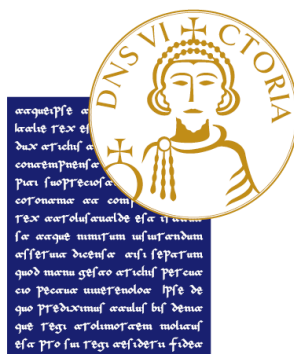


University of Sannio



Department of Engineering

PhD course in Information Technologies for Engineering

Cycle XXXVII

PhD Thesis

Title

Raman Spectroscopy, Surface-Enhanced Raman Spectroscopy (SERS) and Tip-Enhanced Raman Spectroscopy (TERS) for advanced diagnostics

Supervisors:

Prof. Andrea Cusano

Prof. Marco Pisco

PhD Candidate:

Concetta Esposito

ID Number: D50030098

PhD Coordinator:

Prof. Massimiliano Di Penta

Acknowledgements

I would like to thank my supervisors, Professors Andrea Cusano and Marco Pisco, for their constant support and guidance throughout my PhD journey. I am profoundly thankful for your exceptional mentorship, for being exemplary scientific role models, for believing in me, and for making this thesis possible.

I would like to express my deepest gratitude to Dr. Sara Spaziani for transforming this experience into an opportunity for profound professional growth, and for her constant encouragement and support. I am also truly grateful to Drs. Anna Aliberti, Tania Mariastella Caputo, and Angela Maria Cusano for their remarkable human and professional value. I would also like to thank Drs. Sara Ucci and Alessia La Rocca for their kind support and collaboration.

A special thank you goes to my colleagues and dear friends, Federica Piccirillo, Chiara Mulé, and Valeria Iazzetta. Sharing this journey with you in all its facets has been a wonderful experience. I am so grateful for your friendship, your encouragement, and the many moments of shared support that have made this journey truly special.

I am also grateful to the entire Optoelectronics and Photonics Research Group at the University of Sannio for sharing your enthusiasm for research and for constantly inspiring me to grow both professionally and personally. I will never find the right words to thank you enough for everything.

On a more personal note, I would like to express my deepest gratitude to my family, to whom this thesis is dedicated. You mean everything to me — you are my strength and my greatest source of happiness. Without your sacrifices, your unconditional love and support, none of this would have been possible.

List of Publications

Journals:

[1] **A Review of Fluorescent Sensor Development for Ultra-Sensitive Per- and Polyfluoroalkyl compound Analysis in Water: Towards in situ monitoring platforms** Esposito C., Cusano A.M., Caputo T.M., Aliberti A. and Cusano A. *ACS Materials Letters* (2026). <https://doi.org/10.1021/acsmaterialslett.5c01669> (Review)

[2] **Raman Spectroscopy and Artificial Intelligence for Rapid Identification and Classification of Primary Liver Cancer Cells.** Esposito C., Janneh M., Spaziani S., Calcagno V., Bernardi M.L., Iammarino M., Verdone C., Tagliamonte M., Buonaguro L., Pisco M., Aversano L. and Cusano A. *Italian Conference on Optics and Photonics (ICOP)*, Firenze, Italy, pp. 1-4, (2024). doi: 10.1109/ICOP62013.2024.10803665. (Proceedings)

[3] **Artificial Intelligence-assisted Raman Spectroscopy for Liver cancer diagnosis.** Esposito C., Janneh M., Spaziani S., Calcagno V., Bernardi M.L., Iammarino M., Verdone C., Tagliamonte M., Buonaguro L., Pisco M., Aversano L. and Cusano A. *EPJ Web Conf.* Volume 309, 10010, (2024). DOI: 10.1051/epjconf/202430910010. (Conference Paper)

[4] **Assessment of Primary Human Liver Cancer Cells by Artificial Intelligence-Assisted Raman Spectroscopy.** Esposito C., Janneh M., Spaziani S., Calcagno V., Bernardi M.L., Iammarino M., Verdone C., Tagliamonte M., Buonaguro L., Pisco M., Aversano L. and Cusano A. *Cells*; 12(22):2645, (2023). <https://doi.org/10.3390/cells12222645> (Article)

[5] **Using Machine Learning for Classification of Cancer Cells from Raman Spectroscopy.** Aversano, L.; Bernardi, M.L.; Calcagno, V.; Cimitile, M.; Esposito, C.; Iammarino, M.; Pisco, M.; Spaziani, S. and Verdone, C. In *Proceedings of the 3rd International Conference on Deep Learning Theory and Applications - DeLTA*; ISBN 978-989-758-584-5; ISSN 2184-9277, SciTePress, pages 15-24, (2022). DOI: 10.5220/0011142600003277. (Proceedings)

Table of contents

Acknowledgements.....	1
List of Publications	2
Abstract.....	16
Chapter 1: Raman spectroscopy for cancer diagnostics	18
Abstract.....	18
1. Introduction.....	18
2. Raman spectroscopy: Basic Principles and Fundamentals	21
2.1 Raman scattering.....	21
2.2 Interpretation of Raman spectrum.....	26
2.3 Instrumentation	27
3. Fundamentals of Raman Data Analysis.....	28
3.1 Pre-processing of Raman Data.....	28
3.2 From conventional methods to artificial intelligence approaches.....	30
4. Application of Raman spectroscopy in advanced cancer diagnostic.....	34
4.1 Raman spectroscopy <i>in vitro</i>	34
4.2 Raman spectroscopy <i>ex vivo</i>	38
4.3 Raman spectroscopy <i>in vivo</i>	43
5. Challenges in Raman spectroscopy	46
6. Conclusions.....	47
7. References.....	47
Chapter 2: Assessment of Primary Human Liver Cancer Cells by Artificial Intelligence Assisted Raman Spectroscopy	55
Abstract.....	55
1. Introduction.....	55
2. Materials and Methods.....	57
2.1. Sample collection and preparation	57
2.2. Raman Spectroscopy	57
2.3. Raman data pre-processing	58
2.4. Machine learning.....	59
3. Results and Discussion	62
3.1. Morphological analysis	62
3.2. Raman spectroscopy biochemical overview	63
3.3. Dataset.....	65

3.4. Unsupervised Multivariate Analysis	65
3.5. Supervised Multivariate Analysis	66
3.6. CNN-based Classification (Supervised Learning)	70
3.7. Blind prediction of Tumor cells	72
4. Conclusions.....	73
5. References.....	73
Chapter 3: Uncovering biochemical signatures of circulating tumor cells: Raman spectroscopy meets machine learning for non-small cell lung cancer (NSCLC)	76
Abstract.....	76
1. Introduction.....	76
2. Material and Methods	80
2.1 Sample collection and preparation	80
2.2 Raman analysis.....	81
2.3 Raman Data Pre-processing	81
2.4 Machine learning approach: PCA-LDA.....	82
3. Results and Discussion	83
3.1 Morphological Features of NSCLC cell lines and Leukocytes.....	83
3.2 Spectral and biochemical fingerprints of CTC line models and leukocytes	84
3.3 Assessment of heterogeneous biochemical landscapes of CTC line models.....	88
3.4 Cell sample discrimination by PCA-LDA model	90
4. Conclusions.....	95
5. References.....	96
Chapter 4: Surface-enhanced Raman Spectroscopy (SERS) for Advanced Cancer Diagnosis	101
Abstract.....	101
1. Introduction.....	101
2. SERS basic principles.....	102
2.1 Electromagnetic enhancement mechanism	103
2.1.1 SERS Hotspots	105
2.2 Chemical enhancement mechanism	106
3. SERS substrates	106
3.1 Colloidal metallic nanoparticles (NPs)	107
3.2 Metallic nanostructured array	108
3.3 Optical fiber as SERS platform.....	110
3.4 Non-metallic and hybrid nanostructures	110
4. SERS strategies: direct and indirect.....	111

4.1 SERS tags design	112
5. SERS Liquid biopsy and Cancer Bioimaging.....	117
5.1 SERS liquid biopsy	117
5.1.1 Detection of nucleic acids	119
5.1.2 Detection of tumor-exosome vesicles	122
5.1.3 Detection of protein cancer biomarkers	124
5.1.4 Detection of circulating tumor cells (CTCs)	126
5.2 SERS Bioimaging	128
5.2.1 SERS imaging of cancer biomarkers in cells and tissues.....	130
5.2.2 SERS imaging <i>in vivo</i>	133
5.2.3 Multimodal SERS.....	136
6. Optical Fibers (OFs) as innovative SERS platform for cancer diagnostic	138
7. SERS challenges in cancer diagnostics	141
8. Conclusions.....	143
9. References.....	144
Chapter 5: Nanopillar Arrays on Optical Fiber Tip Ends for Surface-enhanced Raman Spectroscopy	154
Abstract.....	154
1. Introduction.....	154
2. Material and methods.....	159
2.1 Fabrication of ZnO thin film	159
2.2 Optical characterization	159
2.3 Fabrication of gold coated ZnO nanopillar onto fiber tip by hydrothermal synthesis	160
2.4 Functionalization of nanostructured SERS optical fiber with BPT	160
2.5 SERS analysis	161
2.6 SERS data pre-processing	161
3. Results and discussion	161
3.1 Characterization of the optical properties of ZnO thin films deposited on optical fibers	161
3.2 SEM of SERS optical fiber architectures.....	163
3.3 Assessment of SERS optical fiber performance under excitation with two distinct laser wavelengths.....	164
3.4 Comparison direct and indirect best configuration for SERS optical fibers.....	170
4. Conclusions.....	171
5. References.....	172

Chapter 6: Nanoscale Analysis of Cancer Cell Membranes and Cell Surface Biomarker Using Tip-Enhanced Raman Spectroscopy (TERS)	176
Abstract	176
1. Introduction	176
2. TERS basic principles	181
3. Materials and Methods	185
3.1 Materials and reagents	185
3.2 Cell culture and preparation	186
3.3 Enzyme-linked immunosorbent assay (ELISA)	186
3.4 Functionalization of AuNP-based SERS Tags	186
3.5 Characterization of SERS tags	187
3.6 TERS setup, measurements and data analysis	187
4. Results and Discussion	188
4.1 TERS label-free investigation of HepG2 cancer cell membrane	188
4.2 TERS for GPC-3 recognition on the cell membrane of HepG2	192
4.2.1 SERS tags profiling	192
4.2.2 Combined TERS and SERS for GPC-3 recognition	196
5. Conclusions	197
6. References	198
Chapter 7: Conclusions	202
7.1 References	206

List of Figures

Chapter 1: Raman spectroscopy for cancer diagnostics

Figure 1. The Raman spectrum provides a unique molecular signature of a cell based on its biochemical constituents, from Ref. [28]	20
Figure 2. Schematic representation of the phenomenon of Rayleigh, anti-Stokes and Stokes Raman scattering, from Ref.[44].	24
Figure 3. CO ₂ vibrational modes and their relative IR and Raman activity, from Ref. [46]	25
Figure 4. Diagram illustrating the key elements of confocal Raman spectroscopy, from Ref. [45]	27
Figure 5. A comprehensive workflow for pre- and post-processing of Raman data, from Ref. [51]	29
Figure 6. Several Raman strategies for in vitro applications. A) Workflow for HER2 protein identification by Raman spectroscopy and Raman imaging, from Ref. [100]. B) Methodological approach combining Raman spectroscopy and Tomographic Phase Microscopy (TPM) to profile and distinguish human colon cancer cells, from Ref. [103]. C)	

Raman analysis of human lung cancer cells (A549) and normal bronchial cells (BEpiC) in both the fingerprint (A and C, respectively) and high-frequency regions (B and D, respectively). The figure also presents differential spectra (A549 vs. BEpiC) for distinct cellular structures—nucleus, mitochondria, endoplasmic reticulum, lipid droplets, cytoplasm, and cell membrane—in both fingerprint and high wavelength regions (E and F, respectively), highlighting the biochemical differences between cancerous and healthy cells, from Ref. [104]. 36

Figure 7. Various Raman spectroscopy strategies employed for ex vivo use. A) Graphical representation of the histopathological diagnosis approach for liver cancer utilizing Raman spectroscopy and a convolutional neural network model, from Ref. [108]. B) Workflow of the combination of Raman spectroscopy with a machine learning approach for colorectal tumor margin assessment, from Ref. [109]. C) Graphical illustration of the process for generating cancer prediction maps from glioblastoma specimens imaged with Raman spectroscopy, from Ref. [110]. D) Raman images comparing cancerous and adjacent normal tissue at specific wavelengths (750, 1585 and 1342 cm^{-1}), alongside optical microscope images, from Ref. [111]. 40

Figure 8. Different in vivo Raman approaches for enhanced cancer diagnosis. A) Workflow outlining the process of brain tumor identification using the Sentry System, from Ref. [114]. B) Diagram of in vivo mouse model analysis combining histology, Raman spectroscopy, and AI-based data interpretation, from Ref. [116]. C) Study overview diagram of the analysis of head and neck tumor biomolecular changes after radiation therapy by Raman spectroscopy an in vivo mouse model, from Ref. [117]. 44

Chapter 2: Assessment of Primary Human Liver Cancer Cells by Artificial Intelligence Assisted Raman Spectroscopy

Figure 1. a) Raman spectra preprocessing steps. Architecture of the classifier based on b) LDA and Tuned-LDA model, c) PCA-LDA model, and d) Convolutional and Recurrent Neural Networks trained using sliding window on Raman Spectra. e) Schematic of the machine learning models used for the blind predictions. 59

Figure 2. Bright-field microscope images of uncultured human Non-Tumor a-c) and Tumor cells d-f) fixed on CaF_2 slides (100 $\times$ magnification; scale bar 2 μm). g) Dimensional analysis of all Non-Tumor and Tumor cells. 62

Figure 3. Top: Averaged Raman spectra of Tumor (red line) and Non-Tumor (black line, offset 10%) cells in the FP region. Solid lines show the average over 180 spectra $\pm$ standard deviation (shaded areas). Bottom: Difference between averaged Raman spectra (black line). The horizontal solid line corresponds to 0 intensity, the two horizontal dashed lines correspond to a threshold of ± 0.025 intensity. Highlighted in yellow, red, and blue are the Raman bands associated with lipids, nucleic acids and proteins, respectively. In orange, the Raman peaks associated with lipids/proteins are shown. 63

Figure 4. PCA of the Raman spectra. PCA 2D data plot distribution of spectra of uncultured Tumor and Non-Tumor cells based on the first 3 PCs components: a) PC1 vs PC2 and b) PC1 vs PC3. The ellipses account for a confidential level of 95% of the data. c) PCA plot associated to the spectra of five uncultured and cultured Tumor cells. The 2D plot is based on the first 2 PCs components: PC1 vs PC2. 65

Figure 5. Loading Plot of principal component 1 (PC1).	66
Figure 6. Learning curve as function of the number of training examples for: a) LDA model and b) LDA model after Hyper-parameters tune optimization methods.	67
Figure 7. Leave-one-out-cross- validation (LOOCV) as a function of the number of PCs.	69
Figure 8. The loss function trend over epochs, during training, for CNN-LSTM models of 22 and 56 layers with different learning rates.....	71

Chapter 3: Uncovering biochemical signatures of circulating tumor cells: Raman spectroscopy meets machine learning for non-small cell lung cancer (NSCLC)

Figure 1. Schematic representation of the Raman data processing workflow. This flowchart outlines the main steps, starting with pre-processing of the data (spectral calibration, spike removal, linear interpolation, baseline correction, normalization and detection of outliers with Hotelling's T2 and Q residual tests), followed by classification with an optimized PCA-LDA model.	82
Figure 2. Bright-field images of non-small cell lung cancer cell lines, including A549 (a), H1299 (b), H1650 (c), H1915 (d) and normal cells, leukocytes (e), in suspension. (60 ×, scale bar 4 μm).....	83
Figure 3. Normalized average Raman spectra of suspended NSCLC lung cancer cell lines (A549, H1299, H1650 and H1915) and leukocytes, each highlighted with different colors. The bands highlighted in red, blue, and yellow correspond to Raman peaks of proteins, nucleic acids, and lipids, respectively, as indicated in the legend. Combinations of these colors reflect non-exclusive assignments, indicating a mixture of biomolecules. P = proteins, NA = nucleic acids, L = lipids, NA/P = nucleic acids/proteins, P/L = proteins/lipids. The Raman spectra have been spatially offset for clarity and the shaded areas indicate the standard deviation for each spectrum.	84
Figure 4. Comparative analysis based on the difference between the normalized average spectrum of leukocytes and that of individual non-small cell lung cancer (NSCLC) cell lines. Spectral differences are presented comparing leukocytes to suspended A549 (a), H1299 (b), H1650 (c), and H1915 (d), respectively. The bands highlighted in red, blue, and yellow correspond to Raman peaks of proteins, nucleic acids, and lipids, respectively, as indicated in the legend. Combinations of these colors reflect non-exclusive assignments, indicating a mixture of biomolecules. P = proteins, NA = nucleic acids, L = lipids, NA/P = nucleic acids/proteins, P/L = proteins/lipids.	88
Figure 5. Subtraction analysis between the normalized average Raman spectra of A549 and H1299 (a), A549 and H1650 (b), H1299 and H1650 (c), A549 and H1915 (d), H1915 and H1299 (e), and H1915 and H1650 (f). The bands highlighted in red, blue, yellow and green correspond to Raman peaks of proteins, nucleic acids, lipids and carbohydrates, respectively, as indicated in the legend. Combinations of these colors reflect non-exclusive assignments, indicating a mixture of biomolecules. P = proteins, NA = nucleic acids, L = lipids, C= carbohydrates, NA/P = nucleic acids/proteins, P/L = proteins/lipids, P/L/C= proteins/lipids/carbohydrates.	89

Figure 6. Evaluation of the number of latent variables to include in the partial least squares model based on two metrics: root mean square error of calibration (RMSEC) and root mean square error of cross-validation (RMSECV) for five cell samples: Leukocytes (a), A549 (b), H1299 (c), H1650 (d), H1915 (e).	91
Figure 7. Outlier detection and removal by plotting Q residuals against Hotelling T^2 with a 95% confidence interval for a PLS model built with 5 latent variables for each cell line: Leukocytes (a), A549 (b), H1299 (c), H1650 (d), H1915 (e). In the plot, each outlier is marked by a numerical label, with the number preceding the dot corresponding to the cell and the number following the dot corresponding to the spectrum number. The percentage of variance explained by the first 5 latent components used to build the model is indicated on the abscissa.	92
Figure 8. PCA 2D plot showing the distribution of spectral data from leukocytes and four suspended non-small cell lung cancer cell lines (A549, H1299, H1650, and H1915) based on the first two principal components (a). The ellipses represent a 95% confidence interval for the data. Loading plot of PC1 and PC2 components and the corresponding Raman band assignments was generated (b). The bands highlighted in red, blue, yellow and green correspond to Raman peaks of proteins, nucleic acids and lipids respectively, as indicated in the legend. Combinations of these colors reflect non-exclusive assignments, indicating a mixture of biomolecules. P = proteins, NA = nucleic acids, L = lipids. The loading plots of PC1 and PC2 are spatially offset for clarity.	93
Figure 9. Three-dimensional scatter plot illustrating the LDA results for leukocytes and four lung cancer cell lines—A549, H1299, H1650, and H1915—represented by red, green, fuchsia, orange, and purple circles, respectively. The plot displays the separation of the five classes based on the first three linear discriminant components (LD1, LD2, and LD3), along with the percentage of variance each component explains. LDA was performed using the first 50 principal components as input.	94
Figure 10. Confusion matrix (a) and classification metrics (b) showing the results of the proposed PCA-LDA model when applied to the test dataset.	95

Chapter 4: Surface-enhanced Raman Spectroscopy (SERS) for Advanced Cancer Diagnosis

Figure 1. The diagram shows the basic phenomena underlying Surface Enhanced Raman Spectroscopy, including a) normal Raman scattering, b) localized surface plasmon resonance (LSPR), and c) the electromagnetic enhancement process in SERS, which involves two-step amplification, from Ref. [27].	104
Figure 2. Illustration of various geometrically distinct gold nanostructures used as SERS substrates, from Ref. [62].	108
Figure 3. Illustration of various methods for fabricating SERS substrates: (1) covalent attachment of nanoparticles to a solid substrate, from Ref. [81], (2) nanosphere lithography, from Ref. [84], (3) electron beam lithography, from Ref. [93].	110
Figure 4. The schematic illustrates two major SERS detection strategies: direct (label-free) and indirect (label-based) approaches. A) Direct SERS detects target molecule interacting with a plasmonic metal surface, while B) Indirect SERS detects target molecules through a	

functionalized SERS tag attached to a plasmonic metal nanoparticle with a Raman reporter molecule. From Ref. [116].	112
Figure 5. Illustration of widely used Raman reporters, from Ref. [120].	114
Figure 6. Overview of A) the different types of targeting agents and of B) the different coupling strategies used to conjugate the targeting agents to the surfaces of the nanotags. From Ref. [115].	115
Figure 7. Schematic representation of SERS liquid biopsy. a) SERS substrates enable molecular fingerprinting for direct and indirect SERS-based applications, including b) the detection of different types of circulating cancer biomarkers (DNA/RNA, EVs, proteins and CTCs) in liquid biopsy for detailed tumor profiling, from Ref. [165]. EVs: exosome vesicles, CTCs: circulating tumor cells.	118
Figure 8. Schematic representation of the different SERS detection strategies for nucleic acid biomarkers. (1) Hierarchical 3D interface arrangement of Au nanocage@Au together with IS-AgMNPs for the detection of miR-21 and miR-31. Illustration of the workflow for (a) the SERS probe and (b) the development of the acquisition unit. (c) The workflow for the detection of two CRC-associated miRNAs, from Ref. [170]. (2) SERS platform for circRNA detection in lung cancer. (a) Preparation of Fe ₃ O ₄ @Ag@DTNB magnetic nanoparticles and of (b) Au@MBA@Ag nanoparticles, (c) SERS detection of circRNA, from Ref. [160].	121
Figure 9. SERS strategies for the detection of cancer-related EVs. (1) PMMA nanocavities array SERS chip (PNAS-chip) for EVs detection in breast cancer. a) The workflow and b-c) the development and characterization of the PNAS-chip, from Ref. [183]; (2) SERS Microfluidic Biosensor for glycan profiling of EVs. a) Glycan profiles on sEVs derived from benign and lung cancer patients, (b) CT and PET images of benign and malignant lung nodules, (c) operating principle of EV-GLYPH assay, (d) alternating current electrohydrodynamic field induces nanomixing fluid flow on the electrode surface, (e) SERS maps of the different signals of three nanotags associated with the expression of LacdiNAc, T antigen and CD81, from Ref. [185].	123
Figure 10. Different SERS approaches for the detection of cancer-related protein biomarkers. (1) Principle of the SERS-LFS strategy for the evaluation of MWA outcomes in HCC patients and characterization of the SERS probe. (a) Collection of blood samples from HCC patients after MWA, (b) SERS immune sensor mechanism – (i) formation of SERS probe/target/antibody complexes on test and control lines. (ii) immune sensor structure enables simultaneous quantification of CCL20 and EGF proteins on two test lines (TL1 and TL2), (c) TEM of SERS nanoparticles, (d-e) optical analysis of SERS substrate and surface charge analysis, respectively, from Ref. [189]; (2) SERS-based microfluidic aptamer sensor for simultaneous detection of two prostate cancer biomarkers. (b) functionalization strategies of SERS nanotags, (c) formation of sandwich immunocomplexes on the multifunctional nanocone array (MNA) platform, from Ref. [191].	125
Figure 11. SERS detection methods for circulating tumor cells. (1) Schematic representation of a dual-modal SERS bioprobe and machine learning approach for CTC detection based on folate receptor recognition, from Ref. [130]; (2) Diagram of SERS-based microfluidic liquid biopsy of prostate cancer. (a) Overview of NFTS fabrication; (b) Step-by-step preparation of two distinct antibody-functionalized SERS nanotags; (c) Functionalization of NFTS with	

capture aptamers and subsequent CTC binding; (d) Integrated workflow on the CMD for multistage CTC capture and in situ single-cell phenotyping via SERS analysis, from Ref. [196]..... 127

Figure 12. (a) Multiplex SERS imaging for analyzing cells and tissues, application of SERS imaging for: (b) different cell samples analysis and tracking therapeutic outcomes and (c) tissue samples investigation for diagnosis and guiding surgical procedures, from Ref. [143]. 129

Figure 13. SERS bioimaging of individual and multiple cancer biomarkers at the cell and tissue levels. (1) Schematic representation of SERS detection of intracellular circSATB2. (a) circSATB2 and probe DNA, (b) SERS nanoprobe design, (c) detection of circSATB2 in living cells, from Ref. [200]; (2) Schematic diagram of a core-shell SERS nanosensor for simultaneous imaging of EGFR on cell membranes and ROS secretion from living cells. (a-b) Schematic representation of the fabrication process of AuNPs@4-MBN@Au@Aptamer/2-MHQ nanosensors and their application in detecting EGFR and ROS on cell membranes, from Ref. [202]; (3) Schematic illustration of a combined morphological and spectroscopic diagnostic method using label-free SERS with a deep learning model to detect HER2 expression in breast cancer tissue from Ref. [203]; (4) Schematic diagram of the SERS-based spectro-histology for distinguishing the clinically relevant biomarkers ER, PR, and HER2, from Ref. [204]..... 131

Figure 14. SERS imaging in vivo. (1) SERS imaging to identify breast cancer phenotypes and evaluate treatment efficacy. (a) Diagram illustrating the synthesis of three types of Raman reporter-adsorbed Au-Ag core-shell nanoparticles and the attachment of PEGylated antibodies to their surface, (b) SERS imaging to detect phenotypic biomarkers on cell membranes, (c) SERS-based evaluation of treatment effects after antitumor drug therapy and surgery, from Ref. [213]; (3) Identification of sentinel lymph nodes in preclinical non-human primate models using ultra-bright NIR-II SERS nanotags, (3.1) Identification of SLN in rhesus monkeys using NIR-II tags, from Ref. [214]..... 134

Figure 15. Different multimodal SERS approaches for cancer diagnostics. (1) Schematic of the FL-SERS biosensor for the detection and visualization of MUC1. (a) Preparation of QD-PDDA-Apt complex and (b) AuNS-cDNA-PEG, (c) Detection and imaging of MUC1 in living cells using FL-SERS biosensor, from Ref. [220]; (2) Bioconjugated SERS nanoparticles for maximal targeting efficiency in breast cancer cell lines. (a) Imaging (brightfield, fluorescence and Raman) showed strong signals in SKBR3 cells incubated with HER2-targeted NPs, (b) quantitative analysis revealed significantly higher intensity of fluorescence and Raman signals in HER2-expressing SKBR3 cells compared to IgG controls, from Ref. [221]; (3) Design of dual-modal SERS-MRI bioprobes for sensitive and specific detection of PD-L1 expression in tumors, from Ref. [222]; (4) Illustration of the synthesis and in vivo evaluation of TetrAuCS for SERS and PA imaging, from Ref. [223]..... 137

Figure 16. SERS-based optical fiber approaches for cancer diagnostics. (1) Schematic illustration of an on-chip SERS assay and an optrode SERS biosensor for the detection of Thyroglobulin (Tg). (a) Functionalization steps in the SERS-based sandwich immunoassay strategy based on bioengineered SERS substrate with capture antibodies to specifically detect Tg molecules, BSA blocking step to prevent non-specific binding and (b)

functionalized gold nanoparticles (f-AuNPs) with a Raman reporter (4-MBA) and detection antibodies to ensure precise Tg recognition, from Ref. [100]; (2) Visual representation detailing the functional mechanism of the fiber-exosome-SERS probe sandwich immunoassay, from Ref. [101]; (3) SERS-enabled fiber tip for detecting intracellular and extracellular pH in living single cells. (A) Comparison between intra- and extracellular pH values in LO2 and HepG2 cells, showing (B) measurement points and (C) corresponding pH values in a histogram, from Ref. [102]; (4) Diagram of fiber-optic-based Raman imaging device integrated with white-light endoscopy and SERS nanoprobe for multiple cancer biomarker detection, from Ref. [103]. 140

Chapter 5: Nanopillar Arrays on Optical Fiber Tip Ends for Surface-enhanced Raman Spectroscopy

Figure 1. The lab-on-fiber vision, from Ref. [6].	155
Figure 2. Indirect interrogation of a SERS optical fiber and nanotechnologies fabrication methods of SERS-active substrates onto a tip of an optical fiber. In an optrode SERS probe, light is guided along the fiber to stimulate a nanostructured tip, with the enhanced Raman signal collected back through the same fiber for analysis (upper part). These nanostructured tips can be created using techniques ranging from basic surface roughening and nanoparticle deposition to advanced lithography and etching methods (lower part). Re-adapted from Ref. [9].	156
Figure 3. Schematic illustration of optical fiber tip fabrication, showing nanostructures formed directly on the tips via top-down or bottom-up methods, or transferred from a flat substrate onto the fiber's end facet, from Ref. [19].	157
Figure 4. ZnO film thickness measurement with profilometer.	162
Figure 5. a) Refractive index characterization of fused silica and ceramic fiber, with b) extinction coefficient analysis of ZnO thin films.	162
Figure 6. Reflection spectroscopy analysis of ZnO thin film on planar and fiber substrates.	163
Figure 7. SEM micrograph of the unmasked sample	163
Figure 8. SEM micrograph of a masked fiber sample.	164
Figure 9. a) Average SERS spectra of BPT obtained from SERS map ($100 \times 100 \mu\text{m}^2$) collected in direct configuration at 785 nm from each SERS-active substrates onto the tip end of unmasked and masked samples. Vertical offsets were applied to the SERS spectra for clearer visualization. b) Statistical distribution of BPT SERS intensities at 1585 cm^{-1} for all fiber types, with error bars representing the standard deviation. Each sample is represented by a distinct color for unambiguous identification.	165
Figure 10. a) Average SERS spectra of BPT obtained from SERS map ($100 \times 100 \mu\text{m}^2$) collected in direct configuration at 633 nm from each SERS-active substrates onto the tip end of all unmasked and masked samples. Vertical offsets were applied to the SERS spectra for clearer visualization. b) Statistical distribution of BPT SERS intensities at 1585 cm^{-1} for all fiber types, with error bars representing the standard deviation. Each sample is represented by a distinct color for unambiguous identification.	166
Figure 11. a) Averaged SERS spectra of BPT, obtained from three single-point measurements acquired in the indirect configuration at 785 nm, are shown for each	

unmasked and masked SERS-active substrate integrated at the fiber tip. Vertical offsets have been applied to the spectra to aid visual clarity. b) Statistical analysis of SERS signal intensities at 1590 cm^{-1} (BPT) and at 800 cm^{-1} (silica reference band) across all fiber types. Error bars represent standard deviations, while distinct colors are used to clearly differentiate between the sample groups.	168
Figure 12. Evaluation of the silica interference on the SERS performance of all SERS fibers. a) Distribution of SERS intensities of BPT and silica at 1590 (black square) and 800 cm^{-1} (red square) at 785 nm and b) 633 nm excitation wavelength. Error bar represent standard deviations	169
Figure 13. a) Averaged SERS spectra of BPT, obtained from three single-point measurements acquired in the indirect configuration at 633 nm, for each unmasked and masked. Vertical offsets have been applied to the spectra to aid visual clarity. b) Statistical analysis of SERS signal intensities at 1590 cm^{-1} (BPT) and at 800 cm^{-1} (silica reference band) across all fiber types. Error bars represent standard deviations, while distinct colors are used to clearly differentiate between the sample groups.	169
Figure 14. Comparison of BPT SERS intensities for various SERS optical fibers under direct and indirect interrogation. a) Plot of intensities distribution at 1585 cm^{-1} at 785 nm excitation in indirect and direct configuration. b) Plot of intensities distribution at 1585 cm^{-1} in indirect and direct configuration at 633 nm. Error bars represent the standard deviations.	171

Chapter 6: Nanoscale Analysis of Cancer Cell Membranes and Cell Surface Biomarker Using Tip-Enhanced Raman Spectroscopy (TERS)

Figure 1. Schematic illustration of the TERS approach, which combines the high sensitivity of SERS—enabled by the metallic tip that amplifies the Raman fingerprints of molecules beneath it—with the high spatial resolution provided by scanning probe microscopy, from Ref. [12].	177
Figure 2. Overview of TERS application, from Ref. [17].	178
Figure 3. Common metal tip geometries for TERS setups include spherical, rod-shaped, truncated, and tapered tips. From Ref. [17].	181
Figure 4. TERS principles. Schematic representation of the enhancement of the optical field at the apex of a sharp metallic tip when illuminated by a laser source in TERS setup (a), from Ref.[45]. Illustration of the lightning rod effect, where a sharp tip concentrates the electric field at its apex, particularly when the incident radiation is polarized along the tip axis, producing a strong local field enhancement (b). E: electric field, k: wave vector of the excitation laser. From Ref. [16].	182
Figure 5. Diagram of the typical TERS setup in side-illumination configuration (a). AFM feedback system in the AFM-TERS system (b). Re-adapted from Ref. [15].	183
Figure 6. Schematic of the various optical configurations used in tip-enhanced Raman spectroscopy. Re-adapted from Ref [48].	184
Figure 7. Schematic illustration of gap-mode TERS, Re-adapted from Ref. [49].	185
Figure 8. Graphical representation of the gold nanoparticle functionalization protocol for GPC3 detection.	187

Figure 9. Atomic force microscopy (AFM) magnitude (Mag) image of HepG2 cells ($40 \times 40 \mu\text{m}$). Cells were scanned in air at room temperature. Scale bar: $10 \mu\text{m}$. The vertical color scale, ranging from dark brown to light yellow, represents regions with progressively stronger to weaker interactions between the sample and the TERS tip. 189

Figure 10. TERS spectra were acquired from five distinct locations on the plasma membrane surface HepG2 cells indicated in the a) and c) Mag AFM images, respectively. AFM Mag images Scale bar: $10 \mu\text{m}$. The vertical color scale, ranging from dark brown to light yellow, represents regions with progressively stronger to weaker interactions between the sample and the TERS tip. The tip positions for cell 1 (a) and cell 2 (c) are marked with colored crosses, each color corresponding to its respective spectrum. The complete sets of spectra for cell 1 (b) and cell 2 (d), together with reference spectra collected outside the cells and at a distance from the membrane, are shown. Characteristic band positions identified in each spectrum are indicated, and the spectra have been vertically offset for clarity. 190

Figure 11. AFM-TERS map of HepG2 Cell 2. a) AFM Mag image of HepG2 cell 2. Scale bar: $10 \mu\text{m}$. The vertical color scale, ranging from dark brown to light yellow, represents regions with progressively stronger to weaker interactions between the sample and the TERS tip. The blue square outlines the $5 \times 5 \mu\text{m}$ region of the cell selected for scanning. b) The corresponding pre-processed TERS map provides simultaneous spatial and biochemical information. The TERS image is generated by mapping the intensity of the Raman band of phenylalanine at 1000 cm^{-1} and the high-wavenumber region ($2800\text{--}3200 \text{ cm}^{-1}$), represented by red and green pixels, respectively. These spectral features correspond to proteins in the case of the phenylalanine band, and to lipids along with protein contributions in the high-wavenumber region. This combined mapping highlights the heterogeneous organization of biomolecules across the cell membrane. 191

Figure 12. Enzyme-linked immunosorbent assay (ELISA) results obtained for GPC-3 concentrations of 0, 10, 50, and 500 ng/mL . Data are expressed as relative fluorescence units (RFU) on the y-axis, while GPC-3 concentration is plotted on the x-axis. The negative control (0 ng/mL) corresponds to phosphate-buffered saline (PBS). 193

Figure 13. Representative SERS spectrum of gold nanoparticles functionalized with 4-MBA and anti-GPC3 antibodies, recorded with a 633 nm excitation laser prior to cell incubation. 194

Figure 14. Bright-field micrograph of HepG2 cells after incubation with gold nanoparticles functionalized with the Raman reporter 4-MBA, an anti-GPC3 antibody for specific targeting, and bovine serum albumin to stabilize the plasmonic nanostructures and prevent nonspecific interactions (a). Image acquired at $100\times$ magnification; scale bar: $2 \mu\text{m}$. Representative SERS spectrum of SERS tags incubated with HepG2 cells (b). The yellow star indicates the specific point where the spectrum was recorded. 195

Figure 15. AFM magnitude image of fixed HepG2 cells is shown in (a). Scale bar: $10 \mu\text{m}$. The vertical color scale, ranging from dark brown to light yellow, represents regions with progressively stronger to weaker interactions between the sample and the TERS tip. TERS spectra were collected at the specific positions marked with colored crosses on the AFM map in (b). 196

List of Tables

Chapter 2: Assessment of Primary Human Liver Cancer Cells by Artificial Intelligence Assisted Raman Spectroscopy

Table 1. Assignment of characteristic Raman bands of Tumor and Non-Tumor cells.....	64
Table 2. Confusion matrix of the LDA model.	67
Table 3. Confusion matrix of the Hyper- parameter tuned LDA models.	68
Table 4. Optimized LDA classification metrics.....	68
Table 5. List of the first 35 PC component and the relative percentages of variance.....	69
Table 6. Training and test results for the CNN-based classifier with different Hyper-parameters, reject options and augmentation strategies.....	71
Table 7. Confusion matrix of the best tuned LSTM-CNN model (last row of Table 3)....	72
Table 8. Classification of samples with different ratios of Tumor and Non-Tumor cell.....	72

Chapter 3: Uncovering biochemical signatures of circulating tumor cells: Raman spectroscopy meets machine learning for non-small cell lung cancer (NSCLC)

Table 1. Assignment of key Raman bands identified in the cell samples under investigation.	85
--	----

Chapter 5: Nanopillar Arrays on Optical Fiber Tip Ends for Surface-enhanced Raman Spectroscopy

Table 1. The average intensity values and corresponding standard deviations of the BPT Raman band at 1585 cm^{-1} were reported for each SERS optical fiber in direct configuration using a 785 nm excitation wavelength.....	166
Table 2. The average intensity values and corresponding standard deviations of the BPT Raman band at 1585 cm^{-1} were reported for each SERS optical fiber in direct configuration using a 633 nm excitation wavelength.....	167
Table 3. The average intensity values and corresponding standard deviations of the BPT Raman band at 1590 cm^{-1} were reported for each SERS optical fiber in indirect configuration using a 785 nm excitation wavelength.....	168
Table 4. The average intensity values and corresponding standard deviations of the BPT Raman band at 1590 cm^{-1} were reported for each SERS optical fiber in indirect configuration using a 633 nm excitation wavelength.....	169
Table 5. Summary of best-performing SERS optical fiber for each configuration and excitation wavelengths.....	172

Chapter 6: Nanoscale Analysis of Cancer Cell Membranes and Cell Surface Biomarker Using Tip-Enhanced Raman Spectroscopy (TERS)

Table 1. TERS band assignments for TERS spectrum recorded from the two HepG2 cells are reported.	190
Table 2. Dynamic light scattering analysis of naked and functionalized nanoparticles. ..	194

Abstract

Cancer is a global health burden, highlighting the limitations of traditional diagnostic approaches, which are often invasive, subjective, and limited in their capacity for timely detection. Addressing these challenges through the development of innovative diagnostic tools that enable accurate, non-invasive and objective assessment is crucial for advancing early diagnosis, optimizing treatment strategies, and improving patient outcomes. This thesis explores the transformative potential of advanced Raman spectroscopy techniques as innovative technologies capable of revolutionizing cancer diagnostics, emphasizing their ability to provide unique biochemical fingerprinting in a non-invasive and real-time manner. Raman spectroscopy is introduced as a label-free technique, with its fundamental principles, instrumentation, and essential data pre-processing steps described in detail. The technique is widely used for the analysis of biological samples—from cells to tissues to biofluids—due to its ability to provide detailed molecular and spectral fingerprints, revealing biochemical composition and enabling deeper understanding of biological systems. Owing to these strengths, Raman spectroscopy can identify subtle biochemical alterations that differentiate healthy cells from cancerous ones, while simultaneously providing insights into the distinct metabolic and genetic profiles characteristic of cancer cells. The discussion also emphasizes the critical role of advanced machine learning algorithms—including both unsupervised and supervised approaches—in interpreting complex spectral data for accurate disease classification. Additionally, the principal applications of Raman spectroscopy in cancer diagnostics are presented across *in vitro*, *ex vivo*, and *in vivo* settings.

The initial part of the research focuses on the application of artificial intelligence-enhanced Raman spectroscopy in two distinct cancer case studies—hepatocellular carcinoma and non-small cell lung cancer—highlighting the broad applicability and versatility of this combined diagnostic strategy. The first study assesses primary human non-tumor and tumor liver cells from a patient with hepatocellular carcinoma using Raman analysis, which successfully differentiates these cellular samples by revealing distinct molecular fingerprints that reflect altered nucleic acid content in cancer cells. For tumor cell identification and discrimination, different linear discriminant analysis (LDA)- and convolutional neural network–long short-term memory (CNN-LSTM)-based models were developed. The LDA-based models achieved classification accuracies close to 90%, while the best CNN-LSTM model reached about 93% at the single-spectrum level.

Raman spectroscopy is further applied in the context of the liquid biopsy, using different suspended non-small cell lung cancer (NSCLC)-derived cell lines as models of circulating tumor cells (CTCs). This methodological approach provides a powerful platform to simulate CTC dissemination in the bloodstream and to reliably distinguish them from healthy leukocytes. Spectral analysis reveals lipid enrichment in cancer cells, while healthy cells exhibit higher levels of nucleic acids and proteins, reflecting their distinct cellular origins and functions. Beyond differentiating malignant from healthy cells, the research provides novel insights into the heterogeneity, and the metabolic and genetic landscapes of different cancer cell lines. By integrating Raman spectroscopy with a PCA-LDA algorithm, the study achieves accurate discrimination between four NSCLC cell lines and leukocytes, reaching a

classification accuracy of 96.40%, highlighting its significant potential in supporting lung cancer diagnostics.

Expanding on diagnostic innovations, the thesis also presents a comprehensive analysis of surface-enhanced Raman spectroscopy (SERS), detailing its electromagnetic and chemical enhancement mechanisms, as well as the strategic design of SERS substrates and functionalized nanotags to optimize sensitivity and specificity. The study further explores SERS applications in biosensing and bioimaging, enabling precise identification of biomarkers in cells, tissues, and even *in vivo*, with particular emphasis on the use of SERS optical fibers as an innovative platform for healthcare applications. In this context, a novel aspect of the research involves the development and evaluation of Au-coated ZnO nanopillar arrays integrated onto optical fiber tips as SERS-active platforms. The study systematically examines how fabrication strategies, structural ordering, and excitation wavelengths affect SERS performance, aiming to optimize sensitivity and reproducibility. These platforms are investigated for their potential in real-time, minimally invasive biomedical applications, highlighting their promise for broader diagnostic use. The research also extends to the application of Tip-enhanced Raman spectroscopy (TERS), outlining its fundamental principles and diagnostic applications. Specifically, a TERS as proof-of-concept study is conducted for nanoscale analysis of cancer cell membranes and, in combination with SERS, for cell surface biomarkers, providing simultaneous biochemical and topographical information. The research demonstrates label-free TERS for characterizing HepG2 hepatocellular carcinoma cell membranes, highlighting their heterogeneity. A hybrid strategy combining TERS with SERS tags is presented for the targeted detection of glypican-3 (GPC-3), a key HCC biomarker, on cell membranes. Functionalized SERS tags specifically bind to overexpressed GPC-3, and gap-mode TERS amplifies the signals, enabling precise nanometric localization of the biomarker.

In conclusion, this thesis underscores the pivotal role of advanced Raman platforms as powerful and reliable diagnostic tools, marking a significant step forward in cancer diagnostics. Their integration with artificial intelligence and nanotechnology represents a key advancement, opening the way to real-time, non-invasive, and highly personalized diagnostic strategies. These developments not only enhance the translational relevance of Raman-based technologies but also lay the foundation for their future implementation in clinical practice, where they hold great promise for improving patient outcomes and advancing the vision of precision oncology.

Chapter 1: Raman spectroscopy for cancer diagnostics

Abstract

Cancer remains one of the leading causes of mortality worldwide, a challenge exacerbated by the shortcomings of conventional diagnostic approaches, which often lack sufficient sensitivity for early detection, lengthy analysis times, variability in result interpretation, and inherent patient risks. In this context, Raman spectroscopy (RS) has emerged as a powerful, non-invasive, and label-free optical diagnostic modality capable of capturing the unique molecular “fingerprints” of biological samples (e.g., cells, tissues) through the Raman effect, which arises from inelastic light scattering by matter. This technique enables detailed characterization of the biochemical composition of cells and tissues, including lipids, proteins, and nucleic acids. To effectively analyze and interpret the complex, high-dimensional data generated by RS, advanced computational approaches—particularly machine learning (ML) and deep learning (DL)—are increasingly utilized, thereby improving diagnostic accuracy and enhancing the objectivity of data interpretation. This chapter provides an overview of the fundamental principles of Raman spectroscopy, examines ML-driven strategies for processing and classifying Raman data, and discusses the major applications of RS in cancer diagnostics across *in vitro*, *ex vivo*, and *in vivo* settings. Ultimately, Raman spectroscopy has the potential to transform precision oncology by integrating molecular-level biochemical insights with clinical decision-making, thereby enabling early cancer detection, objective diagnostics, and personalized patient management.

1. Introduction

Cancer refers to a range of diseases that occur when abnormal cells begin to grow uncontrollably and gain the ability to infiltrate nearby tissues or metastasize to distant parts of the body via the bloodstream or lymphatic system [1]. Cancer remains one of the leading causes of mortality worldwide, with over two million new cases and more than 600,000 deaths expected in the United States alone in 2025 [2]. In the same year, the number of cancer deaths in Europe is expected to be 1.28 million, with age-standardized rates falling to 120.9 per 100,000 in men and 79.1 in women [3]. Conventional cancer diagnosis generally relies on advanced imaging technologies, molecular diagnostic tools and histologic examination by tissue biopsy [4],[5]. Imaging techniques such as X-ray, computed tomography (CT), magnetic resonance imaging (MRI) and positron emission tomography (PET) are crucial in cancer care for detecting tumors, guiding biopsies, assessing disease spread and monitoring treatment response. X-ray imaging uses electromagnetic radiation to generate images internal structures of the body. It is especially useful for highlighting dense tissues, such as bones and certain tumors by measuring how these tissues absorb different amounts of X-rays [6]. CT captures X-rays from various angles and uses computer processing to generate detailed cross-sectional images that allow for quick and accurate detection of abnormalities. However, X-ray and CT are limited by exposure to ionizing

radiation [7]. MRI uses magnetic fields and radiofrequency waves to produce detailed grayscale images of body tissue without exposing patients to ionizing radiation. It excels in producing high-resolution images with excellent soft tissue contrast and allows images to be taken in several planes. However, MRI procedures typically take longer than CT scans, which can limit the number of patients examined and increase the overall cost [8]. Ultrasound is a safe, radiation-free imaging modality that provides high-resolution, real-time views of abnormal tissue and is useful for guiding interventions, although its accuracy depends on operator skill and patient anatomy [9]. PET can identify cancer by detecting biochemical and molecular activities within tumor tissues, enabling earlier diagnosis than the anatomical changes visible on CT or MRI scans [10]. PET-CT is particularly important for detecting the spread of metastases and assessing the response to treatment [5]. Furthermore, radiologic imaging techniques that use high-energy X-rays, such as PET, CT, etc., expose patients to radiation risks, potentially causing cancer, genetic mutations, developmental problems, and side effects like hair loss and skin damage [7], [11]. Although current imaging techniques are essential for cancer diagnosis, they face several challenges, including limited sensitivity in detecting early-stage or small tumors, difficulty in accurately distinguishing between benign and malignant lesions, and the risk of false positive or negative results. Additionally, techniques such as X-ray and CT expose patients to ionizing radiation, while ultrasound and MRI are either highly dependent on operator expertise or associated with high costs and limited accessibility [8]. Other key challenges in cancer imaging involve standardizing protocols, integrating different imaging systems, managing large and complex data sets and ensuring safe, reliable and ethical use of AI in clinical decision-making [12]. These limitations may affect diagnostic accuracy and impact timely treatment planning.

While tissue biopsy is currently the most reliable method for diagnosing tumors and determining cancer subtypes, it is invasive and often difficult to perform, offering only a small snapshot of the overall disease [13]. Limited tissue samples and inconsistencies in immunohistochemical tests can lead to diagnostic errors and less effective treatment decisions. Moreover, repeated biopsies for ongoing surveillance can be physically uncomfortable for patients [14]. Because early-stage tumors are often hard to detect, biopsies may also be insufficient to identify cancer at its earliest, most treatable stages [13].

Molecular diagnostics include methods that analyze genetic, transcriptomic, proteomic and epigenetic alterations and help to categorize cancers, predict a disease progression and track the efficacy of treatments [15], [16]. These less invasive approaches such as liquid biopsy and molecular techniques (immunoassays, immunohistochemistry (IHC), DNA sequencing) have significantly improved the detection of tumor biomarkers in body fluids, tissues and cells, providing critical insights into genetic mutations and protein expression. These biomarkers serve as measurable indicators of normal biological functions or abnormal processes in cancer disease, enabling earlier diagnosis, more accurate prognosis, and more informed treatment decisions, ultimately improving cancer care [17]. Liquid biopsy has great potential, but its reliability depends on the further development of sensitive detection methods, standardized procedures and uniform data analysis [13]. These limitations hinder early detection and effective treatment, emphasizing the need for more accurate, affordable, and non-invasive diagnostic tools to improve cancer treatment outcomes.

In this context, Raman spectroscopy helps to overcome several of these challenges by offering a highly sensitive and chemically specific approach to cancer diagnostics. In particular, RS is based on the Raman effect, a physical phenomenon that occurs when light interacts with the molecules of a substance. When light illuminates matter, most photons are scattered elastically with no change in energy, frequency or wavelength. However, a small fraction of light is scattered inelastically, i.e. the photons gain or lose energy due to interactions with the vibrational modes of the molecules [18], [19]. This energy shift provides a unique spectral "fingerprint" that reflects the molecular composition and structure of the material and enables for highly specific chemical analysis [20], [21]. The specific energy with which molecules vibrate depends on their chemical bonds and the surrounding environment. This allows vibrational spectroscopy to identify and quantify the different components of a mixture and to detect even small differences between samples [22]. RS enables the analysis of different biological samples by acquiring point spectra or spectral images without the need of labeling, while preserving the integrity of the sample due its non-destructive and non-invasive nature [23]. In this regard, Raman spectroscopy can analyze a wide range of biological samples, including live or fixed cells *in vitro*, *ex vivo* tissue sections, *in vivo* biopsies and biofluids (blood, saliva, etc.), facilitating minimally invasive or real-time diagnostics [24], [25], [26]. Changes in the biochemical composition of cells, tissues or body fluids -whether as a result of disease or as a contribution its development - can cause noticeable changes in their Raman spectra. Specifically, biological molecules such as lipids, proteins, and DNA - each with unique chemical and molecular properties - can be easily identified by analyzing the Raman spectrum of the cell as shown in **Figure 1**, with label-free (no staining). Raman spectroscopy is particularly useful for relative quantification of the proportions of broad molecular groups [27]. Variations in molecular signatures can provide valuable insights into disease processes. In addition to its precise molecular detection, Raman scattering can be excited with near-infrared (NIR) light, enabling real-time measurements and making it particularly suitable for biomedical applications such as live tissue imaging [25]. The strength of RS lies in its ability to detect change at the molecular level, making it a valuable tool for disease diagnosis, prognosis as well as evaluating the efficacy of new treatments [26].

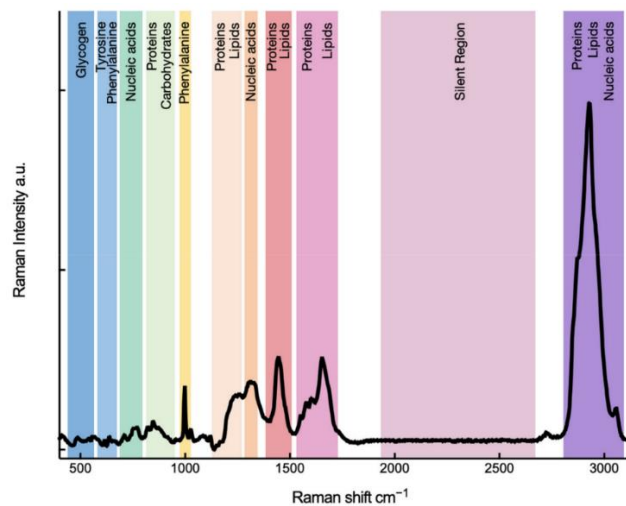


Figure 1. The Raman spectrum provides a unique molecular signature of a cell based on its biochemical constituents, from Ref. [28].

The interpretation of Raman spectral data can be challenging due to its complexity and high dimensionality. Biological spectra often contain overlapping molecular signals and are influenced by the heterogeneous nature of cells and tissues, which complicates analysis. To solve these problems, advanced computer techniques are used to interpret complex optical signals when analyzing Raman data. Methods such as linear regression and multivariate analysis are used to detect molecular signatures with high precision. To streamline this process, artificial intelligence is often integrated either through pre-programmed algorithms or by enabling systems to learn and extract key features from the data [29]. Different types of algorithms are able to manage large, complex datasets, identifying subtle spectral differences that help distinguish between healthy and cancerous cells and tissues, and categorize different cancer types and grades with high accuracy [30], [31], [32]. This integration of ML not only improves diagnostic confidence and reduces human error, but also streamlines clinical workflows and supports more personalized treatment approaches. Although Raman spectroscopy holds great promise for clinical oncology, its broader application still faces hurdles such as weak signal strength and interference from tissue auto-fluorescence [18], [33]. However, integration with other diagnostic tools such as molecular techniques and AI-assisted algorithms could enable a deeper understanding of tumor biology. Given the limitations of current cancer diagnostics, Raman spectroscopy is characterized by its molecular precision, minimal invasiveness and versatility. Combined with machine learning, it can improve interpretation and diagnostic accuracy, supporting earlier detection and better treatment planning. In the future, it will be crucial to make these advanced technologies accessible to a wide audience, which is an important step towards more precise and patient-friendly cancer treatment.

2. Raman spectroscopy: Basic Principles and Fundamentals

2.1 Raman scattering

In 1923, A. Smekal first theorized the phenomenon of inelastic scattering of light, but this remained unconfirmed until 1928, when C.V. Raman and K.S. Krishnan carried out experiments to prove the Raman scattering [34], [35]. It was found that when monochromatic light passed through a transparent medium, a small portion of the scattered light exhibited frequency shifts. This phenomenon confirmed the occurrence of inelastic scattering and established the basis for what was later called the Raman effect [36].

When light interacts with matter, its electromagnetic field can temporarily distort the electron cloud of a molecule, inducing an electric dipole. This process, known as polarization, occurs because the oscillating electric field of the light shifts the negatively charged electrons relative to the positively charged nuclei. This results in a temporary separation of charges within the molecule [37]. The induced electric dipole moment (μ) is directly proportional to the strength of the external electric field (E), with the molecular polarizability (α), as given in the following equation [38]:

$$\mu = \alpha E$$

Here α refers to the polarizability of a molecule, characterizing the degree to which the electron cloud in a molecule or material can be distorted by an external electric field. This concept physically represents how the electric field modifies the electron density from its equilibrium position when no electric field acts on it. Additionally, the electric field E is defined by a vector amplitude E_0 and an oscillation frequency ω_0 . Consequently, E can alternatively be represented as [38]:

$$E = E_0 \cdot \cos(\omega_0 t)$$

In particular, the polarizability of a molecule is not a fixed property, but changes dynamically with the movement of its nuclei. This is because the electron cloud constantly adjusts to the current positions of the nuclei in order to minimize the total energy of the system. As the nuclei move, the electron density is redistributed accordingly, which affects the polarizability of the molecule. To describe this dynamic behavior mathematically, the polarizability is extended as a Taylor series by the equilibrium nuclear configuration of the molecule, which is denoted as q_0 . In this context, Q represents the set of all normal vibrational modes of the molecule, where each mode is described by a coordinate q . Each vibrational mode oscillates at its characteristic frequency ω_R , and the time shift for a given mode can be expressed as follows [38]:

$$q = q_0 \cdot \cos(\omega_R \cdot t)$$

These nuclear vibrations modulate the geometry of the system, which in turn leads to time-dependent changes in polarizability. As a result, the induced dipole moment μ changes with time, leading to [37], [38]:

$$\mu(t) = \alpha_0 \cdot E_0 \cdot \cos(\omega_0 t) + \frac{1}{2} \left(\frac{\partial \alpha}{\partial q} \right)_{q_0} \cdot q_0 \cdot E_0 \cdot \cos[(\omega_0 - \omega_R)t] + \frac{1}{2} \left(\frac{\partial \alpha}{\partial q} \right)_{q_0} \cdot q_0 \cdot E_0 \cdot \cos[(\omega_0 + \omega_R)t]$$

The induced electric dipole moment μ oscillates at three distinct frequencies. The first oscillation corresponds to the excitation frequency ω_0 , which leads to Rayleigh scattering, where the scattered light has the same frequency as the incident light and the polarizability is constant over the time. The other two vibrations are caused by interactions between the external field ω_0 and vibrational modes of the molecule ω_R . In vibrating molecules, the polarizability α fluctuates as the molecular vibrations alter the electron distribution. In particular, the second oscillation corresponds to Stokes scattering, where the scattered light has a lower frequency (red shifted), while the third oscillation corresponds to anti-Stokes scattering, where the scattered light has a higher frequency (blue shifted). The result is that the scattered light contains a combination of the original excitation frequency and shifted frequencies that corresponding to the molecular vibrations. This combination of frequencies provides information about the vibrational properties of the molecule, which can be observed as Rayleigh scattering and inelastic scattering associated with vibrational modes [37].

Although the classical model of inelastic Raman scattering helps to illustrate the basic process, it cannot explain why Stokes and anti-Stokes signals differ in intensity. A quantum mechanical theory is therefore essential for a complete and accurate understanding of these processes. According to quantum theory, light is not considered as a continuous wave but as individual energy quantized packets called photons. Each photon has a specific energy that depends directly on its frequency (ν), as described by the following equation [39]:

$$E = h\nu$$

In this formula, E stands for the energy of the photon, ν for its frequency and h for constant of Planck. When an incident monochromatic electromagnetic radiation with energy $h\omega_0$ interacts with a sample, it can undergo reflection, absorption and scattering in different directions. When an excitation light interacts with a molecule, it can bring the molecule to a higher electronic state. If the excitation energy does not correspond to a specific electronic transition, the interaction between an incident photon and a molecule will cause the molecule to briefly transition to a virtual energy state—a transient, unquantized state (i.e., the state of an induced dipole moment), that is above the vibrational levels of the ground state but is not a stable energy level. This state only exists during the interaction with the excitation source [40].

In Rayleigh scattering, the interaction between light and a molecule is elastic, i.e. there is no net energy exchange - the scattered photons retain the same frequency and energy as the incident light. When a molecule relaxes from an excited virtual state, it can either return to its original energy level by releasing a photon with the same energy as the incident light. Therefore, this process does not provide information about molecular vibrations. In contrast, Raman scattering is an inelastic process in which the energy of the photon changes due to the interaction with the vibrational or rotational states of the molecule, although spontaneous Raman scattering occurs with extremely low probability - only about 1 in 10^8 incident photons is scattered by this process, making the Raman effect naturally very weak [18]. In Stokes scattering, the incident photon transfers some of its energy to the molecule, causing it to vibrate at a higher frequency. The resulting scattered photon has lower energy and frequency ($h\omega_0 - h\omega_R$). In contrast, in anti-Stokes scattering the molecule is initially in an excited vibrational state. It transfers energy to the photon, producing scattered light with higher energy and frequency ($h\omega_0 + h\omega_R$) [18], [41]. Therefore, in Raman scattering, the molecule instead relaxes to a different vibrational energy level, emitting a photon whose energy differs from that of the excitation source by an amount corresponding to the vibrational energy of the molecule, making its vibrational modes visible [19], [40]. Due to the Boltzmann distribution, the vast majority of molecules are in the fundamental vibrational state at room temperature, so Stokes scattering is much more intense than anti-Stokes scattering [19], [40]. Anti-Stokes signals are generally weaker because they rely on molecules being in an excited vibrational state before scattering occurs—a condition that is less common at typical temperatures [42]. For this reason, conventional Raman spectroscopy focuses primarily on the Stokes spectrum, which is more intense and contains detailed information about molecular vibrations [43]. Anti-Stokes signals also reflect vibrational properties, but their lower intensity limits their practical use under standard conditions.

These phenomena are illustrated in **Figure 2**, which shows a Jablonski diagram outlining vibrational and electron energy levels of the molecule. It visualizes how Raman scattering incorporates virtual energy states and highlights the differences between Stokes and anti-Stokes processes in terms of the direction of energy transfer and the initial vibrational state occupation.

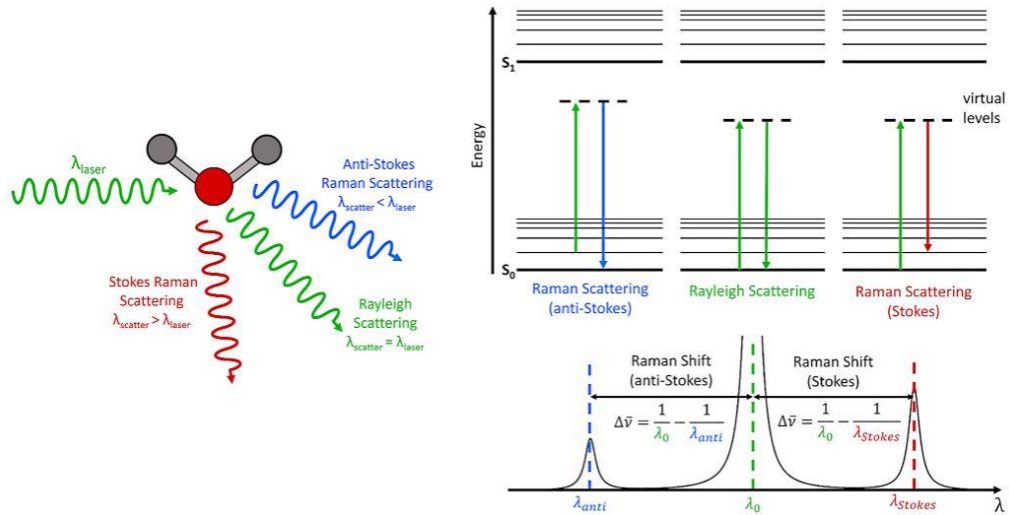


Figure 2. Schematic representation of the phenomenon of Rayleigh, anti-Stokes and Stokes Raman scattering, from Ref.[44].

The intensity of Raman scattered radiation, I_R , is determined by several key factors and can be expressed as follows [39]:

$$I_R \propto \nu^4 I_0 N \left(\frac{\partial \alpha}{\partial Q} \right)^2$$

In this expression, I_0 is the intensity of the incident laser, N is the number of molecules involved in scattering, ν is the frequency of the excitation laser, α is the molecular polarizability, and Q is the vibrational amplitude. The term $(\partial \alpha / \partial Q)$ indicates how polarizability changes with molecular vibration. This equation shows that the Raman signal depends on several important parameters. Since the intensity is proportional to the number of scattering molecules, Raman spectroscopy can be used for quantitative analysis. Since the signal scales with the fourth power of the laser frequency, using an excitation source with a higher frequency (shorter wavelength) or increasing the laser power can significantly enhance the Raman signal.

The wavelength of Raman scattered light depends on the wavelength of the excitation source. Consequently, when different lasers are used, the measured Raman scattering wavelengths vary, making direct comparisons between spectra misleading. For this reason, Raman data are typically expressed as the Raman shift (in cm^{-1}), which represents the difference in energy between the excitation light and the scattered light. The Raman shift is calculated using the following formula [45]:

$$\Delta \nu (\text{cm}^{-1}) = \left(\frac{1}{\lambda_0 (\text{nm})} - \frac{1}{\lambda_1 (\text{nm})} \right) * \frac{(10^7 \text{ nm})}{(\text{cm})}$$

where λ_0 is the wavelength of excitation and λ_1 is the wavelength of scattering, both in nanometers.

Furthermore, only those molecular vibrations that result in a change in polarizability contribute to the Raman effect. For a vibration to be Raman active, the derivative of polarizability with respect to the vibrational coordinate, $(\partial\alpha/\partial Q)$, must be greater than zero: $(\partial\alpha/\partial Q) \neq 0$. The Raman intensity is proportional to the square of this value [39]. If the molecule vibrates and its bond lengths or angles shift, thereby changing the polarizability, Raman scattering can occur. The more the polarizability changes during the vibration, the stronger the scattered signal will be. This is the most important selection rule for Raman spectroscopy and differs from IR spectroscopy, where a vibration must be associated with a change in the permanent dipole moment to be IR active [40]. As shown in **Figure 3**, the nonlinear CO₂ molecule exhibits four basic vibrational modes: asymmetric C–O stretching, symmetric C–O stretching and two degenerate O–C–O bending modes (with the same frequency).

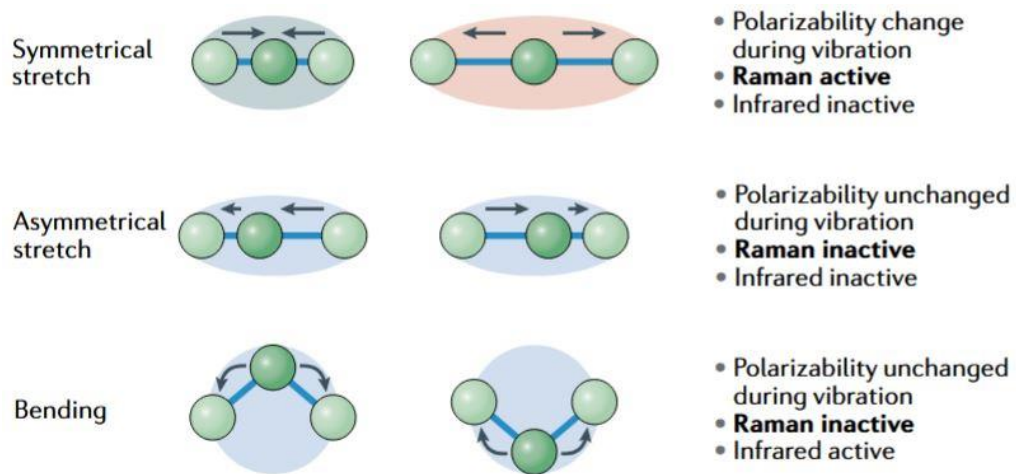


Figure 3. CO₂ vibrational modes and their relative IR and Raman activity, from Ref. [46].

Of these, only the symmetric stretching mode changes the polarizability of the molecule and makes it Raman active. However, this particular mode is not associated with a change in the dipole moment and is therefore inactive in infrared (IR) absorption spectroscopy. This mode is not detected in IR spectroscopy because it does not change the dipole moment [46]. The fundamental difference between the two techniques lies in the underlying physical mechanisms. In IR absorption spectroscopy, a molecule directly absorbs a single infrared photon that provides enough energy to excite a vibrational mode - it is a one-photon absorption process. For a vibration to be IR-active, it must involve a change in the molecular dipole moment during the vibration. In contrast, Raman spectroscopy is based on a two-photon scattering process. In this process, a photon of incident light interacts with the molecule and part of its energy is transferred to (or extracted from) a vibrational mode, resulting in a scattered light with a different frequency. For a vibration to be Raman active, it must cause a change in the polarizability of the molecule during the vibration and not its dipole moment [37].

2.2 Interpretation of Raman spectrum

Raman spectra are generally plotted with the signal intensity or count rate on the vertical axis (y-axis) and the Raman shift on the horizontal axis (x-axis). The Raman shift, expressed in wavenumbers, indicates the frequency difference between the incident laser light and the scattered light and does not depend on the wavelength of the laser itself. The count rate corresponds to how much scattered light is detected at each shift, which reflects the vibrational behavior of molecules in the sample. The interpretation of Raman spectra usually begins with the identification of the functional groups present in the molecule, as each type of bond vibrates at distinct frequencies that appear as specific peaks in the spectrum. These vibrational frequencies are affected by the molecular structure, particularly its symmetry, the polarization of the scattered light and the selection rules that determine which vibrational modes are allowed [45]. Molecular vibrations refer to the slight motions of atoms oscillating around their stable position. These motions are commonly modeled using the harmonic oscillator concept, where the vibrational frequency (ω_R) is influenced by two key parameters: the reduced mass of the atoms involved and the force constant, which represents the strength of the bond. Atoms with lower mass and stronger bonds therefore tend to vibrate at higher frequencies. Notably, the vibrational frequency itself is unaffected by the specific energy levels involved in the transitions. For a molecule consisting of N atoms, there are a total of $3N$ degrees of freedom. Of these, 3 correspond to translational movement and another 3 to rotation—except for linear molecules, where rotation accounts for only 2. The remaining degrees of freedom ($3N-6$ or $3N-5$ for linear molecules) correspond to internal deformations. The different types of molecular vibrations include stretching and deformation movements such as scissoring and wagging, among others [27]. For molecules, vibrations can be classified based on symmetry (symmetric or asymmetric) and orientation (in-plane or out-of-plane). If the angles between the bonds and the plane formed by the other atoms remain unchanged, the vibration is classified as in-plane; otherwise as out-of-plane [19]. Of the total degrees of freedom of a molecule, some are used for translational and rotational vibrations, while the others describe internal vibrations such as stretching, bending and torsion. Only vibrations that significantly change the polarizability of a molecule are Raman-active. Group theory helps predict which modes appear in Raman spectra, with symmetry playing a key role. Raman signals are weak for strongly polar bonds (such as O-H) and strong for bonds with high polarizability changes (such as C=C) [27]. By analyzing the positions of Raman peaks and observing variations in their intensity or shape, it is possible to determine the identity of a material and gain insight into its physical properties. For example, shifts in peak position, broadening or changes in intensity can indicate mechanical stress, variations in crystallinity or changes in concentration. Additionally, comparing spectra from different points of a sample can reveal inconsistencies in the material composition. The vibrational frequency itself is determined by both the mass of the atoms and the strength of the chemical bonds; strong bonds and lighter atoms vibrate at higher frequencies, producing larger Raman shifts, while weaker bonds and heavier atoms produce lower shifts [45]. Each Raman peak represents a unique molecular vibration and serves as a chemical signature. This makes Raman spectroscopy a powerful tool for identifying molecular structures.

2.3 Instrumentation

A Raman spectrometer consists of several key elements: a laser source for excitation, optical filters, a diffraction grating and a charge-coupled (CCD) detector, as depicted in **Figure 4**. Additionally, Raman systems are integrated into microscopes to enable the examination of samples with very high spatial resolution below the diffraction limit. Raman microspectroscopy benefits from the fact that the excitation light is concentrated on a small area, while the scattered photons are efficiently collected by objective lenses with high numerical aperture [24].

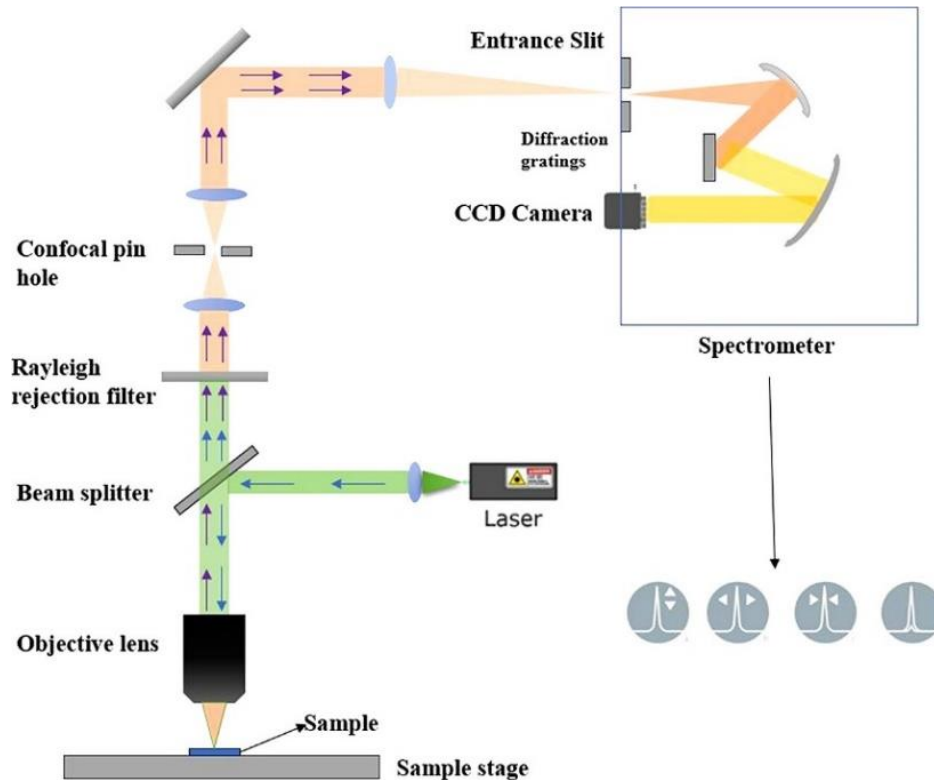


Figure 4. Diagram illustrating the key elements of confocal Raman spectroscopy, from Ref. [45].

