultured on a calcium fluoride (CaF_2) substrate. Prior to Raman

imaging, the culturing medium was replaced by a Tyrode's buffer solution. Raman imaging was performed within 2 hours after the cells were taken out of the incubator.

The cells cultured on the CaF_2 substrate was set at the sample plane of an inverted microscope as a part of a homebuilt line-illumination Raman microscope. The basic configuration of the Raman microscope is described elsewhere.[11,16,17] Briefly, λ = 532 nm laser beam was focused with a cylindrical lens and an objective lens (25×, NA1.1) into a line-shaped beam at the sample. The scattered light was collected by the same objective lens and delivered to the entrance slit of a spectrometer equipped with a grating (600 g/mm). The dispersed light was detected with a cooled CCD camera. The readout rate of the camera was set to 100 kHz for the narrowband and/or low-spectral-resolution measurements, while it was set to 2 MHz for the wideband measurement. The readout chips were reduced from the full dimension (2048 × 2048) to 1200 × 1200 and 1200 × 50 pixels for the wideband and narrowband measurements, respectively. For low-resolution and narrowband measurement, each four pixels of the camera along the wavenumber axis were binned and 1200 × 13 pixels were read out.

A single camera exposure provides a one-dimensional hyperspectral image (x and λ) of the locations irradiated with the line-shaped beam. To obtain a two-dimensional Raman hyperspectral image (x, y, and λ), the cells were scanned by the line-shaped beam with a scan step of 440 nm that corresponds to the pixel size of Raman images along the slit. The laser intensity of the beam at the sample was set at 3.2 $mW/\mu m^2$.

We measured the Raman spectrum of ethanol and calibrating the wavenumber axis by seven bands of the ethanol spectrum.[33] A least-square third-order polynomial function was obtained for approximating the relationship between detector position and wavenumber of the

reference bands. After the wavenumber calibration, we calibrated the spectral responses of detection with a spectrum of light from a standard lamp.

Prior to Raman spectral analysis, the pixels exposed to cosmic rays were processed by 2D median filtering. After the cosmic ray removal, we applied interpolation and decimation to the spectra so that the spectra obtained on different days can be compared along the exactly same wavenumber axis (pixels). Spline curves were used for the interpolation and decimation. For correcting the baseline, we applied iterative alternative-least-square polynomial function fitting to the individual spectra. From the baseline corrected spectra, we reconstructed the Raman images.

From the reconstructed Raman images, we analyzed the mean spectra of nuclei, cytoplasm, and whole cell. Individual mean spectra were normalized by the L2 norms of the wavenumber region of interests.

We applied principal component analysis to the set of normalized spectra. The regression coefficients C_{PC} were calculated by the following equation,

$$C_{PC} = S_{sample} \cdot S_{PC}^{T} \cdot (S_{PC} \cdot S_{PC}^{T})^{-1},$$

where S_{sample} was a spectrum set of MCF-7 or MCF-10A, and S_{PC} was the component loading used.[24] We used the obtained coefficients for classifying MCF-7 and MCF-10A. A linear discriminant curve was drawn at the line where values of probability density of individual datasets equaled each other.

We also calculated the signal-to-background ratio (SBR) of the Raman hyperspectral dataset obtained in the wideband, narrowband and low-spectral-resolution conditions by dividing

the root mean square of the intensity at the region of 1432-1458 cm^{-1} for the pixels corresponding to regions of cells (n > 50) by that of non-cell regions.

Further details of experiments and data analyses can be found in Supporting Information.

Results

Figure 1(a) shows typical nucleus and cytoplasm spectra. The fingerprint region (728-1786 cm^{-1}) and CH region (2823-3017 cm^{-1}) contain different characteristic Raman bands. A band assigned to CH_2 stretching mode at 2856 cm^{-1} represents lipid because CH_2 is largely contained in phospholipid.[16] Lipid also contributes to 1450 cm^{-1} of CH_2 vibrations.[16] There is a contribution of lipid to the 1660 cm^{-1} band assigned to C=C stretching mode.[16] The 1450 and 1660 cm^{-1} bands can also be assigned to protein.[16,35] Protein is also represented by the CH_3 stretching mode at 2941 cm^{-1} and the ring-breathing mode of phenylalanine at 1005 cm^{-1}.[1,35] These molecules are observed in both the nucleus and cytoplasm spectra. In contrast, reduced forms of cytochromes[16] are seen only in the cytoplasm spectrum. In this experiment, the signal from the reduced cytochromes were localized in mitochondria. Typical distributions of the reduced cytochrome (749-760 + 1125-1138 cm^{-1}) in MCF-7 and MCF-10A are shown in **Fig. 1(b)**, together with those of lipid (2851-2860 cm^{-1}) and protein (2924-2946 cm^{-1}). MCF-7 cells contain more cytochrome and protein than MCF-10A cells, while MCF-10A cells contain more lipid than MCF-7 cells. **Figure 1(c)** shows average spectra of nucleus, cytoplasm, and whole cell (nucleus + cytoplasm) regions for two cell types, as well as the averages plus/minus standard deviations. In the cytoplasm spectra, the cytochrome bands at 755 and 1133 cm^{-1} and the protein band at

2941 cm^{-1} are more intense for MCF-7 than MCF-10A. There are also other wavenumber regions where two cell types show differences in the spectra.

To classify two cell types, we first performed principal component analysis (PCA) in order to highlight the variation of the spectral data. We used the loadings that account 90% of the total variance of individual spectra sets by considering the cumulative portions to account for the variance within each dataset (**Fig. S1**). The first to sixth loadings account for 90% of the total variance of the nucleus spectra within all the MCF-7 and MCF-10A cells. Similarly, the first to third loadings explain 90% of the total variance of cytoplasm and whole cell spectra. These loadings and their scores for all the spectra are shown also in **Fig. S1.**

We performed linear discriminant analysis using the selected principal component loadings. **Table 1** summarizes the results. For all cell regions, true classifications exceed 90% of all the 112 spectra.

In the experiments mentioned above, Raman hyperspectral images containing 300,000 spectra were obtained in 24 minutes. Each image typically contains 30 and 10 cells for MCF-7 and MCF-10A respectively. The throughput of cell classification was about 1 cell/min. To increase the throughput of Raman imaging of cells, we reduced the number of readout pixels of the detector. The number of readout pixels for one-dimensional hyperspectral imaging (x and λ) was 1,440,000. At a readout rate of 1 MHz, the readout time reached as long as 1.44 seconds. This accounts for 22% of all the acquisition time (signal accumulation time of 5 seconds + the readout time of 1.44 seconds).

