UC San Diego

UC San Diego Previously Published Works

Title

Cancer Cell Line Classification Using Raman Spectroscopy of Cancer-Derived Exosomes and Machine Learning.

Permalink https://escholarship.org/uc/item/46g9v6jj

Journal Analytical Chemistry, 97(13)

Authors

Villazon, Jorge Dela Cruz, Nathaniel Shi, Lingyan

Publication Date

2025-04-08

DOI

10.1021/acs.analchem.4c06966

Copyright Information

This work is made available under the terms of a Creative Commons Attribution-NonCommercial-NoDerivatives License, available at <u>https://creativecommons.org/licenses/by-nc-nd/4.0/</u>

Peer reviewed



Cancer Cell Line Classification Using Raman Spectroscopy of Cancer-**Derived Exosomes and Machine Learning**

Jorge Villazon, Nathaniel Dela Cruz, and Lingyan Shi*

Cite This: Anal. Chem. 2025, 97, 7289-7298



ACCESS

III Metrics & More

Article Recommendations

S Supporting Information

ABSTRACT: Liquid biopsies are an emerging, noninvasive tool for cancer diagnostics, utilizing biological fluids for molecular profiling. Nevertheless, the current methods often lack the sensitivity and specificity necessary for early detection and real-time monitoring. This work explores an advanced approach to improving liquid biopsy techniques through machine learning analysis of the Raman spectra measured to classify distinct exosome solutions by their cancer origin. This was accomplished by conducting principal component analysis (PCA) of the Raman spectra of exosomes from three cancer cell lines (COLO205, A375, and LNCaP) to extract chemically significant features. This reduced set of features was then utilized to train a linear discriminant analysis (LDA) classifier to predict the source of the exosomes. Furthermore, we investigated differences in the lipid composition in these exosomes by their spectra. This spectral similarity analysis revealed differences in lipid profiles



between the different cancer cell lines as well as identified the predominant lipids across all exosomes. Our PCA-LDA framework achieved 93.3% overall accuracy and F1 scores of 98.2%, 91.1%, and 91.0% for COLO205, A375, and LNCaP, respectively. Our results from spectral similarity analysis were also shown to support previous findings of lipid dynamics due to cancer pathology and pertaining to exosome function and structure. These findings underscore the benefits of enhancing Raman spectroscopy analysis with machine learning, laying the groundwork for the development of early noninvasive cancer diagnostics and personalized treatment strategies. This work potentially establishes the foundation for refining the classification model and optimizing exosome extraction and detection from clinical samples for clinical translation.

INTRODUCTION

Early diagnosis significantly improves the likelihood of successful outcomes of cancer treatments such as radiation and surgery. Despite many improvements in diagnostic technology, nearly half of all cancer cases are still identified only at an advanced stage.¹⁻⁴ Current screening methods like imaging and biopsies have significant drawbacks as they are often expensive, labor-intensive, and invasive. Moreover, they provide limited molecular information for precise characterization and staging that limits their use in early stage diagnosis.^{5,6} Consequently, there is growing interest in developing noninvasive cancer diagnostics that can detect the subtle molecular changes associated with early cancer development.

Liquid biopsies present a noninvasive alternative to traditional biopsies by analyzing cancer-specific biomarkers in bodily fluids like blood, urine, or saliva. They are cost-effective and less labor-intensive and allow for repeated testing with fewer risks and complications. More importantly, they hold the potential to detect cancer at an early stage by identifying circulating cancer biomarkers present long before clinical symptoms or visible tumors develop.^{5–7} However, challenges remain, particularly regarding the biomarkers' specificity. For instance, the prostate-specific antigen, which was once

considered the gold standard for prostate cancer screening, is now recognized as unreliable.^{8,9} Therefore, there is a need for biomarkers that offer more precise biological insights into cancer. Promising biomarkers include those derived directly from tumors, such as circulating tumor DNA, noncoding RNA, extracellular vesicles (EVs), and metabolites.^{4,10–12}

Exosomes refer to EVs ranging from 40 to 160 nm with primary functions including cell-cell communication, waste removal, tissue repair, and immune response modulation. Exosomes are secreted by all cells, including cancer cells. For this reason, exosomes released by cancer cells can be valuable for diagnostics since they harbor tumor-specific molecules typically overexpressed for tumor progression or other cancerrelated activities.^{12–15} Lipids are of particular interest as EVs are structures primarily composed of lipids, and most are also bioactive molecules with roles in signaling pathway regulation,

Received: December 22, 2024 **Revised:** February 9, 2025 Accepted: March 1, 2025 Published: March 27, 2025



induction of tumorigenesis, metastasis, and conferring chemoresistance.^{16–18} Information on the chemical composition of these exosomes can be instrumental not only for cancer detection but also for the characterization of specific cancer types by their unique biochemical profile.

Raman spectroscopy is a powerful tool for facilitating biochemical analysis, known for its high molecular specificity.^{19,20} When light interacts with a sample, most of the scattered light maintains the same frequency as the incident light (Rayleigh scattering); however, a small fraction undergoes a shift in energy, known as Raman scattering. The energy (or frequency shift) between the incident and scattered light is unique to the (ro-)vibrational mode of molecular bonds, providing a detailed chemical fingerprint of the sample.^{21,22} The Raman spectrum of biological materials is typically divided into three distinct regions: (1) the fingerprint region (500- 1800 cm^{-1}), (2) the cell-silent region ($1800-2700 \text{ cm}^{-1}$), and (3) the C-H stretching region $(2700-3100 \text{ cm}^{-1})$.^{23,24} In bioprocessing and biomedical fields, Raman spectroscopy holds several advantages, including being noninvasive, nondestructive, and label-free and requiring minimal sample preparation.²¹⁻²⁵ Recent studies have already shown the usefulness of utilizing Raman spectroscopy to capture cancerspecific chemical signatures from cells and tissue.²⁶⁻²⁸

Computational tools can be leveraged to further enhance the analysis. Machine learning algorithms such as principal component analysis (PCA) and linear discriminant analysis (LDA) can discern patterns in high-dimensional data, improving sensitivity and reducing feature redundancy, given that Raman spectra consist of hundreds to thousands of data points, but not all are relevant for biochemical analysis. PCA is a powerful algorithm that identifies major patterns in data sets and determines the features that maximize variance.^{29,30} To enhance class classification, LDA can be applied following PCA feature extraction, identifying a linear combination of the reduced feature set that maximizes between-class variance while minimizing within-class variance.^{30,31} Machine learning is also useful to reduce subjectivity and mitigate bias from both inter- and intra-analyst variability.

