Journal of Raman Spectroscopy / Volume 54, Issue 6 / p. 596-607

RESEARCH ARTICLE | Full Access

Raman identification of single cell component and FMS-like tyrosine kinase 3-internal tandem duplications subtype for clinical acute myeloid leukemia

Yimeng Wang, Yihui Wu, Mingbo Chi, Mingxin Li, Sujun Gao, Long Su, Wei Han

First published: 31 March 2023
https://doi-org.fgul.idm.oclc.org/10.1002/jrs.6523

Abstract

Acute myeloid leukemia (AML) is a malignant clonal blood disease of hematopoietic stem cells with poor prognosis. Traditional diagnosis of AML depends on the morphology, immunology, cytogenetics, and molecular biology (MICM) classification. As a rapid, label-free and non-destructive detection method, Raman spectroscopy can characterize the molecular information in the biochemical process at the molecular level through peak position, intensity and other information. In this paper, we collected leukemic blast cells from 19 AML patients. Firstly, it was proved that with the help of multivariate analysis methods such as principal component analysis-linear discriminant analysis (PCA-LDA) and multivariate curve resolution-alternating least squares (MCR-ALS), Raman spectroscopy could effectively distinguish AML cells from normal leukocytes, and the accuracy rate was up to 96.71%. The results showed that compared with normal leukocytes, multiple components of information decomposed by the AML spectrum represented abnormal alterations in proteins, nucleic acids, lipids and carbohydrates in leukemia cells. Secondly, the same procedure was used to further detect different types of AML and found that Raman spectroscopy could distinguish AML cells with FMS-like tyrosine kinase 3-Internal Tandem Duplications (FLT3-ITD) mutations or not, and there were differences in spectral characteristics corresponding to nucleic acids and proteins (amino acids). The above results revealed that Raman spectroscopy has great potential for clinical diagnosis and mechanistic study of AML.

1 INTRODUCTION

Acute myeloid leukemia (AML) is the most common heterogeneous hematologic malignancy in adults characterized by highly diverse genetic and epigenetic abnormalities.[1] Although the complete remission rate of young newly diagnosed patients can reach 60–80% after 7 + 3 induction chemotherapy consisting of cytarabine and anthracyclines, the recurrence rate is still more than 50%, elderly patients worse.[2] Therefore, early and accurate diagnosis and classification of AML are extremely important. A variety of factors affect the prognosis of AML, including chromosomal abnormalities, fusion genes and so on. FMS-like tyrosine kinase 3-Internal Tandem Duplications (FLT3-ITD) is one of the mutations in AML that is strongly associated with poor prognosis, increased risk of recurrence and shorter overall survival, accounting for about 30% of all AML cases.[3, 4] Although the current FIT3 inhibitors have improved the prognosis, the treatment situation for leukemia with FLT3-ITD is still very severe. Therefore, the detection and exploration of patients with FLT3-ITD are also crucial.

The current diagnosis of AML is mainly made by morphological evaluation of bone marrow and peripheral blood. Further classification should be based on immunology, cytogenetics, and molecular biology.[5] Conventional methods require highly skilled experts to discriminate, and they are highly subjective, time-consuming and destructive to cells.[6] In addition, the relationship between biochemical indicators and the prognosis of AML is not clear, which limits the diagnosis and treatment of AML.[6] Cancer classification is associated with cancer progression, prognosis and treatment decisions. AML evolution has given rise to cellular phenotypic and genotypic diversity and plasticity, and this random or deterministic shift in the cellular state has the potential to cause drug resistance. At present, the comprehensive analysis of multi-omics needs to strengthen the mapping relationship between cell spatial structure network and component changes. There is still great uncertainty, which is not yet able to give the accurate staging and the most effective chemotherapy regimen in the first diagnosis of patients quickly. Therefore, a method that can effectively identify AML and explore the changes of cellular molecules to find the characteristics corresponding to carcinogenesis and prognosis is desirable. This technology will be faster and less costly, while preserving enough sensitivity and specificity to perform the actual screening.

Raman spectroscopy is a rapidly rising technology. As a photon-based inelastic scattering spectrum, it can reflect the vibration information of the functional groups in the material, providing the type and content information of small molecules (amino acids) and macromolecules (proteins, nucleic acids, lipids), with wide application prospects in biomedical research.[7-9] It requires no reagents or sample conditioning and is fast, highly specific and non-destructive.[10] J. L. Gonzalez-Solis et al. used Raman spectroscopy to detect the serum of healthy people and leukemia patients, and found that there were significant differences in 1338,1447,1532,1556,1587 and 1603 cm^{-1}.[11] Stefano Managò et al. demonstrated that normal B cells and transformed MN60 lymphocyte leukemic cell lines could be identified and discriminated by the unsupervised statistical approach.[12] Hu et al. used surface-enhanced Raman spectroscopy (SERS) to distinguish different subtypes of AML,

confirming the possibility of surface-enhanced Raman spectroscopy detection of acute myeloid leukemia.[13]

In this paper, a large number of single-cell spectral data of AML patients were collected by spontaneous Raman spectroscopy. Combined with multivariate statistical methods of PCA and LDA, a supervised prediction model was established that can distinguish AML cells from healthy cells, and the accuracy of the model was proved by clinical sample tests. Based on this, we performed a more detailed study of the spectra of AML patients to achieve effective identification and investigation of patients with FLT3-ITD mutations. To further explore the association of Raman fingerprinting with AML carcinogenesis and gene mutations, MCR-ALS was used to decompose the spectral components related to biomolecules, and their concentrations. This method solved the problem of overlapping and miscellaneous spectral information and supplemented the molecular basis of Raman diagnostic models. Our method presented nice experimental results for such deeper and more refined classification of spectra, filling the gap of Raman spectroscopy in the research of FLT3-ITD gene mutation scale. Through the analysis of Raman fingerprint big data of newly diagnosed cells in AML patients, it could be diagnosed and analyzed non-destructively at the molecular level, laying the foundation for the combination of multi-omics studies to fully explore the pathogenesis and therapeutic targets of AML.

2 MATERIALS AND METHODS

2.1 Sample collection

To ensure that the differences between different cell types were biological and not caused by different donors, the samples were selected from 6 healthy volunteers and 19 patients with acute myeloid leukemia. For each patient, a bone marrow blood smear was prepared at the recruiting hospital, using manual smears and without prior chemical processing. The hospital also analyzed each patient's blood by morphology, flow cytometry, and gene sequencing to determine the patient's pathological status.[14] All of the participating patients gave informed consent prior to enrolment in the study. This study was approved by the ethics committee of the First Bethune Hospital of Jilin University (NO.2021–347), and conducted in accordance with the Declaration of Helsinki.

2.2 Raman spectroscopic measurements

Raman spectra were acquired by a laboratory-built Raman spectroscopy system. The system was equipped with a 532 nm laser excitation source (Spectra-Physics, Excelsior 532 Single Mode) delivering 40 mW of power on the cells. All spectra were recorded in the 600-1800 cm^{-1} spectral range with a spectrometer whose resolution was about 2 cm^{-1}. The Raman signal was recorded using a CCD detector (Andor Technology, Newton970), cooled at −40°C. Wavenumber calibration of the Raman spectrometer was performed on each measurement

day using the characteristic peak of 520.7 cm^{-1} of the silicon wafer to prevent system errors caused by instrumentation drift.