In principle, if we reduced the number of readout pixels, the readout time could become shorter. This approach becomes realistic if we can classify two cell types by using narrowband

spectra. To confirm this approach, we used narrowband spectra chosen from the wideband spectra obtained in the experiments and performed principal component and regression analysis with it. The results are summarized in **Table S1**. The number of principal components used was determined accounting for 90% of the variance of the dataset. The accuracies of cell classifications using the narrowband spectra (e.g., 1009-1117 cm^{-1} for the nucleus, 1073-1181, 1333-1441, 1527-1635, and 2845-2953 cm^{-1} for the cytoplasm, and 1073-1181, 1397-1505, 1549-1657, and 2845-2953 cm^{-1} for the whole cell) were similar to or even better than the results using the wideband spectra. These results indicate that narrowband spectra contain information sufficient to distinguish two cell types.

Based on the results in **Table S1**, we determined candidate wavenumber regions for narrowband Raman hyperspectral imaging aimed at high-throughput cell classification. We focused on wavenumber regions where accuracies were higher than 95% for any of three cell regions. Such wavenumber regions include 749-857, 987-1203, 1289-1700, and 2823-3017 cm^{-1}. Among them, we exclude the following specific regions from the analysis. The signals at the regions of 749-857, and 1527-1700 cm^{-1} are significantly influenced by Raman scattering of glass, water and/or oxygen, respectively. Additionally, several narrow regions, such as 1019-1029, 1182-1203, 1502-1526, and 2823-2841 cm^{-1}, do not contain significant cell signals, which are relevant from the Raman images (**Fig. S2**). Hence, we chose 1030-1181, 1289-1441, 1397-1501, and 2842-3017 cm^{-1} for the narrowband spectral analysis to achieve high-throughput Raman imaging. These regions contain remarkable bands at 1133 (cytochrome), 1342 (adenine and guanine),[11] 1450 (lipid and protein), 2856 (lipid) and 2941 cm^{-1} (protein). Raman images of these bands are shown in **Figs. 1(b) and S2**.

We also investigated the effects of baseline subtraction to cell classification. In narrowband measurements, a baseline has to be estimated in a narrow wavenumber region, which may provide a result different from the actual baseline. Because the widths of spectra in the regions (1030-1181, 1289-1441, 1397-1501, and 2842-3017 cm^{-1}) are far narrower than that of a typical fluorescence spectrum (> 1000 cm^{-1}), a baseline estimated from high-order (e.g., sixth-order) polynomial function is not adequate for the narrowband spectrum, and consequently Raman bands are denatured after baseline subtraction. We compared the spectral shape between a narrowband spectrum whose baseline was estimated from the wideband spectrum by using the sixth-order polynomial fitting function and one whose baseline was estimated from the narrowband spectrum. Because the bandwidth is small, we used a linear function for estimating baselines from the narrowband spectra. In order to quantify the differences, we calculated the Euclidean distance between two spectra. Prior to this calculation, we subtracted minimal values from individual spectra and normalized the bias-corrected spectra by their L2 norms to make a fair comparison between spectral shapes.

The spectra after baseline subtraction are shown in **Fig. 2(a)**. The spectra at 2842-3017 cm^{-1} are clearly different from the spectra whose baselines were estimated through sixth-order polynomial function fitting to the wideband spectra (Fig. 1(c)). **Figure 2(b)** plots the Euclidean distances for all the 336 spectra from 3 cell regions of 112 cells, as well as their averages and standard deviations. A large difference of the spectra in 2842-3017 cm^{-1} between two baseline subtraction methods was well reflected by this plot. This plot also indicates that the narrow-region baseline subtraction in 1397-1501 cm^{-1} provides spectra, which resemble the ones processed by the wideband baseline subtraction. This region is rich in information on the species of endogenous molecules. It contains remarkable signals from both protein and lipid as shown in

Fig. S2. A weak band at 1427 cm^{-1}, which can be assigned to adenine and guanine,[36] is also found in this region.

We performed line-illumination Raman imaging of MCF-7 and MCF-10A cells in order to demonstrate high-throughput Raman imaging with narrowband detection. We have set the region of interest of the detector to measure 1397-1501 cm^{-1} for acquiring a narrowband Raman hyperspectral image. The image acquisition time (1200 × 250 pixels) was reduced to 15 minutes. **Figure 3(a)** shows typical Raman images (1200 × 250 pixels) reconstructed by average intensity of 1432-1458 cm^{-1} for MCF-7 and MCF-10A. Because the intensity of this region is attributable to lipid and protein, [16] the images reflect the distributions of lipid and protein, highlighting the nucleus and cytoplasm clearly. The representative spectra of nucleus and cytoplasm are shown in **Fig. 3(b)**. They exhibit different characteristics although they look quite similar. The width of the main band at 1450 cm^{-1} is broader for the cytoplasm spectrum than that of the nucleus spectrum. The cytoplasm spectrum exhibits a shoulder band at around 1440 cm^{-1}. We found that this band was intense for lipid-rich regions, which are only found in the cytoplasm but not in the nucleus. In the nucleus spectrum, this shoulder band is not observed. In contrast, the nucleus spectrum shows a weak band at 1427 cm^{-1}. This band can be assigned to nucleotide bases,[36] which are contained densely in the nuclei.

We obtained mean spectra of nucleus and cytoplasm from Raman images of MCF-7 (54 cells) and MCF-10A (53 cells). Averages and standard deviations of these spectra are shown in **Fig. 3(c).** The nucleus spectra of MCF-7 and MCF-10A are quite similar to each other. In contrast, the cytoplasm and whole cell spectra of MCF-10A are broader than those of MCF-7. We performed principal component analysis to the nucleus, cytoplasm, and whole cell spectra. The results are summarized in **Table 2**. As anticipated, the nucleus spectra did not well

discriminate MCF-7 and MCF-10A, while the cytoplasm and whole cell spectra did well. Accurate classification of two cell types with cytoplasm and whole cell spectra needed the same number of loadings as used in the analysis whose results are shown in **Table S1**. Particularly, with the cytoplasm spectra only the first and second loadings discriminated well the two cell types (Table 2). These loadings and their scores for all the spectra are shown in **Fig. 3(d)**. Two cell types are classified along the first loading. The first loading has one positive and two negative peaks. All the MCF-7 spectra have negative scores of the first loading while the MCF-10A spectra are distributed around zero. These results indicate that all the MCF-7 spectra have larger differences between the positive and negative peaks than the MCF-10A spectra. This indication matches with the tendency that the average spectra from MCF-10A are broader than that from MCF-7.

