学 位 論 文 内 容 の 要 旨

博士の専攻分野の名称　博士（総合化学）　氏名 アブドル ハリム ブーヤン

学 位 論 文 題 名

Study on Single Cell Raman Analysis to Enhance Differentiability of Cell Types in Non-homogeneous Environments (不均一環境下における細胞識別性向上に関する 1 細胞ラマン解析研究)

Line illumination Raman microscope extracts the underlying spatial and spectral information of a sample, typically, a few hundred times faster than raster scanning. This enables us to explore a wider range of applications including biological samples such as cells and tissues, which requires moderately intense illumination to prevent sample damage. However, a non-uniform intensity distribution of the laser line illumination may induce some artifacts in the data and lower the accuracy of machine learning models trained to predict sample class membership. Here, using cancerous, and normal human thyroid follicular epithelial cell lines, FTC-133 and Nthy-ori 3-1 lines, whose Raman spectral difference is not so large, we showed the standard preprocessing of spectral analyses widely used for raster scanning microscope introduced some artifacts. To address this issue, we proposed a detrending scheme based on random forest regression, a nonparametric model-free machine learning algorithm, combined with position-dependent wavenumber calibration scheme along illumination line. It was shown that the detrending scheme minimizes the artificial biases arising from non-uniform laser source and significantly enhances the differentiability of the sample states, i.e., cancerous or normal epithelial cells, compared to the standard preprocessing scheme.

In **Chapter 1,** the purpose of the dissertation and background of the study are described.

In **Chapter 2,** a brief description of some preliminary concepts like Principal Component Analysis (PCA), Singular Value Decomposition (SVD), Random Forest (RF) etc. are discussed.

In **Chapter 3,** three preprocessing scheme for the Raman data of FTC-133 and Nthy-ori 3-1 cell lines are emphasized.

In **Chapter 4,** a comparative study of these three schemes to differentiate two phenotypes, FTC-133 and Nthy-ori 3-1 are explained in detail. Figure 1 presents a series of visualization and descriptive statistics estimated on representative Raman images preprocessed by the most standard protocol without position-dependent wavenumber calibration (termed “uncalibrated”) with position-dependent wavenumber calibration (termed “calibrated”) and with the detrending scheme on top of the calibration. Panel (A) shows an uncalibrated Raman image, while panel (B) shows the same image with position-dependent wavenumber calibration, which significantly reduces the artificial spatial correlation of the image. Panel (C) shows the Raman image with the detrending scheme on top of the calibration. Panels (D) and (E) show Pearson

correlation coefficients between the Raman image at each individual wavenumber and the illumination axis and the scanning axis, respectively. Panel (F) shows the averaged with one standard deviation Raman spectrum for the cell region with the three preprocessings. Panel D, without wavenumber calibration, shows high positive correlation at certain Raman shifts, high negative correlation at other Raman shifts, and weak correlation for some Raman shifts. The sign of correlation coefficient is dependent on the definition of the coordinate system. Suppose that a Pearson correlation coefficient is positive about +0.8. This is equally possible to be -0.8 if one inverts the axis from positive to negative in the definition of the coordinate system. However, note that the relative relationship, such that some Raman intensities correlate along one direction (e.g., positively) but the others do along the inverse direction (e.g., negatively) to the chosen axis, holds once the coordinate system is fixed. This spatial correlation is significantly reduced by the position-dependent calibration strategy indicating an apparent wavenumber drift along the illumination axis, which could be attributed to chromatic aberration or changes in physical properties resulting from laser light power variation along this axis. Along the scanning axis some correlation pattern also exists for data without wavenumber calibration as observed in Panel E but with lower amplitude than observed for the illumination axis. It should be noted that, in Panel (F), the averaged cell spectra of this Raman image are almost identical to each other, and the difference between the position-dependent wavenumber calibration and the detrending scheme correction is undetectable by visual inspection. However, the Pearson correlation analysis manifests the further reduction of apparent spatial correlation (c.f., the green and blue lines in Panels (D) and (E)). The absolute value of the Pearson correlation coefficient along the illumination axis for the detrending scheme remains very small but non-zero (<<0.05) although that along the

