



Biochemical fingerprinting of dried blood serum from chronic lymphocytic leukemia patients by Raman spectroscopy: Towards prognostic classification

E. Fazio^a,*, C. Corsaro^a, A. Speciale^b, A. Saija^b, F.L. Salamone^b, M. Mondello^c, R. Sciacotta^d, L. Giordano^d, F. Stagno^d, A. Allegra^d

^a Department of Mathematics and Computer Sciences, Physical Sciences and Earth Sciences, University of Messina, Viale F. Stagno d'Alcontres, 31, Messina, 98166, Italy

^b Department of Chemical Sciences, Biological, Pharmaceutical and Environmental, University of Messina, Viale F. Stagno d'Alcontres, 31, Messina, 98166, Italy

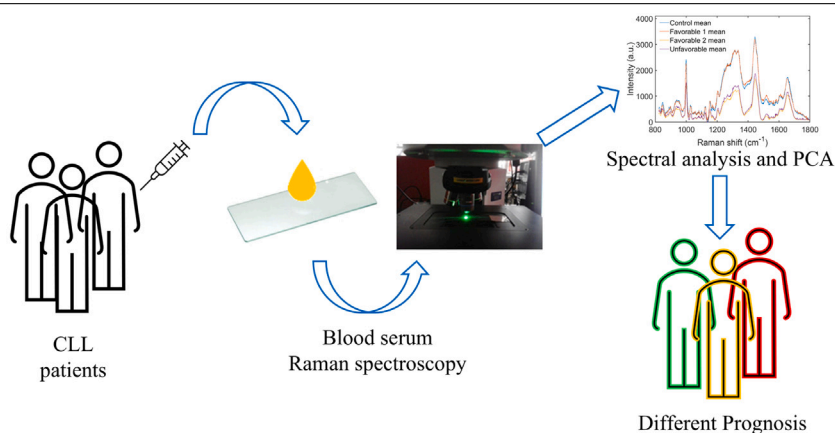
^c Complex Operational Unit of Clinical Pathology, Papardo Hospital, Contrada Papardo, Messina, 98158, Italy

^d Division of Hematology, Department of Human Pathology in Adulthood and Childhood "Gaetano Barresi", University of Messina, Via Consolare Valeria 1, Messina, 98125, Italy

HIGHLIGHTS

- Raman spectroscopy is a promising approach for analyzing blood serum from Chronic Lymphocytic Leukemia patients.
- Chronic Lymphocytic Leukemia serum Raman profiling combined with statistical analyses allows prognostic classification.
- Statistical analyses appear able to identify different subclusters within the favorable prognosis group.

GRAPHICAL ABSTRACT



ARTICLE INFO

Keywords:

Chronic lymphocytic leukemia
Micro-Raman spectroscopy
Serum biomolecular vibrational modes
Label-free biomarker detection
Multivariate analysis
Prognostic differentiation

ABSTRACT

This study presents the application of micro-Raman spectroscopy to the serum of Chronic Lymphocytic Leukemia (CLL) patients for prognostic assessment, building upon preliminary identification by conventional methods. This label-free, rapid vibrational technique reveals distinct biochemical signatures that differentiate favorable from unfavorable prognoses, surpassing insights obtainable through morphology or immunophenotyping alone. Key spectral variations, notably at 1652 and 1205 cm^{-1} , linked to protein secondary structure and amino acid makeup, and the ratio between 1003 cm^{-1} (phenylalanine) and 1344 cm^{-1} (collagen) peaks, reflect metabolic changes correlating with disease progression. Integrating Raman spectra analysis with Principal Component Analysis (PCA) and Partial least squares (PLS) enabled accurate classification, highlighting micro-Raman's innovative potential as a complementary prognostic tool for CLL despite the complexity of spectral interpretation.

* Corresponding author.

E-mail address: enfazio@unime.it (E. Fazio).

1. Introduction

Chronic Lymphocytic Leukemia (CLL) is the most common adult hematological malignancy worldwide, primarily affecting individuals over 65 years of age, who represent more than 40% of all cases. The incidence and prevalence of CLL increase with age, with an average annual prevalence of nearly 5 cases per 100,000 people [1]. CLL arises from the uncontrolled proliferation of abnormal B-lymphocytes in the bone marrow, impairing the production of healthy blood cells. This pathological accumulation leads to clinical manifestations such as anemia, heightened infection susceptibility, and bleeding disorders caused by thrombocytopenia [2]. Malignant lymphocytes infiltrate the bloodstream, lymphatic organs, and bone marrow, displacing normal hematopoietic cells and driving disease progression. Although the precise etiology of CLL remains incompletely understood, somatic genetic mutations in hematopoietic progenitors are recognized contributors to its pathogenesis [3].

Despite progress in characterizing the genomic landscape of CLL, the molecular mechanisms that link genetic alterations to leukemia phenotypes remain only partially elucidated, particularly due to a lack of comprehensive protein expression data. Proteomic studies have identified distinctive signatures in CLL patients, including overexpression of B-cell receptor (BCR) signaling proteins, dysregulated spliceosome components, repression of protein kinase C pathways, and specific proteomic profiles associated with Immunoglobulin Heavy Chain Gene Variable Region (IGHV) mutation status [4,5]. However, the integration of proteomic data with clinical and genetic markers to enhance prognostic accuracy remains a major challenge.

CLL exhibits a highly heterogeneous clinical course; many patients can remain asymptomatic for years, with 30%–50% not requiring immediate treatment [6]. Conversely, a significant subset of patients experiences aggressive disease progression leading to poor outcomes. Clinical, hematological, biochemical, and genetic features profoundly influence disease prognosis and treatment response. Among these, IGHV mutation status and Tumor Protein p53 (TP53) aberrations are the most powerful prognostic markers. IGHV hypermutation defines biologically and clinically distinct CLL subsets, where mutated IGHV (> 2% deviation from germline) correlates with a more indolent disease and better response to therapy [7,8]. TP53 mutation had a more negative impact on prognosis than any other mutations (i.e., Neurogenic locus notch homolog protein 1 (Notch 1), Splicing Factor 3B Subunit 1 (SF3B1), Myeloid Differentiation primary response 88 (MYD88), Baculoviral IAP Repeat Containing 3 (BIRC3)), increasing the risk of death by 4-fold. Patients with TP53 aberration have inferior progression-free survival (PFS) and overall survival (OS) after treatment compared to those with wild-type TP53 [9–11]. Additionally, common chromosomal abnormalities such as del(17p) further stratify high-risk patients. Prognostic scoring systems like the CLL International Prognostic Index (CLL-IPI) combine these markers with clinical parameters to guide patient management [12–14]. Specifically, approximately 80% of all patients with CLL carry at least one of four common chromosomal alterations [15]. In particular deletions of the short arm of chromosome 17 (del(17p)) are found in 5%–8% of chemotherapy-naïve patients; patients with CLL carrying a del(17p) clone show marked resistance against genotoxic chemotherapies [16]. Using the different indicators various prognostic scores have been formulated. For instance, CLL International Prognostic Index (CLL-IPI) is a widely used, weighted scoring system integrating clinical (age, clinical stage), molecular (IGHV, TP53 aberration status), and serologic markers (β 2-microglobulin) for risk stratification of previously untreated CLL patients [International CLL-IPI working group. 2016 [17]].