For the narrowband Raman imaging, the signal accumulation time was reduced from 5 to 3 seconds without sacrificing the SBR **(Table 3)** since the readout noise was reduced by setting the readout rate at 100 kHz. Because of the small number of readout pixels along the wavenumber axis (50 pixels), the readout of 1200 spectra acquired by a single exposure took only 0.6 seconds for the narrowband measurement. This was far shorter than the readout time for acquiring 1200 wideband spectra at 100 kHz (14.4 sec) and was even shorter than that at 2 MHz (0.72 sec). The readout time at 2 MHz for acquiring 1200 narrowband spectra was much shorter (0.03 seconds), but the signal accumulation time for acquiring sufficient signal was much longer than this time, and narrowband imaging does not do much to accelerate measurement at 2 MHz. Rather than using the 2 MHz readout rate, the 100 kHz readout rate and 3 seconds exposure spectral measurement effectively accelerate imaging by 1.6 times, gaining an even higher SBR in comparison to the 2 MHz readout rate and 5 seconds exposure spectral measurement.

To further reduce the number of readout pixels, we applied binning to the detector along the wavenumber axis. By applying binning to each 4 pixels, we reduced the number of pixels for acquiring 1200 narrowband spectra at 1397-1501 cm^{-1} from 1200 × 50 to 1200 × 13 pixels, resulting in the reduction of the image acquisition time from 15 to 13.2 minutes. In addition, a large detector binning is advantageous in improving SBR because the number of photons detected is increased in a single binned pixel in comparison to a non-binned pixel while the readout noise is maintained. With binning of 4 pixels, using a shorter exposure time of 0.7 seconds (less than 25% of the non-binning condition) even increased SBR as shown in Table 3. As a result, the image acquisition time (1200 × 250 pixels) was reduced from 15 to 3.6 minutes, and hence, a wide-field-of-view Raman image (1200 × 1500 pixels) was taken in only 21 minutes. **Figure 4(a)** shows the representative wide-field-of-view Raman images of MCF-7 and MCF-10A obtained with the narrowband detection and the pixel binning.

One of the disadvantages in using pixel binning is to sacrifice the spectral resolution. It was recently reported that low spectral resolution is not an issue in classifying cells[30,31] and tissues.[24,32] We also evaluated the capability of the low-spectral-resolution detection for classifying MCF-7 and MCF-10A. The results of principal component regression analysis for cytoplasm spectra of MCF-7 (100 cells) and MCF-10A (100 cells) are shown in Fig. 4(b). The score distributions and loadings are similar to those of the dataset without the pixel binning (Fig. 3(d)). The linear discriminant analysis using the first and second loading scores classified the two cell types with accuracy of 90.5%.

Discussion

The presented approach can be optimized for improving the image acquisition speed further. One way is to expand the detection region along the entrance slit of the spectrometer. In our current setup, the vertical pixels of the CCD camera were not fully used due to the limitation of illumination and imaging optics (only 1200 of 2048 pixels were used). By using all the pixels of the CCD camera along the slit direction, the number of cells measured per unit time can be increased by 71% whereas an increase in the image acquisition speed is only 13%. It is also possible to extend the area for simultaneous Raman measurement in the direction perpendicular to the slit. Since there are unused detector areas in the narrowband and/or low-spectral-resolution detection modes, we can increase the number of line-illumination in order to enhance the parallel detection capability of Raman spectra.[38] Assuming that 20 illumination lines are used, the two-dimensional Raman hyperspectral image with 1500 lines can be acquired in 4.8 minutes while maintaining the SBR the same as the experimental results in Fig. 4(a), where the total throughput for cell imaging can be increased up to 50 cells/min.

The technique demonstrated in the article can be further enhanced for high-throughput cell classification or screening. In the experiments, we set the number of pixels for binning to 4. This value can be increased as long as cells could be classified. At a larger binning value, we can further decrease the number of readout pixels, i.e., readout time. Additionally, since the number of photons incident to a binned pixel of the detector is increased, the signal accumulation time can also be reduced without diminishing SBR. In considering SBR, we need to also consider the thermal noise that is accumulated by the pixel binning of the detector, but in the experiments the thermal noise was negligibly small with an exposure time of several seconds or less. The SBR can be also improved by increasing the excitation intensity. Although the increase in excitation intensity is limited by cell damage, there is still room to explore since we were able to decrease

the signal accumulation time from 5 to 0.7 seconds at the narrowband and low-spectral-resolution measurement, and for such a short-time exposure, cells can tolerate higher irradiation intensity.

Another approach to increase the throughput is through reduction of the spatial resolution. In the experiments, we used a 25× objective lens and the pixel sizes of the images were 440 nm. To measure a biological cell, of which the size is about 10 μm in diameter, a larger pixel size of an image (e.g., 1 μm) can still show variations of spectra in the intracellular space that can provide information on cell function as well as cell state and cell type. With twice lower spatial resolution, 4-times larger area can be analyzed in a single Raman imaging acquisition. Similarly, sparse sampling can also accelerate measurement.[24,39,40] One possible drawback of low-spatial-resolution and sparse sampling measurements is that the spatial information is sacrificed. It is essential to optimize a spatial resolution for the sample of interest. To aid this issue, sparse sampling Raman microscopy was recently combined with a compressed sensing algorithm so that the loss of spatial information was suppressed while the volume of the data acquired was reduced.[41]

The wavenumber range, the spectral resolution, and the spatial resolution should be carefully chosen to obtain an accurate result in cell classification. To select a wavenumber region, the accuracy of background subtraction and contributions of signals from non-cell components such as optics, substrate for cell culture, and water in a sample should be important. As shown in our experiments, it is possible to minimize contributions of the spectra from non-cell components and attain reliability in the cell classification using narrowband spectra. Indeed, we chose 1397-1501 cm^{-1} because this region is not influenced by water, substrate, and optical components. In

contrast, the wideband spectra analysis actually was influenced by non-cell components. The first loadings of principal component analysis of the wideband spectra show negative bands at 770-830 and 1570-1710 cm^{-1} assigned to optical components and water, respectively (Fig. S1). The first loadings reflect the ratio of the signal from cell components to contributions from non-cell components. A similar issue can be found in analyzing the CH region at 2800-3200 cm^{-1} because the spectra in this wavenumber region are influenced by the intense and complicated water bands. To use this wavenumber, the spectral baseline attributable to water has to be correctly compensated by measuring a relatively wider wavenumber region (e.g., 400 cm^{-1}).

Conclusions

We successfully classified two cell types by narrowband and low-spectral-resolution imaging at 1397-1501 cm^{-1}. This approach improved the throughput of MCF-7/MCF-10A classification by spontaneous Raman hyperspectral imaging up to 10 cells/min with high spatial resolution.