Lasers are well-suited as excitation sources for Raman spectroscopy because they provide intense, monochromatic (single wavelength), polarized light with stable wavelengths (generally 532 nm, 633 nm, 785 nm or 1064 nm) and narrow spectral bandwidths. When working with cells or tissues, the laser power must be carefully controlled to avoid damage such as carbonization or phototoxicity, which also depends on the wavelength, focusing optics and specific sample type. Typical laser powers used range are 10 and 100 mW [24]. While lasers with shorter wavelengths provide stronger Raman signals, they are also more prone to producing fluorescence, which can obscure the spectrum. For this reason, lasers with a wavelength of 785 nm are often used as a compromise between signal strength and fluorescence suppression, while lasers with a wavelength of 1064 nm are preferable for samples with high fluorescence, even if they produce weaker Raman signals [45]. The optical configuration of the spectrometer includes lenses to focus the laser beam and filters to eliminate unwanted light- especially Rayleigh scattered light - to ensure that only the relevant Raman scattered photons reach the spectrometer. Notch and edge filters play an

important role in this light selection. In Raman spectrometers, a diffraction grating is usually used to separate the Raman scattered light. The line density largely determines the spectral resolution and the wavelength range covered. The overall performance also depends on other factors such as the width of the entrance slit, the distance between the grating and the detector (the spectrometer's focal length), and the size of the pixels on the CCD detector [24]. Gratings with higher line densities enhance spectral resolution but decrease light intensity, whereas those with fewer lines per millimeter allow a wider spectral range at the cost of resolution [47]. Detection is performed with cooled CCDs, which are multichannel detectors capable of capturing the Raman spectrum while reducing thermal noise and enhancing the signal-to-noise ratio. The configuration and selection of the laser, filters, grating, and detector are tailored to match the sample's characteristics and the analytical goals. Together, these components function in harmony to generate high-quality Raman spectra, essential for accurate identification and analysis of materials [45].

3. Fundamentals of Raman Data Analysis

3.1 Pre-processing of Raman Data

Raman spectroscopy applied to biological samples produces complex, high-dimensional spectral data that contain detailed biochemical information but are often compromised by low signal-to-noise ratio and artifacts arising from instrumental noise, sample properties (sample heating, fluorescence), and sampling-related factors [48]. A well-structured analysis pipeline (with effective pre- and post-processing steps) is crucial, as Raman spectra often contain diagnostically significant features- such as characteristic marker bands or spectral fingerprints- that reflect the underlying biological state and contribute to a deeper understanding of biomolecular differences between health and disease. Comprehensive analytical workflows are used to accurately identify and interpret these features. This pipeline typically includes the removal of artifacts, normalization of data to ensure standardization between samples, and the application of statistical or machine learning models to transform the preprocessed spectra into meaningful and actionable insights, as illustrated in **Figure 5** [49]. In particular, the intensity of Raman signals is influenced by factors such as laser characteristics, sample properties and temperature. However, the spectra can also be affected by various sources of interference such as instruments, ambient light and computational artifacts. Fluctuations in the photon flow between the sample and the device caused by changes in laser power, sample opacity or instrument conditions lead to multiplicative distortions.

Furthermore, variations in spectral data can occur not only between different replicates of a sample, but also within the same sample when measured at different times. These variations result from inherent biological heterogeneity and dynamics, sample degradation, differences in sample preparation, or other experimental factors that are beyond control [50]. In addition, fluorescence, which is often much stronger than the weak Raman signal, can dominate and obscure spectral features, especially in biological samples [52]. Preprocessing of biological Raman spectra usually involves a number of main steps, such as spectral peak removal, wavenumber calibration, smoothing, background signal subtraction, baseline deviation correction, and normalization of intensity values, all of which aim to remove certain sources

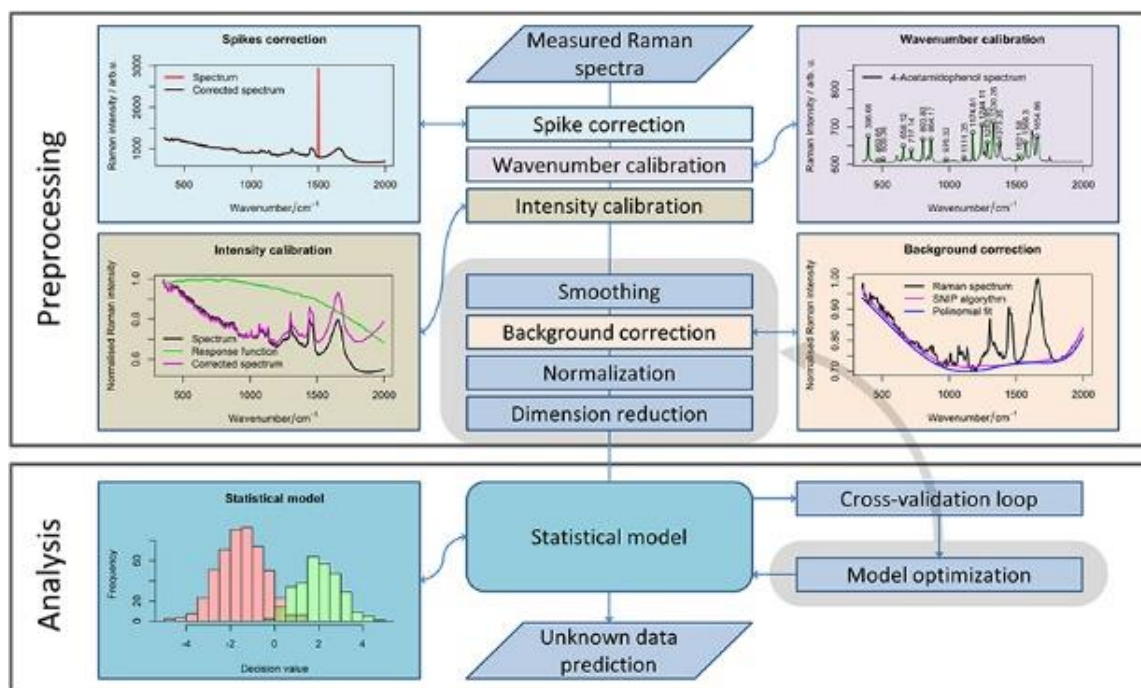


Figure 5. A comprehensive workflow for pre- and post-processing of Raman data, from Ref. [51].

of noise and bias in the data [49]. The first preprocessing step is de-spiking, which deals with sharp, non-molecular artifacts caused by cosmic rays emanating from high-energy particles hitting the CCD detector during spectrum acquisition [52], [53]. Generally, spikes are removed by comparing repeated spectra and replacing unusually high values with lower ones if the difference is greater than the normal noise. Spikes are identified by their sharpness and corrected using a fully automatic despiking procedure or a local interpolation algorithm [54], [55], [56]. Subsequently, in order to ensure comparable spectra, standardization through spectrometer. To ensure comparable spectra, standardization by spectrometer calibration is then required. The measured peak positions are adjusted using standard materials or reference measurements (wavenumber calibration) and the intensity is corrected using a reference standard (intensity calibration) [57]. Smoothing of biological Raman spectra is an optional step for reducing noise and enhancing spectral clarity [50]. Common approaches such as moving average and Savitzky-Golay filters are simple to apply but can sometimes smooth the data too much, resulting in loss or distortion of important spectral details [58]. Nonetheless, it is essential to apply smoothing functions carefully, as the algorithms used for this process can unintentionally remove or distort meaningful spectral information. In fact, more advanced methods - such as Wiener filtering and Voigt profile fitting - offer improved noise suppression while retaining important spectral details [59], [60]. The selection of appropriate parameters plays a crucial role in achieving an optimal balance between noise reduction and feature preservation. Baseline correction is an essential part of preparing Raman spectra for analysis. In the literature, it usually refers either to the removal of the substrate background signal from the sample spectrum or to the removal of the broad fluorescence background inherent in the sample itself [50]. The contribution of the substrate (i.e. the glass) must be removed if it exhibits strong Raman signals, especially if these signals overlap with the characteristic bands of the sample, as this may interfere with the accurate interpretation of the spectrum [61]. However, removing the fluorescence

background in Raman spectra is also crucial, as the auto-fluorescence of biological samples - often caused by inherent biomolecules, the morphology of the sample or Mie scattering by the laser wavelength - can significantly obscure the actual Raman signal. This background interference reduces spectral clarity and poses a major challenge, especially when using multivariate classification methods for disease diagnosis or screening applications [62]. In Raman spectra, the baseline usually changes gradually and smoothly, in contrast to the pronounced sharp peaks of Raman signals. As a result, the baseline can be approximated by a smooth function, which can then be subtracted from the raw data to more clearly identify the underlying Raman peaks and improve the accuracy of the analysis [63]. Various algorithms have been implemented to perform baseline correction, such as derivative processing, asymmetric least squares (ALS), sensitive nonlinear iterative peak (SNIP) algorithms, extended multiplicative signal correction (EMSC), and different types of polynomial fitting [64], [65], [66], [67], [68]. Additionally, Raman spectra can exhibit large intensity variations between different laboratories, instruments or measurement sessions and sometimes even during a single experiment such as mapping or time-dependent studies. These variations are usually caused by factors related to the equipment or experimental setup, such as shifts in laser focusing, thickness or power of the detector [63]. In studies involving molecular changes or subtle biological variations, the choice of the correct normalization method is particularly important as it can significantly affect the ability to detect and highlight meaningful biological signals in the data [69]. Several normalization techniques are commonly used to standardize Raman spectra and make them more comparable across different samples or conditions, such as vector normalization, standard normal variate (SNV), peak intensity normalization, and min-max scaling [70]. In particular, vector normalization adjusts the spectrum so that the sum of the squared intensities equals one while maintaining the relative shape of the spectrum, and it can also be applied independently to the entire spectrum, a particular wavenumber, or a Raman band that is expected to have a constant intensity across all samples [71]. SNV removes scattering effects by centering and scaling each individual spectrum by subtracting the mean of the spectrum and dividing by the standard deviation of the spectrum [72]. In peak normalization, all spectral values are divided by the intensity of a selected Raman peak that serves as a reference point. Although this method is simple, it is not suitable when the peak position varies between samples, such as in cases where structural changes - such as protein denaturation - lead to spectral shifts [73]. In min-max scaling, the spectrum is scaled so that all values lie between a defined minimum and maximum, usually between 0 and 1, while maintaining the shape and equalizing the range. In this case, the minimum value of the spectrum is first subtracted and the resulting values are then divided by the maximum of the adjusted spectrum. While this method is simple and convenient to use, it is generally more susceptible to noise than other normalization methods [71]. Each method is suitable for different analysis objectives and data characteristics.

3.2 From conventional methods to artificial intelligence approaches

Raman spectroscopy captures a detailed biomolecular fingerprint of a sample and reveals unique information about the chemical composition based on molecular vibrations.

However, because Raman spectra are rich in data - often containing signals from many overlapping molecular features - their interpretation can be challenging. The differences between spectra of different sample types are frequently subtle, with only slight variations in peak positions or intensities. These subtle differences are not always easy to see with the naked eye, making it difficult to visually distinguish similar samples without computational analysis or advanced statistical tools. Once properly pre-processed, RS spectral data can be analyzed for a range of biological applications. These include identifying biomolecules in cells, distinguishing between different cell types and measuring the concentrations of biomolecules or pharmaceutical compounds in cells [74]. In this context, chemometric techniques have become essential tools for extracting meaningful information from Raman spectral data. Given the complexity and high dimensionality of Raman spectra, conventional visual analysis is often inadequate. Chemometrics enables researchers to uncover subtle patterns, distinguish between sample types, and quantify molecular components by applying advanced statistical and machine learning methods. These techniques improve the interpretability and analytical performance of Raman spectroscopy, making it indispensable for accurate and efficient data analysis in both research and applications [50]. Furthermore, machine learning encompasses multivariate analysis along with various other statistical techniques and is generally divided into two main categories: unsupervised and supervised learning [74]. In particular, unsupervised algorithms analyze data without using predefined labels or prior knowledge about the samples. This type of approach is mainly used for exploratory purposes to uncover patterns, similarities and differences within the data and to reduce the dimensionality of complex datasets [71], [75]. By revealing similarities or differences between samples, unsupervised methods help to generate hypotheses and guide further, more targeted analyses. Principal component analysis (PCA), hierarchical cluster analysis (HCA) and k-means cluster analysis (KCA) are widely used unsupervised methods that can reveal similarities and differences between spectral data instead of classifying them. A common approach to visualizing single-cell Raman data is to reduce the dimensionality of the spectral matrix by creating new variables, as is done in techniques such as principal component analysis. In PCA, the data is organized as a matrix with the spectral wavenumbers as columns (variables) and the individual spectra as rows (observations). To reduce dimensionality, the data is transformed into uncorrelated variables, called principal components (PC), using orthogonal linear combinations. The first principal component captures the direction of maximum variance and ensures the greatest separation within the data set [25]. After reducing the dimensionality of the dataset, PCA identifies key components that capture the most variance, making it easier to detect trends (such as increases or decreases in Raman peak intensities) or groupings. Specifically, PCA breaks the data into two matrices: loadings and scores. The loadings represent the new variables that capture the most relevant spectral information from biomolecules, while the scores indicate how each sample contributes to these new variables. These scores reflect the relative concentrations of biochemical components, enabling both semi-quantitative analysis and spatial visualization of biomolecule distribution within the cell [74]. Instead of using new variables to extract biochemical information from Raman spectra, cluster analysis (HCA, KCA) groups pixels or objects in the dataset based on shared spectral characteristics. HCA is often applied for analyzing Raman hyperspectral images. Specifically, HCA starts by

computing a symmetrical distance matrix that quantifies the similarity between each pair of n spectra. The method then finds the two spectra with the smallest distance and merges them into an initial cluster. This process repeats until all spectra are grouped into a single cluster after multiple steps [76]. Furthermore, HCA operates using either a top-down (divisive) or bottom-up (agglomerative) strategy to group similar spectra, typically using Euclidean distance and Ward's algorithm. The results are visualized as a dendrogram, where branching points indicate differences between spectra. After clustering, colors are assigned to each group to generate pseudo-color maps that represent different cellular compartments, such as the nucleus, mitochondria and etc. While effective, HCA can be computationally intensive, with large images taking hours to process [25], [76]. K-means cluster analysis organizes data into groups by reducing variation within clusters and increasing differences between them, typically using Euclidean distance. The process begins by choosing K random spectra as initial cluster centers (centroids). Each spectrum is then assigned to the closest centroid, after which the centroids are updated based on the average of assigned spectra. This cycle continues until the cluster assignments no longer change. The result is a division of the data into K clusters, which can be displayed as a false-color map [76].

In contrast, supervised algorithms rely on labeled data—known as classes—to train predictive models. They learn patterns from labeled data during a training phase and then apply this knowledge to classify new, unseen data in a prediction phase [54]. A model can be constructed to learn the fundamental traits of the population and generalize effectively to new, unseen data, yielding comparable prediction errors on both training and test samples [50]. These models can be divided into regression or classification types. Regression models are typically used for calibration and provide quantitative results, whereas classification models produce qualitative outcomes [75]. RS is commonly employed to classify cells according to various criteria, such as different stages of tumor development, distinct cell types, and levels of cell differentiation [77], [78], [79], [80]. This classification process is typically carried out using supervised machine learning algorithms, including Support Vector Machines (SVM), Random Forest (RF), Partial Least Squares Discriminant Analysis (PLS-DA), K-Nearest Neighbors (KNN), Artificial Neural Networks (ANN), and Linear Discriminant Analysis (LDA) [81].

LDA is a supervised learning method that serves as both a classifier and a dimensionality reduction technique. It identifies the optimal linear combinations of features that maximize the separation between class means while minimizing variance within each class [77]. This focus on maximizing inter-class separability and minimizing intra-class variation makes LDA highly effective for both classification and data visualization. LDA is commonly applied on its own or alongside PCA to enhance data analysis and interpretation. In such combined approaches, LDA leverages the principal components derived from PCA to distinguish between different classes. This integrated use of PCA and LDA has been widely adopted in the literature, particularly for cell classification tasks [82], [83], [84]. RF is an ensemble algorithm that constructs multiple random decision trees and aggregates their predictions via voting [81]. Although powerful, RF can struggle with high-dimensional spectral data due to long processing times and correlated variables. These challenges can be addressed by using dimensionality reduction methods, such as PCA, prior to training the RF model [85]. RF has been combined with non-negative matrix factorization to classify

molecular histotypes in breast cancer cells, achieving high levels of accuracy, sensitivity, and specificity in distinguishing cancerous from non-cancerous tissues [86]. For example, the KNN algorithm classifies data by comparing the distances between points and assigning categories based on the majority among the nearest neighbors, without a formal training phase. KNN, along with other machine learning algorithms, has been used to facilitate diagnostic classification in the case study of colon cancer [87]. In contrast, the SVM is a powerful binary classification method suited for nonlinear and high-dimensional data, particularly when sample sizes are small. SVM works by finding the optimal boundary (hyperplane) to separate different groups in the training data. It does this by maximizing the distance between the boundary and the closest data points, called support vectors, to ensure accurate classification [74], [81]. Although its performance depends on the kernel function and it can be inefficient with large datasets, SVM is widely applied in Raman spectroscopy for tumor diagnosis [88], [89]. Collectively, these machine learning methods significantly enhance the analytical capabilities of Raman spectroscopy, enabling fast and precise detection of biochemical alterations in biological specimens. However, the aforementioned conventional techniques are effective for small datasets but face challenges when processing large, complex data. Deep learning models—such as ANNs, which mimic the structure of the human brain using layers of interconnected neurons to process data, and Convolutional Neural Networks (CNNs), a type of ANN specialized for recognizing patterns in structured data like images or spectra - differ from traditional machine learning models [90], [91]. These deep models excel at capturing complex, non-linear relationships in data, enabling them to detect patterns too subtle for conventional methods. As a result, they hold great potential for transforming digital healthcare and biophotonics applications, including Raman spectroscopy [92], [93], [94], [95]. While their success depends on data quality, CNNs are widely favored for their simple design and strong performance in handling extensive, intricate datasets. Moreover, data augmentation techniques—such as Generative Adversarial Networks (GANs)—can enhance deep learning performance even when only limited data are available [96]. Therefore, selecting the most suitable algorithm should be guided by the nature and requirements of the specific problem.

Ensuring the accuracy of machine learning models—particularly those using supervised approaches—is essential to confirm their reliability and prevent misleading results. This is accomplished through validation techniques such as cross-validation, confusion matrices, and ROC curve analysis. Cross-validation involves splitting the data into subsets to train and test the model multiple times. The prediction results from all iterations are then averaged to provide an overall measure of the model's ability to generalize [71]. Depending on how the data is split, cross-validation can take various forms, such as leave-one-out, k-fold, or Monte Carlo cross-validation [97]. The confusion matrix is an essential tool for evaluating the performance of supervised classification models. In binary classification, it is represented as a 2x2 table showing the counts of true positives, true negatives, false positives, and false negatives from the model's validation. This matrix can also be extended to handle multiple classes. Model performance is commonly measured using sensitivity (the rate of correctly identified positives) and 1-specificity (the rate of incorrectly identified negatives). A model with high sensitivity and low 1-specificity is considered highly accurate [74]. The ROC curve further assesses performance by plotting sensitivity against the false positive rate

across thresholds, with the area under the curve (AUROC) indicating how well the model distinguishes classes—values closer to 1 signify better accuracy [74]. These steps are critical for confirming the accuracy and robustness of the analytical results, especially in sensitive applications such as the use of Raman spectroscopy as a diagnostic tool for cancer.

4. Application of Raman spectroscopy in advanced cancer diagnostic

Raman spectroscopy is an advanced diagnostic tool that enables the biochemical characterization of both tissues and individual cells. At the cellular level, Raman spectroscopy provides detailed spectral profiling of individual cells and is commonly used to distinguish healthy cells from cancerous ones, as well as to differentiate between various cancer subtypes. It also supports the development of novel single-cell diagnostic techniques, such as Raman-based cytopathology, which shows strong potential for minimally invasive disease detection [98]. In tissue analysis, it complements conventional histopathology by delivering molecular-level insights through spectral imaging and computational analysis, facilitating detailed chemical mapping of pathological changes [24]. Ongoing technological progress—including improvements in resolution, speed, and data processing—is rapidly expanding its diagnostic utility. Recent innovations have led to the development of compact, handheld Raman instruments and automated spectral interpretation software. For real-time, *in vivo* diagnostics—during surgery or via endoscopic procedures—miniaturized fiber-optic Raman probes have enabled effective cancer detection [99]. These advancements underscore the growing potential of Raman spectroscopy as a practical, real-time tool for cancer diagnosis and intraoperative guidance.

4.1 Raman spectroscopy *in vitro*

Recently, Abramczyk et al. proposed a new immuno-detection technique for assessing human epidermal growth factor receptor 2 (HER2) protein expression in different breast cancer cell types *in vitro*, using Raman spectroscopy and Raman imaging assisted by artificial intelligence [100]. The general workflow is reported in **Figure 6A**. Initially, Raman analysis enabled the identification of molecular features associated with tyrosine phosphorylation - a key post-translational modification in cancer signaling — by capturing specific spectral changes, including the appearance of new peaks, shifts in characteristic bands, and intensity variations. In phosphorylated tyrosine, these alterations are particularly evident in vibrations related to aromatic ring structures and phosphate group stretching modes, allowing differentiation between phosphorylated and non-phosphorylated states. Notably, the technique also detected distinct changes in the vibrational signatures of cytochrome c, especially in bands associated with its methine bridge and heme group, underscoring its sensitivity to phosphorylation. Subsequently, HER2 expression was analyzed in five human breast cell lines representing diverse cancer phenotypes: the normal cell line MCF-10A, the triple-positive MCF-7, the triple-negative MDA-MB-231, HTB-30 (SKBR3) and AU-565, both of which overexpress HER2, with AU-565 also overexpressing additional oncogenes. The Raman intensity at 1618 cm^{-1} , used as a marker for HER2 tyrosine kinase activity, was analyzed to estimate its concentration. This intensity was significantly higher in MCF-7, HTB-30, and AU-565 compared to the normal MCF-10A cells, indicating

elevated levels of tyrosine kinase. In contrast, the triple-negative breast cancer cell line MDA-MB-231 displayed a lower Raman signal than MCF-10A, suggesting reduced HER2 activity. Additionally, the Raman-based measurements showed a strong linear correlation with conventional IHC results for HER2 status, validating the method's accuracy. The study demonstrated that AI-assisted Raman spectroscopy provides a precise, label-free method for HER2 detection, complementing traditional tests and supporting more accurate, personalized breast cancer diagnosis and treatment. In another study, Zhang *et al.* combined Raman spectroscopy with machine learning to distinguish normal breast epithelial cells from various breast cancer subtypes [101]. Six breast cell lines were analyzed by Raman spectroscopy; one normal cell line, MCF-10A, and five cancerous lines, MCF-7 (luminal A), ZR-75-1 (luminal B), BT474 (HER2+), and two triple-negative lines (MDA-MB-231 and SUM149). This selection included the four major subtypes of breast cancer. Multiple Raman spectra were acquired for each cell line to delineate their unique Raman signatures. Two machine learning algorithms, PCA-DFA and PCA-SVM, were trained on this spectral data. Both models performed well in detecting cancerous cells and differentiating between subtypes, with PCA-SVM achieving up to 99% accuracy in distinguishing normal from cancer cells and nearly 94% accuracy in classifying all subtypes. However, the accuracy decreased to approximately 70% when differentiating between two triple-negative subtypes (SUM149 and MDA-MB-231), likely due to their similar biochemical profiles. The study identified clear Raman spectral differences between normal and cancer cells, highlighting higher levels of glycogen, lipids, collagen, and aromatic amino acids in cancerous cells. These molecular distinctions reflect the pathological features of tumors and underscore the diagnostic potential of Raman spectroscopy.

The research by Surmaki *et al.* demonstrated the effectiveness of confocal Raman microspectroscopy as a precise, non-invasive method for identifying primary human bronchial epithelial cells (HBECs) in live-cell environments [102]. The study carefully optimized key experimental factors—including substrate type, cell culture media, laser wavelength, and scanning technique—to improve both spectral quality and cell viability. Line-scan acquisition with a 785 nm laser on quartz substrates provided the best results, balancing speed, resolution, and minimal photodamage. Spectral analysis revealed consistent biochemical differences, especially in peaks associated with DNA, lipids, and proteins. The in-house HBEC line, PAP243, showed greater variability compared to commercial lines, indicating higher biological diversity. Partial least squares discriminant analysis demonstrated excellent classification performance, achieving average sensitivities of 92.4% and specificities of 94.1% for individual spectra, which increased to 96.3% and 95.2% when averaged spectra per cell were utilized. Distinct Raman peaks corresponding to lipids (718, 1264, 1301, 1440, and 1658 cm^{-1}), proteins (641, 1003, 1166/1174, 1239, 1580, 1658, and 1674 cm^{-1}), nucleic acids (784, 828, 1316, and 1458 cm^{-1}), and carbohydrates (881, 944, 1043, and 1085 cm^{-1}) emerged as key molecular signatures for differentiating live lung-derived human cells based on their biochemical composition, emphasizing the potential of this method for accurate cell identification in lung research. In another study, A. Bresci and co-authors combined Raman spectroscopy with tomographic phase microscopy (TPM) to effectively identify and distinguish live colon cancer cells at various stages of disease progression [103]. The experimental design is illustrated in **Figure 6B**. The researchers

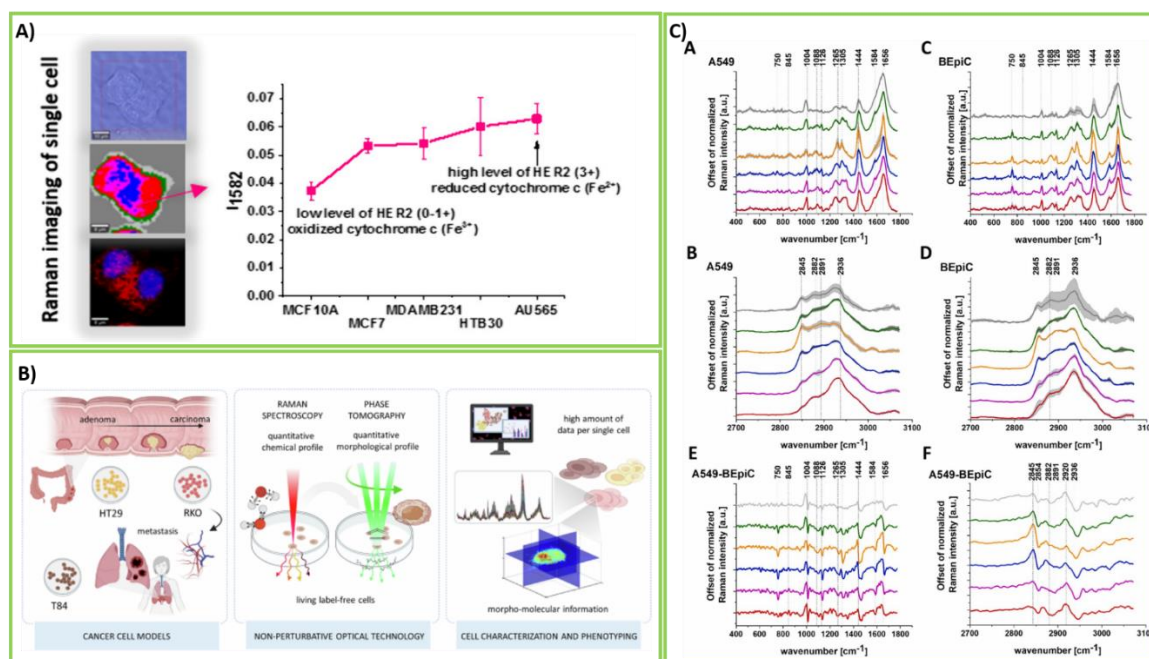


Figure 6. Several Raman strategies for *in vitro* applications. A) Workflow for HER2 protein identification by Raman spectroscopy and Raman imaging, from Ref. [100]. B) Methodological approach combining Raman spectroscopy and Tomographic Phase Microscopy (TPM) to profile and distinguish human colon cancer cells, from Ref. [103]. C) Raman analysis of human lung cancer cells (A549) and normal bronchial cells (BEpIC) in both the fingerprint (A and C, respectively) and high-frequency regions (B and D, respectively). The figure also presents differential spectra (A549 vs. BEpIC) for distinct cellular structures—nucleus, mitochondria, endoplasmic reticulum, lipid droplets, cytoplasm, and cell membrane—in both fingerprint and high wavelength regions (E and F, respectively), highlighting the biochemical differences between cancerous and healthy cells, from Ref. [104].

investigated three distinct cell lines—HT29 (colon adenoma), RKO (colon carcinoma), and T84 (lung metastasis of colon carcinoma)—which have similar appearances but differ in malignancy. RS offered spatially resolved chemical insights into the main cellular constituents, revealing molecular fingerprints of specific intracellular regions. In addition, TPM supplied three-dimensional structural data, such as refractive index maps, including volume, surface area, sphericity, and dry mass. This combined methodology enabled precise, label-free phenotyping of living cancer cell subtypes using a small number of unstained, unaltered live cells. The Raman data highlighted biochemical differences in nucleic acids, proteins, and lipids, while TPM revealed structural variations such as nuclear density and volume. Furthermore, the study used principal component analysis – multiple linear regression to analyze Raman spectra from colon cancer cells, identifying three key spectral components (PC2, PC5, and PC7) that differentiate cell types. Differences in nucleic acid, protein, and lipid content—especially RNA and DNA—were found to be the main chemical markers. The study showed that integrating chemical and morphological data greatly enhances the ability to distinguish cancer cell types. Through multinomial logistic regression modeling, it was found that although RS and TPM can each independently classify cell types, the most accurate and efficient results (100% accuracy) were obtained by combining just two critical features: PC5 from RS and dry mass (D) from TPM. Overall, the integrated RS-TPM approach offers a fast, precise, and minimally invasive tool for cancer cell phenotyping, with strong implications for clinical diagnostics.

In a different approach, Kopec *et al.* used Raman spectroscopy, imaging, and chemometric analysis to identify biochemical differences between normal bronchial epithelial cells

(BEpiC) and lung cancer cells (A549), aiming to discover biomarkers that could help distinguish these cell types and improve understanding of lung cancer progression [104]. Raman spectra and maps are collected and examined using cluster analysis to identify distinct cellular regions (nucleus, mitochondria, endoplasmic reticulum, lipid droplets, cytoplasm and cell membrane) according to their spectral profiles. The study revealed several Raman peaks that may serve as biomarkers to distinguish between normal and cancerous lung cells, as reported in **Figure 6C**. Notably, cancer cells showed increased phosphorylated tyrosine kinase, reflected by a peak at 845 cm^{-1} , and elevated triglyceride synthesis, indicated by a peak at 2845 cm^{-1} . In contrast, reduced cholesterol levels were associated with a band at 1444 cm^{-1} , and lower concentrations of cytochrome c were identified by peaks at 750 cm^{-1} . These results demonstrate the strength of Raman spectroscopy, especially when combined with chemometric techniques like PLS-DA, as a powerful approach for tracking biochemical variations in lung cells. The identified biomarkers reflect significant metabolic and molecular alterations in cancer cells, highlighting the potential of Raman spectroscopy to deepen our understanding of lung cancer at the molecular level. The study of Xia et al. combined Raman spectroscopy with traditional ML models (PLS-DA, SVM) and deep learning models (CNN, SSA-CNN) to discriminate and classify between normal (Hs 578Bst) and five different cancer cell lines (liver cancer cell line, BEL-7404; cervical cancer cell line HeLa; non-small cell lung cancer cell line, A549; human lung squamous cancer cell line, H226; breast cancer cell line, SKBR3) [105]. The study found that all four classification methods—PLS-DA, SVM, CNN, and SSA-CNN—could effectively identify differences between normal and cancerous cells based on Raman spectroscopy data, with SSA-CNN delivering the most accurate results overall. Although models trained on the full spectral range showed a slight improvement in performance, those that focused exclusively on the fingerprint region achieved nearly equivalent accuracy, emphasizing its diagnostic importance in cell classification. In particular, SSA-CNN achieved remarkably high accuracy and consistently outperformed the other models in both two-class (normal vs. one cancer cell line) and multi-class (normal vs. five cancer cell lines) classification tasks. Specifically, SSA-CNN reached the highest level of accuracy, nearing 95%, with a standard error not exceeding 3%, for both the entire spectra and the fingerprint region. In addition, Gradient-weighted class activation mapping (Grad-CAM) was applied to highlight important Raman spectral features, enabling the visualization of key biomolecules such as glucose, proteins, lipids, and nucleic acids. These spectral peaks, which varied between cell lines, reflected unique biochemical signatures associated with each cell type. This analysis emphasized the strength of the SSA-CNN model, not only in accurately classifying cancer cells but also in identifying potential new biomarkers. Overall, the study demonstrated how combining deep learning with spectral visualization offers a promising tool for advancing cancer detection and understanding cellular characteristics. The study conducted by Chen *et al.* investigated the use of single-cell Raman spectroscopy (SCRS) to distinguish leukemic cells from healthy white blood cells and to identify potential biomarkers for acute leukemia [106]. Among healthy cells from three donors, granulocytes, monocytes, B cells, and T cells showed distinct biochemical signatures, with orthogonal partial least squares discriminant analysis (OPLS-DA) achieving 100% classification accuracy. Similarly, leukemic subtypes—myeloblasts, abnormal

promyelocytes, monoblasts, and B-ALL cells—were accurately characterized by Raman spectroscopy. In this case, OPLS-DA effectively distinguished all four subtypes with high area under the curve (AUC) values (≥ 0.9896) and over 90% accuracy. Models were constructed to differentiate leukemic cells from their normal counterparts within the same cellular lineage, elucidating distinct spectral variations, particularly in protein-rich regions at 1003, 1341, and 1579 cm^{-1} . All comparative analyses, encompassing myeloblasts versus granulocytes, exhibited remarkable diagnostic precision, with most models achieving sensitivity and specificity levels exceeding 95%, and select models, such as the abnormal promyelocyte model, approaching optimal performance. Critical discriminatory Raman peaks were identified, with the 1003 cm^{-1} (phenylalanine) and 1341 cm^{-1} (DNA/RNA) peaks recurrently observed across models, while the 1579 cm^{-1} peak, linked to myeloperoxidase (MPO) and porphyrins, was significantly elevated in leukemic granulocytes. This 1579 cm^{-1} peak underwent further validation through semi-quantitative analysis, MPO-specific staining, and transmission electron microscopy (TEM) imaging, substantiating its function as a reliable biomarker for leukemic transformation in granulocytic cells. These results highlight the strong diagnostic potential of Raman spectroscopy for identifying leukemic cell types and characterizing disease-related biochemical changes.

4.2 Raman spectroscopy *ex vivo*

The study carried out by Li *et al.* used Raman spectroscopy to distinguish soft tissue sarcomas from surrounding healthy tissues, such as muscle, dermis, and fat, in an *ex vivo* setting [107]. The research aims to eventually develop a handheld Raman probe for real-time *in vivo* evaluation of tumor margins. By collecting an extensive dataset of over 120,000 Raman spectra from 11 frozen tissue samples collected during surgical procedures from three patients, the researchers successfully identified unique molecular signatures specific to each type of tissue. Significantly, tumor tissues did not exhibit the characteristic Raman peaks at 662, 744, and 752 cm^{-1} , which were found in benign tissues. In contrast, adipose tissue showed lower intensity for the Raman peak at 1002 cm^{-1} . The dermis and fat displayed three distinct peaks at 1343, 1364, and 1396 cm^{-1} , while muscle exhibited a stronger peak at 1343 cm^{-1} relative to the peaks at 1364 and 1396 cm^{-1} . Tumor tissue, however, lacked the peaks at 1364 and 1396 cm^{-1} . The investigation employed two complementary classification strategies: a quantitative approach based on wavenumber thresholding and advanced machine learning models, including a one-dimensional artificial neural network that analyzed spectral intensity only, and a more sophisticated two-dimensional convolutional neural network that incorporated both spectral intensity and wavenumber data. The quantitative approach achieved 83.0% accuracy with high sensitivity (93.3%) and specificity (90.6%). The 1D ANN reached 85.8% accuracy using averaged spectra but showed reduced performance across different patients. The 2D CNN outperformed all other methods, achieving an accuracy of 85.8%, sensitivity of 94.9%, and specificity of 88.3%. Combining the outputs of these methods increased accuracy to 97.6%, significantly improving confidence in tissue classification. The integration of handheld Raman probes, which can efficiently scan larger tissue areas intraoperatively, has been proposed as a significant and time-effective alternative to traditional frozen section analysis. This study strongly supports the feasibility of Raman spectroscopy as a rapid, label-free, and reliable intraoperative

diagnostic tool, with the potential to transform soft tissue sarcoma surgeries by guiding surgeons toward complete tumor excision and reducing the risk of local recurrence. In another study, Huang *et al.* reported an extensive analysis of Raman spectroscopy integrated with deep learning techniques for precise, non-invasive, real-time detection of liver cancer [108]. **Figure.7A** presents a comprehensive overview of the workflow associated with the Raman-based methodology combined with the AI system to achieve this goal. The research involved collecting Raman spectra from tissue samples obtained from 120 patients diagnosed with liver cancer, specifically examining both hepatocellular carcinoma (HCC) and the surrounding non-tumor tissues. The Raman analysis was also extended to the liver cancer subtype intrahepatic cholangiocarcinoma (ICC), as well as tumor stages and cellular differentiation grades. The Raman analysis revealed that ICC tissues displayed more pronounced carotenoid-associated Raman peaks, while HCC tissues exhibited stronger signals linked to amino acids, lipids, and nucleic acids. Tumors in early stages showed increased intensity in carotenoid and tryptophan peaks, while those in advanced stages had higher signals corresponding to nucleic acids and saccharides. Tumors with better differentiation generally demonstrated stronger overall spectral intensity, especially in carotenoid-related regions. These findings were validated through histopathological confirmation using H&E staining. A convolutional neural network (CNN) was then developed using a deep learning system based on a modified VGG-16 architecture. Trained on a dataset of 12,000 Raman spectra collected from 120 pairs of liver tissue samples, the results showed a remarkable diagnostic accuracy of 92.6%, with sensitivity and specificity values of 90.8% and 94.6%, respectively, highlighting the ability to differentiate between cancerous and non-cancerous tissues. Moreover, three other CNN classification models were applied to identify liver cancer subtypes—HCC and ICC—as well as tumor stages and cellular differentiation grades. The accuracy rates for stage and grade classification were lower, at 78.3% and 72.3%, respectively; however, subtype classification demonstrated a robust performance with an accuracy of 82.4%. A tissue metabolomics analysis using liquid chromatography coupled with mass spectrometry (LC-MS) confirmed the findings from Raman spectroscopy, revealing a decrease in essential metabolites including amino acids, lipids, and nucleic acids in cancerous tissues. Raman imaging was also used to generate two- and three-dimensional maps of liver tissues. By applying self-modelling curve resolution (SMCR) and hierarchical cluster analysis (HCA), the researchers depicted the spatial distribution of proteins, lipids, and various biochemical constituents with sub-micrometric precision. These visualizations revealed tumor margins and internal variability that conventional hematoxylin and eosin (H&E) staining techniques often fail to capture. The feasibility of intraoperative diagnosis was demonstrated using a handheld portable Raman system. Although there were slight differences in spectral quality when compared to the confocal Raman system, the portable device similarly assessed the reduced spectral intensities in tumor tissues compared to normal tissues. The ability to distinguish these tissue types suggests its potential for real-time tumor identification during surgery. Consequently, the combination of Raman spectroscopy and artificial intelligence presents considerable promise as an accurate diagnostic tool for liver cancer. Recently, Karnachoriti *et al.* employed Raman spectroscopy combined with machine learning to assess colorectal cancer (CRC) tumor margins [109]. Two Raman platforms were used: a high-resolution micro-

Raman spectrometer for detailed *ex vivo* analysis and a portable Raman system with a fiber optic probe for real-time simulation during intraoperative scenarios. Raman spectra were acquired from four spatially defined regions, tumor core, adjacent margins, intermediate tissue, and distal normal tissue, using samples from 23 CRC patients. After preprocessing, datasets underwent initial exploration via PCA, followed by classification modeling with Partial Least-Squares Discriminant Analysis and eXtreme Gradient Boosting (XGBoost) machine learning techniques. The entire Raman-based method, combined with the machine learning approaches, is illustrated in **Figure 7B**. This study demonstrated the effectiveness of Raman spectroscopy combined with machine learning in distinguishing cancerous, margin and healthy colorectal tissues. Cancer samples showed increased protein signals and decreased collagen and lipid markers, while normal tissues had stronger lipid features. Tumor margins closely resembled cancerous tissue, making them harder to distinguish. Despite this, classification models as PLS-DA and XGBoost achieved high accuracy—86% for cancer vs. normal, 82% for margin vs. normal, and 67% for cancer vs. margin. Metastatic cancer samples showed clearer differentiation at tumor boundaries, improving classification performance. A portable Raman device also proved highly accurate in real-time tissue analysis during surgery, reaching 97% accuracy. Overall, the study supported Raman spectroscopy as a reliable, non-invasive tool for cancer detection and surgical guidance, with further development needed to improve margin classification.

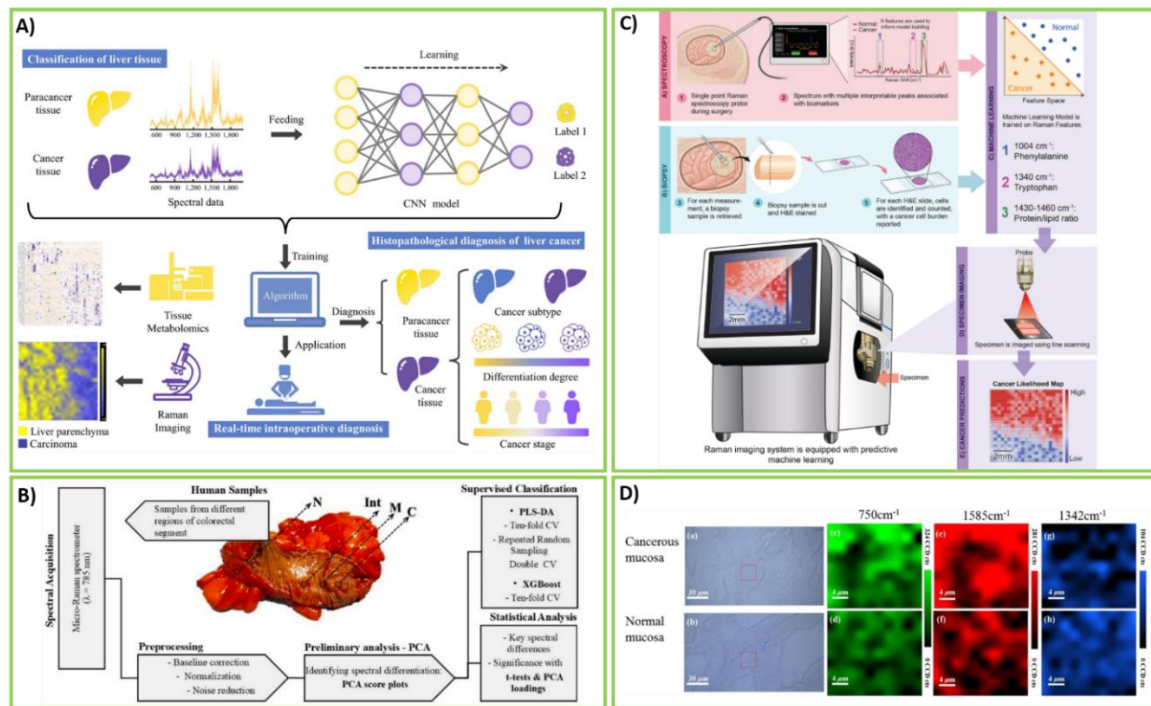


Figure 7. Various Raman spectroscopy strategies employed for *ex vivo* use. A) Graphical representation of the histopathological diagnosis approach for liver cancer utilizing Raman spectroscopy and a convolutional neural network model, from Ref. [108]. B) Workflow of the combination of Raman spectroscopy with a machine learning approach for colorectal tumor margin assessment, from Ref. [109]. C) Graphical illustration of the process for generating cancer prediction maps from glioblastoma specimens imaged with Raman spectroscopy, from Ref. [110]. D) Raman images comparing cancerous and adjacent normal tissue at specific wavelengths (750, 1585 and 1342 cm^{-1}), alongside optical microscope images, from Ref. [111].

Zhu *et al.* introduced an accurate and innovative methodology for detecting glioblastoma infiltration at the cellular level using Raman spectroscopy and machine learning [88]. GBM is an aggressive brain tumor known for its extensive infiltration into adjacent brain parenchyma, making complete surgical excision particularly challenging. Raman analysis differentiated infiltrative murine tumor tissue from normal brain tissue by recognizing unique spectral characteristics, including lower phospholipid-related Raman peaks in tumor zones and higher intensities for nucleic acids and unsaturated fatty acids, indicating the altered metabolic state of cancer cells. These spectral distinctions were confirmed by air flow-assisted desorption electrospray ionization mass spectrometry imaging (AFADESI-MSI) for metabolomics analysis. Raman imaging enabled spatial delineation of biochemical alterations within the tissue, demonstrating a strong correspondence with histopathological characteristics observed in H&E-stained preparations. Notably, the research revealed a robust correlation between Raman spectral intensity and tumor cell density, with specific peaks increasing or decreasing proportionally with the density of GBM cells in a defined tissue area. Machine learning algorithms based on support vector machines were constructed utilizing Raman spectra from murine and human samples. These algorithms achieved remarkable accuracy in differentiating tumor-infiltrated from normal tissue, with AUC values exceeding 95%. A key finding was that Raman spectroscopy could consistently detect glioblastoma infiltration with remarkable sensitivity, identifying as few as three tumor cells within an area of 0.01 mm² in human samples. As the number of tumor cells increased, detection accuracy improved, reaching 100% in areas with at least three cells. This high level of precision makes RS a highly effective tool for defining tumor boundaries during surgery, potentially leading to more complete tumor removal and reduced risk of recurrence. In another investigation, Daoust *et al.* reported the development and preliminary evaluation of a Raman spectroscopy imaging system designed for intraoperative detection of cancer cells at the margins of glioblastoma specimens [110]. The workflow of the entire experimental design is reported in **Figure 7C**. The system features a large field of view of 1 cm² with a spatial resolution of 0.25 mm per pixel, enabling hyperspectral Raman data collection over the specimen surface. A machine learning model, initially trained on in vivo single-point probe data from 24 glioblastoma patients, was employed to analyze spectral features linked to molecular vibrations of phenylalanine, tryptophan, and lipid/protein ratios to identify cancerous tissue. The detection model, based on support vector machines, achieved 90% sensitivity and 95% specificity during cross-validation. It was then applied to nine glioblastoma specimens from five new patients, with subsequent histopathological analysis allowing direct comparison between Raman-derived cancer likelihood maps and histological analysis. The results demonstrated strong concordance, with tumor tissue showing elevated signals at Raman bands associated with phenylalanine (1004 cm⁻¹), tryptophan and collagen (1340 cm⁻¹), alongside a decreased lipid-to-protein ratio in the 1430–1460 cm⁻¹ spectral region. There were no significant differences between tumor core and tumor interphase in the Raman bands used for cancer detection. However, variations were observed in regions linked to phospholipids and nucleic acids. Necrotic tissues displayed higher signals from lipids and proteins and lower signals associated with tryptophan and amino acids. The system successfully detected cancer even in specimens containing necrosis or low cellularity at tumor edges. Pure tumor areas corresponded to high cancer likelihood values above 0.95,

while normal or infiltrated brain tissue exhibited lower values around 0.3 to 0.5, consistent with histological findings. Although Raman spectra from the imaging system and the original point probe were generally consistent, minor systematic differences were noted, particularly near protein bands, likely reflecting differences between *in vivo* and *ex vivo* conditions. Future improvements are needed to enhance light collection efficiency, reduce acquisition time and develop more sophisticated machine learning models to detect pathological features such as invasive tumor margins.

In another investigation, Li *et al.* combined Raman spectroscopy with a deep learning architecture called SENet-LSTM to detect gastric adenocarcinoma (GAC) [111]. Tissue samples from 28 patients were analyzed, yielding a total of 6480 Raman spectra from both cancerous and adjacent healthy gastric tissue. Raman images of cancerous and normal tissues were captured at specific wavelengths of 750, 1585, and 1342 cm^{-1} , as highlighted in **Figure 7D**. The spectral data revealed key molecular differences between the two tissue types, such as increased signals related to cytochrome c and decreased signals associated with compounds like phenylalanine, lipids, collagen, and proteins, indicating cancer-related alterations in cancer. The deep learning model performed exceptionally well, achieving a classification accuracy of 96.20%, sensitivity of 96.48%, specificity of 95.90% and an AUC score of 0.99, demonstrating its reliability in distinguishing healthy from cancerous tissue. To identify the most important spectral features for predictions of the model, a masking technique was applied. This analysis showed that spectral regions between 800–1300 cm^{-1} were key for identifying healthy tissue, while those between 1315–1400 cm^{-1} were most relevant for detecting cancer. These findings matched known chemical markers of gastric cancer, confirming the model's interpretability and relevance, highlighting that the synergy between Raman spectroscopy and deep learning can serve as a powerful tool for identifying gastric cancer with strong potential for clinical use.

Kopec *et al.* investigated the application of Raman spectroscopy and imaging, combined with partial least square discriminant analysis, to differentiate among various types of human brain tumors and breast cancer [112]. Their analysis of tissue samples, including brain tumors with different malignancy levels such as gliosarcoma, oligodendroglioma, meningioma, neurofibroma, and invasive ductal carcinoma of the breast, revealed that Raman imaging provides spatially resolved insights into the biochemical composition of cancerous tissues. The study found that mildly malignant tumors exhibited strong signals from lipids and carotenoids, whereas highly malignant tumors and breast cancers had lower lipid content but higher cytochrome c levels, indicating increased mitochondrial activity. In addition, Raman imaging revealed a considerable heterogeneity in low-grade tumors, in contrast to the greater uniformity and protein-rich composition of high-grade tumors. Significantly, the ratio of cytochrome c (1585 cm^{-1}) to lipids (1444 cm^{-1}) was markedly elevated in malignant tissues. Conversely, the ratios associated with carotenoids (1520/1585 and 1156/1585) were much higher in less malignant tumors, supporting the observation that antioxidant compounds such as carotenoids are reduced in more aggressive cancers. PLS-DA effectively distinguished between different tumor types based on their spectral markers, achieving especially high accuracy for high-grade tumors. Anaplastic oligodendroglioma showed 100% sensitivity and almost 99% specificity, while breast cancer also demonstrated strong diagnostic performance. Even in cases with overlapping spectral features, PLS-DA

enabled clear statistical separation, underscoring the potential of Raman spectroscopy and PLS-DA for cancer diagnostics.

4.3 Raman spectroscopy *in vivo*

The study by Dulude *et al.* evaluated the UltraProbe system, an innovative handheld point Raman probe designed for the rapid, intraoperative identification of soft tissue sarcomas (STS) and their differentiation from normal tissue [113]. Thirty patients with soft tissue sarcomas were recruited, and a Raman probe was used to assess both healthy and tumor tissues. The investigation included different healthy tissue types, such as adipose and muscle tissues analyzed *in vivo*, as well as fibrous and necrotic tissues and multiple sarcoma subtypes *ex vivo*, including well-differentiated liposarcoma (WDLPS), dedifferentiated liposarcoma (DDLPS), leiomyosarcoma (LMS), undifferentiated pleomorphic sarcoma (UPS), solitary fibrous tumor (SFT) and myxofibrosarcoma (MFS). The UltraProbe operated with a 785 nm laser and featured a 21 multi-mode detection fiber optic system, which significantly increased the speed of data acquisition while preserving signal integrity. Machine learning models, particularly Random Forest classifiers, were used to classify tissue samples based on their Raman spectra by grouping tissues according to biologically relevant characteristics with clinical significance. These models demonstrated strong capability in distinguishing liposarcomas (WDLPS and DDLPS) from healthy adipose tissue, achieving high accuracy of 90%, and even higher accuracy of 94% when identifying well-differentiated liposarcoma, which closely resembles healthy adipose tissue. The models also effectively differentiated non-liposarcoma soft tissue sarcomas (LMS, UPS, SFT, MFS) from non-adipose, non-sarcoma tissues such as muscle, fibrotic, and necrotic tissue, with an accuracy of 87%, which is crucial for accurate diagnosis and treatment. The device achieved a mean acquisition time of 276 milliseconds per spectrum, allowing for near real-time application in surgical settings. The research also showed that spectral data from *ex vivo* tissue remained consistent with *in vivo* measurements for up to one hour, supporting its practical use in operating rooms. Although limited by a relatively small and heterogeneous patient cohort and focusing mainly on point measurements rather than comprehensive margin scanning, the UltraProbe showed considerable promise as a real-time, non-invasive diagnostic tool. A study by Ember *et al.* evaluated the Sentry System, a handheld device integrating Raman spectroscopy and machine learning for real-time, dye-free brain tumor detection during surgery [114]. The experimental process design steps were reported in **Figure 8A**. The multicenter trial included 67 patients undergoing resection for glioblastoma, brain metastases, or meningioma, with 976 intraoperative spectral measurements correlated to histopathological biopsy results. The system demonstrated high diagnostic accuracy: 91% for glioblastoma, 97% for metastases, and 96% for meningiomas, with a combined accuracy of 90% across all tumor types. Key Raman spectral features distinguishing tumor tissue from normal brain included differences in proteins and lipids, particularly at 1004 cm^{-1} (phenylalanine), 1299 cm^{-1} (lipids), 1340 cm^{-1} (tryptophan), and 1441 cm^{-1} (a lipid and protein combination). Tumor tissues showed lower lipid content and altered protein signatures compared to healthy brain tissue, consistent with established biological patterns. Compared to traditional intraoperative methods such as MRI or 5-Aminolevulinic Acid (5-ALA) fluorescence, the Sentry System offered a faster, more cost-effective, and reagent-free

approach for real-time tissue assessment during surgery, representing a promising technology for improving the precision of brain tumor resections.

Recently, Y. Liu *et al* explored a new surgical technique that combined real-time Raman spectroscopy with cryoablation to improve the detection and treatment of bladder cancer (BCa) *in vivo* [115]. Standard treatment through transurethral resection of bladder tumor (TURBT) often fails to detect all cancerous tissue, leading to a high rate of recurrence. To address this, a fiber-optic Raman probe that can be inserted during surgery to identify residual cancer was tested, followed by cryoablation to eliminate any remaining tumor cells. Raman spectroscopy combined with discriminant analysis (DA) based on leave-one-out cross-validation, successfully identified biochemical markers unique to bladder cancer and differentiated between invasive and noninvasive forms with high sensitivity (90.5%) and specificity (80.8%). Moreover, the approach was able to discriminate between invasive (T1, T2) and noninvasive (Ta) BCa with a sensitivity and specificity of 83.3% and 87.3%, respectively. In particular, bladder cancer showed intense spectral peaks at 607 cm^{-1} , 677 cm^{-1} , 772 cm^{-1} , 858 cm^{-1} , 1,415 cm^{-1} , 1,582 cm^{-1} , 1,698 cm^{-1} , 1,840 cm^{-1} , and 2,100 cm^{-1} ,

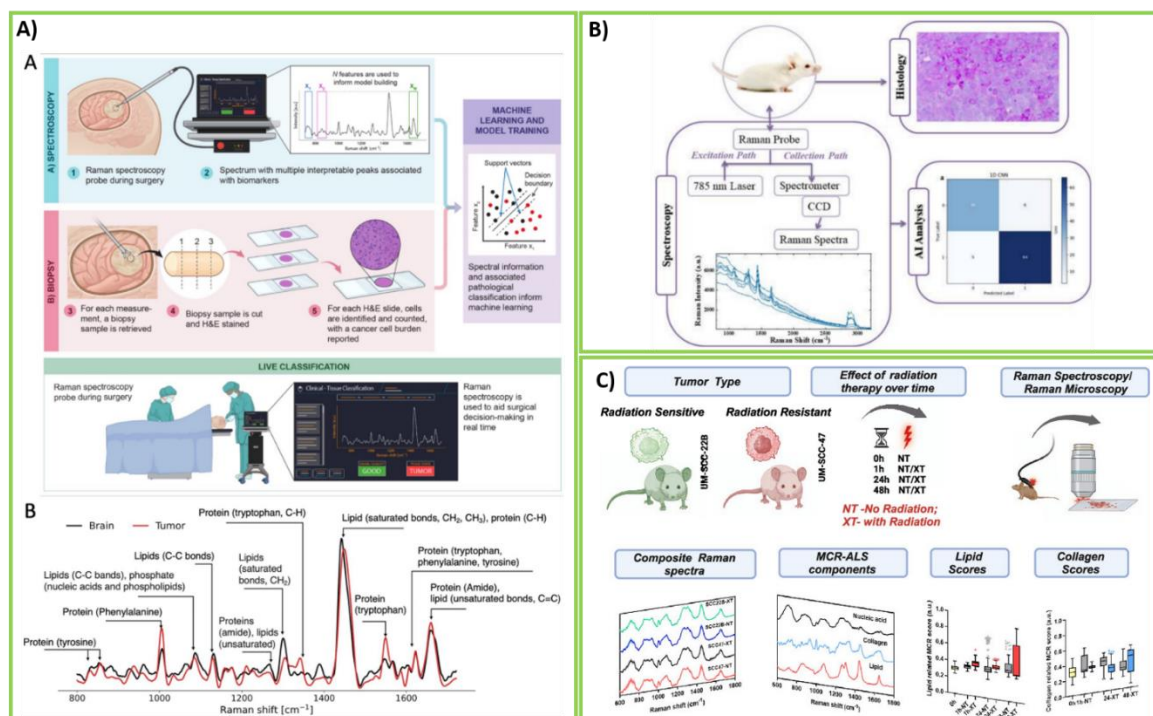


Figure 8. Different *in vivo* Raman approaches for enhanced cancer diagnosis. A) Workflow outlining the process of brain tumor identification using the Sentry System, from Ref. [114]. B) Diagram of *in vivo* mouse model analysis combining histology, Raman spectroscopy, and AI-based data interpretation, from Ref. [116]. C) Study overview diagram of the analysis of head and neck tumor biomolecular changes after radiation therapy by Raman spectroscopy in an *in vivo* mouse model, from Ref. [117].

corresponding mainly to molecular vibrations associated with DNA and proteins. Of 127 patients, data from 74 were used to build a Raman dataset derived from cancerous and healthy tissues. The remaining 53 patients were split into two groups: one received standard TURBT, and the other underwent TURBT with additional Raman scanning and cryoablation if needed. Among the Raman-cryoablation group, 7 patients had residual tumors detected and were treated with cryoablation. Only 2 of these patients experienced recurrence, compared to 8 in the TURBT-only group, a statistically significant reduction ($p = 0.0394$).

The method provides an objective, intraoperative assessment of tumor margins and enables precise treatment of remaining cancer without the risks of further surgical resection. However, despite promising results, the approach has some drawbacks, such as time-intensive data processing, limited flexibility of the Raman probe, and uncertainty about long-term safety. The combined technique could significantly improve bladder cancer management but requires further validation through larger, multicenter trials. Recently, Latka *et al.* developed “*invaScope*”, an innovative endoscopic system that leverages Raman spectroscopy to enable real-time, *in vivo* detection and grading of bladder cancer during surgeries like transurethral resection of bladder tumors (TURBT) [118]. Developed to meet the rigorous standards of the European Medical Device Regulation (MDR 2017/745), the system includes a sterilizable fiber optic probe, a Raman spectrometer compatible with standard endoscopes, and dedicated control software (*invaSoft*) to support clinical use through guided setup and calibration. During testing on 21 patients undergoing tumor resections, Raman spectra were collected from both healthy and suspicious bladder tissues. Of 800 total spectra, 438 were suitable for analysis, with the remainder excluded due to technical issues such as interference from surgical lighting or photoreactive agents like Hexvix. Using PCA-LDA, the system successfully differentiated cancerous from non-cancerous tissue with an overall accuracy of 83%. Early results also suggested the system may distinguish between low-grade and high-grade tumors, although further data will be needed to confirm this capability. This study represents a major step toward using Raman spectroscopy in surgery, demonstrating its technical and regulatory readiness and paving the way for future clinical trials and real-time cancer diagnostics such as optical biopsies and Raman imaging.

In another study, Shu *et al.* introduced a method for real-time, *in vivo* staging of nasopharyngeal carcinoma (NPC) using fiber optic Raman spectroscopy during endoscopy [119]. A total of 5205 Raman spectra were collected from 278 distinct tissue sites across a cohort of 111 subjects, including 50 healthy volunteers and 61 patients diagnosed with various stages of NPC. The spectral data revealed biochemical difference between healthy tissue and cancerous tissue, as well as between early and late stages of NPC. In particular, healthy nasopharyngeal tissue has a significantly lower protein-to-lipid ratio compared to all stages of NPC. This ratio decreases further from early to late stages of NPC, indicating increased lipid accumulation in advanced cancer. For cancer staging, a partial-least-squares linear-discriminant analysis (PLS-LDA) model combined with leave-one-tissue-site-out cross-validation (LOOCV) was developed using the collected spectral data, achieving a classification accuracy of 92.81% for NPC staging. Additionally, a biomolecular Raman model utilizing non-negative linear regression (NNLR) was employed to identify significant biochemical markers in NPC tissues and determine the key biomolecules associated with various NPC stages. This analysis highlighted chondroitin sulfate, glucose, hemoglobin, oleic acid, and triolein as major contributors to the observed spectral changes. These findings demonstrate that fiber optic Raman endoscopy is a highly effective diagnostic tool for real-time, non-invasive cancer staging that provides detailed biochemical insights.

In another research by Kouri *et al.* used a portable Raman probe in combination with a transfer learning model based on a one-dimensional convolutional neural network to characterize and discriminate between control and colorectal tumor animal models using

severe combined immuno-deficient mice (SCID) mice injected with human colorectal cancer cells (HCT 116) to create xenograft tumors, as described in the workflow depicted in **Figure 8B** [116]. The spectral analysis showed that healthy tissues have a greater lipid concentration, whereas malignant tissues exhibit elevated levels of proteins and amino acids, as evidenced by variations in Raman spectral bands. Principal component analysis revealed that significant Raman shifts at 1004, 1657, and 2937 cm^{-1} facilitate differentiation between cancerous and normal tissues. Furthermore, the 1D-CNN employing transfer learning successfully categorized tissue types with an accuracy of 89.9%, demonstrating its efficacy in distinguish cancer tissue from normal one. This approach shows strong potential for clinical implementation, particularly in real-time cancer identification and surgical margin evaluation, setting the stage for future human trials and integration into medical diagnostic technologies.

Additionally, research by Karunakaran *et al.* investigated Raman spectroscopy's ability to detect early biochemical changes in murine tumor models following radiation therapy, comparing radiation-sensitive (UM-SCC-22B) and resistant (UM-SCC-47) head and neck cancer xenografts [117]. The Raman approach used is depicted in **Figure 8C**. A single 2 Gy radiation dose was administrated and *in vivo* Raman spectroscopy measurements were performed at 0, 24 and 48 hours' post-treatment. Spectral analysis using multivariate curve resolution-alternating least squares (MCR-ALS) revealed divergent biomolecular responses between tumor types. In radiation-sensitive tumors, lipid-like spectral contributions increased significantly by 48 hours ($p = 0.0291$), while radiation-resistant tumors displayed a marked increase in collagen-like signals at 24 hours ($p = 0.0125$) compared to the non-irradiated state. *Ex vivo* confocal Raman imaging of tumor sections partially validated these trends, though time-dependent changes were less consistent due to methodological variations in sampling depth and tissue preparation. The findings highlight Raman spectroscopy's capacity to noninvasively identify molecular signatures of treatment response, with lipid and collagen profiles emerging as potential biomarkers for early differentiation of radiation-sensitive and resistant tumors.

5. Challenges in Raman spectroscopy

Raman spectroscopy shows significant potential for the detection of cancer; however, several obstacles remain, particularly in identifying or differentiating specific signals. These challenges may arise from a substance's inherently low Raman activity, the emission of weak Raman signals, or interference caused by strong fluorescence, which can obscure spectral features [120]. Weak signals and fluorescence can often be mitigated using techniques such as Surface-Enhanced Raman Spectroscopy (SERS) or by employing excitation lasers with longer wavelengths [121]. Additionally, because Raman scattering is an inherently rare event—occurring in roughly one out of 10^6 photons—its sensitivity for detecting trace substances or contaminants is limited unless signal-enhancing methods like SERS or beam-enhancement technologies are employed [45]. Other critical challenges include the need for standardized databases and the development of evaluative systems for cancer-associated spectra to reduce inconsistencies arising from different spectrometers and improve the quality and efficiency of spectral acquisition [122]. Variations in laser parameters, integration time and sample preparation significantly affect data consistency,

highlighting the need for standardized protocols and advanced data collection methods [45]. Current data processing techniques are often inadequate for handling the complexity and volume of Raman data. More robust algorithms, potentially deep learning, are being investigated to improve analytical precision and may ultimately assist physicians in cancer diagnosis [123], [124]. Overall, Raman spectroscopy is emerging as a valuable tool for cancer diagnosis, with increasing potential in clinical applications.

6. Conclusions

Raman spectroscopy has emerged as a powerful, label-free, and non-invasive diagnostic technique capable of addressing several limitations of conventional cancer diagnostics. By exploiting the molecular specificity of inelastic light scattering, Raman spectroscopy enables detailed biochemical profiling of cells and tissues, allowing early detection, accurate classification, and real-time monitoring of cancerous transformations. Its compatibility with various biological samples - including cells and tissues - and minimal sample preparation, makes it particularly attractive for clinical translation. The integration of advanced data analysis techniques, particularly machine learning and deep learning algorithms, has further enhanced the diagnostic potential of Raman spectroscopy. These computational tools help extract subtle spectral differences and patterns, enabling accurate discrimination between normal and cancerous tissues, cancer subtypes, and even stages of tumor progression. The combination of Raman spectroscopy with artificial intelligence paves the way for precision oncology, offering a foundation for more personalized, timely, and less invasive diagnostics. Despite current challenges such as weak signal strength and interference from tissue auto-fluorescence, ongoing advancements in instrumentation—such as fiber-optic probes and portable Raman systems—are enabling in vivo, real-time applications. Moreover, the continuous evolution of spectral analysis algorithms and validation strategies is steadily improving robustness, reproducibility, and clinical accuracy. In conclusion, Raman spectroscopy, especially when combined with machine learning, holds immense promise as a next-generation tool for cancer diagnostics. With further refinement, standardization, and clinical validation, it has the potential to revolutionize diagnostic pathways and enhance patient outcomes across a broad spectrum of cancers.

7. References

- [1] D. A. Milner and J. K. Lennerz, "Technology and Future of Multi-Cancer Early Detection," *Life*, vol. 14, no. 7, pp. 1–11, 2024, doi: 10.3390/life14070833.
- [2] R. L. Siegel, T. B. Kratzer, A. N. Giaquinto, H. Sung, and A. Jemal, "Cancer statistics, 2025.," *CA. Cancer J. Clin.*, no. October 2024, pp. 10–45, 2025, doi: 10.3322/caac.21871.
- [3] C. Santucci *et al.*, "European cancer mortality predictions for the year 2025 with focus on breast cancer," *Ann. Oncol.*, vol. 35, no. 3, pp. 308–316, 2024, doi: 10.1016/j.annonc.2023.12.003.
- [4] L. Fass, "Imaging and cancer: A review," *Mol. Oncol.*, vol. 2, no. 2, pp. 115–152, 2008, doi: 10.1016/j.molonc.2008.04.001.
- [5] C. C. Tsai *et al.*, "Diagnostics and Therapy for Malignant Tumors," *Biomedicines*, vol. 12, no. 12, pp. 1–28, 2024, doi: 10.3390/biomedicines12122659.
- [6] P. F. Niederer, "Chapter 1: Diagnostic Medical Imaging: X-ray projection technique, image subtraction method, direct digital x-ray imaging, computed tomography (CT)," *Technol. Heal. Care*, vol. 17, no.

- 5–6, pp. 421–444, 2009, doi: 10.3233/THC-2009-0563.
- [7] S. Hussain *et al.*, “Modern Diagnostic Imaging Technique Applications and Risk Factors in the Medical Field: A Review,” *Biomed Res. Int.*, vol. 2022, 2022, doi: 10.1155/2022/5164970.
 - [8] J. Barentsz *et al.*, “Commonly used imaging techniques for diagnosis and staging,” *J. Clin. Oncol.*, vol. 24, no. 20, pp. 3234–3244, 2006, doi: 10.1200/JCO.2006.06.5946.
 - [9] G. Rongrong, L. Guolan, and F. Baowei, “Ultrasound Imaging Technology for breast cancer detection and management,” *Ultrasound Med. Biol.*, vol. 44, no. 1, pp. 37–70, 2018, doi: 10.1016/j.ultrasmedbio.2017.09.012.ULTRASOUND.
 - [10] A. Rembielak, M. Green, A. Saleem, and P. Price, “Diagnostic and therapeutic imaging in oncology,” *Med. (United Kingdom)*, vol. 44, no. 1, pp. 6–9, 2016, doi: 10.1016/j.mpmmed.2015.10.010.
 - [11] M. S. Linet *et al.*, “Cancer risks associated with external radiation from diagnostic imaging procedures,” *CA. Cancer J. Clin.*, vol. 62, no. 2, pp. 75–100, 2012, doi: 10.3322/caac.21132.
 - [12] J. Rong and Y. Liu, “Advances in medical imaging techniques,” *BMC Methods*, vol. 1, no. 1, pp. 10–12, 2024, doi: 10.1186/s44330-024-00010-7.
 - [13] L. Ma *et al.*, “Liquid biopsy in cancer current: status, challenges and future prospects,” *Signal Transduct. Target. Ther.*, vol. 9, no. 1, p. 336, 2024, doi: 10.1038/s41392-024-02021-w.
 - [14] R. Vaidyanathan, R. H. Soon, P. Zhang, K. Jiang, and C. T. Lim, “Cancer diagnosis: from tumor to liquid biopsy and beyond,” *Lab Chip*, vol. 19, no. 1, pp. 11–34, 2019, doi: 10.1039/c8lc00684a.
 - [15] P. P. Piccaluga, “Introductory Chapter: Molecular Diagnostic in Cancer,” in *Molecular Diagnostics of Cancer*, P. P. Piccaluga, Ed. IntechOpen, 2024. doi: 10.5772/intechopen.1004226.
 - [16] S. Chakraborty, G. Sharma, S. Karmakar, and S. Banerjee, “Multi-OMICS approaches in cancer biology: New era in cancer therapy,” *Biochim. Biophys. Acta - Mol. Basis Dis.*, vol. 1870, no. 5, p. 167120, 2024, doi: 10.1016/j.bbadis.2024.167120.
 - [17] Y. Zhou *et al.*, “Tumor biomarkers for diagnosis, prognosis and targeted therapy,” *Signal Transduct. Target. Ther.*, vol. 9, no. 1, 2024, doi: 10.1038/s41392-024-01823-2.
 - [18] R. R. Jones, D. C. Hooper, L. Zhang, D. Wolverson, and V. K. Valev, “Raman Techniques: Fundamentals and Frontiers,” *Nanoscale Res. Lett.*, vol. 14, 2019, doi: 10.1186/s11671-019-3039-2.
 - [19] M. Prochazka, “Basics of Raman Scattering (RS) Spectroscopy,” in *Surface-Enhanced Raman Spectroscopy: Bioanalytical, Biomolecular and Medical Applications*, Cham: Springer International Publishing, 2016, pp. 7–19. doi: 10.1007/978-3-319-23992-7_2.
 - [20] A. A. Fung and L. Shi, “Mammalian cell and tissue imaging using Raman and coherent Raman microscopy,” *Wiley Interdiscip. Rev. Syst. Biol. Med.*, vol. 12, no. 6, p. e1501, Nov. 2020, doi: 10.1002/wsbm.1501.
 - [21] D. Wang, P. He, Z. Wang, G. Li, N. Majed, and A. Z. Gu, “Advances in single cell Raman spectroscopy technologies for biological and environmental applications,” *Curr. Opin. Biotechnol.*, vol. 64, no. July, pp. 218–229, 2020, doi: 10.1016/j.copbio.2020.06.011.
 - [22] A. P. Mamede *et al.*, “A new look into cancer—a review on the contribution of vibrational spectroscopy on early diagnosis and surgery guidance,” *Cancers (Basel)*, vol. 13, no. 21, 2021, doi: 10.3390/cancers13215336.
 - [23] H. J. Butler *et al.*, “Using Raman spectroscopy to characterize biological materials,” *Nat. Protoc.*, vol. 11, no. 4, pp. 664–687, 2016, doi: 10.1038/nprot.2016.036.
 - [24] J. Popp *et al.*, “Label-free molecular imaging of biological cells and tissues by linear and non-linear Raman spectroscopic approaches,” *Angew. Chem. Int. Ed. Angew. Chem*, vol. 10, no. 10, 2016, [Online]. Available: <http://dx.doi.org/10.1002/anie.201607604>
<http://dx.doi.org/10.1002/ange.201607604>
 - [25] L. A. Austin, S. Osseiran, and C. L. Evans, “Raman technologies in cancer diagnostics,” *Analyst*, vol. 141, no. 2, pp. 476–503, 2016, doi: 10.1039/c5an01786f.