Despite advances in the study of chronic lymphocytic leukemia (CLL), current diagnostic and prognostic methods still face significant limitations, relying on complex, time-consuming, and costly cytogenetic and molecular analyses requiring specialized expertise. This

limitation presents a pressing need for rapid, non-invasive, and cost-effective prognostic tools capable of monitoring disease progression and treatment response in CLL and other hematological malignancies. Traditional approaches, such as blood and bone marrow examinations, are invasive, painful, and may lack sensitivity for early or subtle disease changes. Most Raman studies focus primarily on discrimination between healthy and diseased states or among broad leukemia types. However, the nuanced prediction of disease course and prognosis within CLL subgroups using serum Raman profiling remains underexplored. Additionally, interpreting complex spectral data presents challenges due to the overlapping of biochemical signals and the multifactorial nature of molecular alterations in biological samples. Notably, blood serum is an accessible and minimally invasive sample type, making serum Raman spectroscopy a valuable approach for clinical applications. Prior studies have successfully differentiated healthy individuals from patients with hematopoietic diseases using serum Raman profiles, yet the full potential for prognostic stratification in CLL has not been fully explored [18–26].

The purpose of this work is to address these gaps by presenting an innovative model that combines Raman spectroscopy of blood serum with multivariate analysis, specifically Principal Component Analysis (PCA) and Partial least squares-Discriminant Analysis (PLS-DA), to predict disease progression and prognosis in CLL patients. Our approach captures the entire biochemical signature in the 800–1800 cm^{-1} spectral range, which reflects complex molecular changes associated with disease states. To our knowledge, this is the first study to demonstrate that serum Raman spectral profiling can partially discriminate between different prognostic groups in CLL, overcoming inherent spectral complexity challenges. This study not only extends the potential application of Raman spectroscopy beyond mere disease detection but also lays the groundwork for developing rapid, non-invasive, cost-effective tools to monitor CLL progression and support personalized clinical management. Hence, this work aims to fill the existing void in the literature by leveraging the combined strength of spectral data and advanced analytics to enhance disease stratification accuracy and prognostic prediction in CLL, addressing both clinical and technical shortcomings currently present in spectroscopic techniques.

2. Materials and methods

2.1. Patient population

Patients were selected during a routine visit to the Division of Hematology of the University Hospital of Messina “Policlinico G. Martino”, Italy, in the period between July 2024 and April 2025. The study population included 22 patients with B-CLL, 13 females and 9 males with a median age of 69.22 ± 9.5 years. In all patients, the diagnosis of B-CLL was established according to the National Cancer Institute (NCI) Sponsored Working Group diagnostic criteria for CLL. None of the patients had received specific therapy before the collection of blood samples. According to Rai classification [11], 2 patients were in stage 0, whereas 12 were in stage I, 8 were in stage II. The choice to include only untreated patients in the study explains the absence of patients in advanced stages of disease. An assessment of history and a physical examination was performed as the initial diagnostic procedure. Table S1 reports the following data of each patient: age, sex, data of first diagnosis, IGVH and TP53 gene analysis, deletion of 17p, lymphocytosis, blood platelet count, serum levels of hemoglobin, total proteins, lipids, immunoglobulins (IgA, IgM, IgG), Lactate Dehydrogenase (LDH), β 2-microglobulin. According to results of IGVH and TP53 analysis, the patients could be divided into a good prognosis group (favorable group; mutated IGVH and/or non-mutated TP53) made up of 15 patients, and a bad prognosis group (unfavorable group, non-mutated IGVH and/or non-mutated TP53) made up of 7 patients. Seven sex-matched (5 females and 2 males) and age-matched (65.1 ± 12.7 years) normal subjects were also included in this study

as controls. All patients participating into the study resulted negative for TP53 mutations and del(17p). The study was conducted according to good clinical and laboratory practice rules and the principles of the Declaration of Helsinki. The study was approved by the Local Ethics Committee (L.E.C.) of Messina (L.E.C. prot. n. 37-24 - Resolution of the Extraordinary Commissioner n. 971 of 06/07/2024). Informed written consent was obtained from all the participants. However, to perform our study, the analyzed serum samples were the same as the ones used during the clinical evaluation. None of the patients or of control subjects had symptoms or laboratory signs of active infections, inflammatory diseases, hypertension, diabetes, or kidney failure.

2.2. Serum preparation

For each subject, a whole-blood sample was collected into vacuum tubes and kept at room temperature for 30 min, followed by centrifugation at 2000 rpm for 10 min to ensure serum separation. Serum samples were stored at -80°C in aliquots of 1.5 mL before use.

2.3. Raman measurements and data processing

Patient serum was evenly deposited on a CaF_2 substrate to create two serum drops, 10 μL per drop. The serum drops were naturally dried for 20 min, then Raman measurements were conducted. A micro-Raman spectrometer (XploRA model of Horiba) covering $400\text{--}2000\text{ cm}^{-1}$ range with a spectral resolution of 2 cm^{-1} was used to collect Raman spectra. Raman spectrometer is equipped with a continuous wave (785 nm) diode laser with a power lower than 10 mW and a CCD sensor. In our Horiba XploRA micro-Raman system, confocal optimization was primarily achieved by selecting the appropriate confocal pinhole size. The XploRA system offers three physical confocal pinholes of varying diameters, enabling adjustment of the confocal volume to optimize both spatial and depth resolution for our measurements. For the experiments, a 100 μm pinhole was chosen, providing optimal filtering of the Raman signal from the focal point and enhancing confocality. This optimization was further supported by the use of an Olympus 50 $\times$ objective lens (NA of approximately 0.75), which helped maximize spatial resolution. By using a 785 nm diode laser, the system achieved a confocal resolution of approximately 1 μm laterally and 2 μm axially. Additionally, precise focusing and stage control (facilitated by fine adjustments of the sample position in XYZ via a joystick and motorized stage) contributed for an optimal confocal alignment. Finally, the spot size on the sample is about 1.28 μm . Each Raman spectrum was measured after 2-time averaging with an integration time of 30 s, so a good signal-to-noise (S/N) ratio was granted. Preliminarily, detection points were randomly selected to best capture spectral variability, both inside the “coffee-ring” area and along its outer perimeter. Subsequently, Raman mapping was performed over a 10 μm by 6 μm area with a step size of 2 μm , enabling coverage of a significant portion of the sample. From this, five Raman spectra were chosen as the most representative points exhibiting the greatest variability. After collecting the spectra, first we analyze the variation of the Raman intensities and shifts in the investigated range, due to presence of specific vibrational modes; thus, interpretation of the results was significantly improved by the use of the Principal Component Analysis (PCA). PCA reduces the complexity of high-dimensional metabolomic data by transforming it into principal components that capture the most variance in the data. To improve PCA analysis, spectra were preprocessed as reported in Ref. [21] and here briefly described: a continuous baseline correction was performed on each Raman spectrum using the adaptive iteratively re-weighted penalized least square algorithm (airPLS) and spectra are normalized to the total area. Then, the first principal component (PC1) is determined as the one that shows the largest possible variance with respect to the original data. The second PC is determined as the one having the largest variance compared to the all the other remaining PCs orthogonal to PC1. Then, PLS-DA was performed using

the first four principal component analysis (PCA) scores as predictors. In addition, beyond qualitative classification, Partial Least Squares Regression (PLSR) [27] was used to investigate quantitative associations between Raman spectral features and biochemical components. PCA scores were used as predictors, while serum cholesterol, triglyceride and total protein concentrations were modeled as continuous response variables. More details are reported in the SI.