The presented and discussed approaches to high-speed Raman hyperspectral imaging will boost practical uses of Raman spectral analysis and imaging in biology and medicine for label-free cell and tissue diagnosis[8,39,40,42] as well as a variety of label-free assays such as evaluation of differentiation/undifferentiation,[4,6,7] activation/inactivation,[41,43] and viability[44,45] of cells. For such practical applications, a higher imaging speed is demanded. Further acceleration of Raman imaging and classification can be realized if above-mentioned approaches were combined with coherent Raman imaging techniques, which have enabled fast Raman imaging[18-21] but difficult to obtain high spectral resolution and wideband spectra.

SUPPORTING INFORMATION

The Supporting Information is available free of charge on the ACS Publications website at DOI: xxxxxxxxxxxx. Details of materials and methods, principal component analysis of the wideband spectra dataset (Fig. S1), cell classification using narrowband spectra in various regions (Table S1), and Raman images in candidate regions for narrowband Raman imaging (Fig. S2).

AUTHOR INFORMATION

Corresponding Author

*E-mail: fujita@ap.eng.osaka-u.ac.jp. Fax: +81 6879-7330. Phone: +81 6879-7847.

Author Contributions

Y.K., K.M., and K.F. designed research; Y.K., K.M., and K.H. prepared samples and performed experiments; Y.K. analyzed data; Y.K., H.T., Y.H., and K.F. wrote the paper.

Funding Sources

This work was partially supported by JST/CREST Grant Number JPMJCR1662, Japan (to K.F. and Y.H.) and the Nakatani Foundation (to K.F.).

Notes

The authors declare no competing financial interest.

ACKNOWLEDGMENT

We thank Dr. Tamiki Komatsuzaki of Research Institute for Electronic Science, Hokkaido University, for his advices on the data analysis.

REFERENCES

1. Kumamoto, Y.; Harada, Y.; Takamatsu, T.; Tanaka, H. Label-free molecular imaging and analysis by Raman spectroscopy. *Acta Histochem. Cytochem.* **2018**, *51*, 101-110.

2. Yazdi, Y.; Ramanujam, N.; Lotan, R.; Mitchell, M. F.; Hittelman, W.; Richards-Kortum, R. Resonance Raman spectroscopy at 257 nm excitation of normal and malignant cultured breast and cervical cells. *Appl. Spectrosc.* **1999**, *53*, 82-85.

3. Maquelin, K.; Maquelin, K.; Choo-Smith, L. P.; van Vreeswijk, T.; Endtz, H. P.; Smith, B.; Bennett, R.; Bruining, H. A.; Puppels, G. J. Raman spectroscopic method for identification of clinically relevant microorganisms growing on solid culture medium. *Anal. Chem.* **2000**, *72*, 12-19.

4. Chan, J. W.; Lieu, D. K.; Huser, T.; Li, R. A. Label-free separation of human embryonic stem cells and their cardiac derivatives using Raman spectroscopy. *Anal. Chem.* **2009**, *81*, 1324-1331.

5. Hartmann, K.; Becker-Putsche, M.; Bocklitz, T.; Pachmann, K.; Niendorf, A.; Rösch, P.; Popp, J. A study of Docetaxel-induced effects in MCF-7 cells by means of Raman microspectroscopy. *Anal. Bioanal. Chem.* **2012**, *403*, 745-753.

6. 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.

7. Ilin, Y.; Choi, J. S.; Harley, B. A. C.; Kraft, M. L. Identifying states along the hematopoietic stem cell differentiation hierarchy with single cell specificity via Raman spectroscopy. *Anal. Chem.* **2015**, *87*, 11317-11324.

8. Manago, S.; Valente, C.; Mirabelli, P.; Circolo, D.; Basile, F.; Corda, D.; De Luca, A. C. A reliable Raman-spectroscopy-based approach for diagnosis, classification and follow-up of B-cell acute lymphoblastic leukemia. *Sci. Rep.* **2016**, *6*, 24821.

9. Hashimoto, K.; Andriana, B. B.; Matsuyoshi, H.; Sato, H. Discrimination analysis of excitatory and inhibitory neurons using Raman spectroscopy. *Analyst* **2018**, *143*, 2889-2894.

10. Puppels, G. J.; Grond, M.; Grave, J. Direct imaging Raman microscope based on tunable wavelength excitation and narrow-band emission detection. *Appl. Spectrosc.* **1993**, *47*, 1256–1267.

11. Palonpon, A. F.; Sodeoka, M.; Fujita, K. Molecular imaging of live cells by Raman microscopy. *Curr. Opin. Chem. Biol.* **2013**, *17*, 708-715.

12. 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* **2012**, *109*, 28-32.

13. Huang, C. K.; Hamaguchi, H.; Shigeto, S. In vivo multimode Raman imaging reveals concerted molecular composition and distribution changes during yeast cell cycle. *Chem. Commun.* **2011**, *47*, 9423-9425.

14. Kumamoto, Y.; Taguchi, A.; Smith, N. I.; Kawata, S. Deep ultraviolet resonant Raman imaging of a cell. *J. Biomed. Opt.* **2012**, *17*, 076001.

15. Chernenko, T.; Sawant, R. R.; Miljkovic, M.; Quintero, L.; Diem, M.; Torchilin, V. Raman microscopy for noninvasive imaging of pharmaceutical nano carriers: Intracellular distribution of cationic liposomes of different composition. *Mol. Pharmaceutics* **2012**, *2*, 930-936.

16. 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.

17. Ando, J.; Palonpon, A. F.; Sodeoka, M.; Fujita, K. High-speed Raman imaging of cellular processes. *Curr. Opin. Chem. Biol.* **2016**, *33*, 16-24.

18. Zumbusch, A.; Holtom, G. R.; Xie, X. S. Three-dimensional vibrational imaging by coherent anti-Stokes Raman scattering. *Phys. Rev. Lett.* **1999**, *82*, 4142-4145.

19. Evans, C. L.; Potma, E. O.; Puoris'haag, M.; Côté, D.; Lin, C. P.; Xie, X. S. Chemical imaging of tissue in vivo with video-rate coherent anti-Stokes Raman scattering microscopy. *Proc. Natl. Acad. Sci.* **2005**, *102*, 16807-16812.

20. Saar, B. G.; Freudiger, C. W.; Reichman, J.; Stanley, C. M.; Holtom, G. R.; Xie, X. S. Video-rate molecular imaging in vivo with stimulated Raman scattering. *Science* **2010**, *330*, 1368-1370.