- [26] K. Kong, C. Kendall, N. Stone, and I. Notingher, "Raman spectroscopy for medical diagnostics - From in-vitro biofluid assays to in-vivo cancer detection," *Adv. Drug Deliv. Rev.*, vol. 89, pp. 121–134, 2015, doi: 10.1016/j.addr.2015.03.009.
- [27] D. W. Shipp, F. Sinjab, and I. Notingher, "Raman spectroscopy: techniques and applications in the life sciences," *Adv. Opt. Photonics*, vol. 9, no. 2, p. 315, 2017, doi: 10.1364/aop.9.000315.
- [28] J. Xu *et al.*, "Unveiling cancer metabolism through spontaneous and coherent raman spectroscopy and stable isotope probing," *Cancers (Basel)*, vol. 13, no. 7, pp. 1–24, 2021, doi: 10.3390/cancers13071718.
- [29] F. Lussier, V. Thibault, B. Charron, G. Q. Wallace, and J. F. Masson, "Deep learning and artificial intelligence methods for Raman and surface-enhanced Raman scattering," *TrAC - Trends Anal. Chem.*, vol. 124, 2020, doi: 10.1016/j.trac.2019.115796.
- [30] C. Esposito *et al.*, "Assessment of Primary Human Liver Cancer Cells by Artificial Intelligence-Assisted Raman Spectroscopy," *Cells*, vol. 12, no. 22, 2023, doi: 10.3390/cells12222645.
- [31] M. D'Acunto, R. Gaeta, R. Capanna, and A. Franchi, "Contribution of Raman Spectroscopy to Diagnosis and Grading of Chondrogenic Tumors," *Sci. Rep.*, vol. 10, no. 1, pp. 1–12, 2020, doi: 10.1038/s41598-020-58848-0.
- [32] M. Tang, L. Xia, D. Wei, S. Yan, C. Du, and H.-L. Cui, "Distinguishing Different Cancerous Human Cells by Raman Spectroscopy Based on Discriminant Analysis Methods," *Appl. Sci.*, vol. 7, no. 9, 2017, doi: 10.3390/app7090900.
- [33] B. B. Praveen, P. C. Ashok, M. Mazilu, A. Riches, S. Herrington, and K. Dholakia, "Fluorescence suppression using wavelength modulated Raman spectroscopy in fiber-probe-based tissue analysis," *J. Biomed. Opt.*, vol. 17, no. 7, p. 0770061, 2012, doi: 10.1117/1.jbo.17.7.077006.
- [34] A. Smekal, "Zur Quantentheorie der Dispersion," *Naturwissenschaften*, vol. 11, no. 43, pp. 873–875, 1923, doi: 10.1007/BF01576902.
- [35] C. V. Raman and K. S. Krishnan, "A new type of secondary radiation.," *Nature*, vol. 121, no. 3048, pp. 501–502, 1928, doi: 10.1038/121501c0.
- [36] "Introduction, Basic Theory and Principles," in *Modern Raman Spectroscopy – A Practical Approach*, John Wiley & Sons, Ltd, 2004, pp. 1–21. doi: <https://doi.org/10.1002/0470011831.ch1>.
- [37] D. Cialla-May, M. Schmitt, and J. Popp, "Theoretical principles of Raman spectroscopy," vol. 4, no. 6, 2019, doi: doi:10.1515/psr-2017-0040.
- [38] B. Dietzek, D. Cialla, M. Schmitt, and J. Popp, "Introduction to the Fundamentals of Raman Spectroscopy," in *Confocal Raman Microscopy*, J. Toporski, T. Dieing, and O. Hollricher, Eds. Cham: Springer International Publishing, 2018, pp. 47–68. doi: 10.1007/978-3-319-75380-5_3.
- [39] P. J. Larkin, "Chapter 2 - Basic Principles," in *Infrared and Raman Spectroscopy (Second Edition)*, Second Edi., P. J. Larkin, Ed. Elsevier, 2018, pp. 7–28. doi: <https://doi.org/10.1016/B978-0-12-804162-8.00002-1>.
- [40] B. C. Windom and D. W. Hahn, "Raman Spectroscopy," in *Encyclopedia of Tribology*, Q. J. Wang and Y.-W. Chung, Eds. Boston, MA: Springer US, 2013, pp. 2742–2747. doi: 10.1007/978-0-387-92897-5_1216.
- [41] N. John and S. George, "Chapter 5 - Raman Spectroscopy," in *Spectroscopic Methods for Nanomaterials Characterization*, S. Thomas, R. Thomas, A. K. Zachariah, and R. K. Mishra, Eds. Elsevier, 2017, pp. 95–127. doi: <https://doi.org/10.1016/B978-0-323-46140-5.00005-4>.
- [42] L. Lin, X. Bi, Y. Gu, F. Wang, and J. Ye, "Surface-enhanced Raman scattering nanotags for bioimaging," *J. Appl. Phys.*, vol. 129, no. 19, 2021, doi: 10.1063/5.0047578.
- [43] T. H. Kauffmann, N. Kokanyan, and M. D. Fontana, "Use of Stokes and anti-Stokes Raman scattering for new applications," *J. Raman Spectrosc.*, vol. 50, no. 3, pp. 418–424, 2019, doi: 10.1002/jrs.5523.
- [44] T. Dey, "Microplastic pollutant detection by Surface Enhanced Raman Spectroscopy (SERS): a mini-review," *Nanotechnol. Environ. Eng.*, vol. 8, no. 1, pp. 41–48, 2023, doi: 10.1007/s41204-022-00223-7.

- [45] A. Chandra *et al.*, “Unveiling the Molecular Secrets: A Comprehensive Review of Raman Spectroscopy in Biological Research,” *ACS Omega*, 2024, doi: 10.1021/acsomega.4c00591.
- [46] S. Mosca, C. Conti, N. Stone, and P. Matousek, “Spatially offset Raman spectroscopy,” *Nat. Rev. Methods Prim.*, vol. 1, no. 1, 2021, doi: 10.1038/s43586-021-00019-0.
- [47] R. R. Jones, D. C. Hooper, L. Zhang, D. Wolverson, and V. K. Valev, “Raman Techniques: Fundamentals and Frontiers,” *Nanoscale Res. Lett.*, vol. 14, no. 1, p. 231, 2019, doi: 10.1186/s11671-019-3039-2.
- [48] R. T. Vulchi, V. Morgunov, R. Junjuri, and T. Bocklitz, “Artifacts and Anomalies in Raman Spectroscopy: A Review on Origins and Correction Procedures,” *Molecules*, vol. 29, no. 19, 2024, doi: 10.3390/molecules29194748.
- [49] D. Cialla-May *et al.*, “Raman Spectroscopy and Imaging in Bioanalytics,” *Anal. Chem.*, vol. 94, no. 1, pp. 86–119, 2022, doi: 10.1021/acs.analchem.1c03235.
- [50] S. Guo, J. Popp, and T. Bocklitz, “Chemometric analysis in Raman spectroscopy from experimental design to machine learning–based modeling,” *Nat. Protoc.*, vol. 16, no. 12, pp. 5426–5459, 2021, doi: 10.1038/s41596-021-00620-3.
- [51] O. Ryabchykov, J. Popp, and T. Bocklitz, “Fusion of MALDI spectrometric imaging and Raman spectroscopic data for the analysis of biological samples,” *Front. Chem.*, vol. 6, no. JUL, pp. 1–10, 2018, doi: 10.3389/fchem.2018.00257.
- [52] N. Coca-Lopez, “An intuitive approach for spike removal in Raman spectra based on peaks’ prominence and width,” *Anal. Chim. Acta*, vol. 1295, p. 342312, 2024, doi: <https://doi.org/10.1016/j.aca.2024.342312>.
- [53] D. A. Whitaker and K. Hayes, “A simple algorithm for despiking Raman spectra,” *Chemom. Intell. Lab. Syst.*, vol. 179, no. March, pp. 82–84, 2018, doi: 10.1016/j.chemolab.2018.06.009.
- [54] R. Gautam, S. Vanga, F. Ariese, and S. Umapathy, “Review of multidimensional data processing approaches for Raman and infrared spectroscopy,” *EPJ Tech. Instrum.*, vol. 2, no. 1, 2015, doi: 10.1140/epjti/s40485-015-0018-6.
- [55] O. Ryabchykov *et al.*, “Automatization of spike correction in Raman spectra of biological samples,” *Chemom. Intell. Lab. Syst.*, vol. 155, pp. 1–6, 2016, doi: 10.1016/j.chemolab.2016.03.024.
- [56] A. Maury and R. I. Revilla, “Autocorrelation Analysis Combined with a Wavelet Transform Method to Detect and Remove Cosmic Rays in a Single Raman Spectrum,” *Appl. Spectrosc.*, vol. 69, no. 8, pp. 984–992, 2015, doi: 10.1366/14-07834.
- [57] R. L. McCreery, “Calibration and Validation,” in *Raman Spectroscopy for Chemical Analysis*, (eds J.D. Winefordner and R.L. McCreery), Ed. John Wiley & Sons, Ltd, 2000, pp. 251–291. doi: <https://doi.org/10.1002/0471721646.ch10>.
- [58] S. J. Barton, T. E. Ward, and B. M. Hennelly, “Algorithm for optimal denoising of Raman spectra,” *Anal. Methods*, vol. 10, no. 30, pp. 3759–3769, 2018, doi: 10.1039/c8ay01089g.
- [59] S. Chen, Y. H. Ong, X. Lin, and Q. Liu, “Optimization of advanced Wiener estimation methods for Raman reconstruction from narrow-band measurements in the presence of fluorescence background,” *Biomed. Opt. Express*, vol. 6, no. 7, p. 2633, 2015, doi: 10.1364/boe.6.002633.
- [60] Y. Chen and L. Dai, “Automated decomposition algorithm for Raman spectra based on a Voigt line profile model,” *Appl. Opt.*, vol. 55, no. 15, p. 4085, 2016, doi: 10.1364/ao.55.004085.
- [61] B. D. Beier and A. J. Berger, “Method for automated background subtraction from Raman spectra containing known contaminants,” *Analyst*, vol. 134, no. 6, pp. 1198–1202, 2009, doi: 10.1039/b821856k.
- [62] L. T. Kerr and B. M. Hennelly, “A multivariate statistical investigation of background subtraction algorithms for Raman spectra of cytology samples recorded on glass slides,” *Chemom. Intell. Lab. Syst.*, vol. 158, pp. 61–68, 2016, doi: 10.1016/j.chemolab.2016.08.012.
- [63] B. Barton, J. Thomson, E. Lozano Diz, and R. Portela, “Chemometrics for Raman Spectroscopy Harmonization,” *Appl. Spectrosc.*, vol. 76, no. 9, pp. 1021–1041, 2022, doi:

10.1177/00037028221094070.

- [64] S. Guo, T. Bocklitz, and J. Popp, "Optimization of Raman-spectrum baseline correction in biological application," *Analyst*, vol. 141, no. 8, pp. 2396–2404, 2016, doi: 10.1039/c6an00041j.
- [65] F. Korinth, T. A. Shaik, J. Popp, and C. Krafft, "Assessment of shifted excitation Raman difference spectroscopy in highly fluorescent biological samples," *Analyst*, vol. 146, no. 22, pp. 6760–6767, 2021, doi: 10.1039/d1an01376a.
- [66] S. He *et al.*, "Baseline correction for Raman spectra using an improved asymmetric least squares method," *Anal. Methods*, vol. 6, no. 12, pp. 4402–4407, 2014, doi: 10.1039/c4ay00068d.
- [67] S. Guo *et al.*, "Extended Multiplicative Signal Correction Based Model Transfer for Raman Spectroscopy in Biological Applications," *Anal. Chem.*, vol. 90, no. 16, pp. 9787–9795, 2018, doi: 10.1021/acs.analchem.8b01536.
- [68] J. Ma, M. Zuo, B. Wu, and W. Wang, "An Automated Baseline Correction Method for Raman Spectra Based on Piecewise Polynomial Fitting With Adaptive Window," *J. Raman Spectrosc.*, vol. 56, no. 4, pp. 337–344, 2025, doi: <https://doi.org/10.1002/jrs.6772>.
- [69] A. Fatima, G. Cyril, V. Vincent, J. Stéphane, and P. Olivier, "Towards normalization selection of Raman data in the context of protein glycation: Application of validity indices to PCA processed spectra," *Analyst*, vol. 145, no. 8, pp. 2945–2957, 2020, doi: 10.1039/c9an02155h.
- [70] S. Mostafapour, T. Dörfer, R. Heinke, P. Rösch, J. Popp, and T. Bocklitz, "Investigating the effect of different pre-treatment methods on Raman spectra recorded with different excitation wavelengths," *Spectrochim. Acta - Part A Mol. Biomol. Spectrosc.*, vol. 302, no. March, pp. 1–8, 2023, doi: 10.1016/j.saa.2023.123100.
- [71] O. Ryabchykov, S. Guo, and T. Bocklitz, "Analyzing Raman spectroscopic data," *Phys. Sci. Rev.*, vol. 4, no. 2, pp. 1–16, 2019, doi: 10.1515/psr-2017-0043.
- [72] K. H. Liland, A. Kohler, and N. K. Afseth, "Model-based pre-processing in Raman spectroscopy of biological samples," *J. Raman Spectrosc.*, vol. 47, no. 6, pp. 643–650, 2016, doi: 10.1002/jrs.4886.
- [73] R. Gautam, S. Vanga, F. Ariese, and S. Umapathy, "Review of multidimensional data processing approaches for Raman and infrared spectroscopy," *EPJ Tech. Instrum.*, vol. 2, no. 1, 2015, doi: 10.1140/epjti/s40485-015-0018-6.
- [74] Y. J. Liu, M. Kyne, C. Wang, and X. Y. Yu, "Data mining in Raman imaging in a cellular biological system," *Comput. Struct. Biotechnol. J.*, vol. 18, pp. 2920–2930, 2020, doi: 10.1016/j.csbj.2020.10.006.
- [75] N. M. Ralbovsky and I. K. Lednev, "Raman spectroscopy and chemometrics: A potential universal method for diagnosing cancer," *Spectrochim. Acta - Part A Mol. Biomol. Spectrosc.*, vol. 219, pp. 463–487, 2019, doi: 10.1016/j.saa.2019.04.067.
- [76] M. Hedegaard, C. Matthäus, S. Hassing, C. Krafft, M. Diem, and J. Popp, "Spectral unmixing and clustering algorithms for assessment of single cells by Raman microscopic imaging," *Theor. Chem. Acc.*, vol. 130, no. 4–6, pp. 1249–1260, 2011, doi: 10.1007/s00214-011-0957-1.
- [77] P. M. Conforti, G. Lazzini, P. Russo, and M. D'Acunto, "Raman Spectroscopy and AI Applications in Cancer Grading: An Overview," *IEEE Access*, vol. 12, no. April, pp. 54816–54852, 2024, doi: 10.1109/ACCESS.2024.3388841.
- [78] J. Wen *et al.*, "Detection and Classification of Multi-Type Cells by Using Confocal Raman Spectroscopy," *Front. Chem.*, vol. 9, no. April, pp. 1–8, 2021, doi: 10.3389/fchem.2021.641670.
- [79] J. Geng, W. Zhang, C. Chen, H. Zhang, A. Zhou, and Y. Huang, "Tracking the Differentiation Status of Human Neural Stem Cells through Label-Free Raman Spectroscopy and Machine Learning-Based Analysis," *Anal. Chem.*, vol. 93, no. 30, pp. 10453–10461, Aug. 2021, doi: 10.1021/acs.analchem.0c04941.
- [80] C. Ma *et al.*, "Single cell Raman spectroscopy to identify different stages of proliferating human hepatocytes for cell therapy," *Stem Cell Res. Ther.*, vol. 12, no. 1, pp. 1–12, 2021, doi: 10.1186/s13287-021-02619-9.

- [81] Y. Qi *et al.*, “Recent Progresses in Machine Learning Assisted Raman Spectroscopy,” *Adv. Opt. Mater.*, vol. 2203104, pp. 1–22, 2023, doi: 10.1002/adom.202203104.
- [82] C. Esposito *et al.*, “Assessment of Primary Human Liver Cancer Cells by Artificial Intelligence-Assisted Raman Spectroscopy,” *Cells*, vol. 12, no. 22, Nov. 2023, doi: 10.3390/cells12222645.
- [83] J. Villazon, N. Dela Cruz, and L. Shi, “Cancer Cell Line Classification Using Raman Spectroscopy of Cancer-Derived Exosomes and Machine Learning,” *Anal. Chem.*, 2025, doi: 10.1021/acs.analchem.4c06966.
- [84] M. Lasalvia, V. Capozzi, and G. Perna, “Classifying Raman Spectra of Colon Cells by Principal Component Analysis—Linear Discriminant Analysis and Partial Least Squares—Linear Discriminant Analysis Methods,” *Appl. Sci.*, vol. 15, no. 8, 2025, doi: 10.3390/app15084193.
- [85] T. W. Bocklitz, S. Guo, O. Ryabchykov, N. Vogler, and J. Popp, “Raman Based Molecular Imaging and Analytics: A Magic Bullet for Biomedical Applications!?,” *Anal. Chem.*, vol. 88, no. 1, pp. 133–151, 2016, doi: 10.1021/acs.analchem.5b04665.
- [86] X. Deng *et al.*, “Group and Basis Restricted Non-Negative Matrix Factorization and Random Forest for Molecular Histotype Classification and Raman Biomarker Monitoring in Breast Cancer,” *Appl. Spectrosc.*, vol. 76, no. 4, pp. 462–474, 2022, doi: 10.1177/00037028211035398.
- [87] X. Li, T. Yang, S. Li, D. Wang, Y. Song, and S. Zhang, “Raman spectroscopy combined with principal component analysis and k nearest neighbour analysis for non-invasive detection of colon cancer,” *Laser Phys.*, vol. 26, no. 3, p. 35702, 2016, doi: 10.1088/1054-660X/26/3/035702.
- [88] L. Zhu *et al.*, “Precise Identification of Glioblastoma Micro-Infiltration at Cellular Resolution by Raman Spectroscopy,” *Adv. Sci.*, vol. 2401014, pp. 1–11, 2024, doi: 10.1002/advs.202401014.
- [89] I. Romanishkin *et al.*, “Differentiation of glioblastoma tissues using spontaneous Raman scattering with dimensionality reduction and data classification,” *Front. Oncol.*, vol. 12, no. September, pp. 1–9, 2022, doi: 10.3389/fonc.2022.944210.
- [90] N. M. Ralbovsky and I. K. Lednev, “Towards development of a novel universal medical diagnostic method: Raman spectroscopy and machine learning,” *Chem. Soc. Rev.*, vol. 49, no. 20, pp. 7428–7453, 2020, doi: 10.1039/d0cs01019g.
- [91] J. Acquarelli, T. van Laarhoven, J. Gerretzen, T. N. Tran, L. M. C. Buydens, and E. Marchiori, “Convolutional neural networks for vibrational spectroscopic data analysis,” *Anal. Chim. Acta*, vol. 954, pp. 22–31, 2017, doi: 10.1016/j.aca.2016.12.010.
- [92] Y. Qi, Y. Liu, and J. Luo, “Recent application of Raman spectroscopy in tumor diagnosis: from conventional methods to artificial intelligence fusion,” *Photonix*, vol. 4, no. 1, p. 22, 2023, doi: 10.1186/s43074-023-00098-0.
- [93] N. Blake, R. Gaifulina, L. D. Griffin, I. M. Bell, and G. M. H. Thomas, “Machine Learning of Raman Spectroscopy Data for Classifying Cancers: A Review of the Recent Literature,” *Diagnostics (Basel, Switzerland)*, vol. 12, no. 6, Jun. 2022, doi: 10.3390/diagnostics12061491.
- [94] X. Chen *et al.*, “Applications of Data Characteristic AI-Assisted Raman Spectroscopy in Pathological Classification,” *Anal. Chem.*, vol. 96, no. 16, pp. 6158–6169, Apr. 2024, doi: 10.1021/acs.analchem.3c04930.
- [95] A. M. Fuentes *et al.*, “Raman spectroscopy and convolutional neural networks for monitoring biochemical radiation response in breast tumour xenografts,” *Sci. Rep.*, vol. 13, no. 1, pp. 1–12, 2023, doi: 10.1038/s41598-023-28479-2.
- [96] E. Pavlou and N. Kourkoulis, “Deep adversarial data augmentation for biomedical spectroscopy: Application to modelling Raman spectra of bone,” *Chemom. Intell. Lab. Syst.*, vol. 228, no. August, p. 104634, 2022, doi: 10.1016/j.chemolab.2022.104634.
- [97] J. Allgaier and R. Pryss, “Cross-Validation Visualized: A Narrative Guide to Advanced Methods,” *Mach. Learn. Knowl. Extr.*, vol. 6, no. 2, pp. 1378–1388, 2024, doi: 10.3390/make6020065.
- [98] D. Traynor *et al.*, “Raman spectral cytopathology for cancer diagnostic applications,” *Nat. Protoc.*, vol. 16, no. 7, pp. 3716–3735, 2021, doi: 10.1038/s41596-021-00559-5.

- [99] S. Wachsmann-Hogiu, T. Weeks, and T. Huser, "Chemical analysis in vivo and in vitro by Raman spectroscopy—from single cells to humans," *Curr. Opin. Biotechnol.*, vol. 20, no. 1, pp. 63–73, 2009, doi: 10.1016/j.copbio.2009.02.006.
- [100] H. Abramczyk, J. M. Surmacki, and M. Kopec, "A Novel HER2 Protein Identification Methodology in Breast Cancer Cells Using Raman Spectroscopy and Raman Imaging: An Analytical Validation Study," *J. Med. Chem.*, 2024, doi: 10.1021/acs.jmedchem.4c01591.
- [101] L. Zhang *et al.*, "Raman spectroscopy and machine learning for the classification of breast cancers.," *Spectrochim. Acta. A. Mol. Biomol. Spectrosc.*, vol. 264, p. 120300, Jan. 2022, doi: 10.1016/j.saa.2021.120300.
- [102] J. M. Surmacki, B. J. Woodhams, A. Haslehurst, B. A. J. Ponder, and S. E. Bohndiek, "Raman micro-spectroscopy for accurate identification of primary human bronchial epithelial cells," *Sci. Rep.*, vol. 8, no. 1, pp. 1–11, 2018, doi: 10.1038/s41598-018-30407-8.
- [103] A. Bresci *et al.*, "Label-free morpho-molecular phenotyping of living cancer cells by combined Raman spectroscopy and phase tomography," *Commun. Biol.*, vol. 7, no. 1, pp. 1–13, 2024, doi: 10.1038/s42003-024-06496-9.
- [104] M. Kopec, K. Beton-Mysur, and H. Abramczyk, "Raman imaging and chemometric methods in human normal bronchial and cancer lung cells: Raman biomarkers of lipid reprogramming," *Chem. Phys. Lipids*, vol. 257, no. July, p. 105339, 2023, doi: 10.1016/j.chemphyslip.2023.105339.
- [105] J. Xia, J. Li, X. Wang, Y. Li, and J. Li, "Enhanced cancer classification and critical feature visualization using Raman spectroscopy and convolutional neural networks," *Spectrochim. Acta - Part A Mol. Biomol. Spectrosc.*, vol. 326, no. August 2024, p. 125242, 2025, doi: 10.1016/j.saa.2024.125242.
- [106] X. Cheng *et al.*, "Raman spectroscopy differ leukemic cells from their healthy counterparts and screen biomarkers in acute leukemia," *Spectrochim. Acta - Part A Mol. Biomol. Spectrosc.*, vol. 281, no. November 2021, p. 121558, 2022, doi: 10.1016/j.saa.2022.121558.
- [107] L. Li *et al.*, "Classification of Soft Tissue Sarcoma Specimens with Raman Spectroscopy as Smart Sensing Technology.," *Cyborg bionic Syst. (Washington, D.C.)*, vol. 2021, p. 9816913, 2021, doi: 10.34133/2021/9816913.
- [108] L. Huang *et al.*, "Rapid, label-free histopathological diagnosis of liver cancer based on Raman spectroscopy and deep learning.," *Nat. Commun.*, vol. 14, no. 1, p. 48, Jan. 2023, doi: 10.1038/s41467-022-35696-2.
- [109] M. Karnachoriti *et al.*, "Raman spectroscopy for colorectal tumor margin assessment: A promising tool for real-time surgical delimitation," *Talanta*, vol. 290, no. February, p. 127787, 2025, doi: 10.1016/j.talanta.2025.127787.
- [110] F. Daoust *et al.*, "Preliminary study demonstrating cancer cells detection at the margins of whole glioblastoma specimens with Raman spectroscopy imaging," *Sci. Rep.*, vol. 15, no. 1, pp. 1–15, 2025, doi: 10.1038/s41598-025-87109-1.
- [111] C. Li *et al.*, "Combining Raman spectroscopy and machine learning to assist early diagnosis of gastric cancer," *Spectrochim. Acta - Part A Mol. Biomol. Spectrosc.*, vol. 287, no. P1, p. 122049, 2023, doi: 10.1016/j.saa.2022.122049.
- [112] M. Kopec, M. Błaszczyk, M. Radek, and H. Abramczyk, "Raman imaging and statistical methods for analysis various type of human brain tumors and breast cancers," *Spectrochim. Acta - Part A Mol. Biomol. Spectrosc.*, vol. 262, p. 120091, 2021, doi: 10.1016/j.saa.2021.120091.
- [113] J. P. Dulude *et al.*, "Intraoperative use of high-speed Raman spectroscopy during soft tissue sarcoma resection," *Sci. Rep.*, vol. 15, no. 1, pp. 1–13, 2025, doi: 10.1038/s41598-025-93089-z.
- [114] K. Ember *et al.*, "In situ brain tumor detection using a Raman spectroscopy system—results of a multicenter study," *Sci. Rep.*, vol. 14, no. 1, pp. 1–12, 2024, doi: 10.1038/s41598-024-62543-9.
- [115] Y. Liu, F. Ye, C. Yang, and H. Jiang, "Use of in vivo Raman spectroscopy and cryoablation for diagnosis and treatment of bladder cancer," *Spectrochim. Acta - Part A Mol. Biomol. Spectrosc.*, vol. 308, no. 12, p. 123707, 2024, doi: 10.1016/j.saa.2023.123707.

- [116] M. A. Kouri *et al.*, “Shedding Light on Colorectal Cancer: An In Vivo Raman Spectroscopy Approach Combined with Deep Learning Analysis,” *Int. J. Mol. Sci.*, vol. 24, no. 23, 2023, doi: 10.3390/ijms242316582.
- [117] V. Karunakaran *et al.*, “Investigating In Vivo Tumor Biomolecular Changes Following Radiation Therapy Using Raman Spectroscopy,” *ACS Omega*, 2024, doi: 10.1021/acsomega.4c06096.
- [118] I. Latka *et al.*, “Raman Spectroscopy for Instant Bladder Tumor Diagnosis: System Development and In Vivo Proof-Of-Principle Study in Accordance with the European Medical Device Regulation (MDR2017/745),” *Cancers (Basel)*, vol. 16, no. 18, 2024, doi: 10.3390/cancers16183238.
- [119] C. Shu, W. Zheng, K. Lin, C. M. Lim, and Z. Huang, “Real-time in vivo cancer staging of nasopharyngeal carcinoma patients with rapid fiberoptic Raman endoscopy,” *Talanta*, vol. 259, no. April, p. 124561, 2023, doi: 10.1016/j.talanta.2023.124561.
- [120] C. Krafft and V. Sergo, “Biomedical applications of Raman and infrared spectroscopy to diagnose tissues,” *J. Spectrosc.*, vol. 20, no. 5–6, p. 738186, 2006, doi: <https://doi.org/10.1155/2006/738186>.
- [121] D. Cialla *et al.*, “Surface-enhanced Raman spectroscopy (SERS): progress and trends,” *Anal. Bioanal. Chem.*, vol. 403, no. 1, pp. 27–54, Apr. 2012, doi: 10.1007/s00216-011-5631-x.
- [122] J. D. Rodriguez, B. J. Westenberger, L. F. Buhse, and J. F. Kauffman, “Standardization of Raman spectra for transfer of spectral libraries across different instruments,” *Analyst*, vol. 136, no. 20, pp. 4232–4240, 2011, doi: 10.1039/c1an15636e.
- [123] D. Boateng, “Advances in deep learning-based applications for Raman spectroscopy analysis: A mini-review of the progress and challenges,” *Microchem. J.*, vol. 209, no. January, p. 112692, 2025, doi: 10.1016/j.microc.2025.112692.
- [124] R. Luo, J. Popp, and T. Bocklitz, “Deep Learning for Raman Spectroscopy: A Review,” *Analytica*, vol. 3, no. 3, pp. 287–301, 2022, doi: 10.3390/analytica3030020.

Chapter 2: Assessment of Primary Human Liver Cancer Cells by Artificial Intelligence Assisted Raman Spectroscopy

Abstract

We investigated the possibility of using Raman spectroscopy assisted by artificial intelligence methods to identify liver cancer cells and distinguish them from their Non-Tumor counterpart. To this aim, primary liver cells (40 Tumor and 40 Non-Tumor cells), obtained from resected hepatocellular carcinoma (HCC) tumor tissue and the adjacent non-tumor area (negative control) were analyzed by Raman micro-spectroscopy. Judicious post-processing data analysis was performed to develop a rapid method for discrimination and classification. Three machine learning approaches were simultaneously investigated and successfully used to overcome the overfitting issues associated with large training datasets. The results clearly demonstrate the effectiveness of artificial intelligence (AI) assisted Raman spectroscopy for Tumor cell classification and prediction with an accuracy of nearly 90% of correct predictions at the single-spectrum. (Esposito C, Janneh M, Spaziani S, Calcagno V, Bernardi ML, Iammarino M, Verdone C, Tagliamonte M, Buonaguro L, Pisco M, Aversano L, Cusano A. Assessment of Primary Human Liver Cancer Cells by Artificial Intelligence-Assisted Raman Spectroscopy. *Cells*. 2023 Nov 17;12(22):2645. doi: 10.3390/cells12222645. PMID: 37998378; PMCID: PMC10670489.)

1. Introduction

Liver cancer remains a global health challenge and its incidence is increasing worldwide [1]. It is estimated that liver cancer will affect more than 1 million people annually by 2025 [2]. Epidemiological data also indicate an increase in liver cancer-related deaths, ranking this malignant disease as the third leading cause of death worldwide [3]. Moreover, liver tumors have been predicted to be among the leading causes of death by 2030 [4].

Liver cancer is defined as a heterogeneous group of tumors characterized by different histological, molecular, and clinical features [5]. The most common form of primary liver cancer in adults is hepatocellular carcinoma (HCC), with an incidence rate of 90% [2].

In clinical practice, the traditional diagnosis of HCC relies on various imaging methods such as ultrasonography (US), computed tomography (CT) or magnetic resonance imaging (MRI). However, the unsatisfactory sensitivity, need for standardization, operator-dependent results of US, high cost and time-consuming procedures for CT and MRI make the diagnosis of HCC difficult [6]. In cases where there are no clear CT/MRI findings, liver biopsy is performed.

Tissue samples taken during a liver biopsy must be sent to an analytical laboratory for histopathologic diagnosis, which is performed at the tissue and cellular levels. The gold standard protocol requires that the collected specimens undergo various preparation steps (dehydration, fixation, waxing, staining, slicing, mounting) which require labor and time that typically ranges from 10 hours to 3 days [7]. In addition, the accuracy of pathological diagnosis is influenced by the subjective factors of pathologists [8].

Therefore, a technological approach that allows rapid and “intelligent” analysis and classification of cells and tissues from tissue biopsies would be of utmost importance for improving the diagnosis and characterization of malignant cells.

In this regard, Raman spectroscopy (RS) has emerged as a powerful spectroscopic technique for clinical oncology applications because of its label-free, noninvasive, nondestructive and rapid nature. The Raman spectrum, indeed, represents the “molecular fingerprint” of a biological sample and the Raman bands (i.e. the peaks in the spectrum) are associated with specific chemical information about the cellular components (i.e. lipids, proteins, nucleic acids, carbohydrates). RS has been shown to have high sensitivity and molecular specificity in studying the altered biochemical composition in tumor samples compared to its non-tumor counterpart [9].

Actually, the use of RS in cancer diagnostics is limited by the complexity in the elaboration of the Raman data for identifying vibrational fingerprints, which are useful for distinguishing between oncological and non-pathological samples [10]. Nonetheless, the use of artificially intelligent systems could significantly contribute to appreciate statistically relevant tiny differences in the spectra and support prognostic and diagnostic stages through accurate and in-depth cellular analysis.

Indeed, great efforts have been made using machine learning methods to identify and analyze a characteristic spectral pattern for cell classification. In particular, there is scientific evidence for the application of RS in combination with machine learning approaches to classify different tumor cell types as well as to discriminate between cancer and non-cancer cells [11], [12], [13]. Notably, most machine-learning and Raman spectroscopy-based methods for distinguishing and classifying different human liver cancer cells and non-cancer cells used immortalized cell lines [14], [15]. Although human liver cell lines are commonly used as a model for liver cancer, they represent a homogeneous cell population that lacks some liver-specific functions, has an altered genome, and exhibits atypical cell behavior compared to primary cells [16]. This leads to modeling machine learning methods that suddenly fail when applied to real classifications. On the other hand, the cultivation of primary cells is also not optimal, as only a limited number of cell cycles are observed, leading to a progressive loss of morphological, phenotypic and functional cellular features [17]. Therefore, we focused our studies on uncultured primary human liver cancer and non-cancer cells, which represent the cell systems that better correspond to the real characteristics of the liver under pathological and physiological conditions, respectively.

In this study, we report the first Raman spectral analysis based on characterization and differentiation between uncultured primary human liver cancer and non-cancer cells belonging to an HCC patient. Raman analysis of the two primary cell samples was performed to identify the molecular spectral fingerprint reflecting a different biochemical composition of the two samples.

Two basic machine learning models based on Linear Discriminant Analysis (LDA) were explored to further automate and accelerate data analysis to reveal hidden patterns correlated with pathology.

LDA is a statistical method used to find linear combinations of features that best separate classes of data. PCA is a dimensionality reduction technique used to decrease the dimensionality of the data by projecting it onto a subspace with lower dimensionality

subspace while preserving as much variance as possible. Therefore, we used an LDA model applied directly to the Raman spectra, namely the “Hyper-parameter tuned LDA”, and an LDA model applied to the PCA components, namely the “PCA-LDA” model, to extract the most informative features and feed them into a classifier.

Alternatively, we developed a neural network-based ensemble based on the combination of convolutional and recurrent neural networks. Convolutional Neural Networks (CNNs) are particularly effective at extracting relevant features for recognizing complex patterns [18], such as Raman spectra, while recurrent networks based on Long-Short Term Memory cells (LSTMs) are effective at learning and detecting relationships in sequences [19]. The neural network ensemble, namely the “CNN-LSTM” model, is proposed as an additional model for recognizing and rapidly classifying specific Raman fingerprints of Non-Tumor and Tumor cells, to assist clinicians in diagnosing HCC liver cancer.

The ability to discriminate between cancer and non-cancer cells was assessed by using the three artificial intelligence models to blindly classify unknown cells. The results obtained outline the importance of combining Raman spectroscopy and machine learning to develop a diagnostic tool to improve HCC diagnosis. (Esposito C, Janneh M, Spaziani S, Calcagno V, Bernardi ML, Iammarino M, Verdone C, Tagliamonte M, Buonaguro L, Pisco M, Aversano L, Cusano A. Assessment of Primary Human Liver Cancer Cells by Artificial Intelligence-Assisted Raman Spectroscopy. *Cells*. 2023 Nov 17;12(22):2645. doi: 10.3390/cells12222645. PMID: 37998378; PMCID: PMC10670489.)

2. Materials and Methods

2.1. Sample collection and preparation

Primary human liver cells were provided by the National Cancer Institute “Fondazione Pascale” from tumoral and adjacent non-tumoral tissues from the resected liver of an HCC patient. The protocol was approved by Ethics Committee (approval no. 421/13) After resection, both tissues were dissociated into single-cell suspensions using the GentleMACS dissociator (GentleMACS, Miltenyi Biotec, Bergisch Gladbach, Germany). Afterwards, the fresh uncultured Non-Tumor and Tumor cell populations were seeded onto CaF₂ slides. Subsequently, the samples were fixed with 2% paraformaldehyde (Sigma-Aldrich). Moreover, two additional samples were prepared using the same procedure, mixing different proportions of cancer and non-cancer cells. Specifically, the first additional sample contained 20/80% Non-Tumor/Tumor cells, and the second consisted of a 40/60% of Non-Tumor/Tumor cell fraction (namely MIX1 and MIX2, respectively). Before Raman measurements, samples were washed three times with a solution of H₂O MilliQ and 0.02% sodium azide (Sigma-Aldrich) and air dried. After completion of Raman analysis, samples were stored at +4°C in buffer medium.

2.2. Raman Spectroscopy

Raman measurements were performed using a LabRAM HR Nano (HORIBA Scientific, France). The system was equipped with a frequency doubled Nd:YAG laser operating at an excitation wavelength of 532 nm and providing a maximum output power of about 30 mW.

All measurements were performed with an output power of about 3 mW. A 100× air (Olympus, Japan) objective with a numerical aperture (NA) of 0.90 was used. In the HORIBA system, the collected light is focused with an adjustable pinhole, which was set at 50 μm for all experiments. The backscattered light was spectrally separated using a 300 lines per mm grating and detected with a CCD Camera (Syncerity CCD Deep Cooled Camera, cooled to $-60\text{ }^{\circ}\text{C}$). Raman spectra were recorded in a range of $600 - 1800\text{ cm}^{-1}$ (fingerprint region, FP). Raman spectra were acquired by focusing on the cell nucleus after visual inspection of the bright-field image; at least 5 spectra were acquired for each cell. The integration time and number of accumulations per Raman spectrum were 60.0 s and 2, respectively.

2.3. Raman data pre-processing

Prior to the application of statistical/chemometric methods, a pre-processing procedure is required to correct for any contributions that corrupt the Raman spectra [20].

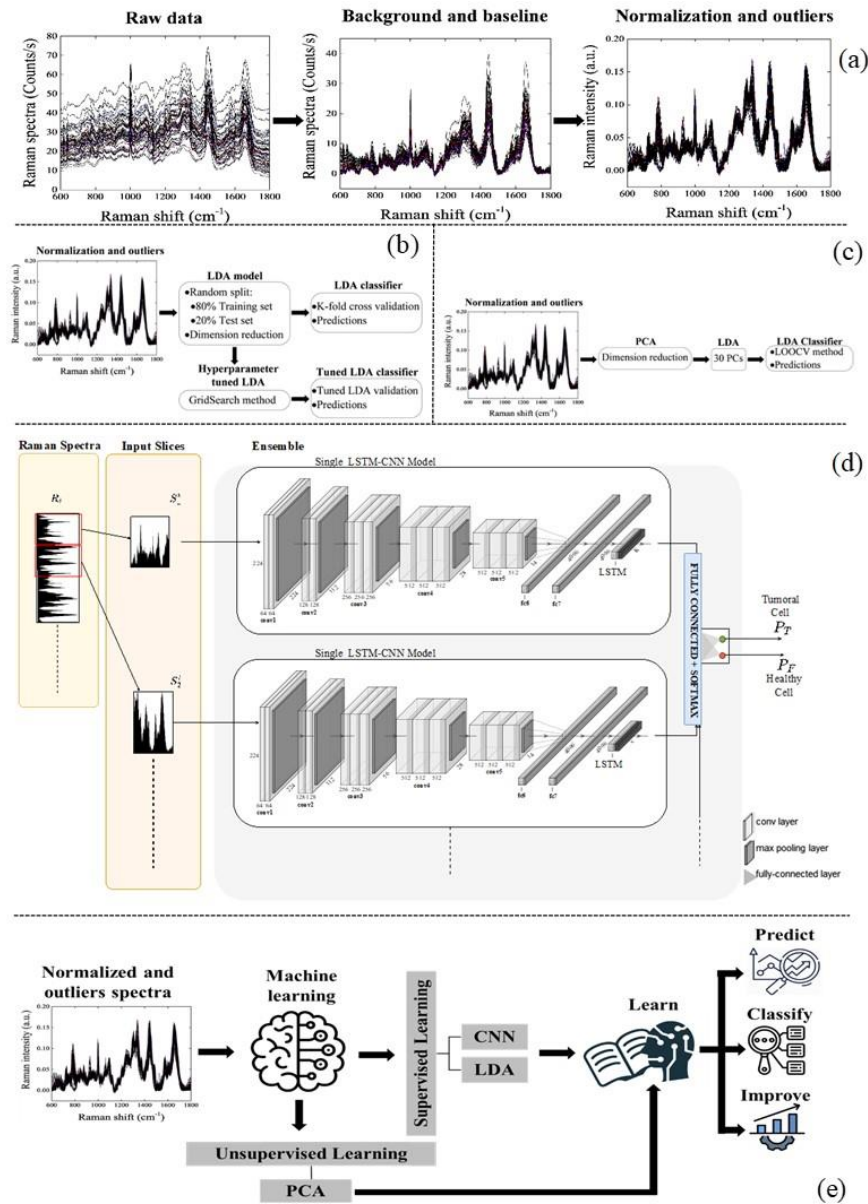


Figure 1. a) Raman spectra preprocessing steps. Architecture of the classifier based on b) LDA and Tuned-LDA model, c) PCA-LDA model, and d) Convolutional and Recurrent Neural Networks trained using sliding window on Raman Spectra. e) Schematic of the machine learning models used for the blind predictions.

To align the spectra, the system was calibrated daily to the spectral line of crystalline silicon at 520.7 cm^{-1} . For background subtraction, a spectrum was recorded on the surface of the CaF_2 substrate and subtracted from all acquired spectra. To remove the fluorescence contribution, the baseline was corrected applying a polynomial fitting (order 3) [21]. Finally, to compare the intensities, all spectra were scaled by applying a vector normalization [22]. Partial least square (PLS) regression in combination with Q-residual and Hotelling's T-squared and a confidence level of 95% were used as metrics to identify outliers in the data set. In **Figure 1a** the preprocessing steps and their effect on the spectra is briefly highlighted.

2.4. Machine learning

Preliminarily Unsupervised Multivariate Analysis (MVA) was used to classify the cells spectra. Specifically, PCA was applied as an unsupervised method to identify relevant differences between Non-Tumor and Tumor cells and reduce data dimensionality.

In addition, LDA was applied as a supervised method to discriminate the data and maximize the variance between the two groups. On the other hand, the predictive performance of the model was evaluated with a learning curve [23], [24] based on k-fold cross-validation. By setting $k=5$, k-fold cross-validation uses 4 folds ($k - 1$) as training datasets and 1 fold for the test dataset, and the test accuracy is calculated after each k-iteration. The average of all results is then used to estimate the performance of the model built on the entire data. To overcome the overfitting, the LDA model was optimised using two different approaches [24].

Hyper-parameter tuned LDA. First, LDA was optimized with a Hyper-parameter tuning based on a grid search method that performs a full search over a given subset of the Hyper-parameter space and selects the useful features for training the algorithm [25], [26]. To test the predictive ability of the LDA model optimized for the Hyper-parameters, k-fold cross-validation ($k = 5$) was again used. **Figure 1b** resumes the main steps involved in the development of the Hyper-parameter tuned LDA model. The confusion matrix was used to evaluate the accuracy, specificity, and sensitivity of the LDA and optimized LDA methods. In particular, the prediction accuracy (Acc) is defined as the proportion of correct predictions to the total number of predictions, can be calculated as follows:

$$Acc = \frac{(TP+TN)}{(TP+TN+FP+FN)}, \quad (1)$$

where true positives (TP) are correctly labelled as positive, false positives (FP) are incorrectly labelled as positive, true negatives (TN) are correctly labelled as negative, and false negatives (FN) are incorrectly labelled as negative. The other metrics used are described as follows:

Sensitivity (SNC): Refers to the actual true positive the model was able to predict and can be explained as

$$SNC = \frac{TP}{TP+FN}, \quad (2)$$

Specificity (SPC): refers to the actual true negative the model was able to predict and can be written as:

$$SPC = \frac{TN}{TN+FP}, \quad (3)$$

Precision (PRC): out of all predicted Positive cases, how many were actually Positive”, or

$$PRC = \frac{TP}{TP+FP}, \quad (4)$$

Finally, F1-Score is a combination of the precision and recall of the model by harmonic mean:

$$F1 = \frac{2*TP}{2*TP+FN+FP}, \quad (5)$$

PCA-LDA. The LDA was alternatively optimized by feeding it with the components obtained from PCA, as schematically shown in **Figure 1c**. In this case, Leave-One-Out-Cross-Validation (LOOCV) was performed to avoid under- or over-fitting due to inappropriate selection of components and to determine the prediction of the error rate associated with the model; accordingly, the first 30 PCs were selected as input to the LDA. Data manipulation and machine learning were performed using Python and Origin 2018 (OriginLab, Northampton, MA, USA).

LSTM/CNN-based classifiers. The CNN-LSTM model combines CNN and LSTM layers to extract features from input data and provide sequence prediction. This model is widely used for activity recognition, image labeling, and video labeling that require visual time series prediction and text annotation generation. The basic architecture of the CNN-LSTM network, including the input layer, visual feature extraction, sequence learning, and output layer, is shown in **Figure 1d**.

The proposed architecture consists of 25 or 56 layers, including an input layer, 22 or 50 CNN layers wrapped with time-distributed layers, an LSTM layer, two or five dense layers, a dropout layer, and an output layer. We ensured that each Raman spectrum was transformed into the appropriate shape before constructing the model by using a sliding window process to extract windows of equal length for each Raman sequence.

Specifically, to feed an ensemble of CNN-based pre-trained networks, each input Raman spectrum is segmented using windows of a fixed size so as to generate k_s slices from each spectrum (referred to as S_j^i where i identifies the spectrum and $j \in \{1, \dots, k_s\}$ is the number of the input channel in the ensemble). Each segment is feed as input to one of the channel of the ensemble, which are jointly trained.

Once all data were properly vectorized, feature extraction was performed using three convolutional layers that automatically extracted features from the input sequences using the ReLu activation function, using 64, 128, and 256 filters in these convolutional layers. The height of kernel was set to 6 and the width of the kernel was set to 4 for the convolution

operation. We wrapped the convolutional layers in a time-distributed wrapper to transform the input data by adding an extra dimension at the end. We used a flatten layer to concatenate all the extracted features and pass them to the LSTM layer. Then, a 100-unit LSTM layer was designed following a dropout layer (0.5) on the fully connected layer. Finally, for binary classification, we used the softmax activation function to specify the outputs. The architecture can be used in combination with augmentation techniques to artificially expand a dataset by creating new samples that are variations of the original samples. In the context of Raman spectra processing, augmentation can be used to generate new samples that have different frequencies and different Raman shift intensities, to increase the size of the dataset and improve the robustness of a classifier. We apply augmentation at the window level.

The first kind of augmentation is to increase or reduce the Raman shift intensity. This involves randomly modifying the Raman shift intensities of a window by a small amount. The shift amount is usually determined by a uniform or normal distribution with a small mean and standard deviation. In this way new spectra can be generated that are similar to the original but have slight differences in the intensity of the peaks (10% in this study). This can help the classifier learn to be more robust to variations in the Raman shift axis that may occur due to instrument calibration, sample preparation, or other factors.

The other augmentation applied in this study is frequency variation. In this case, the frequency axis of a window is randomly shifted by a small amount. Again, the shift factor is typically determined by a uniform or normal distribution with a small mean and standard deviation. In this way, new spectra can be generated that are similar to the original but with slight variations (10% in this study) in the position of the peaks within the window. This can help the classifier learn to be more robust to variations in peaks spectral range, that may occur, for example, due to recalibration of the system.

Both augmentation techniques can be applied independently or together to create a larger and more diverse dataset for training a classifier. In this study, we performed an ablation study in which we applied both techniques independently and together to investigate the effects on the resulting performance of the classifier.

However, it is important to note that the amount and type of augmentation used should be carefully chosen to avoid overfitting the model to the augmented data, which can lead to poor performance on real-world data. In addition, it is important to validate the performance of the classifier on a separate test set containing real-world spectra that were not used for training or augmentation. For this reason, both the validation and test data are never augmented in our study.

When misclassifications can have serious consequences, such as in the medical diagnosis, it is essential to ensure that the classifier's predictions are accurate and reliable. One approach to improving the reliability of a CNN-based classifier is to provide a reject option. With a reject option, the classifier can indicate when it is uncertain about its predictions and when it cannot confidently classify a cell. Instead of forcing a decision, the classifier can reject the cell and request further samples. Introducing a reject option in the proposed LSTM-CNN-based classifier can have a significant impact on its performance, as rejecting cells can reduce the number of false positive classifications, resulting in higher precision. On the other hand, rejecting too many cells can result in even lower recall thus reducing the overall accuracy. By adjusting the decision threshold, it is possible to balance the trade-off between

precision and recall and maximize overall accuracy. In this study we therefore introduced two levels of reject options to test whether they have a positive impact on the actual classifier performance resulting from Hyper-parameter optimization.

To implement a reject option an architecture based on SelectiveNet [27] is adopted. SelectiveNet adds a branch to the network alongside the main prediction branch. This additional branch is responsible for estimating the confidence or uncertainty of the main branch's predictions. If the uncertainty surpasses a certain threshold, the network rejects the prediction. The selective branch can be implemented in various ways, such as a separate neural network, an auxiliary CNN, or a few additional layers in parallel with the main branch. We opted, as shown in [27], to add a few additional layers in parallel with the main branch. This also required a loss function for the SelectiveNet branch that measures the uncertainty or confidence of the predictions. A thresholding mechanism to determine when to reject a prediction has been implemented. If the confidence score is below this threshold, the network rejects the prediction. Adding the reject option can improve the reliability of predictions in situations where the network is uncertain, improving precision. In this work, two thresholds have been compared during assessment (refer to section 3.6 for adopted hyper-parameters ranges).

Blind prediction. To evaluate the predictive ability of the developed artificial intelligence models, we performed a blind prediction on unknown spectra, as schematically shown in **Figure 1e**, not used in model definition. In particular, we selected three sets of cells composed of different percentages of Tumor cells (namely, MIX sets).

3. Results and Discussion

3.1. Morphological analysis

Primary liver cells were obtained from resected HCC tumor tissue and the adjacent non-tumor counterpart. A preliminary morphological analysis was performed by bright-field microscopy. Representative images of uncultured Non-Tumor and Tumor cells are shown in **Figure 2**.

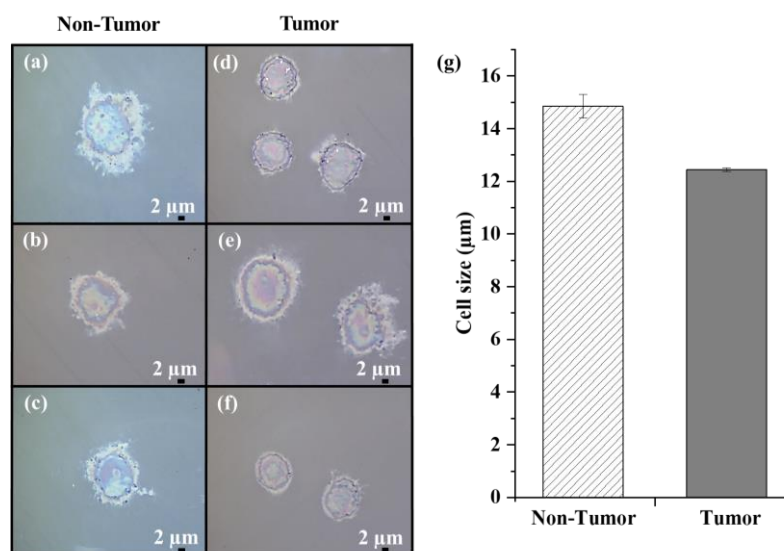


Figure 2. Bright-field microscope images of uncultured human Non-Tumor a-c) and Tumor cells d-f) fixed on CaF₂ slides (100× magnification; scale bar 2 μm). g) Dimensional analysis of all Non-Tumor and Tumor cells.

Differences in the morphological characteristics of the two samples can be appreciated. Hepatocytes exhibited polyhedral or round shapes and a central round nucleus (**Figure 2a-c**). Morphologically, HCC cells showed variable shapes (mostly irregular or round), large nuclei, and increased nuclear/cytoplasmic ratio (**Figure 2d-f**), consistent with their altered and heterogeneous phenotypic characteristics.

We estimated the average cell size of the Non-Tumor and Tumor cell samples and found a value of approximately 15 μm and 12.4 μm , respectively. To note, the primary HCC cells were smaller than their non-tumor counterpart (**Figure 2g**), in accordance with the different size exhibited by cancer cells compared to non-tumor cells [28].

3.2. Raman spectroscopy biochemical overview

The study of the complexity of biomolecular changes occurring between primary Non-Tumor and Tumor cells was performed using RS. For Raman analysis, we analyzed 40 primary liver cells derived from resected HCC tumor tissue and 40 coming from the adjacent non-tumor area (namely, “Non-Tumor” cells). Specifically, Raman spectra were acquired at five different locations of the nucleus of each cell to account for the heterogeneous distribution of cellular components. A total of 200 spectra were acquired for both samples in the 600 - 1800 cm^{-1} spectral fingerprint region, which contains the most informative cellular peaks.

The characteristic average Raman spectra of Non-Tumor and HCC cells are shown in **Figure 3** (top).

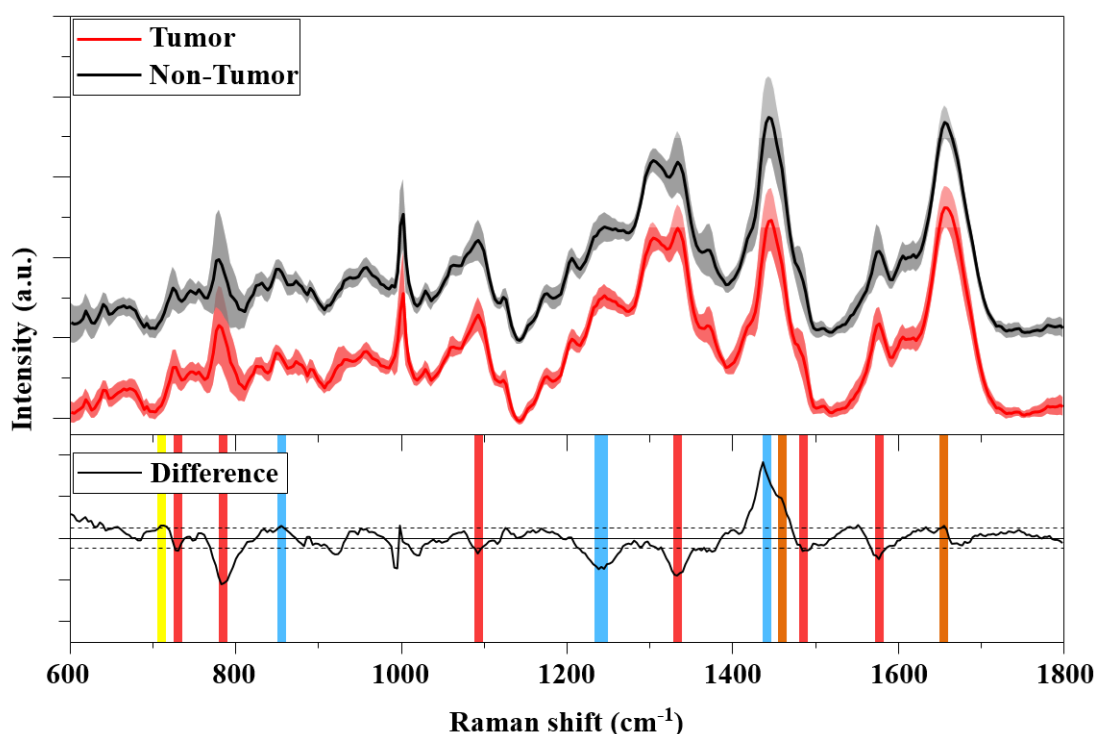


Figure 3. Top: Averaged Raman spectra of Tumor (red line) and Non-Tumor (black line, offset 10%) cells in the FP region. Solid lines show the average over 180 spectra $\pm$ standard deviation (shaded areas). Bottom: Difference between averaged Raman spectra (black line). The horizontal solid line corresponds to 0 intensity, the two horizontal dashed lines correspond to a threshold of ± 0.025 intensity. Highlighted in yellow, red, and blue are the Raman bands associated with lipids, nucleic acids and proteins, respectively. In orange, the Raman peaks associated with lipids/proteins are shown.

Both average Raman spectra show a similar spectral pattern. A detailed band assignment of the vibrational modes of the major biomolecules (nucleic acids, proteins, lipids and carbohydrates) characterizing the cells was summarized in **Table 1**. In particular, both mean spectra had distinct Raman bands at 782, 1094, 1335, 1370 and 1578 cm^{-1} , corresponding to the vibrational modes of the nucleic acid. Other intense peaks were identified at 1004 and 1246 cm^{-1} , corresponding to the phenylalanine and Amide III bands, respectively. A prominent peak at 1444 cm^{-1} , related to proteins/lipids vibrational bonds, and an intense Amide I band at 1656 cm^{-1} were also evident.

To better illustrate the different biochemical composition between the cell classes considered, we show in **Figure 3** (bottom), the difference between primary Tumor and Non-Tumor cells. The analysis revealed an important difference between the cell samples.

In particular, an interesting difference resulting from the nucleic acid at 785 cm^{-1} , the negative peak belonged to HCC cells. In addition, some other bands at 1094, 1335 and 1578 cm^{-1} indicate a greater proportion of DNA in the nuclei of cancer cells compared to non-cancer cells. Difference analysis also showed a positive peak at 1438 cm^{-1} belonging to non-cancer cells. In detail, this Raman band was assigned to CH_2 and CH_3 deformations in normal tissue [29]. In addition, we also found a negative Raman peak at 1240 cm^{-1} in the spectral region 1230-1280 cm^{-1} associated with Amide III.

Table 1. Assignment of characteristic Raman bands of Tumor and Non-Tumor cells.

Raman shift (cm^{-1})	Assignment	Biomolecule
729	A ring br.	Nucleic acid [30]
757	Trp ring br.	Protein [31]
782 - 785	DNA backbone, U, C, T ring br.	Nucleic acid [14]
826	O–P–O str.	Nucleic acid [32]
854	Tyr ring br.	Protein [30]
840 - 860	Polysaccharide structure	Carbohydrates [33]
1004	Phe ring br.	Protein [34]
1031	Phe C–H in-plane bend.	Protein [32]
1064	Skeletal C–C str.	Lipids [32]
1094	Sym. PO_2^- str.	Nucleic acid [34]
1177	Tyr C–H bend.	Protein [30]
1207	Phe, Trp C– C_6H_5 str.	Protein [35]
1240 - 1246	Amide III	Protein [32]
1305	(CH_2) twist.	Lipids/ Protein [36]
1335	A, G ring br., C–H def.	Nucleic acid/ Protein [37]
1370	DNA bases ring br.	Nucleic acid [38]
1438	CH_2 , CH_3 def.	Protein [29]

1444	CH ₂ , CH ₃ def.	Lipids/Protein [39]
1578	A, G ring br.	Nucleic acid [40]
1606	Tyr, Phe C=C bend., C NH ₂	Protein/Nucleic acid [41]
1618	Phe, Tyr and Trp C=C	Protein [42]
1656	C=C str., Amide I	Lipids/Protein [30]

br. = breathing mode, str. = stretching mode, bend. = bending mode, def. = deformation mode, twist. = twisting mode, sym. = symmetric, Phe = phenylalanine, Trp = tryptofan, Tyr = tyrosine, A =adenine, U = uracil, C = cytosine, T = thymine.

Overall, the results showed important spectral differences indicating the higher presence of nucleic acid in the nuclei of cancer cells than in non-cancer cells. This is broadly consistent with other studies in literature. Notably, some authors reported that the peaks assigned to nucleic acids around 782, 1094, 1335 and 1578 cm⁻¹ were mostly assigned to DNA [14], [34], [37], [40]. Many studies have confirmed the association between cancer and an aberrant DNA amount [43], [44]. In particular, abnormalities in the nuclear DNA content of hepatocyte have been described as an important risk factor for their transformation into cancer cells [45]. Moreover, a recent work found a significant alteration in the nuclear ploidy of HCC tumor [46]. In particular, a correlation between nuclear ploidy amplification and HCC aggressiveness and poor prognosis was demonstrated, thus discovering a new marker for HCC classification.

3.3. Dataset

Raman analysis was performed on 40 Non-Tumor and 40 Tumor cells, and 5 Raman spectra were obtained for each cell. After removal of outliers, the resulting dataset consists of 358 Raman spectra, 181 spectra for Tumor cells and 177 spectra for Non-Tumor cells, with 19 and 23 spectra, respectively, considered as outliers.

3.4. Unsupervised Multivariate Analysis

Principal Component Analysis (PCA) was performed on the entire dataset using the Covariance Matrix. The results are shown in **Figure 4a-b**. Starting with PC1 (51.2% of the total variance), the following PCs describe differences in the FP region, which account for progressively smaller proportions of the total variance.

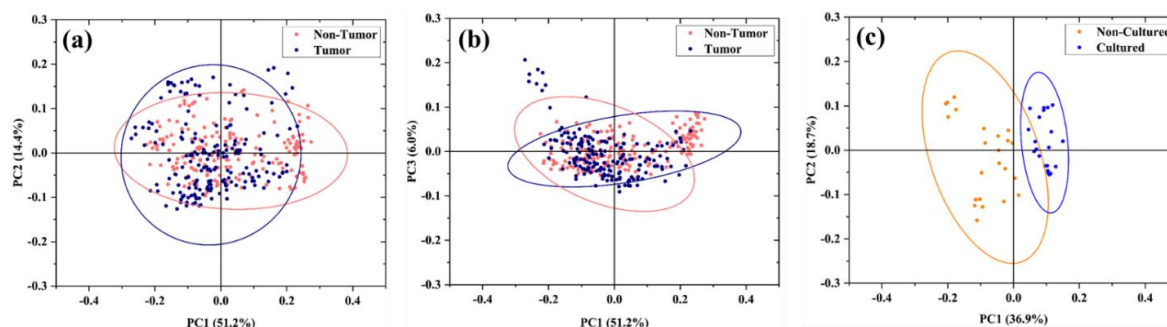


Figure 4. PCA of the Raman spectra. PCA 2D data plot distribution of spectra of uncultured Tumor and Non-Tumor cells based on the first 3 PCs components: a) PC1 vs PC2 and b) PC1 vs PC3. The ellipses account for a confidential level of

95% of the data. c) PCA plot associated to the spectra of five uncultured and cultured Tumor cells. The 2D plot is based on the first 2 PCs components: PC1 vs PC2.

Interestingly, the loading plot of PC1 (see **Figure 5**) shows a high similarity with the spectrum obtained from subtracting the mean spectra of Tumor and Non-Tumor, suggesting that PC1 explains this main difference. The first 3 PCs, accounting for 71.6% of the cumulative variance, are used to build the related 2D scatter plots. As can be seen, the data distribution shows a considerable overlap of the data.

The difficulties in discrimination by PCA analysis can be explained by the intrinsic intra- and intercellular heterogeneity characteristic of primary cells. Indeed, tumors are complex tissues consisting of a heterogeneous cell population composed of the coexistence of Tumor and Non-Tumor cells as well as other cell types (stromal, endothelial, and immune cells) [47].

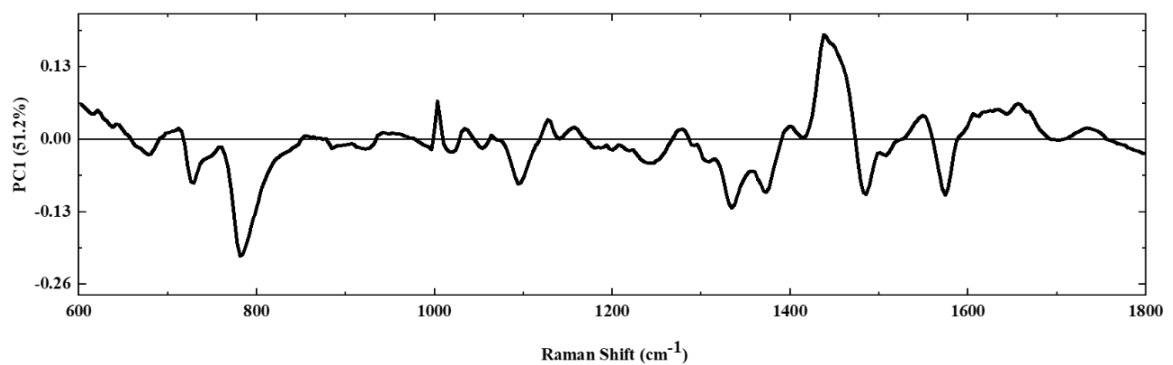


Figure 5. Loading Plot of principal component 1 (PC1).

This characteristic cellular heterogeneity is also manifested at the level of single-cell and cell populations [48].

It is worth mentioning that many scientific studies frequently use cultured primary hepatocytes as in vitro models for HCC [49]. However, cultured cell lines may differ from in vivo Tumor cells just by their homogeneity and could lead to misleading results in relation to the mentioned heterogeneity of primary cells [47]. As an example, we show PCA analysis of a few (five) primary cells compared with the corresponding cultured tumor cells in **Figure 4c**. The 2D scattering plot clearly shows that the cultured cells can have a high degree of similarity, which is due to the artificial growth conditions of the cultured cells, while the corresponding primary Tumor cells can be clustered separately. Therefore, to assess the extent to which Raman spectroscopy can identify Non-Tumor cells and Tumor cells in a realistic scenario, we worked exclusively with primary cells. Considering that PCA alone is not able to fully highlight such subtle differences, in the following we explore different approaches of Supervised Multivariate Analysis to extract the most useful features for distinguishing experimental groups and classifying spectra.

3.5. Supervised Multivariate Analysis

To discriminate the data and maximize the variance between the two groups, LDA was applied as a supervised multivariate analysis directly on Raman spectra. The model was built using the pre-processed data after outliers were removed. To evaluate the accuracy of the LDA model, we used a training test split procedure to generate a random split of 80%

training set and 20% test set. The prediction accuracy of the LDA model was estimated to be 77% (see confusion matrix in **Table 2**).

To understand the reason for the low prediction accuracy, we examine the bias-variance trade-off of the model by plotting the learning model in **Figure 6a**. Ideally, the higher the number of training samples, the better the prediction performance of the model would be. However, as can be seen in **Figure 6a**, the trend of the learning model shows an excessive maximization of the training accuracy, while it is not able to generalize well from the training data to the test data. This effect is known as overfitting and may be related to the quality of the training dataset and, in particular, to the presence

Table 2. Confusion matrix of the LDA model.

Confusion matrix (Hyper-LDA)			
Predicted	Non-Tumor	24 (85.71%)	13 (28.89%)
		4 (14.29%)	32 (71.11%)
Tumor	Non-Tumor	Tumor	
	True		

of noise in the training data. Indeed, an over-fitted model tends to include all features, even those that have a very limited effect or may worsen the final classification. To reduce the impact of overfitting and improve the model accuracy, we optimized the machine learning method using a “Hyper-parameter” optimization. We integrated, as optimization algorithm, the contracting grid search method, to automate the tuning of the Hyper-parameter. In the grid search, starting from a finite set of possible values for each Hyper-parameter, an exhaustive search for the optimal values is performed using k-fold cross-validation [26]. The result is an optimized LDA model that maximizes the weight of the useful features and minimize the weight of the useless ones.

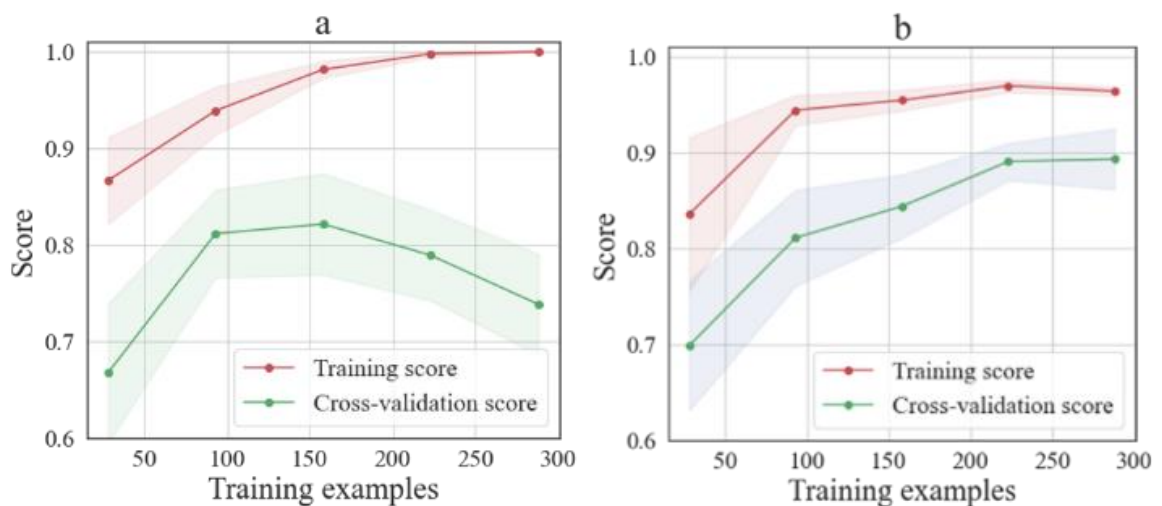


Figure 6. Learning curve as function of the number of training examples for: a) LDA model and b) LDA model after Hyper-parameters tune optimization methods.

Figure 6b shows the learning curve after applying the grid search optimization. The result shows that tuning the Hyper-parameters using grid search improves the LDA model. Indeed,

it can be seen that, after optimization, the error rate decreases in both training and the cross-validation until it reaches a stationary value. The optimized LDA model was validated on the 20% test set.

To determine the number of correct classifications and misclassifications when the optimized LDA model makes predictions for each primary liver cell class, we report the confusion matrix in **Table 3** for the known test set (Non-Tumor = 37 and Tumor = 36). The accuracy was calculated to be 89%. In addition, the classification performance of the model was calculated in terms of sensitivity, specificity, and precision (see **Table 4**). Despite the high similarity between the Raman spectra of Tumor and Non-Tumor cells, the optimized LDA model provides very high accuracy (about 89%).

As further confirmation, we developed another machine learning model exploiting the PC components. We used PCA to reduce the dimensionality of the starting dataset and select only the features that are useful for the model, namely only the PC components that account for a higher percentage of the total variance. The selected PC components were then used to feed the LDA model.