While similar studies have been conducted to distinguish cancerous and normal tissues through Raman spectroscopy and PCA-LDA, research focusing on the classification of different cancer types from Raman spectra is limited.²⁶ Furthermore, much less research has been conducted on promising biomarkers like exosomes. Existing studies on liquid biopsies also do not employ Raman spectroscopy, but rely on techniques like next-generation sequencing and polymerase chain reaction.³² While these methods are highly effective for specific genetic or molecular targets, Raman spectroscopy offers a unique advantage of providing a more comprehensive molecular fingerprint of the entire biochemical landscape, enabling the simultaneous analysis of proteins, lipids, nucleic acids, and other metabolites.³³

This study aims to address this gap by utilizing Raman spectroscopy in conjunction with PCA and LDA to classify various cancer types based on the spectral data of cancerderived exosomes. Furthermore, it explores the lipid composition in cancer cell-derived exosomes through a spectral similarity analysis to provide an overview of the biochemical phenotype of these cancer biomarkers. This innovative approach, displayed in Figure 1, investigates an opportunity to enhance noninvasive cancer diagnostics, with the potential to impact early detection, disease monitoring, and personalized treatment strategies.



Figure 1. Diagram of the workflow for spectral analysis of cancer cellderived exosomes. *Top to bottom*: Spontaneous Raman acquisition of exosome samples, clustering and feature extraction through PCA, classification of exosome identity via LDA of extracted features, and spectral similarity analysis with a library of reference lipid spectra.

EXPERIMENTAL SECTION

Exosome Preparation. Exosomes derived from three cancer cell lines—COLO205, A375, and LNCaP (CellGS, EX301, EX302, and EX305)—were purchased and utilized in this study. COLO205 is a human colorectal adenocarcinoma cell line, A375 is a human melanoma cell line commonly used in skin cancer research due to its highly metastatic nature, and LNCaP is an androgen-sensitive human prostate adenocarcinoma cell line, approximately 20 μ g of the exosome samples was mounted on a 1 mm thick glass slide, where 5 μ L of deionized (DI) water was then added via a pipet, and then sealed with a #1 thickness cover glass.

Spontaneous Raman Spectroscopy. Spontaneous Raman spectra were obtained using a confocal Raman microscope (XploRA PLUS, Horiba) equipped with a 532 nm diode laser source and a 1800 lines/mm grating. The acquisition time is 60 s with an accumulation of 2. The excitation power was approximately 40 mW after passing through a 0.9 NA, 100× objective (MPLN100X, Olympus). The spectral range is from 400 to 3150 cm⁻¹. A total of 143 spectra (COLO205 = 43, A375 = 59, and LNCaP = 41) were collected from the three cell lines. The spectra were preprocessed utilizing baseline correction with asymmetric least-squares with a smoothing parameter (λ) of 5 and asymmetric parameter (p) ranging from 3 to 6. The spectra

were then interpolated to keep the number of features consistent and min-max normalized to mitigate differences due to intensity.

Feature Extraction via PCA. PCA was implemented by using built-in functions from the scikit-learn library. After the minimum number of PCs that met the variance threshold was determined, the weighted sum of loadings for each wave-number was calculated by multiplying the absolute loading of each wavenumber by the explained variance of the corresponding PC and adding these values across all selected PCs.

$$\mu^{(i)} = \frac{1}{m} \sum_{k=1}^{m} |l_k^{(i)}| w_k \tag{1}$$

where $\mu^{(i)}$ is the weighted average for each wavenumber *i*, *m* is the number of PCs utilized, $l_k^{(i)}$ is each wavenumber's PC loading, and w_k is the PC's explained variance ratio. This process enabled the identification of the wavenumbers contributing the variations in the spectral data set. The top 10% of wavenumbers, based on the weighted sum of loadings, were identified as the most relevant features for subsequent analyses.

Classification Using PCA-LDA. As it is a supervised approach, the data set was divided into training, validation, and test sets. Specifically, for each cell line, 50% of the spectra were allocated for training, 10% for validation, and the remaining 40% for testing. LDA was initially applied to the training data to learn the optimal linear combinations of features for class separation. This ratio of training/validation/test sets was chosen due to the size of the total exosome spectra data set with the aim of maximizing the number of spectra classified while maintaining high accuracy. The model was trained on combined data sets from the three cell lines. The analysis was repeated 15 times with performance metrics (F1 score, precision, and recall) averaged across these runs to provide a comprehensive evaluation of the accuracy and robustness of this approach.

Spectral Similarity Analysis. To identify the predominant lipids in each exosome sample, spectral similarity analysis was done using three metrics: (1) mean squared error (MSE), (2) cosine similarity, and (3) cross-correlation. This analysis was done on spectra from samples of 25 lipid subtypes and was restricted to the top 10% of wavenumbers as selected for PCA feature extraction. Each metric was computed between the spectra of each exosome sample and the spectra of the lipid subtypes. The average value of each metric was then calculated for each cell line to summarize the overall spectral similarity. The analysis was performed by using functions from the SciPy and scikit-learn modules.

MSE is calculated as the average squared differences between intensity values of the exosome spectra and lipid spectra, with lower error values indicating greater similarity (eq 2).

$$MSE = \frac{1}{n} \sum_{j=1}^{n} (x_j - x_j^L)^2$$
(2)

where x_j is the spectral intensities for PCA-extracted wavenumbers *j* of a specific data point and x_j^L represents the intensity values of specific lipid *L* spectra for the same wavenumbers, and *n* is the total number of data points. The MSE is a metric of similarity that is scale-dependent, easily skewed by outliers and noise, and highly contingent on the squared Euclidean distance without accounting for differences in spectral shape that are imperative for chemical information.