3. Results and discussion

Figs. 1a–f show Raman spectra of the analyzed dried samples, both prior to (left panels) and following background subtraction (right panels). The spectra show the usual Raman profile of blood serum [24,26,28,29], and the band assignments are summarized in Table 1 according to the literature [30]. However one must note that these assignments are necessarily an oversimplification due to the extensive band overlapping of different types of serum constituents. Despite the rather similarity among Raman spectra for the 3 examined groups (consistent with the generally comparable composition of the compounds present in all the samples), there are differences both in background amount and in the intensity of some prominent bands. It is not excluded that by using the 785 nm wavelength, the background observed in the Raman spectrum is partially due to stray elastic light scattering of the laser rather than arising from specific chemical information [31]. However, different biological sample classes containing varying amounts of high molecular weight constituents, tend to aggregate and so enhance scattering upon drying. Therefore, although the background itself may not serve as a specific chemical marker, we consider the change in background signal as a meaningful “feature” of the spectral data. The background could indirectly reflect the composite biochemical milieu, encompassing contributions from nucleic acids, proteins, and lipids, and thus significantly impacts the spectral feature analysis related to these biomolecular constituents. So, accurate consideration and subtraction of this background from Raman spectra are essential to improve spectral classification performance and facilitate the interpretation of disease-associated biochemical alterations. Fig. 1g illustrates the quantified background intensity removed from each of the three groups. The subtracted background is significantly lower in patients with unfavorable prognosis compared to those with favorable prognosis, and even more so relative to the control group. This spectral difference reflects variations in the biomolecular serum profile between CLL patients and healthy controls, very likely correlating with disease severity and prognosis.

On the other hand, the peaks assigned to the main biomolecules present in the serum of the control subjects show for the most part a greater intensity than that detected in the serum of the patients with an unfavorable prognosis. This is also in agreement with data reported by other authors who employed Raman spectroscopy to examine the serum from patients with hemopoiesis diseases [24,26,28,29]. However, due to the complexity of the biological matrix examined (with up to several thousand of metabolites [40] and the above mentioned band overlapping, we avoided speculation about the assignment of changes in band intensities to variations in the relative amounts of specific serum biomolecules across the different groups of individuals. To integrate multiple potential markers into a single spectroscopic signature, a multivariate approach was employed by preliminarily performing PCA on the Raman spectra. Usually, most of the variance is contained in the first few principal components and, the associated loadings represent, spectral features interpreted as the magnitude of inherent variations allowing spectral discrimination, while scores can be interpreted as the magnitude of the changes. As shown in Fig. 2a, PCA has allowed the visualization of patterns and clustering of samples. Fig. 2b shows the loadings of the PC1 for all the three sets of samples. Particularly, it reveals distinct spectral contributions underlying the separation between unfavorable and control groups highlighted in Fig. 2d. Specifically, the most prominent positive loadings, associated with the unfavorable

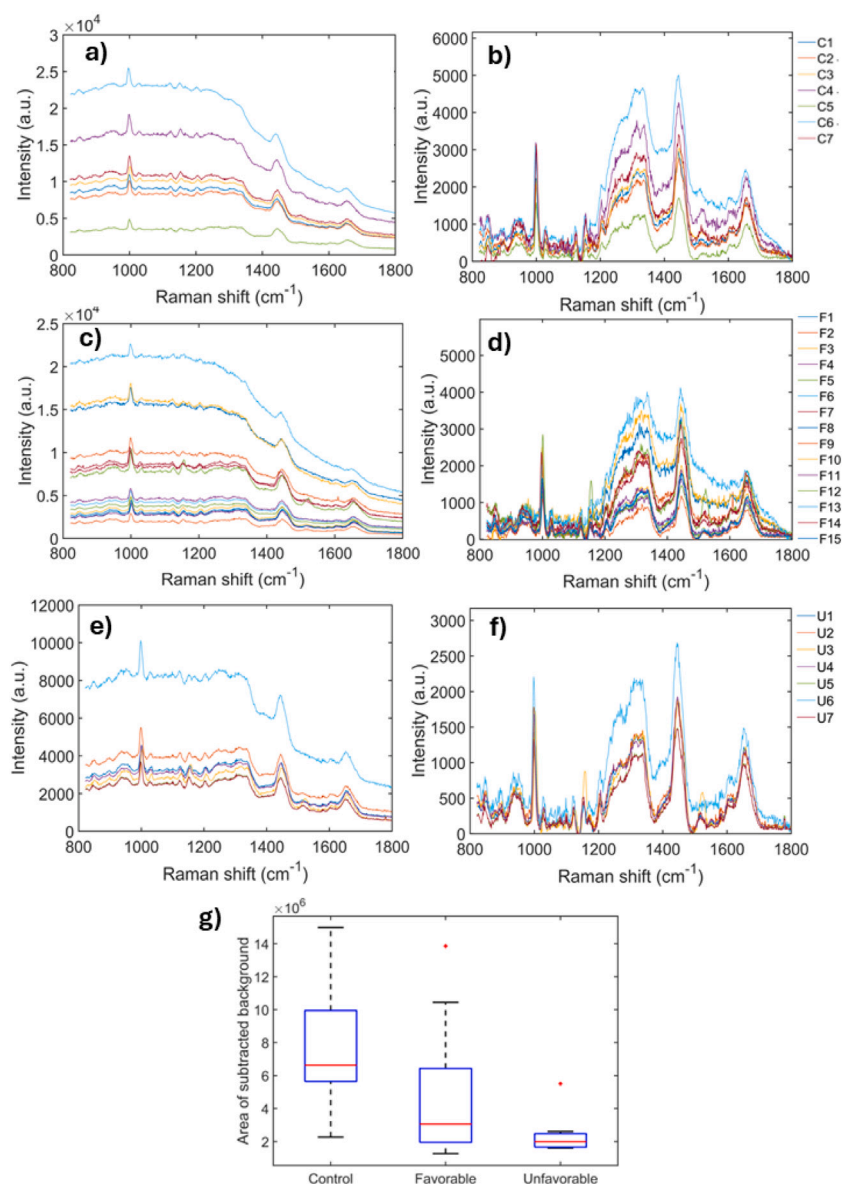


Fig. 1. Raman spectra of investigated samples as acquired (a,c,e) and only after baseline correction (b,d,f); the box plot of the area of the subtracted background (g).

Table 1
Raman bands assignment.

Raman shift (cm ⁻¹)	Mode	Assignment/biomolecule	Ref.
850	$\nu(\text{C}-\text{C})$	Proline, Proteins (collagen)	[21,26,32]
900	δ aromatic ring	DNA- Tryptophan, Tyrosine	[26]
943	$\nu(\text{C}-\text{C})$	Lipids	[26]
950	$\nu(\text{C}-\text{C})$	Proline-Valine	[33]
1003	CH_3 deformation or ring breathing	Phenylalanine	[26,34]
1032	C-H in-plane-bending	Phenylalanine-Protein	[32]
1153	$\delta(\text{C}-\text{C}-\text{H})$	β -carotene	[24,26]
1205	Ring vibration	Tryptophan	[35]
1240–1275	$\delta(\text{N}-\text{H})$, $\nu(\text{C}-\text{N})$	Amide III-Protein	[26,36,37]
1304–1324	(CH_2) twisting	Proteins, Lipids	[26,32]
1344	$\omega(\text{CH}_3/\text{CH}_2)$	Collagen-Lipids	[24,26]
1444	$\delta(\text{CH}_2)$	Amide I-Protein Lipids	[36]
1515	$\nu(\text{C}-\text{C})$	β -Carotene	[38,39]
1580	$\delta(\text{C}-\text{C})$	Phenylalanine	[30]
1600	$\nu(\text{C}=\text{C})$	Tryptophan, Phenylalanine	[32]
1652	$\nu(\text{C}=\text{O})$	Amide I-Protein Collagen	[24,26]