21. Ozeki, Y.; Umemura, W.; Otsuka, Y.; Saoh, S.; Hashimoto, H.; Sumimura, K.; Nishizawa, N.; Fukui, K.; Itoh, K. High-speed molecular spectral imaging of tissue with stimulated Raman scattering. *Nat. Photon.* **2012**, *6*, 845-851.

22. Inoué, S. Foundations of confocal scanned imaging in light microscopy. *Handbook of biological confocal microscopy*, 3rd edition; Springer US: Boston, MA, **2006**; 1-19.

23. Han, Y.; Zhang, A. C.; Lo, Y. H. Review: Imaging technologies for flow cytometry. *Lab. Chip* **2016**, *16*, 4639-4647.

24. Kumamoto, Y.; Harada, Y.; Tanaka, H.; Takamatsu, T. Rapid and accurate peripheral nerve imaging by multipoint Raman spectroscopy. *Sci. Rep.* **2017**, *7*, 845.

25. Huang, W.; Wu, S.; Chen, M.; Sun, L.; Li, Y.; Huang, M.; Huang, S.; Xu, Z.; Chena, R.; Zeng, H. Study of both fingerprint and high wavenumber Raman spectroscopy of pathological nasopharyngeal tissues. *J. Raman Spectrosc.* **2015**, *46*, 537-544.

26. Talari, A. C. S.; Evans, C. A.; Holen, I.; Coleman, R. E.; Rehman, I. U. Raman spectroscopic analysis differentiates between breast cancer cells. *J. Raman Spectrosc.* **2015**, *46*, 421-427.

27. Zhou, X.; Dai, J.; Chen, Y.; Duan, G.; Liu, Y.; Zhang, H.; Wu, H.; Peng, G. Evaluation of the diagnostic potential of ex vivo Raman spectroscopy in gastric cancers: fingerprint versus high wavenumber. *J. Biomed. Opt.* **2016**, *21*, 105002.

28. Duraipandian, S.; Zheng, W.; Ng, J.; Low, J. J. H.; Ilancheran, A.; Huang, Z. In vivo diagnosis of cervical precancer using Raman spectroscopy and genetic algorithm techniques. *Analyst* **2011**, *136*, 4328-4336.

29. Zhao, J.; Zeng, H.; Kalia, S.; Lui, H. Wavenumber selection based analysis in Raman spectroscopy improves skin cancer diagnostic specificity. *Analyst* **2016**, *141*, 1034-1043.

30. Schie, I. W.; Krafft, C.; Popp, J. Cell classification with low-resolution Raman spectroscopy (LRRS). *J. Biophoton.* **2016**, *9*, 994-1000.

31. Schie, I. W.; Rüger, J.; Mondol, A. S.; Ramoji, A.; Neugebauer, U.; Krafft, C.; Popp, J. High-throughput screening Raman spectroscopy platform for label-free cellomics. *Anal. Chem.* **2018**, *90*, 2023-2030.

32. Chen, H.; Li, X.; Broderick, N.; Liu, Y.; Zhou, Y.; Han, J.; Xu, W. Identification and characterization of bladder cancer by low-resolution fiber-optic Raman spectroscopy. *J. Biophoton.* **2018**, *11*, e201800016.

33. Palonpon, A. F.; Ando, J.; Yamakoshi, H.; Dodo, K.; Sodeoka, M.; Kawata, S.; Fujita, K. Raman and SERS microscopy for molecular imaging of live cells. *Nat. Protoc.* **2013**, *8*, 677-692.

34. Lieber, C. A.; Mahadevan-Jansen, A. Automated method for subtraction of fluorescence from biological Raman spectra. *Appl. Spectrosc.* **2003**, *57*, 1363-1367.

35. Ishigaki, M.; Hashimoto, K.; Sato, H.; Ozaki, Y. Non-destructive monitoring of mouse embryo development and its qualitative evaluation at the molecular level using Raman spectroscopy. *Sci. Rep.* **2017**, *7*, 43942.

36. Azan, A.; Untereiner, V.; Gobinet, C.; Sockalingum, G. D.; Breton, M.; Piot, O.; Mir. L. M. Demonstration of the protein involvement in cell electropermeabilization using confocal Raman microspectroscopy. *Sci. Rep.* **2017**, *7*, 40448.

37. Zhang, C.; Huang, K. C.; Rajwa, B.; Li, J.; Yang, S.; Lin, H.; Liao, C. S.; Eakins, G.; Kuang, S.; Patsekin, V.; Robinson, J. P.; Cheng, J. X. Stimulated Raman scattering flow cytometry for label-free single-particle analysis. *Optica* **2017**, *4*, 103-109.

38. Mochizuki, K.; Maeda, S.; Smith, N. I.; Fujita, K. Accelerating Raman imaging by Multi-line illumination. *Proceedings of Focus on Microscopy* **2018**, Singapore.

39. Kong, K.; Rowlands, C. J.; Varma, S.; Perkins, W.; Leach, I. H.; Koloydenko, A. A.; Williams, H. C.; Notingher, I. Diagnosis of tumors during tissue-conserving surgery with integrated autofluorescence and Raman scattering microscopy. *Proc. Natl. Acad. Sci. USA* **2013**, *110*, 15189-15194.

40. Sinjab, F.; Kong, K.; Gibson, G.; Varma, S.; Williams, H.; Padgett, M.; Notingher, I. Tissue diagnosis using power-sharing multifocal Raman micro-spectroscopy and auto-fluorescence imaging. *Biomed. Opt. Express* **2016**, *7*, 2993-3006.

41. Pavillon, N.; Smith, N. I. Compressed sensing laser scanning microscopy. *Biomed. Opt. Express* **2016**, *24*, 30038-30052.