For these reasons, other metrics provide a better assessment of the spectral similarity. In measuring cosine similarity, each spectrum represents a multidimensional vector and the alignment of two vectors is quantified by their cosine angle (eq 3). Unlike MSE, cosine similarity considers spectral shape, emphasizing similarities in features such as peaks and troughs, regardless of variations in intensity. The cosine similarity score ranges from 0 (complete dissimilarity/orthogonality) to 1 (perfect similarity).

$$\cos \theta = \frac{\sum_{j=1}^{n} x_j \cdot x_j^{L}}{\sqrt{\sum_{j=1}^{n} (x_j)^2} \cdot \sqrt{\sum_{j=1}^{n} (x_j^{L})^2}}$$
(3)

The normalized cross-correlation (NCC) coefficient here evaluates the similarity by measuring the displacement of one signal to another independent of intensity using a sliding dot product. NCC accounts for differences not only in terms of shape but also in shifts in the spectra, which is useful to account for shifts due to experimental variations. The NCC score ranges from -1 (perfectly negatively correlated) to 1 (perfectly aligned and correlated). In eq 4, τ represents the shift between the signals/spectra.

$$C(\tau) = \frac{\sum_{j=1}^{n} x_{j} \cdot x_{j-\tau}^{L}}{\sqrt{\sum_{j=1}^{n} (x_{j})^{2}} \cdot \sqrt{\sum_{j=1}^{n} (x_{j}^{L})^{2}}}$$
(4)

RESULTS AND DISCUSSION

Feature Extraction via PCA. Approximately 900 features were acquired upon collecting the spontaneous Raman spectra of the exosomes—each corresponding to a wavenumber between 400 and 3150 cm^{-1} , as seen in Figure 2. By reducing the dimensionality of the exosome spectra data set with PCA, we were able to ascertain the magnitude at which specific wavenumbers contributed to the variance between each individual spectrum. The goal of this was to identify which wavenumbers, and thereby which vibrational modes, were significant in clustering and differentiating the cancerous



Figure 2. Raman spectra from each of the exosome origins. Average spectra (*bold line*) from spectra acquisition across the different cancer cell exosomes: COLO205 (*red*), A375 (*blue*), and LNCaP (*green*). Standard deviation represented by a lighter outline surrounding average spectra. The Raman peaks corresponding to S–S, C–C, and CH₃ stretching are similarly displayed.



Figure 3. Results of PCA. (A) Line graphs of the explained variance ratio (*blue*) and the cumulative explained variance (*orange*) for each PC. (B) 3D scatterplot of the first three PCs of each exosome spectra, labeled by their source: A375 (*blue*), COLO2O5 (*red*), and LNCaP (*green*). (C) 2D scatterplot comparing PC1 against PC2 (*left*) and PC2 against PC3 (*right*).

origins of the exosomes. To optimize the dimensionality reduction of the exosome spectral data set, the number of PCs increased until the cumulative explained variance of the transformed data set was >95%. Clustering showed discernible separation of COLO205 spectra along the first initial components compared to spectra from the A375 and LNCaP exosomes (Figure 3). Conversely, there is a noticeable overlap in the clusters of LNCaP and A375 spectra, suggesting similarities in their spectral features compared to COLO205.

After determining the first five PCs that explained approximately 96% of the data set variance, we then identified the contribution of each wavenumber to these components, utilizing the PCA loadings/coefficients. By using a weighted metric, we quantified the influence of each wavenumber on the variance explained by each component. This weighted average was measured by multiplying the magnitude of each component loading (Figure 4) with the corresponding explained variance. Wavenumbers in the cell-silent region $(1800-2700 \text{ cm}^{-1})$, a region with little to no Raman signal in cells and tissues, exhibited almost negligible contribution to the variance between spectra. Specifically, we found that the first component was primarily characterized by contributions from S-S disulfide stretching (532 cm⁻¹) and disulfide conformations (510-545 cm⁻¹), in particular gauche-gauche-trans and trans-gauche-trans conformations in cystine.34-36 We also found high loading values for 2890 and 2893 cm^{-1} , which correspond to the CH₂ asymmetric stretching region.^{37,38} Meanwhile, the second component expressed greater loadings from wavenumbers in the CH_2 symmetric stretching (2820– 2850 cm⁻¹). Within this band, the peak at 2850 cm⁻¹ is

notable for being commonly used to detect the total lipid concentration in Raman microscopy.^{37–43} Similarly, the wavenumbers that contribute heavily to the third and fifth component include the 2930 cm⁻¹ peak utilized for measuring total protein expression^{37,38,41,42,44} and the CH₃ stretching bands.

By our weighted metric, we measured the greatest contributions to the variance were from wavenumbers in the fingerprint and CH stretching regions. Feature extraction selected the top 10% of wavenumbers utilizing the same weighted metric. The resulting 87 wavenumbers could be grouped into three distinct sets: 31 wavenumbers ranging from 425 to 584 cm⁻¹, 15 from 1068 to 1120 cm⁻¹, and 41 wavenumbers accounting for a significant subsection of the CH stretching region (2808-2957 cm⁻¹). These subsets of the Raman spectra align with the vibrational modes for S-S disulfide stretching and conformation in proteins, the C-C and C–O stretching of saccharides, and the bands for CH₂ and CH₃ stretching.^{35,38,39,45} We thus find that the greatest contributions to the exosome spectral profile are from the C-H stretching of lipids, saturated lipids and proteins, changes in the secondary and tertiary structure of proteins by disulfide bonds, and the organic backbones of saccharides.⁴⁶⁻

Classification Using PCA-LDA. Using these PCAextracted features, we employed LDA for optimal class separation and classification. The PCA-LDA classification algorithm was trained with a training/validation/test split of 50%/10%/40% of the exosome spectral data set. Results from classification were averaged across 15 iterations to account for randomness in data selection and produce more robust



Figure 4. Graph of loadings for first five PCs and PCA-extracted features. Line graphs for the loading per wavenumber in each PC: PC1 (*gold*), PC2 (*teal*), PC3 (*dark pink*), PC4 (*dark red*), and PC5 (*purple*). Black Xs indicate the location of the highest magnitude loadings from each of the five components before weighted averaging.

findings (Figure 5). Across these iterations, the average accuracy achieved for the classification approach was 93.3%. Detailed performance metrics for each cell line were also determined to reveal important insights into the model's classification behavior concerning the three cancer cell lines (Table 1).