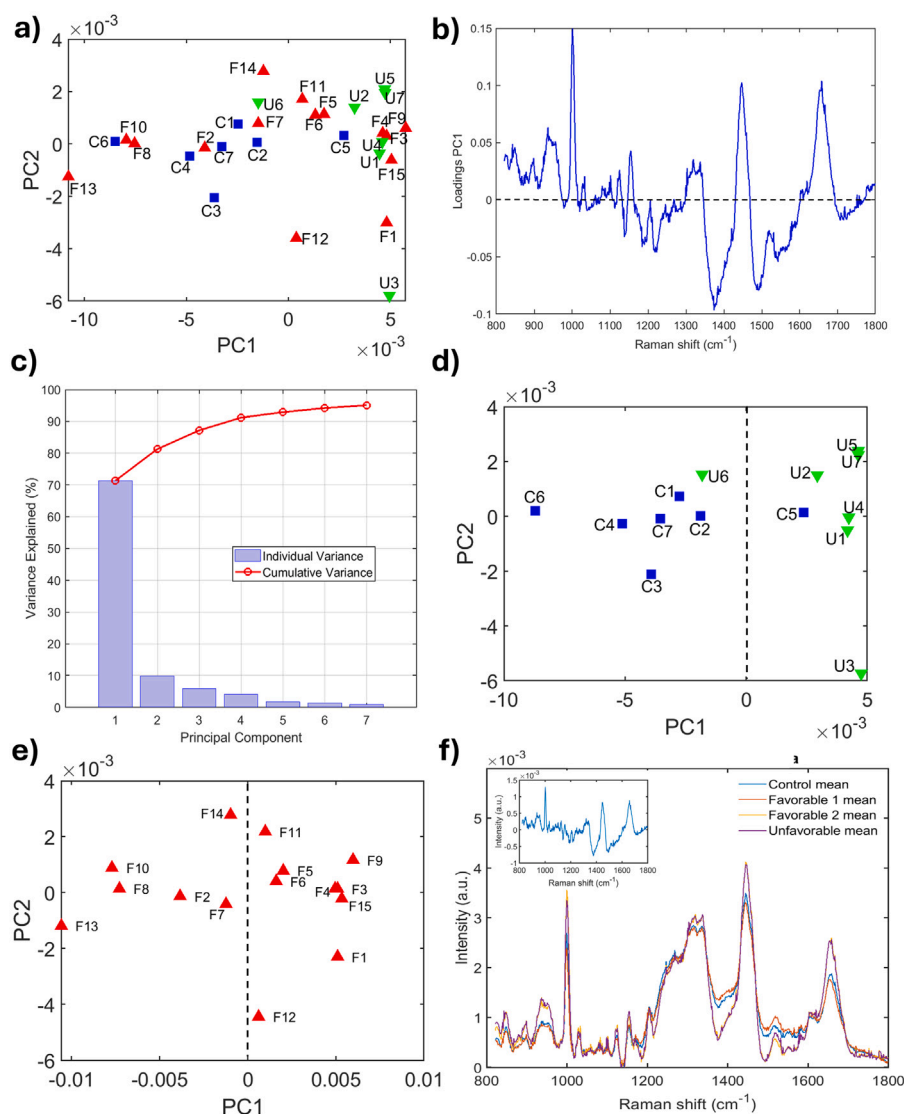


Fig. 2. (a) Results of the PCA analysis in terms of score plots of PC1 vs. PC2. (b) Loadings of the PC1 for all the three sets of samples. (c) PCA scree plot explaining the individual and cumulative variance of the first seven principal components. Score plots PC1 vs. PC2 for (d) control and unfavorable groups and (e) only for favorable samples. (f) Comparison among the averaged Raman spectra of the considered classes, with the inset reporting the difference between the mean spectra of favorable 2 and favorable 1 groups.

prognosis group (non-mutated IGVH), are located at approximately 1003 cm^{-1} , 1444 cm^{-1} , and 1652 cm^{-1} . As reported in Table 1, these bands are respectively assigned to phenylalanine ring breathing, CH_2/CH_3 bending modes (related to both lipid structures and Amide I protein), and C=O stretching vibrations in the protein backbone (Amide I-Protein Collagen). Conversely, in the negative side, the main loadings, associated with the healthy subjects group, are related to the Raman bands centered at about 1350 cm^{-1} and 1510 cm^{-1} , whose vibrational modes are ascribed to CH_3/CH_2 bending (collagen/lipids) and C-C stretching mode of β -Carotene, respectively. Instead, favorable cases were more dispersed and partially overlapped with both other groups (Fig. 2a). In addition, PCA clearly allows the discrimination between healthy subjects and patients with bad prognosis (non-mutated IGVH) (Fig. 2a and Fig. 2d), and the first two principal components (PC1 and PC2) accounted for approximately 81% of the total variance (PC1: 71%, PC2: 10% (Fig. 2c). Specifically, in the PC1-PC2 score space, the patients with the bad prognosis (non-mutated IGVH) tend to cluster on the positive side of PC1, whereas healthy subjects predominantly occupied the negative side (Fig. 2d). More interestingly, PCA reveals that, within the favorable prognosis group (mutated IGVH), two distinct

subgroups cluster separately (Fig. 2e), referred to hereinafter as favorable group 1 and favorable group 2. This finding is easily confirmed by plotting the average Raman spectra (Fig. 2f). In fact, the averaged spectra of the favorable 1 group strictly resemble those of the control group, while the averaged spectra of the favorable 2 group and of the unfavorable group are very similar. This suggests heterogeneity even among patients initially all classified by conventional medical approach, as having a favorable disease outcome. The most relevant differences can be observed in specific spectral regions as shown in the inset of Fig. 2f. The strict similarity with the loadings of PC1 (2b) confirms the spectral bands responsible of the sample discrimination and on which the biochemical investigation should be focused.

To further evaluate the discriminative capability of the Raman spectral data, Partial Least Squares Discriminant Analysis (PLS-DA) was performed using the first four principal component analysis (PCA) scores as predictors, which together accounted for 91.1% of the total variance (Fig. 2c). The initial three-class PLS-DA model, corresponding to the healthy subjects group, all patients with favorable prognosis (mutated IGVH), and patients with bad prognosis (non-mutated IGVH), exhibited limited classification performance, with an overall accuracy

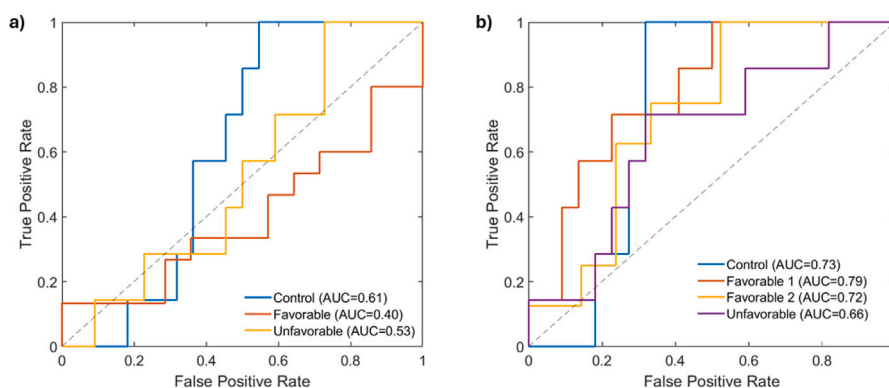


Fig. 3. ROC curves and corresponding AUC values from the analyses of the (a) three and (b) four-class model, respectively.

of 37.9% (see SI for details on the reliability evaluation of the diagnostic test). Class-specific metrics showed low sensitivity for the healthy subjects group (28.6%) and the bad prognosis group (14.3%), while the favorable prognosis group achieved moderate sensitivity (53.3%) but very low specificity (28.6%). Specificity was higher for both the healthy subjects and bad prognosis groups (81.8% each), indicating that most misclassifications involved the favorable prognosis group. Receiver operating characteristic (ROC) curve analysis confirmed this limited discrimination, with areas under the curve (AUC) of 0.61, 0.40, and 0.53 for the healthy subjects group, favorable prognosis group, and bad prognosis group, respectively (Fig. 3a). Summarizing, these preliminary PLS-DA data suggest that the favorable prognosis group is internally heterogeneous because it shows a moderate sensitivity but very low specificity, meaning that the model struggles to correctly and consistently distinguish this group from the others. Additionally, the bad performance in class-specific metrics and the relatively low AUC value for this group indicate variability within the group, leading to overlapping characteristics with both healthy subjects and bad prognosis groups.