16 AML patients and 4 healthy volunteers were used as training set. 100–150 white blood cells were randomly selected from each patient for spectral detection and divided into two groups: cancer group and control group, so as to consider the spectral differences between different cells and the differences between patients. In order to obtain better Raman spectral data and higher detection efficiency, the excitation light was focused on the sample using a 50-fold objective lens of an Olympus microscope, the diameter of the focused spot was about 2 μm, and the integration time of Raman spectrum acquisition was 10 seconds. The measurement parameters and conditions remained consistent during the Raman acquisition. A motorized scanning stage with a coordinate system (Marzhauser, Germany) was installed in the microscope, helping to record the location information of each detected cell in the sample during Raman spectra acquisition. In order to remove the spectra of a small amount of normal cells mixed in the AML samples, we found the field of view of each cell whose Raman spectrum had been acquired under the microscope again after H&E staining of each sample, and a skilled doctor helped to determine the pathological state of the detected cells used morphology method and select the spectral information located on the nucleus for follow-up analysis (Figure 1).

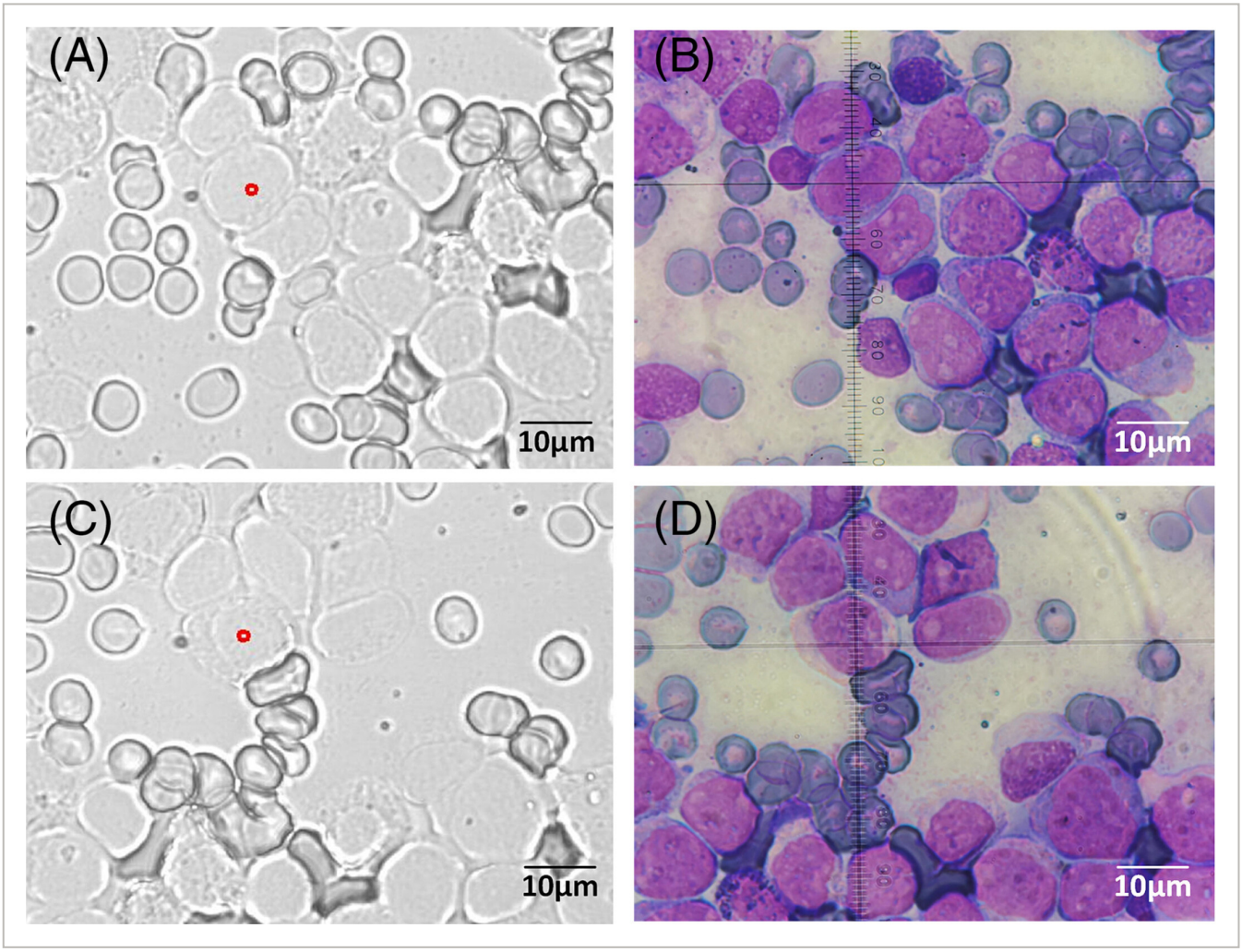

FIGURE 1

Detected cell images before and after HE staining. (A) AML cells image under microscopic and (B) image after H&E staining, (C) normal cells image under microscopic and (D) image after H&E staining.

2.3 Spectral pretreatment

After the detection, the original spectra needed to be preprocessed to improve the spectral quality. For biological cell samples, fluorescence and scattering interferences are commonly present under 532 nm laser irradiation.[15] In addition to reducing their effects by improving instrument performance, conventional digital signal pretreatments such as fluorescence and background removal, as well as noise suppression and baseline correction are still needed. In this study, spectra were smoothed with a Savitzky–Golay filter using a window width of 15 and a cubic polynomial function. The effective separation of the substrate was realized by using the specific scale analysis method based on multi-resolution wavelet transform. The fluorescence background was processed by interpolation fitting based on a cubic spline curve.[16] Finally, the maximum-minimum normalization method was applied to all spectra, and all the intensity of Raman peaks would be within the interval [0–1] to reduce the influence of laser intensity fluctuations and facilitate subsequent data comparison and analysis.[17, 18]

2.4 Spectral analysis

In this study, principal component analysis (PCA) and linear discriminant analysis (LDA) were used to analyze the Raman spectra after pretreatment. As one of the oldest and most widely used multivariate statistical algorithms in data analysis,[19] PCA was used to reduce the dimension of spectral data, eliminate the multicollinearity between variables, and extract the main features of the original data without losing useful information. LDA was a method used to study the boundary between the classes and the probabilities of classification. This method maximized the ratio of "between-class variance" to "within-class variance".[20] It was often used with PCA to form a PCA-LDA classification model for identification and classification.[21] PCA and LDA were performed using Matlab R2016 (MathWorks).

Ten-times cross-validation was a commonly used method to detect the accuracy of the algorithm. All spectral data sets were randomly divided into 10 spectral subsets, and 9 subsets were used as training data in turn. The 10th subset was used as the test data to calculate the accuracy. Repeated this operation and calculated the average to avoid the occurrence of over-learning and under-learning states. Permutation test was also an effective way to validate models in the absence of a large number of samples. The classification labels were randomly permuted and the LDA model was established (n = 10 000) to calculate the accuracy values, which could be compared with the original accuracy values. This method could verify whether the sample classification in two design groups was

significantly superior to any other random classification in any two groups, and evaluate the robustness of the model.[22]

MCR-ALS was a powerful tool aiming to decompose the data matrix of a mixed system into meaningful component and their contributions through a bilinear model.[23] The advantage was that it does not need to be based on a certain model, did not need to know the information of components in advance, and the obtained composition signal and concentration were non-negative.[24] This gave MCR a significant advantage over traditional peak fitting and statistical methods.[25]

The basic model of this algorithm was described as follows:

$$D = CS^T + E$$

in which D was an original matrix composed of Raman spectra of multiple cells from different types of AML, C was the concentration profile matrix, S^T was the resolved spectra matrix, and E was the error matrix. For MCR-ALS analysis, we used a protocol by Felten et al. under Matlab R2016 (MathWorks).[26] In the absence of prior information, singular value decomposition (SVD) combined with PCA could be used to estimate the number of components and SIMPLe-to-use-Interactive Self-Modeling Mixture Analysis (SIMPLISMA) could be used to calculate the initial estimates of the components. After that, it was optimized with the addition of constraints (nonnegativity of spectra and concentration matrices) by alternating least squares until convergence was reached.

