Article

Elucidating the Heterogeneity of Serum Metabolism in Patients with Myelodysplastic Syndrome and Acute Myeloid Leukemia by Raman Spectroscopy

Haoyue Liang, Xiaodong Kong, Haoyu Wang, Yansong Ren, Ertao Liu, Fanfan Sun, Jianwei Qi,* Qiang Zhang,* and Yuan Zhou*

Read Online

ACCESS | Metrics & More | Article Recommendations | Supporting Information

ABSTRACT: Myelodysplastic syndrome (MDS) is difficult to diagnose and classify because it has the potential to evolve into acute myeloid leukemia (AML). Raman spectroscopy and orthogonal partial least squares discrimination analysis (OPLS-DA) are used to systematically analyze peripheral blood serum samples from 33 patients with MDS, 25 patients with AML, and 29 control volunteers to gain insight into the heterogeneity of serum metabolism in patients with MDS and AML. AML patients show unique serum spectral data compared to MDS patients with considerably greater peak intensities of collagen (859 and 1345 cm^{-1}) and carbohydrate (920 and 1123 cm^{-1}) compared to MDS patients. Screening and bioinformatics analysis of MDS- and AML-related genes based on the Gene Expression Omnibus (GEO) database shows that 1459 genes are differentially expressed, and the main signaling pathways are related to Th17 cell differentiation, pertussis, and cytokine receptor interaction. Statistical analysis of serological indexes related to glucose and lipid metabolism shows that patients with AML have increased serum triglyceride (TG) levels and decreased total protein levels. This study provides a spectral basis for the relationship between the massive serological data of patients and the typing of MDS and AML and provides important information for the rapid and early identification of MDS and AML.
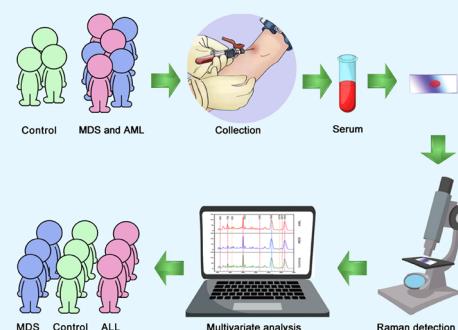
INTRODUCTION

Hematopoietic stem cells are the source of the malignant clonal illness known as acute leukemia.[1,2] On the basis of cell morphology, it may be subdivided into acute myeloid leukemia (AML) and acute lymphocytic leukemia, each of which has its own subgroups. This kind of leukemia is more prevalent in adults and results from a malfunction in the bone marrow's hematopoietic cell proliferation and differentiation.[3−5] AML can be clinically classified into primary AML and secondary AML. Secondary AML does not meet the standard of leukemia at the time of diagnosis but transforms into leukemia as the disease progresses. Secondary AML is more difficult to treat than primary AML, and the prognosis is worse. Myelodysplastic syndrome (MDS), myeloproliferative neoplasms, lymphoma, paroxysmal nocturnal hemoglobinuria (PNH), multiple myeloma, and chronic lymphocytic leukemia may transform into secondary AML.[6−8]

MDS is a collection of diverse myeloid clonal illnesses that manifest as aberrant differentiation and development of myeloid cells and originate in hematopoietic stem cells. Hematopoietic failure, low blood cell counts, and an increased chance of developing AML are all symptoms of this condition.[9,10] Anemia, bleeding, and infection are their classical symptoms, and their pathogenesis is not clear.[11,12] Clinically,

MDS is divided into four distinct subtypes: MDS with single lineage dysplasia (MDS-SLD), MDS with multilineage dysplasia (MDS-MLD), MDS with excess blasts (MDS-EB), and unclassifiable MDS (MDS-U). To further categorize MDS-EB, we distinguish between MDS-EB1 and MDS-EB2. AML may develop from bone marrow failure in patients with MDS.[13,14] Traditionally, MDS is diagnosed based on blood phase and bone marrow phase examination, cytogenetic examination, immunological phenotype analysis, and gene analysis. There is no "gold standard" for the diagnosis of MDS, which is an "exclusive" diagnosis. Its differential diagnosis with AML necessitates detection on multiple platforms.[15,16]

Refractory anemia MDS is easily confused with AML, especially the MDS-EB subtype. Therefore, it is crucial to investigate a low-cost diagnostic tool that may detect myelodysplastic syndromes and AML at an early stage in the
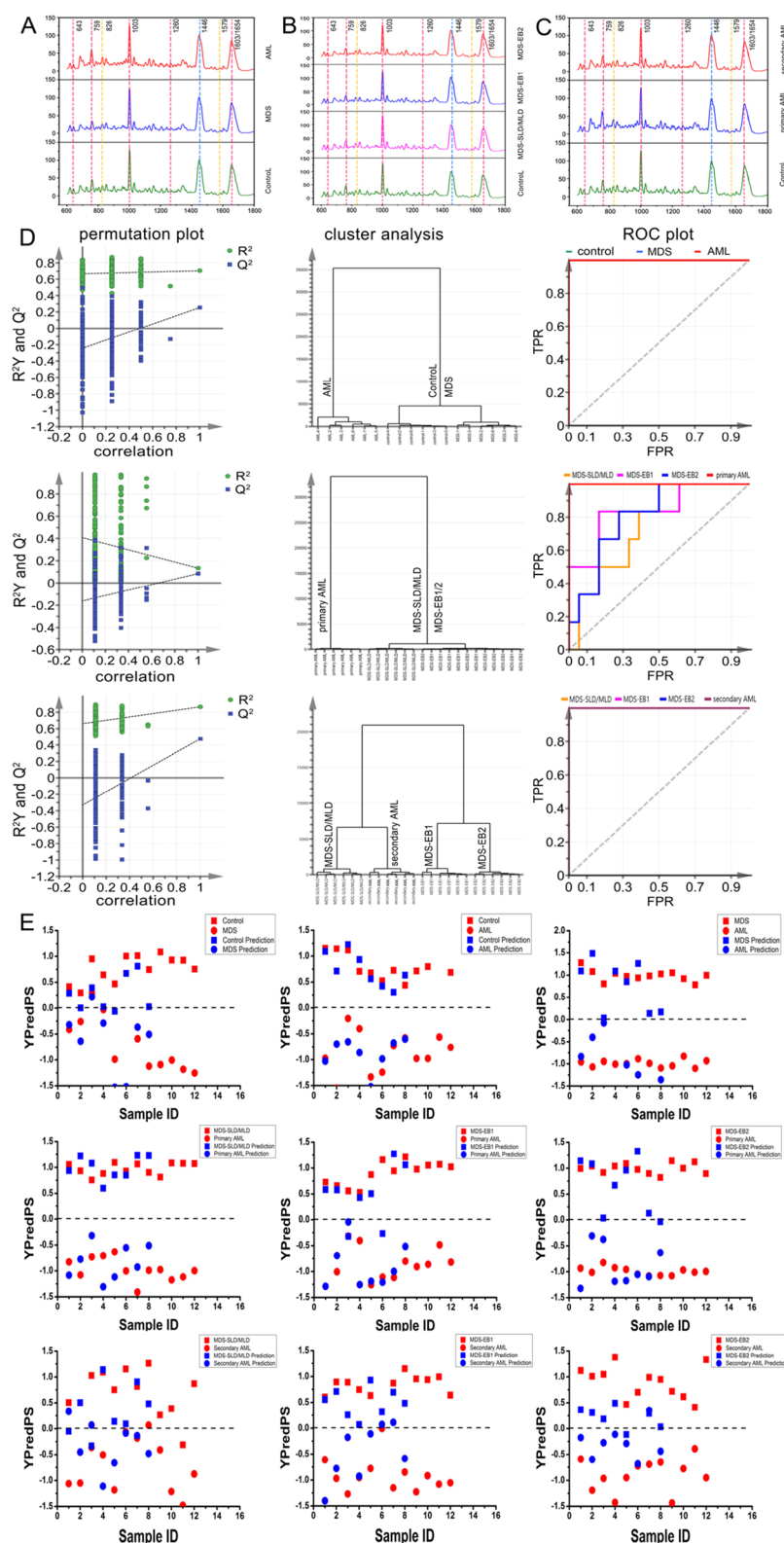
Figure 1. (A) (bottom to top) Mean serum spectra of control subjects, patients with MDS, and patients with AML patients. (B) (bottom to top) Mean serum spectra of control subjects and patients with MDS-SLD/MLD, MDS-EB1, and MDS-EB2. (C) (bottom to top) Mean serum spectra of control subjects and patients with primary AML and secondary AML. (D) Permutation, cluster, and receiver operating characteristic (ROC) curve plots of OPLS discrimination for control vs MDS vs AML (AUC (control) = 1, AUC (MDS) = 1, and AUC (AML) = 1), MDS subtypes vs primary AML (AUC (MDS-SLD/MLD) = 0.75, AUC (MDS-EB1) = 0.842593, AUC (MDS-EB2) = 0.805556, and AUC (primary AML) = 1), and MDS subtypes vs secondary AML (AUC (MDS-SLD/MLD) = 1, AUC (MDS-EB1) = 1, AUC (MDS-EB2) = 1, and AUC (secondary AML) = 1) models. (E) Validation plots of OPLS discrimination for control vs MDS vs AML, MDS subtypes vs primary AML, and MDS subtypes vs secondary AML models.

diagnostic process. In clinical practice, the traditional diagnostic methods of MDS and AML are time-consuming, expensive, and invasive. Moreover, while the association between serological indexes related to glucose and lipid metabolism and the transformation from MDS to AML is not backed by sufficient evidence, studying this association will allow the rapid and early identification of MDS and AML. In this case, developing a rapid identification method of MDS and AML without antibody labeling based on Raman spectroscopy will allow detailed analysis of the massive serological examination data of MDS and AML patients. It will also improve the diagnostic efficiency of diseases, reduce detection costs, and promote further development of rapid and accurate diagnostic methods identifying the transformation of MDS to AML.