Hence, considering the heterogeneity observed within the favorable group, the PLS-DA model was refined by subdividing this class into two subclasses, resulting in a four-class model comprising the healthy subjects group, two favorable subgroups, and the bad prognosis group. Cross-validation of the refined four-class PLS-DA model demonstrated a modest improvement in overall classification performance compared to the three-class model.

Accuracy increased from 37.9% to 41.4%. For the control group, sensitivity rose from 28.6% to 42.9%, while specificity improved from 81.8% to 90.9%. The favorable group 1 showed a sensitivity of 42.9% and specificity of 86.4%, whereas favorable group 2 achieved a sensitivity of 50.0% and a specificity of 66.7% (noting that in the three-class model, the combined favorable class had a sensitivity of 53.3% and specificity of 28.6%). Sensitivity for the bad prognosis group increased from 14.3% to 28.6%, with a slight decrease in specificity from 81.8% to 77.3% (see Supplementary Information for detailed confusion matrices). Receiver operating characteristic (ROC) curve analysis yielded higher AUC values of 0.73, 0.79, 0.72, and 0.66 for the healthy subjects group, favorable prognosis groups 1 and 2, and the bad prognosis group, respectively (Fig. 3b). Furthermore preliminary PLSR trend confirmed quantitative associations between Raman spectral features and sample biochemical components (see Supplementary Information).

Following the evidence remarked by the PCA-PLS analyses, we considered useful to further investigate the specific spectral regions identified by the statistical analysis. Thus, we have estimated the area under the most significant Raman bands for selected spectral regions as well as the ratio of the most intense Raman peaks. Fig. 4 reports the box plot of the area for 5 specific spectral range (912–975 cm^{-1} , 975–1020 cm^{-1} , 1186–1380 cm^{-1} , 1380–1490 cm^{-1} , 1563–1719 cm^{-1} , panels a to e respectively) and the intensity ratio between two average

peaks identified among the most representative ones of the Raman spectrum (1003 cm^{-1} , 1205 cm^{-1} , 1265 cm^{-1} , 1344 cm^{-1} , and 1652 cm^{-1}), calculated for each of the four examined groups (panel f). The spectral areas have been re-normalized for the corresponding spectral width in cm^{-1} allowing a better comparison on the same y-scale. There is a strict similarity between the control and the favorable 1 group, as well as between the unfavorable control and the favorable 2 group, whereas there is an evident difference between the control group and the unfavorable group. These trends are consistent with the spectral similarity (Fig. 2f) observed between favorable 1 and control groups as well as between favorable 2 and unfavorable groups. Also this finding leads us to suppose that in the group of CLL patients with favorable prognosis (as defined on the basis of current diagnostic criteria for CLL), two different subgroups coexist, whose disease could have a different course and progression.

To confirm that the model developed in our study can be used to obtain a spectroscopic signature of the metabolomic serum features of each patient, allowing the definition of several clusters of patients with different disease course, we calculated the ratio between the intensity of specific peaks (Fig. 4f). In particular, we have taken into consideration the peaks at 1003 cm^{-1} (phenylalanine), 1205 cm^{-1} (phenylalanine, tryptophan), 1265 cm^{-1} (Amide III), 1344 cm^{-1} (collagen), and 1652 cm^{-1} (Amide I). The Amide I band (associated with C=O stretching vibrations) and amide III band (combination of C-N stretching and N-H bending vibrations) reflect coupled vibrations of the protein backbone and are commonly used for characterizing changes of protein secondary structure, while the 1003 cm^{-1} and 1205 cm^{-1} bands reflect vibrations of aromatic aminoacids. The intensity ratio (Fig. 4f) of selected Raman peaks is interesting to analyze since they appeared different between control and unfavorable group but similar between control-favorable1 and between favorable 2-unfavorable groups. These changes indicate significant differences among groups in the serum protein profile. For example, the ratio between the peaks centered at 1652 and 1205 cm^{-1} differs of about 1.5 in the two cases, evidencing a marked variation in protein secondary structure and amino acid composition between patient groups. Taken together, these data suggest the existence of alterations in protein variety and conformation in serum from CLL patients, in comparison with healthy subjects, as a consequence of disease-related metabolic changes and the course of the disease. An interesting observation concerns the ratio between the peaks centered at 1003 and 1344 cm^{-1} (assigned to phenylalanine and collagen, respectively), which is lower/higher than 1 for control-favorable1 and favorable 2-unfavorable groups, respectively.

Collagen is one of most abundant proteins in the extracellular matrix of the bone marrow microenvironment, whose influence on normal hematopoiesis and leukemia has largely been accepted [41]. The increase observed in our study in the ratio 1003/1344 cm^{-1} from the value calculated for control/favorable 1 groups to that calculated for favorable2-unfavorable groups might be due to a decrease

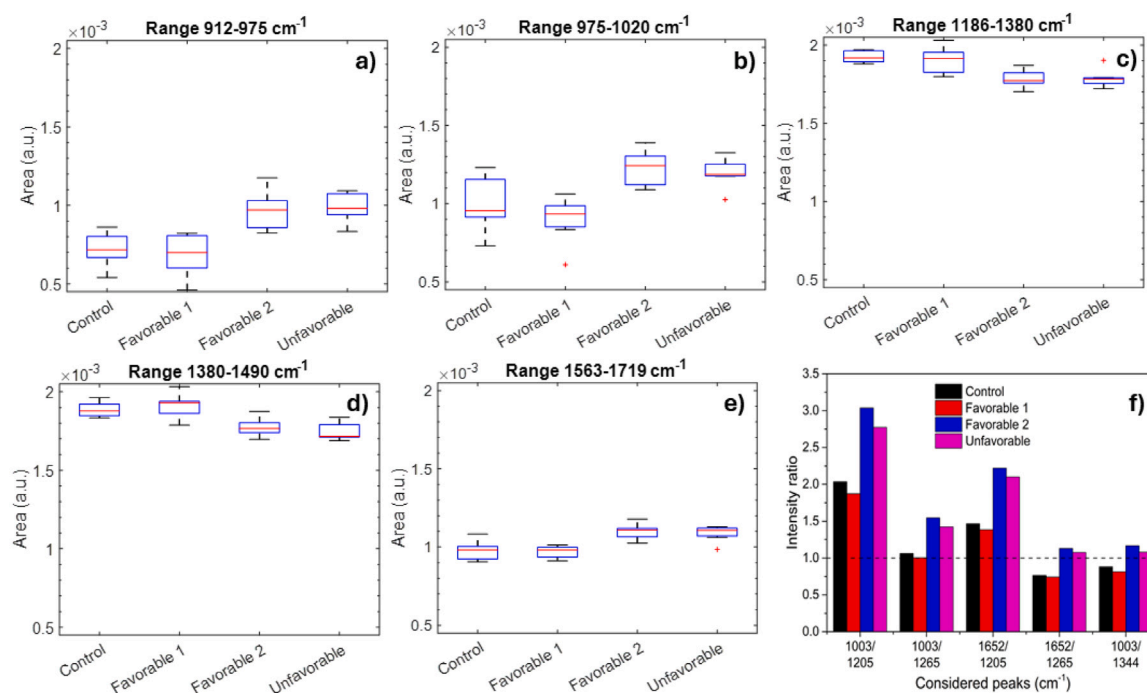


Fig. 4. Box plot of spectral area in the reported ranges (a–e) and intensity ratio between two chosen Raman peaks grouped with respect to the four classified groups (f). The spectral areas reported in panels (a–e) have been re-normalized for the corresponding spectral width in cm^{-1} allowing a better comparison on the same y-scale.