The PCA-LDA approach achieved an average precision of 97.3%, a recall of 99.3%, and an F1 score of 98.2% for COLO205. For the A375 cell line, a precision of 98.8% meant that the classification algorithm performed better at reducing misclassification of exosome spectra from other cell lines as A375 spectra, reducing the number of false positives. However, the lower recall of 84.7% indicated that some A375 exosome spectra were misclassified as belonging to one of the other cell line spectra. For the LNCaP cell line, the recall of 98.8% score



Figure 5. Results of PCA-LDA classification. (A) Scatterplot of the linear determinants of the exosome spectra: A375 (*blue*), COLO205 (*red*), and LNCaP (*green*). Dots represent accurate prediction and xs represent inaccurate prediction. (B) Confusion matrix of PCA-LDA predictions, resulting in an F1 score of 96% for A375, 100% for COLO205, and 94% for LNCaP.

Table 1. Table of Average PCA-LDA Performance Metricsfor Each Cancer Cell Line

	Colo205	A375	LNCaP
precision	97.3	98.8	84.8
recall	99.3	84.7	98.8
F1 score	98.2	91.0	91.1

indicated that LNCaP is more correctly identified as LNCaP across every iteration; however, an average precision of 84.8% points to non-LNCaP spectra being misclassified as LNCaP. This was directly observed across each iteration of the LDA classifier, with A375 spectra misclassified as LNCaP more frequently than with COLO205. In summary, classification metrics indicate that the PCA-LDA classification framework was highly reliable in predicting COLO205 cells with minimal

false positives. The high recall performance suggested that instances of COLO205 spectra were more correctly classified, potentially indicating a greater difference in chemical composition compared to the other two cell lines. Conversely, the recall and precision scores of A375 and LNCaP exosome spectra indicate a closer overlap in the PCA-extracted features between the two. To provide a more comprehensive understanding of the results from this LDA classification, we then utilized spectral similarity analysis to examine the molecular composition of these exosomes.

Spectral Similarity Analysis. To further investigate the differences between exosome classes, spectral similarity analysis between exosome and specific molecule spectra was performed on the features extracted through PCA. In our analysis, we primarily focused on the lipid composition of these exosomes due to the significant role that lipids play in vesicle formation and signaling pathways in cancer progression. Figure 6A displays the regions of the spectra that are utilized for spectral similarity from both exosome and lipid data.

This analysis, conducted with the spectra of 25 lipid subtypes, employed three metrics: MSE, cosine similarity, and the cross-correlation coefficient. By comparison of the scores across all three metrics, the contribution of lipid subtypes could be identified and ranked. Figure 6B displays the metric scores for specific lipid subtypes that demonstrate higher similarity across the different metrics and for each cell line source.

For the COLO205 exosome spectra, phospholipids, triglycerides, cardiolipins, and cholesterol exhibited the closest spectral similarity across all metrics. Furthermore, we found that cosine similarity and cross-correlation yielded identical rankings for the top 10 lipid similarities for COLO205, indicating a minimal shift in spectra. These lipids included cholesterol esters, omega-3 25:5, phosphatidylcholine, phosphatidylcholine 18:1, and phosphatidylethanolamine 18:1. Similarly, for the A375 and LNCaP cell lines, the top 10 lipids were consistent across both cosine similarity and crosscorrelation analyses. The top six lipids-omega-3 25:5, cholesterol, cardiolipin, cholesterol ester, phosphatidylcholine 18:1, and phosphatidylethanolamine 18:1-were identical in both cell lines, including their order. Notably, phospholipids were ranked higher in LNCaP than in A375. This nearidentical lipid spectra similarity profile highlights the remarkably similar Raman spectra of the two, despite coming from distinct cell lines as previously noted in the PCA results, and potentially is a factor in the misclassification by PCA-LDA classification.

The results of the spectral similarity analysis are in line with established findings in exosome and cancer cell biology, with many of the highly ranked lipids reported as being significant to exosome formation or cancer pathology. Cholesterol consistently ranked in the top three for all cell lines and metrics. Lipid composition studies of EVs from different cells including platelets, PC-3 (prostate adenocarcinoma), and Blymphocytes revealed it to be the most predominant lipid followed by phosphatidylcholine.⁵² It also plays a crucial role in maintaining the structural stability and phospholipid bilayer arrangement of the exosomal membrane.53 By enhancing exosomal membrane fluidity, cholesterol enables more effective binding with CD8(+) T cells, allowing exosomes to temporarily suppress T cell activity for immune evasion, as seen in glioblastoma-derived exosomes containing programmed death-ligand 1.54 Disruption of lipid metabolism by



Figure 6. Results of spectral similarity analysis for phospholipids (PLS), cholesterol, cholesterol ester, and cardiolipin. (A) Graphs of Raman spectra from exosomes—A375 (*blue*), COLO205 (*red*), and LNCaP (*green*)—and four different lipid subtypes—*top to bottom*: cholesterol, cholesterol ester, phospholipids, and cardiolipins. Gray vertical lines indicate PCA-extracted wavenumbers along each spectrum for comparison. (B) Tables of the scores for each similarity metric—*top to bottom*: MSE, cosine similarity, and cross-correlation—for each lipid subtype.

cholesterol in immune cells has been shown to hinder cancer cell detection.⁵¹ Cholesterol ester ranks among the top five for all cell lines and metrics aligning with prior reports that cite it as a prevalent cancer marker. It is typically stored in the form of lipid droplets, which cancer cells utilize for metabolism, as supported by high levels of acyl-CoA acyltransferase 1 (ACAT1) and lysosomal acid lipase that facilitate its

conversion and breakdown.^{55,56} The loss of tumor protein 53 (p53) function, due to mutations or deletions in the p53 gene that are observed in many cancers, has also been shown to be linked to excessive cholesterol ester biosynthesis and promotion of hepatocellular carcinoma in mice.⁵⁷ Furthermore, elevated cholesterol esterification has been linked to both the proliferation of leukemia and aggressiveness of prostate cancer cells.^{58,59}

Furthermore, we can also determine specific markers from each cell line. The high degree of similarity between COLO205 and phospholipid spectra aligns with previous findings of glycerophospholipids as the most abundant lipid class in LIM215 colorectal cancer cells and their exosomes, comprising 91.5% and 68.3% of the total lipids, respectively, based on mass spectrometry. Glycerophospholipids, particularly ether-linked species, function as endogenous antioxidants and support membrane fusion, both critical for cancer cell survival and aggressiveness. High similarity with triglycerides is consistent with the result of a meta-analysis linking dyslipidemia (increased triglyceride and cholesterol) and increased colorectal cancer risk. Although the exact biological mechanisms remain unclear, triglycerides may serve as an energy source for neoplastic cells.