Raman spectroscopy has been used for the rapid, non-invasive, and nonlabel identification of hematologic diseases. In a previous study, a bifunctional nano probe based on dopant-driven plasma oxide with surface-enhanced Raman scattering (SERS) was used to distinguish single acute monocyte leukemia THP-1 cells from peripheral blood monocytes with high accuracy. This nano probe could be activated by the biological redox reaction of single cells to produce stimuli for a complementary colorimetric reaction, thus, the prospect of obtaining single-cell-level unified identification accuracy is enhanced, allowing for the precise and cost-effective detection of cancer cells in complex cell samples.[17] Romy et al. differentiated red blood cells (RBCs) from erythroid precursor cells by Raman spectroscopy. These precursor cells were infected by a SARS-CoV-2 variant in vitro. The authors found that differentiated RBCs have impaired hemoglobin biosynthesis, abnormal iron metabolism, high serum ferritin levels, and low serum iron and transferrin levels, which explains the impaired oxygen-binding ability of RBCs in patients with severe COVID-19.[18] The most reliable approach for identifying PNH, an uncommon condition marked by RBC hemolysis and venous thrombosis, is flow cytometry. Kaan et al. devised a technique for analyzing blood samples from participants with and without PNH that combines optical tweezers and Raman spectroscopy (Raman tweezers). Training using support vector machine analysis resulted in an 81.8% specificity and a 78.3% sensitivity for detecting PNH with great accuracy.[19] Gold nanoparticles were used in a cell-free and label-free SERS approach developed by Stacy et al. for the classification of hematopoietic malignancies. When applied to three groups, their linear and quadratic discriminant analyses could only discriminate between them with a 69.8 and 71.4% accuracy, respectively. Their results demonstrate the feasibility of using nanomaterials in translational medicine and pave the way for future research into the noninvasive monitoring of disease development.[20] Multidrug resistance highly correlates with the poor prognosis of chronic myeloid leukemia. In a previous study, laser tweezers Raman spectroscopy (LTRs) was used to isolate adriamycin-resistant chronic myeloid leukemia cells from their parental human chronic myeloid leukemia cell line (K562). This study shows that label-free LTRs analysis combined with multivariate statistical analysis can be applied to rapidly evaluate the chemical resistance status of K562 cells at the single cell level.[21]

The examination of serum from individuals with various forms of MDS and AML has been performed; however, Raman spectroscopy has not been widely employed. Research on the serological indexes, such as glucose and lipid metabolism,

related to the transformation of MDS to AML is also nascent. Therefore, the significance of the present study lies in its ability to differentiate between MDS subtypes, primary AML, and secondary AML populations based on a clinical model established by combining Raman spectroscopy with multivariate analysis. In addition, we also looked for biomarkers of MDS and AML among Raman peaks that play an important role in disease categorization. The findings of this research provide a basis for optimizing the use of data from clinical serological tests for the speedy detection of MDS and AML in their earliest stages.

■ RESULTS

Raman Spectroscopic Analysis of Serum in Patients with MDS, Patients with AML, and Control Subjects. To study the Raman spectra of the sera of patients with MDS and AML and the control group, 173, 197, and 144 Raman spectra of the sera of the control group, MDS group, and AML group were obtained, respectively, including 42 spectra of patients with MDS-SLD/MLD, 60 spectra of patients with MDS-EB1, and 95 spectra of patients with MDS-EB2. The AML group included 114 spectra of patients with primary AML and 30 spectra of those with secondary AML. Figure 1A−C shows the serum Raman spectra of control, MDS, and AML groups, MDS subtypes, and AML subtypes in the range of 600−1800 cm^{-1}, respectively. Refer to Table S1 for the peak position of relevant serum Raman spectra. Figure 1A shows the Raman spectra of control, MDS, and AML samples. Pink, yellow, and blue vertical lines in the figure represent the peaks related to protein (643, 759, 1003, 1260, 1603, and 1654 cm^{-1}), nucleic acid (826 and 1579 cm^{-1}), and lipid (1446 cm^{-1}), respectively. Figure 1B shows the Raman spectra of control, MDS-SLD/MLD, MDS-EB1, and MDS-EB2 samples, and Figure 1C shows the Raman spectra of control, primary AML, and secondary AML samples, all showing similar peak shapes. According to the above spectral patterns and peak positions, it is difficult to identify differences in serum components in the control, MDS, and AML groups. It is necessary to combine the classification model established by orthogonal partial least squares discrimination analysis (OPLS-DA) to further screen peak positions that can effectively identify potential biomarkers for MDS and AML.

Establishment of OPLS-DA Model Based on Raman Spectroscopy to Identify MDS and AML. From the Raman spectra of control, MDS, and AML groups, 18 characteristic spectra (6 in each category) were randomly selected to form three groups of data. From the Raman spectra of MDS-SLD/MLD, MDS-EB1, MDS-EB2, and primary/secondary AML groups, 24 characteristic spectra (6 in each category) were randomly selected to form four groups of data. The sample data were analyzed and contrasted using supervised OPLS-DA. The results of the permutation study indicated that the OPLS-DA model was created and was not overfitted since the intercept of Q2 on the Y axis was negative. The Raman spectra of control, MDS, and AML serum samples, the Raman spectra of MDS-SLD/MLD, MDS-EB1/2, and primary AML serum samples, and the Raman spectra of MDS-SLD/MLD, MDS-EB1, MDS-EB2, and secondary AML serum samples could all be distinguished using cluster analysis in the OPLS-DA model with 100% accuracy. The receiver operating characteristic (ROC) curve indicated high accuracy of discriminant analysis (Figure 1D).
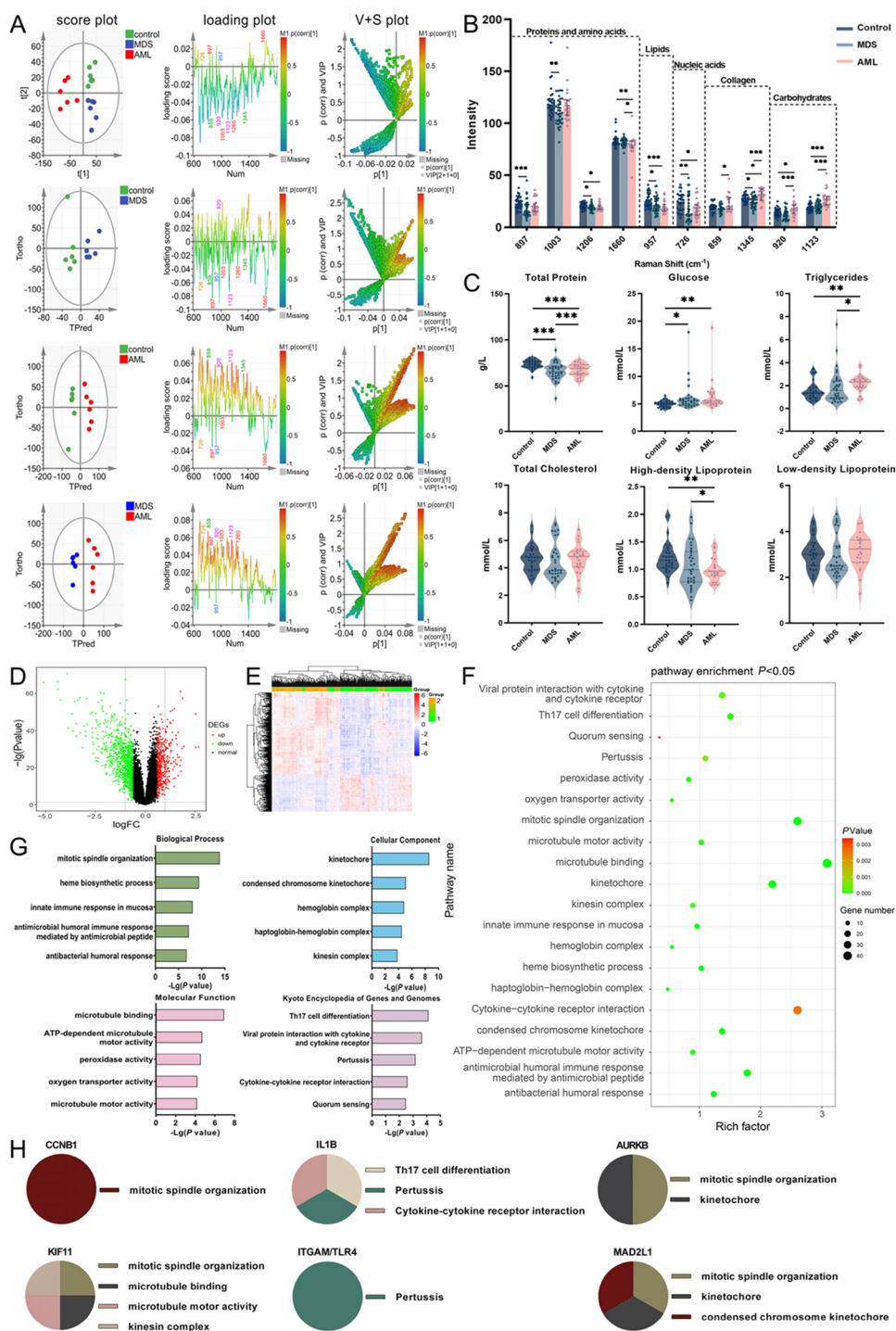
Figure 2. (A) Score plots with 95% Hotelling's confidence ellipses, loading plots, and V + S plots from OPLS models of control vs MDS vs AML, control vs MDS, control vs AML, and MDS vs AML. (B) Statistical analysis of potential biomarkers from OPLS models of control vs MDS vs AML, control vs MDS, control vs AML, and MDS vs AML. (C) Serum biochemical analysis of control, MDS, and AML groups. (D) Volcanic map of DEGs between MDS and AML groups. The X axis represents the multiple changes (logarithmic scale), while the Y axis shows the P value (logarithmic scale). Each symbol represents a different gene, and the red/blue of the symbol classifies the up/downregulated genes under different standards (P value and multiple change threshold). $P < 0.05$ was considered to indicate significance. (E) Heatmap of DEGs between MDS and AML groups. (F) Bubble diagram of functional enrichment analysis of DEGs between MDS and AML groups. The larger the bubble, the more genes enriched in that functional pathway, and the closer the color of the bubble to green, the higher the significance. (G) Biological process, CC, MF, and KEGG enrichment of DEGs between MDS and AML groups. (H) Functions and regulatory signaling pathways of genes most likely involved in the discrimination between MDS and AML.