of serum collagen, correlating with disease progression and poorer prognosis. Similarly, Bai et al. and Liang et al. [23,24] reported that in serum Raman spectra the intensity of peaks assigned to collagen are downregulated in CLL patients when compared with healthy controls. A reduction in serum collagen could indicate a lower bone marrow remodeling activity in high-risk patients, probably due to an increased bone marrow fibrosis. This finding could be interesting given that literature data showed that the grade of bone marrow reticulin fibrosis and short survival were significantly correlated [42]. Although no definitive clinical conclusions can be drawn from these data—partly because serum collagen levels alone do not accurately reflect the amount of collagen in the bone marrow, as they are influenced by other factors such as kidney function, diet, and circadian rhythm—this finding remains intriguing. This is especially relevant given that bone marrow biopsies are no longer routinely required for all patients, according to the updated International Workshop on CLL/National Institutes of Health guidelines [43,44]. Therefore, Raman examination could potentially serve as a valuable new prognostic tool.

We emphasize that the purpose of this work is not to provide a unique and specific clinical explanation for the observed Raman changes (for example, attributing them to the specific protein collagen is just a potential and plausible speculation supported by the literature reports). Rather, it offers a potential interpretation on LCC disease prognosis and stratification, without aiming for a detailed quantification of individual protein species or other biomolecules based solely on Raman peaks.

Ultimately, the outlined results are certainly preliminary; however, they are highly promising and suggest significant potential for advancing prognostic stratification. Thus, in the future, to enhance the utility of Raman spectroscopy in clinical settings, several key areas warrant attention. Among them: (1) expanding sample size and diversity is crucial to improve the statistical power and generalizability of findings; (2) carrying out longitudinal studies that monitor biomolecular changes over time to validate the prognostic value of Raman signatures; (2) integration with established molecular and clinical biomarkers may enhance accuracy and robustness; (3) advanced data analysis techniques, such as machine learning and multivariate statistical models,

to optimize subgroup classification and predictive performance. Finally, conducting prospective clinical trials to evaluate the effect of Raman-based stratification on treatment outcomes and patient management will be crucial for achieving clinical validation.

4. Conclusion

The results, presented and discussed in this paper, indicate that the Raman spectroscopy technique represents a promising approach to analyze dried serum from CLL patients and providing insights into the biochemical alterations associated with the disease. The ability to identify specific spectral markers associated with disease states suggests that Raman spectroscopy could be integrated into routine clinical practices for monitoring disease progression and treatment response in CLL patients. Its application could lead to improved diagnostic accuracy and patient monitoring in clinical practice and then to better therapeutic management strategies.

CRediT authorship contribution statement

E. Fazio: Conceptualization, Supervision, Investigation, Data curation, Writing – review & editing. **C. Corsaro:** Conceptualization, Supervision, Investigation, Data curation, Writing – review & editing. **A. Speciale:** Investigation, Writing – review & editing. **A. Saija:** Conceptualization, Supervision, Writing – review & editing. **F.L. Salamone:** Data curation, Writing – review & editing. **M. Mondello:** Data curation, Writing – review & editing. **R. Sciacotta:** Data curation, Writing – review & editing. **L. Giordano:** Data curation, Writing – review & editing. **F. Stagno:** Data curation, Writing – review & editing. **A. Allegra:** Conceptualization, Supervision, Writing – review & editing.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments and funding

This work was supported by Associazione pro Bambini e Adulti Leucemici (A.B.A.L.) onlus Messina (Italy) (<http://www.abalmessina.it>).

Appendix A. Supplementary data

Supplementary material related to this article can be found online at <https://doi.org/10.1016/j.saa.2025.126961>.

Data availability

Data will be made available on request.