3 RESULTS

3.1 Analysis of Raman spectra characteristics

After the screening, a total of 1648 spectra were obtained, including 1248 AML cells and 400 normal leukocytes. Figure 2A showed the average Raman spectra of normal and AML cells. The Raman spectral peak positions of normal cells and AML cells were basically the same, but their intensities were different, indicating that the molecular composition of the two cells was similar, but the corresponding content might be different. Compared with normal cells, the Raman intensity of AML cells increased at the characteristic peaks of 727, 785, 937, 1126, 1156, 1175, 1340, 1374, 1484, 1556, 1581 cm^{-1}, while the intensity at 829, 854, 1004, 1308, 1450, 1606, 1658 cm^{-1} was lower. The characteristic peak with the largest difference was at 1450 cm^{-1}, which was one of the strongest peaks in the spectrum and corresponded to the CH_2 vibrational mode of lipids, the CH_3 deformation in proteins, and also corresponded to malic acid,[27] an important metabolic intermediate in the TCA cycle. A characteristic peak with high intensity could also be seen at 1658 cm^{-1}, which was caused by the C = O stretching of amide I,[28] while the peak was also excited in unsaturated fatty acids. 1126 cm^{-1} corresponded to the C-N stretching vibration in protein, and in a large number of articles,[29-31] it was identified as an important marker of glucose. What's more, 854 cm^{-1} corresponded to tyrosine, 1003 cm^{-1} corresponded to phenylalanine, and 1308 cm^{-1} corresponded to

amide III in protein. The nucleic acid peaks were located at 785, 1096, and 1581 cm^{-1}, among others, corresponding to the O-P-O backbone structure and bases.

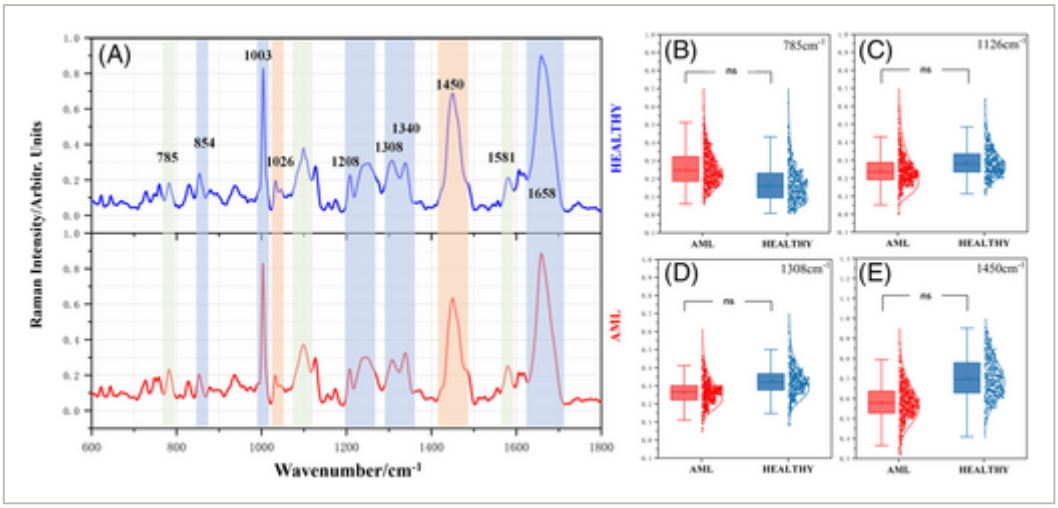

FIGURE 2

Open in figure viewer | PowerPoint

Comparison of Raman spectra of normal cells and AML cells. (A) Average spectra of AML and healthy volunteers (blue background for proteins, green for nucleic acids, yellow for lipids). (B–E) Scatter box patterns of 785, 1126, 1308 and 1450 cm^{-1} peak intensities after normalization.

To further evaluate the differences in spectral information, we plotted box plots of partial peaks. Figure 2B–E showed the peak intensity distribution of 785 cm^{-1} (nucleic acid), 1126 cm^{-1} (protein or glucose), 1308 cm^{-1} (protein), and 1450 cm^{-1} (lipid or protein). Through t-test analysis, we found that the peak intensities between the two types of cells were different, but the P value all greater than 0.05, indicating that only single-peak intensity differences were not sufficient for accurate cell classification. Therefore, univariate analysis was not sufficient to distinguish between AML cells and normal cells, and the multivariate statistical method gave us a better idea.

3.2 Multi-statistical analysis and classification of AML and normal cells

In view of the high dimension of spectrum and the complexity of cell components, it was very important to extract valid information between different cells by multivariate analysis. PCA-LDA multivariate statistical method was used to develop the diagnostic model for classifying the AML and normal cells. Figure 3A showed the PCA scatter plots of the principal component1–3 (PC1–3) concerning AML and normal cells. Points clusters of different colors in the score plot represented different cell groups. The score plots clearly showed AML and normal cells formed 2 well-defined groups, confirming that PCA can effectively distinguish normal cells from AML cells. Observation of PC loadings (Figure 3B) revealed that PC1 was

mainly dominated by peaks such as 754 cm^{-1}(tryptophan), 785 cm^{-1}(nucleic acids), 1003 cm^{-1}(phenylalanine), 1126 cm^{-1}(proteins or glucose), 1484 cm^{-1}(nucleic acids), etc. Peaks such as 1450 cm^{-1} (lipids or proteins), 1658 cm^{-1} (amide I or lipids) were also shown in PC2 and PC3. In order to establish a clear classification model, the spectral data after PCA dimension reduction were brought into the LDA classifier. The LDA scatter plotted in Figure 3C showed a good separation of normal cells and leukemia cells. Figure 3D displayed the classification accuracy of LDA under different PC numbers. When the PC number was 3, the classification accuracy of the model based on ten-times cross validation was about 92.3%. However, due to the low variance ratio, only 60%, it was not enough to represent most features of the spectrum. As the number of selected PCs increases, the classification accuracy tended to rise gradually. When the number of PCs was around 25(variance ratio reached 85%), the classification accuracy tended to be stable, overfitting might occur if the number of PCs continues to increase. Therefore, the model selected the first 25 principal components to comprehensively analyze the training set samples, and the classification accuracy was 96.71%.