A verification model was utilized to confirm the efficacy of the identification approach based on the identification models of control, MDS, and AML developed in accordance with the differences in the amounts of serum components. A training set and a prediction set made up the validation model. The grouping of the spectrum data was identified to validate the
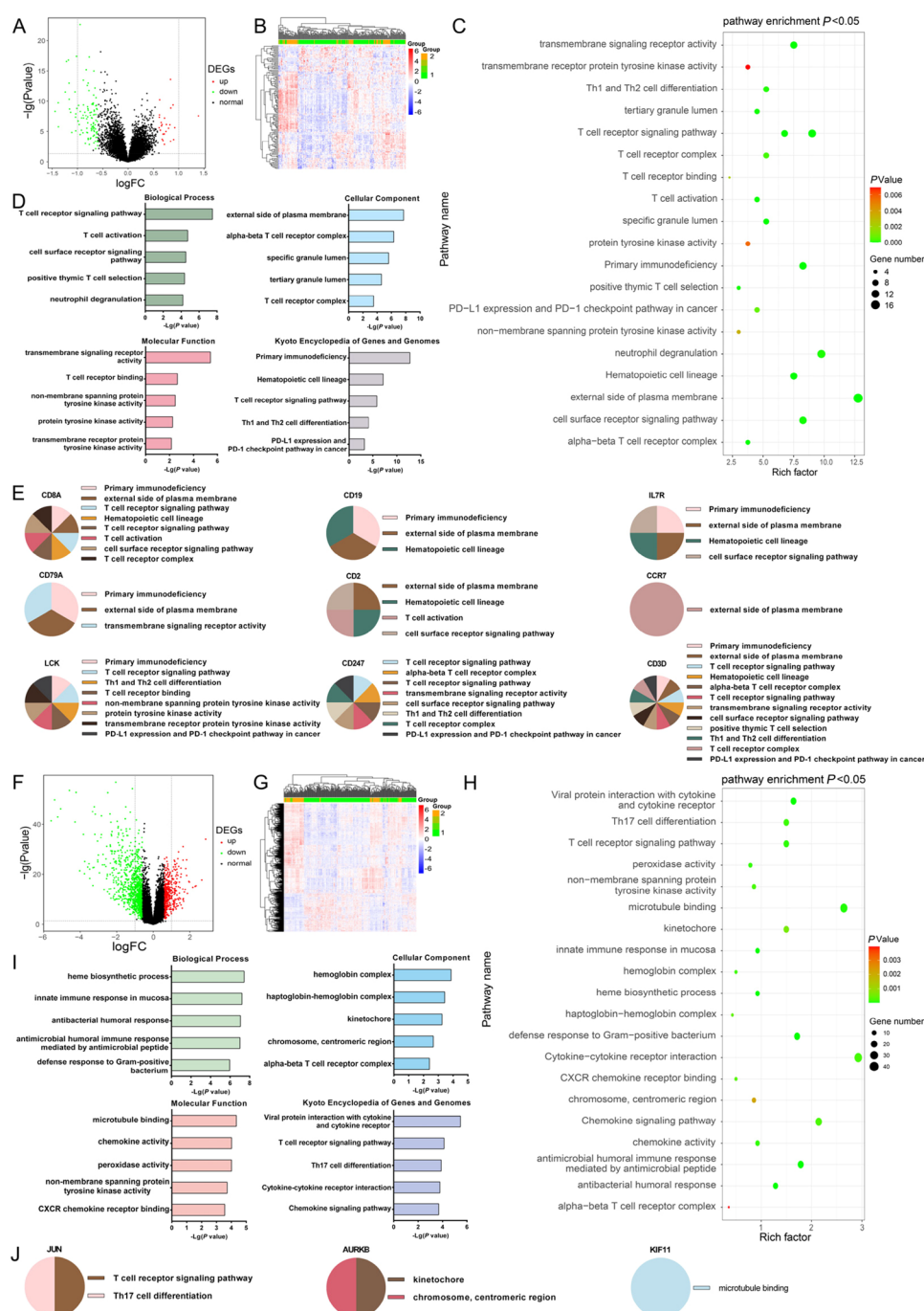
Figure 3. (A) Volcanic map of DEGs between control and MDS groups. (B) Heatmap of DEGs between control and MDS groups. (C) Bubble diagram of functional enrichment analysis of DEGs between control and MDS groups. (D) BP, CC, MF, and KEGG enrichment of DEGs between control and MDS groups. (E) Functions and regulatory signaling pathways of genes most likely involved in the discrimination between control and MDS. (F) Volcanic map of DEGs between control and AML groups. (G) Heatmap of DEGs between control and AML groups. (H) Bubble diagram of functional enrichment analysis of DEGs. (I) BP, CC, MF, and KEGG enrichment of DEGs between control and AML groups. (J) Functions and regulatory signaling pathways of genes most likely involved in the discrimination between control and AML.

model, and the training set contained the spectra of 12 patients from the same group and 12 patients from a different group. The grouping of the spectrum data for the prediction set was left unmarked, and it contained the spectra of an additional eight patients from the same group and eight patients from a different group. Specifically, nine validation models were used to assess the diagnostic model's sensitivity and specificity: control versus MDS, control versus AML, MDS versus AML, MDS-SLD/MLD versus primary AML, MDS-EB1 versus main

AML, MDS-EB2 versus primary AML, and MDS-SLD/MLD versus secondary AML. Classification scores for each spectrum in the training and prediction sets were calculated using the Raman spectral data and the SIMCA-P software. The program awarded classification scores to the two groups in the training set based on how the spectral data were grouped and then assigned classification scores to the two groups in the prediction set based on how closely their spectral data matched the spectral data in the training set. The classification

was considered correct if the first group in the training set and prediction set earned a positive score and the second group earned a negative score. The categorization was thought to be incorrect in all other respects. The diagnostic model's sensitivity and specificity are shown via a point diagram that concurrently shows samples from the training set and the prediction set according to the classification scores of the two sets (Figure 1E). Sensitivity and specificity of the classification and identification model were achieved by using a cut-off value of zero for the predicted value (Tables S4 and S5). Serum sample data from the control group, MDS/AML group, MDS subtypes, main AML, and secondary AML were all characterized using the OPLS-DA model. The nine validation models have a sensitivity range of 75−100% and a specificity range of 92−100%. The OPLS-DA model was then evaluated based on its potential usefulness in the application.

Screening and Validation of Potential Biomarkers. OPLS-DA score plot, loading plot, and V + S plot for the four models (control vs MDS vs AML, control vs MDS, control vs AML, and MDS vs AML) are shown in Figure 2A. The samples may be efficiently divided into three piles in the score plot. We were able to clearly differentiate between the control, the MDS, and the AML groups since all three clustered in opposite halves of the X axis. This finding demonstrates that OPLS-DA can effectively differentiate between serum spectral data from healthy controls, patients with MDS, and patients with AML, providing the foundation for assessing the material properties of the three groups. The Raman peaks that are useful for distinguishing between controls, MDS, and AML were first screened using the loading plot. Peak values associated with protein, nucleic acid, lipids, collagen, and carbohydrates are shown in red, yellow, blue, green, and purple, respectively, in the image. Specifically, there was a correlation between the two depictions of the relationship of sample component content with the component depicted by the peak position on the positive semi axis of the loading plot ordinate also being relatively higher in the population along the positive semi axis of the score plot abscissa. There was a comparable correlative link between the negative halves of the loading plot ordinate and the scoring plot abscissa. Identifying the three groups of samples required the characteristic peaks of protein (897, 1003, 1260, and 1660 cm^{-1}), nucleic acid (726 cm^{-1}), cholesterol/carotenoid (957 cm^{-1}), collagen (859 and 1345 cm^{-1}), and carbohydrate (920 and 1123 cm^{-1}) with the AML group having higher collagen and carbohydrate levels than the control and MDS groups, and lower cholesterol levels. The value added by peak location to the classification model is represented in the V + S graphic by a combination of the VIP and correlation coefficient. Future biomarkers were filtered using the figure. A further study looked at the peak locations that best differentiated the control, MDS, and AML groups to see whether they may serve as biomarkers. The V + S plot organized the Raman peaks from highest to lowest VIP value. Potential markers were identified by screening peaks with VIP > 1.0 and biological importance.

Based on the control vs MDS vs AML model, control, MDS, and AML samples were combined in pairs for OPLS-DA, and OPLS-DA score plot, loading plot, and V + S plot of control vs MDS, control vs AML, and MDS vs AML models were established (Figure 2A). In each of the three score plots, half of the samples fell on the positive X axis, while the other half fell on the negative Y axis. The scatter plot clearly displayed sample grouping, demonstrating OPLS-superior DA's ability to extract

differential information from spectra. The three models were able to distinguish between the two sets of samples, and the developed identification approach could detect variations in the metabolic components of serum samples. The loading plot of the three models showed that the peak intensities of protein (897, 1003, 1260, and 1660 cm^{-1}), nucleic acid (726 cm^{-1}), and cholesterol/carotenoid (957 cm^{-1}) in the control group were higher than those in the MDS group. The peak intensities of protein (897, 1003, and 1660 cm^{-1}), nucleic acid (726 cm^{-1}), collagen (859 cm^{-1}), and cholesterol/carotenoid (957 cm^{-1}) in the control group were higher than those in the AML group, while the peak intensities of collagen (859 and 1345 cm^{-1}) and carbohydrate (920 and 1123 cm^{-1}) were lower than those in the AML group. The peak intensities of protein (897, 1003, and 1260 cm^{-1}), nucleic acid (726 cm^{-1}), collagen (859 cm^{-1}), and carbohydrate (920 and 1123 cm^{-1}) in the AML group were higher than those in the MDS group, while the peak intensity of cholesterol/carotenoid (957 cm^{-1}) was lower than that in the MDS group. To identify possible biomarkers in the control, MDS, and AML models, the V + S plot of the three models might produce a list of Raman peak locations in the order of VIP values from high to low (Figure 3A). Peaks having biological importance and VIP > 1.0 were explored for as possible indicators. Important Raman peak positions affecting the sample classification were discovered in the four models of control vs MDS vs AML, control vs MDS, control vs AML, and MDS vs AML. Relevant parameters, such as VIP (VIP > 1.0), correlation coefficient, load, and distance, from the center in the V + S diagram were also cautiously being considered. Raman peak locations without a discernible change were discarded during the following biomarker verification step, which removed them from the biomarker range.