References

- [1] A. Redaelli, B. Laskin, J. Stephens, M. Botteman, C. Pashos, The clinical and epidemiological burden of chronic lymphocytic leukaemia, *Eur. J. Cancer Care* 13 (3) (2004) 279–287, <http://dx.doi.org/10.1111/j.1365-2354.2004.00489.x>.
- [2] S.K.R. Mukkamalla, A. Taneja, D. Malipeddi, S.R. Master, Chronic lymphocytic leukemia, StatPearls [Internet] 2023 Mar 7 (2025) Treasure Island (FL): StatPearls Publishing. <https://www.ncbi.nlm.nih.gov/books/NBK470433/>.
- [3] D.A. Landau, S.L. Carter, P. Stojanov, A. McKenna, K. Stevenson, M.S. Lawrence, C. Sougnez, C. Stewart, A. Sivachenko, L. Wang, Y. Wan, W. Zhang, S.A. Shukla, A. Vartanov, S.M. Fernandes, S. Saksena, K. Cibulski, B. Tesar, S. Gabriel, N. Hacohen, M. Meyerson, E.S. Lander, D. Neuberg, J.R. Brown, G. Getz, C.J. Wu, Evolution and impact of subclonal mutations in chronic lymphocytic leukemia, *Cell* 152 (2013) 714–726, <http://dx.doi.org/10.1016/j.cell.2013.01.019>.
- [4] F. Meier-Abt, J. Lu, E. Cannizzaro, M.F. Pohly, S. Kummer, S. Pfammatter, L. Kunz, B.C. Collins, F. Nadeu, K.S. Lee, P. Xue, M. Gwerder, M. Roiss, J. Hüllelin, S. Scheinost, S. Dietrich, E. Campo, W. Huber, R. Aebersold, T. Zenz, The protein landscape of chronic lymphocytic leukemia, *Blood* 138 (24) (2021) 2514–2525, <http://dx.doi.org/10.1182/blood.2020009741>.
- [5] J. Lu, M. Böttcher, T. Walther, D. Mougiakakos, T. Zenz, W. Huber, Energy metabolism is co-determined by genetic variants in chronic lymphocytic leukemia and influences drug sensitivity, *Haematologica* 104 (9) (2019) 1830–1840, <http://dx.doi.org/10.3324/haematol.2018.203067>.
- [6] M.A. Hoechstetter, C.-M. Wendtner, Clinical trials in early-stage CLL: what has been learned and what's next? *Leuk. Lymphoma* 66 (3) (2024) 378–388, <http://dx.doi.org/10.1080/10428194.2024.2422839>.
- [7] R.N. Damle, T. Wasi, F. Fais, F. Ghiotto, A. Valetto, S.L. Allen, A. Buchbinder, D. Budman, K. Dittmar, J. Kollitz, S.M. Lichtman, P. Schulman, V.P. Vinciguerra, K.R. Rai, M. Ferrarini, N. Chiorazzi, Ig V gene mutation status and CD38 expression as novel prognostic indicators in chronic lymphocytic leukemia, *Blood* 94 (1999) 1840–1847.
- [8] T.J. Hamblin, Z. Davis, A. Gardiner, D.G. Oscier, F.K. Stevenson, Unmutated Ig V(H) genes are associated with a more aggressive form of chronic lymphocytic leukemia, *Blood* 94 (1999) 1848–1854.
- [9] M. Mihara, S. Erster, A. Zaika, O. Petrenko, T. Chittenden, P. Pancoska, U.M. Moll, P53 has a direct apoptogenic role at the mitochondria, *Mol. Cell* 11 (2003) 577–590, [http://dx.doi.org/10.1016/j.s1097-2765\(03\)00050-9](http://dx.doi.org/10.1016/j.s1097-2765(03)00050-9).
- [10] S. Stilgenbauer, T. Zenz, D. Winkler, A. Bühler, R.F. Schlenk, S. Groner, R. Busch, M. Hensel, U. Dührsen, J. Finke, P. Dreger, U. Jäger, E. Lengfelder, K. Hohloch, U. Söling, R. Schlag, M. Kneba, M. Hallek, H. Döhner, Subcutaneous alemtuzumab in fludarabine-refractory chronic lymphocytic leukemia: Clinical results and prognostic marker analyses from the CLL2H study of the German chronic lymphocytic leukemia study group, *J. Clin. Oncol.* 27 (24) (2009) 3994–4001, <http://dx.doi.org/10.1200/jco.2008.21.1128>.
- [11] M. Hallek, B.D. Cheson, D. Catovsky, F. Caligaris-Cappio, G. Dighiero, H. Döhner, P. Hillmen, M.J. Keating, E. Montserrat, K.R. Rai, T.J. Kipps, Guidelines for the diagnosis and treatment of chronic lymphocytic leukemia: a report from the International Workshop on Chronic Lymphocytic Leukemia updating the National Cancer Institute–Working Group 1996 guidelines, *Blood* 111 (12) (2008) 5446–5456, <http://dx.doi.org/10.1182/blood-2007-06-093906>.
- [12] S. Stilgenbauer, A. Schnaiter, P. Paschka, T. Zenz, M. Rossi, K. Döhner, A. Bühler, S. Böttcher, M. Ritgen, M. Kneba, D. Winkler, E. Tausch, P. Hoth, J. Edelmann, D. Mertens, L. Bullinger, M. Bergmann, S. Kless, S. Mack, U. Jäger, N. Patten, L. Wu, M.K. Wenger, G. Fingerle-Rowson, P. Lichter, M. Cazzola, C.M. Wendtner, A.M. Fink, K. Fischer, R. Busch, M. Hallek, H. Döhner, Gene mutations and treatment outcome in chronic lymphocytic leukemia: results from the CLL8 trial, *Blood* 123 (21) (2014) 3247–3254, <http://dx.doi.org/10.1182/blood-2014-01-546150>.
- [13] D. Rossi, S. Rasi, V. Spina, A. Bruscaggin, S. Monti, C. Ciardullo, C. Deambrogi, H. Khiabani, R. Serra, F. Bertoni, F. Forconi, L. Laurenti, R. Marasca, M. Dal-Bo, F.M. Rossi, P. Bulian, J. Nomdedeu, G. Del Poeta, V. Gattei, L. Pasqualucci, R. Rabadan, R. Foà, R. Dalla-Favera, G. Gaidano, Integrated mutational and cytogenetic analysis identifies new prognostic subgroups in chronic lymphocytic leukemia, *Blood* 121 (8) (2013) 1403–1412, <http://dx.doi.org/10.1182/blood-2012-09-458265>.
- [14] I. Negara, C. Tomuleasa, S. Buruiana, D.G. Efremov, Molecular subtypes and the role of TP53 in diffuse large B-cell lymphoma and Richter syndrome, *Cancers* 16 (12) (2024) 2170, <http://dx.doi.org/10.3390/cancers16122170>.
- [15] M. Hallek, Chronic lymphocytic leukemia: 2025 update on the epidemiology, pathogenesis, diagnosis, and therapy, *Am. J. Hematol.* 100 (3) (2025) 450–480, <http://dx.doi.org/10.1002/ajh.27546>.
- [16] M.d.L.L. Chauffaille, I. Zalberg, W.G. Barreto, I. Bendit, Detection of somatic TP53 mutations and 17p deletions in patients with chronic lymphocytic leukemia: a review of the current methods, *Hematol. Transfus. Cell Ther.* 42 (3) (2020) 261–268, <http://dx.doi.org/10.1016/j.htct.2020.05.005>.
- [17] International CLL-IPI working group, An international prognostic index for patients with chronic lymphocytic leukaemia (CLL-IPI): a meta-analysis of individual patient data, *Lancet Oncol.* 17 (6) (2016) 779–790, [http://dx.doi.org/10.1016/s1470-2045\(16\)30029-8](http://dx.doi.org/10.1016/s1470-2045(16)30029-8).
- [18] H. Liang, R. Shi, H. Wang, Y. Zhou, Advances in the application of Raman spectroscopy in haematological tumours, *Front. Bioeng. Biotechnol.* 10 (2023) 1–16, <http://dx.doi.org/10.3389/fbioe.2022.1103785>.
- [19] M. Rizzo, C. Corsaro, S. Marrara, V. Crupi, S. Conoci, F. Neri, E. Fazio, Raman spectral analyses to investigate the physiological and metabolic development of a 3D hepatocellular carcinoma model, *Spectrochim. Acta Part A: Mol. Biomol. Spectrosc.* 343 (2025) 126564, <http://dx.doi.org/10.1016/j.saa.2025.126564>.
- [20] S. Yonezawa, T. Haruki, K. Koizumi, A. Taketani, Y. Oshima, M. Oku, A. Wada, T. Sato, N. Masuda, J. Tahara, N. Fujisawa, S. Koshiyama, M. Kadowaki, I. Kitajima, S. Saito, Establishing monoclonal gammopathy of undetermined significance as an independent pre-disease state of multiple myeloma using Raman spectroscopy, dynamical network biomarker theory, and energy landscape analysis, *Int. J. Mol. Sci.* 25 (3) (2024) 1570, <http://dx.doi.org/10.3390/ijms25031570>.
- [21] M. Bonsignore, S. Trusso, C. De Pasquale, G. Ferlazzo, A. Allegra, V. Innao, C. Musolino, D. Franco, L. Maria De Plano, S.P.P. Guglielmino, F. Neri, E. Fazio, A multivariate analysis of Multiple Myeloma subtype plasma cells, *Spectrochim. Acta Part A: Mol. Biomol. Spectrosc.* 258 (2021) 119813, <http://dx.doi.org/10.1016/j.saa.2021.119813>.
- [22] M. Russo, L. Tirinato, F. Scionti, M.L. Coluccio, G. Perozziello, C. Riillo, V. Mollace, S. Gratter, N. Malara, M.T. Di Martino, G. Viglietto, P. Tagliaferri, P. Tassone, M. Rossi, P. Candeloro, Raman spectroscopic stratification of multiple myeloma patients based on exosome profiling, *ACS Omega* 5 (47) (2020) 30436–30443, <http://dx.doi.org/10.1021/acsomega.0c03813>.
- [23] H. Liang, Z. Cao, Y. Ren, Y. Li, H. Wang, F. Sun, M. Xue, G. Zhu, Y. Zhou, Raman spectroscopy and bioinformatics-based identification of key genes and pathways capable of distinguishing between diffuse large B cell lymphoma and chronic lymphocytic leukemia, *Front. Immunol.* 16 (2025) 1–22, <http://dx.doi.org/10.3389/fimmu.2025.1516946>.
- [24] Y. Bai, Z. Yu, S. Yi, Y. Yan, Z. Huang, L. Qiu, Raman spectroscopy-based biomarker screening by studying the fingerprint characteristics of chronic lymphocytic leukemia and diffuse large B-cell lymphoma, *J. Pharm. Biomed. Anal.* 190 (2020) 113514, <http://dx.doi.org/10.1016/j.jpba.2020.113514>.
- [25] H. Liang, X. Kong, Y. Ren, H. Wang, E. Liu, F. Sun, G. Zhu, Q. Zhang, Y. Zhou, Application of serum Raman spectroscopy in rapid and early discrimination of aplastic anemia and myelodysplastic syndrome, *Spectrochim. Acta Part A: Mol. Biomol. Spectrosc.* 302 (2023) 123008, <http://dx.doi.org/10.1016/j.saa.2023.123008>.
- [26] A.M.F. Lima, C.R. Daniel, M.T.T. Pacheco, P.L. de Brito, L. Silveira, Discrimination of leukemias and non-leukemic cancers in blood serum samples of children and adolescents using a Raman spectral model, *Lasers Med. Sci.* 38 (2022) 22, <http://dx.doi.org/10.1007/s10103-022-03681-2>.
- [27] S. Wold, M. Sjöström, L. Eriksson, PLS-regression: a basic tool of chemometrics, *Chemometr. Intell. Lab. Syst.* 58 (2) (2001) 109–130, [http://dx.doi.org/10.1016/s0169-7439\(01\)00155-1](http://dx.doi.org/10.1016/s0169-7439(01)00155-1).
- [28] J.L. González-Solis, J.C. Martínez-Espinosa, J.M. Salgado-Román, P. Palomares-Anda, Monitoring of chemotherapy leukemia treatment using Raman spectroscopy and principal component analysis, *Lasers Med. Sci.* 29 (3) (2014) 1241–1249, <http://dx.doi.org/10.1007/s10103-013-1515-y>.
- [29] A.M. da Silva, F.S.A. de Siqueira e Oliveira, P.L. de Brito, L.S. Jr., Spectral model for diagnosis of acute leukemias in whole blood and plasma through Raman spectroscopy, *J. Biomed. Opt.* 23 (10) (2018) 1, <http://dx.doi.org/10.1117/1.jbo.23.10.107002>.
- [30] X. Chen, X. Li, J. Xie, H. Yang, A. Liu, Non-invasive discrimination of multiple myeloma using label-free serum surface-enhanced Raman scattering spectroscopy in combination with multivariate analysis, *Anal. Chim. Acta* 1191 (2022) 339296, <http://dx.doi.org/10.1016/j.aca.2021.339296>.
- [31] F. Bonnier, A. Mehmood, P. Knief, A.D. Meade, W. Hornebeck, H. Lambkin, K. Flynn, V. McDonagh, C. Healy, T.C. Lee, F.M. Lyng, H.J. Byrne, In vitro analysis of immersed human tissues by Raman microspectroscopy, *J. Raman Spectrosc.* 42 (5) (2010) 888–896, <http://dx.doi.org/10.1002/jrs.2825>.