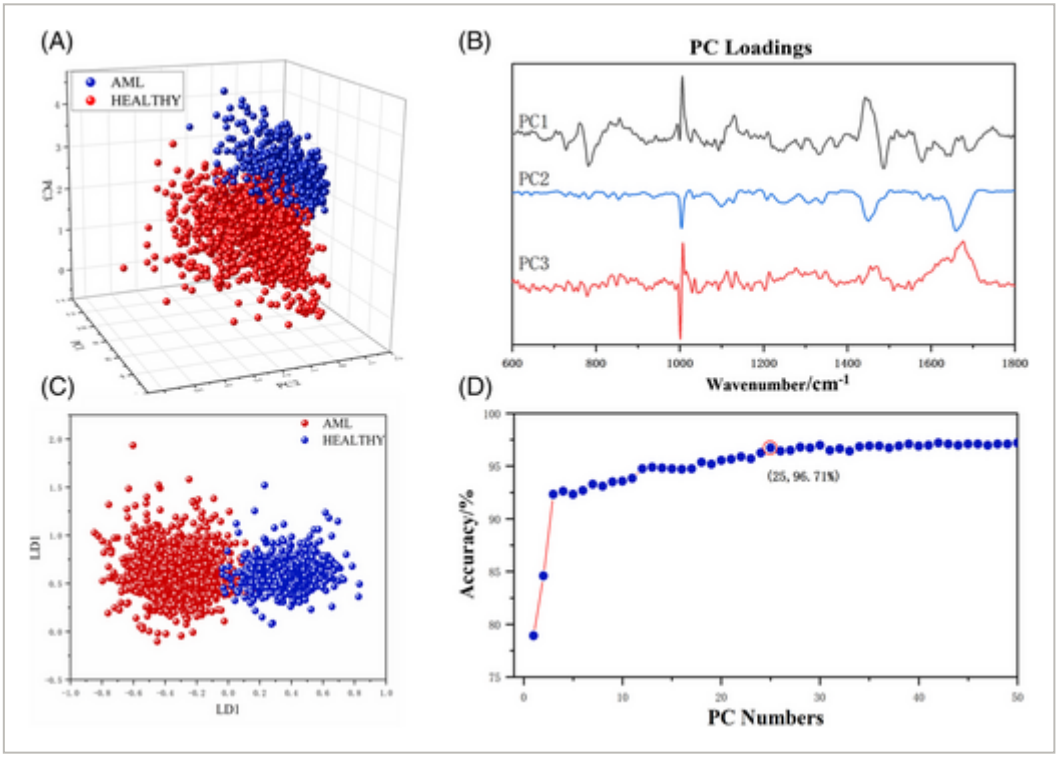

FIGURE 3

Open in figure viewer | PowerPoint

Results of PCA-LDA analysis. (A) PCA scatter plots, (B) LDA scatter plots, (C) PCs loadings of first 3 principal components, (D) classification accuracy of LDA under different PC numbers.

In order to verify the predictive ability of the model, we collected five more patients, who were not involved in the modeling, used as the prediction set, and the diagnosis results were shown in Figure 4. The identification accuracy was 94.2%, the sensitivity was 95.5% (86/90),

and the specificity was 91.7% (55/60). Calculating the diagnostic results for each patient, we found that the accuracy of our test could reach 90% and above for both healthy volunteers and AML patients, indicating that our diagnostic model could differentiate the AML patients and healthy volunteers well, with the accuracy of 100% (5/5), and effectively assist in the diagnosis of AML disease in the clinic. Although PCA-LDA models could cluster spectra and provide molecular information to some extent, these PCs showed features that included both positive and negative information and had overlapping wavenumbers, which made it hard to interpret the spectra in a biologically meaningful way.

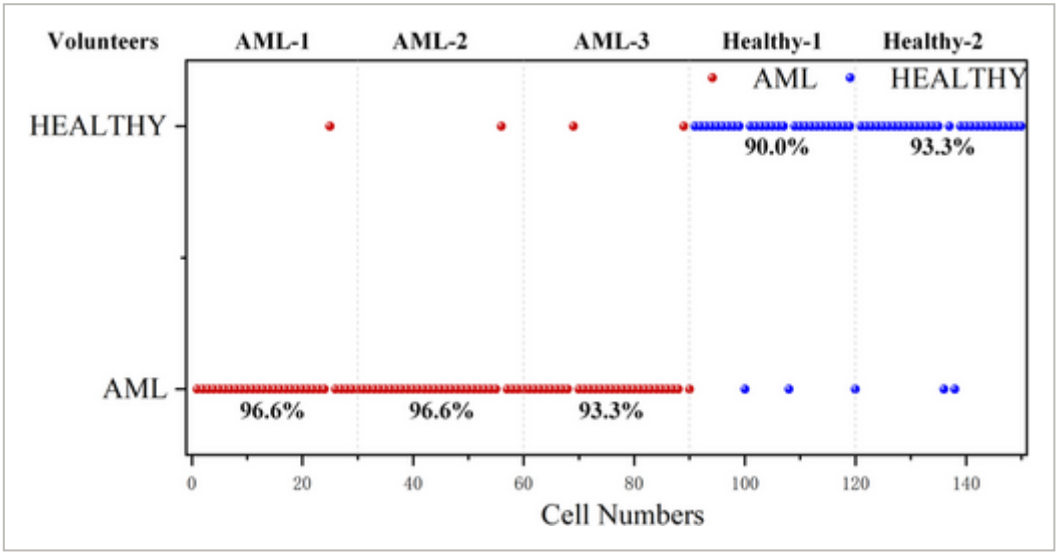

FIGURE 4

Open in figure viewer | PowerPoint

The diagnosis results of five volunteers as prediction set. The accuracy of AML patients and were 96.6%(29/30), 96.6% (29/30), 93.3%(28/30), respectively. The accuracy of healthy volunteers were90%(27/30), 93.3%(28/30), respectively.

3.3 Multivariate curve resolution analysis

In order to explore the molecular changes in acute myeloid leukemia cells and to establish a reasonable basis for successful discrimination by statistical methods such as LDA, we attempted MCR-ALS analysis to obtain multiple meaningful molecular components and their respective concentrations. A total of 200 AML cells and 200 normal cells were randomly selected for this analysis. The spectral profiles of three components extracted from the MCR-ALS were shown in Figure 5A–C. Table 1 showed the assignment of Raman peaks of each component.

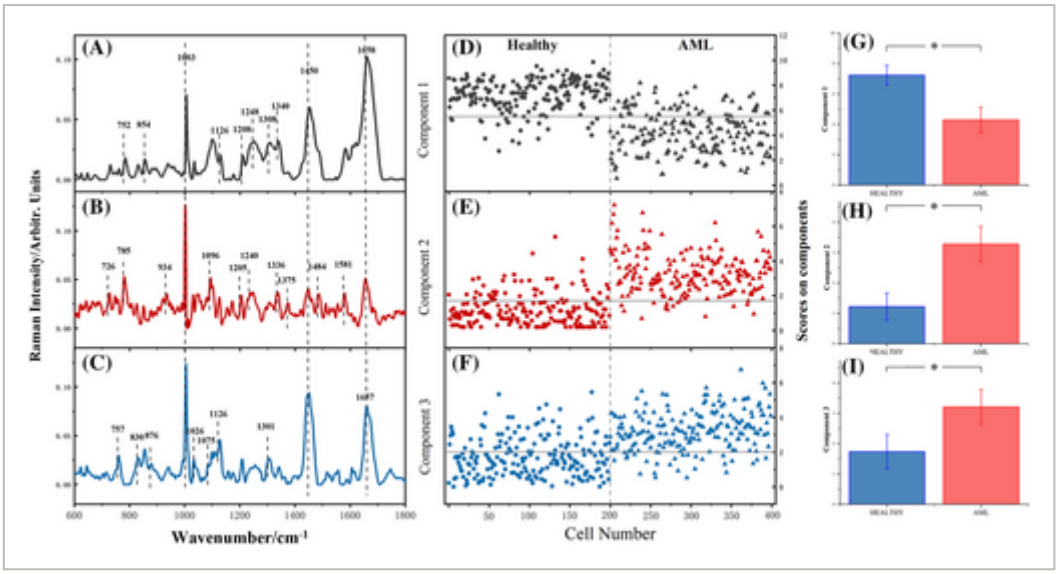

FIGURE 5

Open in figure viewer | PowerPoint

Using MCR-ALS analysis of Raman spectra from AML and healthy cells. Three components were decomposed and their respective MCR loadings (A–C) and scores (D–F) are plotted. (G–H) Average abundance histogram of three components. P-values obtained by t-test to determine the significant difference between the two groups ('*' means $p < 0.05$, 'ns' means no significance).

TABLE 1. Raman spectra peak assignments of MCR components obtained from the MCR-ALS analysis in Figure 5A–C.