The statistical analysis of the peak locations for the Raman characteristics with VIP > 1.0 for control vs MDS, control vs MDS vs AML, control vs AML, and MDS vs AML is shown in Figure 2B. Figure 2C shows the six types of peripheral blood biochemical indexes of total protein (TP), glucose, triglyceride (TG), total cholesterol (TC), high-density lipoprotein (HDL), and low-density lipoprotein (LDL) in control, MDS, and AML groups. The peak intensities of representative proteins (897, 1003, 1206, and 1660 cm^{-1}), nucleic acid (726 cm^{-1}), collagen (1345 cm^{-1}), and cholesterol/carotenoid (957 cm^{-1}) in the control group were significantly higher than those in the MDS group. The peak intensities of cholesterol/carotenoid (957 cm^{-1}) and nucleic acid (726 cm^{-1}) in the control group were significantly higher than those in the AML group, while the peak intensities of collagen (1345 cm^{-1}) and carbohydrate (920 and 1123 cm^{-1}) were significantly lower than those in the AML group. The peak intensities of collagen (859 and 1345 cm^{-1}) and carbohydrate (920 and 1123 cm^{-1}) in the MDS group were significantly lower than those in the AML group. The aforementioned findings supported the serological findings for TP, hyperglycemia, and TC, which showed substantial variations across groups (Figure 2C).

Bioinformatics was utilized to search the differentially expressed genes between MDS and AML, control and MDS, and control and AML. Based on Raman spectroscopy and multiparameter analysis, the differentially expressed genes' (DEGs) biological roles and regulatory pathways were investigated. DEGs were screened from the GSE15061 chip data based on certain screening conditions. There were 456, 26, and 418 upregulated and 1003, 107, and 981 down-regulated DEGs in the comparisons of MDS and AML, control
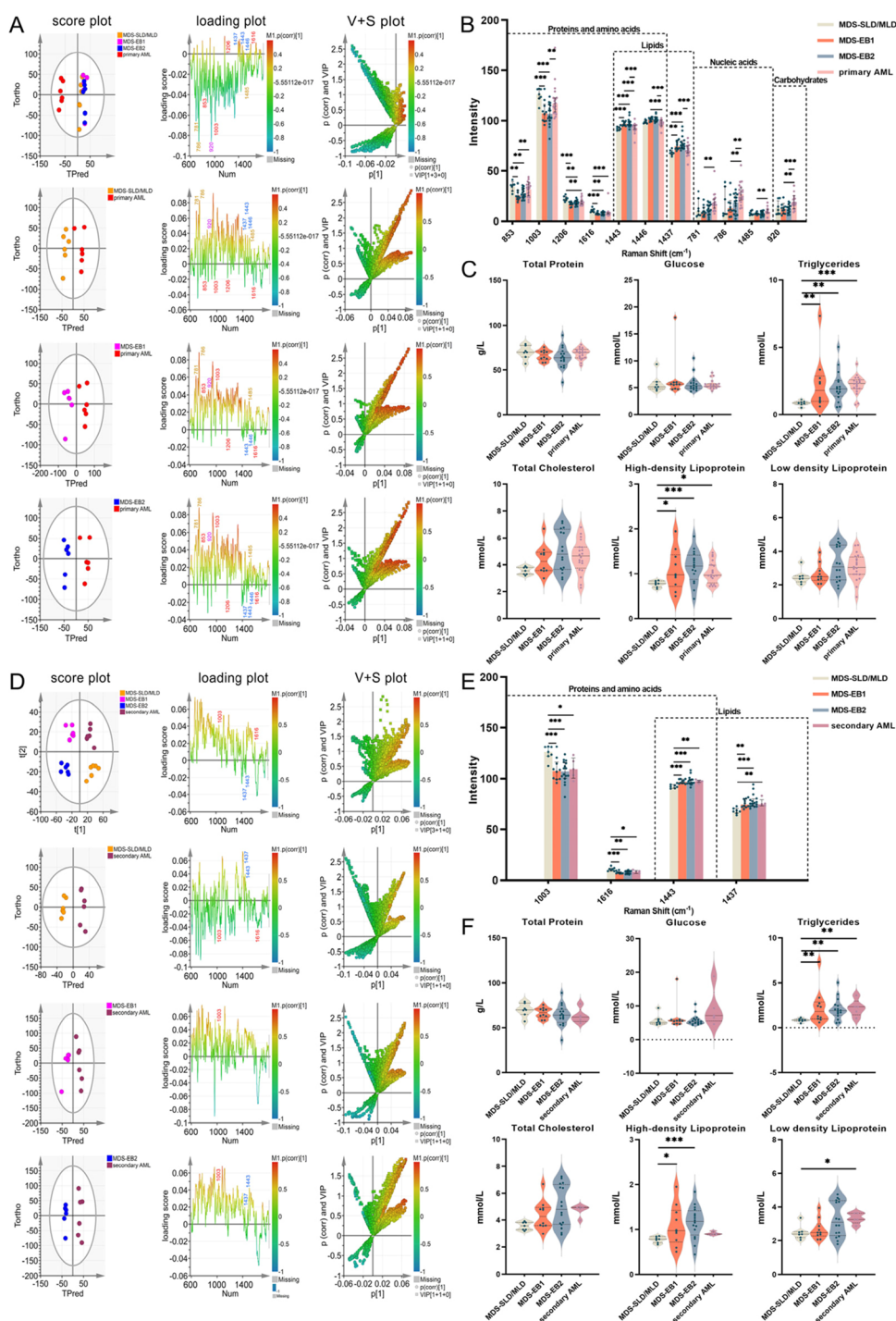
Figure 4. (A) Score plots with 95% Hotelling's confidence ellipses, loading plots, and V + S plots from OPLS models of MDS subtypes vs primary AML, MDS-SLD/MLD vs primary AML, MDS-EB1 vs primary AML, and MDS-EB2 vs primary AML. (B) Statistical analysis of potential biomarkers from OPLS models of MDS subtypes vs primary AML, MDS-SLD/MLD vs primary AML, MDS-EB1 vs primary AML, and MDS-EB2 vs primary AML. (C) Serum biochemical analysis of MDS-SLD/MLD, MDS-EB1, MDS-EB2, and primary AML. (D) Score plots with 95% Hotelling's confidence ellipses, loading plots, and V + S plots from OPLS models of MDS subtypes vs secondary AML, MDS-SLD/MLD vs secondary AML, MDS-EB1 vs secondary AML, and MDS-EB2 vs secondary AML. (E) Statistical analysis of potential biomarkers from OPLS models of MDS subtypes vs secondary AML, MDS-SLD/MLD vs secondary AML, MDS-EB1 vs secondary AML, and MDS-EB2 vs secondary AML. (F) Serum biochemical analysis of MDS-SLD/MLD, MDS-EB1, MDS-EB2, and secondary AML.

and MDS, and control and AML, respectively (Tables S6–S11). The red or blue spots on the volcano figure (Figures 3D and 4A,F) reflect strongly elevated or downregulated genes, respectively. The respective heat maps are shown in Figures 2E and 3B,G. Functional analysis of MDS and AML showed that

these DEGs were associated with Th17 cell differentiation, pertussis, and cytokine receptor interaction pathways.

DAVID was used for gene ontology (GO) and pathway enrichment analysis of common DEGs. The DEGs of MDS and AML were significantly enriched in three Kyoto

Encyclopedia of Genes and Genomes (KEGG) pathways (Th17 cell differentiation, pertussis, and cytokine receptor interaction), one GO-BP (mitotic spindle organization), three GO-CC (kinetochore, condensed chromosome kinetochore, and kinesin complex), and two GO-MF (microtubule binding and microtubule motor activity). The top five GO functions were obtained after arranging the *P* values from small to large (Figure 2F,G and Table S12).

DEGs in the control and MDS groups were significantly enriched in five KEGG pathways (hematopoietic cell lineage, T cell receiver signaling pathway, Th1 and Th2 cell differentiation, PD-L1 expression, and PD-1 checkpoint pathway in cancer), four GO-BP (T cell receiver signaling pathway, T cell activation, cell surface receiver signaling pathway, and positive thymic T cell selection), two GO-CC (alpha−beta T cell receptor complexα-βT and T cell receptor complex), and five GO-MF (transmembrane signaling receiver activity, T cell receiver binding, nonmembrane spanning protein tyrosine kinase activity, protein tyrosine kinase activity, and transmembrane receiver protein tyrosine kinase activity). The top five GO functions were screened after arranging the *P* values from small to large (Figure 3C,D and Table S13).

DEGs of the control and AML groups were significantly enriched in two KEGG pathways (T cell receiver signaling pathway and Th17 cell differentiation), no GO-BP, three GO-CC (kinetochore, chromosome, and central region), and one GO-MF (microtubule binding). The top five GO functions were screened after arranging the *P* values from small to large (Figure 3H,I and Table S14).

When the results of the three groups of differential gene analysis were combined, the GO-BP was related to mitotic spindle and T cell receptor signal transduction, CC was kinetochore and chromosome, and MF mainly involved microtubule. In KEGG analysis, regulatory signals mainly involve the T cell signaling pathway, especially the Th17 cell differentiation pathway.

Based on the signal pathway analysis of DEG, the protein−protein interaction (PPI) network was used to identify key candidate genes. The PPI network of MDS and AML had 1323 nodes and 27,566 interaction pairs. If the topological score of a node was high, it was regarded as a key node of the network and the degree values of the top 10 genes were determined (Figure S6A and Table S15). The hub genes were confirmed by using the cytoHubba plugin as *CDK1, CCNB1, IL1B, CCNA2, ITGAM, AURKB, TOP2A, KIF11, TLR4,* and *MAD2L1.* The data showed that there might be a strong interaction between them. When the results of the top five GO functions were combined, among the top 10 genes, *KIF11, AURKB, CCNB1,* and MAD2L1 were involved in mitotic spindle organization; *AURKB* and *MAD2L1* mainly related to kinetochore; and *MAD2L1* mainly related to condensed chromosome kinetochore, and *KIF11* mainly related to kinesin complex. *KIF11* plays a major role in the MFs of microtubule binding and microtubule motor activity. *IL1B* is involved in Th17 cell differentiation and cytokine receptor interaction signal pathway. *ITGAM, IL1B,* and *TLR4* are involved in the pertussis signaling pathway (Figure 2H).