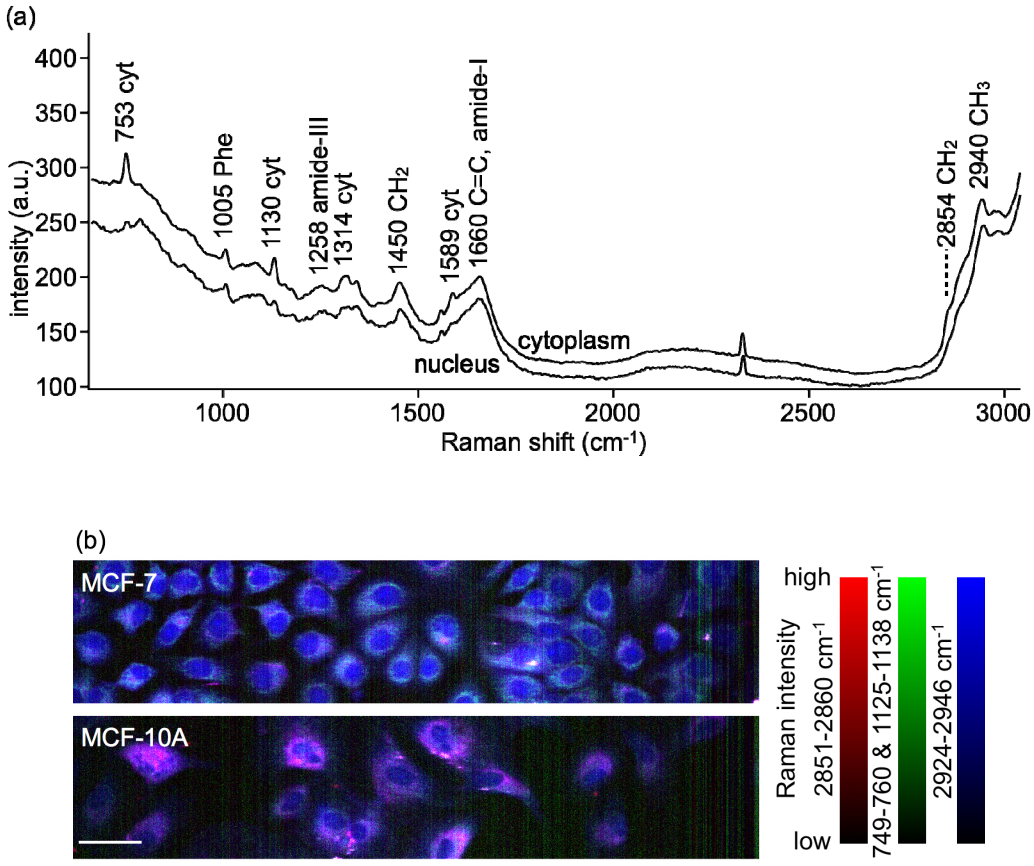
42. Kong, K.; Kendall, C.; Stone, N.; Notingher, I. Raman spectroscopy for medical diagnostics — From in-vitro biofluid assays to in-vivo cancer detection. *Adv. Drug Deliv. Rev.* **2015**, *89*, 121-134.

43. 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.

44. Verrier, S.; Notingher, I.; Polak, J. M.; Hench, L. L. In situ monitoring of cell death using Raman microspectroscopy. *Biopolymers* **2004**, *74*, 157-162.

45. Ohira, S.; Tanaka, H.; Harada, Y.; Minamikawa, T.; Kumamoto, Y.; Matoba, S.; Yaku, H.; Takamatsu, T. Label-free detection of myocardial ischaemia in the perfused rat heart by spontaneous Raman spectroscopy. *Sci. Rep.* **2017**, *7*, 42401.

FIGURES



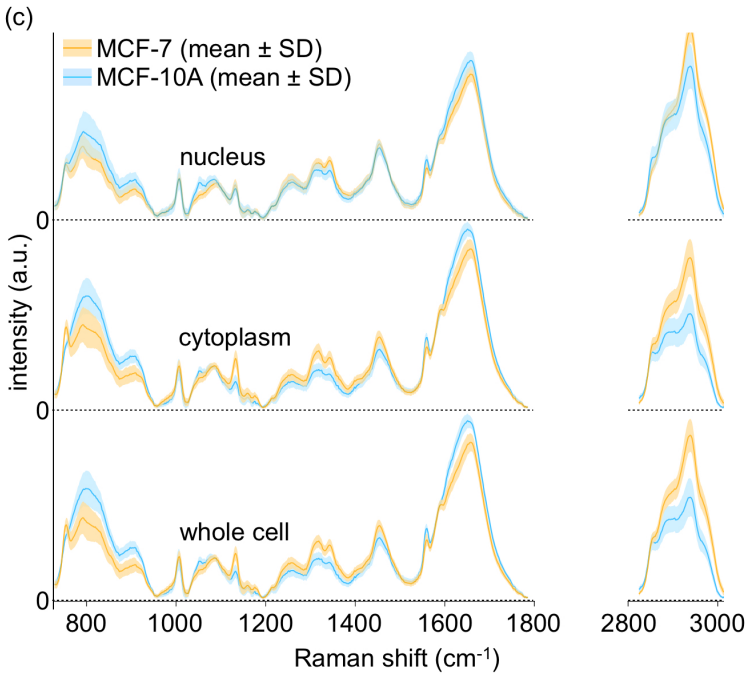


Figure 1. Experimental results of Raman hyperspectral imaging of MCF-7 and MCF-10A. (a) Representative spectra of the cell lines. Assignments and wavenumbers of remarkable vibration bands are shown. (b) Representative Raman images (1200 × 250 pixels) reconstructed with intensity of cytochrome, lipid, and protein bands. The scale bar shows 50 μm. In producing these images, singular value decomposition (SVD) was not applied to the spectra for removing the noise. (c) Mean spectra of the nucleus, cytoplasm, and whole cell of 60 MCF-7 and 52 MCF-10A and their standard deviations.