Component 1			Component 2			Component 3	
Peak	Assignment		Peak	Assignment		Peak	Assignment
752	Tryptophan	Proteins	726	A	Nucleic Acids	757	MPO; Cytochrome C
822	Tyrosine	Proteins	785	T; C; U; O-P-O stretching	Nucleic Acids	830	Diester symmetric stretching
854	Proline; Tyrosine	Proteins	1003	phenylalanine	Proteins	1003	phenylalanine
934	C–C stretching	Proteins	1096	O-P-O symmetric stretching	Nucleic Acids	1026	Lactic acid
1003	phenylalanine	Proteins	1205	A; T	Nucleic Acids	1057	Triacylglycerols

Component 1			Component 2			Component 3	
Peak	Assignment		Peak	Assignment		Peak	Assignment
1126	C-N stretching	Proteins	1240	RNA	Nucleic Acids	1075	Triacylglycerols, C-O

Note: Assignments based on refs.[12, 30, 32-37]

According to the spectral peak assignment, component 1 could correspond to most of the proteins and amino acids in cells. The peaks such as 758, 822, 854, 1003, 1208 cm^{-1} related to tryptophan, tyrosine, and phenylalanine, 934 cm^{-1} related to C-C vibration of protein, 1340, 1450 cm^{-1} related to C-H deformation vibration of protein, 1126 cm^{-1} related to C-N stretching vibration of protein, 1208, 1340 cm^{-1} related to amide III, and 1658 cm^{-1} related to amide I, were all appeared in component 1. It was observed in Figure 5D,G that the concentration of component 1 in AML cells was lower, indicating that the content of protein and amino acids in AML cells was reduced compared with normal leucocyte. The characteristic changes of spectra might be related to the abnormal active catabolism of amino acids during cell carcinogenesis. Component 2 may correspond to the mixture of nucleic acid and a small number of proteins associated with nucleic acid. Component 2 showed multiple peaks at 726, 785, 1208, 1336, 1375, 1484 cm^{-1} excited by bases and 785, 1096, 1581 cm^{-1} excited by O-P-O skeleton structure of DNA or RNA, and also 1003 cm^{-1} corresponding to amino acids, with clear different concentration values in the two cell samples (Figure 5E,H). Although there was some protein interference in component 2, it can still help us in understanding the trend of nucleic acid content in the cell. The nucleic acids content of AML cells was higher than that of normal cells, which could be used as one of the effective features to identify AML cells in blood smears. Component 3 was related to lipids and carbohydrates in cells, such as, 1026, 1126 cm^{-1} corresponding to lactic acid and glucose, 830, 1057, 1075, 1301, 1450 cm^{-1} related to a variety of lipid components, and also 1003 cm^{-1} related to protein. The component 3 increased slightly in AML cells (Figure 5F,I).

Unlike the method of PCA, MCR-ALS provided a more detailed and biological explanation for the spectral data, and better helped to describe the content of intracellular substances and molecular changes in the process of carcinogenesis from the macromolecular scales such as proteins, nucleic acid, lipids, and carbohydrates.

3.4 Identification of FLT3-ITD cells based on Raman spectroscopy

On the basis of successful differentiation of AML and healthy leukocytes, we could further observe the spectra of AML cells with specific gene mutations. The average spectra of AML patients with FLT3-ITD mutation were shown in Figure 6A. Compared with leukemia cells

without FLT3-ITD mutation, the Raman intensity of FLT3-ITD mutant cells decreased at the characteristic peaks of 645, 829, 854, 1126, 1155 and 1208 cm^{-1}, and the intensity of Raman peaks at 785, 1033, 1096, 1484, 1581 and 1605 cm^{-1} increased.

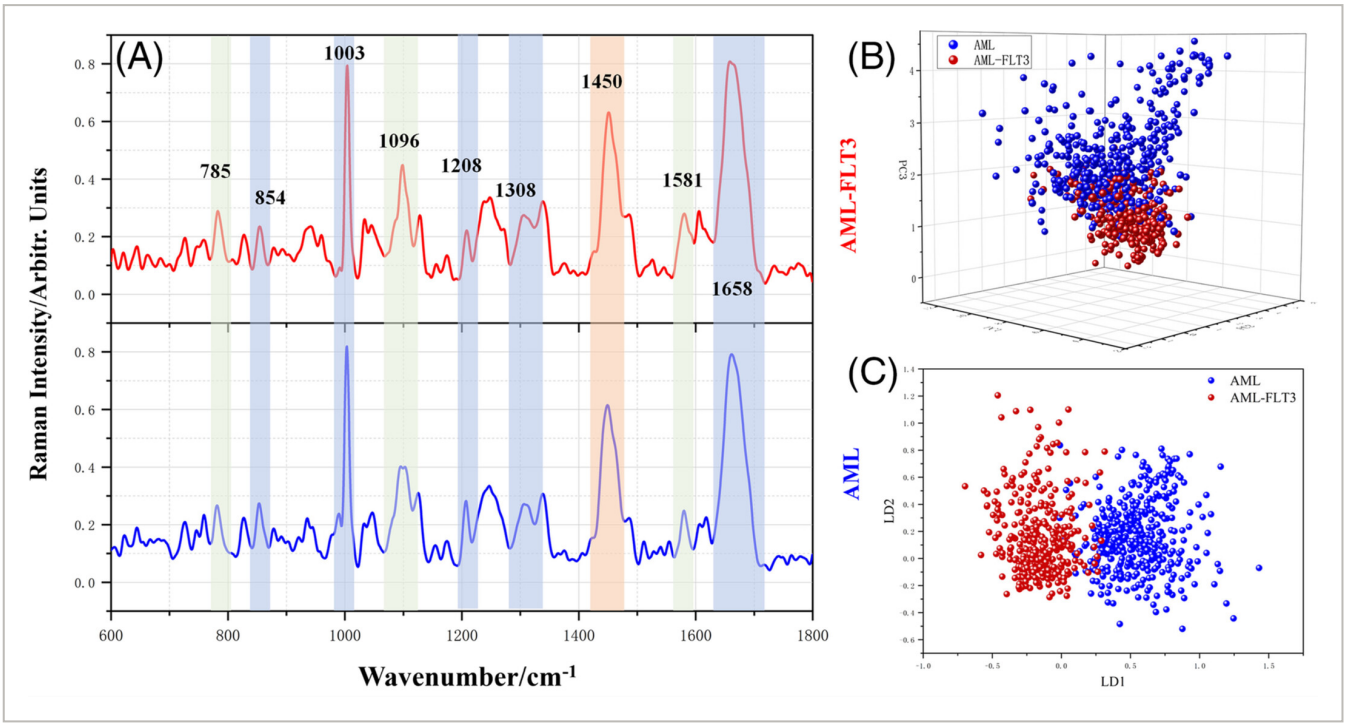

FIGURE 6

Open in figure viewer | PowerPoint

Analysis of Raman spectra of AML cells with and without FLT3-ITD mutation (A) average spectra comparison. (B–C) Results of PCA-LDA analysis.

PCA was used to reduce the dimension of 730 cells of two types of patients with FLT3-ITD mutation or not, and the three-dimensional scatter plot of PCA was obtained as Figure 6B. The LDA classification result (Figure 6C) was further shown, and the classification accuracy calculated by ten-times cross-validation was up to 90.2%. The permutation test (P value <0.001) proved that results achieved were valid (Figure 7) for the classification model built for the FLT3-ITD mutation. Such results indicated that the high accuracy of the classifier was unlikely to have occurred by chance.