The PPI network of control and MDS had 83 nodes and 702 interaction pairs. If the topological score of a node was high, it was regarded as a key node of the network and the degree values of the top 10 genes were determined (Figure S6B, Table S16). The hub genes were confirmed by using the cytoHubba plugin as *CD8a, CD19, IL7R, CD79A, CD2, CCR7, LCK,*

PAX5, CD247, and *CD3D.* The data showed that there might be a strong interaction between them. When the functional results of the top five GO genes were combined, among the top 10 genes, *CD8A, CD247,* and *CD3D* were involved in T cell receiver signaling pathway; *CD2, CD8A, CD247, IL7R,* and *CD3D* were involved in cell surface receiver signaling pathway; *CD3D* was involved in positive thymic T cell selection; *CD3D, CD2, CD79A, CD8A, CD19, CCR7,* and *IL7R* mainly related to the external side of the plasma membrane; *CD247* and *CD3D* mainly related to the alpha−beta T cell receiver complex; *CD79A, CD247,* and *CD3D* mainly related to the MF of transmembrane signaling receiver activity; *LCK* is mainly involved in the MF of T cell receiver binding; *LCK* mainly relates to the MF of nonmembrane spanning protein tyrosine kinase activity, protein tyrosine kinase activity, and transmembrane receiver protein tyrosine kinase activity; *CD79A, LCK, CD8A, CD19, IL7R,* and *CD3D* are involved in the primary immunodeficiency signaling pathway; *CD2, CD8A, CD19, IL7R,* and *CD3D* are involved in the hematopoietic cell lineage signaling pathway; *LCK, CD8A, CD247,* and *CD3D* are involved in the regulation of T cell receiver signaling pathway; *LCK, CD247,* and *CD3D* are involved in the regulation of Th1 and Th2 cell differentiation signaling pathway; and *LCK, CD247,* and *CD3D* are involved in the regulation of PD-L1 expression and PD-1 checkpoint pathway in cancer (Figure 3E).

The PPI network of control and AML had 1251 nodes and 21,942 interaction pairs. If the topological score of a node was high, it was regarded as a key node of the network and the degree values of the top 10 genes were determined (Figure S6C and Table S17). The hub genes were confirmed by using the cytoHubba plugin as *ITGAM, CD8A, CDK1, JUN, CCNB1, CCNA2, CD44, MMP9, KIF11,* and *AURKB.* The data showed that there might be a strong interaction between them. When the results of the top five GO functions were combined, among the top 10 genes, *AURKB* was mainly related to kinetochore, chromosome, and central region; *KIF11* was mainly involved in the molecular function of microtubule binding, and *JUN* was involved in T cell receiver signaling pathway and Th17 cell differentiation signaling pathway (Figure 3J).

Based on the screening of potential biomarkers of MDS and AML, the MDS subtypes model was used to analyze differences in the levels of components in the serum samples and screen potential classification markers of MDS subtypes and primary AML. Figure 4A shows the OPLS-DA score plot, loading plot, and V + S plot of MDS subtypes vs primary AML, MDS-SLD/MLD vs primary AML, MDS-EB1 vs primary AML, and MDS-EB2 vs primary AML. In the score plot, the MDS subtypes and primary AML samples were clearly grouped. The MDS subtypes group was located in the positive half of the *X* axis, and the primary AML group was located in the negative half of the *X* axis, showing that the MDS subtypes group was distinguished from the primary AML group. This result shows that OPLS-DA can well distinguish the serum spectral data of patients with MDS-SLD/MLD, MDS-EB1, MDS-EB2, and primary AML, which provides conditions for analyzing the material characteristics of the four groups. The loading plot was used to preliminarily screen the Raman peaks that contribute to the MDS subtypes vs the primary AML identification model. The red, yellow, blue, green, and purple peak numbers in the figure relate to protein, nucleic acid, lipid, collagen, and carbohydrate, respectively. The characteristic peaks of protein (853, 1003, 1206, and 1616 cm^{-1}), lipid

(1437, 1443, and 1446 cm^{-1}), nucleic acid (781, 786, and 1485 cm^{-1}), and carbohydrate (920 cm^{-1}) played important roles in the identification of five groups of samples; the main AML group had greater quantities of nucleic acid and carbohydrates than the MDS subtypes group. Potential biomarkers were further screened using the V + S plot, and the significance of peak locations was then analyzed to determine the peak positions that can effectively identify potential biomarkers for MDS subtypes and primary AML (Figure 4A).

Based on the MDS subtypes vs primary AML model, samples of the primary AML and three groups of MDS subtypes were combined for OPLS-DA and the OPLS-DA score plot, loading plot, and V + S plot of MDS-SLD/MLD vs primary AML, MDS-EB1 vs primary AML, and MDS-EB2 vs primary AML were established (Figure 4A). The two groups of samples in the three score plots were located on the positive and negative half of the X axis, respectively. Sample clustering in the scatter plot was obvious, indicating that the three models could well identify the two groups of samples in the model. As seen in the loading plot, the peak intensities of nucleic acid (781, 786, and 1485 cm^{-1}) and carbohydrate (920 cm^{-1}) in the primary AML group were generally higher than those in the MDS subtypes group and the peak intensities of lipid (1437, 1443, and 1446 cm^{-1}) were higher than those in the MDS-SLD/MLD group and lower than those in the MDS-EB1 and MDS-EB2 groups. The V + S plot provides the main basis for determining the potential classification markers in MDS subtypes and primary AML models (Figure 4A). Peaks with biological importance were chosen as the peak range of putative biomarkers from the list of peaks with VIP > 1.0 produced from the V + S plot. In MDS subtypes vs primary AML, MDS-SLD/MLD vs primary AML, MDS-EB1 vs primary AML, and MDS-EB2 vs primary AML models, the same method was used to screen and verify biomarkers.

Figure 4B shows the statistical analysis of characteristic Raman peaks with VIP > 1.0 of MDS subtypes vs primary AML, MDS-SLD/MLD vs primary AML, MDS-EB1 vs primary AML, and MDS-EB2 vs primary AML models. Figure 4C shows six types of peripheral blood biochemical indexes of TP, glucose, TG, TC, HDL, and LDL in the four groups of MDS-SLD/MLD, MDS-EB1, MDS-EB2, and primary AML. Statistical analysis showed that the peak intensity of representative protein (853, 1003, 1206, and 1616 cm^{-1}) in the MDS-SLD/MLD group was significantly higher than that in the MDS-EB1, MDS-EB2, and primary AML groups, while the peak intensity of representative lipid (1437, 1443, and 1446 cm^{-1}) was significantly lower than that in the MDS-EB1, MDS-EB2, and primary AML groups. These results were consistent with those obtained for the serological indexes, such as TG, TC, HDL, and LDL, with significant differences between groups (Figure 4C). The peak intensities of representative nucleic acid (781, 786, and 1485 cm^{-1}) and carbohydrate (920 cm^{-1}) in the MDS-EB1 and MDS-EB2 groups were significantly lower than those in the primary AML group.

Based on the results of screening for potential classification markers of MDS subtypes and primary AML, the MDS subtypes model was used to screen the potential classification markers of MDS subtypes and secondary AML. Figure 4D shows the OPLS-DA score plot, loading plot, and V + S plot. The four groups of samples in the score plot were clearly grouped and distributed in the four quadrants of the graph,

reflecting that the MDS subtypes and secondary AML groups have been distinguished. This result shows that OPLS-DA can well distinguish the serum spectral data of patients with MDS-SLD/MLD, MDS-EB1, MDS-EB2, and secondary AML, which provides conditions for analyzing the material characteristics of the four groups. Raman peaks contributing to the MDS subtypes versus the secondary AML detection model were first screened using the loading plot. Protein, nucleic acid, lipid, collagen, and carbohydrates are represented by the red, yellow, blue, green, and purple peak numbers, respectively, in the image. The lipid content of the secondary AML group was larger than that of the MDS subtypes group, and the protein characteristic peaks (1003 and 1616 cm^{-1}) and lipid peaks (1437 and 1443 cm^{-1}) played major roles in the identification of the four groups of samples. Further screening of prospective biomarkers was performed using the V + S plot, and the importance of typical peak positions was verified in future analysis to discover the peak positions that may effectively identify potential biomarkers for the control group and MDS subtypes (Figure 4D).

Based on the MDS subtypes vs secondary AML model, secondary AML and three groups of MDS subtypes samples were combined for OPLS-DA and OPLS-DA score plot, loading plot, and V + S plot (Figure 4D) of MDS-SLD/MLD vs secondary AML, MDS-EB1 vs secondary AML, and MDS-EB2 vs secondary AML models were established. The two groups of samples in the three score plots were located on the positive and negative half of the X axis, respectively, and the sample clustering was obvious in the scatter plot. Therefore, it may be considered that the two models had good discrimination ability for the two groups of samples in the model. It is seen from the loading plot that the intensity of the peak position (1003 cm^{-1}) representing protein in the secondary AML group was higher than that in the MDS-EB1 and MDS-EB2 groups but lower than that in the MDS-SLD/ MLD group. The V + S plot provides the main basis for determining the potential classification markers in the MDS subtypes and secondary AML models (Figure 4D). According to the V + S plot, a list of peaks with VIP > 1.0 was derived and peaks with biological significance were selected as the peak range of potential biomarkers according to the literature. In MDS subtypes vs secondary AML, MDS-SLD/MLD vs secondary AML, MDS-EB1 vs secondary AML, and MDS-EB2 vs secondary AML models, the same method was used to screen and verify biomarkers.

Figure 4E shows the statistical analysis of VIP > 1.0 characteristic Raman peak positions of MDS subtypes vs secondary AML, MDS-SLD/MLD vs secondary AML, MDS-EB1 vs secondary AML, and MDS-EB2 vs secondary AML. Figure 4F shows the six types of peripheral blood biochemical indexes of TP, glucose, TG, TC, HDL, and LDL in the MDS-SLD/MLD, MDS-EB1, MDS-EB2, and secondary AML groups. Statistical analysis showed that the peak intensity of representative protein (1003 and 1616 cm^{-1}) in the MDS-SLD/MLD group was significantly higher than that in the MDS-EB1, MDS-EB2, and secondary AML groups, while the peak intensity of representative lipid (1437 and 1443 cm^{-1}) was significantly lower than that in the MDS-EB1, MDS-EB2, and secondary AML groups. These results were consistent with results for the serological indexes, such as TG, TC, HDL, and LDL, with significant differences between groups (Figure 4F).