Table 3. Confusion matrix of the Hyper- parameter tuned LDA models.

Confusion matrix (Hyper-parameter LDA)		
Predicted	Non-Tumor	Tumor
	31 (93.94%)	6 (15.00%)
	2 (6.06%)	34 (85.00%)
		True
		Non-Tumor
		Tumor

Table 4. Optimized LDA classification metrics.

Merit	Scores (%)
Sensitivity	94.4
specificity	84
Precision	85
F1-score	89

The selection of the optimal number of PC components is not trivial, since a smaller number of components can cause the loss of important features, while a larger number of PC components would not solve the overfitting issue. Accordingly, a LOOCV for the classification model was performed using different numbers of PCs from 5 to 60, and an optimum value of 30 PC components was selected, accounting for 94.3% of cumulative variance and discharging the PC components accounting for less than 0.085% of total variance (see **Figure 7** and **Table 5**, respectively).

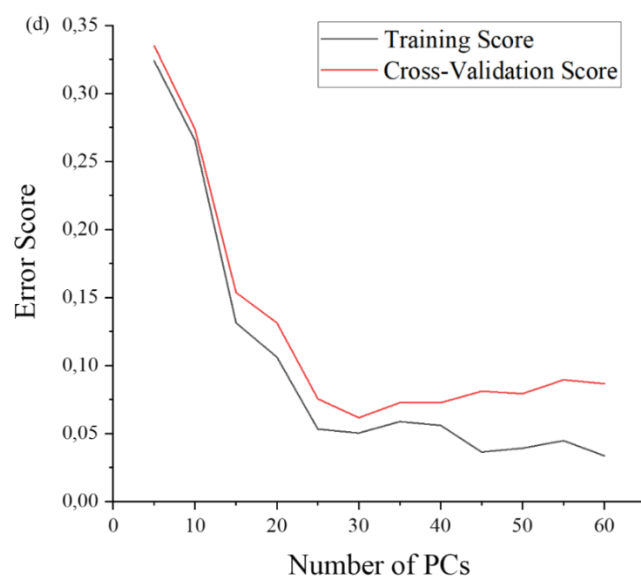


Figure 7. Leave-one-out-cross- validation (LOOCV) as a function of the number of PCs.

Table 5. List of the first 35 PC component and the relative percentages of variance.

Principal Component Number	Percentage of Variance (%)	Cumulative (%)
1	51.284	51.284
2	14.309	65.593
3	6.017	71.610
4	5.044	76.654
5	4.097	80.750
6	2.442	83.193
7	2.282	85.475
8	1.677	87.152
9	1.198	88.350
10	0.869	89.219
11	0.749	89.968
12	0.683	90.651
13	0.518	91.169
14	0.504	91.673
15	0.308	91.981
16	0.289	92.270
17	0.245	92.515
18	0.232	92.747
19	0.199	92.946
20	0.170	93.117
21	0.166	93.283

22	0.146	93.429
23	0.134	93.562
24	0.123	93.686
25	0.116	93.802
26	0.106	93.908
27	0.096	94.005
28	0.094	94.098
29	0.087	94.185
30	0.087	94.272
31	0.085	94.356
32	0.083	94.439
33	0.079	94.519
34	0.077	94.596
35	0.076	94.671

3.6. CNN-based Classification (Supervised Learning)

In the neural models based on the CNN-LSTM network shown in **Figure 1**, the training phase is applied directly to the Raman spectra, which are partitioned into fixed-sizes windows. The training phase also includes the Hyper-parameter optimization phase, which had the task of determining:

- the optimal number of layers, which was chosen between 22 and 56;
- the most appropriate learning rate in the interval $[0.0001, 0.05]$;
- the best dropout chosen in the interval $[0.1, 0.25]$;
- the threshold of the reject option (in the interval $\{0.4, 0.8\}$);
- the type of augmentation to apply to the training set (selected from the set $\{\text{No augmentation, Frequency, Value, Both}\}$);
- the windows are chosen in order to obtain $k_s = 3$ with no overlap;
- a cross-validation approach with $k=5$ folds is used.

The results of the CNN-LSTM trained model are shown in **Table 6**. As can be seen from the table, the absolute best results for the largest network are obtained by applying a lower learning rate. This naturally also leads to a longer training time. However, it is interesting to note that the network with 25 layers, although achieving lower performance in the best case, is more competitive exceeding the 0.8 F1-score, for very lower epoch values than the network with 56 layers (≈ 80 epochs compared to ≈ 250), clearly indicating a trade-off, between training times and maximum performance, to be considered in this context. This is also confirmed by the trend of the loss function when epochs vary during training. This trend, shown in **Figure 8**, clearly indicates that as the number of layers increases and the learning rate decreases, the model requires a greater number of epochs to achieve the same level of performance. It is also interesting to analyze the impact of augmentation and reject options.

From the **Table 6**, it can be seen that the greatest effect is obtained by applying both types of augmentations, since individually they have little effect on the results. On the other hand, applying both types together allow the classifier to be more robust and to generalize better to real samples that have never been submitted to the network. This result is consistent for both the 25 and 56 models, but the impact is lower for the larger net.

As regards the reject option, the best threshold value is equal to 0.8 (high threshold). In this case, consistently and regardless of the use of the augmentation, the impact of the reject option is equal to $\approx 3\%$ of the overall performance (F1-score).

In summary, compared with the PCA-LDA approach, the CNN-LSTM model achieves slightly better F1-score but requires a higher training effort.

Table 6. Training and test results for the CNN-based classifier with different Hyper-parameters, reject options and augmentation strategies.

# Layers	LR	Dropout	Reject option	Augmentation	Prec	Rec	F1
25	0.001	0.15	LOW (0.2)	NO	0.816	0.816	0.816
			LOW (0.2)	YES (freq)	0.822	0.835	0.828
			LOW (0.2)	YES (value)	0.832	0.815	0.823
			LOW (0.2)	YES (both)	0.872	0.872	0.872
			HIGH (0.8)	NO	0.825	0.818	0.821
			HIGH (0.8)	YES (freq)	0.834	0.838	0.836
			HIGH (0.8)	YES (value)	0.847	0.848	0.847
			HIGH (0.8)	YES (both)	0.899	0.899	0.899
			LOW (0.2)	NO	0.826	0.826	0.826
			LOW (0.2)	YES (freq)	0.843	0.848	0.845
56	0.01	0.2	LOW (0.2)	YES (value)	0.865	0.877	0.871
			LOW (0.2)	YES (both)	0.899	0.899	0.899
			HIGH (0.8)	NO	0.902	0.921	0.911
			HIGH (0.8)	YES (freq)	0.912	0.904	0.908
			HIGH (0.8)	YES (value)	0.912	0.915	0.913
			HIGH (0.8)	YES (both)	0.943	0.917	0.930

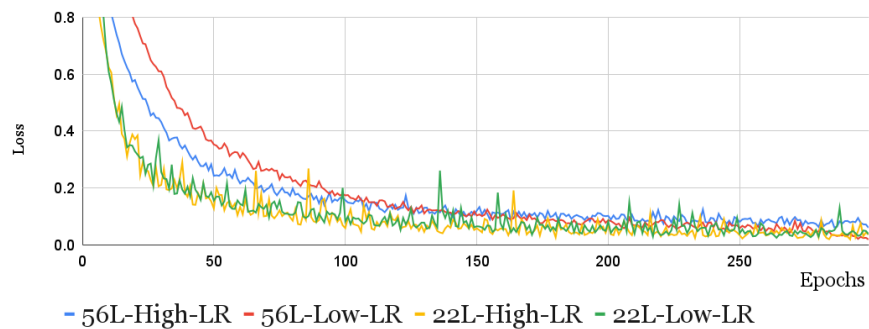


Figure 8. The loss function trend over epochs, during training, for CNN-LSTM models of 22 and 56 layers with different learning rates.

This is also confirmed by the confusion matrix presented in **Table 7**, which shows that the best neural model outperforms the best tuned PCA-LDA model for both true positives and false positives being equivalent for false negatives.

Table 7. Confusion matrix of the best tuned LSTM-CNN model (last row of **Table 3**).

Confusion matrix (LSTM-CNN)		
Predicted	Non-Tumor	Tumor
	33 (91.67%)	2 (5.40%)
	3 (8.33%)	35 (94.59%)
		True

3.7. Blind prediction of Tumor cells

In order to assess the performances of the developed artificial intelligence models, we prepared and analyzed cells with different ratios of Tumor to Non-Tumor cells and compared the obtained results with their nominal percentages. Specifically, we analyzed samples with the following ratios of Tumor to Non-Tumor cells: 5 to 0, 4 to 1, and 3 to 2. Raman spectra were acquired, pre-processed, and classified with the same setup instrument, procedures, and parameters used to build the training model.

Table 8 shows that, for the sample of 45 spectra containing only tumor cells, we correctly classified 40 and 39 spectra, with a percentage of correct predictions at the single-spectrum level of 89% and 87% for PCA-LDA and Hyper-parameter tuned LDA, respectively. For sample with a 4-to-1 Tumor/Non-Tumor cell ratio with a total number of 136 spectra, the percentage of spectra classified as a tumor are 82% and 80% for PCA-LDA and Hyper-parameter tuned LDA, respectively. For the sample with a 3-to-2 Tumor/Non-tumor cell ratio with a total number of 93 spectra, the percentages are 58% and 62% for PCA-LDA and Hyper-parameter tuned LDA, respectively. These results confirm the predictive ability of the two models.

As for the CNN-LSTM models, the last two rows in **Table 8** give the results for the three blind sets, as was done for PCA-LDA model. As can be seen from the table, all models trained using the Raman spectra are able to identify satisfactorily the proposed cell mixtures.

Table 8. Classification of samples with different ratios of Tumor and Non-Tumor cell.

	Tum (% Tum)	MIX1 (% Tum)	MIX2 (% Tum)
Nominal value	100.00%	80.00%	60.00%
Hyper-parameter tuned LDA	87.00%	80.10%	62.30%
PCA-LDA	89.00%	82.40%	58.00%
CNN-LSTM-22	91.60%	82.76%	58.33%
CNN-LSTM-56	92.70%	81.67%	61.54%

LDA proved to be particularly effective for MIX1, where 80% of the samples are cancerous, while for the other two mixes the 56-layer CNN-LSTM network gave a result significantly closer to the nominal values. It is worth noting that the performance difference between the best 25-layer and 56-layer neural networks is small, which can guide the choice in contexts with limited computer resources.

4. Conclusions

In this work, we analyzed uncultured primary human liver cancer by Raman spectroscopy. Specifically, 40 primary liver cells derived from resected HCC tumor tissue and 40 coming from the adjacent area of the HCC lesion were analyzed. Preliminarily, the cells were analyzed morphologically and spectrally. Morphological differences in shape and size were observed. In addition, the Raman spectra obtained provide detailed biochemical information about the studied sample, without the need for sample preparation. Indeed, key spectral differences were found, revealing a higher presence of nucleic acid in the nuclei of cancer cells than in non-cancer cells. Unsupervised Multivariate Analysis using principal components was performed, but clear differentiation of cells was not possible using these simple methods. Therefore, specific artificially intelligent routines were developed to analyze the same cells. Both Hyper-parameter tuned LDA and PCA-LDA methods were actually used to classify cells and make blind predictions for further cell datasets. In addition, a neural based approach integrating Convolutional and Recurrent neural networks (i.e., a CNN-LSTM model) was chosen to investigate whether this type of neural networks is more suitable for this task. An accuracy of nearly 90% was obtained for PCA-LDA and ≈ 0.93 for the best CNN-LSTM with 56 layers. These results confirm that the models are able to make predictions and classifications at single spectrum level, indicating the possibility of rapidly studying and classifying a HCC primary cell. We have shown that the synergy between Raman spectroscopy and machine learning in primary cancer cells can be used to discriminate between Tumor and Non-Tumor cells from liver tissues. This approach could provide an effective analytical tool for cancer diagnosis and rapid intraoperative classification. Further studies are currently underway to expand the training dataset and generalize the model.

5. References

- [1] J. M. Llovet *et al.*, “Immunotherapies for hepatocellular carcinoma,” *Nat. Rev. Clin. Oncol.*, vol. 19, no. 3, pp. 151–172, 2022, doi: 10.1038/s41571-021-00573-2.
- [2] J. M. Llovet *et al.*, “Hepatocellular carcinoma,” *Nat. Rev. Dis. Prim.*, vol. 7, no. 1, 2021, doi: 10.1038/s41572-020-00240-3.
- [3] H. Sung *et al.*, “Global Cancer Statistics 2020: GLOBOCAN Estimates of Incidence and Mortality Worldwide for 36 Cancers in 185 Countries,” *CA. Cancer J. Clin.*, vol. 71, no. 3, pp. 209–249, 2021, doi: 10.3322/caac.21660.
- [4] L. Rahib, B. D. Smith, R. Aizenberg, A. B. Rosenzweig, J. M. Fleshman, and L. M. Matrisian, “Projecting cancer incidence and deaths to 2030: The unexpected burden of thyroid, liver, and pancreas cancers in the united states,” *Cancer Res.*, vol. 74, no. 11, pp. 2913–2921, 2014, doi: 10.1158/0008-5472.CAN-14-0155.
- [5] D. Sia, A. Villanueva, S. L. Friedman, and J. M. Llovet, “Liver Cancer Cell of Origin, Molecular Class, and Effects on Patient Prognosis,” *Gastroenterology*, vol. 152, no. 4, pp. 745–761, 2017, doi: 10.1053/j.gastro.2016.11.048.
- [6] W. Wang and C. Wei, “Advances in the early diagnosis of hepatocellular carcinoma,” *Genes Dis.*, vol. 7, no. 3, pp. 308–319, 2020, doi: 10.1016/j.gendis.2020.01.014.
- [7] S. A. Taqi, S. A. Sami, L. B. Sami, and S. A. Zaki, “A review of artifacts in histopathology,” *J. oral Maxillofac. Pathol.*, vol. 22, no. 2, 2018, doi: 10.4103/jomfp.JOMFP_125_15.

- [8] T. T. Brunyé, E. Mercan, D. L. Weaver, and J. G. Elmore, "Accuracy is in the eyes of the pathologist: The visual interpretive process and diagnostic accuracy with digital whole slide images," *J. Biomed. Inform.*, vol. 66, pp. 171–179, 2017, doi: 10.1016/j.jbi.2017.01.004.
- [9] S. Elumalai, S. Managó, and A. C. De Luca, "Raman microscopy: Progress in research on cancer cell sensing," *Sensors (Switzerland)*, vol. 20, no. 19, pp. 1–19, 2020, doi: 10.3390/s20195525.
- [10] N. Blake, R. Gaifulina, L. D. Griffin, I. M. Bell, and G. M. H. Thomas, "Machine Learning of Raman Spectroscopy Data for Classifying Cancers : A Review of the Recent Literature," *Diagnostics (Basel)*, vol. 12, no. 6, 2022.
- [11] J. Wen *et al.*, "Detection and Classification of Multi-Type Cells by Using Confocal Raman Spectroscopy," *Front. Chem.*, vol. 9, no. April, pp. 1–8, 2021, doi: 10.3389/fchem.2021.641670.
- [12] L. Zhang *et al.*, "Raman spectroscopy and machine learning for the classification of breast cancers," *Spectrochim. Acta - Part A Mol. Biomol. Spectrosc.*, vol. 264, 2022, doi: 10.1016/j.saa.2021.120300.
- [13] Q. He, W. Yang, W. Luo, S. Wilhelm, and B. Weng, "Label-Free Differentiation of Cancer and Non-Cancer Cells Based on Machine-Learning-Algorithm-Assisted Fast Raman Imaging," *Biosens.*, vol. 12, no. 4, 2022, doi: 10.3390/bios12040250.
- [14] T. Tolstik *et al.*, "Discrimination and classification of liver cancer cells and proliferation states by raman spectroscopic imaging," *Analyst*, vol. 139, no. 22, pp. 6036–6043, 2014, doi: 10.1039/c4an00211c.
- [15] O. Ryabchykov *et al.*, "Raman spectroscopic investigation of the human liver stem cell line HepaRG," *J. Raman Spectrosc.*, vol. 49, no. 6, pp. 935–942, 2018, doi: 10.1002/jrs.5392.
- [16] M. Richter, O. Piwocka, M. Musielak, I. Piotrowski, W. M. Suchorska, and T. Trzeciak, "From Donor to the Lab: A Fascinating Journey of Primary Cell Lines," *Front. Cell Dev. Biol.*, vol. 9, no. July, pp. 1–11, 2021, doi: 10.3389/fcell.2021.711381.
- [17] R. E. Schwartz, H. E. Fleming, S. R. Khetani, and S. N. Bhatia, "Pluripotent stem cell-derived hepatocyte-like cells," *Biotechnol. Adv.*, vol. 32, no. 2, pp. 504–513, 2014, doi: 10.1016/j.biotechadv.2014.01.003.
- [18] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "ImageNet Classification with Deep Convolutional Neural Networks," *Commun. ACM*, vol. 60, no. 6, pp. 84–90, 2017, doi: 10.1145/3065386.
- [19] K. Greff, R. K. Srivastava, J. Koutník, B. R. Steunebrink, and J. Schmidhuber, "LSTM: A Search Space Odyssey," *IEEE Trans. Neural Networks Learn. Syst.*, vol. 28, no. 10, pp. 2222–2232, 2017, doi: 10.1109/TNNLS.2016.2582924.
- [20] T. Bocklitz, A. Walter, K. Hartmann, P. Rösch, and J. Popp, "How to pre-process Raman spectra for reliable and stable models?," *Anal. Chim. Acta*, vol. 704, no. 1–2, pp. 47–56, 2011, doi: 10.1016/j.aca.2011.06.043.
- [21] C. A. Lieber and A. Mahadevan-Jansen, "Automated method for subtraction of fluorescence from biological Raman spectra," *Appl. Spectrosc.*, vol. 57, no. 11, pp. 1363–1367, 2003.
- [22] R. Gautam, S. Vanga, F. Ariese, and S. Umapathy, "Review of multidimensional data processing approaches for Raman and infrared spectroscopy," *EPJ Tech. Instrum.*, vol. 2, no. 1, 2015, doi: 10.1140/epjti/s40485-015-0018-6.
- [23] C. Perlich, "Learning Curves in Machine Learning," in *Encyclopedia of Machine Learning*, C. Sammut and G. I. Webb, Eds. Boston, MA: Springer, Boston, MA, 2011, pp. 577–580. doi: 10.1007/978-0-387-30164-8_452.
- [24] X. Ying, "An Overview of Overfitting and its Solutions," in *Journal of Physics: Conference Series*, 2019, vol. 1168, no. 2. doi: 10.1088/1742-6596/1168/2/022022.
- [25] L. Yang and A. Shami, "On hyperparameter optimization of machine learning algorithms: Theory and practice," *Neurocomputing*, vol. 415, pp. 295–316, 2020, doi: 10.1016/j.neucom.2020.07.061.
- [26] D. S. Soper, "Greed is good: Rapid hyperparameter optimization and model selection using greedy k-fold cross validation," *Electron.*, vol. 10, no. 16, 2021, doi: 10.3390/electronics10161973.
- [27] Y. Geifman and R. El-Yaniv, "SelectiveNet: A Deep Neural Network with an Integrated Reject Option," 2019. [Online]. Available: <https://api.semanticscholar.org/CorpusID:59316904>
- [28] B. Qiuhui Lia, Kiera Rycaja, Xin Chena, and Dean G. Tanga, "Cancer stem cells and cell size: A causal link?," *Semin Cancer Biol.*, vol. 35, pp. 191–199, 2015, doi: 10.1016/j.semcancer.2015.07.002.
- [29] X. Bi, B. Rexer, C. L. Arteaga, M. Guo, and A. Mahadevan-Jansen, "Evaluating HER2 amplification status and acquired drug resistance in breast cancer cells using Raman spectroscopy," *J. Biomed. Opt.*, vol. 19, no. 2, 2014, doi: 10.1117/1.jbo.19.2.025001.
- [30] I. Notingher, "12 - Characterisation using Raman micro-spectroscopy," in *Tissue Engineering Using Ceramics and Polymers*, 1st ed., J. E. G. Aldo R. Boccaccini, Ed. Woodhead Publishing Series in Biomaterials, 2007, pp. 248–266.
- [31] Y. Tan, S. O. Konorov, H. G. Schulze, J. M. Piret, M. W. Blades, and R. F. B. Turner, "Comparative study using Raman microspectroscopy reveals spectral signatures of human induced pluripotent cells more closely resemble those from human embryonic stem cells than those from differentiated cells," *Analyst*, vol. 137, no. 19, pp. 4509–4515, 2012, doi: 10.1039/c2an35507h.
- [32] N. Stone, C. Kendall, J. Smith, P. Crow, and H. Barr, "Raman spectroscopy for identification of epithelial cancers," *Faraday Discuss.*, vol. 126, no. 1, pp. 141–157, 2004, doi: 10.1039/b304992b.

- [33] C. Ma *et al.*, “Single cell Raman spectroscopy to identify different stages of proliferating human hepatocytes for cell therapy,” *Stem Cell Res. Ther.*, vol. 12, no. 1, pp. 1–12, 2021, doi: 10.1186/s13287-021-02619-9.
- [34] J. M. Surmacki, B. J. Woodhams, A. Haslehurst, B. A. J. Ponder, and S. E. Bohndiek, “Raman micro-spectroscopy for accurate identification of primary human bronchial epithelial cells,” *Sci. Rep.*, vol. 8, no. 1, pp. 1–11, 2018, doi: 10.1038/s41598-018-30407-8.
- [35] J. Guo, B. Du, M. Qian, W. Cai, Z. Wang, and Z. Sun, “Raman spectroscopic identification of normal and malignant hepatocytes,” *Chinese Opt. Lett.*, vol. 7, pp. 60–63, Jan. 2009, doi: 10.3788/COL20090701.0060.
- [36] K. Czamara, K. Majzner, M. Z. Pacia, K. Kochan, A. Kaczor, and M. Baranska, “Raman spectroscopy of lipids: A review,” *J. Raman Spectrosc.*, vol. 46, no. 1, pp. 4–20, 2015, doi: 10.1002/jrs.4607.
- [37] S. Qiu *et al.*, “Raman profile alterations of irradiated human nasopharyngeal cancer cells detected with laser tweezer Raman spectroscopy,” *RSC Adv.*, vol. 10, no. 24, pp. 14368–14373, 2020, doi: 10.1039/d0ra01173h.
- [38] S. Managò *et al.*, “A reliable Raman-spectroscopy-based approach for diagnosis, classification and follow-up of B-cell acute lymphoblastic leukemia,” *Sci. Rep.*, vol. 6, 2016, doi: 10.1038/srep24821.
- [39] E. Szafraniec *et al.*, “Raman spectroscopy-based insight into lipid droplets presence and contents in liver sinusoidal endothelial cells and hepatocytes,” *J. Biophotonics*, vol. 12, no. 4, pp. 1–11, 2019, doi: 10.1002/jbio.201800290.
- [40] J. Li *et al.*, “Label-free Raman imaging of live osteosarcoma cells with multivariate analysis,” *Appl. Microbiol. Biotechnol.*, vol. 103, no. 16, pp. 6759–6769, 2019, doi: 10.1007/s00253-019-09952-3.
- [41] J. M. Surmacki, I. Quiros-Gonzalez, and S. E. Bohndiek, “Evaluation of Label-Free Confocal Raman Microspectroscopy for Monitoring Oxidative Stress In Vitro in Live Human Cancer Cells,” *Antioxidants (Basel)*, vol. 11, no. 3, 2022, doi: 10.3390/antiox11030573.
- [42] H. J. Byrne *et al.*, “Vibrational Spectroscopy: Disease Diagnostics and Beyond,” in *Optical Spectroscopy and Computational Methods in Biology and Medicine*, M. Baranska, Ed. Dordrecht: Springer Netherlands, 2014, pp. 355–399. doi: 10.1007/978-94-007-7832-0_13.
- [43] K. Tarao *et al.*, “DNA synthesis activities of hepatocytes from noncancerous cirrhotic tissue and of hepatocellular carcinoma (HCC) cells from cancerous tissue can predict the survival of hepatectomized patients with HCC,” *Cancer*, vol. 71, no. 12, pp. 3859–3863, 1993, doi: 10.1002/1097-0142(19930615)71:12<3859::aid-cncr2820711211>3.0.co;2-v.
- [44] U. Ben-David and A. Amon, “Context is everything: aneuploidy in cancer,” *Nat. Rev. Genet.*, vol. 21, no. 1, pp. 44–62, 2020, doi: 10.1038/s41576-019-0171-x.
- [45] Y. Koike *et al.*, “DNA content of hepatocytes in various stages of liver cirrhosis,” *Liver*, vol. 5, no. 3, pp. 156–161, 1985, doi: 10.1111/j.1600-0676.1985.tb00231.x.
- [46] M. Bou-Nader *et al.*, “Polyploidy spectrum: A new marker in HCC classification,” *Gut*, vol. 69, no. 2, pp. 355–364, 2020, doi: 10.1136/gutjnl-2018-318021.
- [47] A. Strickaert *et al.*, “Cancer heterogeneity is not compatible with one unique cancer cell metabolic map,” *Oncogene*, vol. 36, no. 19, pp. 2637–2642, 2017, doi: 10.1038/onc.2016.411.
- [48] R. A. Rocha, J. M. Fox, P. G. Genever, and Y. Hancock, “Biomolecular phenotyping and heterogeneity assessment of mesenchymal stromal cells using label-free Raman spectroscopy,” *Sci. Rep.*, vol. 11, no. 1, Feb. 2021, doi: 10.1038/s41598-021-81991-1.
- [49] G. R. Romualdo *et al.*, “In Vivo and In Vitro Models of Hepatocellular Carcinoma: Current Strategies for Translational Modeling,” *Cancers (Basel)*, vol. 13, no. 21, Nov. 2021, doi: 10.3390/cancers13215583.

Chapter 3: Uncovering biochemical signatures of circulating tumor cells: Raman spectroscopy meets machine learning for non-small cell lung cancer (NSCLC)

Abstract

Lung cancer remains the most prevalent cause of cancer-related mortality worldwide, highlighting the pressing need for improved diagnostic methodologies. This chapter introduces a label-free diagnostic approach that combines Raman spectroscopy (RS) with a Principal Component Analysis–Linear Discriminant Analysis (PCA–LDA) machine learning model to identify and characterize circulating tumor cell (CTC) models of Non-Small Cell Lung Cancer (NSCLC), while effectively distinguishing them from leukocytes, the main interferents in blood. Distinct Raman spectral and biochemical signatures were obtained for four NSCLC cell lines (A549, H1299, H1650, H1915) in suspension, revealing significant biochemical diversity and intrinsic heterogeneity among them. The optimized PCA–LDA model achieved an accuracy of classification of 96.4% in discriminating cancerous from normal cells and in differentiating between NSCLC subtypes. This integrated RS–ML platform provides a rapid, non-invasive, and highly reliable method for real-time CTC detection and molecular characterization, offering critical insights into tumor metabolism and advancing early diagnosis and personalized treatment in lung cancer care.

1. Introduction

Globally, lung cancer represents the most significant oncological challenge, accounting for both the highest incidence among all cancer types and the leading cause of cancer mortality. In 2022, there were approximately 2.48 million newly diagnosed cases and about 1.8 million deaths attributed to this malignancy [1]. Epidemiological projections are concerning, with estimates indicating that lung cancer incidence and mortality could reach 4-62 million new cases and 3-55 million deaths, respectively, by 2050 [2]. Lung cancer is widely recognized for its marked heterogeneity, encompassing a diverse range of clinical presentations, pathological features, and molecular profiles. It is primarily classified into two major types: non-small cell lung cancer (NSCLC), which accounts for roughly 85% of all cases, and small cell lung cancer (SCLC), comprising the remaining 15%. NSCLC is further categorized into subtypes such as adenocarcinoma, squamous cell carcinoma, and large cell carcinoma [3]. Adenocarcinoma is the most prevalent subtype, followed by squamous cell carcinoma, both associated with high mortality [4], [5]. Large cell carcinoma, a less common NSCLC variant is characterized by its lack of differentiation, as it does not display the specific microscopic features of either adenocarcinoma or squamous cell carcinoma. This poorly differentiated nature contributes to its aggressive behavior and resistance to standard chemotherapy [6], [7].

Lung cancer poses a critical healthcare challenge, as most cases are not detected until advanced stages, when treatment options are limited and survival rates decline significantly. Modern lung cancer diagnosis employs a comprehensive approach combining screening and imaging methods, such as chest X-ray, low-dose computed tomography (LDCT), and

magnetic resonance imaging (MRI) [7]. Chest X-ray is a commonly used, accessible, and low-risk initial test for lung cancer, but its limited ability to distinguish benign from malignant findings necessitates further imaging for accurate diagnosis [8]. LDCT has proven effective in reducing lung cancer mortality, but its effectiveness depends on screening frequency. Shorter intervals improve early detection but increase radiation exposure, costs, and false positives, making it important to balance benefits and risks [9]. MRI offers diagnostic performance comparable to LDCT, with lower false-positive rates and no radiation exposure. However, it is often considered a secondary option due to the impact of respiratory motion on image quality and overall accuracy [8].

Tissue biopsy remains the gold standard for definitive diagnosis. The main advantages of tissue biopsy include established diagnostic protocols and comprehensive histological characterization. However, significant limitations include its invasive nature, potential complications influenced by patient health, and the need for highly skilled medical personnel [10]. Additional factors limiting tissue-based approaches include difficulty accessing certain tumor locations, time-consuming processing that delays diagnosis, static sampling that cannot track tumor changes, and compromised molecular characterization from inadequate or poor-quality DNA specimens [11], [12], [13]. Comprehensive genomic testing is strongly recommended for patients with advanced or metastatic NSCLC to identify targetable genetic alterations that guide precision therapy and clinical trial eligibility. Key biomarkers include EGFR, ALK, KRAS, ROS1, BRAF, NTRK1/2/3, MET exon 14 skipping, RET, and HER2. Testing can be performed using tissue specimens, circulating tumor DNA from blood, or both, with parallel testing helping to minimize diagnostic delays. ALK and ROS1 rearrangements also play a vital role in selecting appropriate personalized targeted therapies [14]. Furthermore, next-generation sequencing (NGS) enables concurrent evaluation of numerous genetic markers, providing a more rapid and precise approach compared to individual gene analysis for enhancing personalized therapy in non-small cell lung cancer [15].

In recent years, liquid biopsy has emerged as a promising complementary diagnostic method, offering a minimally invasive and less expensive alternative to traditional tissue biopsy. By analyzing blood or other bodily fluids, liquid biopsy enables the detection of cancer-associated biomolecules and cellular components, such as circulating tumor DNA, RNA, exosomes or circulating tumor cells [16]. This approach allows real-time monitoring of tumor dynamics, facilitates comprehensive molecular profiling, and can be adopted when tissue availability is low or there is insufficient material for analysis [12]. It provides exceptional patient tolerance through simple minimally invasive procedures, eliminates procedural risks, enables serial monitoring for real-time disease assessment, captures comprehensive molecular information from multiple tumor sites, and accelerates clinical decision-making through rapid result availability [17]. These advantages may increase molecular diagnosis rates and improve access to targeted therapies.

Most lung cancer patients are diagnosed at advanced stages of disease progression, typically after metastatic lesions have developed in distant anatomical sites coinciding with symptom onset [18]. While multiple biomarkers demonstrate therapeutic value, inherent constraints limit their effectiveness. Circulating tumor DNA does not capture the full spectrum of tumor diversity, whereas RNA-based markers—including total RNA, cell-free RNA, microRNAs,

and long non-coding RNAs—suffer from molecular instability and significant measurement variability, compromising their clinical reliability and implementation [19]. Circulating tumor cells shed from tumor sites into the bloodstream offer valuable diagnostic information by mirroring the properties of primary and metastatic cancers. CTCs, found individually or in clusters within the systemic and lymphatic circulation, are extraordinarily scarce in advanced NSCLC patients [20]. The detection challenge is substantial: among the millions of peripheral blood mononuclear cells (10^6 – 10^7 per mL), only 1–10 CTCs can typically be found in the same volume [16]. In addition, peripheral blood CTC enumeration serves as a powerful predictor of lung cancer patient outcomes. Patients with higher CTC numbers at diagnosis and post-treatment typically experience worse clinical outcomes. Studies reveal that even minimal CTC presence—as few as one cell per 7.5 mL—can predict disease progression in locally advanced patients receiving chemotherapy or targeted therapies [21]. Such extreme rarity necessitates the development of highly sophisticated and sensitive isolation techniques. These cells serve as dynamic biomarkers carrying genetic and molecular information that reflects tumor biology, genotype, and invasiveness. Their clinical utility includes early disease detection, metastatic risk prediction, and real-time therapeutic monitoring [20]. However, detection faces significant challenges due to their extreme rarity in the bloodstream, particularly during early disease stages, short circulation half-life, phenotypic changes through epithelial-to-mesenchymal transition that enhance survival but complicate capture, and heterogeneous nature with variable molecular expression patterns [10].

Current CTC detection methods fall into two main categories: marker-based identification and physical separation techniques. Marker-based approaches target surface proteins such as EpCAM, found on epithelial cells, and CD45, present on leukocytes. The FDA-approved CellSearch® system (Veridex), which is based on EpCAM detection and blood cell exclusion (positive selection), has shown limited effectiveness in this context, particularly due to the downregulation or absence of EpCAM during epithelial-mesenchymal transition (EMT). Alternatively, negative depletion strategies focus on removing non-tumor cells using markers like CD45 [22]. In contrast, physical separation methods rely on intrinsic cellular traits such as size and density, offering a marker-independent approach. Systems like ISET® and ScreenCell® use size-based filtration and have demonstrated higher sensitivity in detecting CTCs in NSCLC. However, they may miss smaller or more deformable CTCs with atypical morphology [16]. While marker-based techniques benefit from standardized protocols and regulatory approval, they are less effective at capturing phenotypically altered CTCs. Physical methods improve recovery but may lack specificity or damage the cells [23]. Advancements in both strategies are essential to enhance sensitivity, speed, and clinical utility of CTC detection.

As highlighted in Chapters 1 and 2, Raman spectroscopy overcomes the limitations of current diagnostics, offering a label-free, rapid, and non-destructive approach ideally suited for detecting CTCs in lung cancer [24]. Recent studies have demonstrated the strong potential of Raman spectroscopy as a non-invasive diagnostic tool for lung cancer detection, classification, and molecular profiling. Zheng *et al.* showed that Raman spectroscopy can detect biochemical changes in lung tissues, including elevated lipid levels and altered protein-to-lipid and nucleic acid-to-lipid ratios in cancerous tissues [25]. This technique

successfully differentiated cancerous tissues from normal ones and distinguished between lung cancer subtypes, achieving diagnostic accuracies of 99.31% for identifying cancer and 92.20% for classifying adenocarcinoma versus squamous carcinoma. Furthermore, Raman spectroscopy was combined with a PCA/SVM algorithm to accurately detect EGFR mutations (L858R and E19del) in lung adenocarcinoma tissues, achieving 87.8% accuracy and high ROC values, demonstrating strong potential for clinical molecular diagnosis [26]. In a separate study, Kopec *et al.* employed Raman spectroscopy, imaging, and chemometric analysis to differentiate between normal lung epithelial cells (BEpiC) and cancerous lung cells (A549), along with their respective organelles [27]. The research identified several key Raman biomarkers linked to phosphorylation, lipid metabolism, oxidative phosphorylation, cholesterol levels, and cytochrome activity. Cancerous cells showed increased phosphorylation and *de novo* triglyceride synthesis, while exhibiting reduced cholesterol and cytochrome c levels compared to normal cells. Chemometric analysis demonstrated high sensitivity and specificity, with cross-validation sensitivities for A549 cells ranging from 87.5% to 100%, and specificities between 61.8% and 94.7% across different organelles. These results highlight the strong potential of this method for distinguishing between cell types and their organelles. In a different approach, Sun *et al.* developed a novel diagnostic method that combines Raman spectroscopy with PLS-LDA to accurately differentiate small cell and non-small cell lung cancer (SCLC vs NSCLC) at the cellular and tissue level [28]. This method showed excellent performance, with 100% accuracy in cell samples and 93.75% accuracy in tissue samples. The biochemical comparison revealed notable variations in amino acids, nucleic acids, and collagen between the two lung cancer types, supporting its potential to enhance diagnosis and guide treatment. In another study, Wen *et al.* focused on utilizing confocal Raman spectroscopy to detect live cells and employed advanced statistical methods to differentiate multiple cancer cell lines originating from different organs, including lung cancer cells (A549) [29]. By combining Raman spectroscopy data with classification techniques such as SVM, LDA and QDA, the study achieved classification accuracies exceeding 95%, demonstrating its effectiveness for early identification of circulating tumor cells and determining the origin of cancerous tissues. Furthermore, several research efforts have utilized surface-enhanced Raman spectroscopy (SERS) combined with sophisticated analytical methods to distinguish lung cancer cells from healthy cells and to differentiate among cancer subtypes. In one study, SERS paired with intracellular SERS nanoprobe and support vector machine (SVM) classification successfully identified three NSCLC subtypes (H1229, H460, A549) from leukocytes, reaching up to 95% accuracy in mixed samples and 100% accuracy when separating NSCLC cells from healthy leukocytes [30]. Another investigation conducted by Fang *et al.* used label-free SERS alongside SVM to classify NSCLC (H1975, H1703, H661), SCLC (H446) and healthy cells, including blood cells and immortalized lung cells (16HBE) achieving nearly 99% accuracy for NSCLC versus healthy cells and perfect accuracy for distinguishing SCLC from both normal cells and NSCLC [31]. Notable variations in nucleic acid-related Raman peaks were observed between NSCLC and SCLC cells. Furthermore, a SERS-enabled nanosensor was developed to detect lung cancer metastasis by analyzing healthy, primary, metastatic cells and patient blood samples, achieving full diagnostic sensitivity (100%) in preliminary tests through machine learning-based identification of metastatic

molecular signatures [32]. These findings demonstrate the powerful capability of Raman spectroscopy to precisely identify and characterize tumor cells, effectively distinguish them from normal leukocytes, and monitor cancer progression by analyzing the molecular signatures of circulating tumor cells. This underscores its potential as a sensitive and non-invasive tool for lung cancer diagnosis.

In this chapter, a Raman spectroscopy-based strategy was employed to establish a distinct spectral and biochemical fingerprint capable of characterizing and differentiating four non-small cell lung cancer (NSCLC) cell lines—A549, H1299, H1650, and H1915—analyzed in suspension as circulating tumor cell (CTC) models. These malignant cells were compared against normal leukocytes, considered the main source of interference in CTC detection. Subsequently, RS was further exploited to gain deeper insights into the metabolic landscape of the various cancer phenotypes, revealing remarkable biochemical variations among them. To enhance classification accuracy, a combined principal component analysis–linear discriminant analysis (PCA–LDA) model was applied, allowing precise discrimination not only between leukocytes and tumor cells but also among the individual NSCLC cell lines based on their specific molecular signatures. This analytical framework offers a potential decision-support tool for clinicians, fostering more objective and reproducible interpretation of diagnostic data while minimizing subjective bias. Importantly, the integrated Raman–PCA–LDA approach proposed represents the first comprehensive molecular characterization of this specific panel of NSCLC cell lines. It provides novel and detailed insights into their biochemical diversity, allowing clear distinction from normal cells. Overall, the Raman-PCA-LDA approach establishes a robust optical platform for the identification and classification of various non-small cell lung cancer CTC models, combining high diagnostic accuracy with a deeper understanding of tumor cell metabolism and heterogeneity.

2. Material and Methods

2.1 Sample collection and preparation

Four distinct non-small cell lung cancer (NSCLC) cell lines—A549, H1299, H1650, and H1915—as well as leukocytes isolated from the buffy coat of a healthy donor, were provided by IRCCS Istituto Romagnolo per lo Studio dei Tumori (IRST) “Dino Amadori”. Leukocytes were extracted from the buffy coat using density gradient centrifugation to ensure clean separation from other blood components. All cell types were cultured and then fixed in suspension in 2% paraformaldehyde (Sigma-Aldrich, Saint Louis, MO, USA) according to a standardized fixation protocol to preserve cellular integrity. After fixation, the cells were thoroughly washed and re-suspended in phosphate-buffered saline (PBS). Each cell type was aliquoted into separate tubes, each containing 1×10^6 cells in 1 ml of PBS, and stored at 4°C until Raman analysis. For measurements, 50 µl of each cell suspension was diluted with 150 µl of PBS and gently deposited onto calcium fluoride (CaF₂) slides, placed in Petri dishes. Raman measurements were performed after the cells had settled onto the Raman substrate by gravity, which occurred approximately one hour after deposition, ensuring stable and optimal Raman signal acquisition.

2.2 Raman analysis

Raman analysis of cellular samples was performed using a confocal Raman microscope system (LabRAM HR Nano system, Horiba Jobin Yvon S.A.S., Longjumeau Cedex, France) equipped with a frequency-doubled Nd:YAG laser operating at an excitation wavelength of 532 nm. Before data acquisition, the instrument was calibrated using a standard silicon wafer at 520.7 cm^{-1} as a reference point to ensure spectral accuracy. Sample illumination and signal collection were performed using a $60\times$ water immersion objective (Olympus, Tokyo, Japan) with a numerical aperture (NA) of 1.00, enabling high spatial resolution and efficient light collection from cellular specimens in suspension. To enhance the quality and resolution of the collected spectra, a confocal pinhole of $50\text{ }\mu\text{m}$ was used to limit out-of-focus light, while the acquisition parameters were optimized with an acquisition time of 30 seconds and two accumulations per spectrum, minimizing noise and improving the signal-to-noise ratio. The backscattered light was spectrally separated using a 300 lines per mm grating and detected with a CCD camera (Syncerity CCD Deep Cooled Camera, cooled to $-60\text{ }^{\circ}\text{C}$). For each cell type, five Raman spectra were acquired per individual cell, providing multiple measurements to capture spectral variability and ensure statistical robustness. Spectra were acquired across two specific spectral ranges: the fingerprint region (FP) ($400\text{--}1800\text{ cm}^{-1}$) and the high-wavelength (HW) region ($2800\text{--}3800\text{ cm}^{-1}$), encompassing the key vibrational modes most representative of the molecular constituents within the cells. This systematic approach yielded a comprehensive dataset of 625 spectra in total (125 spectra per cell line), facilitating accurate, reproducible comparisons across different lung cancer cell lines and leukocytes for downstream classification analysis.

2.3 Raman Data Pre-processing

The pre-processing of Raman spectra followed a carefully structured workflow to ensure data quality, consistency, and suitability for classification. The first step involved the spectral calibration performed with Labspec 6 software, where each spectrum was individually aligned by shifting its x-axis so that the reference peak (phenylalanine at 1002 cm^{-1}) is positioned correctly, thereby compensating for instrumental drifts and environmental fluctuations.

After spikes removal, linear interpolation was applied to account for differences in the x-axis sampling of individual spectra. Although all spectra were aligned at 1002 cm^{-1} , their spectral grids were not identical; thus, interpolation onto a common x-axis was required to ensure consistency and allow subsequent machine learning analysis. Next, data cutting was performed to restrict the analysis to two well-defined Raman shift regions: $460\text{--}1800\text{ cm}^{-1}$ and $2800\text{--}3024\text{ cm}^{-1}$, thereby eliminating spectral areas with excessive noise or irrelevant information, particularly at the spectrum's boundaries. This helped prevent analytical artifacts and focused on biologically relevant signals. After cutting, baseline correction was performed using polynomial fitting implemented in MATLAB, based on Lieber and Mahadevan-Jansen (2003) [33]. A third-order polynomial was applied in the fingerprint region and a zero-order (constant) fit in the high-wavelength region. A simple least-squares polynomial fit cannot efficiently reproduce the fluorescent background, since Raman peaks bias the fit. Therefore, an iterative algorithm (100 iterations) was applied, where points

above the fit were reassigned to the original spectrum, progressively excluding Raman peaks. The final baseline was then estimated and subtracted, yielding spectra oscillating around zero.

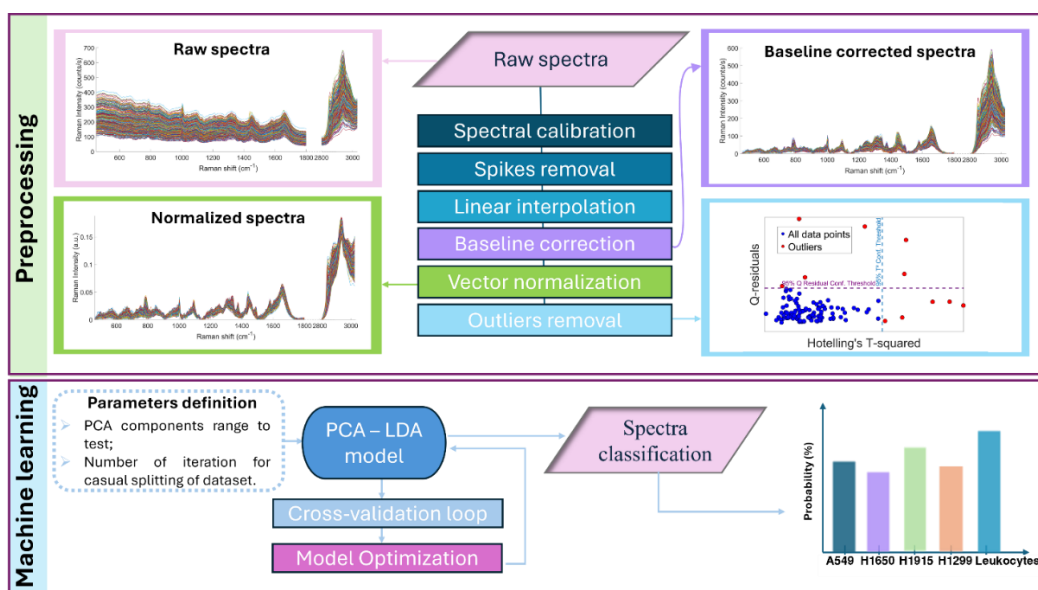


Figure 1. Schematic representation of the Raman data processing workflow. This flowchart outlines the main steps, starting with pre-processing of the data (spectral calibration, spike removal, linear interpolation, baseline correction, normalization and detection of outliers with Hotelling's T² and Q residual tests), followed by classification with an optimized PCA-LDA model.

Once corrected, the spectra were subjected to vector normalization, which scaled the intensity of each spectrum to unit length, compensating for differences in signal intensity that could arise from varying sample concentrations or measurement conditions. Finally, outlier detection was carried out using a Hotelling's T² versus Q-residuals test. In this approach, a scatterplot was generated with Hotelling's T² values on the x-axis and residuals on the y-axis. All samples falling outside the 95% confidence interval of this plot were considered outliers and removed. To calculate the Q residuals and Hotelling's T² values, Partial Least Squares (PLS) regression was used to model the data, with the number of latent components determined by RMSEC and RMSECV. Pre-processing steps are schematized in **Figure 1**. This comprehensive pre-processing pipeline was performed with MATLAB (MathWorks, Inc., United States) ensuring that the Raman data were clean, standardized, and optimally prepared for the subsequent PCA-LDA classification analysis.

2.4 Machine learning approach: PCA-LDA

In this study, a PCA-LDA algorithm was used for dimensionality reduction and classification of Raman spectral data, as shown in **Figure 1**. PCA first reduced the high-dimensional data to a set of uncorrelated principal components (PCs) capturing the most variance, and LDA was then applied to these scores to maximize separation between cell classes. To ensure accuracy and generalizability, a computational framework was developed for hyper-parameter tuning, focusing on optimizing the number of PCs and accurately assessing model performance. Using iterated Monte Carlo cross-validation (100 iterations) with dynamic 80-20 train-test splits, the number of PCs was optimized within a range of 30–50, with 48 PCs

ultimately selected to achieve the highest average cross-validated accuracy. The confusion matrix, along with performance metrics such as Precision, Recall, and F1-score for each class, was used to evaluate the performance of the chosen model and assess the accuracy, specificity, and sensitivity of the PCA-LDA method. Origin 2018 was used to perform PCA and MATLAB software was used to implement the PCA-LDA model.

3. Results and Discussion

3.1 Morphological Features of NSCLC cell lines and Leukocytes

The four tumor cell lines studied are cancer epithelial cells and were analyzed in suspension to more closely simulate the physiological state of circulating tumor cells found in the bloodstream *in vivo*. These conditions induced a pronounced shift in NSCLC cell morphology, from their usual flattened, spread-out shape in adherent culture to a rounder form in suspension for A549, H1299, H1650, H1915, respectively, as reported in **Figure 2a-d**.

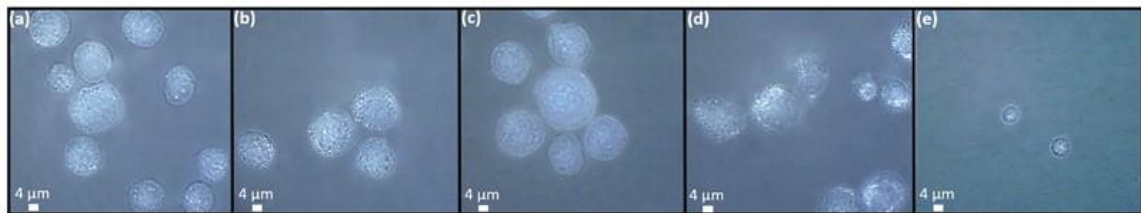


Figure 2. Bright-field images of non-small cell lung cancer cell lines, including A549 (a), H1299 (b), H1650 (c), H1915 (d) and normal cells, leukocytes (e), in suspension. (60 ×, scale bar 4 μm).

The suspension state caused a significant morphological change in the NSCLC cell lines, transforming their typical shape in adherent conditions into a more spherical form. This rounder morphology closely resembles that of CTCs observed in patient blood samples and supports the use of suspension models for studying tumor cell behavior during metastasis. In comparison, leukocytes, which are non-cancerous immune cells, naturally grow in suspension and consistently exhibit a round shape, as illustrated in **Figure 2e**. These cells are a typically round shape and are substantially smaller than NSCLC cells. This pronounced size difference, together with distinct morphological features, underscores the biological divergence between immune cells and epithelial tumor cells, reflecting their different developmental origins and functional roles. These findings are consistent with previous studies which reported that leukocytes measure less than 10 μm in diameter, whereas cancer cell lines commonly used in CTC research range from 15 to 25 μm, indicating a clear size disparity between the two cell types [36]. Supporting this, M. Hosokawa *et al.* demonstrated that lung cancer CTCs are larger than surrounding leukocytes [37]. Overall, these results emphasize the relevance of using suspension culture to model CTC behavior and validate that both morphological and dimensional distinctions between cancer cells and leukocytes are important for developing effective CTC analysis.

3.2 Spectral and biochemical fingerprints of CTC line models and leukocytes

To identify and compare the spectral fingerprints of different circulating tumor cell line models and leukocytes, Raman analysis was conducted on 25 individual cells per sample. For each cell, five Raman spectra were acquired, focusing on two key spectral regions: 460–1800 cm^{-1} and 2800–3024 cm^{-1} , corresponding to the fingerprint and high-wavelength regions, respectively. The fingerprint region reveals distinct vibrational signatures of biomolecules such as DNA, proteins, and lipids. The high-wavelength region captures vibrations related to lipids and proteins, providing additional molecular information [38]. Together, these regions contribute to a comprehensive biochemical profile of cells under investigation. The collected spectra were pre-processed to remove fluorescence interference and normalize intensity, after which the spectra from each sample were averaged to generate representative spectral profiles. The averaged spectra are presented in **Figure 3**, highlighting the distinct biochemical signatures of each cell type. Raman spectra showed that leukocytes have a distinct spectral signature compared to tumor cells, with notable differences in both the intensity and shape of several Raman peaks (**Figure 3**).

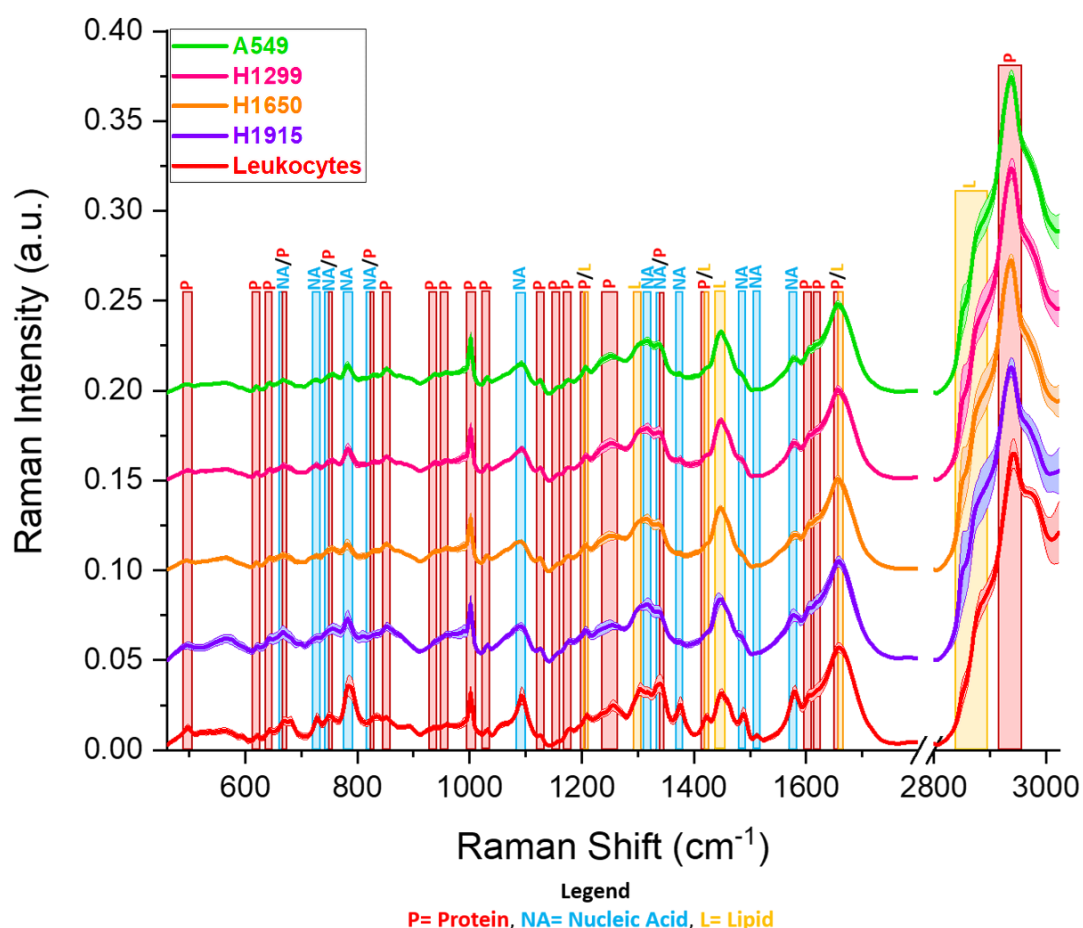


Figure 3. Normalized average Raman spectra of suspended NSCLC lung cancer cell lines (A549, H1299, H1650 and H1915) and leukocytes, each highlighted with different colors. The bands highlighted in red, blue, and yellow correspond to Raman peaks of proteins, nucleic acids, and lipids, respectively, as indicated in the legend. Combinations of these colors reflect non-exclusive assignments, indicating a mixture of biomolecules. P = proteins, NA = nucleic acids, L = lipids, NA/P = nucleic acids/proteins, P/L = proteins/lipids. The Raman spectra have been spatially offset for clarity and the shaded areas indicate the standard deviation for each spectrum.

In particular, the averaged spectrum of leukocytes exhibited a different shape of the Raman bands at 1093 and 1304-1340 cm^{-1} in the FP region. Additionally, a pronounced shoulder at approximately 2980 cm^{-1} was observed in the HW region, which was less evident in tumor cells. Compared to cancer cells, leukocytes also displayed significantly stronger Raman signals at characteristic peaks associated with nucleic acids (NA) and proteins (P), including 783 cm^{-1} (NA), 1093 cm^{-1} (NA), 1255 cm^{-1} (P), 1375 cm^{-1} (NA), 1421 cm^{-1} (P/L), and 1482 cm^{-1} (NA). A complete list of all identified Raman bands was provided in **Table 1**.

Table 1. Assignment of key Raman bands identified in the cell samples under investigation.

Raman shift (cm^{-1})	Assignment	Biomolecule	Reference
498	S—S disulphide str.	P	[39]
576-580	Amide IV, C=O bend.	P	[40]
618-622	C—C twist.	P	[41], [42]
643	Tyr, C—C twist.	P	[43]
654-657	Ala	P	[44]
668	G, T, Tyr, G bk. in RNA	NA/P	[41]
682	G, ring br.	NA	[45]
727-731	A	NA	[45]
748	T	NA	[43]
752-755	Trp, sym. br.	P	[43]
762	T	NA	[44]
783-790	U, C, T, ring br.	NA	[42], [45]
828-831	O—P—O str./Tyr, ring br.	NA/P	[41], [42]
834	PO ₂ asym str./Tyr/Diester sym str.	NA/P/L	[46]
845-848	Tyr, CH bend.	P	[44]
852-859	Tyr, ring br.	P	[41], [47]
862	Trp	P	[44]
872	C—C—N ⁺ sym. str.	L	[41]
937	C—C-str. alpha-helix, Pro, Val	P	[43]
951	Val	P	[44]
958-968	Tyr	P	[43], [45]
982	C—C head groups	L	[42]
1002-1005	Phe	P	[41]
1032	Phe, C—N str.	P	[41]
1079	C—C str./phospholipids/ C—O str.	P/L/C	[42]
1093-1096	DNA/RNA, PO ₂ ⁻ , sym str.	NA	[40]

1126	C—N str.	P	[41]
1156-1159	C—C/C—N str.	P	[41], [42]
1176-1179	Tyr, C—H bend.	P	[42]
1183	Nucleic acid base, C—N	NA	[47]
1202	Gly, Pro, Amide III, CH ₂ wag.	P	[48]
1206-1209	Phe, Amide III, =CH def.	P/L	[46]
1212	Phe, Tyr	P	[47]
1252-1258	Amide III	P	[49]
1304	CH ₂ twist.	L	[42]
1317-1320	G	NA	[40]
1337-1343	A, G/ Trp	NA/P	[42], [29]
1375	T,A,G	NA	[45]
1401-1404	Protein	P	[50]
1421-1424	CH def., CH ₃	P/L	[51]
1434-1450	phospholipids	L	[40], [52]
1482-1491	A, G	NA	[41], [43], [49]
1510-1514	A, ring br.	NA	[45]
1545-1552	Trp	P	[44]
1574--1580	A, G	NA	[41], [42]
1605	Tyr, Phe C=C	P	[41], [42]
1618/1628	Tyr, Trp, C=C Tyr	P	[42], [48]
1643	Amide I, C=O str.	P	[40]
1656-1659	Amide I α helix, C=C str.	P/L	[41], [42]
1665-1678	Amide I	P	[45], [52]
2846-2892	CH ₂ str.	L	[53], [27]
2916-2932	CH ₂ str.	L	[54], [55] , [56]
2938-2951	Sym. CH ₃ str.	P	[53], [57]
2972	Protein	P	[40]
2983-2988	Asym. CH ₃ str.	P	[53]

sym.= symmetric, asym. = asymmetric, str.= stretching mode, bend. = bending mode, twist. = twisting mode, def.= deformation mode, wag. = wagging mode, ring br.= ring breathing, bk= backbone, Ala = alanine, Gly= glycine, Phe = phenylalanine, Pro= proline, Trp = tryptofan, Tyr = tyrosine, Val = valine, A = adenine, C = cytosine, G= guanine, T = thymine, U = uracil.

To further investigate the biochemical differences between tumor cells and healthy immune cells, the mean spectral subtraction of tumor cells and leukocytes was performed, as shown in **Figure 4**. This spectral difference analysis enabled to highlight the specific Raman bands

that contribute most significantly to distinguishing cancerous from non-cancerous cells. In all plots, leukocytes consistently exhibited stronger Raman signals at the Raman shifts corresponding to the vibrational modes of nucleic acids, including 790, 1093, 1375, 1491, and 1580 cm^{-1} . These signals indicate of a high nuclear-to-cytoplasmic ratio typical of immune cells. Supporting this observation, Casabella *et al.* demonstrated that Jurkat cells—representative of white blood cells—exhibited pronounced Raman peaks corresponding to nucleic acids at 783, 1095 and 1378 cm^{-1} . They proposed that the relatively higher nucleic acid signal compared to lipid content in these cells could be attributed, at least in part, to their small size and the predominance of nuclear material, as lymphocytes typically possess a large nucleus relative to their cytoplasmic volume [58]. Moreover, leukocytes demonstrated enhanced protein content associated with peaks at 498, 1212, 1255–1258 (amide III) and 1671–1675 cm^{-1} (amide I), suggesting a relative protein enrichment in the FP region. Additionally, leukocytes exhibited greater intensity of protein-associated bands at 2988 cm^{-1} compared to the H1650 cell line (**Figure 4c**), and in the 2951–2986 cm^{-1} range relative to H1915 cells (**Figure 4d**). Furthermore, all examined lung cancer cell lines showed elevated intensities at lipid-associated Raman bands, including 1437 cm^{-1} , along with prominent peaks in the HW region (2852–2924 cm^{-1}) corresponding to the CH_2 stretching mode. These lipid-enriched spectral features reflect high membrane synthesis and increased lipid metabolism, processes often associated with cancer cell proliferation, membrane remodeling, and metabolic reprogramming.

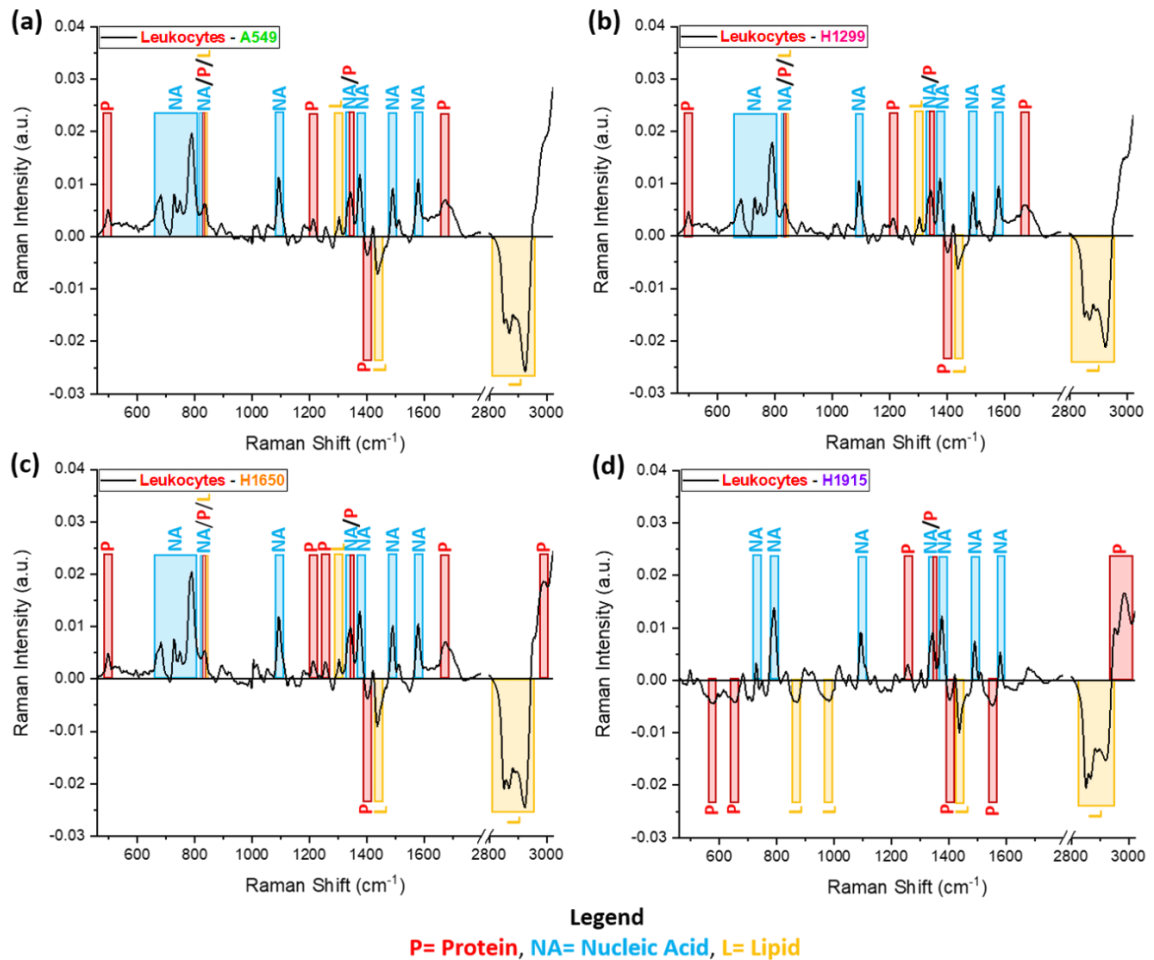


Figure 4. Comparative analysis based on the difference between the normalized average spectrum of leukocytes and that of individual non-small cell lung cancer (NSCLC) cell lines. Spectral differences are presented comparing leukocytes to suspended A549 (a), H1299 (b), H1650 (c), and H1915 (d), respectively. The bands highlighted in red, blue, and yellow correspond to Raman peaks of proteins, nucleic acids, and lipids, respectively, as indicated in the legend. Combinations of these colors reflect non-exclusive assignments, indicating a mixture of biomolecules. P = proteins, NA = nucleic acids, L = lipids, NA/P = nucleic acids/proteins, P/L = proteins/lipids.

In lung cancer, mutations in KRAS drive a lipogenic transcriptional program that promotes *de novo* fatty acid synthesis in both cell cultures and *in vivo*. Similarly, altered lipid metabolism is a feature of NSCLC with EGFR mutations, where EGFR signaling increases the synthesis of monounsaturated fatty acids, thereby facilitating tumor growth and progression [60]. The marked biochemical difference, involving the predominance of nucleic acids and proteins in leukocytes and the elevated lipid content in NSCLC cells, demonstrates the effectiveness of Raman spectroscopy in distinguishing immune cells from tumor cells based on their unique molecular profiles. This distinction is especially important for the label-free detection and characterization of rare circulating tumor cells within heterogeneous biological samples. These molecular contrasts also arise from their distinct cellular origins: leukocytes derive from hematopoietic immune lineages, whereas NSCLC cells stem from epithelial tumor lineages, resulting in fundamentally different biochemical compositions that are readily captured by Raman spectral analysis [61], [62]. Notably, Premachandran *et al.* used a SERS-based nanosensor to identify molecular differences between healthy, primary, and metastatic lung cancer cells, noting elevated fatty acids, reduced nucleic acids, and protein variability in cancer cells compared to healthy cells [32]. They also observed distinct clustering of primary and metastatic populations through principal component analysis, with increased nucleic acids and reduced proteins in metastatic cells (H1915) compared to primary cancer cells. In contrast, the simpler Raman spectroscopy approach proposed here provides similar insights into CTC heterogeneity and effective discrimination from normal cells, without the need to develop complex nanostructures, highlighting its practicality and cost-efficiency.

3.3 Assessment of heterogeneous biochemical landscapes of CTC line models

In addition to the pronounced differences between leukocytes and cancer cells, Raman spectral subtraction revealed distinct biochemical variations among the NSCLC cell lines, as reported in **Figure 5**, highlighting their molecular heterogeneity and metabolic landscapes. Specifically, A549 cells displayed elevated signals corresponding to proteins and lipids, with prominent protein-related bands at 937, 1002 cm^{-1} , and lipid-associated bands at 1447 and 2848–2927 cm^{-1} . In contrast, H1299 cells were enriched in nucleic acid signals, showing stronger peaks at 786, 1096, 1375, 1488, and 1580 cm^{-1} . These differences indicate distinct metabolic profiles between the two cell lines, with A549 favoring protein and lipid synthesis, while H1299 exhibit a higher nucleic acid content (**Figure 5a**). H1299 cells, derived from metastatic lymph node tissue of a lung cancer patient, exhibits high invasive and proliferative attitude and display mesenchymal-like properties, including reduced epithelial marker expression, in contrast to the more epithelial phenotype of weakly invasive A549 cells, which are derived from primary lung carcinoma [63], [64], [65]. Supporting these observations, the biochemical analysis showed that A549 cells are enriched

in proteins and lipids, indicative of active biosynthetic processes, whereas H1299 cells contain higher levels of nucleic acids, suggesting enhanced proliferative activity and more aggressive cellular behavior.

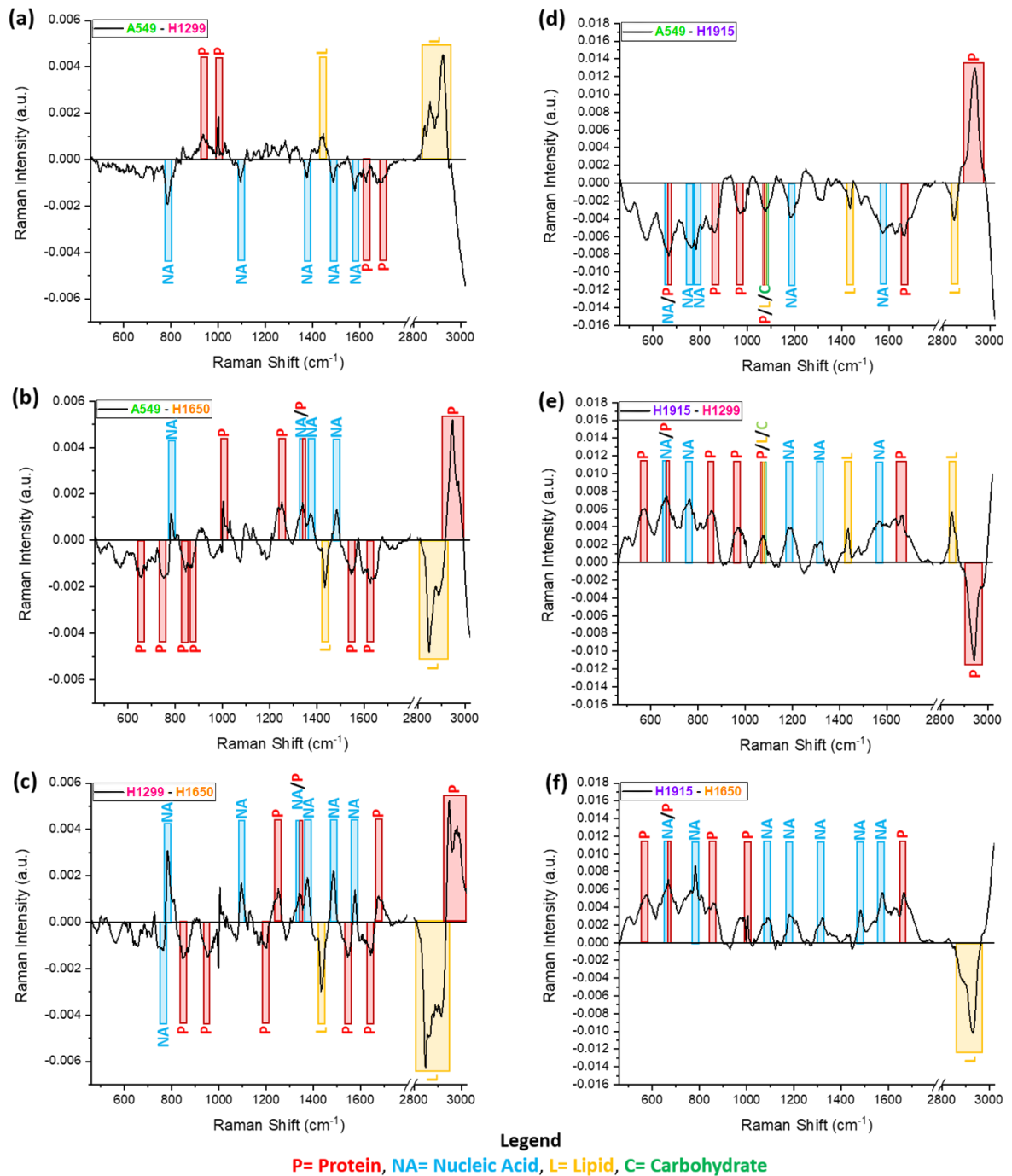


Figure 5. Subtraction analysis between the normalized average Raman spectra of A549 and H1299 (a), A549 and H1650 (b), H1299 and H1650 (c), A549 and H1915 (d), H1915 and H1299 (e), and H1915 and H1650 (f). The bands highlighted in red, blue, yellow and green correspond to Raman peaks of proteins, nucleic acids, lipids and carbohydrates, respectively, as indicated in the legend. Combinations of these colors reflect non-exclusive assignments, indicating a mixture of biomolecules. P = proteins, NA = nucleic acids, L = lipids, C= carbohydrates, NA/P = nucleic acids/proteins, P/L = proteins/lipids, P/L/C= proteins/lipids/carbohydrates.