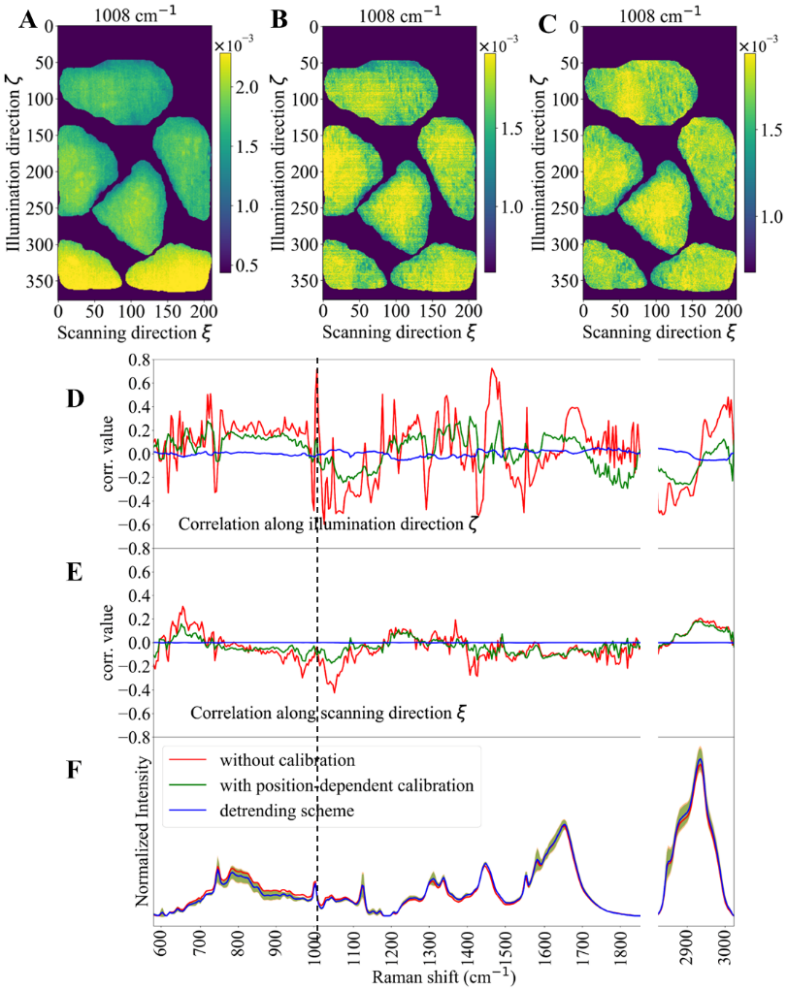

Figure 1: Correlation analysis for three preprocessing schemes.

scanning axis is almost zero (<<0.001). This reflects the fact that the apparent spatial correlation is more pronounced along the illumination axis than the scanning axis.

To evaluate the quality of the three different preprocessing schemes including with/without the position-dependent wavenumber calibration and the detrending scheme on the top of the position-dependent wavenumber calibration, on the Raman images, a comparative analysis of the classification performance of the two cell lines was conducted. Additionally, to provide a visual representation of the effect of the three preprocessing strategies on the data, the average single cell Raman spectra are projected into a low-dimensional space. Figure 2A, B, and C visualize the projection of sixty average single cell spectra in a low-dimensional space by performing classical multidimensional scaling (MDS). This low dimensional representation manifests that the detrending scheme clearly enhances the differentiability between FTC-133 and Nthy-ori 3-1, as shown in Figure 2C (c.f., Figure 2A and B). Figure 2D and E shows the box-and-whisker plot of 25 cross-validated accuracies of random forest classifier models in predicting of FTC-133/Nthy-ori 3-1 for the three different preprocessing schemes. Figure 2D shows the accuracy when considering average single-cell Raman spectra, while Figure 2E is obtained by considering all the spectra belonging to cells. From these figures, it is evident that a proper wavenumber calibration adapted for line-illumination microscopes and/or a detrending scheme is essential to stabilize the performance of the classifiers. Indeed, the average accuracy increases progressively from uncalibrated data to detrended data, while the standard deviation of the accuracy decreases. This trend emphasizes that our preprocessing method improves the stability of random forest classifier by reducing the number of outliers.

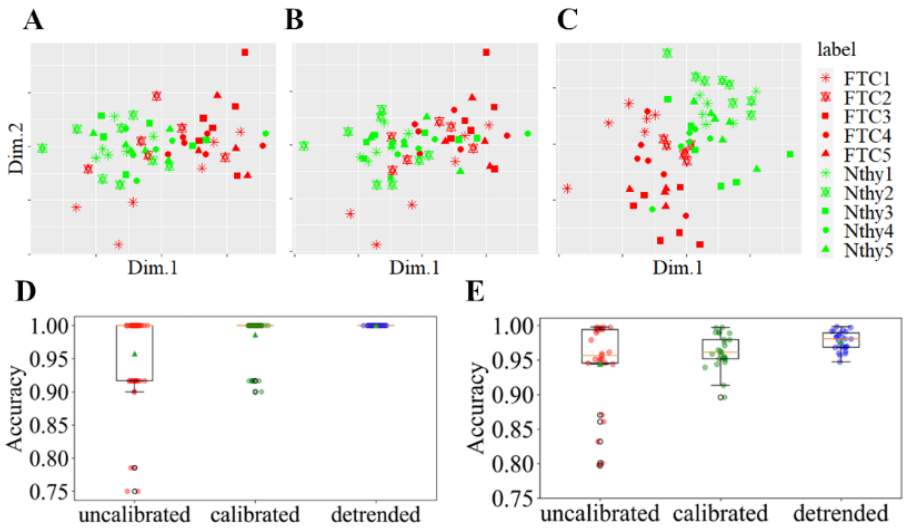

Figure 2: MDS projections and cross validation accuracies of three schemes.

In **Chapter 5**, all the important results are summarized, and future plans are described.