DISCUSSION

In this study, OPLS-DA could well distinguish control subjects, patients with MDS, and patients with AML based on serum spectral data. The results of this exploratory study confirm the feasibility of interpreting the heterogeneity of serum metabolism in patients with MDS and AML based on serum Raman spectroscopy and multivariate analysis, suggesting the utility of serum Raman spectroscopy in the detection of MDS, AML, and biomarker mining. The peaks representing protein (853, 1003, 1206, and 1616 cm^{-1}), lipid (1437, 1443, and 1446 cm^{-1}), collagen (859 and 1345 cm^{-1}), and carbohydrate (920 and 1123 cm^{-1}) can be used as potential biomarkers to identify MDS and AML. A decrease in peripheral blood cells in patients with MDS due to bone marrow failure may cause hypoproteinemia secondary to anemia, putting these patients at risk of developing hypoproteinemia. Due to the proliferation of malignant cells, the demand for protein and amino acids in patients with AML is higher than that in patients in the MDS group, and therefore the risk of hypoproteinemia in patients with AML is higher.

High TG levels indicate impaired cellular energy metabolism. Since damaged cells cannot efficiently burn sugar, the body must maintain high levels of TG in the blood.[22−24] The peripheral serum TG level of patients with AML was higher than that in patients with MDS and control subjects, suggesting the risk of hyperlipidemia in patients with AML. Compared with the MDS-SLD/MLD group, the peak intensity representing lipid (1437, 1443, and 1446 cm^{-1}) in primary AML and secondary AML groups was significantly higher, consistent with the results of serum biochemical analysis. This may indicate an abnormality in cellular energy metabolism in patients with AML, which is consistent with our previous findings. According to our previous report on the analysis of AML subtypes, the analysis of bone marrow supernatant of patients with AML suggested that the presence of leukemia cells in the bone marrow microenvironment resulted in lower serum levels of TC, HDL, and LDL in patients with AML than in control subjects, indicating that AML patients had lipid metabolism disorder.[25]

Statistical analysis revealed that the 957 cm^{-1} peak intensity in the AML and MDS groups was lower than that in the control group with the difference being significant between the AML and the control groups. The 957 cm^{-1} peak represents both lipid and β-carotenoids. We speculate that the antioxidant level in patients with AML and MDS is low, which is not conducive to free radical scavenging, while the high carotenoid level in healthy individuals inhibits tumors, such as AML and MDS.[26−28] In the remission phase of leukemia, the plasma carotenoid level increases, which may also act as a self-protection mechanism.[29]

Based on the screening of MDS- and AML-related genes in the Gene Expression Omnibus (GEO) database, we retrieved ITGAM, IL1B, and TLR4 genes as key differential genes, which are involved in the pertussis signaling pathway. Pertussis is a type of disease that can trigger a lymphoid leukemia reaction. Pertussis toxin can inhibit T cell movement and displacement by inhibiting the activation of G protein or "guanylate binding protein." It is a special protein related to transmembrane signal transduction in the cell membrane. The peak intensity of MDS-EB1 in primary AML and MDS subtypes was significantly different at 1485 cm^{-1}. The peak at 1485 cm^{-1} is related to guanine metabolism. Patients with

MDS-EB1 have insufficient T cell function and carry a high risk of transforming into AML. Abnormal nucleic acid metabolism in the MDS-EB1 metabolic microenvironment may be related to the transformation potential of MDS to AML, but further study is necessary in this regard.

Using bioinformatics analysis, we also found that DEGs between MDS and AML related to the pathway of Th17 cell differentiation. Th17 is a T cell subset that shows heterogeneity and plasticity in different immune environments, closely related to the occurrence of autoimmune diseases and tumors. The metabolic microenvironment critically influences the differentiation and function of Th17.[30] The serine/threonine kinase Akt signaling pathway is necessary for the peripheral induction of Th17.[31,32] Serine and threonine are hydroxyaliphatic side chain amino acids. The difference in the peak position of lipids (1437, 1443, and 1446 cm^{-1}) in the blood microenvironment of MDS and AML may be related to the transformation from MDS to AML.[33,34]

AML affects the differentiation and proliferation of hematopoietic cells due to clonal abnormalities, while MDS affects only the differentiation of hematopoietic cells due to clonal abnormalities. Therefore, in theory, MDS can be regarded as "pre-leukemia." In the OPLS-DA model, the MDS and AML groups could be well identified and cluster analysis yielded good results. The results of Raman spectrum analysis of serum samples agreed well with those of serum biochemical analysis results, but due to the influence of sample size, serum TP, glucose, TC, and LDL levels did not differ significantly, indicating that the Raman spectrum detection of serum was more sensitive and could capture subtle differences in the component levels that cannot otherwise be captured by conventional serological detection.

The differential diagnosis between MDS and AML is challenging because massive serological data related to glucose and lipid metabolism and corresponding clinical diagnostic indicators are lacking. The present study highlights the need for further mining serological detection data. Statistical analysis of larger samples will help to identify new methods for the rapid and early identification of different types of MDS and AML. Raman spectroscopy, which is both quick and cost-effective, will lay a solid foundation for the development of big data-aided diagnosis and application software in the future.

In conclusion, the close combination of serum Raman spectroscopy and OPLS-DA can scientifically detect differences in the structure, components, and content of biomacromolecules in the serum of patients with different types of MDS and AML. Compared with bone marrow puncture, Raman spectroscopy analysis of serum samples is less traumatic and easily accepted by patients, which is conducive to follow-up research. The identification of different types of MDS and AML has far-reaching scientific research and clinical significance.

LIMITATIONS OF THE STUDY

However, the patient's medical histories, drug histories, and smoking and drinking habits were not thoroughly explored in this research, which might have influenced the outcomes. Larger sample size and the use of consistent techniques of data collection are therefore required to enhance the reliability of samples. And although the control group, MDS subgroup, and AML subgroup all had about the same percentage of male and female participants, the age distribution was however different. The control group had individuals in the age range of 12−69

years, the MDS-SLD/MLD group had patients who were 19–65 years old, the MDS-EB1 group had patients who were 34–77 years old, the MDS-EB2 group had patients of age 37–73 years, the primary AML group had patients of age 5–64 years, and the secondary AML group had patients who were 34–74 years old. The difference in age distribution among the six groups was statistically significant (Tables S2 and S3). However, the metabonomic analysis and statistical methods used to analyze the indicators basically met the needs for comparing the differences between samples in the study, have prompt value for clinical practice guidelines, and also provide a reference for further follow-up research.

CONCLUSIONS

To our knowledge, this work is the first application of serum Raman spectroscopy to interpret the heterogeneity of serum metabolism in patients with MDS and AML. Preliminary analysis of Raman spectra showed specific biomolecular differences between MDS and AML, which may be caused by changes in the patients' body metabolism. Peak collagen (859 and 1345 cm^{-1}) and carbohydrate (920 and 1123 cm^{-1}) intensities were substantially lower in the MDS group as compared to the AML group. Screening and bioinformatics analysis of MDS- and AML-related genes based on the GEO database revealed 1459 DEGs, and GO function was mainly related to mitotic spindle organization. The main cell components were kinetochore, condensed chromosome kinetochore, and kinesin complex, and the main molecular function was related to microtubule binding and microtubule motor activity. KEGG analysis of the DEGs between MDS and AML revealed the main signaling pathways as Th17 cell differentiation, pertussis, and cytokine receptor interaction. The 10 hub genes screened in the PPI network were *CDK1*, *CCNB1*, *IL1B*, *CCNA2*, *ITGAM*, *AURKB*, *TOP2A*, *KIF11*, *TLR4*, and *MAD2L1*.

Furthermore, the peak intensity of representative proteins (853, 1003, 1206, and 1616 cm^{-1}) in the MDS-SLD/MLD group was significantly higher than that in the MDS-EB1, MDS-EB2, and primary AML groups, while the peak intensity of representative lipid (1437, 1443, and 1446 cm^{-1}) was significantly lower than that in the MDS-EB1, MDS-EB2, and primary AML groups. The peak intensities of representative nucleic acid (781, 786, and 1485 cm^{-1}) and carbohydrate (920 cm^{-1}) in the MDS-EB1 and MDS-EB2 groups were significantly lower than those in the primary AML group. The peak intensity of representative protein (1003 and 1616 cm^{-1}) in the MDS-SLD/MLD group was significantly higher than that in the MDS-EB1, MDS-EB2, and secondary AML group, while the peak intensity of representative lipid (1437 and 1443 cm^{-1}) was significantly lower than that in the MDS-EB1, MDS-EB2, and secondary AML group. In particular, combined with the statistical analysis of serological indexes related to glucose and lipid metabolism, the peaks at 920 and 1123 cm^{-1} closely related to glucose can be used as potential biomarkers for the identification of MDS; peaks at 853, 1003, 1206, and 1616 cm^{-1} closely related to TP, TG, TC, HDL, and LDL, and 1437, respectively, can be used to identify AML; and peaks at 1443 and 1446 cm^{-1} can be used as potential biomarkers for the early diagnosis of MDS-SLD/MLD. The results of this exploratory study indicate the potential of applying Raman spectrum serum analysis as a clinical tool for the noninvasive detection and screening of potential biomarkers to identify MDS and AML. The potential

correlation between the massive serological examination information of patients with MDS and AML and the classification and prognosis of the disease was determined. However, the sample size of this study was small; therefore, the results may not be accurate. Considering that patients and samples with different types of MDS and AML are rare and the results are interpretable, we will conduct a more detailed prospective study in the next step to test the feasibility of identifying biomarkers related to MDS and AML.