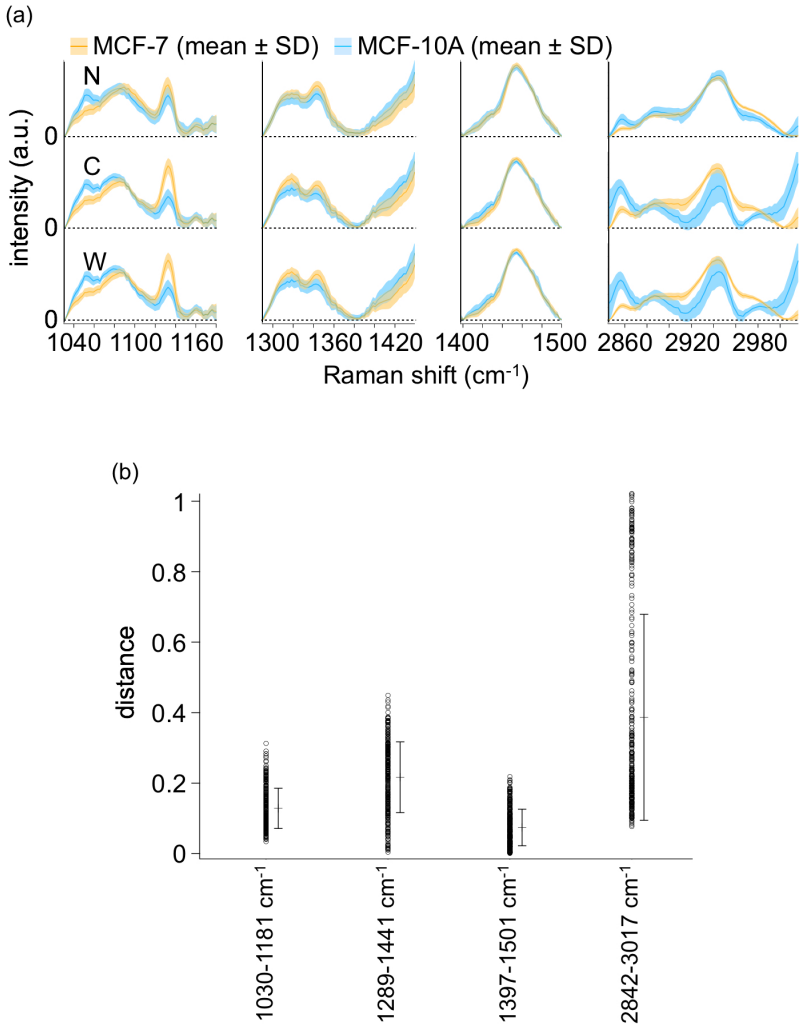


Figure 2. (a) Raman spectra of candidate regions for narrowband imaging. The spectra were shown after subtraction of the baseline estimated by fitting a linear function. N, C, and W indicate nucleus, cytoplasm, and whole cell, respectively. (b) Euclidean distances between the spectra after base-line corrections using polynomial (wideband) and linear (narrow region) functions, which were calculated using the 336 spectra from the 112 different cells. Averages and standard deviations of the plots are also shown.

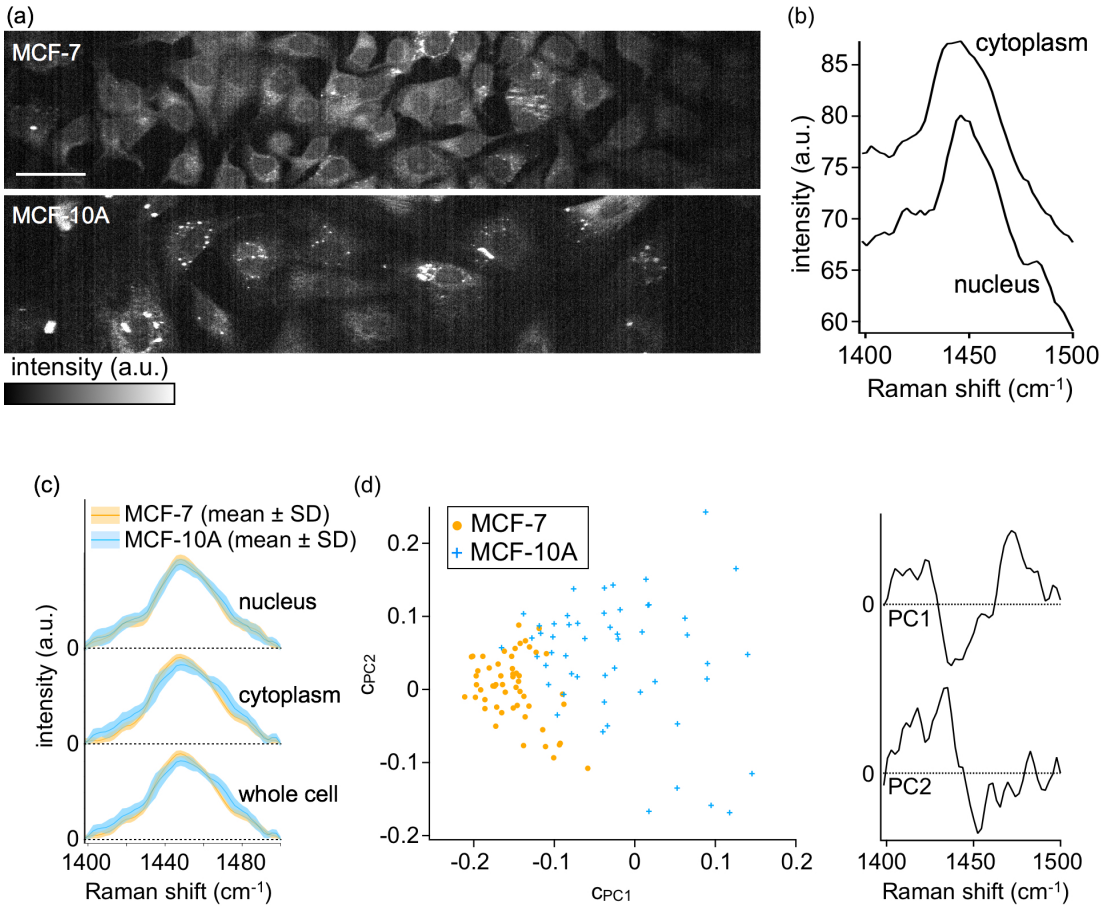


Figure 3. Narrowband imaging with the detection window of 1397-1501 cm^{-1} for high-throughput classification of MCF-7 and MCF-10A. (a) Typical Raman images (1200 × 250 pixels) reconstructed by the average intensity of 1432-1458 cm^{-1}. The scale bar shows 50 μm. In producing these images, SVD was not applied to the spectra. (b) Representative spectra of nucleus and cytoplasm of the cell lines. (c) Mean spectra of the nucleus, cytoplasm, and whole cell of 54 MCF-7 and 53 MCF-10A and their standard deviations. (d) Results of principal component and regression analysis. The regression coefficients to the first and second loadings, c_{PC1} and c_{PC1}, are plotted for all the cytoplasm spectra at the left. At the right, the first and second loadings are shown.