The similarity between LNCaP exosome and phospholipid spectra agrees with findings showing that glycerophospholipids are abundant in LNCaP cell-derived exosomes. This is likely due to prostate cancer cells' reliance on lipid metabolism, particularly glycerophospholipid metabolism, as well as autophagy and the ferroptosis pathway, rather than glycolysis.^{60–62} Additionally, a lipidomic study using quantitative mass spectrometry on prostate cancer cells reported heightened levels of all classes of phospholipids including phosphatidylethanolamine. Phospholipids like phosphatidylethanolamine are closely associated with cancer cell proliferation through their involvement in Akt-mediated signaling pathways, further supporting the detection of phosphatidylethanolamine 18:1 in LNCaP exosomes and other cell lines.⁶³

In addition to being one of the most abundant components of mammalian membranes, phosphatidylcholine, a type of phospholipid, is produced in large quantities because its breakdown products are essential for meeting the demands of rapidly growing cancer cells. These metabolites promote DNA repair, lipid droplet synthesis, and autophagy, among other intracellular processes that support drug resistance.⁶⁴ Oxidized phosphatidylcholines were also found to facilitate epithelialmesenchymal transition (EMT) that increases cancer cells' metastatic potential.⁶⁵ Cardiolipin is another lipid suggested to be present in high amounts in all three of the exosomes based on similarity metrics. This type of phospholipid is a characteristic lipid of the mitochondria. Cardiolipin plays an important role in the biosynthesis of acetyl-CoA, which has been implicated in several activities for cancer progression.⁶⁶ Acetylation and activation of the Smad2 transcription factor is facilitated by an increase in intracellular acetyl-CoA, eventually resulting in EMT and induction of metastasis in breast cancer cells.⁶⁷ Furthermore, global histone acetylation with accumulation of acetyl-CoA also contributes to pancreatic tumorigenesis.⁶⁸ High levels of cardiolipins have similarly been observed in thyroid carcinoma and breast cancer epithelial cells.^{66,69}

Interestingly, omega-3 25:5 emerges as the most predominant lipid in the A375 and LNCaP cell lines based on spectral similarity. While most studies have concentrated on the therapeutic effects of omega-3 fatty acids, their exact mechanisms remain not fully understood. Furthermore, the role of omega-3 fatty acids in cancer progression has not been well-studied. Additional research is needed to validate the levels of omega-3 fatty acids in cells and exosomes and to elucidate the potential roles that they may play in cancer progression. This finding suggests that expanding the lipid library for analysis, by including even more specific compounds, could provide additional insights and potentially shift the prominence of omega-3 25:5 as identified.

The relative composition of these lipids varies significantly among exosomes, depending on the cell line, subtype, physiological stage, and associated biological activity, as lipids are not only key structural components but also integral constituents of the overall cargo. This variation exemplifies how metabolic reprogramming is a hallmark of cancer.

Limitations and Future Work. Due to the breadth of this study, it is important to consider the limitations of our data set and analysis. The lipid profiles presented in this work are characterized by spectral similarity to other reference Raman spectra rather than more direct characterization methods for lipidomic analysis, and are constrained by the lipid spectral library that contains only 25 lipid subtypes, which may not fully represent the complete lipid composition. Exosomes also consist of a wide range of molecules, including lipids, proteins, and miRNA, that all contribute to the signals in the Raman spectrum. This molecular diversity presents challenges as overlapping signals can make it hard to isolate and accurately attribute distinct spectral features to lipids. Despite these limitations, the results offer valuable context regarding which lipids may be overexpressed in these exosomes, emphasizing their roles in maintaining the exosomal structural integrity and supporting cancer-related activities.

While the PCA-LDA classification framework demonstrated robust overall performance, these results are constrained by the small sample size used, which may lead to the challenges in classification, especially for LNCaP and A375. For COLO205, limited data may lead to an overestimation of the model's accuracy. Thus, a larger spectral library is needed to simultaneously confirm the robustness of these findings and improve the classification performance for LNCaP and A375. Nevertheless, the overall high accuracy and robust performance of the model across all three cell lines underscore the effectiveness of the approach for classification of cancer exosomes based on their Raman spectrum.

Future work could explore improving the classification of A375 and LNCaP exosomes using more advanced machine learning algorithms such as support vector machines or deep learning models like convolutional neural networks for better feature extraction and pattern recognition from the Raman spectra. Additionally, it is crucial to extend this methodology to encompass other cancer types, including lung, pancreatic, and brain cancers, while also focusing on the classification of cancer subtypes. For instance, breast cancer displays a high degree of heterogeneity with multiple subtypes and phenotypes.^{41,70} The spectra of the five molecular subtypes may present greater challenges in differentiation, potentially making them even more difficult to distinguish than the cases of A375 and LNCaP. Successfully classifying breast cancer molecular subtypes using Raman spectroscopy combined with machine learning could significantly enhance the identification of these subtypes, which currently relies heavily on resource-expensive immunohistochemistry.⁷¹ Raman spectroscopy provides a

label-free alternative to the current methodology that pairs well with emerging diagnostic techniques.

Building upon these initial findings for clinical translation, the next step would be to apply this methodology to exosomes derived from clinical samples, such as bodily fluids. While this study utilized commercially purchased purified exosomes, exploring the processes of exosome extraction and purification from bodily fluids will be crucial for clinical applications. This entails optimizing purification methods, Raman spectroscopy, and machine learning techniques to accommodate the complexities of these samples, which can significantly impair sensitivity. Addressing challenges such as the low concentrations of exosomes in liquid samples may be achieved through advanced methods like surface-enhanced Raman spectroscopy, coherent anti-Stokes Raman spectroscopy, or stimulated Raman scattering, which amplify spectral signals and enhance detection capabilities.^{22,72-75} This integration could ultimately pave the way for more effective, early noninvasive cancer diagnostic tools.

CONCLUSIONS

The classification metrics and an overall classification accuracy of 93.3% of the PCA-LDA model indicate that the classification framework can effectively identify differences among the latent features of the Raman spectra of the COLO205, A375, and LNCaP exosomes. Utilizing a more robust learning algorithm for classification that considers nonlinearities in the relationship between specific features of the spectrum and increasing the size and range of samples in the spectral data set could be made to improve the accuracy and sensitivity crucial for clinical practice. This study also provides insights and a comparative analysis of the lipid composition of COLO205, A375, and LNCaP exosomes for the first time, addressing the scarcity of studies that explore the lipid profiles of these cancer cell lines and their exosomes. The findings from this study have profound implications for translational research. By elucidating distinct spectral signatures associated with exosomes of specific cancer types, we have enhanced diagnostic capabilities for cancer detection and staging. The utilization of exosomes and Raman spectroscopy for noninvasive diagnostics, particularly for liquid biopsies, presents a promising approach for early malignancy detection by revealing subtle biochemical changes linked to cancer progression and for continuous monitoring and personalized treatment.