MATERIALS AND METHODS

Sample Collection. The research included 87 participants, including 44 men and 43 females, with age ranging from 5 to 77 years, all of whom were recruited in 2021 from the Hospital of Blood Diseases, Chinese Academy of Medical Sciences (Institute of Hematology, Chinese Academy of Medical Sciences). There were 33 patients in the MDS group, including seven patients with MDS-SLD/MLD, 10 patients with MDS-EB1, and 16 patients with MDS-EB2. There were 25 patients with AML, including 20 patients with primary AML and five patients with secondary AML. The control group comprised 29 healthy individuals.

All patients in the experimental group underwent examination by blood phase, bone marrow phase, cytogenetics, immunological phenotype analysis, and gene analysis, and the results were confirmed by experienced hematology experts. The primary diagnosis of these cases is hematopoietic diseases, namely MDS or AML. The Chinese Academy of Medical Sciences' Hospital for Blood Diseases Ethics Committee gave its approval to this work (KT2020016-EC-2). All participating healthy individuals and patients provided written informed consent. Serum samples were obtained from patients with MDS, patients with AML, and individuals from the control group. All subjects underwent routine serum biochemical testing at the clinical testing center of the Hospital of Blood Diseases, Chinese Academy of Medical Sciences (Institute of Hematology, Chinese Academy of Medical Sciences).

General Biochemical Data. Serum samples obtained from the subjects who fasted for 10 h were used for the analysis of TP, TC, glucose, TG, LDL, and HDL in the peripheral blood using an automatic biochemical analyzer.

Raman Spectroscopic Analysis of Peripheral Blood Serum. On a quartz slide, we deposited 5.0 μL of serum for analysis using a confocal Raman spectrometer HORIBA Xplora Raman microscope equipped with a 785 nm laser, 40 MW of output power, a 40× objective lens, and an XYZ three-dimensional specimen platform. Imaging was performed using a 40× (0.75 numerical aperture) Nikon lens, approximately 2 × 2 μm spot size range to receive laser beam irradiation, the output power of 10 MW, the single integration time of 250 s, measurement range of 600–1800 cm^{-1}, and resolution of 1 cm^{-1}. From 5 to 10, various spots were measured for each group. To have a baseline, we measured the Raman spectra of the quartz slide. Labspec6 was used to perform tasks, including smoothing, background subtraction, and baseline correction on the collected data. Spectra were all calibrated relative to the Raman signal at 1450 cm^{-1}.

Diagnostic Model Establishment Based on Raman Spectral Data Analysis. Patients with MDS, AML patients, and healthy controls all had their blood Raman spectra analyzed using supervised OPLS-DA performed in SIMCA 14.1. The OPLS model's efficacy was measured by the goodness of fit metrics R^2 and Q^2. Under the null hypothesis,
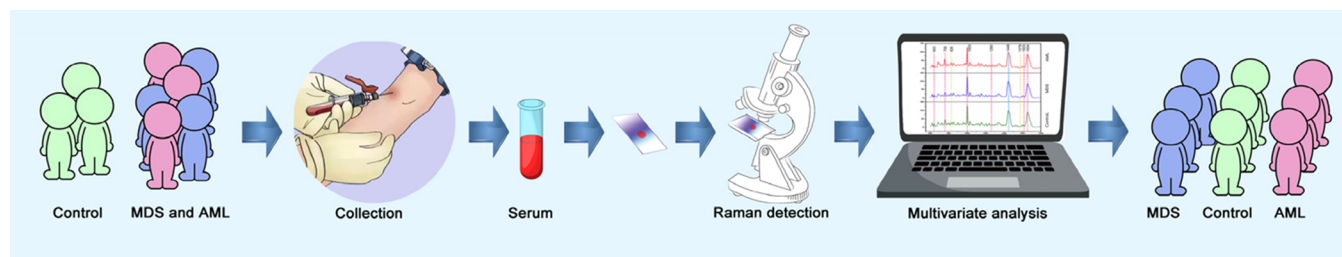
Figure 5. Schematic diagram of serum biochemical analysis by Raman spectroscopy. The data set was analyzed by multivariate statistical analysis using SIMCA 14.1.

we randomly changed the Y matrix 200 times and resampled the data to see how well the model held up. Raman peaks of statistical significance were identified as potential biomarkers using a classification model that included cluster analysis and V + S analysis. On the basis of the V + S plot's correlation coefficient, loadings, and distance from the center, the peak site of a potential biomarker was selected as a VIP > 1.0. The collected candidates for biomarkers were subjected to a significance test, and those that had a P value lower than 0.05 were deemed to have high clinical utility. Processing of pertinent data was done in Origin. Statistics were analyzed using IBM SPSS Statistics 20, and graphics were created with GraphPad Prism 5. Figure 5 depicts the conceptual framework for this study, which includes the collection of serum samples, the detection of Raman spectra, multivariate analysis, and the construction of the identification model.

Bioinformatics Analysis. From the GEO database, we obtained and downloaded gene chip GSE15061. DEGs were compared and identified between control and MDS, control and AML, and MDS and AML using the R program and the Bioconductor software package. P 0.01 and |logFC| > 0.6 were used to filter the DEGs. Following this, the screened DEGs were subjected to GO function annotation and KEGG signal pathway enrichment analysis using DAVID6.8. Cytoscape was used to mine the main genes that play an important part in the biological processes connected to alterations in the control, MDS, and AML groups using the PPI network analysis provided using the STRING online analysis tool.

Statistical Analysis. To analyze the data, we utilized SPSS 26.0. We compared the frequencies using the chi-square test (also called Fisher's exact test). The mean ± standard deviation is used to represent data with a normal distribution, whereas the median is used to describe data without a normal distribution (25th−75th percentile). The comparison of data groups with normal distributions was performed using one-way ANOVA. The pairwise comparison of homogenous variance groups was performed using the least significant difference approach. The pairwise comparison of heterogeneous variance groups was conducted using Tamhane's T2 approach. The nonparametric Kruskal−Wallis test technique was used to compare groups of data with nonconforming normal distributions. A statistically significant difference was defined as $P < 0.05$.

ASSOCIATED CONTENT

Supporting Information

The Supporting Information is available free of charge at https://pubs.acs.org/doi/10.1021/acsomega.2c06170.

Statistical analysis of potential biomarkers from OPLS models and serum biochemical analysis of control vs

MDS subtypes and control vs AML subtypes, permutation, cluster, ROC, and validation plots of OPLS discrimination for the control, MDS and AML groups, Raman spectra of serum samples from the control, MDS, and AML groups, degree value of screened key differential candidate genes for control, MDS and AML groups, peak assignments for Raman spectra of serum, clinical data for the MDS, AML and control groups, clinical data of the control subjects, patients with MDS, and patients with AML subtypes, sensitivity and specificity of control, MDS and AML diagnostic models, DEGs identified in GSE15061, top 5 data of gene ontology and KEGG signaling pathway enrichment analysis, and PPI network degree value top 10 list of DEGs (PDF)

AUTHOR INFORMATION

Corresponding Authors

Jianwei Qi − *State Key Laboratory of Experimental Hematology, National Clinical Research Center for Blood Diseases, Haihe Laboratory of Cell Ecosystem, Institute of Hematology & Blood Diseases Hospital, Chinese Academy of Medical Sciences & Peking Union Medical College, Tianjin 300020, China*; Email: qijianwei@ihcams.ac.cn

Qiang Zhang − *Department of Geriatrics, Tianjin Geriatrics Institute, Tianjin Medical University General Hospital, Tianjin 300052, China*; Email: zhangqiangyulv@163.com

Yuan Zhou − *State Key Laboratory of Experimental Hematology, National Clinical Research Center for Blood Diseases, Haihe Laboratory of Cell Ecosystem, Institute of Hematology & Blood Diseases Hospital, Chinese Academy of Medical Sciences & Peking Union Medical College, Tianjin 300020, China*; Email: yuanzhou@ihcams.ac.cn

Authors

Haoyue Liang − *State Key Laboratory of Experimental Hematology, National Clinical Research Center for Blood Diseases, Haihe Laboratory of Cell Ecosystem, Institute of Hematology & Blood Diseases Hospital, Chinese Academy of Medical Sciences & Peking Union Medical College, Tianjin 300020, China*; orcid.org/0000-0002-3766-7605

Xiaodong Kong − *Department of Geriatrics, Tianjin Geriatrics Institute, Tianjin Medical University General Hospital, Tianjin 300052, China*

Haoyu Wang − *State Key Laboratory of Experimental Hematology, National Clinical Research Center for Blood Diseases, Haihe Laboratory of Cell Ecosystem, Institute of Hematology & Blood Diseases Hospital, Chinese Academy of Medical Sciences & Peking Union Medical College, Tianjin 300020, China*

Yansong Ren − *State Key Laboratory of Experimental Hematology, National Clinical Research Center for Blood Diseases, Haihe Laboratory of Cell Ecosystem, Institute of Hematology & Blood Diseases Hospital, Chinese Academy of Medical Sciences & Peking Union Medical College, Tianjin 300020, China*

Ertao Liu − *State Key Laboratory of Experimental Hematology, National Clinical Research Center for Blood Diseases, Haihe Laboratory of Cell Ecosystem, Institute of Hematology & Blood Diseases Hospital, Chinese Academy of Medical Sciences & Peking Union Medical College, Tianjin 300020, China*

Fanfan Sun − *State Key Laboratory of Experimental Hematology, National Clinical Research Center for Blood Diseases, Haihe Laboratory of Cell Ecosystem, Institute of Hematology & Blood Diseases Hospital, Chinese Academy of Medical Sciences & Peking Union Medical College, Tianjin 300020, China*

Complete contact information is available at:
https://pubs.acs.org/10.1021/acsomega.2c06170

■ REFERENCES

(1) Fang, C.; Rao, S.; Crispino, J. D.; Ntziachristos, P. Determinants and role of chromatin organization in acute leukemia. *Leukemia* **2020**, *34*, 2561−2575.

(2) Lorentino, F.; Labopin, M.; Ciceri, F.; Vago, L.; Fleischhauer, K.; Afanasyev, B.; Kröger, N.; Cornelissen, J. J.; Lovira, M.; Meijer, E.; Vitek, A.; Elmaagacli, A.; Blaise, D.; Ruggeri, A.; Chabannon, C.; Nagler, A.; Mohty, M. Acute Leukemia Working Party (ALWP) and Cellular Therapy and Immunobiology Working Party (CTIWP) of the European Society for Blood and Marrow Transplantation (EBMT). *Leukemia* **2021**, *35*, 585−594.