- [32] C. Esposito, M. Janneh, S. Spaziani, V. Calcagno, M.L. Bernardi, M. Iammarino, C. Verdone, M. Tagliamonte, L. Buonaguro, M. Pisco, L. Aversano, A. Cusano, Assessment of primary human liver cancer cells by artificial intelligence-assisted Raman spectroscopy, *Cells* 12 (22) (2023) 2645, <http://dx.doi.org/10.3390/cells12222645>.
- [33] S. Feng, R. Chen, J. Lin, J. Pan, G. Chen, Y. Li, M. Cheng, Z. Huang, J. Chen, H. Zeng, Nasopharyngeal cancer detection based on blood plasma surface-enhanced Raman spectroscopy and multivariate analysis, *Biosens. Bioelectron.* 25 (11) (2010) 2414–2419, <http://dx.doi.org/10.1016/j.bios.2010.03.033>.
- [34] L. Huang, H. Sun, L. Sun, K. Shi, Y. Chen, X. Ren, Y. Ge, D. Jiang, X. Liu, W. Knoll, Q. Zhang, Y. Wang, Rapid, label-free histopathological diagnosis of liver cancer based on Raman spectroscopy and deep learning, *Nat. Commun.* 14 (2023) 48, <http://dx.doi.org/10.1038/s41467-022-35696-2>.
- [35] D. Lin, Y. Wang, T. Wang, Y. Zhu, X. Lin, Y. Lin, S. Feng, Metabolite profiling of human blood by surface-enhanced Raman spectroscopy for surgery assessment and tumor screening in breast cancer, *Anal. Bioanal. Chem.* 412 (7) (2020) 1611–1618, <http://dx.doi.org/10.1007/s00216-020-02391-4>.
- [36] P.N. Paramitha, R. Zakaria, A. Maryani, Y. Kusaka, B.B. Andriana, K. Hashimoto, H. Nakazawa, S. Kato, H. Sato, Raman study on lipid droplets in hepatic cells co-cultured with fatty acids, *Int. J. Mol. Sci.* 22 (14) (2021) 7378, <http://dx.doi.org/10.3390/ijms22147378>.
- [37] T. Tolstik, C. Marquardt, C. Matthäus, N. Bergner, C. Bielecki, C. Krafft, A. Stallmach, J. Popp, Discrimination and classification of liver cancer cells and proliferation states by Raman spectroscopic imaging, *Analyst* 139 (22) (2014) 6036–6043, <http://dx.doi.org/10.1039/c4an00211c>.
- [38] R. Ullah, S. Khan, F. Farman, M. Bilal, C. Krafft, S. Shahzad, Demonstrating the application of Raman spectroscopy together with chemometric technique for screening of asthma disease, *Biomed. Opt. Express* 10 (2) (2019) 600, <http://dx.doi.org/10.1364/boe.10.000600>.
- [39] V. Mieites, M.G. Fernández-Manteca, I. Santiuste Torcida, F. Madrazo Toca, M.J. Marín Vidalled, O.M. Conde, Assessment of blood serum stability with Raman spectroscopy and explanatory AI, *Spectrochim. Acta Part A: Mol. Biomol. Spectrosc.* 340 (2025) 126297, <http://dx.doi.org/10.1016/j.saa.2025.126297>.
- [40] N. Psychogios, D.D. Hau, J. Peng, A.C. Guo, R. Mandal, S. Bouatra, I. Sinelnikov, R. Krishnamurthy, R. Eisner, B. Gautam, N. Young, J. Xia, C. Knox, E. Dong, P. Huang, Z. Hollander, T.L. Pedersen, S.R. Smith, F. Bamforth, R. Greiner, B. McManus, J.W. Newman, T. Goodfriend, D.S. Wishart, The human serum metabolome, *PLoS ONE* 6 (2) (2011) e16957, <http://dx.doi.org/10.1371/journal.pone.0016957>.
- [41] C. Zanetti, D.S. Krause, “Caught in the net”: the extracellular matrix of the bone marrow in normal hematopoiesis and leukemia, *Exp. Hematol.* 89 (2020) 13–25, <http://dx.doi.org/10.1016/j.exphem.2020.07.010>.
- [42] T. Tadmor, L. Shvidel, A. Aviv, R. Ruchlemer, O. Bairey, M. Yuklea, Y. Herishanu, A. Braester, N. Levene, F. Vernea, J. Ben-Ezra, J. Bejar, A. Polliack, Significance of bone marrow reticulin fibrosis in chronic lymphocytic leukemia at diagnosis: A study of 176 patients with prognostic implications, *Cancer* 119 (10) (2013) 1853–1859, <http://dx.doi.org/10.1002/cncr.27930>.
- [43] B.D. Cheson, J.M. Bennett, M. Grever, N. Kay, M.J. Keating, S. O'Brien, K.R. Rai, National Cancer Institute-sponsored Working Group guidelines for chronic lymphocytic leukemia: revised guidelines for diagnosis and treatment, *Blood* 87 (1996) 4990–4997.
- [44] M. Hallek, B.D. Cheson, D. Catovsky, F. Caligaris-Cappio, G. Dighiero, H. Döhner, P. Hillmen, M. Keating, E. Montserrat, N. Chiorazzi, S. Stilgenbauer, K.R. Rai, J.C. Byrd, B. Eichhorst, S. O'Brien, T. Robak, J.F. Seymour, T.J. Kipps, iwCLL guidelines for diagnosis, indications for treatment, response assessment, and supportive management of CLL, *Blood* 131 (25) (2018) 2745–2760, <http://dx.doi.org/10.1182/blood-2017-09-806398>.