ASSOCIATED CONTENT

Data Availability Statement

Code and exosome Raman spectra data are available at https://github.com/lingyanshi2020/exosome_spectra.

Supporting Information

The Supporting Information is available free of charge at https://pubs.acs.org/doi/10.1021/acs.analchem.4c06966.

Table of PCA-extracted Raman Bands and corresponding Chemical Bonds (PDF)

AUTHOR INFORMATION

Corresponding Author

Lingyan Shi – Shu Chien-Gene Lay Department of Bioengineering, University of California San Diego, La Jolla, California 92093, United States; Aiiso Yufeng Li Family Department of Chemical and Nano Engineering, Department of Electrical and Computer Engineering, Institute of Engineering in Medicine, and Synthetic Biology Institute, University of California San Diego, La Jolla, California 92093, United States; Email: l2shi@ucsd.edu

Authors

- Jorge Villazon Shu Chien-Gene Lay Department of Bioengineering, University of California San Diego, La Jolla, California 92093, United States; © orcid.org/0009-0006-8296-5894
- Nathaniel Dela Cruz Department of Chemical and Biomolecular Engineering, Korea Advanced Institute of Science and Technology (KAIST), Daejeon 34141, Republic of Korea; orcid.org/0009-0005-9639-9072

Complete contact information is available at: https://pubs.acs.org/10.1021/acs.analchem.4c06966

Author Contributions

The manuscript was written through contributions of all authors and all authors have given approval to the final version of the manuscript.

Notes

The authors declare no competing financial interest.

ACKNOWLEDGMENTS

The authors would like to acknowledge the lab group members that contributed to data acquisition and discussion, in particular Dr. Yajuan Li and Menghan Li. We acknowledge support from the UCSD Startup funds, NIH U54DK134301, NIH R01GM149976, NIH U01AI167892, NIH SR01NS111039, NIH R21NS125395, NIH U54CA132378, CZI DAF2023-328667, and Sloan Research Fellow Award.

REFERENCES

 S. Alshuhri, M.; Al-Musawi, S. G.; Al-Alwany, A. A.; Uinarni, H.; Rasulova, I.; Rodrigues, P.; Alkhafaji, A. T.; Alshanberi, A. M.; Alawadi, A. H.; Abbas, A. H. *Pathol., Res. Pract.* **2024**, 253, 154996.
 Whitaker, K. *Lancet Oncol.* **2020**, 21 (1), 6–8.

(3) Hawkes, N. *BMJ* **2019**, *364*, 1408.

(4) Crosby, D.; Bhatia, S.; Brindle, K. M.; Coussens, L. M.; Dive, C.; Emberton, M.; Esener, S.; Fitzgerald, R. C.; Gambhir, S. S.; Kuhn, P.; Rebbeck, T. R.; Balasubramanian, S. *Science* **2022**, *375* (6586), No. eaay9040.

- (5) Milner, D. A., Jr; Lennerz, J. K. Life 2024, 14 (7), 833.
- (6) Foser, S.; Maiese, K.; Digumarthy, S. R.; Puig-Butille, J. A.; Rebhan, C. *Clin. Chem.* **2024**, *70* (1), 27–32.
- (7) Connal, S.; Cameron, J. M.; Sala, A.; Brennan, P. M.; Palmer, D.
- S.; Palmer, J. D.; Perlow, H.; Baker, M. J. *J. Transl. Med.* **2023**, *21* (1), 118.
- (8) Punglia, R. S.; D'Amico, A. V.; Catalona, W. J.; Roehl, K. A.; Kuntz, K. M. N. Engl. J. Med. **2003**, 349 (4), 335–342.
- (9) Merriel, S. W. D.; Pocock, L.; Gilbert, E.; Creavin, S.; Walter, F. M.; Spencer, A.; Hamilton, W. *BMC Med.* **2022**, *20*, 54.

(10) Cescon, D. W.; Bratman, S. V.; Chan, S. M.; Siu, L. L. Nat. Cancer **2020**, 1 (3), 276–290.

- (11) Tatischeff, I. Cancers 2023, 15 (5), 1456.
- (12) Zhou, B.; Xu, K.; Zheng, X.; Chen, T.; Wang, J.; Song, Y.; Shao, Y.; Zheng, S. Signal Transduction Targeted Ther. **2020**, *5*, 144.
- (13) Maia, J.; Čaja, S.; Strano Moraes, M. C.; Couto, N.; Costa-Silva, B. Front. Cell Dev. Biol. 2018, 6, 18.

(14) Zhang, W.; Peng, P.; Shen, K. Zhongguo Yixue Kexueyuan Xuebao **2016**, 38 (4), 480–483.

(15) Wang, Z.; Wang, Q.; Qin, F.; Chen, J. Front. Cell Dev. Biol. 2024, 12, 1344705.

(16) Kalluri, R.; LeBleu, V. S. Science **2020**, 367 (6478), No. eaau6977.

(17) Luo, X.; Zhao, X.; Cheng, C.; Li, N.; Liu, Y.; Cao, Y. *Exp. Mol. Med.* **2018**, 50 (9), 1–10.

(18) Vasseur, S.; Guillaumond, F. Oncogenesis 2022, 11 (1), 46.

(19) Tu, Q.; Chang, C. Nanomed. Nanotechnol. Biol. Med. 2012, 8 (5), 545-558.

(20) Choo-Smith, L.-P.; Edwards, H. G. M.; Endtz, H. P.; Kros, J. M.; Heule, F.; Barr, H.; Robinson, J. S., Jr.; Bruining, H. A.; Puppels, G. J. *Biopolymers* **2002**, *67* (1), 1–9.

(21) Butler, H. J.; Ashton, L.; Bird, B.; Cinque, G.; Curtis, K.; Dorney, J.; Esmonde-White, K.; Fullwood, N. J.; Gardner, B.; Martin-Hirsch, P. L.; Walsh, M. J.; McAinsh, M. R.; Stone, N.; Martin, F. L. *Nat. Protoc.* **2016**, *11* (4), 664–687.