(3) Kayser, S.; Levis, M. J. Advances in targeted therapy for acute myeloid leukaemia. *Br. J. Haematol.* **2018**, *180*, 484−500.

(4) Kayser, S.; Levis, M. J. Updates on targeted therapies for acute myeloid leukaemia. *Br. J. Haematol.* **2022**, *196*, 316−328.

(5) Kantarjian, H. M.; Jain, N.; Garcia-Manero, G.; Welch, M. A.; Ravandi, F.; Wierda, W. G.; Jabbour, E. J. The cure of leukemia through the optimist's prism. *Cancer* **2022**, *128*, 240−259.

(6) Menssen, A. J.; Walter, M. J. Genetics of progression from MDS to secondary leukemia. *Blood* **2020**, *136*, 50−60.

(7) Dunbar, A. J.; Rampal, R. K.; Levine, R. Leukemia secondary to myeloproliferative neoplasms. *Blood* **2020**, *136*, 61−70.

(8) Sun, L.; Babushok, D. V. Secondary myelodysplastic syndrome and leukemia in acquired aplastic anemia and paroxysmal nocturnal hemoglobinuria. *Blood* **2020**, *136*, 36−49.

(9) Bauer, M.; Vaxevanis, C.; Heimer, N.; Al-Ali, H. K.; Jaekel, N.; Bachmann, M.; Wickenhauser, C.; Seliger, B. Expression, Regulation and Function of microRNA as Important Players in the Transition of MDS to Secondary AML and Their Cross Talk to RNA-Binding Proteins. *Int. J. Mol. Sci.* **2020**, *21*, 7140.

(10) Goel, H.; Rahul, E.; Gupta, I.; Chopra, A.; Ranjan, A.; Gupta, A. K.; Meena, J. P.; Viswanathan, G. K.; Bakhshi, S.; Misra, A.; Hussain, S.; Kumar, R.; Singh, A.; Rath, A. S.; Sharma, A.; Mittan, S.; Tanwar, P. Molecular and genomic landscapes in secondary & therapy related acute myeloid leukemia. *Am. J. Blood Res.* **2021**, *11*, 472−497.

(11) Fenaux, P.; Platzbecker, U.; Ades, L. How we manage adults with myelodysplastic syndrome. *Br. J. Haematol.* **2020**, *189*, 1016−1027.

(12) Garcia-Manero, G.; Chien, K. S.; Montalban-Bravo, G. Myelodysplastic syndromes: 2021 update on diagnosis, risk stratification and management. *Am. J. Hematol.* **2020**, *95*, 1399−1420.

(13) AlMozain, N.; Mashi, A.; Alneami, F.; Al-Omran, A.; Bakshi, N.; Owaidah, T.; Khalil, S.; Khogeer, H.; Hashmi, S.; Al-Sweedan, S.; Morris, D.; AlNounou, R. Spectrum of myelodysplastic syndrome in patients evaluated for cytopenia(s). A Report from a Reference Centre in Saudi Arabia. *Hematol. Oncol. Stem Cell Ther.* **2022**, *15*, 39−44.

(14) Strupp, C.; Nachtkamp, K.; Hildebrandt, B.; Giagounidis, A.; Haas, R.; Gattermann, N.; Bennett, J. M.; Aul, C.; Germing, U. New proposals of the WHO working group (2016) for the diagnosis of myelodysplastic syndromes (MDS): Characteristics of refined MDS types. *Leuk. Res.* **2017**, *57*, 78−84.

(15) Estey, E.; Hasserjian, R. P.; Döhner, H. Distinguishing AML from MDS: a fixed blast percentage may no longer be optimal. *Blood* **2022**, *139*, 323−332.

(16) Ogawa, S. Genetics of MDS. *Biosens. Bioelectron.* **2019**, *133*, 1049−1059.

(17) Zhang, B. Y.; Yin, P.; Hu, Y.; Szydzik, C.; Khan, M. K.; Xu, K.; Thurgood, P.; Mahmood, N.; Dekiwadia, C.; Afrin, S.; Yang, Y.; Ma, Q.; McConville, C. F.; Khoshmanesh, K.; Mitchell, A.; Hu, B.; Baratchi, S.; Ou, J. Z. Highly accurate and label-free discrimination of single cancer cell using a plasmonic oxide-based nanoprobe. *Biosens. Bioelectron.* **2022**, *198*, No. 113814.

(18) Kronstein-Wiedemann, R.; Stadtmüller, M.; Traikov, S.; Georgi, M.; Teichert, M.; Yosef, H.; Wallenborn, J.; Karl, A.; Schütze, K.; Wagner, M.; El-Armouche, A.; Tonn, T. SARS-CoV-2 Infects Red Blood Cell Progenitors and Dysregulates Hemoglobin and Iron Metabolism. *Stem Cell Rev. Rep.* **2022**, *18*, 1809−1821.

(19) Soysal, K. B.; Parlatan, S.; Mastanzade, M.; Ozbalak, M.; Yenerel, M. N.; Unlu, M. B.; Basar, G.; Parlatan, U. Raman tweezers as an alternative diagnostic tool for paroxysmal nocturnal hemoglobinuria. *Anal. Methods* **2021**, *13*, 3963−3969.

(20) Grieve, S.; Puvvada, N.; Phinyomark, A.; Russell, K.; Murugesan, A.; Zed, E.; Hassan, A.; Legare, J. F.; Kienesberger, C. P.; Pulinilkunnil, T.; Reiman, T.; Scheme, E.; Brunt, R. K. Nanoparticle surface-enhanced Raman spectroscopy as a noninvasive, label-free tool to monitor hematological malignancy. *Nanomedicine* **2021**, *16*, 2175−2188.

(21) Zhang, Q.; Ye, M.; Wang, L.; Jiang, D.; Yao, S.; Lin, D.; Chen, Y.; Feng, S.; Yang, T.; Hu, J. Characterization of Drug Resistance in Chronic Myeloid Leukemia Cells Based on Laser Tweezers Raman Spectroscopy. *Appl. Spectrosc.* **2021**, *75*, 1296−1304.

(22) Sun, J.; Lou, Y.; Zhu, J.; Shen, H.; Zhou, D.; Zhu, L.; Yang, X.; Xie, X.; Li, L.; Huang, X.; Zhu, M.; Zheng, Y.; Xie, W.; Ye, X.; Jin, J.;

Zhu, H. H. Hypertriglyceridemia in Newly Diagnosed Acute Promyelocytic Leukemia. *Front. Oncol.* **2020**, *10*, No. 577796.

(23) Pabst, T.; Kortz, L.; Fiedler, G. M.; Ceglarek, U.; Idle, J. R.; Beyoğlu, D. The plasma lipidome in acute myeloid leukemia at diagnosis in relation to clinical disease features. *BBA Clin.* **2017**, *7*, 105−114.

(24) Morel, S.; Amre, D.; Teasdale, E.; Caru, M.; Laverdière, C.; Krajinovic, M.; Sinnett, D.; Curnier, D.; Levy, E.; Marcil, V. Dietary Intakes Are Associated with HDL-Cholesterol in Survivors of Childhood Acute Lymphoblastic Leukaemia. *Nutrients* **2019**, *11*, 2977.

(25) Liang, H.; Cheng, X.; Dong, S.; Wang, H.; Liu, E.; Ru, Y.; Li, Y.; Kong, X.; Gao, Y. Rapid and non-invasive discrimination of acute leukemia bone marrow supernatants by Raman spectroscopy and multivariate statistical analysis. *J. Pharm. Biomed. Anal.* **2022**, *210*, No. 114560.

(26) da Silva, A. M.; de Siqueira, E.; Oliveira, F.; de Brito, P.; Silveira, L. Spectral model for diagnosis of acute leukemias in whole blood and plasma through Raman spectroscopy. *J. Biomed. Opt.* **2018**, *23*, 1−11.

(27) Laudisio, D.; Castellucci, B.; Barrea, L.; Pugliese, G.; Savastano, S.; Colao, A.; Muscogiuri, G. Mediterranean diet and breast cancer risk: a narrative review. *Minerva Endocrinol.* **2021**, *46*, 441−452.

(28) Chaudhary, P.; Sharma, A.; Singh, B.; Nagpal, A. K. Bioactivities of phytochemicals present in tomato. *J. Food Sci. Technol.* **2018**, *55*, 2833−2849.

(29) González-Solís, J. L.; Martínez-Espinosa, J. C.; Salgado-Román, J. M.; Palomares-Anda, P. Monitoring of chemotherapy leukemia treatment using Raman spectroscopy and principal component analysis. *Lasers Med. Sci.* **2014**, *29*, 1241−1249.

(30) Yu, C. Y.; Peng, W. M.; Oldenburg, J.; Hoch, J.; Bieber, T.; Limmer, A.; Hartmann, G.; Barchet, W.; Eis-Hübinger, A. M.; Novak, N. Human plasmacytoid dendritic cells support Th17 cell effector function in response to TLR7 ligation. *J. Immunol.* **2010**, *184*, 1159−1167.

(31) DiToro, D.; Harbour, S. N.; Bando, J. K.; Benavides, G.; Witte, S.; Laufer, V. A.; Moseley, C.; Singer, J. R.; Frey, B.; Turner, H.; Bruning, J.; Darley-Usmar, V.; Gao, M.; Conover, C.; Hatton, R. D.; Frank, S.; Colonna, M.; Weaver, C. T. Insulin-Like growth factors are key regulators of T helper 17 regulatory T cell balance in autoimmunity. *Immunity* **2020**, *52*, 650−667.

(32) Li, P.; Ji, M.; Park, J.; Bunting, K. D.; Ji, C.; Tse, W. Th17 related cytokines in acute myeloid leukemia. *Front. Biosci.* **2012**, *17*, 2284−2294.

(33) Han, Y.; Ye, A.; Bi, L.; Wu, J.; Yu, K.; Zhang, S. Th17 cells and interleukin-17 increase with poor prognosis in patients with acute myeloid leukemia. *Cancer Sci.* **2014**, *105*, 933−942.

(34) Abousamra, N. K.; Salah El-Din, M.; Helal, R. Prognostic value of Th17 cells in acute leukemia. *Med. Oncol.* **2013**, *30*, 732.