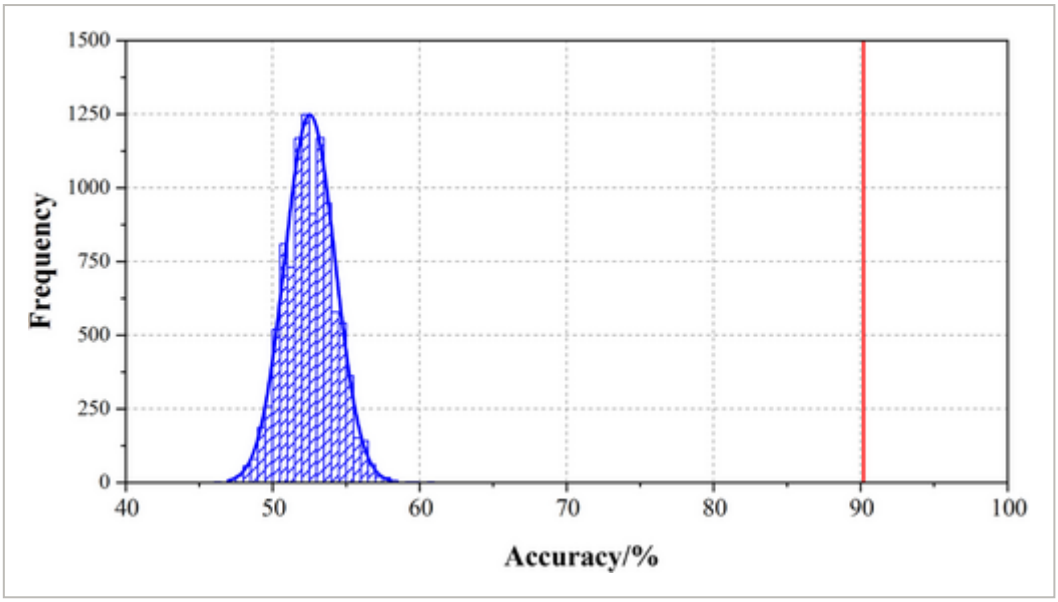

FIGURE 7

Open in figure viewer | PowerPoint

Histogram of the accuracy in the analysis of 10 000 permutation test applied for the PCA-LDA model, the red vertical line corresponds to the accuracy of the model when correctly labeled.

Similarly, in order to further explore the molecular changes in AML cells with different mutations, we randomly selected 100 cells with FLT3-ITD and 100 cells without FLT3-ITD, and used the MCR algorithm to extract three components, related to proteins (amino acids), nucleic acids, lipids and carbohydrates (Figure 8A–C). It could be seen from Figure 8D–I that in cells with FLT3-ITD mutation, the content of component 1 related to protein and amino acids increased, and the concentration of component 2 related to nucleic acids also increased accordingly, while the content of component 3 related to lipids, carbohydrates and a small amount of protein were not significantly different. In summary, Raman spectroscopy combined with multivariate statistical algorithms had the ability to distinguish between cancer cells in AML patients with and without FLT3-ITD mutations and assist in understanding intracellular molecular changes.

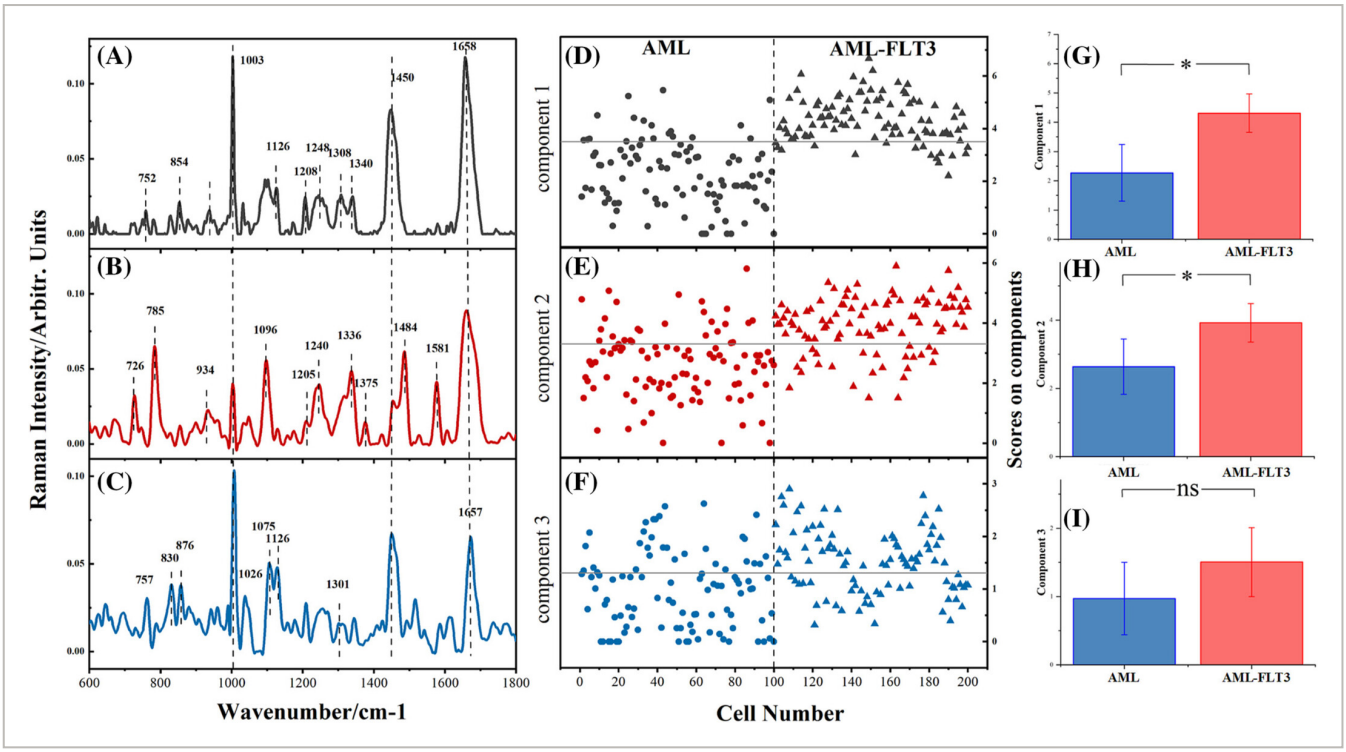

FIGURE 8

Open in figure viewer | PowerPoint

Using MCR-ALS analysis of Raman spectra from AML without FLT3-ITD and AML with FLT3-ITD. Three components were decomposed and their respective MCR loadings (A–C) and scores (D–E) are plotted. (G–H) Average abundance histogram of three components. P-values obtained by t-test to determine the significant difference between the two groups ('*' means $p < 0.05$, 'ns' means no significance).

4 DISCUSSION

Thousands of proteins, lipids, nucleic acids, and carbohydrates exist in cells and tissues. The composition and content of these molecules in cells will change with the state of cells, such as the development of diseases.[38] Alterations in intracellular molecules such as abnormal expression of genes and metabolic disorders are widely accepted hallmarks of cancer,[39] and are monitored by many pathways and methods. Raman spectroscopy is one of the most cutting-edge tools in probing cellular components and is a powerful complementary tool for AML diagnosis as it provides more comprehensive detection of substance types and can probe cellular changes from a variety of components such as proteins, amino acids, nucleic acids, and lipids.

This study showed that it was feasible to identify AML based on Raman spectroscopy and multivariate statistical analysis, and could further distinguish and explore FLT3-ITD mutant cells. The combination of multivariate statistical algorithms was necessary, because the complexity of the biological spectrum was extremely strong, and the high heterogeneity of cell samples led to spectral uncertainty. The spectrum also had certain absolute intensity

fluctuations during the measurement process, which was mainly due to system errors (such as laser power, acquisition time, acquisition efficiency, spatial resolution, etc.) and experimental errors (such as focal depth, measurement position, etc.). Therefore, univariate method was difficult to identify AML cells.