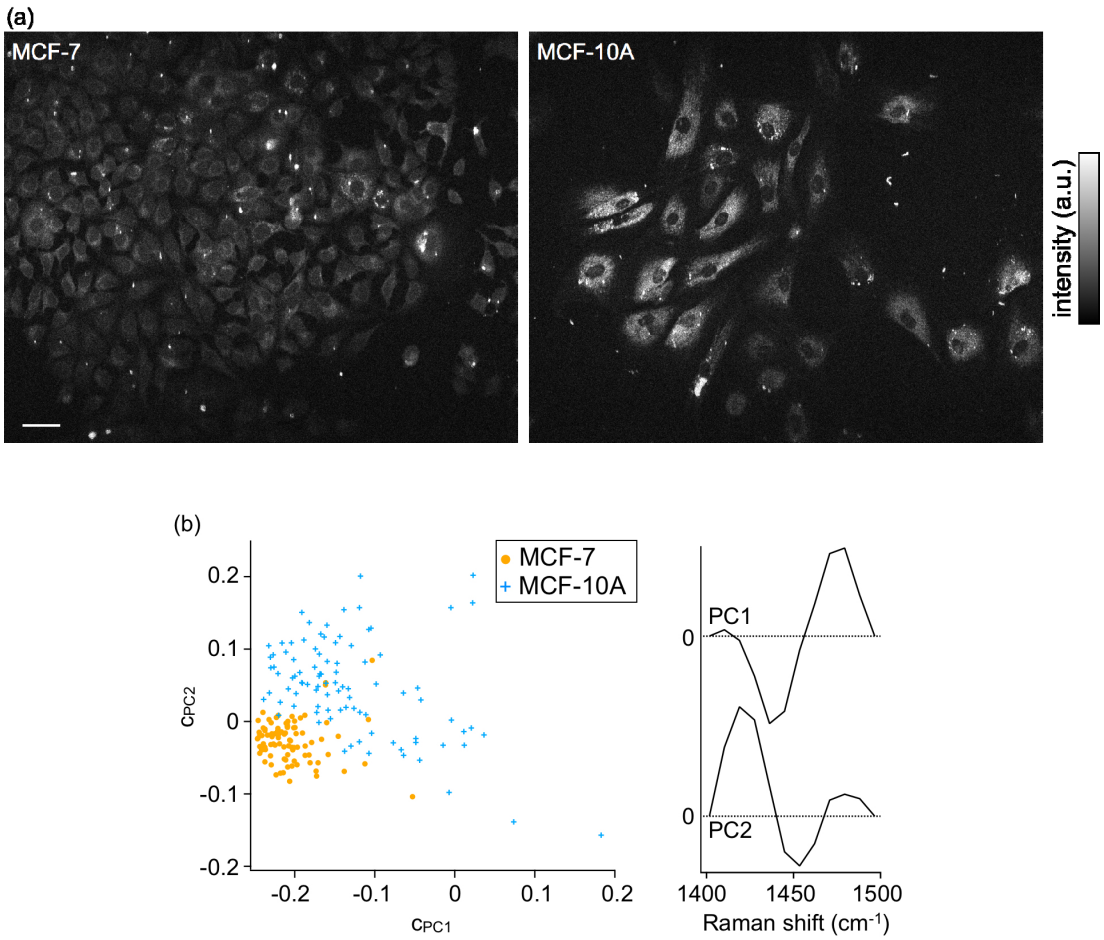


Figure 4. Narrowband (1397-1501 cm^{-1}) and low-spectral-resolution (the binning of 4 pixels) imaging of MCF-7 and MCF-10A. (a) Typical Raman images (1200 × 1500 pixels) reconstructed by average intensity of 1432-1458 cm^{-1}. The scale bar shows 50 μm. In producing these images, SVD was not applied to the spectra. (b) Results of principal component and regression analysis. The regression coefficients to the first and second loadings, c_{PC1} and c_{PC1}, are plotted for all the cytoplasm spectra for 100 MCF-7 and 100 MCF-10A cells at the left. At the right, the first and second loadings are shown.

TABLES

Table 1. Results of principal component and regression analysis for cell classification using wavenumbers in the region: 728-1786 & 2823-3017 cm^{-1}.

cell region	PC used	cumulative portion (%)	TP	FN	TN	FP	(TP+TN)/ (TP+FN+TN+FP)	TP/ (TP+FN)	TN/ (TN+FP)
nucleus	1-6	90.3	60	0	50	2	0.982	1	0.962
cytoplasm	1-3	92.7	56	4	49	3	0.938	0.933	0.942
whole cell	1-3	93.3	57	3	50	2	0.955	0.95	0.962

PC: principal component, TP: true positive, FN: true negative, TN: true negative, and FP: false positive. "Positive" and "negative" means positive and negative predictions of the cancer cell lines, respectively.

Table 2. Results of principal component regression and linear discriminant analysis using the 1397-1501 cm^{-1} spectra measured by narrowband Raman hyperspectral imaging

cell region	PC used	cumulative portion (%)	TP	FN	TN	FP	(TP+TN)/ (TP+FN+TN+FP)	TP/ (TP+FN)	TN/ (TN+FP)
nucleus	1-15	90.4	48	6	44	9	0.860	0.889	0.830
cytoplasm	1-9	91.3	54	0	45	8	0.925	1	0.849
whole cell	1-8	90.5	53	1	48	5	0.944	0.981	0.906
nucleus	1-10	81.8	49	5	41	12	0.841	0.907	0.774
cytoplasm	1-5	82.6	54	0	47	6	0.944	1	0.887
whole cell	1-4	78.7	53	1	43	10	0.897	0.981	0.811
cytoplasm	1-2	68.5	54	0	42	11	0.897	1	0.792

PC: principal component, TP: true positive, FN: true negative, TN: true negative, and FP: false positive. "Positive" and "negative" means positive and negative predictions of the cancer cell lines, respectively.

Table 3. Speed and SBR for wideband, narrowband, and/or low-resolution Raman hyperspectral imaging in different conditions.

spectral region (cm^{-1})	readout rate (MHz)	readout pixels (/line)	readout time (s/line)	exposure time (s/line)	readout + exposure time (m/image)	SBR[a]	readout noise (rms)[b]	Detector binning
1397-1501	0.1	1200 × 50	0.6	3	15	1.77 (MCF-7) 1.43 (MCF-10A)	4.8	1
536-3132	0.1	1200 × 1200	14.4	3	72.5	1.77 (MCF-7) 1.43 (MCF-10A)	4.8	1
536-3132	2	1200 × 1200	0.72	5	23.8	1.52 (MCF-7) 1.28 (MCF-10A)	8.9	1
1397-1501	2	1200 × 50	0.03	5	21.0	1.52 (MCF-7) 1.28 (MCF-10A)	8.9	1
1397-1501	2	1200 × 50	0.03	3	12.6	1.31 (MCF-7) 1.17 (MCF-10A)	8.9	1
1397-1501	0.1	1200 × 13	0.156	0.7	3.6	3.38 (MCF-7) 2.43 (MCF-10A)	4.8	4

[a]The values of SBR for the second and fourth rows were adopted from the first or third rows since the readout rates, exposure times, and detector binning for the second and fourth rows were the same as those of the first or third rows. The value of SBR for the fifth row was calculated by using the value of the fifth row and compensating the difference of the exposure time.

[b]The values of readout noise for the second, fourth, fifth, and sixth rows were adopted from the first or third rows since the readout rates for the second, fourth, fifth, and sixth rows were the same as those of the first or third rows.

TOC graphic