Furthermore, differential analysis between A549 and H1650 cells revealed that A549 cells are enriched in proteins and nucleic acids, as indicated by stronger Raman peaks at 1002,

1252, 2948, and 2972 cm^{-1} (proteins) and 783, 1375, and 1485 cm^{-1} (nucleic acids). In contrast, H1650 cells exhibited a predominance of lipid-related signals, particularly at 1434, 2848, and 2887–2892 cm^{-1} , suggesting differences in lipid metabolism between the two cell lines (**Figure 5b**). Wang *et al.* investigated EGFR mutations in lung adenocarcinoma tissues using Raman spectroscopy and DNA sequencing, developing a PCA/SVM model for diagnostic purposes [26]. PC1 and PC2 loadings derived from PCA of wild-type EGFR tissues were compared with reference spectra of DNA, RNA, and proteins obtained from A549 (wild-type EGFR), H1650 (E19del mutation), and H1975 (L858R mutation) cell lines. Their analysis showed that E19del-mutant cells (H1650) had reduced nucleic acid content compared to wild-type EGFR tissue, consistent with our findings that also demonstrated lower nucleic acid levels in H1650 relative to A549, emphasizing consistency across different approaches.

In addition, similar to A549 cells, H1299 cells displayed a more pronounced presence of nucleic acids and proteins, with elevated peaks at 783, 1096, 1375, 1485, and 1577 cm^{-1} (nucleic acids) and 1678, 2948, and 2983 cm^{-1} (proteins), compared to H1650 cells. This further supports the observation that H1650 is more enriched in lipids, as indicated by strong signals at 1434 and 2848–2916 cm^{-1} (**Figure 5c**). Furthermore, comparison between A549 and H1915 cells revealed that A549 exhibited higher protein content, as indicated by the Raman peak at 2938 cm^{-1} , whereas H1915 showed relative enrichment in both lipids (1437, 2848 cm^{-1}) and nucleic acids (762, 783, 1183, 1574 cm^{-1}). The biochemical profile of H1915 may reflect a higher proliferative capacity (**Figure 5d**). Specifically, H1915 cells originate from a brain metastasis in a 61-year-old woman with poorly differentiated large cell lung carcinoma [66]. Recent studies have shown that both H1915 and H1299 non-small cell lung cancer cell lines exhibit high levels of HOXB9 expression, while A549 cells display low expression. HOXB9, frequently upregulated in NSCLC, plays a key role in driving epithelial-to-mesenchymal transition (EMT), thereby enhancing tumor cell motility and invasiveness [67]. This research supports the greater aggressiveness and metastatic potential observed in H1915 and H1299 cells relative to A549 cells. Moreover, H1915 exhibited higher levels of nucleic acids (762, 1183, 1320, 1574 cm^{-1}) and lipids (1437, 2846 cm^{-1}) compared to H1299, which retained stronger protein-associated signal at 2943 cm^{-1} . These differences may reflect variations in protein abundance between the two cell lines, as well as the greater aggressiveness of H1915, characterized by elevated nucleic acid and lipid content (**Figure 5e**). In contrast, H1915 cells were characterized by elevated levels of nucleic acids (783, 1093, 1183, 1320, 1482, and 1574 cm^{-1}) and proteins (576, 859, 1005, 1668 cm^{-1}) compared to H1650, reinforcing the lipid-dominant nature (2887 and 2932 cm^{-1}) of H1650 (**Figure 5f**). These nuanced differences among cancer cell lines highlight the capability of Raman spectroscopy to resolve not only cell type distinctions but also subtle biochemical variations among tumor subtypes, which may have implications for understanding tumor heterogeneity, drug responsiveness, or metastatic potential.

3.4 Cell sample discrimination by PCA-LDA model

Accurate pre-processing of Raman data is essential for obtaining reliable and reproducible results, as it removes instrumental artifacts and ensures that observed spectral differences

reflect true biological variations rather than noise or interferences. In particular, sharp intensity spikes from cosmic ray artifacts, which are unrelated to the sample's molecular composition, were eliminated first. This was followed by interpolation in order to align all spectra onto a common wavenumber axis, correcting for small instrumental deviations and promoting consistency across all Raman measurements [68]. Next, a baseline adjustment was performed to remove the fluorescence contribution from the spectra [69]. Normalization was then applied, as a crucial pre-processing step to correct for intensity differences arising from variations in sample thickness, density, laser focus, optical path length, as well as fluctuations in laser power and detector sensitivity [70]. This adjustment ensured that the spectra could be reliably compared across different measurements. After normalization, outlier removal was applied to exclude spectra that significantly deviate from expected patterns due to noise, contamination, or measurement errors, ensuring that aberrant data do not compromise the overall analysis. Outliers were identified using Hotelling's T^2 versus Q -residuals tests with a 95% confidence interval based on a partial least squares model. Overfitting in the PLS model was evaluated by assessing its fit to the training data and the reliability of predictions on unseen data using two metrics: the root mean square error of calibration (RMSEC) and the root mean square error of cross-validation (RMSECV). RMSEC measured the model's error on the calibration (training) data, while RMSECV estimated the model's predictive error on data not used in training, typically through cross-validation. In particular, these metrics assessed how well the model fit the training data and how reliably it predicted new data across different numbers of latent variables (LVs), respectively [71].

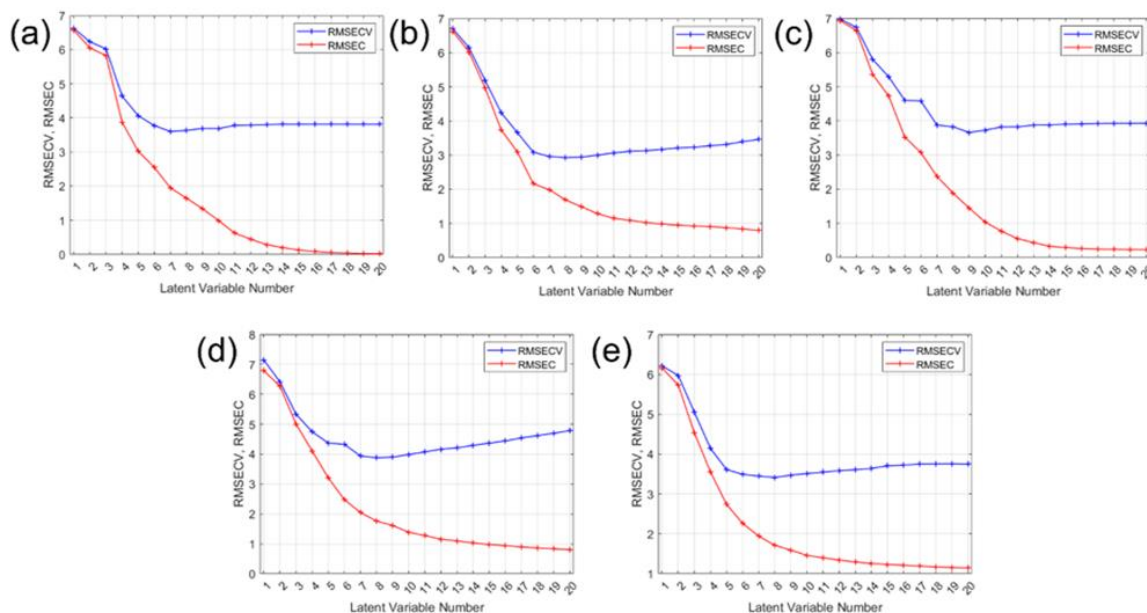


Figure 6. Evaluation of the number of latent variables to include in the partial least squares model based on two metrics: root mean square error of calibration (RMSEC) and root mean square error of cross-validation (RMSECV) for five cell samples: Leukocytes (a), A549 (b), H1299 (c), H1650 (d), H1915 (e).

A large gap between a low RMSEC and a significantly higher RMSECV indicates overfitting, where the model performs well on training data but poorly generalizes to new samples. To detect this, RMSEC and RMSECV values are plotted against the number of

latent variables used in the PLS model. Initially, both errors decline; however, if RMSEC continues to decrease while RMSECV plateaus or increases, this suggests that additional complexity is not improving predictive performance and may instead be capturing noise. The optimal model complexity is determined at the point where RMSECV is minimized or remains unchanged, balancing model accuracy with generalizability. In this study, as shown in **Figure 6**, analysis of RMSEC and RMSECV trends indicated that five latent variables provided the best balance, avoiding overfitting while ensuring reliable predictions. Furthermore, the Q statistic measured how well each sample fit within the model, while Hotelling's T^2 assessed the distance of sample scores from the model center, detecting unusual score patterns [72]. Samples that fell outside the confidence limits on this combined plot were classified as outliers and removed to maintain model integrity, as illustrated in **Figure 7**. Together, these pre-processing steps improve spectral quality, preserve important biochemical features, and enable the reliable detection of subtle molecular differences between the cellular samples investigated, supporting precise and consistent biomedical analysis. After removing outliers from the spectral dataset, PCA was performed to reduce the high-dimensional Raman spectra into a smaller set of orthogonal components that retained most of the variance.

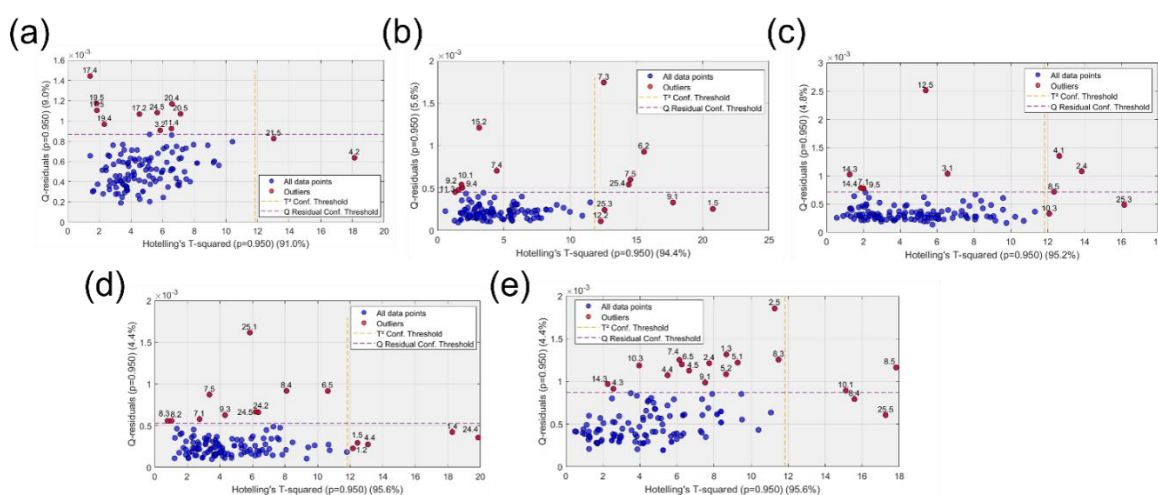


Figure 7. Outlier detection and removal by plotting Q residuals against Hotelling T^2 with a 95% confidence interval for a PLS model built with 5 latent variables for each cell line: Leukocytes (a), A549 (b), H1299 (c), H1650 (d), H1915 (e). In the plot, each outlier is marked by a numerical label, with the number preceding the dot corresponding to the cell and the number following the dot corresponding to the spectrum number. The percentage of variance explained by the first 5 latent components used to build the model is indicated on the abscissa.

As shown in **Figure 8a**, the first two principal components (PC1 and PC2) together explained 85.8% of the total variance. Among these, PC1 provided the most distinct separation between healthy and cancerous cell spectra, highlighting key biochemical differences between leukocytes and cancer cell lines. Interpretation of the PC1 loading plot revealed that positive PC1 values were primarily associated with Raman bands characteristic of lipids (1437, 2852, and 2887 cm^{-1}), which were more prominent in the cancer cell lines. Conversely, negative PC1 values corresponded to bands assigned to nucleic acids (786, 1093, 1375, 1491, and 1580 cm^{-1}) and proteins (1678 cm^{-1}), whose content was higher in leukocytes. These observations were consistent with previous comparative analyses of averaged normalized spectra from leukocytes and cancer cells. In contrast, PC2 did not

facilitate clear discrimination among the cancer cell lines. The Raman spectra of A549, H1299, and H1650 cells tended to cluster along the positive axis of PC2, with no distinct separation among them.

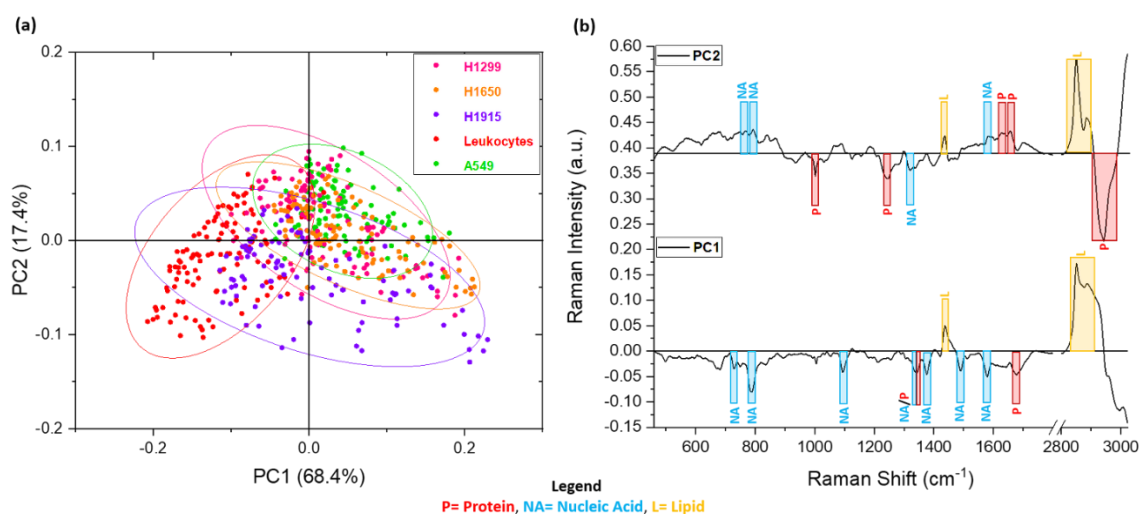


Figure 8. PCA 2D plot showing the distribution of spectral data from leukocytes and four suspended non-small cell lung cancer cell lines (A549, H1299, H1650, and H1915) based on the first two principal components (a). The ellipses represent a 95% confidence interval for the data. Loading plot of PC1 and PC2 components and the corresponding Raman band assignments was generated (b). The bands highlighted in red, blue, yellow and green correspond to Raman peaks of proteins, nucleic acids and lipids respectively, as indicated in the legend. Combinations of these colors reflect non-exclusive assignments, indicating a mixture of biomolecules. P = proteins, NA = nucleic acids, L = lipids. The loading plots of PC1 and PC2 are spatially offset for clarity.

Only partial separation of the H1915 cell line was observed relative to the others. Although the loading plot for PC2 (**Figure 8b**) identified additional Raman bands of interest, the overlap among cell line spectra made interpretation less conclusive. Overall, the inability of PCA to achieve well-defined clustering and distinct separation among the cancer cell lines highlighted its limitations as a standalone classification tool in this context. Consequently, more advanced classification methods were required to achieve accurate discrimination, particularly among the different NSCLC cell lines.

PCA outputs served as inputs to a supervised classifier, linear discriminant analysis, which aimed to maximize both intra- and inter-class separation. To ensure optimal performance, the PCA-LDA pipeline underwent rigorous parameter tuning through an optimized algorithm. This algorithm systematically tested different numbers of principal components, within a safe range of 30–50, and various training dataset splits to identify the configuration yielding the highest classification accuracy while avoiding overfitting. In a preliminary phase, iterated Monte Carlo cross-validation (20 iterations) was used to explore the relationship between the number of PCs and overfitting risk, revealing that retaining 30–50 PCs (explaining 90–95% of variance) achieved good model stability without overfitting. Because the optimal number of PCs depended on the specific training dataset used—due to the randomness of data splitting—a more comprehensive evaluation was implemented. The model was subjected to 100 iterations of 5-fold cross-validation, where in each iteration the dataset was randomly split into 80% training and 20% testing. This robust, iterated approach provided a statistically reliable estimate of model performance and allowed the algorithm to determine the best number of PCs for each split. For each principal component (within the

30–50 range), model accuracy was recorded across iterations, and the configuration with the highest average accuracy was selected. The final PCA-LDA model retained 48 PCs, explaining 98.38% of the total variance, and achieved a cross-validated accuracy of 96.25% on the training set and 96.40% accuracy on the independent test set. This thorough optimization process ensured both high performance and model generalizability across different data subsets.

Ultimately, the PCA-LDA model with the highest average cross-validated accuracy was selected and validated. The model's effectiveness was further confirmed through a PCA-LDA scatter plot (**Figure 9**), where each group of cells was projected onto the first three discriminant components, with distinct colors representing each category. The scatter plot revealed the clustering pattern of the samples, with clear separation indicating that the data contain sufficient discriminatory features to distinguish among the five groups. The percentages of variance explained by LD1, LD2, and LD3 indicate how much of the class-differentiating information is captured in this reduced space. In this context, the four NSCLC cancer cell lines formed well-separated clusters compared to healthy cells in the LDA feature space. This visualization highlighted the model's capacity to distinguish between cancerous and healthy cells based on their Raman spectral profiles, as illustrated in **Figure 9**.

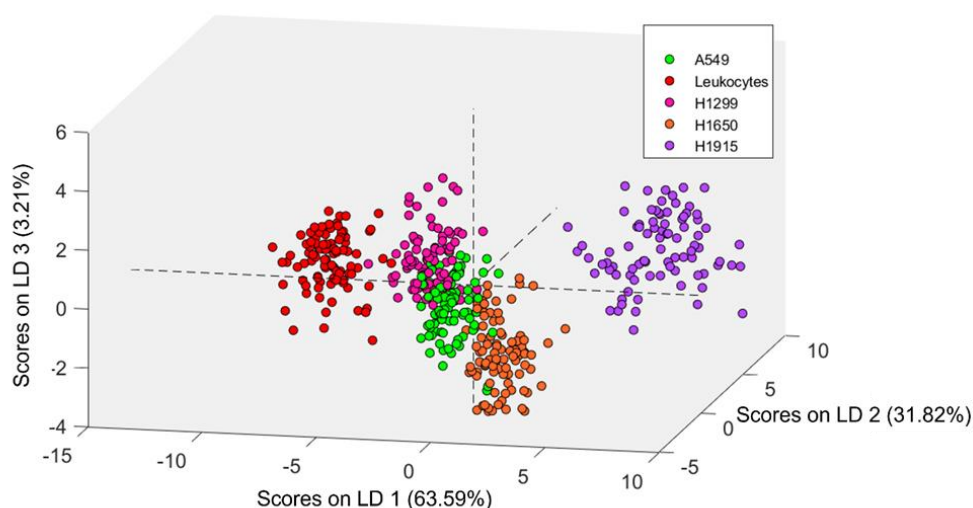


Figure 9. Three-dimensional scatter plot illustrating the LDA results for leukocytes and four lung cancer cell lines—A549, H1299, H1650, and H1915—represented by red, green, fuchsia, orange, and purple circles, respectively. The plot displays the separation of the five classes based on the first three linear discriminant components (LD1, LD2, and LD3), along with the percentage of variance each component explains. LDA was performed using the first 50 principal components as input.

The classification achieved by PCA-LDA emphasized the fundamental biological differences between leukocytes and cancer cells, while also capturing the metabolic heterogeneity among the various lung cancer cell lines. For a comprehensive evaluation, the model's predictive performance on the final test dataset was assessed using a confusion matrix and three key classification metrics: precision, recall, and F1-score, as shown in **Figure 10a-b**. Precision measured the proportion of true positives among all predicted positives, recall captured the proportion of actual positives correctly identified, and the F1-score provided a balanced measure that combined both precision and recall [73].

The confusion matrix provided a visual assessment of how well the algorithm distinguished between the five categories—leukocytes and the four lung cancer cell lines (A549, H1299,

H1650, and H1915), as reported in **Figure 10a**. Each cell in the matrix reflects how many samples from a given true class were assigned to each predicted class. High values along the diagonal indicate that the model consistently predicted the correct category, while lower values or off-diagonal entries represent misclassifications. The confusion matrix revealed a high rate of correct classifications with minimal misclassifications, indicating strong model generalization.

The classification metrics summarize the model’s performance numerically, reporting measures such as precision, recall, F1-score, and overall accuracy. These results provide a quantitative assessment of reliability, showing not only that the model achieves high accuracy but also that it maintains a balanced ability to correctly identify samples from all

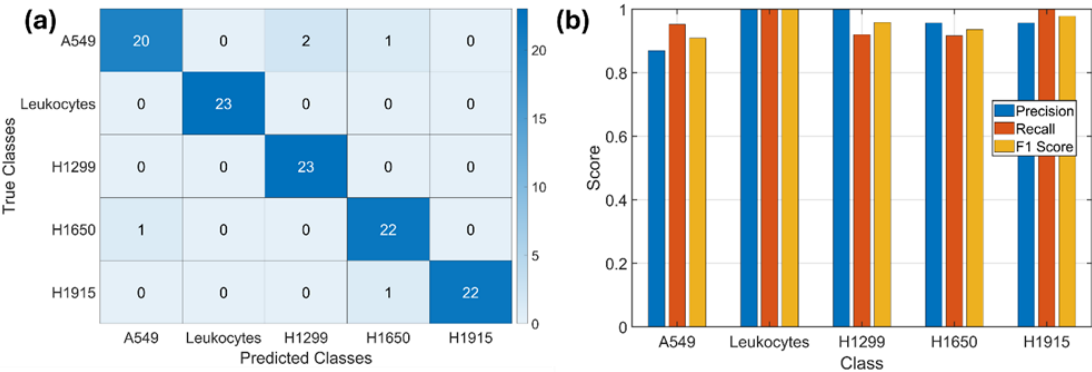


Figure 10. Confusion matrix (a) and classification metrics (b) showing the results of the proposed PCA-LDA model when applied to the test dataset.

five categories without biasing predictions toward any one class (**Figure 10b**).

Overall, these comprehensive results highlight the powerful integration of Raman spectroscopy with advanced data analysis techniques, enabling precise, label-free, and non-invasive classification of different cell types. Importantly, the ability to discriminate non-small cell lung cancer cell lines—used here as models for circulating tumor cells—from healthy leukocytes, as well as to distinguish between different cancer cell lines, underscores the clinical relevance of this approach. Given the critical role of CTCs in cancer progression, metastasis, and treatment monitoring, accurate identification and classification of these rare cells could significantly enhance early detection and diagnosis of lung cancer, supporting more personalized and timely therapeutic interventions.

4. Conclusions

This chapter demonstrated the significant potential of combining label-free Raman spectroscopy with advanced machine learning, particularly the PCA-LDA algorithm, as an effective and non-invasive diagnostic approach for lung cancer. By successfully differentiating four non-small cell lung cancer cell lines from normal leukocytes with an exceptional classification accuracy of 96.40%, the research provided unprecedented insights into the biochemical heterogeneity of these cancer cells, even when cultured in suspension to mimic CTCs in the bloodstream. The distinct molecular fingerprints captured by Raman analysis showed that leukocytes are rich in nucleic acids and proteins, while NSCLC cells have elevated lipid content, highlighting fundamental biological differences and the

metabolic reprogramming characteristic of cancer progression. Furthermore, the methodology's ability to resolve subtle biochemical variations among different NSCLC subtypes is crucial for understanding tumor heterogeneity and could have implications for personalized treatment strategies. Overall, this innovative approach offers a rapid, minimally invasive, and highly accurate method for the detection and characterization of circulating tumor cells, paving the way for improved early diagnosis, more precise disease monitoring, and ultimately, better clinical decision-making and patient outcomes in lung cancer management.

5. References

- [1] “International Agency for Research on Cancer. Global Cancer Observatory: cancer today. World Health Organization.” <https://gco.iarc.who.int/today> (accessed Jul. 06, 2025).
- [2] J. Zhou, Y. Xu, J. Liu, L. Feng, J. Yu, and D. Chen, “Global burden of lung cancer in 2022 and projections to 2050: Incidence and mortality estimates from GLOBOCAN,” *Cancer Epidemiol.*, vol. 93, no. September, p. 102693, 2024, doi: 10.1016/j.canep.2024.102693.
- [3] K. Inamura, “Lung cancer: understanding its molecular pathology and the 2015 WHO classification,” *Front. Oncol.*, vol. 7, no. AUG, pp. 1–7, 2017, doi: 10.3389/fonc.2017.00193.
- [4] A. A. Thai, B. J. Solomon, L. V. Sequist, J. F. Gainor, and R. S. Heist, “Lung cancer,” *Lancet*, vol. 398, no. 10299, pp. 535–554, 2021, doi: 10.1016/S0140-6736(21)00312-3.
- [5] S. C. M. Lau, Y. Pan, V. Velcheti, and K. K. Wong, “Squamous cell lung cancer: Current landscape and future therapeutic options,” *Cancer Cell*, vol. 40, no. 11, pp. 1279–1293, 2022, doi: 10.1016/j.ccell.2022.09.018.
- [6] R. Ruiz-Cordero and W. P. Devine, “Targeted Therapy and Checkpoint Immunotherapy in Lung Cancer,” *Surg. Pathol. Clin.*, vol. 13, no. 1, pp. 17–33, 2020, doi: 10.1016/j.path.2019.11.002.
- [7] F. H. Tang *et al.*, “Recent advancements in lung cancer research: a narrative review,” *Transl. Lung Cancer Res.*, vol. 14, no. 3, pp. 975–990, 2025, doi: 10.21037/tlcr-24-979.
- [8] A. Panunzio and P. Sartori, “Lung Cancer and Radiological Imaging,” *Curr. Radiopharm.*, vol. 13, no. 3, pp. 238–242, 2020, doi: 10.2174/1874471013666200523161849.
- [9] H. L. Lancaster, M. A. Heuvelmans, and M. Oudkerk, “Low-dose computed tomography lung cancer screening: Clinical evidence and implementation research,” *J. Intern. Med.*, vol. 292, no. 1, pp. 68–80, 2022, doi: 10.1111/joim.13480.
- [10] D. Di Capua, D. Bracken-Clarke, K. Ronan, A. M. Baird, and S. Finn, “The liquid biopsy for lung cancer: State of the art, limitations and future developments,” *Cancers (Basel)*, vol. 13, no. 16, pp. 1–22, 2021, doi: 10.3390/cancers13163923.
- [11] A. V. Ospina, “Overview of the Role of Liquid Biopsy in Non-small Cell Lung Cancer (NSCLC),” *Clin. Oncol.*, vol. 36, no. 10, pp. e371–e380, 2024, doi: 10.1016/j.clon.2024.07.004.
- [12] U. Malapelle, M. Tiseo, A. Vivancos, J. Kapp, M. J. Serrano, and M. Tiemann, “Liquid Biopsy for Biomarker Testing in Non-Small Cell Lung Cancer: A European Perspective,” *J. Mol. Pathol.*, vol. 2, no. 3, pp. 255–273, 2021, doi: 10.3390/jmp2030022.
- [13] L. Bonanno *et al.*, “Liquid biopsy and non-small cell lung cancer: are we looking at the tip of the iceberg?,” *Br. J. Cancer*, vol. 127, no. 3, pp. 383–393, 2022, doi: 10.1038/s41416-022-01777-8.
- [14] G. J. Riely *et al.*, “Non-Small Cell Lung Cancer, Version 4.2024, NCCN Clinical Practice Guidelines in Oncology,” *J. Natl. Compr. Cancer Netw.*, vol. 22, no. 4, pp. 249–274, 2024, doi: 10.6004/jnccn.2204.0023.
- [15] C. Lazzari *et al.*, “Next generation sequencing in non-small cell lung cancer: Pitfalls and opportunities,” *Diagnostics*, vol. 10, no. 12, pp. 1–12, 2020, doi: 10.3390/diagnostics10121092.
- [16] M. A. Molina-Vila *et al.*, “Liquid biopsy in non-small cell lung cancer,” *Front. Med.*, vol. 3, no. DEC, pp. 1–8,

2016, doi: 10.3389/fmed.2016.00069.

- [17] J. Zhou, A. Kulasinghe, A. Bogseth, K. O'Byrne, C. Punyadeera, and I. Papautsky, "Isolation of circulating tumor cells in non-small-cell-lung-cancer patients using a multi-flow microfluidic channel," *Microsystems Nanoeng.*, vol. 5, no. 1, 2019, doi: 10.1038/s41378-019-0045-6.
- [18] F. Ren, Q. Fei, K. Qiu, Y. Zhang, H. Zhang, and L. Sun, "Liquid biopsy techniques and lung cancer: diagnosis, monitoring and evaluation," *J. Exp. Clin. Cancer Res.*, vol. 43, no. 1, pp. 1–26, 2024, doi: 10.1186/s13046-024-03026-7.
- [19] D. Janjua, A. Chaudhary, U. Joshi, T. Tripathi, V. K. Jaggi, and A. C. Bharti, "Redefining cancer care: harnessing circulating tumor cells' potential for improved diagnosis and prognosis," *Cancer Cell Int.*, vol. 25, no. 1, 2025, doi: 10.1186/s12935-025-03883-y.
- [20] T. A. Allen, "The Role of Circulating Tumor Cells as a Liquid Biopsy for Cancer: Advances, Biology, Technical Challenges, and Clinical Relevance," *Cancers (Basel)*, vol. 16, no. 7, 2024, doi: 10.3390/cancers16071377.
- [21] A. B. Volovetsky *et al.*, "Prognostic Value of the Number of Circulating Tumor Cells in Patients with Metastatic Non-Small Cell Lung Cancer," *Micromachines*, vol. 16, no. 4, 2025, doi: 10.3390/mi16040470.
- [22] C. S. Dai, A. Mishra, J. Edd, M. Toner, S. Maheswaran, and D. A. Haber, "Circulating tumor cells : Blood-based detection , molecular biology , and clinical applications," *Cancer Cell*, pp. 1–24, 2025, doi: 10.1016/j.ccell.2025.07.008.
- [23] H. Qian, Y. Zhang, J. Xu, J. He, and W. Gao, "Progress and application of circulating tumor cells in non-small cell lung cancer," *Mol. Ther. Oncolytics*, vol. 22, no. September, pp. 72–84, 2021, doi: 10.1016/j.omto.2021.05.005.
- [24] M. B. Fenn, P. Xanthopoulos, G. Pyrgiotakis, S. R. Grobmyer, P. M. Pardalos, and L. L. Hench, "Raman spectroscopy for clinical oncology," *Adv. Opt. Technol.*, vol. 2011, 2011, doi: 10.1155/2011/213783.
- [25] Q. Zheng *et al.*, "Raman spectroscopy as a potential diagnostic tool to analyse biochemical alterations in lung cancer," *Analyst*, vol. 145, no. 2, pp. 385–392, 2020, doi: 10.1039/c9an02175b.
- [26] L. Wang *et al.*, "Evaluation of Raman spectroscopy for diagnosing EGFR mutation status in lung adenocarcinoma," *Analyst*, vol. 139, no. 2, pp. 455–463, 2013, doi: 10.1039/c3an01381b.
- [27] M. Kopec, K. Beton-Mysur, and H. Abramczyk, "Raman imaging and chemometric methods in human normal bronchial and cancer lung cells: Raman biomarkers of lipid reprogramming," *Chem. Phys. Lipids*, vol. 257, no. July, p. 105339, 2023, doi: 10.1016/j.chemphyslip.2023.105339.
- [28] W. Sun, L. Liao, R. Zhu, and Y. Wang, "Precise diagnosis of Small Cell and Non-Small Cell Lung Cancer based on Raman spectroscopy," *Photodiagnosis Photodyn. Ther.*, p. 104747, 2025, doi: 10.1016/j.pdpdt.2025.104747.
- [29] J. Wen *et al.*, "Detection and Classification of Multi-Type Cells by Using Confocal Raman Spectroscopy," *Front. Chem.*, vol. 9, no. April, pp. 1–8, 2021, doi: 10.3389/fchem.2021.641670.
- [30] Y. Zhang *et al.*, "Identification and distinction of non-small-cell lung cancer cells by intracellular SERS nanoprobes," *RSC Adv.*, vol. 6, no. 7, pp. 5401–5407, 2016, doi: 10.1039/C5RA21758J.
- [31] X. Fang, S. Li, Q. Fu, P. Wang, X. Wu, and Y. Zhang, "Label-free identification of lung cancer cells from blood cells based on surface-enhanced Raman scattering and support vector machine," *Optik (Stuttg.)*, vol. 248, no. October, p. 168157, 2021, doi: 10.1016/j.ijleo.2021.168157.
- [32] S. Premachandran, A. K. Dhinakaran, S. Das, K. Venkatakrishnan, B. Tan, and M. Sharma, "Detection of lung cancer metastasis from blood using L-MISC nanosensor: Targeting circulating metastatic cues for improved diagnosis," *Biosens. Bioelectron.*, vol. 243, no. August 2023, p. 115782, 2024, doi: 10.1016/j.bios.2023.115782.
- [33] C. A. Lieber and A. Mahadevan-Jansen, "Automated method for subtraction of fluorescence from biological Raman spectra," *Appl. Spectrosc.*, vol. 57, no. 11, pp. 1363–1367, 2003.
- [34] H. He and H. Maruta, "2 - Oncogenicity of PAKs and Their Substrates," in *Paks, Rac/Cdc42 (p21)-activated Kinases*, H. Maruta, Ed. Oxford: Elsevier, 2013, pp. 23–51. doi: <https://doi.org/10.1016/B978-0-12-407198-8.00002-3>.

- [35] S. Ju *et al.*, “Detection of circulating tumor cells: opportunities and challenges,” *Biomark. Res.*, vol. 10, no. 1, pp. 1–25, 2022, doi: 10.1186/s40364-022-00403-2.
- [36] S. J. Hao, Y. Wan, Y. Q. Xia, X. Zou, and S. Y. Zheng, “Size-based separation methods of circulating tumor cells,” *Adv. Drug Deliv. Rev.*, vol. 125, pp. 3–20, 2018, doi: 10.1016/j.addr.2018.01.002.
- [37] M. Hosokawa *et al.*, “Size-Based Isolation of Circulating Tumor Cells in Lung Cancer Patients Using a Microcavity Array System,” *PLoS One*, vol. 8, no. 6, 2013, doi: 10.1371/journal.pone.0067466.
- [38] L. E. Masson *et al.*, “Dual excitation wavelength system for combined fingerprint and high wavenumber Raman spectroscopy,” *Analyst*, vol. 143, no. 24, pp. 6049–6060, 2018, doi: 10.1039/c8an01989d.
- [39] J. M. Surmacki, I. Quiros-Gonzalez, and S. E. Bohndiek, “Evaluation of Label-Free Confocal Raman Microspectroscopy for Monitoring Oxidative Stress In Vitro in Live Human Cancer Cells,” *Antioxidants (Basel)*, vol. 11, no. 3, 2022, doi: 10.3390/antiox11030573.
- [40] J. M. Surmacki, B. J. Woodhams, A. Haslehurst, B. A. J. Ponder, and S. E. Bohndiek, “Raman micro-spectroscopy for accurate identification of primary human bronchial epithelial cells,” *Sci. Rep.*, vol. 8, no. 1, pp. 1–11, 2018, doi: 10.1038/s41598-018-30407-8.
- [41] R. J. Swain, S. J. Kemp, P. Goldstraw, T. D. Tetley, and M. M. Stevens, “Assessment of Cell Line Models of Primary Human Cells by Raman Spectral Phenotyping,” *Biophys. J.*, vol. 98, no. 8, pp. 1703–1711, 2010, doi: 10.1016/j.bpj.2009.12.4289.
- [42] I. Notingher, S. Verrier, S. Haque, J. M. Polak, and L. L. Hench, “Spectroscopic study of human lung epithelial cells (A549) in culture: Living cells versus dead cells,” *Biopolym. - Biospectroscopy Sect.*, vol. 72, no. 4, pp. 230–240, 2003, doi: 10.1002/bip.10378.
- [43] U. Neugebauer, J. H. Clement, T. Bocklitz, C. Krafft, and J. Popp, “Identification and differentiation of single cells from peripheral blood by Raman spectroscopic imaging,” *J. Biophotonics*, vol. 3, no. 8–9, pp. 579–587, 2010, doi: 10.1002/jbio.201000020.
- [44] H. Nawaz, A. Garcia, A. D. Meade, F. M. Lyng, and H. J. Byrne, “Raman micro spectroscopy study of the interaction of vincristine with A549 cells supported by expression analysis of bcl-2 protein,” *Analyst*, vol. 138, no. 20, pp. 6177–6184, 2013, doi: 10.1039/c3an00975k.
- [45] N. Chaudhary, T. N. Q. Nguyen, D. Cullen, A. D. Meade, and C. Wynne, “Discrimination of immune cell activation using Raman micro-spectroscopy in an in-vitro & ex-vivo model,” *Spectrochim. Acta - Part A Mol. Biomol. Spectrosc.*, vol. 248, p. 119118, 2021, doi: 10.1016/j.saa.2020.119118.
- [46] S. Managò, P. Mirabelli, M. Napolitano, G. Zito, and A. C. De Luca, “Raman detection and identification of normal and leukemic hematopoietic cells,” *J. Biophotonics*, vol. 11, no. 5, p. e201700265, May 2018, doi: 10.1002/jbio.201700265.
- [47] A. J. Hobro, Y. Kumagai, S. Akira, and N. I. Smith, “Raman spectroscopy as a tool for label-free lymphocyte cell line discrimination,” *Analyst*, vol. 141, no. 12, pp. 3756–3764, 2016, doi: 10.1039/c6an00181e.
- [48] G. Pezzotti, “Raman spectroscopy in cell biology and microbiology,” *J. Raman Spectrosc.*, vol. 52, no. 12, pp. 2348–2443, 2021, doi: 10.1002/jrs.6204.
- [49] M. Lasalvia, G. Perna, and V. Capozzi, “Raman spectroscopy of human neuronal and epidermal cells exposed to an insecticide mixture of chlorpyrifos and deltamethrin,” *Appl. Spectrosc.*, vol. 68, no. 10, pp. 1123–1131, 2014, doi: 10.1366/13-07299.
- [50] N. Pavillon and N. I. Smith, “Non-invasive monitoring of T cell differentiation through Raman spectroscopy,” *Sci. Rep.*, vol. 13, no. 1, pp. 1–10, 2023, doi: 10.1038/s41598-023-29259-8.
- [51] S. Managò *et al.*, “A reliable Raman-spectroscopy-based approach for diagnosis, classification and follow-up of B-cell acute lymphoblastic leukemia,” *Sci. Rep.*, vol. 6, 2016, doi: 10.1038/srep24821.
- [52] A. Borek-Dorosz *et al.*, “Raman-based spectrophenotyping of the most important cells of the immune system,” *J. Adv. Res.*, vol. 41, pp. 191–203, 2022, doi: 10.1016/j.jare.2021.12.013.
- [53] D. W. Shipp, F. Sinjab, and I. Notingher, “Raman spectroscopy: techniques and applications in the life sciences,” *Adv. Opt. Photonics*, vol. 9, no. 2, p. 315, 2017, doi: 10.1364/aop.9.000315.

- [54] C. Aksoy and F. Severcan, "Role of vibrational spectroscopy in stem cell research," *Spectrosc. (New York)*, vol. 27, no. 3, pp. 167–184, 2012, doi: 10.1155/2012/513286.
- [55] K. Lin, D. L. P. Cheng, and Z. Huang, "Optical diagnosis of laryngeal cancer using high wavenumber Raman spectroscopy," *Biosens. Bioelectron.*, vol. 35, no. 1, pp. 213–217, 2012, doi: 10.1016/j.bios.2012.02.050.
- [56] N. Mehta *et al.*, "Multimodal Label-Free Monitoring of Adipogenic Stem Cell Differentiation Using Endogenous Optical Biomarkers," *Adv. Funct. Mater.*, vol. 31, no. 43, pp. 1–16, 2021, doi: 10.1002/adfm.202103955.
- [57] J. Surmacki, J. Musial, R. Kordek, and H. Abramczyk, "Raman imaging at biological interfaces: Applications in breast cancer diagnosis," *Mol. Cancer*, vol. 12, no. 1, 2013, doi: 10.1186/1476-4598-12-48.
- [58] S. Casabella, P. Scully, N. Goddard, and P. Gardner, "Automated analysis of single cells using Laser Tweezers Raman Spectroscopy," *Analyst*, vol. 141, no. 2, pp. 689–696, 2016, doi: 10.1039/c5an01851j.
- [59] K. Eltayeb, S. La Monica, M. Tiseo, R. Alfieri, and C. Fumarola, "Reprogramming of Lipid Metabolism in Lung Cancer: An Overview with Focus on EGFR-Mutated Non-Small Cell Lung Cancer," *Cells*, vol. 11, no. 3, 2022, doi: 10.3390/cells11030413.
- [60] C. M. Dowling, H. Zhang, T. N. Chonghaile, and K.-K. Wong, "Shining a light on metabolic vulnerabilities in non-small cell lung cancer.," *Biochim. Biophys. acta. Rev. cancer*, vol. 1875, no. 1, p. 188462, Jan. 2021, doi: 10.1016/j.bbcan.2020.188462.
- [61] K. D. Sutherland and A. Berns, "Cell of origin of lung cancer," *Mol. Oncol.*, vol. 4, no. 5, pp. 397–403, 2010, doi: 10.1016/j.molonc.2010.05.002.
- [62] G. Litwack, "Chapter 21 - Blood and Lymphatic System," in *Human Biochemistry*, G. Litwack, Ed. Boston: Academic Press, 2018, pp. 681–708. doi: <https://doi.org/10.1016/B978-0-12-383864-3.00021-1>.
- [63] H. J. Huang *et al.*, "Identification of potential descriptors of water-soluble fullerene derivatives responsible for antitumor effects on lung cancer cells via QSAR analysis," *Comput. Struct. Biotechnol. J.*, vol. 19, pp. 812–825, 2021, doi: 10.1016/j.csbj.2021.01.012.
- [64] C. H. Li *et al.*, "Cytoplasmic aryl hydrocarbon receptor regulates glycogen synthase kinase 3 beta, accelerates vimentin degradation, and suppresses epithelial–mesenchymal transition in non-small cell lung cancer cells," *Arch. Toxicol.*, vol. 91, no. 5, pp. 2165–2178, 2017, doi: 10.1007/s00204-016-1870-0.
- [65] L. Zu *et al.*, "Establishment and characterization of a novel highly malignant lung cancer cell line ZX2021H derived from a metastatic lymph node lesion," *Thorac. Cancer*, vol. 13, no. 8, pp. 1199–1210, 2022, doi: 10.1111/1759-7714.14385.
- [66] S. M. Nolte *et al.*, "A cancer stem cell model for studying brain metastases from primary lung cancer," *J. Natl. Cancer Inst.*, vol. 105, no. 8, pp. 551–562, 2013, doi: 10.1093/jnci/djt022.
- [67] H. S. Zheng *et al.*, "HOXB9 enhances the ability of lung cancer cells to penetrate the blood-brain barrier," *Aging (Albany. NY)*, vol. 13, no. 4, pp. 4999–5019, 2021, doi: 10.18632/aging.202324.
- [68] D. Hutsebaut, P. Vandenabeele, and L. Moens, "Evaluation of an accurate calibration and spectral standardization procedure for Raman spectroscopy," *Analyst*, vol. 130, no. 8, pp. 1204–1214, 2005, doi: 10.1039/b503624k.
- [69] B. Barton, J. Thomson, E. Lozano Diz, and R. Portela, "Chemometrics for Raman Spectroscopy Harmonization," *Appl. Spectrosc.*, vol. 76, no. 9, pp. 1021–1041, 2022, doi: 10.1177/00037028221094070.
- [70] A. Fatima, G. Cyril, V. Vincent, J. Stéphane, and P. Olivier, "Towards normalization selection of Raman data in the context of protein glycation: Application of validity indices to PCA processed spectra," *Analyst*, vol. 145, no. 8, pp. 2945–2957, 2020, doi: 10.1039/c9an02155h.
- [71] A. S. Luna and J. S. de Gois, "Chapter Seven - Application of Chemometric Methods Coupled With Vibrational Spectroscopy for the Discrimination of Plant Cultivars and to Predict Physicochemical Properties Using R," in *Vibrational Spectroscopy for Plant Varieties and Cultivars Characterization*, vol. 80, J. Lopes and C. Sousa, Eds. Elsevier, 2018, pp. 165–194. doi: <https://doi.org/10.1016/bs.coac.2018.03.008>.
- [72] G. Nepomuceno, C. V. C. Junho, M. S. Carneiro-Ramos, and H. da Silva Martinho, "Tyrosine and Tryptophan vibrational bands as markers of kidney injury: a renocardiac syndrome induced by renal ischemia and reperfusion study," *Sci. Rep.*, vol. 11, no. 1, pp. 1–13, 2021, doi: 10.1038/s41598-021-93762-z.

- [73] M. Yu *et al.*, “A general framework for qualitative analysis of Raman spectroscopy based on deep learning,” *Microchem. J.*, vol. 199, no. January, p. 109990, 2024, doi: 10.1016/j.microc.2024.109990.

Chapter 4: Surface-enhanced Raman Spectroscopy (SERS) for Advanced Cancer Diagnosis

Abstract

The complex nature of cancer and the limitations of current diagnostic techniques require the development of innovative technologies with high sensitivity, specificity, multiplexing capability, and non-invasiveness, which are essential for a precise cancer diagnosis. In this regard, a promising bioanalytical technology is Surface-enhanced Raman Spectroscopy (SERS). This chapter provides an overview of the fundamental principles underlying the SERS phenomenon and reported the main types of plasmonic nanostructures employed as active substrates to enhance diagnostic sensitivity and accuracy. Additionally, it explores the design of various SERS-based bioassays and highlights innovative diagnostic platforms, such as SERS-integrated optical fibers. Recent advances in SERS applications for cancer biomarker detection and bioimaging are reviewed through selected key studies. Finally, the chapter addresses the current challenges and future perspectives of SERS in the field of cancer diagnostics.

1. Introduction

The global impact of a widespread, biologically complex and life-threatening disease such as cancer highlights the importance of detecting it at an early clinical stage to enable timely and accurate diagnosis. Identification of cancer at an advanced stage limits therapeutic interventions and contributes significantly to cancer-related mortality [1]. In addition, the variability in genetic and epigenetic profiles that occurs within a single tumor and between tumors from different patients –known as intra- and inter-tumor heterogeneity - poses a significant challenge to cancer diagnosis [2], [3]. In this context, the ability to quickly and accurately identify cancer biomarkers - biological molecules of different types, such as proteins, peptides, nucleic acids, etc. - which can provide valuable information about cancer, is crucial. Cancer biomarkers can be located on the cell membrane or within the cells or circulate in body fluids such as blood, saliva, urine and cerebrospinal fluid [4]. In particular, changes in the structure or function of cancer biomarkers provide valuable information about the presence, molecular mechanisms and type of cancer and serve not only as reliable diagnostic indicators but also as potential therapeutic targets. Accurate identification of cancer biomarkers can greatly assist clinicians in the rapid and effective development and application of treatment strategies, ultimately leading to better patient outcomes and enabling more precise and targeted treatment of the disease [5], [6].

As reported in the previous chapters, cancer diagnosis relies on various modalities, including imaging, tissue biopsy, and biomarker detection. While imaging techniques like X-ray, CT, MRI, and PET provide structural, metabolic, and functional insights, they have limitations such as radiation exposure, high cost, or procedural challenges [7]. Tissue biopsy with histopathology and immunohistochemistry (IHC) is highly accurate but invasive and less suitable for early detection or routine screening [8]. Biomarker detection using ELISA, PCR, NGS, and MS is constrained by complexity, low biomarker levels, high cost, and need for

skilled operators [9], [10], [11], [12]. Additionally, the complexity of the body fluid, the low concentration of cancer biomarkers and the need to identify multiple tumor-related indicators require necessitate the use of techniques that ensure a very high level of specificity, sensitivity and multiplexing capability. To overcome these challenges in cancer diagnosis, nanotechnology-based methods offer promising solutions. They have great potential to improve our understanding of cancer biology and facilitate their effective application in healthcare [13]. In this context, Surface-enhanced Raman spectroscopy is a highly sensitive technique for identifying the unique fingerprint of molecules by enhancing Raman scattering using specially designed metallic nanostructures. This improves the sensitivity of traditional Raman spectroscopy and enables the detection of low concentrations of biomolecules even in a complex matrix [14]. Additionally, recent advances in highly specialized nanofabrication device and techniques have enabled the production of sophisticated plasmonic nanomaterials, which are crucial for improving SERS in clinical applications [15]. Furthermore, bioengineering of SERS substrates enables the development of specialized SERS tags, i.e., nanoparticles functionalized to more precisely identify cancer-related molecules and distinguish them from other interfering factors, thereby improving the specificity of SERS [16]. Additionally, the SERS multiplexing capability enables the simultaneous analysis of multiple biomarkers in a single assay, making it particularly useful in areas such as cancer diagnosis [17]. This combination of high sensitivity and specificity, multiplexing, non-destructiveness and the ability to provide rapid results makes SERS especially valuable in the clinical setting, where rapid and accurate information on cancer biomarkers is essential for timely medical decisions.

This chapter presents an in-depth look at SERS principles, focusing on the two main phenomena that enhance the Raman signal: electromagnetic and chemical amplification. The development and characteristics of high performance SERS substrates are also discussed, highlighting their critical role in signal amplification and the significant influence of substrate design on enhancement capabilities. Direct and indirect SERS-based strategies for cancer biomarker detection are reviewed, along with an analysis of the respective advantages and limitations of each approach. Recent advancements in SERS biosensing and imaging are outlined, including the use of SERS for the detection and/or mapping of various cancer biomarkers in biological fluids, tumor cells, and tissues, as well as in combination with complementary techniques. Finally, the chapter addresses the current challenges associated with the clinical application of SERS technologies. SERS offers a non-invasive, sensitive and efficient method for detecting cancer at its earliest stages, making it a promising alternative to traditional diagnostic methods and opens up new avenues for personalized medicine and real-time monitoring of cancer progression.

2. SERS basic principles

SERS was first observed by Fleischmann in 1974, who attributed the enhanced Raman signal of pyridine on rough silver electrodes to a surface effect [18], [19]. Later, Van Duyne and Creighton confirmed and explained SERS as arising from both increased electric fields at metal surfaces and metal-molecule interactions [20], [21], [22]. Moskovits further highlighted the role of plasmon resonance in metallic nanostructures in enhancing SERS

signals [23]. Since then, extensive research has focused on elucidating the mechanisms behind SERS.

SERS is a powerful technique that amplifies the inherently weak Raman signal of a sample using plasmonic nanostructures. SERS relies on two main interactions: light-matter and light-metal, fundamentals to understand the phenomenology of SERS. Specifically, the light-matter interaction arises from the Raman effect—the inelastic scattering of light by the sample—which have been discussed in detail in Chapter 1. Raman scattering occurs in particular when incident photons with frequency ω_0 interact with matter and a small fraction of them are scattered inelastically (1 in 10^7 - 10^8) with different frequencies [22], [24]. Raman scattering is inherently weak due to its low spontaneous occurrence, limiting signal intensity. Traditional enhancement methods, such as increasing laser power or focusing the beam, can improve signal but risk damaging the sample [25], [26]. The discovery that molecules near metal nanostructures experience dramatically enhanced Raman signals has greatly increased the sensitivity of the technique and expanded its potential applications. Two theories have been proposed that contribute to the enhancement of the physical phenomenon of Raman scattering: the electromagnetic (EM) and the chemical (CM) mechanisms. The scattered intensities of Raman (I_{Raman}) and SERS (I_{SERS}) are related by the following equation (1) as described in [22]:

$$I_{\text{SERS}} = EF_{\text{SERS}} \times I_{\text{Raman}} = EF_{\text{EM,SERS}} \times EF_{\text{CM,SERS}} \times I_{\text{Raman}} \quad (1)$$

The term EF_{SERS} denotes the total SERS enhancement factor, which reflects the gain caused by the substrate. The total SERS enhancement is the product of two contributing factors: the electromagnetic ($EF_{\text{EM,SERS}}$) and the chemical enhancement ($EF_{\text{CM,SERS}}$).

The electromagnetic (EM) mechanism is the primary contributor to SERS enhancement, providing increases of 10^4 - 10^{11} , while the chemical (CE) mechanism contributes more modestly, typically 10–100. The EM principles are relatively well understood, whereas the details of CE amplification remain less clear [27]. This section will outline the key concepts behind both EM and CE mechanisms.

2.1 Electromagnetic enhancement mechanism

The electromagnetic (EM) mechanism in SERS enhances the Raman signal by amplifying both incident and scattered fields. When incident light illuminates plasmonic nanostructures, it excites the conduction electrons at the metal-dielectric surface, inducing collective oscillations known as surface plasmons (SPs). In this way, the collective oscillation of the free surface conduction electrons is induced and a resonance is generated (**Figures 1a-b**) [28], [29]. SPs are collective oscillations of conduction electrons, forming a negatively charged cloud around a lattice of positive ions. They occur only in materials with a negative real dielectric constant and a small positive imaginary dielectric constant, meaning that not all metals can support surface plasmon resonance [30], [31]. Gold and silver are the most commonly used materials, although other metals such as copper and aluminum also exhibit plasmon resonance [32]. Specifically, plasmons in plasmonic nanostructures with a radius much smaller than the wavelength of light, such as metallic nanoparticles, are called localized surface plasmons (LSPs). LSPs do not propagate, which means that their oscillations are confined to the particle itself [30]. The oscillatory behavior of conduction

electrons in metallic nanoparticles is driven by incoming electromagnetic excitation, which facilitates the separation of charge [33]. Then, when the frequency of the incident light matches the oscillation frequency of the electrons, a strong electromagnetic field is generated near the surface of the plasmonic nanostructure [22]. This event leads to localized surface plasmon resonance (LSPR), the optical phenomenon underlying the EM effect (**Figure 1b**). The excitation is enhanced when the wavelength of the incident radiation coincides with the LSPR peak, resulting in a red-shifted wavelength [32]. It has been found that the optimal excitation wavelength should be slightly shorter than the LSPR wavelength [33].

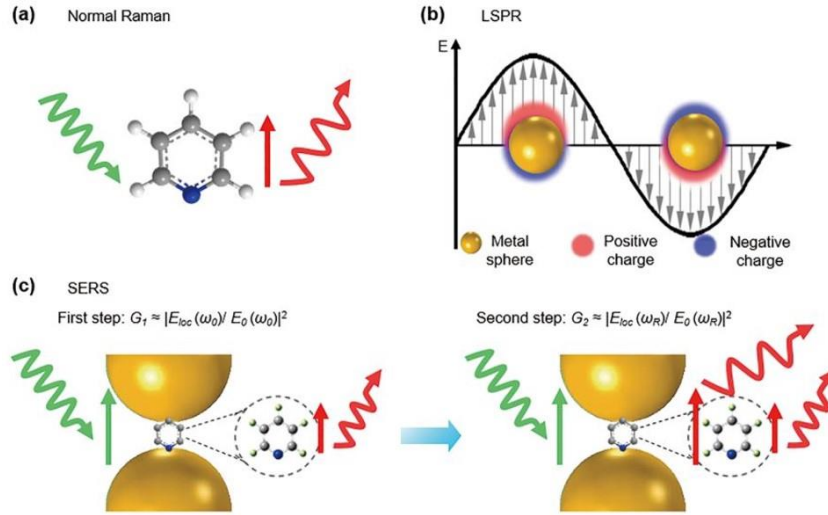


Figure 1. The diagram shows the basic phenomena underlying Surface Enhanced Raman Spectroscopy, including a) normal Raman scattering, b) localized surface plasmon resonance (LSPR), and c) the electromagnetic enhancement process in SERS, which involves two-step amplification, from Ref. [27].

Furthermore, the resonance wavelength and sensitivity depend on nanoparticle composition, size, shape, surrounding dielectric environment, and inter-particle distance [25], [38]. The SERS process involves two different stages of amplification. The first step consists of the amplification of the local incident field by LSPR, as shown in **Figure 1c**. In this context, gold nanoparticles (AuNPs) act as optical receiving antennas that convert the light from the far field into a localized near field around the nanoparticles, thus amplifying the incident light. In the subsequent phase, the AuNPs act as transmitting optical antennas to transfer the Raman signal from the near field to the far field, thus increasing the scattering photons in the Raman process [32], [33], [39]. In this way, the molecules near the metallic surface experience an intense electromagnetic field leading to their polarization and an increased induced dipole moment, which contributes to the enhancement of the intensity of Raman scattering [22]. Thus, the overall SERS enhancement is determined by both the excitation and emission fields. The enhancement factor (EF) in SERS quantifies the increase in a molecule's Raman signal due to the presence of plasmonic nanostructures, reflecting the efficiency of signal amplification [34]. The EF is significantly influenced by several factors, including the excitation wavelength, the SERS nanostructures and the analyte under investigation. Theoretically, the EF associated with the electromagnetic field is typically quantified using the equation (2) [35]:

$$EF_{EM,SERS} = G_1(\omega_0)G_2(\omega_R) = \frac{|E_{loc}(\omega_0)|^2}{|E_0(\omega_0)|^2} \frac{|E_{loc}(\omega_R)|^2}{|E_0(\omega_R)|^2} \approx \frac{|E_{loc}(\omega_R)|^4}{|E_0(\omega_0)|^4} \quad (2)$$

Where $G_1(\omega_0)$ and $G_2(\omega_R)$ represent the enhancement factor of the electromagnetic field for excitation and emission light, respectively. Moreover, ω_0 and ω_R indicate the incident and scattered Raman frequencies, while E_{loc} and E_0 denote the local electric fields with and without the presence of nanoparticles, respectively [35]. Additionally, the $EF_{EM, SERS}$ can be assumed to be approximately proportional to the fourth power of the enhancement of the local electric field when the wavelengths of the incident laser and the Raman scattering signal are close each other [27]. In addition, experimental $EF_{EM, SERS}$ measurements can be described by the following equation (3) [36]:

$$EF_{EM, SERS} = \frac{I_{SERS}(\omega_R)/N_{Surf}}{I_{Raman}(\omega_R)/N_{Vol}} \quad (3)$$

Here the SERS intensity ($I_{SERS}(\omega_R)$) is normalized by the number of molecules adsorbed on the nanoparticle (N_{Surf}) and compared to the unamplified Raman intensity ($I_{Raman}(\omega_R)$), which is normalized by the number of molecules in the scattering volume (N_{Vol}). However, the accurate measurement of molecular adhesion, scattering volume and surface area of the nanoparticle is a practical challenge [36].

2.1.1 SERS Hotspots

The electromagnetic enhancement factor of a single spherical nanoparticle is relatively low ($\sim 10^1$ – 10^3), insufficient for high-sensitivity applications such as cancer diagnostics [29]. However, when nanoparticles are closely spaced, they form nanogaps known as hotspots, where LSPR generates an intense electromagnetic field induced around the plasmonic nanostructure [37]. Hot spots typically form at junctions or gaps between noble metal nanoparticles, at sharply curved regions like corners or tips (the 'lightning rod' effect), or on surfaces with nanometer-scale roughness and other fine structural features [38]. In these nano-sized areas, the SERS enhancement can achieve magnitudes of 10^7 to 10^9 , which facilitates the detection of single molecules [29]. The local electromagnetic enhancement in SERS strongly depends on the gap size between nanoparticles, with optimal enhancement occurring at 2–10 nm [39]. Gaps smaller than 1 nm are limited by quantum effects, such as electron tunneling, which reduce field strength [40]. The intensity of the electric field enhancement in SERS hotspots decreases exponentially as the distance from the core of the hotspot increases along the surface of the nanoparticle [41]. Despite covering only about 1% of the substrate, the density of hotspots has a major impact on the Raman signal intensity, with greater density leading to stronger enhancement [22], [42]. The random distribution of hotspots presents a significant challenge for achieving consistent and reliable quantitative SERS detection, as this irregularity leads to variations in both the enhancement factor and Raman signal intensity across measurements [38].

To improve reproducibility and accuracy in SERS, it is essential to design substrates with a uniform and well-controlled distribution of hotspots. A single molecule positioned precisely within a hotspot can generate a signal equivalent to that of hundreds of randomly distributed molecules. In fact, about 80% of the total SERS signal originates from less than 1% of molecules located near hotspots, while molecules farther away contribute minimally [43].

Although electromagnetic enhancement predominantly determines SERS sensitivity, chemical enhancement also contributes to a comprehensive understanding of the overall mechanism.

2.2 Chemical enhancement mechanism

The second major mechanism contributing to SERS is chemical enhancement (CE), predominantly governed by charge transfer (CT) interactions [44]. While electromagnetic enhancement dominates in noble metal substrates such as gold and silver, CE is the prevailing mechanism in semiconductor-based SERS systems [37]. Although the nature of the chemical mechanism is still widely debated and not fully understood, it is generally believed to be based on the formation of a charge transfer complex between the metal and the molecule, determining the modification of polarizability of the molecule [22]. Specifically, the chemisorption of a molecule onto a metal surface induces the formation of new electronic states, altering the electronic properties of the adsorbed molecule and enabling additional electronic transitions within the surface complex [45], [46]. The interaction between incident light and the metal generates electron–hole pairs, which transfer energy to nearby molecular bonds, initiating the Raman process before returning the energy to the metal for scattering [22]. The electromagnetic effect is the dominant contributor to SERS, producing much stronger enhancement than the chemical mechanism. Unlike CM, which is molecule-specific and yields modest enhancement factors (10–100) over very short ranges, EM is analyte-independent, amplifies signals for all nearby molecules, and depends primarily on their distance from the metal surface. CM, in contrast, relies on changes in molecular polarizability and requires molecules to be within a few angstroms of the substrate [22], [47].

3. SERS substrates

The role of SERS in cancer diagnostics depends heavily on the effectiveness of the substrates used to amplify Raman signals. Ideal SERS substrates must be reproducible, stable, sensitive, uniform, cost-effective, easy to fabricate, non-toxic, and biocompatible. The choice and design of nanomaterials—based on size, composition, geometry, and surface functionalization—are key to achieving optimal SERS performance [48]. Therefore, the optimization of these parameters is a crucial aspect in the development of SERS substrates for diagnostic applications. The fabrication of SERS substrates is pivotal in determining the sensitivity, reproducibility, and overall performance of the technique. Two main fabrication strategies are involved in the development of SERS substrates: top-down and bottom-up. Top-down approaches involve the development of nanostructures from bulk materials, offering highly ordered and reproducible hotspot distributions [49], [50], [51]. Despite their ability to produce uniform signal enhancement, these methods are costly, time-consuming, require complex processes using advanced equipment, and offer limited analysis for larger biomolecules at nano-gaps level posing significant obstacles to large-scale production [51], [52], [53]. In contrast, bottom-up techniques assemble nanostructures from atoms or molecules in solution [49]. While these methods are more cost-effective and scalable, they suffer from nanoparticle aggregation and variability in hotspot distributions, which

compromises reproducibility [51], [54]. The following subsections provide an overview of the key SERS substrates commonly employed for cancer diagnostic applications, highlighting their types and functionalities.

3.1 Colloidal metallic nanoparticles (NPs)

Silver and gold spherical nanoparticles have been the predominant choice for SERS-based applications, due to their favorable plasmonic properties in the visible to near-infrared range, which directly impact the efficiency of Raman signal enhancement [55], [56]. Metal nanoparticles are produced using either top-down or bottom-up approaches. Top-down techniques, such as mechanical milling, thermal processing, and laser ablation, fragment bulk materials into nanosized particles but often lead to irregular shapes and modified surface properties. In contrast, bottom-up chemical reduction methods create nanoparticles by reducing metal ions in solution with the help of reducing and stabilizing agents. While this approach is simple and provides better control over particle formation, it can be limited by the toxicity, expense, and impurities of the chemicals involved [57]. In SERS application, gold nanoparticles (AuNPs) are widely chosen for their excellent chemical stability, whereas silver nanoparticles (AgNPs) exhibit a broader plasmon resonance range (300–1200 nm vs. Au's 500–1200 nm) and higher signal enhancement at a lower cost, but they are less stable due to their tendency to oxidize, compromising their long-term durability and functional efficiency [22], [58]. As a result, Au nanostructures are preferred for biomedical SERS applications due to their durability and favorable optical properties [59], [60]. However, the enhancement of SERS signals depends not only on the choice of nanomaterials but also on their shape, size, geometry, and aggregation, which influence hotspot formation and signal amplification. In terms of size, smaller nanoparticles (10–100 nm) generally generate stronger plasmonic effects and higher SERS enhancement, with those around 50 nm being preferred for their optimal balance of enhancement, stability, and ease of fabrication [34]. In contrast, metallic nanoparticles larger than 100 nm show reduced SERS enhancement due to charge redistribution and non-radiative transfer [61]. Despite their widely used, the moderate field enhancement provided by spherical metallic nanoparticles, can be improved by optimizing nanogaps and increasing hotspot density through the design of anisotropic and branched nanostructures [62]. Specifically, advancements in technology have enabled the development of a wide range of structures, as illustrated in **Figure 2**, including anisotropic nanostructures such as nanoflowers [63], nanordos [64], triangular nanoplates [65], nanostars [66], nanocages [67], nanorings [68], nanodisks [69], nanobypiramids [70], which are characterized by sharp edges or corners, which can generate highly intense hotspots. Interestingly, Tian et al. found that nanoparticle shape strongly affects SERS performance, with nanostars producing the highest enhancement due to their sharp tips and abundant electromagnetic hotspots [71].

Additionally, core-shell nanostructures are a type of composite nanomaterial used as SERS substrates, consisting of a central core and an outer shell, made of two different materials. In Au–Ag core-shell SERS systems, the synergistic interaction between the gold core and silver shell generates strong and tunable LSPR ranging from approximately 410 nm (Ag) to 520 nm (Au), leading to enhanced SERS sensitivity [72], [73]. Specifically, the gold core provides excellent chemical stability and facilitates surface functionalization, while the

silver shell contributes intense plasmonic activity, delivering superior optical and sensing performance compared to single-metal nanoparticles [74]. The characteristics of core-shell nanoparticles can be finely adjusted by modifying parameters such as

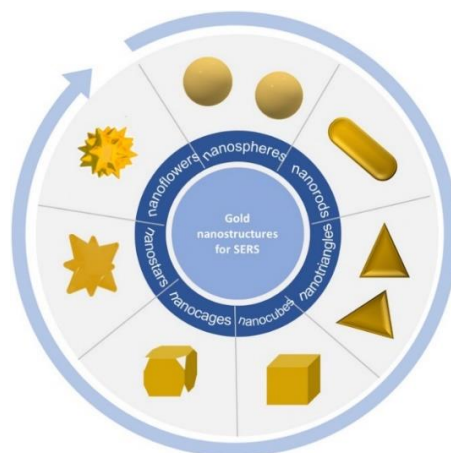


Figure 2. Illustration of various geometrically distinct gold nanostructures used as SERS substrates, from Ref. [62].

the materials used for the core and shell, the thickness of each layer, and the ratio between them, enhancing versatility, stability and biocompatibility [75]. For instance, García-Ramírez et al. utilized Au–Ag nanostars as an advanced SERS sensing platform for α -fetoprotein detection, reaching a detection limit of 16.73 ng/mL. The exceptional plasmonic properties of these Au–Ag nanostructures enabled highly sensitive and rapid biomarker analysis, underscoring their promise for early cancer diagnosis applications [76]. Moreover, although individual nanoparticles produce limited SERS effects, their interactions generate multiple hotspots, greatly enhancing SERS signals. This principle is widely used in diagnostic SERS applications through the design of controlled nanoparticle assemblies. One commonly employed strategy is the dynamic regulation of hotspots mediated by DNA [77]. In this regard, metallic nanoparticles are functionalized with specific DNA strands and a reporter molecule. When complementary DNA sequences are present, the nanoparticles cross-link to form organized networks. Within these networks, plasmonic coupling between nanoparticles enhances the SERS response of the reporter molecule, boosting sensitivity and detection efficiency [78].

3.2 Metallic nanostructured array

Enhancing SERS signals through engineered nanostructures requires not only optimized morphology but also stable and reproducible platforms. Since nanoparticles in solution are often unstable, immobilizing them on solid platform using chemical tethering methods, involving the covalent bonding of nanoparticles to the substrate surface, ensures better stability and uniform nanoparticle distribution [22], [31], [79]. Traditional SERS substrates made from planar dielectric support such as silicon and glass provide stable, uniform signals, while the development of metal nanostructures on these supports allows large, well-controlled, and reproducible nanostructure arrays [31]. In this regard, SERS-active substrates were created by functionalizing the surface with silanes, silicon-containing compounds with reactive groups, terminated with thiol (–SH) (see **Figure 3.1**) or amine (–NH₂) groups, which facilitate strong covalent binding of gold nanoparticles [80], [81].

Nanosphere lithography (NSL) is another bottom-up approach for producing nanostructure arrays by depositing metal onto self-assembled monolayers of polystyrene or silica nanospheres, which act as masks to form well-ordered metal nanostructures [82], [83]. By exploiting the self-assembly of polystyrene nanospheres at the air–water interface, Managò et al. fabricated three different efficient SERS planar substrates, which were also integrated onto optical fiber tips: close-packed gold-coated arrays, triangular nanoprisms created after sphere removal, and sparse arrays obtained by air plasma downsizing dense nanospheres, as reported in **Figure 3.2** [84]. Nanosphere lithography is a versatile, cost-effective method for producing nanopores and nano-island arrays, offering high precision, reproducibility, and tunable control over pattern size, shape, and arrangement [85], [86], [87], [88]. In another template-assisted fabrication approach, block copolymer nanopatterning was used to develop nanostructured arrays with precisely controlled nanogaps [89]. Post metal deposition on block copolymer–derived polystyrene templates self-assembly yields dense, periodic metal nanogap arrays over large areas, reaching a LOD of 100 nM for adenine detection.

Advanced nanofabrication methods, such as electron beam lithography (EBL) enable precise nano-patterning and control of SERS-active nanostructures. EBL offers nanometer-scale resolution and reproducibility [90], [91]. It operates by directing a focused electron beam (typically 10–50 keV) onto a solid substrate, commonly SiO_x/Si wafers coated with an electron-sensitive resist such as PMMA or PEG. The interaction between the electron beam and the resist induces chemical modifications that define the desired pattern. In the case of positive resists like PMMA, the exposed regions are removed, while for negative resists, the exposed regions undergo crosslinking, remaining intact, as depicted in **Figure 3.3** [92], [93]. Through careful control of exposure and resist type, EBL enables the fabrication of precise nanoscale features for SERS substrates. The various methods for SERS nanostructured array fabrication are reported in **Figure 3**.

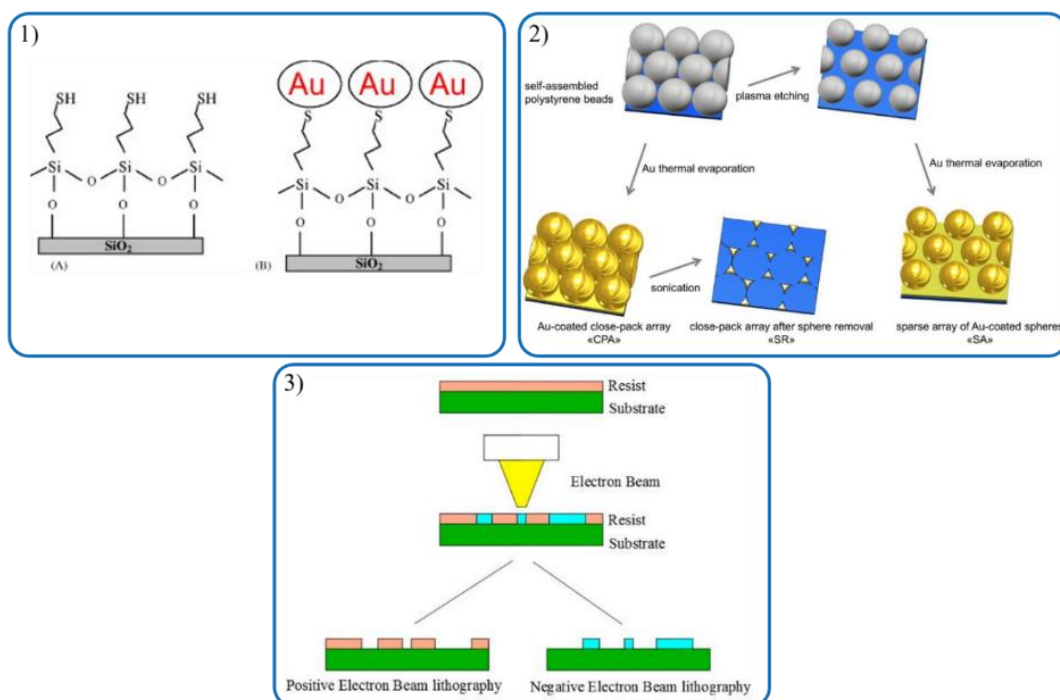


Figure 3. Illustration of various methods for fabricating SERS substrates: (1) covalent attachment of nanoparticles to a solid substrate, from Ref. [81], (2) nanosphere lithography, from Ref. [84], (3) electron beam lithography, from Ref. [93].