In the diagnostic model of AML, we found significant differences in the distribution of protein, nucleic acid, glucose and lipid between AML and normal cells. The decrease of component 1 showed in Figure **5A** in AML cells might be related to the abnormal active catabolism of amino acids during cell carcinogenesis, such as glutamine and arginine. Amino acids played a vital role in cells. Compared with normal cells, many amino acids were absorbed in cancer cells and participated in synthesis and metabolic pathways. Their breakdown not only provided the carbon and nitrogen sources needed to maintain protein synthesis or energy metabolism for cell growth, but also played an important role in regulating the various biological processes of cancer through specific signaling networks, in particular the PI3K/AKT /mTOR signaling pathway.[40] Jones et al. found that the amino acid metabolism level in the leukemia blast cells of newly diagnosed AML patients was active,[41] which was consistent with the results of our experiment. A major feature of cell carcinogenesis was the enhancement of reproductive capacity. DNA was related to cell division and reproduction, and RNA was related to cell protein synthesis and growth. Therefore, increased nucleic acid was the material basis for rapid tumor growth. The change of glycolysis was the main sign of malignant growth, so was AML. Cunningham and Kohno examined 124 patients with acute leukemia using fluorodeoxyglucose (FDG) -PET/CT and found that cells from 101 AML patients showed high glucose uptake.[42] The spectral peaks in component 3, such as 1126 cm^{-1} related to glucose and 830, 1026 cm^{-1} corresponding to lactic acid, were slightly up-regulated in AML, which meant that Raman spectroscopy had the ability to detect the change of glucose metabolism in cancer cells. Lipid-related peaks such as 1075, 1450, 1657 cm^{-1} in component 3 indicated the presence of abnormal lipid metabolism in AML cells. Lipid was a highly complex biological molecule. It not only had the function of energy source, but also provided fatty acid building block for cell membrane biogenesis and signaling pathway regulation.[43] Therefore, in-depth study of cellular lipids was helpful for the diagnosis and follow-up monitoring of AML. At the same time, we found that the lipid peak overlaps with the protein before analysis, which also showed the analytical ability of MCR-ALS in the face of multiple structural spectral peaks overlap.

In acute leukemia, the most common type of FLT3 mutation was the intracellular tandem replication of JMD (FLT3-ITD). FMS-like tyrosine kinase 3 (FLT3) was a proto-oncogene involved in the key steps of hematopoietic cell proliferation, differentiation and survival. In recent years, FLT3 had become an important marker of different hematological malignancies, especially in acute myeloid leukemia. FLT3 gene mutation was related to the clinical prognosis, treatment and survival of patients.[44] According to the experimental results, we found that the content of proteins and amino acids increased in FLT3-ITD mutant cells. This might be because this extended Flt3 receptor lost the ability of automatic

inhibition, resulting in structural autophosphorylation, activating PI3K/AKT, RAS/ERK and STAT5 pathways.[45] Similarly, studies have found that the content of nucleic acid in leukemia cells in FLT3-ITD was more, which was consistent with the increase of component 2 concentration corresponding to nucleic acid in the experimental results. Indicating that FLT3-ITD had a greater tendency to proliferate by providing nucleotides synthesized from DNA and RNA.[46]

Using Raman spectroscopy combined with multivariate analysis algorithms, an AML diagnosis model with an accuracy rate higher than 90% could be developed. To further improve the accuracy, we need to collect more samples and enrich the database to eliminate the errors caused by individual differences. Moreover, the purity of the sample is also very important. How to separate and accurately enrich different types of cells is the key to further improving the accuracy of the model.

5 CONCLUSION

In conclusion, our study demonstrated that Raman spectroscopy was a powerful and promising tool of the clinical diagnosis for AML. By utilizing PCA-LDA, the spectral data could distinguish leukemic cells from normal cells with the accuracy up to 96.7%. In the classification model based on Raman spectroscopy, the components and content obtained by MCR-ALS reflected the spectral differences between normal cells and AML cells. The Raman spectral characteristics of these components were similar to those of proteins, nucleic acids, lipids, carbohydrates, etc. The changes of the content of these components could help to explore the intracellular activities and associate with the abnormal amino acids decomposition, rapid proliferation and other features during carcinogenesis. In addition, cells with Flt3-ITD mutations could be further identified with an accuracy of 90.2%. With the accumulation of a large number of leukemia and normal Raman spectral data, as well as the combined use of multi-omics studies, the pathogenesis of leukemia will be further revealed, and the evaluation of leukemia and treatment options will be more accurate and effective to achieve early diagnosis and precision medicine.

ACKNOWLEDGEMENTS

This work was supported by the National Natural Science Foundation of China (U21A20395, 61727813), Research Equipment Development Project of Chinese Academy of Sciences (YJKYYQ20210035) and Bethune Medical Engineering and Instruments Centre Fund (BQEGCZX2021010).

REFERENCES

1 M. G. Yu, H. Y. Zheng, *Chin. Med. J.* 2017, **130**, 211.

CAS | PubMed | Web of Science® | Google Scholar

2 H. Dombret, C. Gardin, *Blood* 2016, **127**, 53.

CAS | PubMed | Web of Science® | Google Scholar

3 F. A. Lagunas-Rangel, V. Chavez-Valencia, *Med. Oncol.* 2017, **34**, 114.

PubMed | Web of Science® | Google Scholar

4 R. Friedman, *Biochim. Biophys. Acta Rev. Cancer* 2022, **1877**, 188666.

CAS | PubMed | Web of Science® | Google Scholar

5 Y. Ma, H. X. Tong, X. Deng, Y. Zhao, Z. G. Liu, J. H. Zhang, *Zhongguo Shi Yan Xue Ye Xue Za Zhi* 2009, **17**, 12.

CAS | PubMed | Google Scholar

6 H. Liang, X. Cheng, S. Dong, H. Wang, E. Liu, Y. Ru, Y. Li, X. Kong, Y. Gao, *J. Pharm. Biomed. Anal.* 2022, **210**, 114560.

CAS | PubMed | Web of Science® | Google Scholar

7 Q. Tu, C. Chang, *Nanomedicine* 2012, **8**, 545.

CAS | PubMed | Web of Science® | Google Scholar

8 L. Li, J. Yang, J. Wei, C. Jiang, Z. Liu, B. Yang, B. Zhao, W. Song, *Light: Sci. Appl.* 2022, **11**, 1.

PubMed | Web of Science® | Google Scholar

9 S. Bai, K. Sugioka, *Light: Adv. Manufact.* 2021, **2**, 186.

Google Scholar

10 S. Elumalai, S. Manago, A. C. De Luca, *Sensors (Basel)* 2020, **20**, 19.

Google Scholar

11 J. L. Gonzalez-Solis, J. C. Martinez-Espinosa, J. M. Salgado-Roman, P. Palomares-Anda, *Lasers Med. Sci.* 2014, **29**, 1241.

PubMed | Web of Science® | Google Scholar

12 S. Manago, P. Mirabelli, M. Napolitano, G. Zito, A. C. De Luca, *J. Biophotonics* 2018, **11**, e201700265.

PubMed | Web of Science® | Google Scholar

13 M. Ye, Y. Chen, Y. Wang, L. Xiao, Q. Lin, H. Lin, Z. Duan, S. Feng, Y. Cao, J. Zhang, J. Li, J. Hu, *Spectrochim. Acta A Mol. Biomol. Spectrosc.* 2022, **271**, 120865.

CAS | PubMed | Web of Science® | Google Scholar

14 T. Happillon, V. Untereiner, A. Beljebbar, C. Gobinet, S. Daliphard, P. Cornillet-Lefebvre, A. Quinquenel, A. Delmer, X. Troussard, J. Klossa, M. Manfait, *Analyst* 2015, **140**, 4465.