(22) Shi, L.; Fung, A. A.; Zhou, A. Quant Imaging Med. Surg. 2020, 11 (3), 1078–1101.

(23) Camp Jr, C. H.; Cicerone, M. T. Nat. Photonics 2015, 9 (5), 295–305.

(24) Lee, M.; Tipping, W. J. 15 - Stimulated Raman Scattering for Cell and Tissue Imaging. *Neurophotonics and Biomedical Spectroscopy*; Alfano, R. R., Shi, L., Eds.; *Nanophotonics*; Elsevier, 2019; pp 401– 425.

(25) Esmonde-White, K. A.; Cuellar, M.; Lewis, I. R. Anal. Bioanal. Chem. 2022, 414 (2), 969.

(26) Ralbovsky, N. M.; Lednev, I. K. Spectrochim. Acta, Part A 2019, 219, 463-487.

(27) Byrne, H. J.; Baranska, M.; Puppels, G. J.; Stone, N.; Wood, B.; Gough, K. M.; Lasch, P.; Heraud, P.; Sulé-Suso, J.; Sockalingum, G. D. *Analyst* **2015**, *140* (7), 2066–2073.

(28) Cui, S.; Zhang, S.; Yue, S. J. Healthcare Eng. 2018, 2018 (1), 1–11.

(29) Beattie, J. R.; Esmonde-White, F. W. L. Appl. Spectrosc. 2021, 75 (4), 361-375.

(30) Odhiambo Omuya, E.; Onyango Okeyo, G.; Waema Kimwele, M. *Expert Syst. Appl.* **2021**, *174*, 114765.

(31) Hayati, R.; Munawar, A. A.; Lukitaningsih, E.; Earlia, N.; Karma, T.; Idroes, R. Case Stud. Chem. Environ. Eng. 2024, 9, 100552.

(32) Cheung, A. H.-K.; Chow, C.; To, K.-F. J. Thorac. Dis. 2018, 10 (S14), S1645-S1651.

(33) Dodo, K.; Fujita, K.; Sodeoka, M. J. Am. Chem. Soc. 2022, 144 (43), 19651–19667.

(34) Kuhar, N.; Sil, S.; Umapathy, S. Spectrochim. Acta, Part A 2021, 258, 119712.

(35) Stone, N.; Kendall, C.; Smith, J.; Crow, P.; Barr, H. Faraday Discuss. 2004, 126 (0), 141–157.

(36) Greig, J. C.; Tipping, W. J.; Graham, D.; Faulds, K.; Gould, G. W. *Analyst* **2024**, *149* (19), 4789–4810.

(37) Howell, N. K.; Arteaga, G.; Nakai, S.; Li-Chan, E. C. J. Agric. Food Chem. 1999, 47 (3), 924–933.

(38) Bresson, S.; Bormann, D.; Khelifa, B. Vib. Spectrosc. **1998**, *16* (2), 163–171.

(39) Wang, C.-H.; Huang, C.-C.; Lin, L.-L.; Chen, W. J. Raman Spectrosc. 2016, 47 (8), 940–947.

(40) El Hadri, M.; Achahbar, A.; El Khamkhami, J.; Khelifa, B.; Faivre, V.; Cong, T. T.; Bougrioua, F.; Bresson, S. *Vib. Spectrosc.* **2013**, *64*, 78–88.

(41) Fu, D.; Yu, Y.; Folick, A.; Currie, E.; Farese, R. V., Jr; Tsai, T.-H.; Xie, X. S.; Wang, M. C. J. Am. Chem. Soc. **2014**, 136 (24), 8820– 8828.

(42) Jia, H.; Yue, S. J. Phys. Chem. B 2023, 127 (11), 2381.

(43) Li, Z.; Nguyen, C.; Jang, H.; Hoang, D.; Min, S.; Ackerstaff, E.; Koutcher, J. A.; Shi, L. *Biomed. Opt. Express* **2023**, *14* (11), 5764–5780.

(44) Li, Y.; Munoz-Mayorga, D.; Nie, Y.; Kang, N.; Tao, Y.; Lagerwall, J.; Pernaci, C.; Curtin, G.; Coufal, N. G.; Mertens, J.; Shi, L.; Chen, X. *Cell Metab.* **2024**, *36* (6), 1351–1370e8.

(45) Zhang, W.; Li, Y.; Fung, A. A.; Li, Z.; Jang, H.; Zha, H.; Chen, X.; Gao, F.; Wu, J. Y.; Sheng, H.; Yao, J.; Skowronska-Krawczyk, D.; Jain, S.; Shi, L. *Nat. Commun.* **2024**, *15* (1), 1599.

(46) Freudiger, C. W.; Pfannl, R.; Orringer, D. A.; Saar, B. G.; Ji, M.; Zeng, Q.; Ottoboni, L.; Ying, W.; Waeber, C.; Sims, J. R.; De Jager, P. L.; Sagher, O.; Philbert, M. A.; Xu, X.; Kesari, S.; Xie, X. S.; Young, G. S. *Lab. Invest.* **2012**, *92* (10), 1492–1502.

(47) Wiercigroch, E.; Szafraniec, E.; Czamara, K.; Pacia, M. Z.; Majzner, K.; Kochan, K.; Kaczor, A.; Baranska, M.; Malek, K. Spectrochim. Acta, Part A **2017**, 185, 317–335.

(48) Robinson, P. J.; Kanemura, S.; Cao, X.; Bulleid, N. J. J. Biol. Chem. 2020, 295 (8), 2438.

(49) Biswas, N.; Waring, A. J.; Walther, F. J.; Dluhy, R. A. Biochim. Biophys. Acta, Biomembr. 2007, 1768 (5), 1070–1082.

(50) Qin, M.; Wang, W.; Thirumalai, D. Proc. Natl. Acad. Sci. U.S.A. **2015**, 112 (36), 11241–11246.

(51) Kaminský, J.; Kapitán, J.; Baumruk, V.; Bednárová, L.; Bour, P. J. Phys. Chem. A **2009**, 113 (15), 3594–3601.

(52) Ghadami, S.; Dellinger, K. Front. Mol. Biosci. 2023, 10, 1198044.