3.3 Optical fiber as SERS platform

Until now, either colloidal metallic nanoprobe in suspension or nanostructured planar surfaces have been extensively used as SERS substrates. In contrast to these conventional SERS systems, optical fibers offer possess several key advantages, such as their small size, low production costs, remote analyte detection, ease of handling, and multimodal properties and on-site sensing [94], [95]. Furthermore, their flexibility, adaptable shape and compatibility with portable technologies make them well-suited for incorporation into medical instruments such as catheters and guidewires used in minimally invasive procedures [96], [97].

A standard optical fiber is cylindrical in shape and consists of a core with a higher refractive index encased in a cladding with a lower refractive index that guides light through the core. An outer protective layer improves flexibility and helps to absorb mechanical stresses. The core and cladding are usually made from silica, with the refractive index contrast achieved by doping—commonly using germanium for the core or fluorine for the cladding [98]. The integration of metallic nanostructures on the fibers enhance light–matter interactions through plasmonic resonance and generate strong SERS signals [99]. Different optical fiber architectures can be tailored into advanced SERS-based platforms such as optrodes, fiber-integrated endoscopic devices and imaging systems [100], [101], [102], [103], [104]. Wang et al. developed nanoscale fiber tips through chemical etching, coated with silver nanoparticles and Raman reporters using laser-induced deposition [102]. Highly specialized hollow-core anti-resonant fibers (HcARFs) were fabricated by stacking and thermally drawing silica capillaries to form a microstructured fiber, whose inner surfaces was chemically functionalized as sensing platform. These innovations also support applications such as miniature fiber-optic Raman imaging devices for clinical endoscopy [103]. These configurations enable highly sensitive, real-time, and minimally invasive clinical diagnostics and intraoperative applications.

3.4 Non-metallic and hybrid nanostructures

Beyond metal-based substrates, semiconductor materials (e.g., TiO_2 , ZnO) and hybrid nanostructures have also shown promise in SERS applications [105], [106]. Semiconductors enhance signals via charge transfer interactions, offering benefits like chemical stability and biocompatibility [107]. Xu et al. pioneered the use of a pure TiO_2 SERS substrate, synthesized via the sol-gel method, for detecting CTCs, providing a novel strategy for non-noble metal semiconductor SERS bioprobes in liquid biopsies with a detection limit of 5×10^{-7} M [108]. Additionally, a biocompatible 3D ZnO quantum probe fabricated with multi-photon ionization mechanism allowed label-free, single-cell cancer detection, exhibiting SERS enhancement up to $\sim 10^6$ and enabling biomolecular sensing of DNA, RNA, proteins, and lipids on a nano-dendrite platform [109]. Additionally, combining various nanomaterials resulted in hybrid nanostructures, such as semiconductor nanostructures decorated with Au or Ag nanoparticles. In this regard, Cheng et al. fabricated ZnO nanorods decorated with silver nanoparticles on cellulose paper, as inexpensive SERS platform that amplified Raman

signals from serum samples, allowing fast, one-drop blood detection of liver cancer [110]. In another work, Lee et al. developed gold-coated ZnO nanorod substrate by hydrothermal method as biosensing chip capable of monitoring living cells using signal enhancement from gold clustering at the nanorods head [111]. Overall, non-metallic and hybrid nanostructures provide low-cost and highly sensitive SERS substrates, making them a promising platform for SERS diagnostic applications.

4. SERS strategies: direct and indirect

The detection of cancer biomarkers and pathological changes in cells and tissues can be performed using two types of SERS approaches: direct (label-free) and indirect (label-based).

Direct SERS strategies involve recording the spectral signature of a biofluid, cell or tissue after the interaction of the biological substance with the SERS substrate. The nature of the interaction between the metallic surface and the biological element plays a key role in the overall process. This strategy is particularly suitable for molecules with aromatic structures (thiols, N-heterocyclic aromatic compounds, etc.) and unsaturated bonds, as they exhibit chemisorption on metallic surfaces and strong SERS signal [112]. Generally, label-free SERS approaches use untargeted strategies to capture the biochemical signature of a sample without the use of biological recognition element. These methods do not focus on detection of specific biomarkers and avoid the need for functionalization strategies. In fact, the main contribution to SERS signaling is provided by molecules that have greater affinity for the SERS substrate. Therefore, this label-free method is especially effective for comprehensive cancer screening, early-detection and identification of potential biomarkers [52]. However, the use of bare nanoparticles is complicated by opsonization, a phenomenon that causes a multilayer of proteins to form around the nanoparticles in biological fluids. This process decreases signal intensity, reduces selectivity and impairs the circulation time of these particles [113]. Furthermore, biological samples, such as blood and tissue, consist of a variety of components that can interfere with direct SERS detection, making the identification of specific cancer biomarkers difficult. The presence of overlapping spectral signatures of various biomarkers exacerbates this problem and reduces the specificity of direct SERS applications in cancer diagnostics. Despite the ability of SERS to enhance Raman signals, the detection of low-abundance biomarkers within intricate biological matrices poses a considerable challenge, which may undermine the sensitivity required for early cancer detection [114]. Alternatively, indirect SERS approaches use targeted strategies that mainly involve SERS tags modified with a Raman reporter that has unique spectral properties. In particular, these tags are combined with a targeting agent (antibody, aptamer, etc.) to accurately identify specific biological targets. The interaction of the SERS tag with the molecule of interest is fundamental for its identification and localization as it leads to the production of a distinctive signal by the Raman reporter in SERS analysis. Therefore, in this approach, the SERS signal does not originate directly from the target molecule, but from the Raman reporter, which is indirectly linked to the presence of the target molecule. Furthermore, variations in the SERS spectrum as well as shifts in the position or changes in the intensity of the Raman peaks associated with the Raman reporter are directly related to the biological target under investigation. Specifically, targeted SERS allows for the detection

of low levels of cancer biomarkers, which is essential for early diagnosis and detailed tumor profiling. Furthermore, targeted SERS can be used for multiplexing, enabling the simultaneous detection of multiple biomarkers, which is critical for understanding tumor heterogeneity. Therefore, this approach is particularly suitable for SERS imaging to study cancer cells and tissues and to identify membrane and circulating cancer biomarkers, thereby supporting biomarker-specific imaging and diagnostics [6], [115]. Direct and indirect strategies and the key element of SERS tags were reported in **Figure 4**.

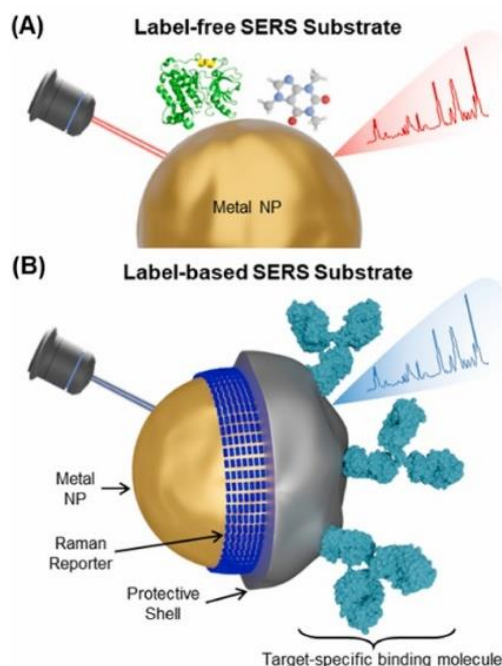


Figure 4. The schematic illustrates two major SERS detection strategies: direct (label-free) and indirect (label-based) approaches. A) Direct SERS detects target molecule interacting with a plasmonic metal surface, while B) Indirect SERS detects target molecules through a functionalized SERS tag attached to a plasmonic metal nanoparticle with a Raman reporter molecule. From Ref. [116].

Indeed, each of these strategies has its own advantages and disadvantages. In this context, label-free SERS technologies have several advantages over the label-based methods, such as their simplicity, faster results and the ability to detect a wide range of analytes. However, it is typically less sensitive and specific than the targeted approach [117]. Instead, indirect SERS offers the primary advantage that SERS tags do not depend on the cross-section of the target molecule or its affinity to the metal surface, which is one of the disadvantages of label-free methods. Furthermore, the label-based SERS approach is widely used in cancer diagnostics due to its multiplexing potential and versatility in SERS assay development [118]. Additionally, the use of bio-recognition elements improves both sensitivity and specificity, facilitating the detection of low levels of target biomolecules and reducing the background interference [13], [114].

4.1 SERS tags design

The SERS tag plays a crucial role in the indirect SERS strategy, which is why it is important to clarify the function of the individual main components: Raman reporter (RR), protective layer and the biorecognition element, apart from the SERS substrates as signal amplifiers, which are discussed in the section on SERS substrates.

Raman reporter molecules are essential components of SERS tags as their SERS response provides a characteristic SERS fingerprint for indirect identification of specific targets. Ideal Raman reporters have i) high Raman cross-sections able to produce strong SERS signals, ii) simple spectra to reduce the overlap of the peaks in multiplexing application, iii) resistance to photobleaching, and iv) the choice of the structure of reporter molecule is crucial for effective attachment to the surfaces of gold or silver, as the enhancement of SERS is significantly influenced by the proximity of the reporter molecule to the nanoparticles [119]. Furthermore, Raman reporters are categorized into non-resonant and resonant reporters based on their ability to undergo electronic transitions under visible light. Non-resonant reporters lack such transitions under visible light, resulting in weaker signals, while resonance only occurs under ultraviolet light [120]. Despite these considerations, these compounds offer several benefits: i) they are typically aromatic compounds characterized by an aromatic group responsible for their significant cross-section and a thiol group that has a high affinity for gold and silver, ii) they can form stable self-assembled monolayers, iii) they are not affected by fluorescence interference and iv) they provide a distinct Raman profile with some clear Raman peaks, making them ideal candidates for multiplexing [121]. Typical examples of non-resonant reporters are 4-aminothiophenol (4-ATP), 4-mercaptobenzoic acid (4-MBA), 5,5-dithiobis (2-nitrobenzoic acid) (DNTB), etc. [122]. Another important aspect for a Raman reporter is to minimize the background noise and improve the clarity of SERS signal, which is indirectly related to the biological target. Recently, new Raman reporters, namely silent-region SERS reporters with functional groups with triple bonds, such as alkynes and nitriles, or carbon–deuterium bonds, exhibited a distinct vibrational frequency ($2000\text{--}2200\text{ cm}^{-1}$) in the silent region ($1800\text{--}2800\text{ cm}^{-1}$) of biological samples [121]. The presence of Raman peaks in the biological silent region helps to minimize background interference and enhances the clarity of signals detected by the nanotags. In particular, 5-ethynyl-2-deoxyuridine (EdU), 4-ethynylbenzenethiol, phenylthiocyanate, 4-mercaptobenzonitrile are examples of representative silent Raman reporter molecules, among several others [123]. The image of the different types of Raman reporters is shown in **Figure 5**. Additionally, resonant Raman reporters are organic dyes, often fluorescent, that undergo electronic transitions in the visible to near-infrared spectrum, creating a resonance effect. This phenomenon is called Surface-Enhanced Resonance Raman Scattering (SERRS) [124]. Resonant reporters exhibit a high SERS signal and sensitivity, which is increased by 10^3 to 10^4 times [125], [126]. SERRS typically involves several resonant molecules such as fluorescein, rhodamine 6G (R6G), crystal Violet (CV), malachite Green (MG) and various cyanines, such as 3,3'-diethylthiatricarbocyanine (DTTC) and indocyanine green (IG) [120]. The application of resonant molecules is associated with several disadvantages, including i) fluorescence background associated with fluorescent dye, which is not always completely quenched near a metallic surface, ii) the generation of more intricate SERS spectra due to a higher number of vibrational modes, which limits multiplexing capabilities, iii) the prolonged laser irradiation can lead to photodegradation, reducing the sensitivity of SERRS nanotags [125], [127]. However, SERRS tags with NIR Raman reporters enable highly sensitive *in vivo* detection. Nonetheless, their weak interaction with plasmonic nanoparticles limits their stability.

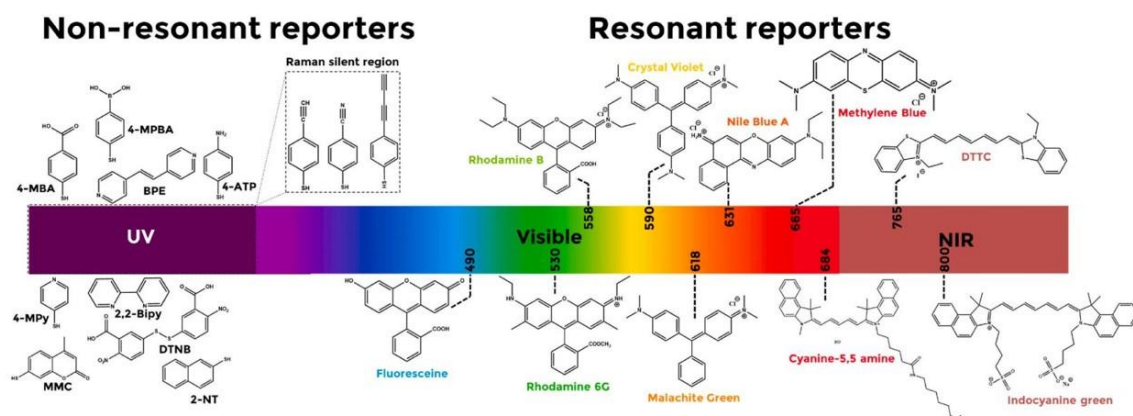


Figure 5. Illustration of widely used Raman reporters, from Ref. [120].

Encapsulating the dyes in a shell or cavity enhances stability and sensitivity, making them more effective for SERS applications [16].

Additionally, in the context of cancer diagnostics and biomedical applications, a coating layer is typically required to prevent i) aggregation, ii) interaction with aspecific molecule present in a complex matrix, iii) degradation that may occur in the biochemical environment, iv) biotoxicity and v) dissociation of the Raman reporter from the metallic surface, thus increasing the chemical and physical stability of the SERS tags [114], [128]. In certain approaches, a coating is also applied that acts as a reliable substrate for the binding of a targeting agent [129]. However, agents of different nature can form coating layers such as biomolecules, polymers or silicon dioxide (SiO_2) [113]. Among biological coatings, bovine serum albumin (BSA), a small protein, is commonly used due to its biocompatibility and stability [130]. The use of polymer coatings, such as the biocompatible polymer polyethylene glycol (PEG), enhances stability and enables tailored functionalization. This approach allows for the covalent bonding of the polymer to the SERS substrate as well as exposure of chemical groups that can interact strongly with biological targets. Furthermore, a silica shell based on biocompatible SiO_2 provides a stable and mechanically durable coating that protects against aggregation [131]. An important aspect of the bioengineering of SERS tags is the integration of biorecognition elements, which enables the specific binding of cancer biomarkers in biofluids or at cell and tissue level, reducing non-specific interactions with other molecules. The process of targeting cells and tissues involves identifying and attaching to specific proteins that are either unique to the target cells or present in larger quantities on them [131]. The SERS tags can be functionalized with different categories of targeting agents, such as small molecules, peptides, proteins, aptamers, and molecularly imprinted polymers (MIPs), as shown in **Figure 6**. Folic acid, which is classified as a vitamin, and urea, is small molecules commonly used to label the folic acid receptor and prostate-specific membrane antigen (PSMA) for the cancer cell detection [132], [133], [134]. Alternatively, small peptides that consist of amino acids sequences such as arginine-glycine-aspartic (RGD) or arginine-glycine-aspartic acid-phenylalanine-cysteine (RGDFC) can be used to target integrin receptors that are notably overexpressed in cancer cells. [135], [136]. In this context bioconjugation of SERS tags with the epidermal growth factor (EGF) peptide is implemented to facilitate the detection of the epidermal growth factor receptor (EGFR), prominently expressed on the membrane surface of circulating tumor cells [124]. Transferrin, a protein targeting agent, has been identified as

a crucial component of the transferrin (TF)-PEG-BGERTs SERS nanoprobe that enables real-time visualization of the transferrin receptor (TfR) on the membranes of cancer cells [137].

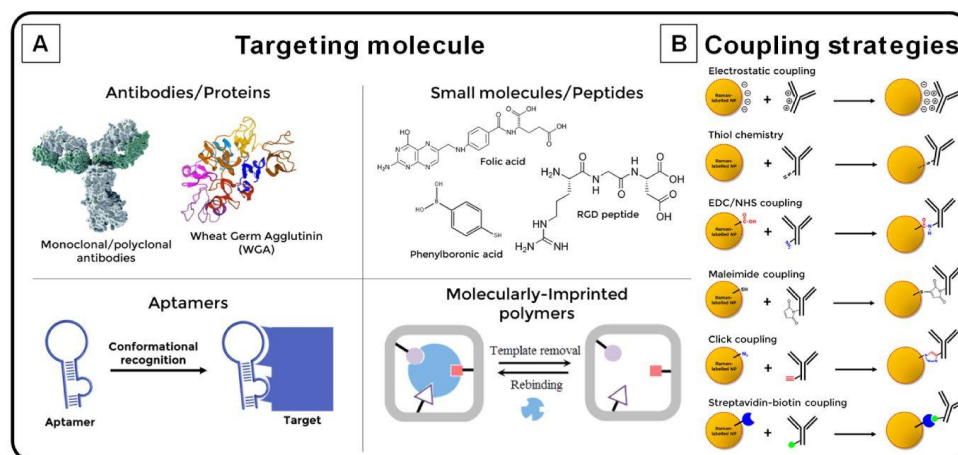


Figure 6. Overview of A) the different types of targeting agents and of B) the different coupling strategies used to conjugate the targeting agents to the surfaces of the nanotags. From Ref. [115].

Antibodies are the first choice for targeting proteins in SERS nanotags due to their exceptional specificity and strong affinity for antigens. A SERS-immunoassay involves antibodies as selective recognition agents to identify target analytes for quantitative analysis in a sample. The amplification effects of metal surfaces are combined with the high specificity of the interactions between antibodies and antigens [138]. Therefore, the integration of antibodies into SERS tags enhances detection sensitivity, allowing for the identification of low concentrations of cancer biomarkers and significantly improving SERS-based cancer diagnostics [139]. Additionally, SERS-based immunoassays are classified into competitive and non-competitive types (also referred to as sandwich immunoassay). In non-competitive assays, the target is captured between immobilized capture antibodies on SERS substrate surface and detection antibodies on nanoparticles functionalized with Raman reporter, resulting in a signal intensity that is directly proportional with the target concentration [129]. This method is typically used to detect macromolecules such as proteins that have more than one epitope [140]. In contrast, competitive immunoassays, labeled antigens acts as probes that compete to the target for antibody binding, leading to a signal intensity that is inversely proportional with the target concentration. This method is particularly advantageous for the detection of small molecules such as peptides, ions and hormones that contain a single binding site. Compared to competitive immunoassays, sandwich immunoassays generally have higher sensitivity, a wider dynamic range and enhanced reproducibility [129]. In order to achieve effective and precise binding to their target molecules, it is also crucial that the antibodies are tightly bound to the plasmonic nanostructures. Conjugation of antibodies to SERS probes can be achieved by covalent or non-covalent strategies. In particular, covalent binding between antibodies and nanoparticle surfaces can be achieved by various methods, including (i) chemisorption using thiol derivatives, (ii) 1-ethyl-3-(3-dimethylaminopropyl)carbodiimide/N-hydroxysuccinimide (EDC/NHS) chemistry, (iii) maleimide coupling, (iv) click chemistry, (v) streptavidin-biotin coupling, (vi) silane chemistry on silica-coated tags [131], [141]. In

particular, antibodies do not have thiol groups, which necessitates the use of anchoring and cross-linking elements for conjugation. Anchoring agents usually have a thiol group that binds to the surface of metallic nanoparticles and a functional group that reacts with antibodies using EDC/NHS chemistry or interacts with crosslinkers to facilitate antibody attachment. Crosslinking can act either as anchoring agents or as bifunctional molecules, with one end binding to the metal/shell structure and the other end attaching to a target-specific ligand. Together, these elements ensure a stable and efficient covalent conjugation of antibodies to SERS probes [142], [143]. Therefore, covalent binding provides higher stability and reliability, requiring more complex chemical methodologies to maintain antibody conformation and orientation [144]. In addition, the non-covalent binding process is characterized by its simplicity and low cost, as the adsorption of antibodies to SERS probes occurs by hydrophobic and electrostatic forces without the need for linkers. Despite its advantages, non-covalent binding has several limitations: (i) it is weak in nature, (ii) it requires high antibody concentrations to form antibody–gold nanoparticle conjugates, (iii) the orientation of antibodies on the nanoparticle surface is unpredictable, (iv) it promotes nanoparticle aggregation, and (v) the attachment is influenced by pH changes, making it susceptible to antibody displacement by other molecules in biological samples [142], [145], [141]. Although the use of antibodies as biological detection elements is widespread, it faces several challenges. These include the high cost, particularly for monoclonal antibodies, sensitivity in handling, problems associated with epitope presentation and the potential for cross-reactivity when multiple targets are present [113], [146]. Aptamers are short single-stranded oligonucleotide sequences of DNA or RNA that have a high affinity for specific targets, making them a useful targeting agent. The Systematic Evolution of ligands by exponential enrichment (SELEX) technique enables the development of aptamers that have high binding affinity to various targets such as proteins, cells or even small molecules [147]. Aptamers interact with their target molecule via mechanisms such as hydrogen bonding, van der Waals interactions, and aromatic stacking [148]. Their advantages include stability at room temperature, ease of chemical modification and a smaller size compared to antibodies [149]. Aptamers face several challenges, mainly due to the limited number of targets for which they are designed. Furthermore, these molecules are usually identified in relation to soluble proteins, which are morphologically different from membrane-bound proteins. Such differences can result in significant alterations in the presentation of epitopes [113]. A broad selection of aptamers have been developed for specific targeting of numerous biomarkers of interest [150], [151].

Molecularly imprinted polymers represent a class of synthetic materials designed to mimic the selective binding properties of natural biological recognition elements, such as antibodies. They are used extensively due to their high affinity, selectivity, stability, strong mechanical properties and resistance to extreme temperature, pressure, and pH conditions. The formation of MIPs on the surface of nanoparticles involves the polymerization of functional and cross-linking monomers in a porogenic solvent, in the presence of target molecule, which is the template. After removal of the molecular template, the polymer retains cavities or binding sites that may the template's size, shape and functionality of the template, allowing selective binding of the target molecule. Compared to monoclonal antibodies, MIPs are not only easier and less expensive to synthesize but also do not require

preclinical animal involvement [152]. Despite the various advantages associated with MIPs, there are notable limitations, such as the time-consuming nature of their preparation and challenges in accessing the binding sites [153]. However, numerous studies have highlighted the significance of MIPs as targeting agents in the field of cancer diagnostics [154], [155].

5. SERS Liquid biopsy and Cancer Bioimaging

5.1 SERS liquid biopsy

Liquid biopsy is a minimally invasive technique that uses various body fluids (blood, cerebrospinal fluid, urine, saliva, etc.) to identify molecular changes associated with circulating tumor-derived biomarkers [156]. Although tissue biopsy is considered the gold standard for tumor diagnosis, this method is invasive, presents difficulties in sample collection, requires a considerable amount of time and is not well-suited for the continuous monitoring of the disease, particularly in the early detection phase. In this regard, liquid biopsy represents a simpler, less invasive and risky alternative to tissue biopsy, emerging as a promising diagnostic tool for different diagnostic applications such as early detection, prediction of treatment efficacy, evaluation of tumor recurrence and prognosis, as reported in **Figure 7** [157], [158]. The main circulating tumor biomarkers detected through liquid biopsy include circulating tumor DNA (ctDNA), microRNA (miRNAs), extracellular vesicles (EVs), protein and circulating tumor cells (CTCs) [156]. One of the key challenges in utilizing circulating biomarkers lies in their very low concentrations within biofluids. In recent years, the combination of molecular amplification techniques with SERS has proven highly effective in boosting Raman signal intensity by amplifying and increasing the detectable amount of these scarce targets, thereby enhancing overall detection sensitivity. Specifically, hybridization chain reaction (HCR) and catalytic hairpin assembly (CHA) are enzyme-free DNA-based signal amplification techniques. In HCR system, the initiator (I) binds to two or more hairpins DNA structures (commonly H1 and H2), causing them to open and hybridize with each other, resulting in an extended, nicked double-stranded DNA structure, effectively amplifying the DNA signal. This process continues automatically until all hairpins are consumed. In CHA, the reaction is more controlled: each amplification cycle requires the continuous presence of the initiator, which allows the reaction to self-regulate. Both methods can be integrated SERS by functionalizing the hairpins with Raman reporter molecules or metal nanoparticles, significantly boosting the detectable signal [159]. These enzyme-free strategies are now widely employed to enhance the sensitivity of SERS-based biosensing platforms.

Currently, alongside the application of bioengineered SERS substrates for the identification of circulating tumor biomarkers, a variety of strategies have been developed to combine SERS with other technologies to achieve higher specificity, sensitivity and reproducibility compared to conventional SERS substrates. Specifically, integrating SERS with magnetic beads, microfluidic systems, and paper-based detection assays enables the optimization of SERS performance for liquid biopsy applications.

In magnetic-bead SERS assays, magnetic beads—typically consisting of a Fe_3O_4 core—are combined with metallic nanoparticles to capture and isolate cancer biomarkers from complex biological fluids [130], [160], [161], [162], [163]. Two main configurations are commonly

employed. In the core-shell nanostructure, a bioengineered metallic shell surrounding the magnetic core amplifies Raman signals for enhanced detection sensitivity, while the magnetic core allows efficient isolation and separation of target molecules from biofluids [34]. In an alternative approach, the target molecule is initially captured by ligand-functionalized magnetic beads, after which SERS-engineered nanotags bind to the biomarker, forming a sandwich complex that can be magnetically separated and analyzed via SERS [139].

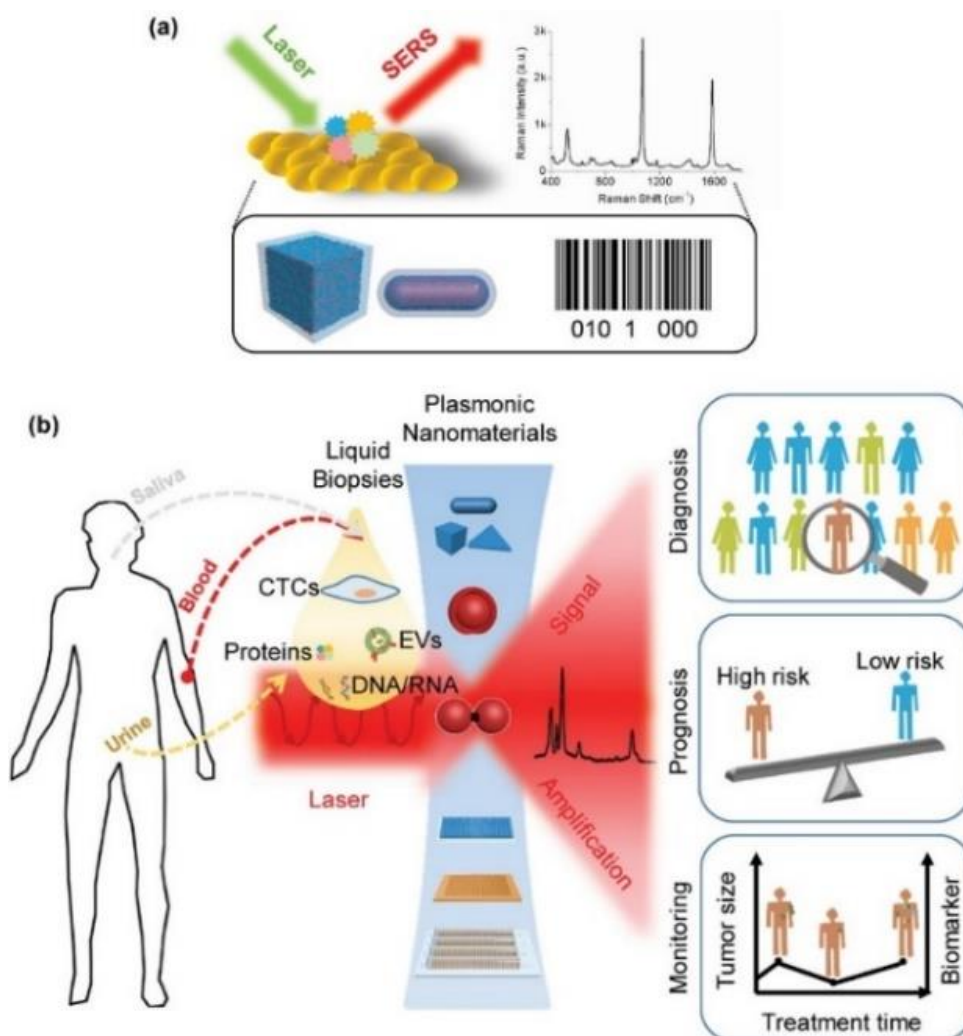


Figure 7. Schematic representation of SERS liquid biopsy. a) SERS substrates enable molecular fingerprinting for direct and indirect SERS-based applications, including b) the detection of different types of circulating cancer biomarkers (DNA/RNA, EVs, proteins and CTCs) in liquid biopsy for detailed tumor profiling, from Ref. [165]. EVs: exosome vesicles, CTCs: circulating tumor cells.

Microfluidic chips are microscale devices with precisely fabricated micro-channels designed to separate and analyze biomolecules according to their physical or biochemical properties, often employing antibody-based capture [34]. They provide accurate fluid control, use minimal amounts of samples and reagents, and are both compact and cost-efficient. SERS can be readily integrated into these devices through functionalized SERS substrates or SERS nanotags [139], [164]. Combining SERS with microfluidics enables highly effective liquid biopsy platforms, allowing high-throughput, real-time analysis and multiplexed detection of cancer biomarkers in small sample volumes.

Another integrated system, involve the combination of SERS with paper-based assay, a preferred platform for designing point of care tests due to their cost-effectiveness, biocompatibility,

adaptability, and portability. Later flow assays (LFAs) are the most common paper-based diagnostic platform capable of detecting and quantifying analytes in complex samples within 5–30 minutes. Their ease of use and quick results have made them widely adopted in clinical settings for identifying specific antigens and antibodies [166]. Generally, in LFAs, capillary action drives a liquid sample across a nitrocellulose membrane, where detection and capture antibodies linked to gold nanoparticles, form complexes with target analytes at the test line, producing a visible signal, while control lines verify test accuracy. Incorporating SERS nanotags into paper-based assays greatly enhances sensitivity, allowing precise quantification of analytes through the SERS signal generated at the illuminated test line—an achievement not possible with semi-quantitative conventional colorimetric methods typically used in conventional LFA assays [139].

This section examines recent advances in the development of SERS-based and integrated platforms, to design increasingly efficient and ultrasensitive methods for detecting circulating biomarkers—including nucleic acids, EVs, proteins, and CTCs—thus contributing to significant progress in the field of liquid biopsy.

5.1.1 Detection of nucleic acids

Circulating free nucleic acids (cfNAs) are emerging as powerful diagnostic biomarkers, enabling real-time monitoring of tumor status. Alterations in cell-free DNA (cfDNAs) and regulatory RNAs, including microRNA (miRNAs) and circular RNAs (circRNAs), offer key information on tumor progression and underlying molecular pathways [167].

Circulating cell-free DNA (cfDNA) is released from both normal and malignant cells in the bloodstream. The part of cfDNA that originates exclusively from tumor cells is called circulating cell-free tumor DNA (ctDNA) and consists of fragments of single or double DNA strands. ctDNA can originate from primary tumors, CTCs or metastases and is characterized by tumor-specific genetic alterations (point mutations, changes in methylation pattern, etc.), revealing the tumor's molecular profile [168], [169]. In their work, Wu *et al.* designed a lab-on-a-chip system for the detection of the epidermal growth factor receptors (EGFR) E746-A750 mutation in ctDNA from lung cancer patients [170]. The microfluidic device integrates two modules: a filter moat that removes blood cells from whole blood, and a mutation sensor module containing a SERS substrate made of Au@Ag nanorods functionalized with Raman reporters (DTNB) and hairpin DNA probes specific to the EGFR mutation. The system used a “signal-off” sensing strategy, where binding of the target DNA opens the hairpin probe, reducing the SERS signal intensity. It achieved high sensitivity (down to 100 fM) and completed detection in 35 minutes. Clinical validation showed lower SERS signals in patient samples with the mutation compared to controls, confirming both sensitivity and specificity and highlighting the platform's promise for point-of-care lung cancer diagnostics. Although ctDNA holds significant potential for cancer management, its clinical implementation is limited by technical variability and the low concentration of ctDNA in patient samples. To fully realize its potential, standardized testing protocols, and large-scale clinical studies are needed to confirm its reliability across different cancer types [171].

Unlike ctDNA mutation analysis, cell-free RNA (cfRNA) profiles represent early and dynamic changes in carcinogenesis. Non-coding RNAs, especially microRNAs (miRNAs), have been widely studied as promising non-invasive biomarkers for cancer detection [172]. MicroRNAs are 20–25 nucleotides in length, that control gene expression at the post-transcriptional level. Their dysregulation is common in cancer, often leading to reduced suppression of oncogenes, which contribute to cancer development. Furthermore, miRNA profiling has greatly improved the molecular classification of tumors, allowing differentiation based on the specific cancer type and associated prognosis [173], [174]. miR-21, a known oncomiR, is up-regulated in various cancers and contributes to tumor development, highlighting its potential as a molecular biomarker [175]. Recently, Tan *et al.* implemented a dual-signal SERS biosensor for the sensitive and accurate detection of miR-21 in breast cancer [176]. The SERS nanostructures consisted of spindle-shaped array of gold nanoparticles (SAu), fabricated by self-assembly at a liquid-liquid interface (water/hexane), and subsequent transfer to a silicon wafer. The SERS array was bioengineered using ROX-labeled hairpin DNA probe for miR-21 detection, along with DTNB as a secondary Raman reporter to serve as an internal reference for signal stability and calibration. In the presence of miR-21, the target bound the DNA probe, opening the hairpin and reducing the ROX SERS signal, while the DTNB signal remains unchanged, contributing to accurate detection. The biosensor exhibited a wide detection range (0.1–105 pM), a low detection limit of 0.046 pM, and excellent specificity for miR-21. It showed reliable reproducibility and has been validated on human serum samples, offering potential for clinical cancer diagnostics. However, single miRNA tests often provide limited diagnostic precision, whereas panels of multiple miRNAs have demonstrated greater accuracy and improved diagnostic performance [177]. In their work, Wu *et al.* developed a 3D biosensor based on hierarchical array for the simultaneous detection of miRNA associated with colorectal cancer (CRC)-related miRNAs, miR-21 and miR-31 [161]. The system used Ag-coated magnetic nanoparticles (AgMNPs) to enrich target miRNAs, each containing a Raman reporter as an internal standard and DNA capture complementary for half of the sequence to miRNA targets. Complementary SERS probes consisted in silver-gold nanocages AuNC@AuNPs labeled with distinct Raman reporters (NTP and 2-NT) bound the other halves of the miRNA sequences, forming a sandwich hybridization complex as shown in **Figure 8.1**.

The subsequent addition of signal amplification probes (SA probes) linked to AuNC@AuNPs (without raman reporter) triggered the self-assembly of 3D AuNC@AuNP clusters on AgMNPs, resulting in strongly enhanced SERS signals and improved detection sensitivity. The biosensor exhibited remarkable sensitivity, with detection limits of 3.46 aM for miR-21 and 6.49 aM for miR-31, along with high specificity and reproducibility. When evaluated using clinical serum samples, it effectively distinguished colorectal cancer patients from healthy individuals, even in AFP-negative cases, and successfully identified different tumor stages, highlighting its promise as a powerful tool for early CRC diagnosis and staging.

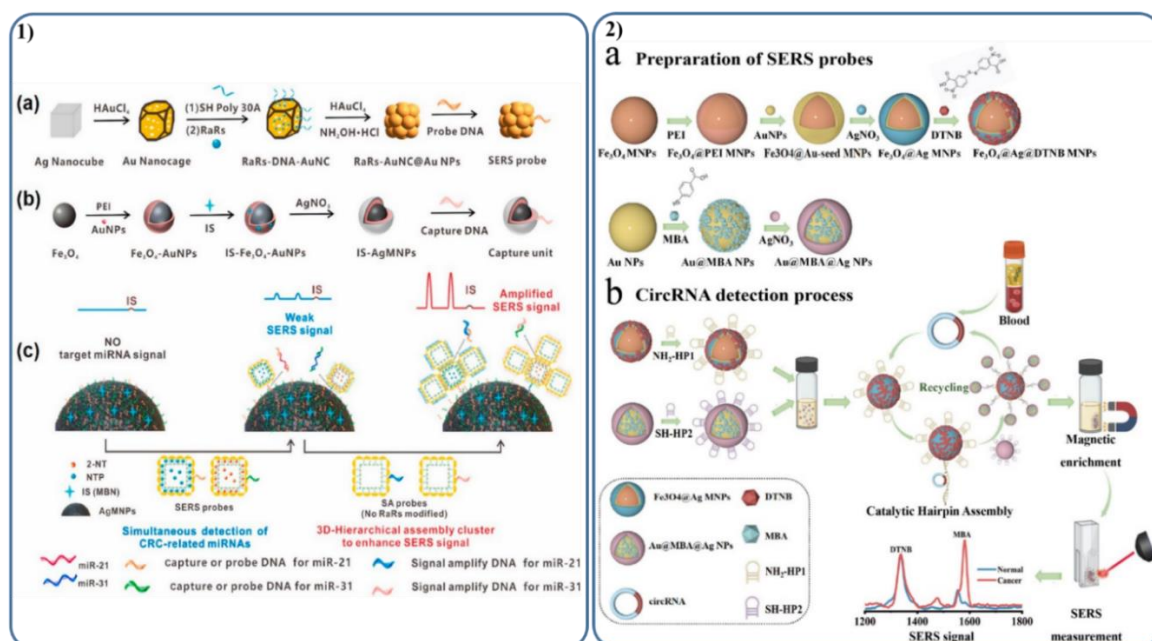


Figure 8. Schematic representation of the different SERS detection strategies for nucleic acid biomarkers. (1) Hierarchical 3D interface arrangement of Au nanocage@Au together with IS-AgMNP for the detection of miR-21 and miR-31. Illustration of the workflow for (a) the SERS probe and (b) the development of the acquisition unit. (c) The workflow for the detection of two CRC-associated miRNAs, from Ref. [170]. (2) SERS platform for circRNA detection in lung cancer. (a) Preparation of Fe₃O₄@Ag@DTNB magnetic nanoparticles and of (b) Au@MBA@Ag nanoparticles, (c) SERS detection of circRNA, from Ref. [160].

Among cell-free RNA, circular RNAs (circRNAs) have recently attracted considerable interest as biomarkers, given their widespread presence in tissues, cells, and bodily fluids, as well as their specific expression depending on tissue type and developmental stage. Numerous circRNAs have been discovered to function as oncogenes or tumor suppressors, and their abnormal expression in cancers highlights their involvement in tumor development and malignant cellular processes [178]. In this regard, Xu *et al.* developed a highly sensitive SERS biosensor for detecting circVAPA, a circular RNA associated with lung cancer, by integrating catalytic hairpin assembly (CHA) for signal amplification with magnetic nanoparticle-based enrichment [160]. The system employed Fe₃O₄@Ag@DTNB-HP1 magnetic nanoparticles and Au@MBA@Ag-HP2 core-shell SERS nanoprobe, each functionalized with DNA hairpin probes (HP1 and HP2, respectively) designed to recognize the unique back-splicing site of circVAPA. In the CHA amplification, circVAPA bound HP1, opening its hairpin structure due to higher binding affinity. Subsequently, HP2, having a stronger complementarity to HP1 than circRNA, displaced the bound circVAPA, releasing it for additional reaction cycles. This repeated assembly led to the formation of a stable HP1-HP2 complex, which brings the magnetic and SERS nanoparticles in close proximity, creating a core-satellite nanostructure (**Figure 8.2**), enhancing the Raman signal of MBA while keeping the DTNB internal standard stable. The platform achieved 5.5 attomolar-level detection limit, high specificity against similar RNAs, and successfully distinguished tumor subtypes and pre- versus post-operative states in clinical samples. Combined with conventional biomarkers, it achieved an AUC of 0.859, demonstrating strong potential as a non-invasive tool for early lung cancer diagnosis and monitoring.

5.1.2 Detection of tumor-exosome vesicles

Exosomes are small extracellular vesicles (30–150 nm) secreted by various cell types secreted and play a crucial role in the communication between cells [179]. Tumor-derived extracellular vesicles contain a variety of molecular components such as proteins, RNA and DNA, reflecting the molecular state of their cancerous origin, making them valuable for assessing tumor progression, treatment response, and prognosis [180]. Since cancerous exosomes represent only a small fraction of the exosomes present in body fluids, the development of highly sensitive and specific detection methods is a fundamental prerequisite for the further development of cancer diagnostics based on the identification of exosomes [181]. To this end, various SERS strategies have been used to accurately detect tumor-derived exosomes with high sensitivity and specificity.

Recent research indicates that tumor-derived exosomes (TEXs) carry immune checkpoint proteins such as programmed death-ligand 1 (PD-L1), which suppress T cell activity and promote tumor invasion and metastasis [182]. However, exosomal PD-L1 can also be produced by non-tumor cells, leading to inaccurate cancer diagnosis. In this regard, Su *et al.* developed a SERS-based immunoassay combined with immunomagnetic separation to detect PD-L1-tumor derived exosomes (TEXs^{PD-L1}) and differentiate them from non-tumor exosomes in breast cancer [162]. The strategy employed EpCAM aptamer-functionalized magnetic nanoparticles to selectively capture TEXs^{PD-L1} expressing the tumor-associated antigen EpCAM, absent in non-tumor EVs. Detection was subsequently achieved using SERS nanoprobe—gold nanoparticles conjugated with PD-L1 aptamers and the Raman reporter 4-NTP—enabling specific identification of the captured TEXs^{PD-L1}. This sandwich complex enabled highly sensitive detection, translating TEXs^{PD-L1} presence into a SERS signal of 4-NTP at 1344 cm⁻¹, reaching a limit of detection (LOD) of 347 particles/μL. Testing on plasma samples from breast cancer patients and healthy controls demonstrated a strong correlation between Raman intensity and TEXs^{PD-L1} levels, achieving an AUC of 0.988. Therefore, the development of a highly sensitive and multiplexed sEV analysis method is essential to enhance detection efficiency, reflect the complexity of tumor-derived signals, and support earlier, more accurate cancer diagnosis. In this regard, Wang *et al.* developed a PMMA nanocavity array SERS chip (PNAS chip) for the multiplexed detection of different breast cancer tumor-derived exosomes from whole blood [183]. The PNAS chip enabled single-exosome entry and CD63 aptamer-based capture, providing both size-selective recognition and isolation. Following capture, three distinct SERS probes—silver nanoparticles functionalized with different Raman reporters (4-MBA, DTNB, MMC) and DNA aptamers targeting the marker proteins PD-L1, HER2, and EpCAM—were introduced for detection, as described in **Figure 9.1**. The SERS probes bound to exosomal surface proteins in proportion to their expression levels, reflecting differences in HER2, EpCAM, and PD-L1 profiles across exosome types. Integration of the PNAS chip with machine learning enabled precise profiling of tumor-derived exosomes from whole blood samples of various breast cancer subtypes and healthy individuals, achieving accuracy rates of up to 95%. In another work, Chen *et al.* also introduced a multiplexing strategy to simultaneously target different tumor-derived exosomes biomarkers [184]. They developed a SERS-multichannel microfluidic extracellular vesicle assay (S-MMEV) to capture and analyze EVs from ovarian cancer cell cultures or patient serum directly on-chip. The system employed a

multichannel gold-coated microfluidic chip functionalized with anti-CD63 antibodies to capture EVs, while detection was achieved using Au@Ag core-shell nanoparticles labeled with MGITC and antibodies targeting six EV biomarkers: two general EV markers (CD9, CD81) and four ovarian cancer-associated markers (EpCAM, EGFR, CD24, CA125). After binding to EV biomarkers according to their expression levels, the SERS probes revealed differences in protein profiles among various exosomes. Validated with EVs-derived ovarian cancer cell lines and clinically tested on patient samples, the platform detected as few as 10 particles/mL. Notably, the comparison between healthy controls and ovarian cancer patients (stages I-II) showed an AUC of 0.9467 achieved by combining six-marker panel.

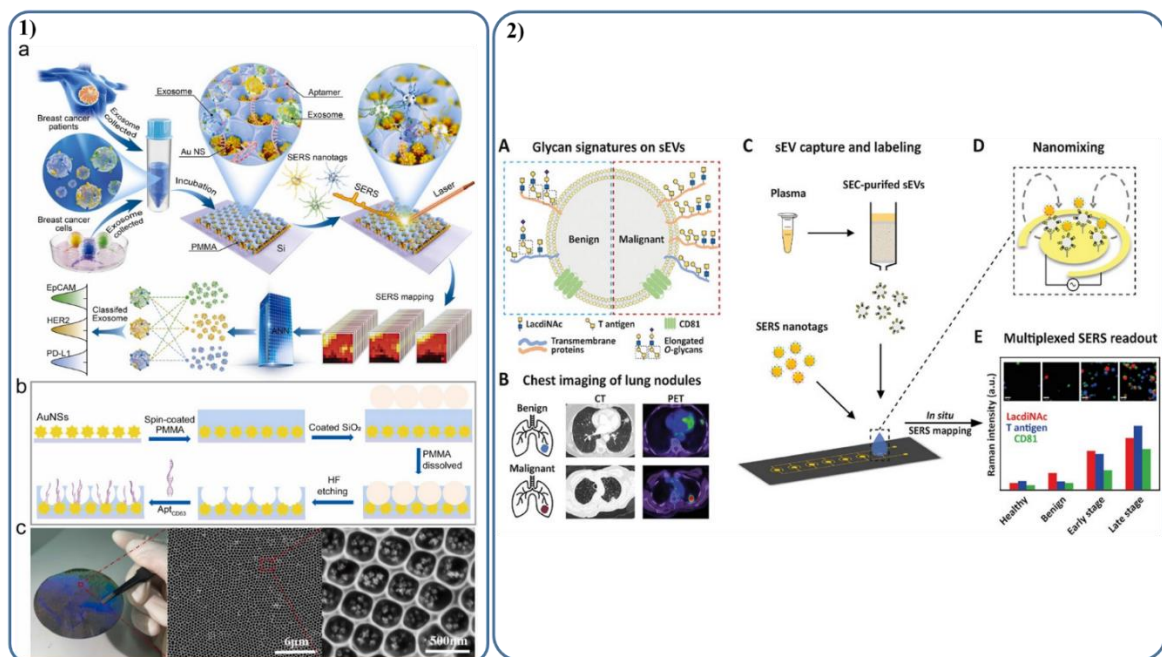


Figure 9. SERS strategies for the detection of cancer-related EVs. (1) PMMA nanocavities array SERS chip (PNAS-chip) for EVs detection in breast cancer. a) The workflow and b-c) the development and characterization of the PNAS-chip, from Ref. [183]; (2) SERS Microfluidic Biosensor for glycan profiling of EVs. a) Glycan profiles on sEVs derived from benign and lung cancer patients, (b) CT and PET images of benign and malignant lung nodules, (c) operating principle of EV-GLYPH assay, (d) alternating current electrohydrodynamic field induces nanomixing fluid flow on the electrode surface, (e) SERS maps of the different signals of three nanotags associated with the expression of LacdiNAc, T antigen and CD81, from Ref. [185].

In addition to proteins, EV membranes also contain glycans—carbohydrate chains attached to proteins or lipids—that are crucial for cell signaling and often altered in cancer, making them useful markers for EV isolation and tumor profiling [186]. LacdiNAc and T antigen are cancer-associated glycans, with LacdiNAc implicated in tumor progression and T antigen commonly overexpressed in cancer cells. Recognizing their relevance, Zhou *et al.* developed a microfluidic SERS-based biosensor for glycan profiling of extracellular vesicles (EVs) in non-small cell lung cancer (NSCLC) [185]. The system integrated microfluidics with SERS tags to capture and analyze EV glycan patterns (see **Figure 9.2**). Tumor-associated EVs were first captured using MUC1 antibody-functionalized electrodes, while gold nanoparticles conjugated with specific lectins (WFA and PNA) and an anti-CD81 antibody—each labeled with distinct Raman reporters—enabled selective detection of LacdiNAc, T antigen, and the sEV marker CD81, respectively. SERS signal intensities of the three different SERS probes reflected sEV quantity and biomarker expression, allowing

detailed glycosylation profiling in NSCLC. The assay achieved a LOD of 1×10^6 sEVs/mL and successfully distinguished NSCLC-derived sEVs from benign and healthy controls based on glycan signatures. Clinical plasma tests confirmed diagnostic utility, with strong ROC performance for T antigen (AUC = 0.87), LacdiNAc (AUC = 0.68), highlighting its potential for NSCLC diagnosis.

5.1.3 Detection of protein cancer biomarkers

The identification of protein tumor markers in body fluids is a key step in the detection and monitoring of cancer by liquid biopsy. These proteins provide crucial insights into the molecular characteristics, classification and progression of cancer. Additionally, changes in protein levels can indicate genomic mutations, disease stages or response to treatment [187]. Actually, various SERS strategies have been developed to improve the sensitivity, specificity and multiplexing capabilities for the detection of single and multiple protein cancer-related biomarkers. In particular, SERS-based platforms for detecting protein cancer biomarkers commonly utilize SERS immunoassays, where immuno-functionalized SERS tags specifically bind to the target biomarker. The resulting Raman signal—both its intensity and spectral features—allows for precise identification and quantification of the biomarker [16]. For example, Chen et al. developed a highly sensitive and selective plasmonic dual-aptamer immunosandwich assay (PISA) to detect soluble fibroblast activation protein (FAP), a key pan-cancer biomarker involved in extracellular tumor remodeling [188]. The assay used a gold-coated substrate functionalized with a glycan-targeting aptamer (Seq2T) to capture FAP, along with silver-based SERS nanotags labeled with the Raman reporter PATP and a second aptamer (A1) that binds a different site on FAP. The formation of the immunosandwich upon FAP binding allowed quantification by monitoring the SERS signal of PATP. The method demonstrated exceptional sensitivity, detecting FAP at concentrations as low as 10 pg/mL. In serum tests, it successfully distinguished gallbladder cancer patients from healthy controls, providing results consistent with standard ELISA but in a faster and more cost-effective manner.

Recently, the performance of sandwich-type immunoassays for detecting and quantifying circulating cancer-related protein biomarkers has been enhanced by integrating them with paper-based platforms and microfluidic technology, enabling improved portability, efficiency, and analytical precision. In terms of integrating paper-based approaches with SERS, Su *et al.* introduced a combined SERS and lateral flow immunoassay platform for the simultaneous detection and quantification of two serum biomarkers, CCL20 and EGF, which were closely associated with the therapeutic outcomes of microwave ablation (MWA) in hepatocellular carcinoma (HCC) patients [189]. The device featured a nitrocellulose membrane with two test lines—TL1 for CCL20 and TL2 for EGF—where immobilized monoclonal antibodies specifically captured their respective targets, along with a control line for assay validation. The conjugate pad contained gold nanostar (AuSt) probes functionalized with the Raman reporter *p*-NTP and antibodies against CCL20 and EGF. Upon sample addition, target biomarkers bound to their SERS probes and were subsequently captured at the corresponding test lines, forming sandwich immunocomplexes, as shown in **Figure 10.1**. The SERS signals obtained from TL1 and TL2 provide a highly sensitive and multiplexed readout, enabling the simultaneous quantification of both CCL20 and EGF

biomarkers in a single assay, achieving a LOD of 0.064 ng/mL for CCL20 and 0.090 ng/mL for EGF. Clinical tests from serum samples from healthy individuals and HCC patients undergoing complete or partial MWA, showed that complete MWA increases CCL20 levels while decreasing EGF levels, in contrast to incomplete MWA, which causes a slight decrease in CCL20 and a significant increase in EGF. These results, confirmed by ELISA data, highlight the exceptional sensitivity, rapid detection and portability of the biosensor in effectively monitoring MWA results.

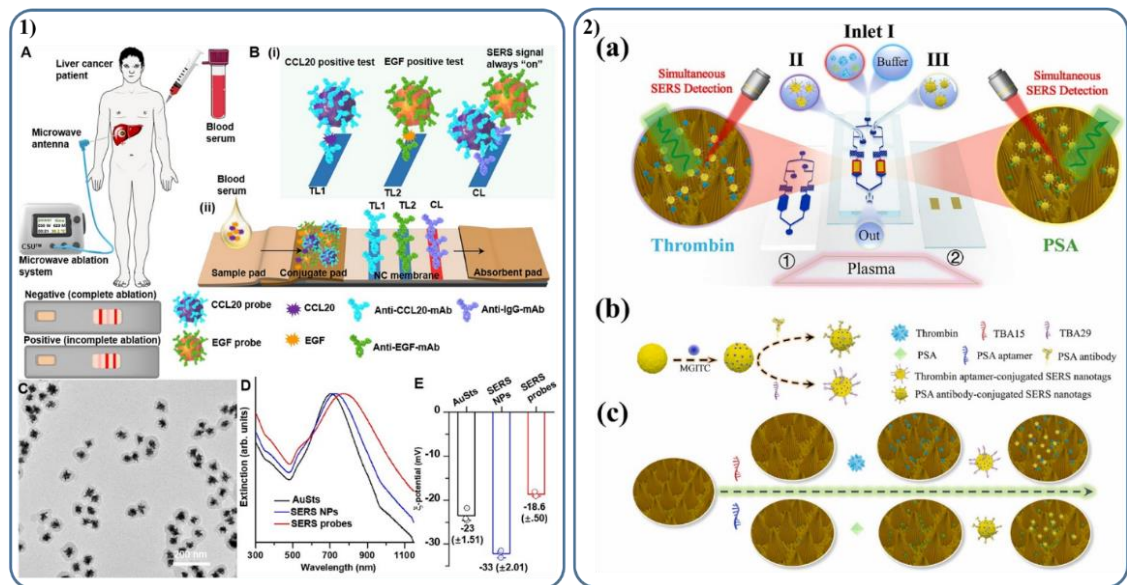


Figure 10. Different SERS approaches for the detection of cancer-related protein biomarkers. (1) Principle of the SERS-LFS strategy for the evaluation of MWA outcomes in HCC patients and characterization of the SERS probe. (a) Collection of blood samples from HCC patients after MWA, (b) SERS immune sensor mechanism – (i) formation of SERS probe/target/antibody complexes on test and control lines. (ii) immune sensor structure enables simultaneous quantification of CCL20 and EGF proteins on two test lines (TL1 and TL2), (c) TEM of SERS nanoparticles, (d-e) optical analysis of SERS substrate and surface charge analysis, respectively, from Ref. [189]; (2) SERS-based microfluidic aptamer sensor for simultaneous detection of two prostate cancer biomarkers. (b) functionalization strategies of SERS nanotags, (c) formation of sandwich immunocomplexes on the multifunctional nanocone array (MNA) platform, from Ref. [191].

In a separate study, Peng *et al.* employed an immunosandwich SERS-lateral flow assay approach to detect prostate-specific antigen (PSA), a key biomarker for prostate cancer, achieving a detection limit of 0.08 ng/mL. The platform demonstrated reliable performance in clinical serum, with recoveries of 96–105% and relative standard deviation (RSDs) below 10% [190]. In a notable advancement, Lu *et al.* developed a microfluidic sensor integrated a multifunctional SERS nanocone array (MNA), enabling the simultaneous detection of PSA and thrombin—two key biomarkers of prostate cancer [191]. The gold-coated MNA, functionalized with PSA and thrombin aptamers, functioned as both a SERS substrate and a capture platform. The device also employed two SERS nanotags, functionalized with MGITC and either PSA antibodies or thrombin aptamers. Upon addition of the target proteins and nanotags, the MGITC SERS signal was recorded, providing a direct readout of both protein levels. This configuration allows for the simultaneous, highly sensitive detection of PSA and thrombin, achieving detection limits of 0.01 ng/mL and 0.01 nM, respectively, while maintaining excellent reproducibility, with a RSDs of 7.37%. This platform offered a powerful approach for early prostate cancer detection, enabling multiplexed analysis and higher sensitivity than the single-PSA detection reported by Peng

et al., underscoring the advantage of integrating nanostructured surfaces with microfluidic platforms for enhanced biomarker detection. Functionalized with PSA and thrombin capture aptamers, the gold-coated MNA functioned both as a SERS substrate and a biomarker capture system. The device employed two SERS nanotags, functionalized with MGITC and either PSA antibodies or thrombin aptamers, as shown in **Figure 10.2**. Upon addition of the target proteins and nanotags, the MGITC SERS signal was recorded, providing a direct readout of both protein levels. This configuration allows for the simultaneous, highly sensitive detection of PSA and thrombin, achieving detection limits of 0.01 ng/mL and 0.01 nM, respectively, while maintaining excellent reproducibility, with a RSDs of 7.37%. This platform offered a powerful approach for early prostate cancer detection, enabling multiplexed analysis and higher sensitivity than the single-PSA detection reported by Peng *et al.*, underscoring the advantage of integrating nanostructured surfaces with microfluidic platforms for enhanced biomarker detection.

5.1.4 Detection of circulating tumor cells (CTCs)

Distant metastases are the leading cause of cancer-related deaths in solid tumors, with circulating tumor cells playing a key role in their development. CTCs are rare cells that enter the bloodstream and allow the cancer to spread to distant organs [192]. CTCs are an important cancer biomarker as they provide information about tumor development, metastasis, evaluation of treatment outcomes and patient prognosis. Furthermore, CTCs are extremely rare in peripheral blood, with most cancer patients exhibiting only about 1–10 CTCs per 10 mL, surrounding from billions of normal blood cells. Furthermore, their short circulation half-life (around 1–2.4 hours) makes their detection and analysis particularly challenging [193], [194]. Consequently, various enrichment technologies have been developed in order to use as selection criteria the physical properties of CTCs—such as size, deformability, electric charge, and density—or their biological characteristics, including the expression of key surface biomarkers. The combination of SERS with these enrichment strategies markedly enhances the sensitivity of biomarker detection and minimize potential interference. To detect non-small cell lung cancer CTCs and separate them from white blood cells (WBCs), Lv *et al.* utilized a physical enrichment method based on micropore membrane filtration with aptamer-based SERS nanoprobe [195]. Although membrane filtration is simple and cost-effective method for size-based CTC capture, its specificity is limited by WBCs of similar size. To overcome this, AuNPs SERS nanoprobe functionalized with the Raman reporter 4-MBA and an aptamer targeting A549 cells were introduced, enhancing specificity (**Figure 11.1**). The microporous parylene membrane was engineered with customized pore sizes, allowing rapid capture of SERS-tagged CTCs via gravity filtration in under 10 seconds, thereby minimizing WBC retention. Using Raman signal variations of 4-MBA at 1075 cm^{-1} , the SERS bio-probe effectively distinguished A549 cells from WBCs. Tests with simulated blood samples containing 20 A549 cells demonstrated recovery and detection rates exceeding 90%, highlighting the method's efficiency in detecting low levels of CTCs.

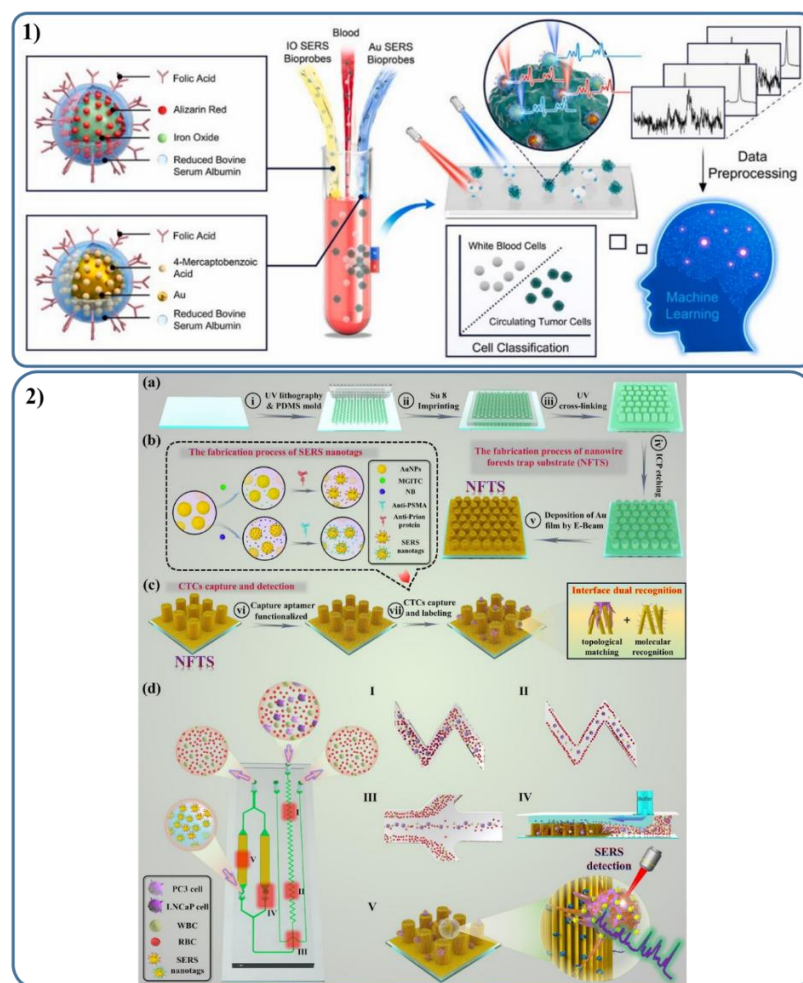


Figure 11. SERS detection methods for circulating tumor cells. (1) Schematic representation of a dual-modal SERS bioprobe and machine learning approach for CTC detection based on folate receptor recognition, from Ref. [130]; (2) Diagram of SERS-based microfluidic liquid biopsy of prostate cancer. (a) Overview of NFTS fabrication; (b) Step-by-step preparation of two distinct antibody-functionalized SERS nanotags; (c) Functionalization of NFTS with capture aptamers and subsequent CTC binding; (d) Integrated workflow on the CMD for multistage CTC capture and in situ single-cell phenotyping via SERS analysis, from Ref. [196].

A significant advancement was achieved by Zhang *et al.*, who developed a highly sensitive method for detecting CTCs from various cancer types using dual-modal SERS in combination with machine learning [130]. Their approach addressed not only the distinction between CTCs and blood cells but also consider the heterogeneity in the expression of surface biomarker, such as folate acid receptor (FR) among tumor cells of different origin (HeLa, MDA-MB-231 and A549). Dual-modal SERS bioprobes, consisting of magnetic and gold nanoprobe, were co-incubated with blood cells and CTCs exhibiting varying levels of FR expression. The magnetic nanoparticles enabled efficient CTC separation, while both types of bioprobes generated SERS signals through their respective Raman reporter molecules (4-MBA and AR), which produced distinct spectral peaks, while folate acid functionalization enhanced selective folate receptor targeting on tumor cells. Machine learning algorithms, including principal component analysis and Random Forest, were applied to Raman spectra to construct an effective CTC identification model. The method successfully detected CTCs at concentrations as low as 2 cells/mL, achieving a 98% detection rate and 90% classification accuracy across HeLa, MDA-MB-231, A549, and

WBCs, which holds significant promise for early cancer diagnosis and personalized medicine.

A key challenge with CTCs is also their pronounced heterogeneity with phenotypes ranging from epithelial to mesenchymal or hybrid states, which influence motility, invasiveness, and adhesion [197]. This heterogeneity is also evident at molecular level in the variable expression of epithelial/mesenchymal surface markers, which complicates their detection. In such cases, multiplexed analyses that detect multiple biomarkers simultaneously improve diagnostic accuracy by providing a comprehensive view of the cancer's molecular profile. In this regard, Zhan *et. al* proposed an innovative dual detection and multiplexed analysis of heterogeneous prostate cancer CTCs based on SERS-integrated microfluidic device [196]. The system employed a nanowire forest trap substrate (NFTS) and gold SERS nanoprobe as SERS substrates. The first component was bioengineered to mimic the natural cellular environment, enabling CTCs to anchor via their adhesion properties, while gold SERS nanoprobe allowed phenotype-specific detection. Specifically, the SERS tags targeted protein proteins in epithelial (PC3) cells and PSMA in mesenchymal (LNCaP) cells, with their distinct Raman reporters (MGITC or NB) producing unique SERS signals for differentiation. SERS analysis revealed that epithelial cells exhibited strong MGITC signals and weak NB signals, whereas mesenchymal cells showed the opposite pattern. The schematic of the combined SERS-integrated microfluidic strategy is reported in **Figure 11.2**. The approach ensured dual mechanical and molecular specificity in CTC detection and classification based on their molecular profiles. The device achieved up to 93.6% CTC capture efficiency from prostate cancer patient blood, with SERS signals reflecting disease stage—mesenchymal markers in advanced cases and epithelial markers in early stages—providing molecular insights to guide precise clinical decisions.

5.2 SERS Bioimaging

Bioimaging refers to the real-time visualization of cells, tissues, organs, and whole organisms, as well as various cellular functions. The main goal of bioimaging techniques is to achieve minimal invasiveness while observing cellular activities [198]. As a real-time and non-invasive optical technique, SERS is a promising bioimaging tool that provides deep insights into intracellular and cellular surface biomarkers, molecular interactions, dynamic processes, and microenvironments [16]. Furthermore, advanced multiplex tissue imaging enables the spatial mapping of numerous biomarkers at the single cell level, allowing detailed investigation of intricate cancerous tissue landscapes [199]. In contrast to fluorescence bioimaging, where are used fluorophores as markers to investigate biomolecules and cellular structures, SERS bioimaging provides superior spatial resolution (< 0.5 micrometers in the visible range), enabling high-resolution imaging of sample [6]. In addition, SERS methods significantly reduce the laser powers used, avoiding light-induced cell damage and photobleaching effects (gradual loss of fluorescence). Furthermore, SERS spectra have a narrower peak width of 0.1 nm, in contrast to the spectral overlap of broad emission profiles typically associated with fluorescence (bandwidths of 20-80 nm), which increases their suitability for analytical multiplexing applications. Moreover, SERS nanotags offer excellent photostability, optical stability, exceptional sensitivity, specificity,

multiplexing capability, showing better performance than fluorescent probes [6], [16], [143]. Due to the considerable advantages of SERS, numerous research groups have developed several SERS-based bioimaging strategies to detect and monitor the distribution of cancer biomarkers in cells and tissues. These approaches often involve modifying SERS probes with specific recognition elements to identify single intracellular or surface biomarkers, enhancing detection sensitivity and specificity. In particular, intracellular imaging, facilitated by endocytic uptake of SERS probes, allows real-time monitoring of biomarker expression and subcellular activity. Surface biomarker imaging, on the other hand, aids pathological assessment by precisely delineating tumor margins for surgical guidance. Multiplex SERS imaging further enables the simultaneous mapping of multiple cancer markers with high signal specificity [24]. SERS-based bioimaging is a powerful technique for detecting cancer at an early stage, tracking treatment progress and tailoring therapies to individual patients, as shown schematically in **Figure 12**.

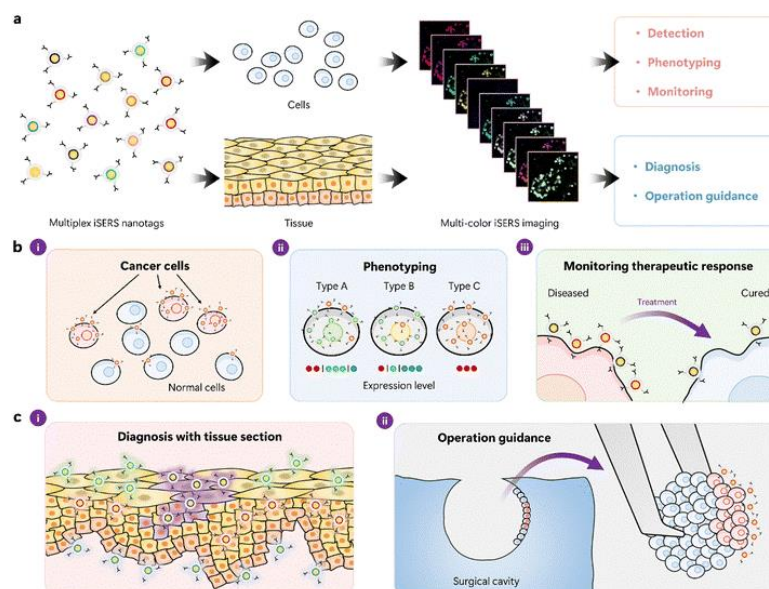


Figure 12. (a) Multiplex SERS imaging for analyzing cells and tissues, application of SERS imaging for: (b) different cell samples analysis and tracking therapeutic outcomes and (c) tissue samples investigation for diagnosis and guiding surgical procedures, from Ref. [143].

This section presents the latest advances in SERS-based bioimaging, highlighting its exceptional capabilities at both cellular and tissue scales. It examines the detection of single and multiple biomarkers—whether intracellular or surface-expressed—through individual and multiplexed approaches, enhancing diagnostic precision, revealing complex biomarker interactions, and enabling more accurate predictions of therapeutic responses. Moreover, recent developments in SERS for *in vivo* tumor targeting and real-time surgical guidance are discussed, showcasing its potential to transform cancer diagnosis and treatment. The integration of multimodal strategies that further augment SERS performance is also considered, emphasizing its growing role as a versatile and powerful platform in modern cancer diagnostics.

5.2.1 SERS imaging of cancer biomarkers in cells and tissues

In SERS bioimaging, targeting a single or multiple cancer biomarkers involves using SERS to detect and visualize a specific molecular marker or panels of biomarkers associated with cancer cells and tissues. For diagnostic applications, it is often essential not only to confirm the presence of a biomarker but also to quantify its abundance. High-resolution SERS imaging can provide detailed information on the spatial distribution and concentration of biomarkers, from intracellular components to cell-surface markers, offering valuable insights into cancer behavior and progression. In this context, Xu *et al.* reported a novel dual-signal SERS strategy for the rapid and sensitive detection of the intracellular biomarker circSATB2, a lung cancer-associated circular RNA [200]. The SERS nanoprobe consisted of AuNPs integrated with two Raman reporters, 4-MBN (internal standard, 2223 cm^{-1}) and ROX (signal reporter, 1507 cm^{-1}), separated by a gold shell. The SERS tags were modified with a double-stranded DNA complex (S1/S2), where the DNA strand S1 specifically hybridized with circSATB2 and DNA strand S2 was labeled with the Raman reporter ROX, enabling circRNA quantification. Upon cellular uptake, the binding of circSATB2 to S1 led to the displacement of S2, resulting in a decrease in the ROX Raman signal. The entire strategy is reported in **Figure 13.1**. Using SERS mapping, circSATB2 biosensor was successfully detected and quantified in normal lung epithelial cells (BEAS-2B) and lung cancer cell lines (H1650 and HCC827) showed Raman intensity (I_{1507}/I_{2223}) significantly lower in cancer cells, indicating higher circSATB2 expression. The SERS-based biosensor achieved a LOD for circSATB2 of 0.043 pM, highlighting its potential for lung cancer diagnosis.