CAS | PubMed | Web of Science® | Google Scholar

15 D. I. Ellis, R. Goodacre, *Analyst* 2006, **131**, 875.

CAS | PubMed | Web of Science® | Google Scholar

16 X. X. Han, Y. H. Wu, M. B. Chi, S. J. Gao, Q. Wang, *J. Raman Spectrosc.* 2020, **51**, 2255.

CAS | Web of Science® | Google Scholar

17 Y. Zhao, Y. Zhang, X. Liang, *J. Raman Spectrosc.* 2021, **6**, 52.

Google Scholar

18 Y. Dai, W. Li, L. Wang, C. Luo, Q. Huang, L. Pang, *Appl. Spectrosc.* 2021, **12**, 75.

Google Scholar

19 K. Iwasaki, A. Araki, C. M. Krishna, R. Maruyama, T. Yamamoto, H. Noothalapati, *Int. J. Mol. Sci.* 2021, **22**, 2.

Web of Science® | Google Scholar

20 X. Zhuang, H. Z. Pei, T. Li, J. Huang, Y. Guo, Y. Zhao, M. Yang, D. Zhang, Z. Chang, Q. Zhang, L. Yu, C. He, L. Zhang, Y. Pan, C. Chen, Y. Chen, *Front. Oncol.* 2022, **12**, 931462.

PubMed | Web of Science® | Google Scholar

21 A. Amin, N. Ghouri, S. Ali, M. Ahmed, M. Saleem, J. Qazi, *J. Raman Spectrosc.* 2017, **48**, 705.

CAS | Web of Science® | Google Scholar

22 A. Mahadevan-Jansen, G. R. Lloyd, J. Hutchings, L. M. Almond, H. Barr, C. Kendall, N. Stone, W. Petrich, *Biomed. Cal Vib. Spectrosc. V: Adv. Res. Ind.* 2012, **8219**, 65.

Google Scholar

23 A. De Juan, J. Jaumot, R. Tauler, *J. Anal. Methods* 2014, **6**, 4964.

Web of Science® | Google Scholar

24 I. A. Matveeva, O. O. Myakinin, Y. A. Khristoforova, I. A. Bratchenko, E. N. Tupikova, V. P. Zakharov, O. A. Romanovskii, Y.V. Kistenev, Fourth International Conference on Terahertz and Microwave Radiation: Generation, Detection, and Applications 2020.

Google Scholar

25 S. Kim, W. Kim, A. Bang, J. Y. Song, J. H. Shin, S. Choi, *Anal. Methods* 2021, **13**, 3249.

CAS | PubMed | Web of Science® | Google Scholar

26 J. Felten, H. Hall, J. Jaumot, R. Tauler, A. de Juan, A. Gorzsas, *Nat. Protoc.* 2015, **10**, 217.

CAS | PubMed | Web of Science® | Google Scholar

27 S. Chen, C. Wang, R. Zhu, S. Zhu, G. Zhang, *Nanomedicine (Lond.)* 1873, **2021**, 16.

Google Scholar

28 N. Kuhar, S. Sil, S. Umapathy, *Spectrochim. Acta A Mol. Biomol. Spectrosc.* 2021, **258**, 119712.

CAS | PubMed | Web of Science® | Google Scholar

29 Z. Gabazana, L. Sitole, *Spectrochim. Acta A Mol. Biomol. Spectrosc.* 2021, **248**, 119256.

CAS | PubMed | Web of Science® | Google Scholar

30 J. W. Kang, Y. S. Park, H. Chang, W. Lee, S. P. Singh, W. Choi, L. H. Galindo, R. R. Dasari, S. H. Nam, J. Park, *Sci. Adv.* 2020, **6**, 5206.

Web of Science® | Google Scholar

31 Y. Park Sang, S. Ahn, H. Chang, W. Lee, S. N. Hyun, *Annu. Int. Conf. IEEE Eng. Med. Biol. Soc.* 2020, **2020**, 6139.

PubMed | Google Scholar

32 S. Kar, S. V. Jaswandkar, K. S. Katti, J. W. Kang, R. Paulmurugan, D. Liepmann, R. Venkatesan, D. R. Katti, *Sci. Rep.* 2022, **12**, 8050.

CAS | PubMed | Web of Science® | Google Scholar

33 L. Zhang, C. Li, D. Peng, X. Yi, S. He, F. Liu, X. Zheng, W. E. Huang, L. Zhao, X. Huang, *Spectrochim. Acta A Mol. Biomol. Spectrosc.* 2022, **264**, 120300.

CAS | PubMed | Web of Science® | Google Scholar

34 N. Feuerer, D. A. Carvajal Berrio, F. Billing, S. Segan, M. Weiss, U. Rothbauer, J. Marzi, K. Schenke-Layland, *Biomedicines* 2022, **10**, 5.

Web of Science® | Google Scholar

35 M. Marro, C. Nieva, A. de Juan, A. Sierra, *Anal. Chem.* 2018, **90**, 5594.

CAS | PubMed | Web of Science® | Google Scholar

36 Z. Duan, Y. Chen, M. Ye, L. Xiao, Y. Chen, Y. Cao, Y. Peng, J. Zhang, Y. Zhang, T. Yang, W. Liu, S. Feng, J. Hu, *FASEB J* 2022, **36**, e22416.

CAS | PubMed | Web of Science® | Google Scholar

37 Y. Chen, P. Jiang, S. Lei, X. Chen, S. Yao, D. Jiang, D. Lin, X. Jia, J. Hu, *J. Biophotonics* 2022, **15**, e202200117.

CAS | PubMed | Web of Science® | Google Scholar

38 M. Kopec, A. Imiela, H. Abramczyk, *Sci. Rep.* 2019, **9**, 166.

CAS | PubMed | Web of Science® | Google Scholar

39 R. J. DeBerardinis, N. S. Chandel, *Sci. Adv.* 2016, **2**, e1600200.

PubMed | Web of Science® | Google Scholar

40 Y. Mesbahi, T. N. Trahair, R. B. Lock, P. Connerty, *Front. Oncol.* 2022, **12**, 807266.

PubMed | Web of Science® | Google Scholar

41 C. L. Jones, B. M. Stevens, A. D'Alessandro, J. A. Reisz, R. Culp-Hill, T. Nemkov, S. Pei, N. Khan, B. Adane, H. Ye, A. Krug, D. Reinhold, C. Smith, J. DeGregori, D. A. Pollyea, C. T. Jordan, *Cancer Cell* 2018, **34**, 724.

CAS | PubMed | Web of Science® | Google Scholar

42 I. Cunningham, B. Kohno, *Am. J. Hematol.* 2016, **91**, 379.

PubMed | Web of Science® | Google Scholar

43 M. Soltani, Y. Zhao, Z. Xia, M. H. Ganjalikhani, A. V. Bazhin, *Front. Oncol.* 2021, **11**, 767026.

PubMed | Web of Science® | Google Scholar

44 S. Moharram, *Role of FLT3 in acute myeloid leukemia: molecular mechanisms and therapeutic opportunities*, Department of Laboratory Medicine, Faculty of Medicine, Lund University, Sweden, 2021.

Google Scholar

45 J. Cheng, L. Qu, J. Wang, L. Cheng, Y. Wang, *Mol. Med. Rep.* 2018, **17**, 2885.

CAS | PubMed | Web of Science® | Google Scholar

46 B. Stockard, T. Garrett, J. Guingab-Cagmat, S. Meshinchi, J. Lamba, *Sci. Rep.* 2018, **8**, 5534.

PubMed | Web of Science® | Google Scholar

Download PDF