(53) Wang, W.; Zhu, N.; Yan, T.; Shi, Y.-N.; Chen, J.; Zhang, C.-J.; Xie, X.-J.; Liao, D.-F.; Qin, L. Cell Commun. Signaling **2020**, 18 (1), 119.

(54) Ricklefs, F. L.; Alayo, Q.; Krenzlin, H.; Mahmoud, A. B.; Speranza, M. C.; Nakashima, H.; Hayes, J. L.; Lee, K.; Balaj, L.; Passaro, C.; Rooj, A. K.; Krasemann, S.; Carter, B. S.; Chen, C. C.; Steed, T.; Treiber, J.; Rodig, S.; Yang, K.; Nakano, I.; Lee, H.; Weissleder, R.; Breakefield, X. O.; Godlewski, J.; Westphal, M.; Lamszus, K.; Freeman, G. J.; Bronisz, A.; Lawler, S. E.; Chiocca, E. A. *Sci. Adv.* **2018**, *4* (3), No. eaar2766.

(55) Xiao, M.; Xu, J.; Wang, W.; Zhang, B.; Liu, J.; Li, J.; Xu, H.; Zhao, Y.; Yu, X.; Shi, S. *Exp. Mol. Med.* **2023**, *55* (9), 1982–1995.

(56) Xia, W.; Wang, H.; Zhou, X.; Wang, Y.; Xue, L.; Cao, B.; Song, J. Front. Pharmacol **2023**, *14*, 928821.

(57) Li, Y.; Karin, M.; Prochownik, E. V. *Explor Target Antitumor Ther.* **2023**, *4* (5), 1122–1127.

(58) Mulas, M. F.; Abete, C.; Pulisci, D.; Pani, A.; Massidda, B.; Dessi, S.; Mandas, A. *Cell Proliferation* **2011**, *44* (4), 360–371.

(59) Yue, S.; Li, J.; Lee, S.-Y.; Lee, H. J.; Shao, T.; Song, B.; Cheng, L.; Masterson, T. A.; Liu, X.; Ratliff, T. L.; Cheng, J.-X. *Cell Metab.* **2014**, *19* (3), 393–406.

(60) Lydic, T. A.; Townsend, S.; Adda, C. G.; Collins, C.; Mathivanan, S.; Reid, G. E. *Methods* **2015**, *87*, 83–95.

(61) Yao, X.; Tian, Z. Cancer Causes Control **2015**, 26 (2), 257–268. (62) Yi, X.; Li, Y.; Hu, X.; Wang, F.; Liu, T. J. Cancer **2021**, 12 (10), 2893–2902.

(63) Sorvina, A.; Bader, C. A.; Caporale, C.; Carter, E. A.; Johnson, I. R. D.; Parkinson-Lawrence, E. J.; Simpson, P. V.; Wright, P. J.; Stagni, S.; Lay, P. A.; Massi, M.; Brooks, D. A.; Plush, S. E. *Oncotarget* **2018**, 9 (85), 35541–35552.

(64) Saito, R. F.; Andrade, L. N. S.; Bustos, S. O.; Chammas, R. Front. Immunol. 2022, 13, 768606.

(65) Seok, J. K.; Hong, E.-H.; Yang, G.; Lee, H. E.; Kim, S.-E.; Liu, K.-H.; Kang, H. C.; Cho, Y.-Y.; Lee, H. S.; Lee, J. Y. *Cells* **2021**, *10* (3), 558.

(66) Ahmadpour, S. T.; Mahéo, K.; Servais, S.; Brisson, L.; Dumas, J.-F. Int. J. Mol. Sci. 2020, 21 (21), 8031.

(67) Rios Garcia, M.; Steinbauer, B.; Srivastava, K.; Singhal, M.; Mattijssen, F.; Maida, A.; Christian, S.; Hess-Stumpp, H.; Augustin, H. G.; Müller-Decker, K.; Nawroth, P. P.; Herzig, S.; Berriel Diaz, M. *Cell Metab.* **2017**, *26* (6), 842–85565.

(68) Carrer, A.; Trefely, S.; Zhao, S.; Campbell, S. L.; Norgard, R. J.; Schultz, K. C.; Sidoli, S.; Parris, J. L. D.; Affronti, H. C.; Sivanand, S.; Egolf, S.; Sela, Y.; Trizzino, M.; Gardini, A.; Garcia, B. A.; Snyder, N. W.; Stanger, B. Z.; Wellen, K. E. *Cancer Discovery* **2019**, *9* (3), 416– 435.

(69) DeHoog, R. J.; Zhang, J.; Alore, E.; Lin, J. Q.; Yu, W.; Woody, S.; Almendariz, C.; Lin, M.; Engelsman, A. F.; Sidhu, S. B.; Tibshirani, R.; Suliburk, J.; Eberlin, L. S. *Proc. Natl. Acad. Sci. U.S.A.* **2019**, *116* (43), 21401–21408.

(70) Turner, K. M.; Yeo, S. K.; Holm, T. M.; Shaughnessy, E.; Guan, J.-L. Am. J. Physiol.: Cell Physiol. 2021, 321 (2), C343-C354.

(71) Lopez-Tarruella, S.; Del Monte-Millán, M.; Roche-Molina, M.; Jerez, Y.; Echavarria Diaz-Guardamino, I.; Herrero López, B.; Gamez Casado, S.; Marquez-Rodas, I.; Alvarez, E.; Cebollero, M.; Massarrah, T.; Ocaña, I.; Arias, A.; García-Sáenz, J. A..; Moreno Anton, F.; Olier Garate, C.; Moreno Muñoz, D.; Marrupe, D.; Lara Álvarez, M. A..; Enrech, S.; Bueno Muiño, C.; Martín, M. *Breast Cancer Res. Treat.* **2024**, 203 (1), 163–172.

(72) Han, X. X.; Rodriguez, R. S.; Haynes, C. L.; Ozaki, Y.; Zhao, B. *Nat. Rev. Methods Primers* **2021**, *1* (1), 1–17.

(73) Haroon, M.; Tahir, M.; Nawaz, H.; Majeed, M. I.; Al-Saadi, A. A. Photodiagn. Photodyn. Ther. **2022**, 37, 102690.

(74) Li, S.; Li, Y.; Yi, R.; Liu, L.; Qu, J. Front. Phys. **2020**, 8, 598420. (75) Hu, F.; Shi, L.; Min, W. Nat. Methods **2019**, 16 (9), 830-842.