While intracellular biomarkers provide important insights, examining the expression of cell surface receptors is equally critical, as many of these receptors are frequently overexpressed in cancer cells and serve as key mediators of signaling pathways that drive tumor growth, invasion, and metastasis. Assessing their differential expression can therefore offer valuable diagnostic and therapeutic insights in cancer research. In this context, Spaziani *et al.* proposed a SERS-Raman screening technique, aiming to quantify HER2 levels in different breast cancer cell lines (BCs) and simultaneously investigate associated metabolic changes linked to HER2 expression and membrane distribution [201]. SERS mapping was performed using 40 nm gold nanoparticles functionalized with 4-MBA (Raman reporter) and trastuzumab (TZ), the HER2-targeting antibody. Three BC cell lines with varying HER2 expression—SKBR3 (high), MDA-MB-468 (low), and MCF10A (non-tumorigenic)—were incubated with the AuNPs-4-MBA-TZ nanotags. This approach enabled direct detection and quantification of HER2 at the single-cell level, allowing clear differentiation of cells based on HER2 expression. By combining SERS with RNA interference to inhibit HER2, the study reduced background signals and provided deeper insights into the biological characteristics of HER2-positive cells. This strategy simulates the effects of HER2-targeted therapies and represents an innovative method to enhance diagnostic precision and evaluate therapeutic efficacy in HER2-positive breast cancer. While understanding the expression and distribution of a single cancer biomarker is important, cancer is a complex, multifactorial disease, and relying on a single protein often provides limited sensitivity and specificity.

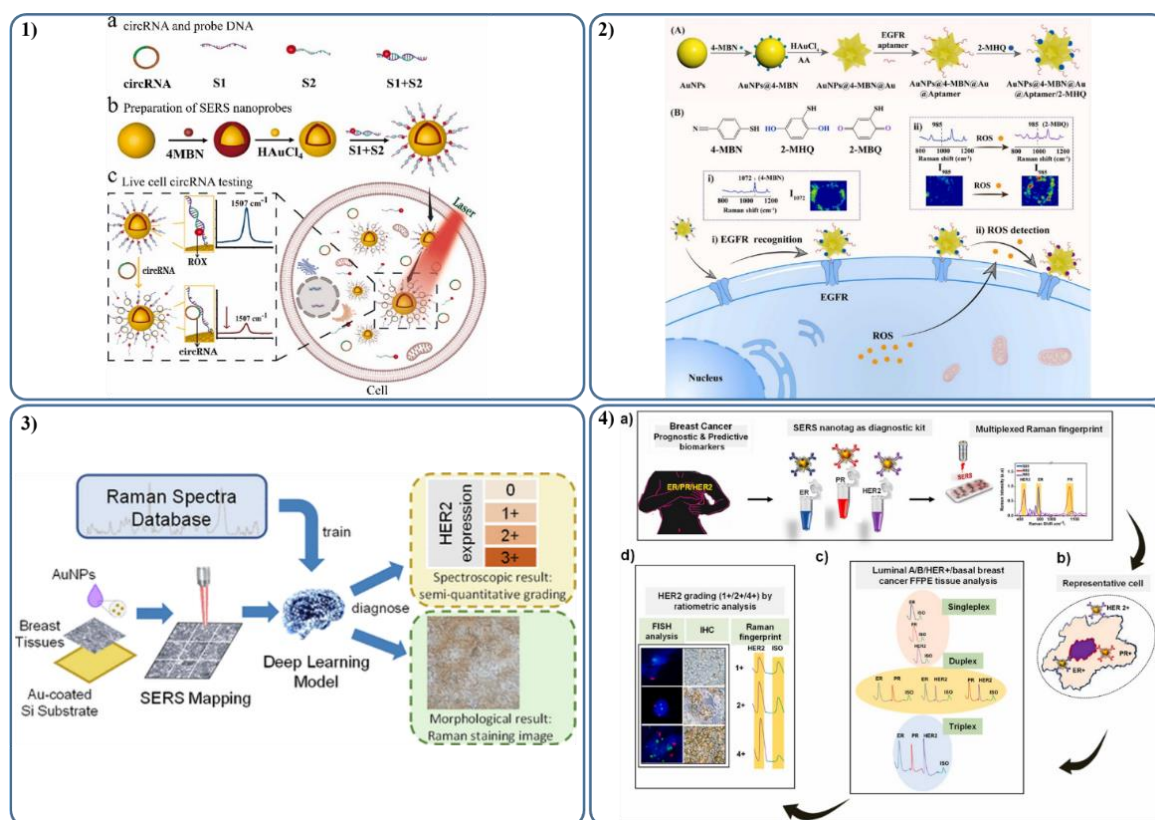


Figure 13. SERS bioimaging of individual and multiple cancer biomarkers at the cell and tissue levels. (1) Schematic representation of SERS detection of intracellular circSATB2. (a) circSATB2 and probe DNA, (b) SERS nanoprobe design, (c) detection of circSATB2 in living cells, from Ref. [200]; (2) Schematic diagram of a core-shell SERS nanosensor for simultaneous imaging of EGFR on cell membranes and ROS secretion from living cells. (a-b) Schematic representation of the fabrication process of AuNPs@4-MBN@Au@Aptamer/2-MHQ nanosensors and their application in detecting EGFR and ROS on cell membranes, from Ref. [202]; (3) Schematic illustration of a combined morphological and spectroscopic diagnostic method using label-free SERS with a deep learning model to detect HER2 expression in breast cancer tissue from Ref. [203]; (4) Schematic diagram of the SERS-based spectro-histology for distinguishing the clinically relevant biomarkers ER, PR, and HER2, from Ref. [204].

Combining it with profiling of additional biomarkers can offer complementary insights into tumor biology, enhancing both diagnostic accuracy and prognostic assessment. In this context, the multiplexing capability of SERS enables simultaneous detection of multiple cancer biomarkers, offering deeper insights into tumor cell mechanisms. Notably, Chen *et al.* recently investigated the interplay between different biomarkers, such as the overexpressed surface protein EGFR, associated with abnormal cell proliferation, and ROS, linked to oxidative stress [202]. They developed multifunctional core-shell SERS nanoprobe capable of simultaneous dual imaging of EGFR and ROS in living cells. The SERS nanoprobe integrated three key components: (1) an embedded Raman reporter, 4-MBN (1072 cm^{-1}), serving as an internal standard and EGFR signal reference; (2) EGFR-specific aptamers for targeted molecular recognition; and (3) a ROS-responsive molecule, 2-mercaptohydroquinone (2-MHQ, 985 cm^{-1}), which undergoes oxidation to quinone under ROS exposure, resulting in distinct SERS spectral changes. The SERS probe design and mechanism are shown in **Figure 13.2**. When applied to MCF-7 breast cancer cells, these nanoprobe enabled real-time, concurrent monitoring of EGFR expression and ROS production, revealing a positive correlation between the two biomarkers—suggesting ROS-mediated upregulation of EGFR and reciprocal ROS generation promoted by EGFR activation. This system demonstrated the capability of SERS-based platforms to elucidate

dynamic relationships among interrelated cancer biomarkers. While cell models serve as valuable *in vitro* tools for investigating biomarker expression and unraveling cancer-related cellular mechanisms, transitioning to tissue-level detection introduces a more relevant context.

Tissues encompass diverse cell types and a complex extracellular matrix that collectively reproduce the structural and biochemical heterogeneity of the *in vivo* microenvironment. This complexity allows for a more comprehensive and realistic assessment of biomarker localization, distribution, and interaction within their native surroundings. Consequently, extending SERS-based detection from cells to tissues represents a critical step toward translating this technology into clinically meaningful applications and understanding biomarker dynamics in real biological systems.

In their study, Mo *et al.* combined label-free SERS imaging with a deep learning model to assess HER2 expression and distribution in breast cancer tissue samples (see **Figure 13.3**) [203]. Tissue samples were prepared by depositing them onto gold-coated silicon SERS substrate, followed by the addition of AuNP-based SERS nanoprobes. A HER2-specific spectral database was created with collected spectra from purified HER2 protein, HER2-positive breast cancer tissues, and normal breast tissues. A ResNet-based deep learning model was trained to accurately recognize HER2-specific signals within complex tissue spectra. During test, the model analyzed tissue SERS maps to quantify HER2 expression by counting HER2-positive spectra and generating Raman-staining images that provided additional morphological information. Traditionally, HER2 status is assessed by immunohistochemistry (IHC), which labels proteins with antibodies and classifies expression from 0 to 3+ (0 = negative, 1+ = low expression, 2+ = requires FISH testing, 3+ = positive). These assessments rely on morphological evaluation and can be affected by subjective interpretation, leading to observer variability [205]. Validation of the SERS approach using 40 clinical samples (10 each of IHC 0, 1+, 2+, and 3+) showed up to 95% concordance with IHC, demonstrating high sensitivity, quantitative capability, and label-free potential for clinical breast cancer assessment. Currently, artificial intelligence is widely applied to analyze spectroscopic medical data objectively, supporting clinicians in cancer diagnosis and treatment response prediction [206], [207], [208].

Recently, SERS-based tissue analysis has evolved from single-biomarker detection to multiplexed monitoring, improving diagnostic accuracy, predicting treatment responses more effectively, and enabling personalized medicine by analyzing multiple molecular pathways simultaneously. In this regard, Li *et al.* reported a super-multiplexed SERS imaging approach for mapping immune checkpoint networks (ICPs) in breast cancer tissues, as schematically shown in **Figure 13.4** [209]. Immune checkpoints regulate immune responses, and tumors exploit inhibitory checkpoints like PD-L1 to suppress T cell activity. Blocking these pathways can restore antitumor immunity, forming the basis of cancer immunotherapies [210], [211]. The Li research team engineered 10 super-multiplexed VASERS nanoprobes for detecting key immune checkpoint proteins, bioengineering silver nanoparticles with Raman reporters, PDA coating, and antibody targeting. Each nanoprobe produced a unique Raman signal in Raman silent region ($1800\text{--}2800\text{ cm}^{-1}$), avoiding spectral overlapping and enabling precise multiplexed imaging of ICPs. The VASERS nanoprobes were incubated with breast cancer tissue from three different patients and the resulting

expression patterns of ICPs were analyzed and interpreted by integrating SERS and correlation network analysis, contributing to the advancement of immune therapy strategies for breast cancer. In the study by Murali et al., SERS multiplexing strategy was applied at both the cellular and tissue levels [204]. They reported a novel SERS-based spectro-histology approach using bioengineered gold nanoparticles to simultaneously detect multiple breast cancer biomarkers, including estrogen receptor (ER), progesterone receptor (PR), and HER2. Gold nanoparticles were modified with distinct Raman reporters: crystal violet (CV, 440 cm^{-1}), 4-MBA (1084 cm^{-1}), and squaraine di-lipoic acid (SDL, 580 cm^{-1}), coated with a carboxy-PEG layer for stability and conjugated with antibodies specifically targeting HER2, PR and ER. A non-targeted control SERS tag was also prepared using a Raman reporter (DTNB) and a non-specific IgG antibody. *In vitro* testing involved incubating these SERS nanoprobe with three breast cancer cell lines: MCF-7 (ER+/PR+), SKBR-3 (HER2+), and MDA-MB-231 (triple negative). Detection of distinct Raman peaks in SERS maps allowed selective recognition and differentiation of breast cancer cell lines. Validation on FFPE clinical tissue sections used ratiometric evaluation to reduce false positives using non-targeted control SERS tags. The SERS analysis for tissue specimens showed sensitivity and specificity rates of 95%/92%, 88%/85%, and 75%/67% for the detection of individual biomarkers (singleplex), a combination of two biomarkers (duplex) and of all three biomarkers (triplex), respectively. Furthermore, HER2 grading of 4+, 2+, and 1+ tissue samples were semi-quantitatively assessed using ratiometric SERS intensity profiling, consistent with FISH results. Collectively, these works underscore the promise of SERS in advancing accurate and sensitive cancer diagnostics.

5.2.2 SERS imaging *in vivo*

In addition to the detection of cancer biomarkers in solutions, on cell membranes or at the intracellular level, SERS nanotags can also be used for *in vivo* tumor targeting. These probes are functionalized with tumor recognition elements and injected into animal models for "active" targeting. The use of SERS nanoprobe for *in vivo* tumor targeting offers significant advantages as the unique SERS fingerprint enables accurate tracking in complex biological environments or use SERS imaging to differentiate tumors containing SERS probes from surrounding normal tissues by using the characteristic Raman signals of reporter molecules [129]. SERS tags are also becoming a valuable tool for precise image-guided surgery. They help surgeons to accurately identify tumor margins and detect microscopic lesions that are often overlooked by current imaging techniques [16]. This advance addresses challenges in cancer surgery, such as incomplete resection or damage to healthy tissue, and improves outcomes by enabling more accurate tumor removal and reducing the risk of recurrence and metastasis.

In 2008, Quian et al. were pioneers in using bioengineered gold nanoparticles as SERS nanoprobe for *in vivo* tumor targeting of EGFR detection in xenograft tumor models [212]. Since then, SERS imaging has expanded its potential for *in vivo* applications. In this context, Qiu *et al.* utilized SERS imaging to detect surface biomarkers and assess treatment response in breast cancer cells and xenotransplanted mouse models, respectively [213]. Three Au@Ag core-shell SERS nanoprobe with dual-layer Raman reporters (MGITC, RBITC,

DTDC) were antibody-conjugated to target EGFR, HER2, and IGF-1, were designed (Figure 14.1).

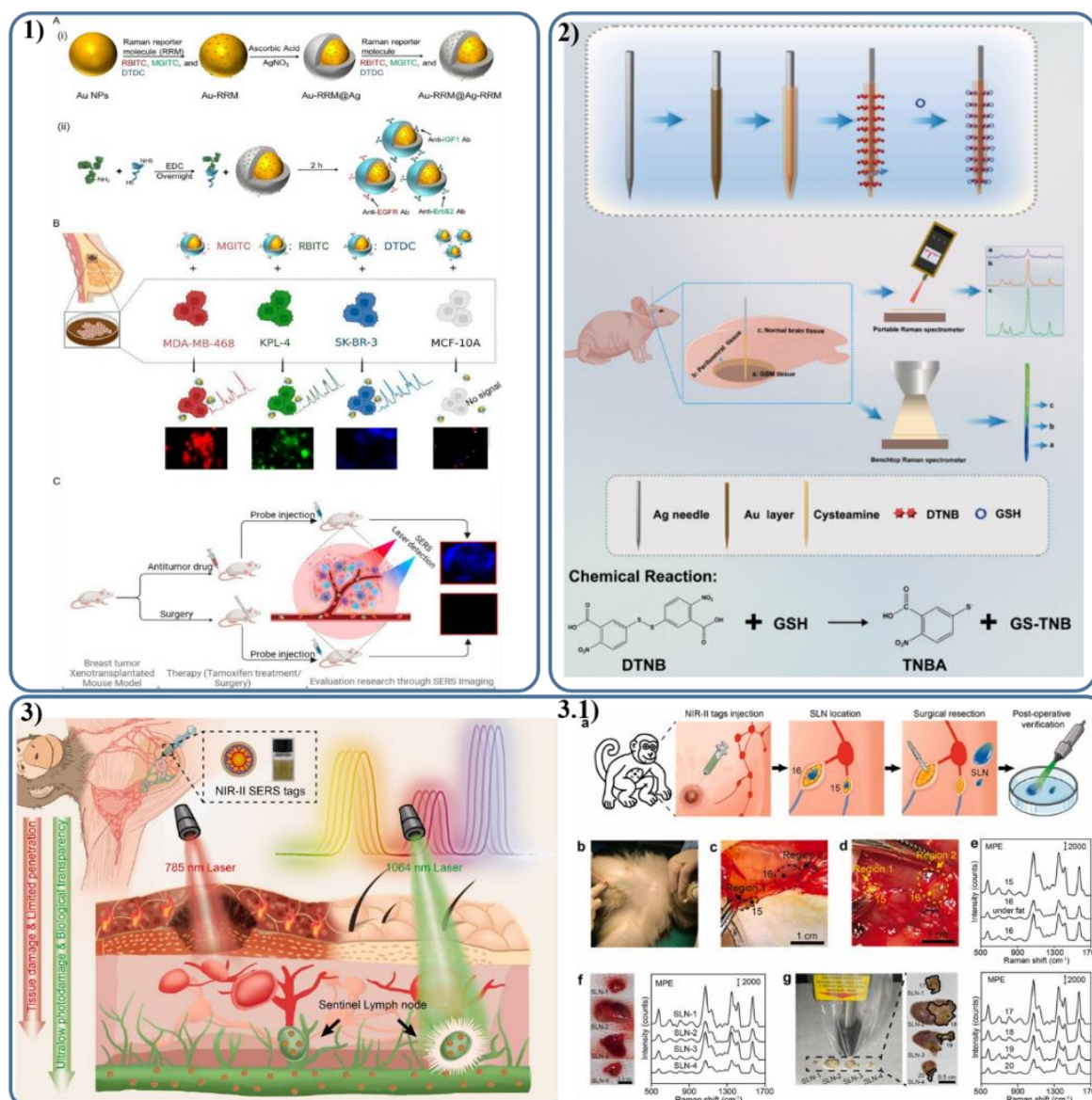


Figure 14. SERS imaging *in vivo*. (1) SERS imaging to identify breast cancer phenotypes and evaluate treatment efficacy. (a) Diagram illustrating the synthesis of three types of Raman reporter-adsorbed Au-Ag core-shell nanoparticles and the attachment of PEGylated antibodies to their surface, (b) SERS imaging to detect phenotypic biomarkers on cell membranes, (c) SERS-based evaluation of treatment effects after antitumor drug therapy and surgery, from Ref. [213]; (3) Identification of sentinel lymph nodes in preclinical non-human primate models using ultra-bright NIR-II SERS nanotags, (3.1) Identification of SLN in rhesus monkeys using NIR-II tags, from Ref. [214].

Average SERS signal intensities of nanoprobes incubated with normal (MCF-10A) and breast cancer cell lines (MDA-MB-468, SK-BR-3, KPL-4) revealed high EGFR expression in MDA-MB-468, HER2 in SK-BR-3, and IGF-1 in KPL-4, while all three markers showed minimal expression in normal cells. SERS imaging of tumor location was also employed to assess therapeutic efficacy in a xenograft breast tumor model. Administration of IGF1-targeted SERS probes revealed decreased tumor signals following tamoxifen treatment (anti-tumor drug) and surgical resection in comparison to negative control, demonstrating the potential of multiplexing SERS imaging for accurate breast cancer diagnosis and therapy monitoring. Eremina *et al.* also reported an advanced multiplexing SERS approach using 26

SERS nanoprobe (AuNPs@Raman reporter@silica shell) for precise molecular profiling of cultured cells, human cancer tissue sections, and *in vivo* mouse models [215]. *In vivo*, the multiplexing capabilities of a combination of 26 SERS tags were investigated after their simultaneous injection into the tail vein of a nude mouse, demonstrating that high-complexity SERS imaging enables improved localization of targeted biomarkers in a living organism.

Active *in vivo* SERS targeting represents a versatile technique for simultaneous, real-time detection of multiple cancer biomarkers and for guiding surgical interventions. By employing bioengineered SERS nanoprobe that preferentially localize within tumor tissues, this method allows accurate mapping of tumor boundaries and biomarker expression during “plasmonic surgery”. The incorporation of SERS imaging into intraoperative workflows improves the precision of tumor removal, reduces the risk of residual malignancy, and offers significant potential for advancing patient care [216]. In this regard Zhang *et al.*, a SERS-active microneedle (DTNB@Au@Ag) was used for the detection of glutathione (GSH) — which is highly expressed in tumor cells to protect them from ROS — in glioblastoma (GBM) tissues [217]. When the SERS microneedle contacted GBM tissues, SERS imaging visualized GSH activity through a decrease in DTNB Raman signal, caused by GSH-induced degradation of the Raman reporter and release of TNBA from the microneedle surface, thereby mapping biomarker distribution (**Figure 14.2**). The SERS-microneedle provided high sensitivity (LOD: 0.037 mM), rapid detection (12 min), and a broad linear range (0.1–15 mM). Moreover, it allows GSH mapping and accurate tissue differentiation in both animal models and patient samples, demonstrating its promise for intraoperative GBM diagnosis.

While notable progress has been made in developing SERS-active nanoplateforms for *in vivo* applications, biosafety concerns persist because nanotags often accumulate in non-target organs, including the liver and kidneys. Subcutaneous injection, especially for sentinel lymph node imaging, is a safer alternative, but further research in non-human primates is needed to account for anatomical differences. The laser safety issues can be addressed by using the NIR-II window (1000-1700 nm), which offers higher maximum permissible exposure (MPE) limits, better light safety and deeper tissue penetration, making the imaging process safer and more effective [218]. In this regard, the research of Deng *et al.*, focused on the development of ultra-bright near-infrared II surface-enhanced Raman scattering (NIR-II SERS) nanotags for the identification of breast sentinel lymph nodes (SLNs) in animal models [214]. Specifically, the NIR-II SERS nanotags were constructed with a core of petal gold surrounded by Raman reporter molecules (IR-1048), an outer silver shell and inner mesoporous silica layers (**Figure 14.3**). Core-shell gap-enhanced Raman tags (GERTs) have successfully demonstrated their efficacy in SERS imaging as tracers for SLNs after subcutaneous injection in rabbits and rhesus monkeys. This research represents a groundbreaking success in identifying multiple SLNs in a non-human primate model using Raman tags as a navigation tool. The results of this study confirm the safety and biocompatibility of NIR-II SERS nanotags and highlight their promising role in clinical practice for accurate SLN localization, which is essential for surgical procedures and postoperative monitoring. Therefore, by enabling highly precise tumor targeting and real-time surgical guidance, SERS is profoundly transforming the landscape of *in vivo* cancer

diagnostics, offering unprecedented accuracy in tumor delineation and paving the way for more effective, image-guided therapeutic interventions.

5.2.3 Multimodal SERS

The development of SERS imaging technology for cancer detection is focused on optimizing it for practical application in the clinical setting. Efforts are being made to improve its sensitivity, precision and ease of use so that it can be used effectively in the diagnosis, monitoring and treatment of cancer. The goal is to move from a research-focused tool to one that can be used routinely in medical practice, enabling reliable, non-invasive cancer detection in various healthcare settings. To achieve this, SERS has been combined with other imaging modalities such as fluorescence, photoacoustics (PA) and MRI, leading to the development of advanced multimodal systems that can offer synergistic platforms for more effective cancer diagnosis and therapy. This integrated approach enables comprehensive analysis at multiple levels and combines the strengths of each technique to compensate for the limitations of the others.

Notably, SERS is an advanced bioimaging platform that combines molecular-level ultrasensitivity with high multiplexing potential. Unlike fluorescence imaging or MRI, which exhibit limited molecular specificity, SERS provide precise biomolecular detection with cellular-level spatial resolution ($<10\ \mu\text{m}$). Its near-infrared (NIR) excitation supports deep tissue penetration (up to $\sim 7\ \text{mm}$), effectively bridging the gap between imaging depth and resolution. Advanced imaging techniques such as NIR-II fluorescence and photoacoustic imaging (PAI) enable deeper centimeter-scale penetration (1–5 cm) but remain constrained by reduced sensitivity and multiplexing performance [219]. SERS uniquely combines molecular specificity, high spatial resolution, and moderate depth penetration—positioning it as a transformative platform for next-generation molecular diagnostics and biomedical imaging. In particular, combining SERS with the previously mentioned imaging modalities enhances both tissue penetration and molecular imaging, yielding a more complete and precise diagnostic platform. In this context, Wang *et al.* developed a dual-mode fluorescence–SERS (FL-SERS) biosensor for real-time imaging and quantitative detection of MUC1 in different living cells [220]. The system integrated quantum dots labeled with the Raman reporter PDDA and a MUC1-specific aptamer (QD-PDDA-Apt) with gold nanostars (AuNSs) functionalized with complementary DNA. In the absence of MUC1, the biosensor shows strong fluorescence and weak SERS signals. Upon MUC1 binding, the aptamer's hairpin structure opened, allowing hybridization with complementary DNA on gold nanostars to form a satellite structure. This led to fluorescence quenching (threefold reduction) and enhanced SERS signals due to AuNS hotspots, as shown in **Figure 15.1**. Hence, the variations in fluorescence and SERS responses enabled the biosensor to sensitively detect and simultaneously image MUC1 in different cell types, thereby distinguishing MCF-7 breast cancer cells from HepG2 liver cancer cells based on their differential MUC1 expression. Furthermore, the biosensor showed high sensitivity for MUC1 detection, with LOD of 1.19 fg/mL (FL) and 1.16 fg/mL (SERS).

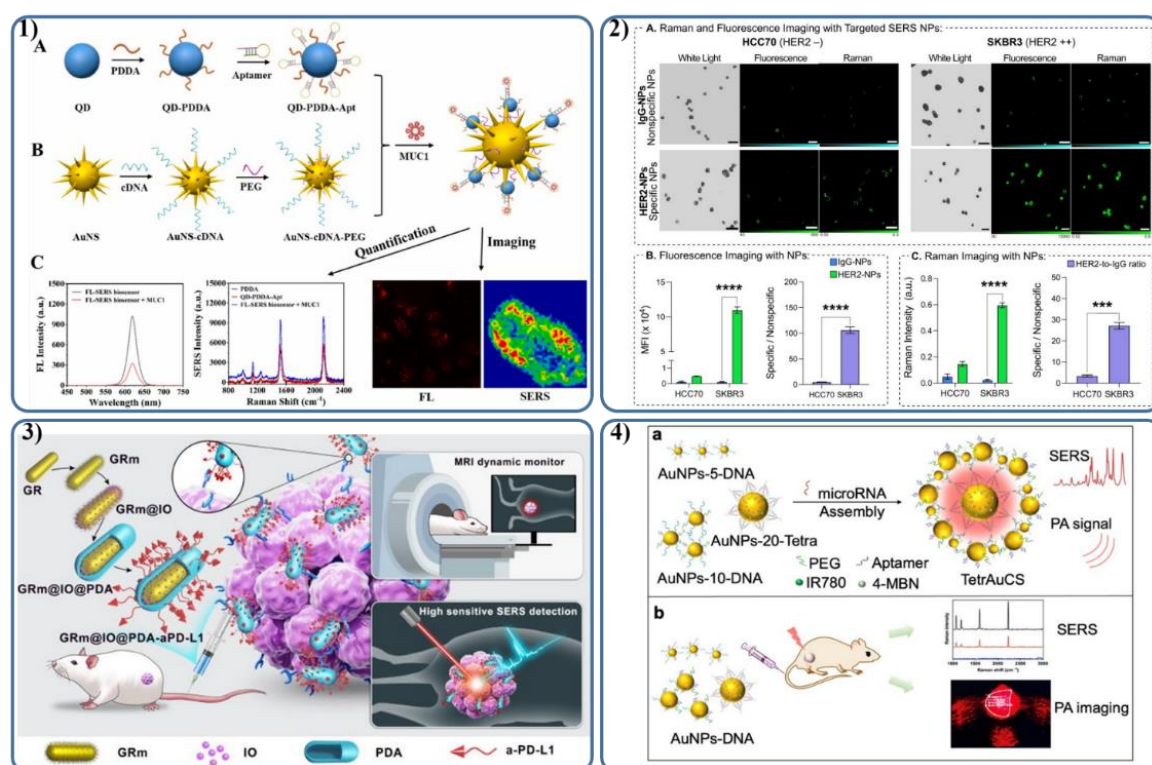


Figure 15. Different multimodal SERS approaches for cancer diagnostics. (1) Schematic of the FL-SERS biosensor for the detection and visualization of MUC1. (a) Preparation of QD-PDDA-Apt complex and (b) AuNS-cDNA-PEG, (c) Detection and imaging of MUC1 in living cells using FL-SERS biosensor, from Ref. [220]; (2) Bioconjugated SERS nanoparticles for maximal targeting efficiency in breast cancer cell lines. (a) Imaging (brightfield, fluorescence and Raman) showed strong signals in SKBR3 cells incubated with HER2-targeted NPs, (b) quantitative analysis revealed significantly higher intensity of fluorescence and Raman signals in HER2-expressing SKBR3 cells compared to IgG controls, from Ref. [221]; (3) Design of dual-modal SERS-MRI bioprobes for sensitive and specific detection of PD-L1 expression in tumors, from Ref. [222]; (4) Illustration of the synthesis and in vivo evaluation of TetrAuCS for SERS and PA imaging, from Ref. [223].

To perform multiplexed imaging, Bagheri *et al.* proposed a dual-mode FL-SERS strategy to for efficient active targeting of EGFR and HER2, overexpressed in breast cancer cell lines [221]. The FL-SERS probes, which consist of 60 nm AuNPs, were labelled with the Raman reporter BPE, protected with a silica coating, and further functionalized with the fluorescent dye DyLight650-maleimide and monoclonal antibodies specific to target EGFR and HER2. Control nanoparticles functionalized with non-specific IgG antibodies were used to evaluate non-target binding. Following incubation with cells, fluorescence-SERS imaging results revealed selective binding of the FL-SERS probes to biomarker-overexpressing cell lines, such as SKBR3 (high HER2) and A431 (high EGFR), while control cell lines, including HCC70 (low HER2) and MDA-MB-231 (low EGFR), showed minimal non-specific interactions, as shown in **Figure 15.2**. Fluorescence imaging enabled rapid visualization, whereas Raman imaging provided precise molecular mapping, confirming the specificity and the contrast agent activity of the FL-SERS probes. Additionally, other research groups have explored the integration of the deep tissue penetration offered by MRI as a powerful anatomical imaging technique with the molecular sensitivity of SERS, providing an important resource for the detection of molecular-level changes associated with cancer biomarkers [224].

In their study, Pan *et al.* developed a combined dual SERS-MRI imaging strategy using gold nanorods iron oxide (GRm@IO) to detect and monitor PD-L1 expression *in vitro* and *in vivo*

models of triple negative breast cancer (TNBC) [222]. The bioengineered SERS probes consisted of gold nanorods modified with the Raman reporter 4-MBA, conjugated with iron oxide nanoparticles, coated with a polydopamine (PDA) layer to enhance biocompatibility, and functionalized with PD-L1 antibodies (aPD-L1) for targeted detection of PD-L1 (**Figure 15.3**). Single-cell SERS imaging analysis *in vitro* revealed that MDA-MB-231 TNBC cells exhibited high PD-L1 expression, producing stronger SERS signals compared to MCF-7 cells with lower PD-L1 levels. MRI confirmed the accumulation of the probes in TNBC tumors, enabling real-time *in vivo* tracking of PD-L1 expression. Overall, this dual SERS–MRI approach demonstrates a synergistic and accurate strategy for tumor imaging in TNBC. Another promising imaging technique is photoacoustic imaging (PAI), which is characterized by its ability to generate high-resolution optical contrast images in biological tissue at a depth of several centimeters [225]. The combination of SERS and PAI results in a unique dual approach that effectively combines the high molecular specificity and sensitivity of SERS with the ability of PAI to image deep tissues, enabling precise biochemical detection and high-resolution imaging capabilities. In this regard, Tan *et al.* developed a SERS-PAI dual-imaging strategy for *in vitro* and *in vivo* breast cancer, enabling MiR-21 detection via SERS and tumor tissue visualization via PAI [223]. The SERS-PAI system consisted of a self-assembly of the DNA tetrahedral structure-linked gold nanoparticle (AuNPs). Specifically, the SERS-PAI platform (TetrAuCS) was engineered as a DNA-assembled core–satellite system, in which a 20 nm central AuNP serves as the core, surrounded by two layers of smaller satellite AuNPs sized 10 and 5 nm, enabling precise assembly and enhanced plasmonic interactions. The design incorporated IR780 for tumor photoacoustic imaging, a Raman reporter 4-MBN coating on the core, and surface-bound aptamers. The presence of MiR-21 induced the self-assembly of AuNPs SERS nanoprobe of different dimensions into the core–satellite TetrAuCS architecture, as shown in **Figure 15.4**. The SERS signal of 4-MBN at 2228 cm^{-1} was measured against varying MiR-21 concentrations. An “off-to-on” SERS response was observed in the presence of MiR-21, with a linear correlation between signal intensity and MiR-21 concentration. The LOD was determined to be 1.97 pM. The miRNA imaging capability of TetrAuCSs was tested in living cells, showing a time-dependent increase in 4-MBN SERS signal in MiR-21–high MCF-7 cells, while MiR-21–negative MCF-10A cells exhibited no signal. Furthermore, the TetrAuCS produced strong SERS signals in mice bearing MCF-7 tumors, confirming MiR-21-induced accumulation at tumor sites. Additionally, The PAI performance of TetrAuCSs with IR780 was evaluated for *in vivo* MiR-21 detection, showing signal enhancement via LSPR effects. Incorporation of a PTK-7-targeted aptamer TetrAuCSs@IR780 improved cellular uptake and tumor penetration, producing stronger and deeper photoacoustic signals than IR780 alone, demonstrating potential for tumor imaging. These advancements demonstrate the increasing significance of SERS in diagnostic applications, progressively bridging the gap between laboratory research and clinical implementation.

6. Optical Fibers (OFs) as innovative SERS platform for cancer diagnostic

The integration of SERS with microstructured optical fibers has led to the development of advanced fiber-optic SERS sensors, significantly expanding their analytical capabilities.

These sensors enable applications ranging from *in vitro* to *in vivo* studies, allowing highly sensitive detection of biomarkers in biological fluids, tissue biopsies, and living organisms. A SERS-based optical fiber sensor that enables sensitive and selective detection of biomarkers in biological fluids or during surgery would be a transformative tool for personalized medicine. By facilitating minimally invasive, real-time diagnostics, such a platform could guide treatment decisions, optimize therapeutic outcomes, and reduce patient risk. In this regard, Spaziani *et al.* developed a SERS-based biosensing platform using a sandwich strategy with planar and fiber-optic substrates for detecting thyroglobulin (Tg) *in vitro* and in FNAB washout fluids from thyroid cancer patients. [100]. The SERS platform consisted of two SERS substrates: 1) a periodic array of gold nanoislands fabricated by nanosphere lithography directly onto planar on-chip and fiber-tip surfaces, subsequently functionalized with DSP and anti-Tg capture antibodies, and (2) gold nanoparticles labeled with the Raman reporter 4-MBA and conjugated with anti-Tg detection antibodies (**Figure 16.1**). In the presence of Tg, a SERS sandwich complex formed, positioning the Raman reporter between the SERS-active surface and nanoprobe, generating a signal proportional to Tg concentration. An evaluation of the detection capability of the on-chip platform was performed with different Tg concentrations (0.1, 1, 10 and 100 ng/mL) in buffer solution, resulting in a LOD of only 7 pg/mL. Unlike conventional planar SERS substrates, the fiber-optic SERS sensor offers the potential for integration into a biopsy needle, representing a promising step toward clinical application.

In another work, Xia *et al.* presented an advanced SERS-optical fiber biosensor specifically designed to detect single cancer exosomes via SERS sandwich approach [101]. The optical fiber sensor, a hollow-core anti-resonant fiber (HCARF) was functionalized with a CD63-specific aptamer to capture exosomes, while Ag nanoparticle SERS nanoprobe labeled with 4-MBN and an HER2-specific aptamer enabled selective detection, as shown in **Figure 16.2**. The sensor exhibited a linear Raman response for exosome concentrations from 8 to 1600 particles/ μ L, with a LOD of 2.2 particles/ μ L using only 253 nL of sample. It showed high specificity for HER2-positive SKBR3 exosomes and was further adapted for multiplex detection using SERS nanoprobe engineered with different Raman reporters (4-Mpy, 2NAT, 4-NBT, MGITC) to target other protein cancer biomarkers, such as ER, MUC1, EGFR, and EpCAM. This multiplexing capability enabled profiling of exosomes from various cancer cell lines and distinguishing between cancer types based on the surface protein signatures of exosomes, highlighting the platform's strong potential for cancer diagnosis. Interestingly, Khetani *et al.* reported another type of SERS-optical fiber sensor, specifically developed for cancer cell detection [226]. Specifically, a portable, cost-effective SERS-based hollow core photonic crystal fiber (HC-PCF) was realized for leukemia cells (HL60 cell line) detection for acute myeloid leukemia (AML), providing a practical alternative to complex and costly techniques such as flow cytometry and PCR.

The detection approach enhanced Raman signals from leukemia cells using silver nanoparticles co-injected into a modified SERS-optical fiber with a 20 μ m core tailored to cell size. The fiber design and 785 nm excitation wavelength ensured efficient light-cell interaction. The HC-PCF sensor integrated an optical system with a fluidic mechanism, featuring an H-shaped pressure-controlled channel for precise introduction and removal of cell samples. The SERS-HC-PCF sensor showed intense spectral signatures of cell

components such as nucleic acids, proteins and lipids. Using PCA, the platform distinguished between live, apoptotic, and necrotic cells, while partial least square regression (PLSR) showed a strong linear correlation between cell concentration and Raman intensity. The sensor detected as few as 300 cells/mL, outperforming flow cytometry sensitivity limits. Overall, this SERS–HC-PCF platform offers a rapid, sensitive, and low-cost tool for leukemia diagnosis and real-time monitoring of cell health. Notably, SERS-based optical fibers enable also the detailed investigation of intracellular extracellular biological processes, providing high-resolution insights at the single-cell level. Regarding to this, Wang *et al.* presented a minimally invasive SERS- based fiber-optic sensor for real-time

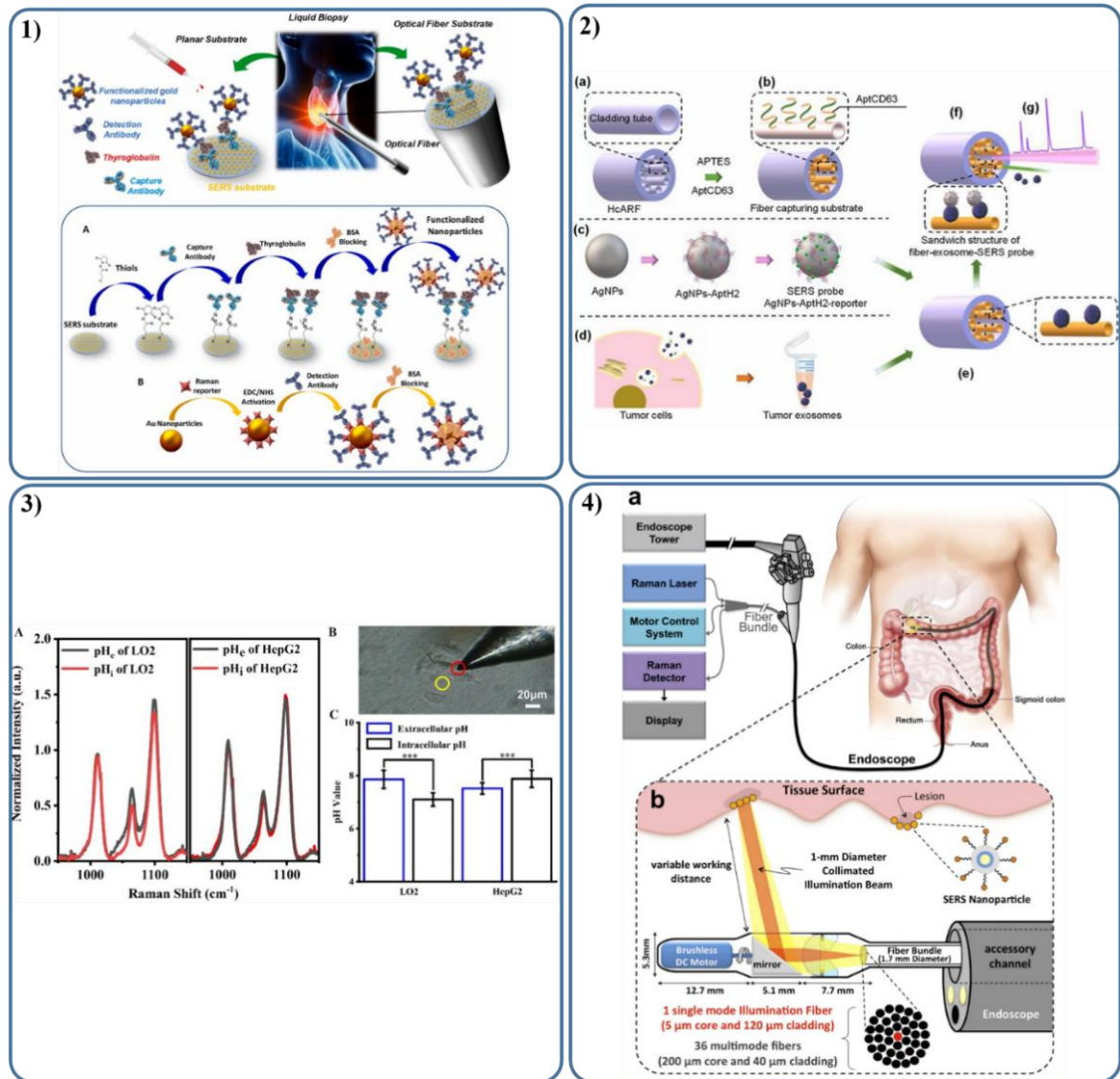


Figure 16. SERS-based optical fiber approaches for cancer diagnostics. (1) Schematic illustration of an on-chip SERS assay and an optrode SERS biosensor for the detection of Thyroglobulin (Tg). (a) Functionalization steps in the SERS-based sandwich immunoassay strategy based on bioengineered SERS substrate with capture antibodies to specifically detect Tg molecules, BSA blocking step to prevent non-specific binding and (b) functionalized gold nanoparticles (f-AuNPs) with a Raman reporter (4-MBA) and detection antibodies to ensure precise Tg recognition, from Ref. [100]; (2) Visual representation detailing the functional mechanism of the fiber-exosome-SERS probe sandwich immunoassay, from Ref. [101]; (3) SERS-enabled fiber tip for detecting intracellular and extracellular pH in living single cells. (A) Comparison between intra- and extracellular pH values in LO2 and HepG2 cells, showing (B) measurement points and (C) corresponding pH values in a histogram, from Ref. [102]; (4) Diagram of fiber-optic-based Raman imaging device integrated with white-light endoscopy and SERS nanoprobe for multiple cancer biomarker detection, from Ref. [103].

monitoring of intracellular and extracellular pH at the single-cell level [102]. The research focused on exploring pH differences between healthy and cancerous cells avoiding damage to living cells. The sensor featured a fiber tip decorated with silver nanoparticles and functionalized with the pH-sensitive Raman reporter 4-mercaptopyridine (4-Mpy). Calibration in buffered solutions demonstrated reliable pH detection between 5.01 and 9.10. Applied to living cells, the sensor distinguished normal liver cells (LO2) from liver cancer cells (HepG2) by detecting expected and reversed pH gradients, respectively (**Figure 16.3**). It also revealed subcellular pH heterogeneity in breast (MCF-7), liver (HepG2), and gastric (SGC7901) cancer cells, with significant nucleus–cytoplasm differences in MCF-7 and HepG2 but not in SGC7901. Viability assays confirmed minimal cellular damage, highlighting the sensor’s suitability for non-destructive, high-precision pH measurements in living cells.

Fiber-optic Raman systems allow *in vivo* detection of SERS nanoprobe signals and facilitate real-time imaging during surgical procedures. These compact, user-friendly devices—including endoscopic and handheld probes—overcome the limitations of conventional microscopy and support both scanning and wide-field imaging. In endoscopic settings, fiber-optic Raman spectroscopy provides minimally invasive access to deep tissues while delivering precise molecular information for diagnosis and monitoring [17]. In this regard, Garai *et al.* developed a Raman imaging device based on the integration of fiber optic system with standard endoscope, perform multiplexing targeting by using SERS nanoprobe in a pilot human clinical study [103]. This technology offered molecular insights that complement traditional white-light endoscopy and improve the early detection of gastrointestinal lesions, as depicted in **Figure 16.4**. The device had a fiber-optic system with a laser (785 nm) that transmits light through a single-mode fiber, while 36 flexible multi-mode fibers collect the Raman signals, making it compatible with clinical endoscopes. The system also incorporates a rotating mirror that performs a circumferential scan and captures the Raman signals for detailed analysis. A hybrid algorithm processes the data to detect and quantify different SERS nanoparticle signals. *In vitro*, *ex vivo* and *in vivo* tests in animal models and initial human trials had shown high sensitivity, effective multiplex detection and the ability to integrate 3D surface mapping. Although the nanoparticles have not yet received clinical approval, the system is very promising for improving the detection of gastrointestinal cancer.

7. SERS challenges in cancer diagnostics

SERS has proven to be a powerful technique in advanced cancer diagnostics, but several critical challenges hinder its full clinical implementation. In this context, the stability of SERS tags is an important issue. For example, plasmonic nanoparticles that provide SERS enhancement are highly sensitive to environmental factors such as temperature fluctuations, interactions with ions and changes in pH, leading to agglomeration and reduced functionality. In addition, this instability affects the long-term storage and overall reliability of SERS tags, making them unsuitable for long-term use in the field [31]. To overcome these problems, the stability of SERS tags needs to be improved so that they are more reliable in challenging biological environments such as body fluids, cells and tissues that contain proteins, enzymes and other molecules that can degrade, interfere or interact non-specifically

with the tags. This can be achieved through various strategies, including i) surface modification to make the nanoparticles more resistant to external influences, ii) optimization of storage conditions to minimize the risk of agglomeration, and iii) use of polymer encapsulation techniques, such as coating the nanoparticles with protective materials or with an additional metal layer on the outer surface to prevent competitive adsorption of cellular molecules on the Raman tags while achieving enhanced signal amplification for the molecules between core and shell, iv) improving the calibration curve by simulating the cellular environment to increase reliability [27], [31]. Another important issue is the control of the signal-to-noise ratio, which affects the accuracy and clarity of the detected SERS signals and affects the reproducibility of the SERS substrates. These performance variations make it difficult to obtain consistent and reproducible results across different experiments or samples and limit the wider use of SERS in analysis and diagnostics. To address this problem, strategies such as reducing noise from impurities and interfering species, standardizing SERS protocols, fabricating or synthesizing uniform SERS substrates, and optimizing nanostructure design — by adjusting size, shape, coupling modes, and LSPR properties — can help maximize signal amplification [227]. In addition, the biosafety and biocompatibility of SERS tags are two other important aspects, especially for *in vivo* applications. SERS tags usually consist of nanoparticles, which may pose risks depending on their size, surface properties and materials. Although gold is generally considered safe, its toxicity can vary depending on the surface ligands, while silver nanoparticles are known to produce ROS that can damage DNA. Despite the common use of gold, there are concerns about its slow removal from the body as it can lead to prolonged exposure to ROS. Although SERS tags generally have low toxicity, their gradual removal remains a challenge [228]. Biocompatibility refers to the ability of SERS tags to interact with biological systems without causing adverse effects or immune reactions. In contrast, biosafety is about reducing potential toxicological risks. Various strategies can be used to improve biocompatibility and biosafety, such as surface functionalization that introduces a biocompatible coating, modifications of charge density, hydrophilicity and hydrophobicity, and the use of naturally derived cell membranes. The use of these methods not only reduces toxicity and adverse effects, but also improves biodistribution and pharmacokinetic properties [17]. However, another important aspect is the size of the nanotag, a factor that should be carefully considered when planning experiments for bioimaging. Since the exceptional optical properties of plasmonic nanoparticles vary with size, it is important to choose the right size of nanoparticles to ensure successful detection and imaging results. In particular, the plasmonic behavior of individual Au/Ag nanoparticles intensifies with increasing particle size, which ranges from 2 to 100 nm [229]. The choice of the appropriate size is influenced by several factors such as i) cellular uptake, ii) the effect of enhanced permeability and retention (EPR) and iii) biodistribution and circulation. Regarding the interaction between cells and nanoparticles, numerous *in vitro* studies indicate that optimal cellular uptake occurs mainly in the range of 10 to 60 nanometers [230]. In particular, the EPR effect (indirect tumor targeting) is a phenomenon due to the atypical organization of the vascular endothelial cells of tumor vessels, which, due to their small size, allows more efficient infiltration of nanoparticles into cancer cells compared to healthy tissue. In addition, the impaired lymphatic drainage near the tumor contributes to the nanoparticles remaining longer in the

tumor microenvironment. This phenomenon enables the accumulation of plasmonic nanoparticles in tumor areas. Furthermore, the maximization of the EPR effect is achieved by using probes with a size between 10-200 nanometers [231]. The impact of nanoparticle size on biodistribution, half-life in the bloodstream, cellular uptake, tumor penetration and targeting capabilities is considerable and must be carefully considered in biomedical applications. Especially for *in vivo* applications, it must be considered that nanoparticles smaller than 10 nm are usually rapidly eliminated by the renal system. Nanoparticles larger than 100 nm, on the other hand, are recognized by macrophages of the mononuclear phagocytic system, which leads to their phagocytosis and subsequent deposition in certain organs [230]. A major obstacle to the use of visible light-responsive SERS nanoprobe *in vivo* is the strong absorption and scattering of light by biological tissue, as well as interference from background auto-fluorescence, which attenuates SERS signals. This makes the NIR-I window (650–900 nm) more suitable for *in vivo* imaging and therapy, although it still has limitations in terms of tissue penetration depth compared to the NIR-II window (1000–1700 nm). Recent advances in NIR-II SERS imaging have improved tissue penetration and resolution by reducing light scattering, absorption and auto-fluorescence [17]. The growing demand for portable devices for cancer diagnosis is also facing obstacles due to the need for qualified professionals and the high cost of conventional Raman devices. While compact Raman devices exist, they are often costly, less efficient and have limited capabilities. To overcome these challenges, smaller, less expensive and more powerful SERS devices need to be developed for real-time, on-site testing, which would make SERS technology more widely accessible for clinical use [31]. These challenges highlight the importance of continued research and innovation to realize the full potential of SERS-based diagnostic methods. By improving reliability, these advances will pave the way for successful clinical applications and contribute significantly to the clinical acceptance of SERS-based diagnostics.

8. Conclusions

This chapter reports on the latest advances in the use of SERS for modern cancer diagnostics. It provides an overview of the key elements of the SERS technique, including its fundamental principles, the SERS substrates used, and the development and bioengineering of SERS nanotags. In addition, the chapter explores how SERS is used in biosensing and bioimaging for cancer detection, highlighting both the challenges and significant benefits of using this powerful diagnostic tool in healthcare.

SERS offers remarkable potential as a highly sensitive and specific tool for cancer diagnostics, enabling the detection and spatial mapping of important cancer biomarkers with molecular precision. Its multiplexed imaging capability combined with chemical specificity makes it a powerful platform for identifying tumor heterogeneity and delineating cancer regions. Despite these advantages, several challenges need to be overcome to fully exploit their clinical utility. These include the reproducibility of nanoparticle synthesis, the stability and functionalization of SERS probes in biological environments, and the standardization of protocols for data acquisition and analysis. Issues such as limited tissue penetration, immune responses and regulatory hurdles remain key barriers to clinical translation. Advances in the development of SERS substrates, biocompatibility and the integration of SERS with

complementary imaging techniques promise to overcome these challenges. As research progresses, SERS has the potential to become a transformative diagnostic technology, enabling real-time, non-invasive cancer detection and paving the way for personalized medicine. By bridging the gap between laboratory innovation and clinical application, SERS could significantly improve early cancer detection, treatment planning and monitoring, ultimately improving patient outcomes.

9. References

- [1] Y. Zhou *et al.*, “Diagnosis of cancer as an emergency: A critical review of current evidence,” *Nat. Rev. Clin. Oncol.*, vol. 14, no. 1, pp. 45–56, 2017, doi: 10.1038/nrclinonc.2016.155.
- [2] M. De Palma and D. Hanahan, “The biology of personalized cancer medicine: Facing individual complexities underlying hallmark capabilities,” *Mol. Oncol.*, vol. 6, no. 2, pp. 111–127, 2012, doi: 10.1016/j.molonc.2012.01.011.
- [3] S. Rath, D. Chakraborty, J. Pradhan, M. Imran Khan, and J. Dandapat, “Epigenomic interplay in tumor heterogeneity: Potential of epidrugs as adjunct therapy,” *Cytokine*, vol. 157, no. July, p. 155967, 2022, doi: 10.1016/j.cyto.2022.155967.
- [4] T. Thenrajan, S. Alwarappan, and J. Wilson, “Molecular Diagnosis and Cancer Prognosis—A Concise Review,” *Diagnostics*, vol. 13, no. 4, 2023, doi: 10.3390/diagnostics13040766.
- [5] T. C. Dakal *et al.*, “Emerging methods and techniques for cancer biomarker discovery,” *Pathol. Res. Pract.*, vol. 262, no. April, pp. 0–2, 2024, doi: 10.1016/j.prp.2024.155567.
- [6] M. Blanco-Formoso and R. A. Alvarez-Puebla, “Cancer diagnosis through sers and other related techniques,” *Int. J. Mol. Sci.*, vol. 21, no. 6, 2020, doi: 10.3390/ijms21062253.
- [7] A. Pulumati, A. Pulumati, B. S. Dwarakanath, A. Verma, and R. V. L. Papineni, “Technological advancements in cancer diagnostics: Improvements and limitations,” *Cancer Rep.*, vol. 6, no. 2, pp. 1–17, 2023, doi: 10.1002/cnr2.1764.
- [8] M. M. U. D. Malik, M. M. Alqahtani, I. Hadadi, I. Kanbayti, Z. Alawaji, and B. A. Aloufi, “Molecular Imaging Biomarkers for Early Cancer Detection: A Systematic Review of Emerging Technologies and Clinical Applications,” *Diagnostics*, vol. 14, no. 21, pp. 1–24, 2024, doi: 10.3390/diagnostics14212459.
- [9] C. Nie, I. Shaw, and C. Chen, “Application of microfluidic technology based on surface-enhanced Raman scattering in cancer biomarker detection: A review,” *J. Pharm. Anal.*, vol. 13, no. 12, pp. 1429–1451, 2023, doi: 10.1016/j.jpha.2023.08.009.
- [10] S. Hosseini, P. Vázquez-Villegas, M. Rito-Palomares, and S. O. Martínez-Chapa, “Advantages, Disadvantages and Modifications of Conventional ELISA,” in *Enzyme-linked Immunosorbent Assay (ELISA): From A to Z*, Singapore: Springer Singapore, 2018, pp. 67–115. doi: 10.1007/978-981-10-6766-2_5.
- [11] Y. Zhou *et al.*, “Tumor biomarkers for diagnosis, prognosis and targeted therapy,” *Signal Transduct. Target. Ther.*, vol. 9, no. 1, 2024, doi: 10.1038/s41392-024-01823-2.
- [12] T. Shegekar, S. Vodithala, and A. Juganavar, “The Emerging Role of Liquid Biopsies in Revolutionising Cancer Diagnosis and Therapy,” *Cureus*, vol. 15, no. January 2010, 2023, doi: 10.7759/cureus.43650.
- [13] L. Guerrini and R. A. Alvarez-Puebla, “Surface-enhanced raman spectroscopy in cancer diagnosis, prognosis and monitoring,” *Cancers (Basel)*, vol. 11, no. 6, pp. 1–15, 2019, doi: 10.3390/cancers11060748.
- [14] X. X. Han, R. S. Rodriguez, C. L. Haynes, Y. Ozaki, and B. Zhao, “Surface-enhanced Raman spectroscopy,” *Nat. Rev. Methods Prim.*, vol. 1, no. 1, p. 87, 2022, doi: 10.1038/s43586-021-00083-6.
- [15] T. B. Rouf, “Application of SERS Nanoparticles for Imaging for Fast, Sensitive and Selective Cancer Screening: A Review,” *Significances Bioeng. Biosci.*, vol. 6, no. 1, 2023, doi: 10.31031/sbb.2023.06.000629.
- [16] H. Liu, X. Gao, C. Xu, and D. Liu, “SERS Tags for Biomedical Detection and Bioimaging,” *Theranostics*, vol. 12, no. 4, pp. 1870–1903, 2022, doi: 10.7150/thno.66859.
- [17] H. Chang, “In vivo surface-enhanced Raman scattering techniques: nanoprobe, instrumentation, and applications,” *Light Sci. Appl.*, pp. 11–22, 2025, doi: 10.1038/s41377-024-01718-5.
- [18] M. Fleischmann, P. J. Hendra, and A. J. McQuillan, “Raman spectra of pyridine adsorbed at a silver electrode,” *Chem. Phys. Lett.*, vol. 26, no. 2, pp. 163–166, 1974, doi: [https://doi.org/10.1016/0009-2614\(74\)85388-1](https://doi.org/10.1016/0009-2614(74)85388-1).
- [19] J. Langer *et al.*, “Present and future of surface-enhanced Raman scattering,” *ACS Nano*, vol. 14, no. 1, pp. 28–117, 2020, doi: 10.1021/acsnano.9b04224.
- [20] D. L. Jeanmaire and R. P. Van Duyne, “Surface raman spectroelectrochemistry: Part I. Heterocyclic, aromatic, and aliphatic amines adsorbed on the anodized silver electrode,” *J. Electroanal. Chem. Interfacial Electrochem.*, vol. 84, no. 1, pp. 1–20, 1977, doi: [https://doi.org/10.1016/S0022-0728\(77\)80224-6](https://doi.org/10.1016/S0022-0728(77)80224-6).
- [21] M. G. Albrecht and J. A. Creighton, “Anomalous intense Raman spectra of pyridine at a silver electrode,” *J.*

- Am. Chem. Soc.*, vol. 99, no. 15, pp. 5215–5217, Jun. 1977, doi: 10.1021/ja00457a071.
- [22] R. Goel, S. Chakraborty, V. Awasthi, V. Bhardwaj, and S. Kumar Dubey, “Exploring the various aspects of Surface enhanced Raman spectroscopy (SERS) with focus on the recent progress: SERS-active substrate, SERS-instrumentation, SERS-application,” *Sensors Actuators A Phys.*, vol. 376, p. 115555, 2024, doi: <https://doi.org/10.1016/j.sna.2024.115555>.
- [23] M. Moskovits, “Surface roughness and the enhanced intensity of Raman scattering by molecules adsorbed on metals,” *J. Chem. Phys.*, vol. 69, pp. 4159–4161, 1978, [Online]. Available: <https://api.semanticscholar.org/CorpusID:96650879>
- [24] C. Andreou, S. A. Kishore, and M. F. Kircher, “Surface-enhanced Raman spectroscopy: A new modality for cancer imaging,” *J. Nucl. Med.*, vol. 56, no. 9, pp. 1295–1299, 2015, doi: 10.2967/jnumed.115.158196.
- [25] R. R. Jones, D. C. Hooper, L. Zhang, D. Wolverson, and V. K. Valev, “Raman Techniques: Fundamentals and Frontiers,” *Nanoscale Res. Lett.*, vol. 14, no. 1, p. 231, 2019, doi: 10.1186/s11671-019-3039-2.
- [26] R. Panneerselvam *et al.*, “Surface-enhanced Raman spectroscopy: Bottlenecks and future directions,” *Chem. Commun.*, vol. 54, no. 1, pp. 10–25, 2017, doi: 10.1039/c7cc05979e.
- [27] C. Zong *et al.*, “Surface-Enhanced Raman Spectroscopy for Bioanalysis: Reliability and Challenges,” *Chem. Rev.*, vol. 118, no. 10, pp. 4946–4980, 2018, doi: 10.1021/acs.chemrev.7b00668.
- [28] K. M. Kosuda, J. M. Bingham, K. L. Wustholz, R. P. Van Duyne, and R. J. Groarke, *Nanostructures and surface-enhanced raman spectroscopy*, vol. 1–5, no. August 2015. Elsevier Ltd., 2019. doi: 10.1016/B978-0-12-803581-8.00611-1.
- [29] H. Ma *et al.*, “Surface-Enhanced Raman Spectroscopy: Current Understanding, Challenges, and Opportunities,” *ACS Nano*, vol. 18, no. 22, pp. 14000–14019, Jun. 2024, doi: 10.1021/acsnano.4c02670.
- [30] V. Amendola, R. Pilot, M. Frascioni, O. M. Maragò, and M. A. Iati, “Surface plasmon resonance in gold nanoparticles: A review,” *J. Phys. Condens. Matter*, vol. 29, no. 20, 2017, doi: 10.1088/1361-648X/aa60f3.
- [31] Y. Wu, Y. Wang, T. Mo, and Q. Liu, “Surface-enhanced Raman scattering-based strategies for tumor markers detection: A review,” *Talanta*, vol. 280, no. April, p. 126717, 2024, doi: 10.1016/j.talanta.2024.126717.
- [32] S. Unser, I. Bruzas, J. He, and L. Sagle, “Localized surface plasmon resonance biosensing: Current challenges and approaches,” *Sensors (Switzerland)*, vol. 15, no. 7, pp. 15684–15716, 2015, doi: 10.3390/s150715684.
- [33] B. P. Majee and A. K. Mishra, “Fundamentals and Applications of Surface Enhanced Raman Spectroscopy,” in *Modern Techniques of Spectroscopy: Basics, Instrumentation, and Applications*, D. K. Singh, M. Pradhan, and A. Materny, Eds. Singapore: Springer Singapore, 2021, pp. 185–208. doi: 10.1007/978-981-33-6084-6_7.
- [34] Y. J. Liu *et al.*, “SERS probes for cancer detection: Innovations and applications,” *TrAC - Trends Anal. Chem.*, vol. 193, no. September, p. 118467, 2025, doi: 10.1016/j.trac.2025.118467.
- [35] S. Y. Ding *et al.*, “Nanostructure-based plasmon-enhanced Raman spectroscopy for surface analysis of materials,” *Nat. Rev. Mater.*, vol. 1, no. 6, 2016, doi: 10.1038/natrevmats.2016.21.
- [36] E. Petryayeva and U. J. Krull, “Localized surface plasmon resonance: Nanostructures, bioassays and biosensing-A review,” *Anal. Chim. Acta*, vol. 706, no. 1, pp. 8–24, 2011, doi: 10.1016/j.aca.2011.08.020.
- [37] J. Langer *et al.*, “Present and Future of Surface-Enhanced Raman Scattering,” *ACS Nano*, vol. 14, no. 1, pp. 28–117, Jan. 2020, doi: 10.1021/acsnano.9b04224.
- [38] Z. Huang, J. Peng, L. Xu, and P. Liu, “Development and Application of Surface-Enhanced Raman Scattering (SERS),” *Nanomaterials*, vol. 14, no. 17, 2024, doi: 10.3390/nano14171417.
- [39] C. Ghanashyam, R. K. Sinha, and A. Bankapur, “Surface-Plasmon-Polaritons for Reversible Assembly of Gold Nanoparticles, In Situ Nanogap Tuning, and SERS,” *Small Methods*, vol. 8, no. 1, pp. 1–11, 2024, doi: 10.1002/smt.202301086.
- [40] G. Hajisalem, M. S. Nezami, and R. Gordon, “Probing the quantum tunneling limit of plasmonic enhancement by third harmonic generation,” *Nano Lett.*, vol. 14, no. 11, pp. 6651–6654, 2014, doi: 10.1021/nl503324g.
- [41] X. Wang, S. C. Huang, S. Hu, S. Yan, and B. Ren, “Fundamental understanding and applications of plasmon-enhanced Raman spectroscopy,” *Nat. Rev. Phys.*, vol. 2, no. 5, pp. 253–271, 2020, doi: 10.1038/s42254-020-0171-y.
- [42] P. G. Etchegoin and E. C. Le Ru, “A perspective on single molecule SERS: Current status and future challenges,” *Phys. Chem. Chem. Phys.*, vol. 10, no. 40, pp. 6079–6089, 2008, doi: 10.1039/b809196j.
- [43] E. C. Le Ru, P. G. Etchegoin, and M. Meyer, “Enhancement factor distribution around a single surface-enhanced Raman scattering hot spot and its relation to single molecule detection,” *J. Chem. Phys.*, vol. 125, no. 20, pp. 1–14, 2006, doi: 10.1063/1.2390694.
- [44] J. I. Gersten, R. L. Birke, and J. I. Lombardi, “Theory of Enhance I Light Scattering from Molecules Adsorbed at the Metal-Solution Interface,” *Phys. Rev. Lett.*, vol. 43, no. 2, pp. 147–150, 1979, doi: 10.1103/PhysRevLett.43.147.
- [45] L. Xia *et al.*, “Visualized method of chemical enhancement mechanism on SERS and TERS,” *J. Raman Spectrosc.*, vol. 45, no. 7, pp. 533–540, 2014, doi: 10.1002/jrs.4504.
- [46] L. Guerrini and D. Graham, “Molecularly-mediated assemblies of plasmonic nanoparticles for Surface-Enhanced

- Raman Spectroscopy applications,” *Chem. Soc. Rev.*, vol. 41, no. 21, pp. 7085–7107, 2012, doi: 10.1039/c2cs35118h.
- [47] C. Lin *et al.*, “Recent development of surface-enhanced Raman scattering for biosensing,” *J. Nanobiotechnology*, vol. 21, no. 1, pp. 1–37, 2023, doi: 10.1186/s12951-023-01890-7.
- [48] N. Chauhan, K. Saxena, R. Rawal, L. Yadav, and U. Jain, “Advances in surface-enhanced Raman spectroscopy-based sensors for detection of various biomarkers,” *Prog. Biophys. Mol. Biol.*, vol. 184, no. August, pp. 32–41, 2023, doi: 10.1016/j.pbiomolbio.2023.08.003.
- [49] P. Slepíčka, N. S. Kasálová, J. Siegel, Z. Kolská, and V. Švorčík, “Methods of gold and silver nanoparticles preparation,” *Materials (Basel)*, vol. 13, no. 1, 2020, doi: 10.3390/ma13010001.
- [50] A. Shiohara, Y. Wang, and L. M. Liz-Marzán, “Recent approaches toward creation of hot spots for SERS detection,” *J. Photochem. Photobiol. C Photochem. Rev.*, vol. 21, pp. 2–25, 2014, doi: 10.1016/j.jphotochemrev.2014.09.001.
- [51] P. P. P. Kumar, S. Kaushal, and D. K. Lim, “Recent advances in nano/microfabricated substrate platforms and artificial intelligence for practical surface-enhanced Raman scattering-based bioanalysis,” *TrAC - Trends Anal. Chem.*, vol. 168, no. September, p. 117341, 2023, doi: 10.1016/j.trac.2023.117341.
- [52] D. Cialla-May *et al.*, “Biomedical SERS - the current state and future trends,” *Chem. Soc. Rev.*, vol. 53, no. 18, 2024, doi: 10.1039/d4cs00090k.
- [53] K. Srivastava, H. Le-The, J. J. A. Lozeman, A. van den Berg, W. van der Stam, and M. Odijk, “Prospects of nano-lithographic tools for the fabrication of surface-enhanced Raman spectroscopy (SERS) substrates,” *Micro Nano Eng.*, vol. 23, no. May, p. 100267, 2024, doi: 10.1016/j.mne.2024.100267.
- [54] R. Chen *et al.*, “Micro-/nanostructures for surface-enhanced Raman spectroscopy: Recent advances and perspectives,” *Adv. Colloid Interface Sci.*, vol. 331, no. April, p. 103235, 2024, doi: 10.1016/j.cis.2024.103235.
- [55] R. Pilot, R. Signorini, C. Durante, L. Orian, M. Bhamidipati, and L. Fabris, “A review on surface-enhanced Raman scattering,” *Biosensors*, vol. 9, no. 2, 2019, doi: 10.3390/bios9020057.
- [56] A. Sultangaziyev, A. Ilyas, A. Dyussupova, and R. Bukasov, “Trends in Application of SERS Substrates beyond Ag and Au, and Their Role in Bioanalysis,” *Biosensors*, vol. 12, no. 11, 2022, doi: 10.3390/bios12110967.
- [57] P. G. Jamkhane, N. W. Ghule, A. H. Bamer, and M. G. Kalaskar, “Metal nanoparticles synthesis: An overview on methods of preparation, advantages and disadvantages, and applications,” *J. Drug Deliv. Sci. Technol.*, vol. 53, no. July, p. 101174, 2019, doi: 10.1016/j.jddst.2019.101174.
- [58] D. Barman, L. Saikia, and B. K. Sarma, “Wide range tunability of surface plasmon resonance of nanostructured Ag deposited by off-axis magnetron sputtering,” *Mater. Today Proc.*, vol. 65, pp. 2581–2585, 2022, doi: 10.1016/j.matpr.2022.04.789.
- [59] G. M. Das, S. Managò, M. Mangini, and A. C. De Luca, “Biosensing using sers active gold nanostructures,” *Nanomaterials*, vol. 11, no. 10, 2021, doi: 10.3390/nano11102679.
- [60] N. John and A. T.M, “New trends in gold nanostructure-based SERS substrate: From fundamental to biomedical applications,” *Vib. Spectrosc.*, vol. 124, no. November 2022, p. 103477, 2023, doi: 10.1016/j.vibspec.2022.103477.
- [61] J. Karlo, S. S. Razi, M. S. Phaneeswar, A. Vijay, and S. P. Singh, “Metallic Nanoparticles for Surface-Enhanced Raman Scattering Based Biosensing Applications,” *Photochem*, vol. 4, no. 4, pp. 417–433, 2024, doi: 10.3390/photochem4040026.
- [62] Á. I. López-Lorente, “Recent developments on gold nanostructures for surface enhanced Raman spectroscopy: Particle shape, substrates and analytical applications. A review,” *Anal. Chim. Acta*, vol. 1168, 2021, doi: 10.1016/j.aca.2021.338474.
- [63] S. Chen, M. Lv, J. Fan, Y. Huang, G. Liang, and S. Zhang, “Bioorthogonal surface-enhanced Raman scattering flower-like nanoprobe with embedded standards for accurate cancer cell imaging,” *Anal. Chim. Acta*, vol. 1246, no. July 2022, p. 340895, 2023, doi: 10.1016/j.aca.2023.340895.
- [64] O. Liba, E. D. Sorelle, D. Sen, and A. De La Zerda, “Contrast-enhanced optical coherence tomography with picomolar sensitivity for functional in vivo imaging,” *Sci. Rep.*, vol. 6, no. March, pp. 1–12, 2016, doi: 10.1038/srep23337.
- [65] Y. Huang, A. R. Ferhan, Y. Gao, A. Dandapat, and D. H. Kim, “High-yield synthesis of triangular gold nanoplates with improved shape uniformity, tunable edge length and thickness,” *Nanoscale*, vol. 6, no. 12, pp. 6496–6500, 2014, doi: 10.1039/c4nr00834k.
- [66] Y. C. Ou *et al.*, “Diagnosis of immunomarkers: In vivo via multiplexed surface enhanced Raman spectroscopy with gold nanostars,” *Nanoscale*, vol. 10, no. 27, pp. 13092–13105, 2018, doi: 10.1039/c8nr01478g.
- [67] S. E. Skrabalak *et al.*, “Gold Nanocages: Synthesis, Properties, and Applications,” *Acc. Chem. Res.*, vol. 41, no. 12, pp. 1587–1595, Dec. 2008, doi: 10.1021/ar800018v.
- [68] T. H. Chow, Y. Lai, X. Cui, W. Lu, X. Zhuo, and J. Wang, “Colloidal Gold Nanorings and Their Plasmon Coupling with Gold Nanospheres,” *Small*, vol. 15, no. 35, pp. 1–12, 2019, doi: 10.1002/sml.201902608.
- [69] J. S. Wi, J. Park, H. Kang, D. Jung, S. W. Lee, and T. G. Lee, “Stacked Gold Nanodisks for Bimodal Photoacoustic

- and Optical Coherence Imaging,” *ACS Nano*, vol. 11, no. 6, pp. 6225–6232, 2017, doi: 10.1021/acsnano.7b02337.
- [70] D. Chateau *et al.*, “From gold nanobipyramids to nanojavelins for a precise tuning of the plasmon resonance to the infrared wavelengths: Experimental and theoretical aspects,” *Nanoscale*, vol. 7, no. 5, pp. 1934–1943, 2015, doi: 10.1039/c4nr06323f.
- [71] F. Tian, F. Bonnier, A. Casey, A. E. Shanahan, and H. J. Byrne, “Surface enhanced Raman scattering with gold nanoparticles: effect of particle shape,” *Anal. Methods*, vol. 6, no. 22, pp. 9116–9123, 2014, doi: 10.1039/C4AY02112F.
- [72] G. Awiaz, J. Lin, and A. Wu, “Recent advances of Au@Ag core-shell SERS-based biosensors,” *Exploration*, vol. 3, no. 1, 2023, doi: 10.1002/EXP.20220072.
- [73] L. H. dos Santos, F. D. Kiss, A. C. Ferraz, and R. Miotto, “Ag@Au and Au@Ag Core-Shell Nanoparticles: Insights on the Growth Process and Interactions with Small Molecules,” *J. Phys. Chem. C*, vol. 128, no. 25, pp. 10751–10760, 2024, doi: 10.1021/acs.jpcc.4c03044.
- [74] L. Wang, X. Gao, J. Han, J. Yuan, Z. Wang, and X. Ma, “Core-shell SERS nanoprobe for ROS detection and imaging during apoptosis,” *Food Biosci.*, vol. 67, no. February, p. 106339, 2025, doi: 10.1016/j.fbio.2025.106339.
- [75] A. Yilmaz and M. Yilmaz, “Bimetallic core-shell nanoparticles of gold and silver via bioinspired polydopamine layer as surface-enhanced Raman spectroscopy (SERS) platform,” *Nanomaterials*, vol. 10, no. 4, 2020, doi: 10.3390/nano10040688.
- [76] J. I. García-Ramírez *et al.*, “SERS-Based Immunoassay for α -Fetoprotein Biomarker Detection Using an Au-Ag Nanostars Platform,” *Biosensors*, vol. 15, no. 9, pp. 1–13, 2025, doi: 10.3390/bios15090632.
- [77] Y. Wang *et al.*, “Dynamically controllable hot spots in DNA-derived hydrogel scaffold SERS substrate for exosome recognition using DNA self-assembly amplification,” *Chem. Eng. J.*, vol. 496, no. April, p. 154270, 2024, doi: 10.1016/j.cej.2024.154270.
- [78] D. Cialla *et al.*, “Surface-enhanced Raman spectroscopy (SERS): progress and trends,” *Anal. Bioanal. Chem.*, vol. 403, no. 1, pp. 27–54, Apr. 2012, doi: 10.1007/s00216-011-5631-x.
- [79] J. Peng, Y. Song, Y. Lin, and Z. Huang, “Introduction and Development of Surface-Enhanced Raman Scattering (SERS) Substrates: A Review,” *Nanomaterials*, vol. 14, no. 20, 2024, doi: 10.3390/nano14201648.
- [80] P. A. Mosier-Boss, “Review of SERS substrates for chemical sensing,” *Nanomaterials*, vol. 7, no. 6, 2017, doi: 10.3390/nano7060142.
- [81] O. Péron, E. Rinnert, M. Lehaitre, P. Crassous, and C. Compère, “Detection of polycyclic aromatic hydrocarbon (PAH) compounds in artificial sea-water using surface-enhanced Raman scattering (SERS),” *Talanta*, vol. 79, no. 2, pp. 199–204, 2009, doi: 10.1016/j.talanta.2009.03.043.
- [82] O. Guselnikova *et al.*, “New Trends in Nanoarchitected SERS Substrates: Nanospaces, 2D Materials, and Organic Heterostructures,” *Small*, vol. 18, no. 25, 2022, doi: 10.1002/sml.202107182.
- [83] M. Pisco *et al.*, “Nanosphere lithography for optical fiber tip nanoprobe,” *Light Sci. Appl.*, vol. 6, no. 5, pp. 1–12, 2017, doi: 10.1038/lsa.2016.229.
- [84] S. Managò *et al.*, “Tailoring lab-on-fiber SERS optrodes towards biological targets of different sizes,” *Sensors Actuators, B Chem.*, vol. 339, no. July 2020, 2021, doi: 10.1016/j.snb.2020.129321.
- [85] L. Štolcová, V. Peksa, J. Proška, and M. Procházka, “Gold film over very small (107 nm) spheres as efficient substrate for sensitive and reproducible surface-enhanced Raman scattering (SERS) detection of biologically important molecules,” *J. Raman Spectrosc.*, vol. 49, no. 3, pp. 499–505, 2018, doi: 10.1002/jrs.5317.
- [86] P. Colson, C. Henrist, and R. Cloots, “Nanosphere lithography: A powerful method for the controlled manufacturing of nanomaterials,” *J. Nanomater.*, vol. 2013, 2013, doi: 10.1155/2013/948510.
- [87] M. Domonkos and A. Kromka, “Nanosphere Lithography-Based Fabrication of Spherical Nanostructures and Verification of Their Hexagonal Symmetries by Image Analysis,” *Symmetry (Basel)*, vol. 14, no. 12, 2022, doi: 10.3390/sym14122642.
- [88] C. L. Haynes and R. P. Van Duyne, “Nanosphere Lithography: A Versatile Nanofabrication Tool for Studies of Size-Dependent Nanoparticle Optics,” *J. Phys. Chem. B*, vol. 105, no. 24, pp. 5599–5611, Jun. 2001, doi: 10.1021/jp010657m.
- [89] H. M. Jin *et al.*, “Ultralarge Area Sub-10 nm Plasmonic Nanogap Array by Block Copolymer Self-Assembly for Reliable High-Sensitivity SERS,” *ACS Appl. Mater. Interfaces*, vol. 10, no. 51, pp. 44660–44667, 2018, doi: 10.1021/acsami.8b17325.
- [90] W. Yue *et al.*, “Electron-beam lithography of gold nanostructures for surface-enhanced Raman scattering,” *J. Micromechanics Microengineering*, vol. 22, no. 12, 2012, doi: 10.1088/0960-1317/22/12/125007.
- [91] N. A. Cinel, S. Cakmakyan, S. Butun, G. Ertas, and E. Ozbay, “E-Beam lithography designed substrates for surface enhanced Raman spectroscopy,” *Photonics Nanostructures - Fundam. Appl.*, vol. 15, pp. 109–115, 2015, doi: 10.1016/j.photonics.2014.11.003.
- [92] M. Fan, G. F. S. Andrade, and A. G. Brolo, “A review on the fabrication of substrates for surface enhanced Raman spectroscopy and their applications in analytical chemistry,” *Anal. Chim. Acta*, vol. 693, no. 1–2, pp. 7–25, 2011, doi: 10.1016/j.aca.2011.03.002.

- [93] H. Shahali, J. Hasan, H. Wang, T. Tesfamichael, C. Yan, and P. K. D. V. Yarlagadda, "Evaluation of particle beam lithography for fabrication of metallic nano-structures," *Procedia Manuf.*, vol. 30, pp. 261–267, 2019, doi: 10.1016/j.promfg.2019.02.038.
- [94] S. E. Mowbray and A. M. Amiri, "A brief overview of medical fiber optic biosensors and techniques in the modification for enhanced sensing ability," *Diagnostics*, vol. 9, no. 1, 2019, doi: 10.3390/diagnostics9010023.
- [95] C. Leitão *et al.*, "Cost-Effective Fiber Optic Solutions for Biosensing," *Biosensors*, vol. 12, no. 8, 2022, doi: 10.3390/bios12080575.
- [96] K. Bartnik, M. Koba, and M. Śmietana, "Advancements in optical fiber sensors for in vivo applications – A review of sensors tested on living organisms," *Meas. J. Int. Meas. Confed.*, vol. 224, no. May 2023, 2024, doi: 10.1016/j.measurement.2023.113818.
- [97] J. Zhang, S. Chen, T. Gong, X. Zhang, and Y. Zhu, "Tapered Fiber Probe Modified by Ag Nanoparticles for SERS Detection," *Plasmonics*, vol. 11, no. 3, pp. 743–751, 2016, doi: 10.1007/s11468-015-0105-1.
- [98] O. Bratash, A. Buhot, L. Leroy, and E. Engel, "Optical fiber biosensors toward in vivo detection.," *Biosens. Bioelectron.*, vol. 251, no. February, p. 116088, 2024, doi: 10.1016/j.bios.2024.116088.
- [99] R. K. Gangwar *et al.*, "Optical fiber SERS sensors: Unveiling advances, challenges, and applications in a miniaturized technology," *Coord. Chem. Rev.*, vol. 510, no. April, p. 215861, 2024, doi: 10.1016/j.ccr.2024.215861.
- [100] S. Spaziani *et al.*, "SERS assisted sandwich immunoassay platforms for ultrasensitive and selective detection of human Thyroglobulin," *Biosens. Bioelectron.*, vol. 233, no. December 2022, pp. 0–7, 2023, doi: 10.1016/j.bios.2023.115322.
- [101] Z. Xia *et al.*, "Giant Enhancement of Raman Scattering by a Hollow-Core Microstructured Optical Fiber Allows Single Exosome Probing," *ACS Sensors*, vol. 8, no. 4, pp. 1799–1809, 2023, doi: 10.1021/acssensors.3c00131.
- [102] J. Wang, Y. Geng, Y. Shen, W. Shi, W. Xu, and S. Xu, "SERS-active fiber tip for intracellular and extracellular pH sensing in living single cells," *Sensors Actuators, B Chem.*, vol. 290, no. April, pp. 527–534, 2019, doi: 10.1016/j.snb.2019.03.149.
- [103] E. Garai *et al.*, "A real-time clinical endoscopic system for intraluminal, multiplexed imaging of surface-enhanced Raman scattering nanoparticles," *PLoS One*, vol. 10, no. 4, pp. 1–16, 2015, doi: 10.1371/journal.pone.0123185.
- [104] S. Jeong *et al.*, "Fluorescence-Raman Dual Modal Endoscopic System for Multiplexed Molecular Diagnostics," *Sci. Rep.*, vol. 5, pp. 13–16, 2015, doi: 10.1038/srep09455.
- [105] X. Wang, E. Zhang, H. Shi, Y. Tao, and X. Ren, "Semiconductor-based surface enhanced Raman scattering (SERS): from active materials to performance improvement," *Analyst*, vol. 147, no. 7, pp. 1257–1272, 2022, doi: 10.1039/d1an02165f.
- [106] H. Sun, S. Cong, Z. Zheng, Z. Wang, Z. Chen, and Z. Zhao, "Metal-Organic Frameworks as Surface Enhanced Raman Scattering Substrates with High Tailorability," *J. Am. Chem. Soc.*, vol. 141, no. 2, pp. 870–878, 2019, doi: 10.1021/jacs.8b09414.
- [107] B. Yang *et al.*, "Recent Development of SERS Technology: Semiconductor-Based Study," *ACS Omega*, vol. 4, no. 23, pp. 20101–20108, 2019, doi: 10.1021/acsomega.9b03154.
- [108] Y. Xu *et al.*, "A TiO₂-based bioprobe enabling excellent SERS activity in the detection of diverse circulating tumor cells," *J. Mater. Chem. B*, vol. 10, no. 20, pp. 3808–3816, 2022, doi: 10.1039/d2tb00464j.
- [109] R. Haldavnekar, K. Venkatakrishnan, and B. Tan, "Non plasmonic semiconductor quantum SERS probe as a pathway for in vitro cancer detection," *Nat. Commun.*, vol. 9, no. 1, pp. 1–18, 2018, doi: 10.1038/s41467-018-05237-x.
- [110] N. Cheng, J. Fu, D. Chen, S. Chen, and H. Wang, "An antibody-free liver cancer screening approach based on nanoplasmonics biosensing chips via spectrum-based deep learning," *NanoImpact*, vol. 21, p. 100296, 2021, doi: https://doi.org/10.1016/j.impact.2021.100296.
- [111] S. Lee, S. H. Lee, B. Paulson, J. C. Lee, and J. K. Kim, "Enhancement of local surface plasmon resonance (LSPR) effect by biocompatible metal clustering based on ZnO nanorods in Raman measurements," *Spectrochim. Acta - Part A Mol. Biomol. Spectrosc.*, vol. 204, pp. 203–208, 2018, doi: 10.1016/j.saa.2018.06.045.
- [112] Y. Zhang, Y. R. Zhen, O. Neumann, J. K. Day, P. Nordlander, and N. J. Halas, "Coherent anti-Stokes Raman scattering with single-molecule sensitivity using a plasmonic Fano resonance," *Nat. Commun.*, vol. 5, pp. 1–7, 2014, doi: 10.1038/ncomms5424.
- [113] L. Fabris, "SERS Tags: The Next Promising Tool for Personalized Cancer Detection?," *ChemNanoMat*, vol. 2, no. 4, pp. 249–258, 2016, doi: 10.1002/cnma.201500221.
- [114] K. Yuan, B. Jurado-Sánchez, and A. Escarpa, "Nanomaterials meet surface-enhanced Raman scattering towards enhanced clinical diagnosis: a review," *J. Nanobiotechnology*, vol. 20, no. 1, pp. 1–28, 2022, doi: 10.1186/s12951-022-01711-3.
- [115] L. Guerrini, N. Pazos-Perez, E. Garcia-Rico, and R. Alvarez-Puebla, "Cancer characterization and diagnosis with SERS-encoded particles," *Cancer Nanotechnol.*, vol. 8, 2017, doi: 10.1186/s12645-017-0031-3.
- [116] L. Vázquez-Iglesias *et al.*, "SERS sensing for cancer biomarker: Approaches and directions," *Bioact. Mater.*, vol.

- 34, no. September 2023, pp. 248–268, 2024, doi: 10.1016/j.bioactmat.2023.12.018.
- [117] M. Constantinou, K. Hadjigeorgiou, and S. Abalde-Cela, “Label-Free Sensing with Metal Nanostructure-Based Surface-Enhanced Raman Spectroscopy for Cancer Diagnosis,” *ACS Appl. Nano Mater.*, vol. 5, no. 9, pp. 12276–12299, 2022, doi: 10.1021/acsanm.2c02392.
 - [118] M. A. Tahir, N. E. Dina, H. Cheng, V. K. Valev, and L. Zhang, “Surface-enhanced Raman spectroscopy for bioanalysis and diagnosis,” *Nanoscale*, vol. 13, no. 27, pp. 11593–11634, 2021, doi: 10.1039/d1nr00708d.
 - [119] Y. Li, Z. Wang, X. Mu, A. Ma, and S. Guo, “Raman tags: Novel optical probes for intracellular sensing and imaging,” *Biotechnol. Adv.*, vol. 35, no. 2, pp. 168–177, 2017, doi: 10.1016/j.biotechadv.2016.12.004.
 - [120] A. Verdin, C. Malherbe, and G. Eppe, “Designing SERS nanotags for profiling overexpressed surface markers on single cancer cells: A review,” *Talanta*, vol. 276, no. March, p. 126225, 2024, doi: 10.1016/j.talanta.2024.126225.
 - [121] L. Lin, X. Bi, Y. Gu, F. Wang, and J. Ye, “Surface-enhanced Raman scattering nanotags for bioimaging,” *J. Appl. Phys.*, vol. 129, no. 19, 2021, doi: 10.1063/5.0047578.
 - [122] M. Usman *et al.*, “Recent advances in surface enhanced Raman spectroscopy for bacterial pathogen identifications,” *J. Adv. Res.*, vol. 51, pp. 91–107, 2023, doi: 10.1016/j.jare.2022.11.010.
 - [123] X. Liu, X. Liu, P. Rong, and D. Liu, “Recent advances in background-free Raman scattering for bioanalysis,” *TrAC - Trends Anal. Chem.*, vol. 123, p. 115765, 2020, doi: 10.1016/j.trac.2019.115765.
 - [124] G. McNay, D. Eustace, W. E. Smith, K. Faulds, and D. Graham, “Surface-enhanced Raman scattering (SERS) and surface-enhanced resonance raman scattering (SERRS): A review of applications,” *Appl. Spectrosc.*, vol. 65, no. 8, pp. 825–837, 2011, doi: 10.1366/11-06365.
 - [125] O. E. Eremina, S. Schaefer, A. T. Czaja, S. Awad, M. A. Lim, and C. Zavaleta, “Multiplexing potential of NIR resonant and non-resonant Raman reporters for bio-imaging applications,” *Analyst*, vol. 148, no. 23, pp. 5915–5925, 2023, doi: 10.1039/d3an01298k.
 - [126] S. Harmsen, M. A. Wall, R. Huang, and M. F. Kircher, “Cancer imaging using surface-enhanced resonance Raman scattering nanoparticles,” *Nat. Protoc.*, vol. 12, no. 7, pp. 1400–1414, Jul. 2017, doi: 10.1038/nprot.2017.031.
 - [127] C. Qi *et al.*, “Nonfluorescent Near-Infrared Surface-Enhanced Resonance Raman Nanoprobes with Ultrahigh Brightness and Synergistic Photothermal Effect,” *ACS Appl. Mater. Interfaces*, vol. 16, no. 49, pp. 67333–67343, Dec. 2024, doi: 10.1021/acsami.4c15017.
 - [128] D. Cialla-May *et al.*, “Recent advances of surface enhanced Raman spectroscopy (SERS) in optical biosensing,” *TrAC - Trends Anal. Chem.*, vol. 181, no. PA, p. 117990, 2024, doi: 10.1016/j.trac.2024.117990.
 - [129] Z. Wang, S. Zong, L. Wu, D. Zhu, and Y. Cui, “SERS-Activated Platforms for Immunoassay: Probes, Encoding Methods, and Applications,” *Chem. Rev.*, vol. 117, no. 12, pp. 7910–7963, 2017, doi: 10.1021/acs.chemrev.7b00027.
 - [130] C. Zhang *et al.*, “Machine learning assisted dual-modal SERS detection for circulating tumor cells,” *Biosens. Bioelectron.*, vol. 268, no. February 2024, p. 116897, 2025, doi: 10.1016/j.bios.2024.116897.
 - [131] L. Fabris, “Gold-based SERS tags for biomedical imaging,” *J. Opt. (United Kingdom)*, vol. 17, no. 11, p. 114002, 2015, doi: 10.1088/2040-8978/17/11/114002.
 - [132] X. Wu *et al.*, “Improved SERS nanoparticles for direct detection of circulating tumor cells in the blood,” *ACS Appl. Mater. Interfaces*, vol. 7, no. 18, pp. 9965–9971, 2015, doi: 10.1021/acsami.5b02276.
 - [133] M. He *et al.*, “Octahedral silver oxide nanoparticles enabling remarkable SERS activity for detecting circulating tumor cells,” *Sci. China Life Sci.*, vol. 65, no. 3, pp. 561–571, 2022, doi: 10.1007/s11427-020-1931-9.
 - [134] M. Li, S. R. Banerjee, C. Zheng, M. G. Pomper, and I. Barman, “Ultrahigh affinity Raman probe for targeted live cell imaging of prostate cancer,” *Chem. Sci.*, vol. 7, no. 11, pp. 6779–6785, 2016, doi: 10.1039/c6sc01739h.
 - [135] D. Rist *et al.*, “Cancer Cell Targeting, Magnetic Sorting, and SERS Detection through Cell Surface Receptors,” *ACS Sensors*, vol. 8, no. 12, pp. 4636–4645, Dec. 2023, doi: 10.1021/acssensors.3c01625.
 - [136] C. D. L. De Albuquerque and Z. D. Schultz, “Super-resolution SERS Imaging of Single Particles in Cells,” *Anal. Chem.*, vol. 92, no. 13, pp. 9389–9398, 2020, doi: 10.1021/acs.analchem.0c01864.Super-resolution.
 - [137] Q. Zhang, J. Li, P. Tang, X. Lu, J. Tian, and L. Zhong, “Dynamic imaging of transferrin receptor molecules on single live cell with bridge gaps-enhanced raman tags,” *Nanomaterials*, vol. 9, no. 10, 2019, doi: 10.3390/nano9101373.
 - [138] G. Geka *et al.*, “SERS Immunosensors for Cancer Markers Detection,” *Materials (Basel)*, vol. 16, no. 10, 2023, doi: 10.3390/ma16103733.
 - [139] N. Lyu, A. Hassanzadeh-Barforoushi, L. M. Rey Gomez, W. Zhang, and Y. Wang, “SERS biosensors for liquid biopsy towards cancer diagnosis by detection of various circulating biomarkers: current progress and perspectives,” *Nano Converg.*, vol. 11, no. 1, 2024, doi: 10.1186/s40580-024-00428-3.
 - [140] N. S. Greenspan, “Epitopes, paratopes, and other topes 30 years on: Understanding what we are talking about,” *Hum. Immunol.*, vol. 84, no. 9, pp. 429–438, 2023, doi: 10.1016/j.humimm.2023.06.006.
 - [141] M. H. Jazayeri, H. Amani, A. A. Pourfatollah, H. Pazoki-Toroudi, and B. Sedighimoghaddam, “Various methods of gold nanoparticles (GNPs) conjugation to antibodies,” *Sens. Bio-Sensing Res.*, vol. 9, pp. 17–22, 2016, doi: 10.1016/j.sbsr.2016.04.002.

- [142] H. J. Ryu, W. K. Lee, Y. H. Kim, and J. S. Lee, "Interfacial interactions of SERS-active noble metal nanostructures with functional ligands for diagnostic analysis of protein cancer markers," *Microchim. Acta*, vol. 188, no. 5, 2021, doi: 10.1007/s00604-021-04807-z.
- [143] N. Choi, Y. Zhang, Y. Wang, and S. Schlücker, "iSERS: from nanotag design to protein assays and ex vivo imaging," *Chem. Soc. Rev.*, vol. 53, no. 13, pp. 6675–6693, 2024, doi: 10.1039/d3cs01060k.
- [144] J. P. Oliveira, A. R. Prado, W. J. Keijok, P. W. P. Antunes, E. R. Yapuchura, and M. C. C. Guimarães, "Impact of conjugation strategies for targeting of antibodies in gold nanoparticles for ultrasensitive detection of 17 β -estradiol," *Sci. Rep.*, vol. 9, no. 1, pp. 1–8, 2019, doi: 10.1038/s41598-019-50424-5.
- [145] S. Kumar, J. Aaron, and K. Sokolov, "Directional conjugation of antibodies to nanoparticles for synthesis of multiplexed optical contrast agents with both delivery and targeting moieties," *Nat. Protoc.*, vol. 3, no. 2, pp. 314–320, 2008, doi: 10.1038/nprot.2008.1.
- [146] S. S. Farid, "Process economics of industrial monoclonal antibody manufacture," *J. Chromatogr. B Anal. Technol. Biomed. Life Sci.*, vol. 848, no. 1, pp. 8–18, 2007, doi: 10.1016/j.jchromb.2006.07.037.
- [147] M. Sheraz, X. F. Sun, Y. Wang, J. Chen, and L. Sun, "Recent Developments in Aptamer-Based Sensors for Diagnostics," *Sensors*, vol. 24, no. 23, 2024, doi: 10.3390/s24237432.
- [148] K. Rahimizadeh *et al.*, "Development of cell-specific aptamers: Recent advances and insight into the selection procedures," *Molecules*, vol. 22, no. 12, pp. 1–13, 2017, doi: 10.3390/molecules22122070.
- [149] M. Muhammad and Q. Huang, "A review of aptamer-based SERS biosensors: Design strategies and applications," *Talanta*, vol. 227, no. February, p. 122188, 2021, doi: 10.1016/j.talanta.2021.122188.
- [150] W. Lv *et al.*, "Efficient detection of single circulating tumor cell in blood using Raman mapping based on Aptamer-SERS bio-probe coupled with micropore membrane filtration," *Talanta*, vol. 267, no. September 2023, p. 125220, 2024, doi: 10.1016/j.talanta.2023.125220.
- [151] C. Dong *et al.*, "DNA walker-powered SERS logic aptasensors for accurate imaging and identification of tumor cells," *Sensors Actuators B Chem.*, vol. 401, no. October 2023, p. 134922, 2024, doi: 10.1016/j.snb.2023.134922.
- [152] F. Canfarotta, A. Cecchini, and S. Piletsky, "Nano-sized Molecularly Imprinted Polymers as Artificial Antibodies," in *Molecularly Imprinted Polymers for Analytical Chemistry Applications*, W. Kutner and P. S. Sharma, Eds. The Royal Society of Chemistry, 2018, p. 0. doi: 10.1039/9781788010474-00001.
- [153] B. Gao, Q. Niu, and R. Du, "Preparation and recognition performance of cytosine alkaloid-imprinted material prepared using novel surface molecular imprinting technique," *J. Sep. Sci.*, vol. 33, no. 9, pp. 1338–1348, 2010, doi: 10.1002/jssc.200900762.
- [154] D. Yin *et al.*, "Surface-enhanced Raman scattering imaging of cancer cells and tissues via sialic acid-imprinted nanotags," *Chem. Commun.*, vol. 51, no. 100, pp. 17696–17699, 2015, doi: 10.1039/c5cc05174f.
- [155] Y. Lu, Q. Liu, B. Fu, P. Li, and W. Xu, "Label-free MIP-SERS biosensor for sensitive detection of colorectal cancer biomarker," *Talanta*, vol. 258, no. March, p. 124461, 2023, doi: 10.1016/j.talanta.2023.124461.
- [156] S. N. Lone *et al.*, "Liquid biopsy: a step closer to transform diagnosis, prognosis and future of cancer treatments," *Mol. Cancer*, vol. 21, no. 1, pp. 1–22, 2022, doi: 10.1186/s12943-022-01543-7.
- [157] S. Connal *et al.*, "Liquid biopsies: the future of cancer early detection," *J. Transl. Med.*, vol. 21, no. 1, pp. 1–18, 2023, doi: 10.1186/s12967-023-03960-8.
- [158] L. Ma *et al.*, "Liquid biopsy in cancer current: status, challenges and future prospects.," *Signal Transduct. Target. Ther.*, vol. 9, no. 1, p. 336, 2024, doi: 10.1038/s41392-024-02021-w.
- [159] Y. Xie *et al.*, "Nucleic acid-mediated SERS Biosensors: Signal enhancement strategies and applications," *Biosens. Bioelectron.*, vol. 282, no. April, p. 117519, 2025, doi: 10.1016/j.bios.2025.117519.
- [160] L. Xu *et al.*, "A catalytic hairpin assembly-assisted magnetic SERS platform with core-satellite structure for ultra-sensitive circRNA detection in early lung cancer screening," *Chem. Eng. J.*, vol. 514, no. January, 2025, doi: 10.1016/j.cej.2025.163309.
- [161] J. Wu *et al.*, "3D hierarchic interfacial assembly of Au nanocage@Au along with IS-AgMNPs for simultaneous, ultrasensitive, reliable, and quantitative SERS detection of colorectal cancer related miRNAs," *Biosens. Bioelectron.*, vol. 248, no. October 2023, p. 115993, 2024, doi: 10.1016/j.bios.2023.115993.
- [162] N. Su *et al.*, "Specific isolation and quantification of PD-L1 positive tumor derived exosomes for accurate breast cancer discrimination via aptamer-functionalized magnetic composites and SERS immunoassay," *Talanta*, vol. 281, no. August 2024, pp. 1–9, 2025, doi: 10.1016/j.talanta.2024.126956.
- [163] J. Xiong *et al.*, "DNA walker-powered ratiometric SERS cytosensor of circulating tumor cells with single-cell sensitivity," *Biosens. Bioelectron.*, vol. 213, no. March, p. 114442, 2022, doi: 10.1016/j.bios.2022.114442.
- [164] B. V. Saghezi, K. Mansouri, and R. Hallaj, "Microfluidic Systems in Cancer Diagnosis: Advancing Exosome and CTC Detection From Laboratory Research to Clinical Application," *Nano Sel.*, pp. 1–20, 2025, doi: 10.1002/nano.202400108.
- [165] J. Wang, K. M. Koo, Y. Wang, and M. Trau, "Engineering State-of-the-Art Plasmonic Nanomaterials for SERS-Based Clinical Liquid Biopsy Applications," *Adv. Sci.*, vol. 6, no. 23, 2019, doi: 10.1002/adv.201900730.
- [166] K. M. Koczula and A. Gallotta, "Lateral flow assays," *Essays Biochem.*, vol. 60, no. 1, pp. 111–120, 2016, doi:

- 10.1042/EBC20150012.
- [167] Y. Sun, Y. Tang, L. Yu, Y. Wang, S. Peng, and X. Zhou, "Circulating nucleic acids as liquid biopsies for disease prediction, screening and diagnosis," *Sci. China Chem.*, vol. 66, no. 11, pp. 3006–3023, 2023, doi: 10.1007/s11426-023-1727-9.
 - [168] N. Lyu *et al.*, "Multiplex detection of ctDNA mutations in plasma of colorectal cancer patients by pcr/sers assay," *Nanotheranostics*, vol. 4, no. 4, pp. 224–232, 2020, doi: 10.7150/ntno.48905.
 - [169] K. Yi, X. Wang, S. K. Filippov, and H. Zhang, "Emerging ctDNA detection strategies in clinical cancer theranostics," *Smart Med.*, vol. 2, no. 4, pp. 1–17, 2023, doi: 10.1002/smmd.20230031.
 - [170] L. Wu *et al.*, "A hand-powered SERS-microfluidic chip for circulating tumor DNA detection from whole blood," *Sensors Actuators B Chem.*, vol. 401, no. July 2023, p. 135081, 2024, doi: 10.1016/j.snb.2023.135081.
 - [171] A. Bartolomucci *et al.*, "Circulating tumor DNA to monitor treatment response in solid tumors and advance precision oncology," *npj Precis. Oncol.*, vol. 9, no. 1, pp. 1–19, 2025, doi: 10.1038/s41698-025-00876-y.
 - [172] S. Toden and A. Goel, "Non-coding RNAs as liquid biopsy biomarkers in cancer," *Br. J. Cancer*, no. January, 2022, doi: 10.1038/s41416-021-01672-8.
 - [173] H.-N. Wang, B. M. Crawford, A. M. Fales, M. L. Bowie, V. L. Seewaldt, and T. Vo-Dinh, "Multiplexed Detection of MicroRNA Biomarkers Using SERS-Based Inverse Molecular Sentinel (iMS) Nanoprobes," *J. Phys. Chem. C. Nanomater. Interfaces*, vol. 120, no. 37, pp. 21047–21050, Sep. 2016, doi: 10.1021/acs.jpcc.6b03299.
 - [174] V. Armand-Labit and A. Pradines, "Circulating cell-free microRNAs as clinical cancer biomarkers," *Biomol. Concepts*, vol. 8, no. 2, pp. 61–81, 2017, doi: 10.1515/bmc-2017-0002.
 - [175] D. Bautista-Sánchez *et al.*, "The Promising Role of miR-21 as a Cancer Biomarker and Its Importance in RNA-Based Therapeutics," *Mol. Ther. Nucleic Acids*, vol. 20, no. 22, pp. 409–420, 2020, doi: 10.1016/j.omtn.2020.03.003.
 - [176] H. S. Tan, T. Wang, J. M. Han, M. Liu, and S. S. Li, "Dual-signal SERS biosensor based on spindle-shaped gold array for sensitive and accurate detection of miRNA 21," *Sensors Actuators B Chem.*, vol. 403, no. October 2023, p. 135157, 2024, doi: 10.1016/j.snb.2023.135157.
 - [177] A. U. Khan, M. Ali, and M. A. Wahab, "Diagnostic Potential of Cross-Specimen microRNA Panels as Biomarkers for Colorectal Cancer: A Systematic Review and Meta-analysis," *Mol. Diagnosis Ther.*, 2025, doi: 10.1007/s40291-025-00805-6.
 - [178] S. Wang *et al.*, "Circular RNAs in body fluids as cancer biomarkers: the new frontier of liquid biopsies," *Mol. Cancer*, vol. 20, no. 1, pp. 1–10, 2021, doi: 10.1186/s12943-020-01298-z.
 - [179] M. Yáñez-Mó *et al.*, "Biological properties of extracellular vesicles and their physiological functions," *J. Extracell. Vesicles*, vol. 4, no. 2015, pp. 1–60, 2015, doi: 10.3402/jev.v4.27066.
 - [180] B. Zhou *et al.*, "Application of exosomes as liquid biopsy in clinical diagnosis," *Signal Transduct. Target. Ther.*, vol. 5, no. 1, 2020, doi: 10.1038/s41392-020-00258-9.
 - [181] D. Yu *et al.*, "Exosomes as a new frontier of cancer liquid biopsy," *Mol. Cancer*, vol. 21, no. 1, pp. 1–33, 2022, doi: 10.1186/s12943-022-01509-9.
 - [182] Y. Wu *et al.*, "Tumor-derived exosomal PD-L1: a new perspective in PD-1/PD-L1 therapy for lung cancer," *Front. Immunol.*, vol. 15, no. March, pp. 1–11, 2024, doi: 10.3389/fimmu.2024.1342728.
 - [183] T. Wang *et al.*, "One-Step Capture and Profiling of Single Exosomes from Whole Blood Using Nanocavities through a Cascade Capture Strategy," *Sensors Actuators B Chem.*, p. 137315, 2025, doi: <https://doi.org/10.1016/j.snb.2025.137315>.
 - [184] X. Chen *et al.*, "Sensitive phenotyping of serum extracellular vesicles on a SERS-microfluidic platform for early-stage clinical diagnosis of ovarian carcinoma," *Biosens. Bioelectron.*, vol. 267, no. September 2024, p. 116724, 2025, doi: 10.1016/j.bios.2024.116724.
 - [185] Q. Zhou *et al.*, "Glycan Profiling in Small Extracellular Vesicles with a SERS Microfluidic Biosensor Identifies Early Malignant Development in Lung Cancer," *Adv. Sci.*, vol. 2401818, pp. 1–12, 2024, doi: 10.1002/advs.202401818.
 - [186] V. Vrablova *et al.*, "Glycosylation in extracellular vesicles: Isolation, characterization, composition, analysis and clinical applications," *Biotechnol. Adv.*, vol. 67, no. June, p. 108196, 2023, doi: 10.1016/j.biotechadv.2023.108196.
 - [187] N. Feliu, M. Hassan, E. Garcia Rico, D. Cui, W. Parak, and R. Alvarez-Puebla, "SERS Quantification and Characterization of Proteins and Other Biomolecules," *Langmuir*, vol. 33, no. 38, pp. 9711–9730, 2017, doi: 10.1021/acs.langmuir.7b01567.
 - [188] J. Chen, P. Ma, J. Xu, M. Zang, and W. Li, "Glycosylation-Targeting Aptamer for the Feasible Construction of a Dual Aptamer-Based Plasmonic Immunosandwich Assay in Cancer Diagnostics," *Anal. Chem.*, 2024, doi: 10.1021/acs.analchem.4c03770.
 - [189] X. Su *et al.*, "SERS lateral flow strip detection of serum biomarkers for noninvasive assessment of operative microwave ablation outcomes of unresectable hepatocellular carcinoma," *Chem. Eng. J.*, vol. 485, no. October 2023, p. 149833, 2024, doi: 10.1016/j.cej.2024.149833.

- [190] S. Peng *et al.*, “Portable SERS-based lateral flow immunoassay strips with self-calibration for detection of a prostate cancer biomarker,” *Sensors Actuators B Chem.*, vol. 401, no. November 2023, p. 135012, 2024, doi: 10.1016/j.snb.2023.135012.
- [191] Y. Lu *et al.*, “Multifunctional nanocone array as solid immunoassay plate and SERS substrate for the early diagnosis of prostate cancer on microfluidic chip,” *Sensors Actuators B Chem.*, vol. 376, no. PA, p. 133046, 2023, doi: 10.1016/j.snb.2022.133046.
- [192] D. J. Smit and K. Pantel, “Circulating tumor cells as liquid biopsy markers in cancer patients,” *Mol. Aspects Med.*, vol. 96, no. February, p. 101258, 2024, doi: 10.1016/j.mam.2024.101258.
- [193] C. Alix-Panabières and K. Pantel, “Challenges in circulating tumour cell research,” *Nat. Rev. Cancer*, vol. 14, no. 9, pp. 623–631, 2014, doi: 10.1038/nrc3820.
- [194] X. Li, Y. Li, W. Shao, Z. Li, R. Zhao, and Z. Ye, “Strategies for enrichment of circulating tumor cells,” *Transl. Cancer Res.*, vol. 9, no. 3, pp. 2012–2025, 2020, doi: 10.21037/tcr.2020.01.17.
- [195] W. Lv *et al.*, “Efficient detection of single circulating tumor cell in blood using Raman mapping based on Aptamer-SERS bio-probe coupled with micropore membrane filtration,” *Talanta*, vol. 267, no. April 2023, p. 125220, 2024, doi: 10.1016/j.talanta.2023.125220.
- [196] C. Zhan *et al.*, “SERS-based microfluidic liquid biopsy of prostate cancer: New bioinspired dual recognition interface for single-cell detection,” *Chem. Eng. J.*, vol. 516, no. March, p. 164059, 2025, doi: 10.1016/j.cej.2025.164059.
- [197] D. Burčák, J. Macko, N. Podrojková, J. Demeterová, M. Stano, and A. Oriňák, “Role of Cell Adhesion in Cancer Metastasis Formation: A Review,” *ACS Omega*, vol. 10, no. 6, pp. 5193–5213, 2025, doi: 10.1021/acsomega.4c08140.
- [198] P. Yadav and C. C. Mandal, “Bioimaging: Usefulness in Modern Day Research,” in *Practical Approach to Mammalian Cell and Organ Culture*, T. K. Mukherjee, P. Malik, and S. Mukherjee, Eds. Singapore: Springer Nature Singapore, 2023, pp. 1205–1230. doi: 10.1007/978-981-19-1731-8_23-1.
- [199] A. Bollhagen and B. Bodenmiller, “Highly Multiplexed Tissue Imaging in Precision Oncology and Translational Cancer Research,” *Cancer Discov.*, vol. 14, no. 11, pp. 2071–2088, 2024, doi: 10.1158/2159-8290.CD-23-1165.
- [200] L. Xu *et al.*, “A dual-signal biosensor based on surface-enhanced Raman spectroscopy for high-sensitivity quantitative detection and imaging of circRNA in living cells,” *Biosens. Bioelectron.*, vol. 268, no. October 2024, p. 116875, 2025, doi: 10.1016/j.bios.2024.116875.
- [201] S. Spaziani *et al.*, “Combined SERS-Raman screening of HER2-overexpressing or silenced breast cancer cell lines,” *J. Nanobiotechnology*, vol. 22, no. 1, pp. 1–17, 2024, doi: 10.1186/s12951-024-02600-7.
- [202] H. Y. Chen *et al.*, “A multi-functional core-shell surface-enhanced Raman scattering nanosensor for simultaneous imaging of epidermal growth factor receptor on cell membranes and ROS secreted from living cells,” *Sensors Actuators B Chem.*, vol. 422, no. July 2024, p. 136636, 2025, doi: 10.1016/j.snb.2024.136636.
- [203] W. Mo *et al.*, “Combined Morphological and Spectroscopic Diagnostic of HER2 Expression in Breast Cancer Tissues Based on Label-Free Surface-Enhanced Raman Scattering,” *Anal. Chem.*, vol. 95, no. 5, pp. 3019–3027, 2023, doi: 10.1021/acs.analchem.2c05067.
- [204] V. P. Murali *et al.*, “A clinically feasible diagnostic spectro-histology built on SERS-nanotags for multiplex detection and grading of breast cancer biomarkers,” *Biosens. Bioelectron.*, vol. 227, no. November 2022, p. 115177, 2023, doi: 10.1016/j.bios.2023.115177.
- [205] Y. Che, F. Ren, X. Zhang, L. Cui, H. Wu, and Z. Zhao, “Immunohistochemical HER2 Recognition and Analysis of Breast Cancer Based on Deep Learning,” *Diagnostics*, vol. 13, no. 2, 2023, doi: 10.3390/diagnostics13020263.
- [206] C. Esposito *et al.*, “Assessment of Primary Human Liver Cancer Cells by Artificial Intelligence-Assisted Raman Spectroscopy,” *Cells*, vol. 12, no. 22, 2023, doi: 10.3390/cells12222645.
- [207] A. Horta-Velázquez, F. Arce, E. Rodríguez-Sevilla, and E. Morales-Narváez, “Toward smart diagnostics via artificial intelligence-assisted surface-enhanced Raman spectroscopy,” *TrAC - Trends Anal. Chem.*, vol. 169, no. September, 2023, doi: 10.1016/j.trac.2023.117378.
- [208] H. Z. He *et al.*, “Integrating surface-enhanced Raman scattering with machine learning: Pioneering a comprehensive diagnostic and therapeutic platform for cancer management,” *Interdiscip. Med.*, vol. 3, no. 5, 2025, doi: 10.1002/INMD.20250037.
- [209] J. Li, F. Liu, X. Bi, and J. Ye, “Imaging immune checkpoint networks in cancer tissues with supermultiplexed SERS nanoprobes,” *Biomaterials*, vol. 302, no. September, p. 122327, 2023, doi: 10.1016/j.biomaterials.2023.122327.
- [210] X. He and C. Xu, “Immune checkpoint signaling and cancer immunotherapy,” *Cell Res.*, vol. 30, no. 8, pp. 660–669, 2020, doi: 10.1038/s41422-020-0343-4.
- [211] X. Lin *et al.*, *Regulatory mechanisms of PD-1/PD-L1 in cancers*, vol. 23, no. 1. 2024. doi: 10.1186/s12943-024-02023-w.
- [212] X. Qian *et al.*, “In vivo tumor targeting and spectroscopic detection with surface-enhanced Raman nanoparticle tags,” *Nat. Biotechnol.*, vol. 26, no. 1, pp. 83–90, 2008, doi: 10.1038/nbt1377.

- [213] C. Qiu *et al.*, “Highly sensitive surface-enhanced Raman scattering (SERS) imaging for phenotypic diagnosis and therapeutic evaluation of breast cancer,” *Chem. Eng. J.*, vol. 459, no. September 2022, p. 141502, 2023, doi: 10.1016/j.cej.2023.141502.
- [214] B. Deng *et al.*, “Sentinel lymph node identification using NIR-II ultrabright Raman nanotags on preclinical models,” *Biomaterials*, vol. 308, no. January, p. 122538, 2024, doi: 10.1016/j.biomaterials.2024.122538.
- [215] O. E. Eremina, A. T. Czaja, A. Fernando, A. Aron, D. B. Eremin, and C. Zavaleta, “Expanding the Multiplexing Capabilities of Raman Imaging to Reveal Highly Specific Molecular Expression and Enable Spatial Profiling,” *ACS Nano*, vol. 16, no. 7, pp. 10341–10353, 2022, doi: 10.1021/acsnano.2c00353.
- [216] B. Chen, J. Gao, H. Sun, Z. Chen, and X. Qiu, “Plasmonic surgery: SERS-active smart nanotools reshaping intraoperative decision-making and precision therapy,” *Crit. Rev. Oncol. Hematol.*, vol. 214, no. July, p. 104838, 2025, doi: 10.1016/j.critrevonc.2025.104838.
- [217] M. Zhang *et al.*, “Rapid Glutathione Analysis with SERS Microneedles for Deep Glioblastoma Tissue Differentiation,” *Anal. Chem.*, vol. 96, no. 25, pp. 10200–10209, 2024, doi: 10.1021/acs.analchem.4c00483.
- [218] A. N. S. Institute and L. I. o America, “American national standard for safe use of lasers.” Laser Institute of America Orlando, FL, USA, 2007.
- [219] B. Chen, J. Gao, H. Sun, Z. Chen, and X. Qiu, “Innovative applications of SERS in precision medicine: In situ and real-time live imaging,” *Talanta*, vol. 294, no. March, p. 128225, 2025, doi: 10.1016/j.talanta.2025.128225.
- [220] X. Wang *et al.*, “A fluorescence and SERS dual-mode biosensor for quantification and imaging of Mucin1 in living cells,” *Biosens. Bioelectron.*, vol. 270, no. November 2024, p. 116964, 2025, doi: 10.1016/j.bios.2024.116964.
- [221] P. Bagheri *et al.*, “A Systematic Approach toward Enabling Maximal Targeting Efficiency of Cell Surface Proteins with Actively Targeted SERS Nanoparticles,” *ACS Appl. Mater. Interfaces*, vol. 16, no. 13, pp. 15847–15860, 2024, doi: 10.1021/acsami.3c18959.
- [222] T. Pan *et al.*, “Gold nanorods with iron oxide dual-modal bioprobes in SERS-MRI enable accurate programmed cell death ligand-1 expression detection in triple-negative breast cancer,” *APL Bioeng.*, vol. 7, no. 2, pp. 1–10, 2023, doi: 10.1063/5.0152846.
- [223] Y. Tan *et al.*, “DNA Assembly of Plasmonic Nanostructures Enables In Vivo SERS-Based MicroRNA Detection and Tumor Photoacoustic Imaging,” *Anal. Chem.*, vol. 95, no. 30, pp. 11236–11242, 2023, doi: 10.1021/acs.analchem.3c00775.
- [224] M. Rahman, “Magnetic Resonance Imaging and Iron-oxide Nanoparticles in the era of Personalized Medicine,” *Nanotheranostics*, vol. 7, no. 4, pp. 424–449, 2023, doi: 10.7150/ntno.86467.
- [225] A. B. E. Attia *et al.*, “A review of clinical photoacoustic imaging: Current and future trends,” *Photoacoustics*, vol. 16, no. November, p. 100144, 2019, doi: 10.1016/j.pacs.2019.100144.
- [226] A. Khetani, A. Momenpour, E. I. Alarcon, and H. Anis, “Hollow core photonic crystal fiber for monitoring leukemia cells using surface enhanced Raman scattering (SERS),” *Biomed. Opt. Express*, vol. 6, no. 11, p. 4599, 2015, doi: 10.1364/boe.6.004599.
- [227] X. S. Zheng, I. J. Jahn, K. Weber, D. Cialla-May, and J. Popp, “Label-free SERS in biological and biomedical applications: Recent progress, current challenges and opportunities,” *Spectrochim. Acta - Part A Mol. Biomol. Spectrosc.*, 2018, doi: 10.1016/j.saa.2018.01.063.
- [228] L. A. Lane, X. Qian, and S. Nie, “SERS Nanoparticles in Medicine: From Label-Free Detection to Spectroscopic Tagging,” *Chem. Rev.*, vol. 115, no. 19, pp. 10489–10529, 2015, doi: 10.1021/acs.chemrev.5b00265.
- [229] J. Dolai, K. Mandal, and N. R. Jana, “Nanoparticle Size Effects in Biomedical Applications,” *ACS Appl. Nano Mater.*, vol. 4, no. 7, pp. 6471–6496, 2021, doi: 10.1021/acsanm.1c00987.
- [230] N. Hoshyar, S. Gray, H. Han, and G. Bao, “The effect of nanoparticle size on in vivo pharmacokinetics and cellular interaction,” *Nanomedicine (Lond.)*, vol. 11, no. 6, pp. 673–692, Mar. 2016, doi: 10.2217/nnm.16.5.
- [231] Z. Chen, R. K. Kankala, L. Long, S. Xie, A. Chen, and L. Zou, “Current understanding of passive and active targeting nanomedicines to enhance tumor accumulation,” *Coord. Chem. Rev.*, vol. 481, no. 668, p. 215051, 2023, doi: 10.1016/j.ccr.2023.215051.

Chapter 5: Nanopillar Arrays on Optical Fiber Tip Ends for Surface-enhanced Raman Spectroscopy

Abstract

This chapter explores the design, fabrication, and characterization of gold-coated ZnO nanopillar arrays on optical fiber tips for surface-enhanced Raman spectroscopy (SERS) for potential utility in cancer diagnostics. Using hydrothermal synthesis, both ordered (masked) and disordered (unmasked) nanostructures were fabricated. The effects of nanostructure arrangement and excitation wavelength on plasmonic behavior, signal reproducibility and SERS signal intensity were systematically assessed, employing biphenyl-4-thiol (BPT) as a reference molecule under both direct and remote interrogation configurations with 633 nm and 785 nm lasers. By tuning structural ordering and excitation wavelength, hydrothermally grown Au-coated ZnO nanopillar optrodes can be adapted to meet diverse operational requirements, marking a crucial step towards the future development of lab-on-fiber SERS devices for cancer diagnostics.

1. Introduction

Surface-enhanced Raman spectroscopy is emerging as a powerful technique in biomedicine and cancer diagnostics, providing highly sensitive, specific detection and accurate biochemical fingerprinting of biological samples. Conventional SERS sensors are limited by their bulky, rigid design, fabrication challenges, and poor flexibility, which restrict their use in real-time, in-situ monitoring [1]. Designing SERS substrates that combine stability, reproducibility, and high sensitivity with affordability and robustness for clinical use continues to be a significant issue [2]. Inefficient light delivery and collection further reduce sensitivity and signal stability, underscoring the need for more versatile and innovative SERS substrates. In recent years, combining SERS with advanced platforms such as optical fiber sensors have opened new frontiers in biomedicine, offering a range of advantages over conventional approaches. This integration enables rapid, real-time, and portable detection with exceptional sensitivity and low detection limits. Optical fibers bring additional benefits, including resistance to electromagnetic interference, multiplexing capacity, low costs and remote sensing functionality, all of which enhance their suitability for biomedical applications. Their small, flexible, and hair-like structure and biocompatibility also allows for precise positioning and minimally invasive use, making them particularly promising for *in vivo* point-of-care diagnostics when incorporated into medical catheters or needles [3].

In particular, the integration of functional materials into optical fibers has paved the way for a new paradigm known as *lab-on-fiber* technology [4], [5], [6]. This approach goes beyond the traditional role of optical fibers as simple light-guiding elements, by embedding micro- and nanoscale structures or components directly onto the fiber platform (**Figure 1**). As a result, standard optical fibers can be transformed into versatile, multifunctional devices capable of performing sensing, diagnostic, and analytical tasks within a compact and highly efficient architecture [7].

SERS optical fiber sensors are designed based on various geometries and configurations to optimize light-matter interaction and signal collection.

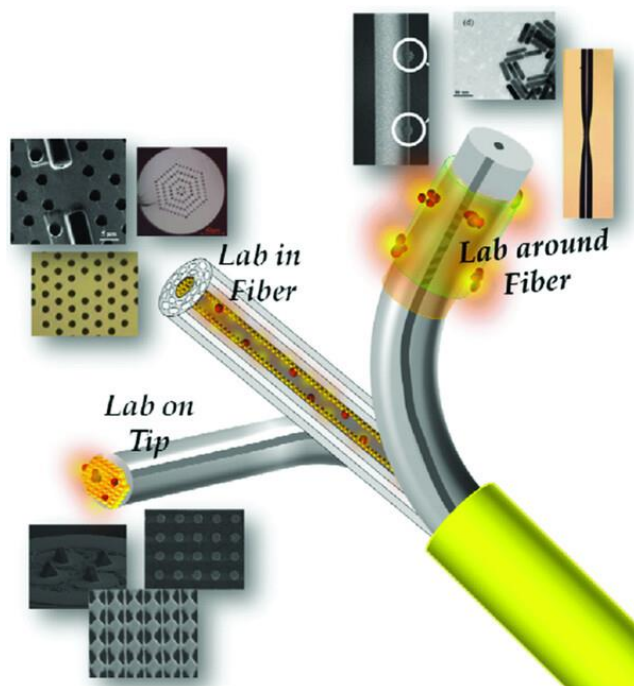


Figure 1. The lab-on-fiber vision, from Ref. [6].

In optical fiber-based SERS platforms, light-matter interactions can be engineered to occur at different regions of the fiber, depending on the integration of functional materials. Specifically, Lab-on-fiber technology can be categorized into three main types, depending on where the functionalization occurs (**Figure 1**). In *lab-in-fiber*, the functional material is confined to the fiber's internal channels or hollow cores, facilitating interactions inside the fiber itself [1]. In *lab-around-fiber*, materials or structures are applied to the fiber's sidewalls, enabling interactions along the fiber cladding. In *lab-on-tip* places functional materials at the fiber end facet, concentrating excitation and Raman scattering at the probe tip [4]. These configurations extend the capabilities of optical fibers for sensing, analysis, and other multifunctional applications.

Among these, SERS optical fibers with nanostructures at the end of a tip, commonly referred to as optrodes, are particularly innovative. This type of SERS optrode essentially functions as an optical fiber probe whose end facet is integrated with plasmonic nanomaterials or nanostructures, creating SERS-active fiber tips through either top-down or bottom-up fabrication approaches. It serves simultaneously as a SERS-active substrate, a waveguide for excitation light, and a collector of the enhanced Raman signals, as shown in **Figure 2** [8], [9].

Generally, on the tip of the optical fiber, nanostructures are engineered in the form of metallic nanoparticles or patterned metallic surfaces, which serve as efficient SERS-active sites. When illuminated, these nanostructures generate strong electromagnetic hotspots that amplify the Raman scattering of nearby molecules. Because the structures are directly integrated onto the fiber tip, the excitation light and the scattered Raman signal naturally overlap within the same confined region, leading to highly efficient signal collection. This

intimate spatial coupling between enhancement sites and the optical pathway maximizes the overall sensitivity and performance of the fiber-based SERS probe. However, coating the flat end face of an optical fiber with nanoparticles offers a simpler method compared to fabricating structured fibers, but it limits the sensing area and reduces detection sensitivity [10].

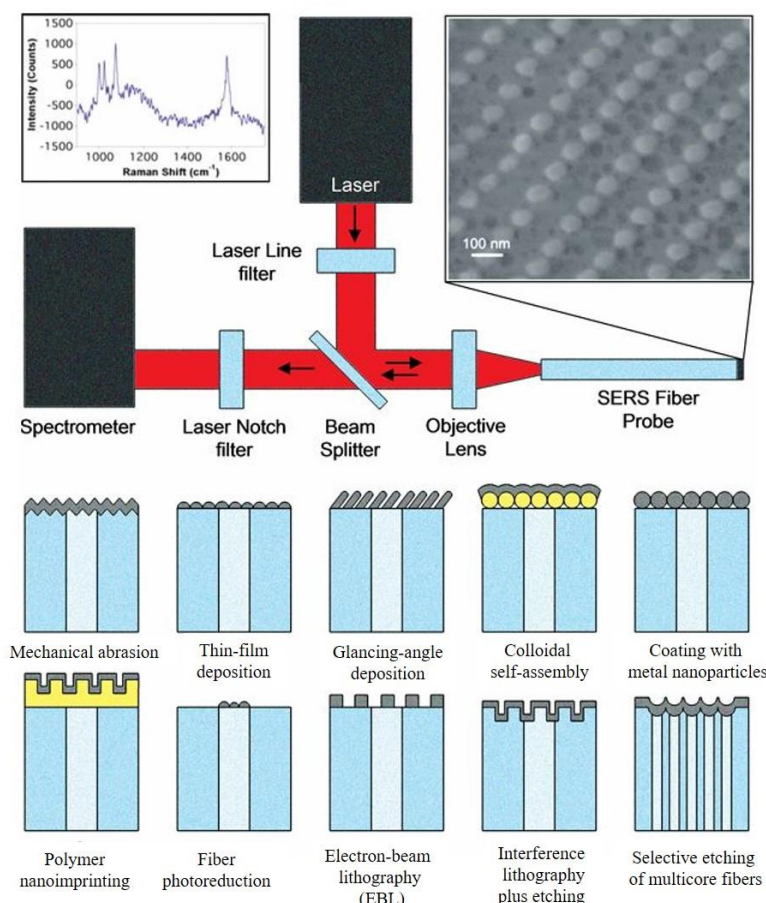


Figure 2. Indirect interrogation of a SERS optical fiber and nanotechnologies fabrication methods of SERS-active substrates onto a tip of an optical fiber. In an optrode SERS probe, light is guided along the fiber to stimulate a nanostructured tip, with the enhanced Raman signal collected back through the same fiber for analysis (upper part). These nanostructured tips can be created using techniques ranging from basic surface roughening and nanoparticle deposition to advanced lithography and etching methods (lower part). Re-adapted from Ref. [9].

Regarding their fabrication, SERS-active fiber tips can be achieved through either top-down or bottom-up strategies (**Figure 3**). Top-down methods—such as focused ion beam milling [11], chemical etching [12], femtosecond laser ablation [13] and different lithography methodologies like electron beam lithography [14], nanoimprint lithography [15], enabling precise control of nanostructures and optical properties, but they are often expensive, technically demanding, and prone to fabrication challenges. Bottom-up techniques, including nanoparticle self-assembly [16], [17], 3D direct laser writing [18], are more straightforward and cost-efficient, though they generally offer less stability and structural precision [19], [20]. Interestingly, nanosphere lithography (NSL) is a bottom-up technique that uses the self-assembly of nanospheres to form masks for low-cost nanoscale patterning. It offers a rapid and low-cost alternative to conventional lithography techniques, but its practical application and scalability depend on producing uniform, defect-free monolayers

over large areas [21]. Overall, top-down fabrication emphasizes accuracy and tunability, whereas bottom-up methods prioritize simplicity and scalability [22]. Recent studies demonstrate that Lab on fiber biosensors are capable of achieving ultra-low detection limits, down to the pM and fM range, thus paving the way for a new generation of *in vivo* liquid and tissue biopsy techniques as described in the previous chapter.

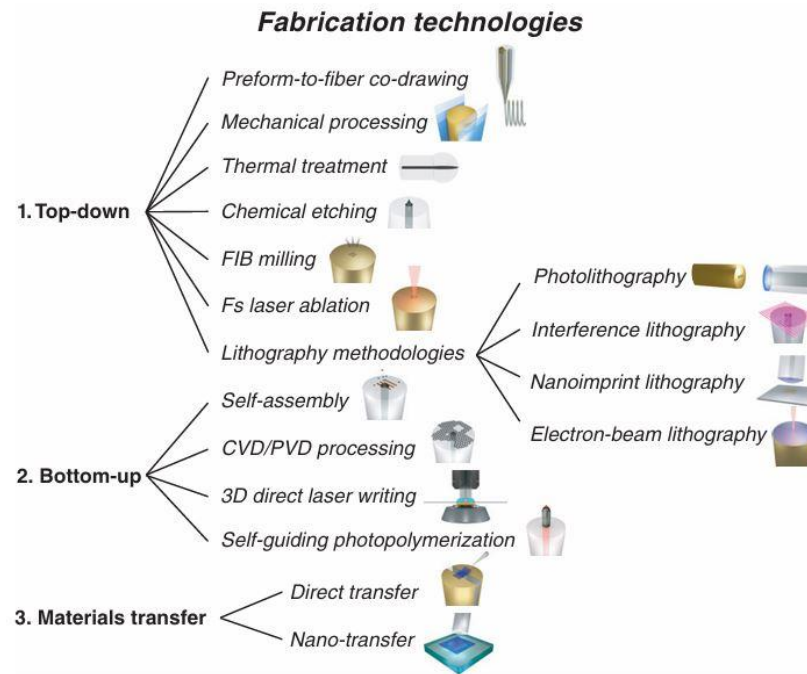


Figure 3. Schematic illustration of optical fiber tip fabrication, showing nanostructures formed directly on the tips via top-down or bottom-up methods, or transferred from a flat substrate onto the fiber's end facet, from Ref. [19].

One of the primary challenges in advancing fiber-based SERS probes is the development of stable, reproducible, and highly sensitive nanostructured substrates at the fiber tip through methods that are both reliable and cost-efficient. The transfer of planar nanostructures onto such small, delicate platforms often leads to mechanical instabilities, which in turn reduce light-coupling efficiency and compromise the resonant properties of the sensing interface, major obstacles to scalability and large-scale production [4]. Although the flat end-facet of an optical fiber is, in principle, compatible with standard nanofabrication approaches, the fiber's small diameter, high aspect ratio, and intrinsic fragility pose significant difficulties in processing and handling [7]. Additionally, the fiber itself introduces unwanted background signals, including auto-fluorescence and Raman scattering from its glass matrix, which can obscure the analyte signal and diminish overall spectral quality. Addressing these issues requires novel fabrication strategies that go beyond conventional methods, enabling both reduced production costs and high-throughput scalability. Equally critical is the judicious choice of fiber types and the optimization of their coupling with Raman spectrometers, ensuring efficient signal collection while minimizing background interference [2]. Additionally, fiber optrodes with disordered nanostructures often exhibit high variability, resulting in low uniformity of SERS hotspots and limited reproducibility. Conversely, fiber SERS optrodes with carefully engineered, ordered nanostructures on the fiber tip establish a consistent and uniform pattern of hot spots, resulting in stronger signal stability, higher sensitivity, and more reliable, reproducible sensing performance [8].

Together, these improvements are essential for realizing high-performance SERS optrodes with practical applicability.

Recent researches have focused on designing noble metal–semiconductor hybrid SERS nanostructures based on ZnO, aiming for synergistic advanced applications in SERS. In particular, semiconductors provide chemical enhancement but weak Raman signals, while noble metals like Au or Ag offer strong electromagnetic enhancement. When high-dielectric semiconductors such as ZnO or TiO₂ are coated with noble metals, the plasmon resonance shifts into the visible/near-IR range, creating hotspots and better matching probe molecules. This synergy greatly boosts SERS efficiency [23]. Specifically, zinc oxide (ZnO) is a biocompatible, stable, and abundant n-type semiconductor whose nanostructured forms offer outstanding electrical, optical, and mechanical properties. With features like a wide bandgap, strong exciton-binding energy, high surface-to-volume ratio, and different shapes, ZnO is highly versatile and especially promising for advanced applications such as efficient biosensors [24], [25]. While physical techniques can create high-quality ZnO nanostructures, they are costly, complex, and often demand high temperatures and advanced equipment. In contrast, chemical methods, including hydrothermal growth of ZnO nanostructures provide a more practical option, being simpler, scalable, efficient, and cost-effective, which makes them well-suited for large-scale and industrial applications [26]. Hydrothermal synthesis outperforms traditional solid-phase methods by reducing particle-handling challenges while providing high product purity, rapid reaction times, and precise control over particle size and morphology. Utilizing high-temperature water as an inexpensive and eco-friendly solvent, it enables the production of a variety of nanostructures [27].

The fabrication of hybrid SERS substrates with hydrothermally grown ZnO nanostructures subsequently covered with noble metals have been recently explored from different research groups. One approach involved fabricating Ag-decorated ZnO nanoplates through hydrothermal growth of the ZnO followed by chemical reduction of silver nitrate, producing a SERS substrate capable of detecting methylene blue with a high enhancement factor of 6.2×10^6 and a detection limit of 10^{-9} M, while maintaining good reproducibility [28]. Another study developed a three-dimensional Ag/ZnO/Au hybrid structure for ultrasensitive and reusable DNA detection, combining thermal evaporation of an ultrathin Ag film, hydrothermal growth of ZnO nanorods, and sputtering of Au nanoparticles. This substrate achieved an enhancement factor of 1×10^{10} , with LODs down to 10^{-16} M for rhodamine 6-G and 10 ng/μL for lambda DNA [29]. These studies demonstrate the adaptability of ZnO-based hybrid nanostructures for creating highly sensitive, reliable, and long-lasting sensing platforms. However, they primarily focused on substrates such as Si/SiO₂ or quartz slides, rather than directly onto optical fiber tips.

This chapter provides a detailed investigation into the development and evaluation of Au-coated ZnO nanopillars as SERS-active platforms. To this end, the study focused on integrating gold-coated ZnO nanopillar nanostructures onto optical fiber tips through hydrothermal synthesis, examining both ordered (masked) and disordered (unmasked) growth approaches to assess how nanostructure arrangement and wavelength selection affect plasmonic behavior, signal intensity and reproducibility.

First, ZnO nanopillar arrays were hydrothermally grown on fiber tips, followed by gold deposition to generate plasmonic hotspots. The multimode fibers were then functionalized

with biphenyl-4-thiol (BPT) to form a self-assembled monolayer, which represented a standard reference to assess the SERS fiber responses. Morphological analysis was conducted using SEM to evaluate the effects of masking on nanopillar nanostructures. To gain deeper insights, comparative statistical analyses of characteristic Raman band intensities were performed, alongside SERS mapping collection, enabling spatial visualization of signal uniformity and quantitative assessment of SERS performance across different fiber architectures. This investigation provided a robust framework for correlating structural and chemical modifications with the resulting plasmonic activity and SERS performance of the nanostructured optical fibers.

2. Material and methods

2.1 Fabrication of ZnO thin film

Low-hydroxyl commercial multimodal optical fibers (Thorlabs, USA model: FP200ERT) with a pure silica core and hard polymer cladding (200/225 μm) and featuring a numerical aperture (NA) of 0.50 were employed. The fibers were jacketed with a 500 μm Tefzel protective layer to improve mechanical robustness and reduce handling-induced damage. The fiber ends were terminated with a 2.5 mm zirconia based-ceramic patch cord. In order to create ZnO nanopillar on the fiber tip, a seed layer of ZnO must be deposited on the optical fiber tip. Therefore, ZnO thin films were deposited on the fiber tip by using electron beam evaporation at room temperature under a high vacuum of 10^{-7} torr, ensuring minimal contamination and precise control of deposition, with a film thickness of 30 nm targeted to obtain optically thin films suitable for ellipsometric analysis and interference-based reflection measurements. The ZnO thin films were deposited on both zirconia based-ceramic patch cord optical fiber and planar fused silica (25 mm in diameter and 1 mm thickness), selected for its well-known optical transparency and stability, making it a reference material for thin film studies. The thickness of the deposited films was measured using a profilometer.

2.2 Optical characterization

The optical properties of the ZnO thin films were investigated using spectroscopic ellipsometry and reflection spectroscopy. Ellipsometric modeling was performed by fixing the ZnO bandgap energy at 3.15 eV, as established in literature, to constrain the fitting process [30], [31]. For films deposited on fused silica, a standard oscillator-based dispersion model was employed. The model was adjusted to match the experimental ellipsometry data, yielding film thickness, surface roughness, and refractive index values. For the zirconia substrates, a modified Sellmeier model was used to describe the substrate dispersion, which was then integrated into the ZnO thin film fitting process. This allowed determination of the refractive index and thickness of the ZnO layer when deposited on the ceramic fiber tips. Complementary optical characterization was performed using reflection spectroscopy with an Avantes system. The experimental setup included a deuterium–halogen light source (AvaLight-D(H)-S) to provide broadband illumination, a high-resolution spectrometer (AvaSpec-ULS3648) controlled via dedicated software, and a fiber-coupled reflection probe for signal collection. A micro-positioner ensured accurate alignment between the reflection

probe and the ZnO-coated sample surface, with a fixed working distance of 3 mm. The integration time was set to 1.4 ms, and each spectrum was averaged over 100 acquisitions to improve signal-to-noise ratio. A thin gold film was used as the reflectance reference standard, due to its stable and highly reflective properties across the UV–visible spectral range.

2.3 Fabrication of gold coated ZnO nanopillar onto fiber tip by hydrothermal synthesis

Following the deposition of ZnO thin film, ZnO nanopillars were fabricated on optical fiber tips using different fabrication steps to produce both masked and unmasked nanostructures. In the masked fabrication process, the fiber tip- surface was first coated with a monolayer of colloidal crystals (MCC) formed by nanospheres of polystyrene dissolved in ethanol (40:60 ratio) to generate self-assembled nanospheres in a close-packed array (CPA757). The coated fibers were then immersed in a sodium dodecyl sulfate (SDS) solution for six days, with the surfactant promoting a more uniform deposition. The resulting layer was annealed at 105 °C for 10 minutes. Next, a silica-based mask was only deposited in the interstitial spaces between nanospheres, using a diluted tetraethyl orthosilicate (TEOS) solution (prepared with TEOS, 0.1 M HCl, and ethanol). The solution was spin-coated at 2000 rpm for 5 s and subsequently dried at 60 °C for 10 minutes. Selective removal of the nanospheres was achieved by rinsing with toluene for 3 minutes, followed by ethanol washing and nitrogen drying. Finally, a second thermal annealing at 100 °C for 15 minutes was carried out. After these steps, hydrothermal growth of ZnO nanopillar was performed in sealed vials containing $\text{Zn}(\text{NO}_3)_2$ and 50 mM hexamethylenetetramine (HMTA) at 60°C for varying times. For the unmasked sample, the fiber was directly subjected to the hydrothermal growth of ZnO nanopillar in sealed vials containing $\text{Zn}(\text{NO}_3)_2$ and 50 mM HMTA at 60°C.

Gold nanopillar structures were consistently obtained by uniformly covering the fibers with a 30 nm-thick gold layer. Subsequently, the terminal ends of the optical fibers, where no SERS nanostructures were deposited, were polished to ensure uniform length and a smooth surface prior to remote SERS measurements. The polishing procedure was carried out sequentially using abrasive pads of decreasing roughness, applying each step for 30 seconds: first with a 9 μm pad, followed by 6 μm , 1 μm , and finally a finisher pad, maintaining a polishing angle of 0.5°. After polishing, the fibers were cleaned in an ultrasonic bath using pure ethanol and deionized water to remove any residual debris before SERS measurements. The fiber length used in all experiments was ~50 cm.

2.4 Functionalization of nanostructured SERS optical fiber with BPT

Biphenyl-4-thiol (BPT, Sigma-Aldrich, St. Louis, MO, USA) was dissolved in ethanol to obtain a 1 mM solution. The optical fiber probe tips were functionalized by immersion in this solution at 4 °C overnight, followed by gentle rinsing with pure ethanol to remove unbound molecules. This procedure ensures the formation of a uniform self-assembled monolayer of BPT on the optical fiber tips.

2.5 SERS analysis

SERS analysis was conducted by using a LabRAM HR Nano (Horiba Jobin Yvon S.A.S., Longjumeau Cedex, France). For the SERS measurements, the functionalized optical fiber probe was mounted in a fiber holder and fixed on a stage adapter for spectral acquisition. Both 633 nm and 785 nm excitation lasers were employed to investigate the SERS response under two configurations: direct and indirect (remote) interrogation. In particular, in the direct configuration, the laser was focused directly onto the SERS-active fiber tip, and the Raman scattering was collected externally through the same objective. Conversely, in the remote configuration, the excitation beam was focused by the same objective into the opposite end of the optical fiber, allowing the laser to propagate through the fiber and excite the nanostructured tip, with the scattered signal then guided back along the fiber.

In the direct mode, measurements were performed using a 50× objective (Olympus, Tokyo, NA= 0.5), with optimized acquisition parameters of 1 s integration time, single accumulation, 300 μm hole size and a spectral window from 600 to 1800 cm^{-1} . Laser powers were set to 25% (1.05 mW) at 633 nm and 10% (2.08 mW) at 785 nm. The collected backscattered signal was directed onto a diffraction grating with 300 lines per millimeter, which dispersed the light into its spectral components. The resulting spectra were then captured by a Sincerity deep-cooled CCD detector, maintained at $-60\text{ }^{\circ}\text{C}$. SERS maps were carried out acquiring $100 \times 100\text{ }\mu\text{m}^2$ regions with a step size of 3 μm , resulting in 1156 collected spectra for each sample.

In the indirect mode, the setup comprised a 10× objective (Olympus, Tokyo, NA= 0.25) with an excitation power of 100% (7.05 mW for 633 nm and 30.1 mW for 785 nm), a 600 μm hole, 1 s as integration time and single accumulation. In this configuration, three single point spectral measurements were collected for each sample. Calibration of the system was performed with a silicon standard at 520.7 cm^{-1} before SERS measurement.

2.6 SERS data pre-processing

All SERS spectra were pre-processed by removing randomly spikes and performing baseline correction with a sixth-order polynomial fit (up to 200 points) to achieve reliable background subtraction. These operations were carried out using the Labspec 6 software package.

3. Results and discussion

3.1 Characterization of the optical properties of ZnO thin films deposited on optical fibers

Profilometer analysis showed that the ZnO thin film deposited on fused silica sample reached an average thickness of $28.7 \pm 0.05\text{ nm}$ (**Figure 4**), which is very close to the intended 30 nm, demonstrating the high precision of electron beam evaporation.

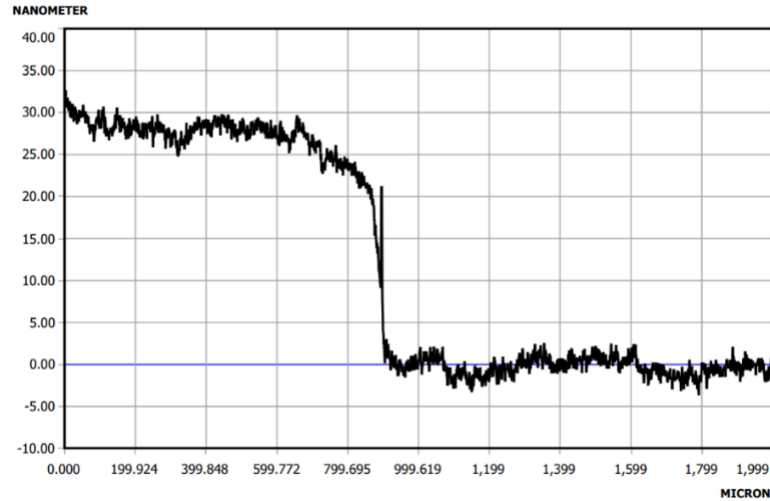


Figure 4. ZnO film thickness measurement with profilometer.

Spectroscopic ellipsometry offered a deeper understanding of the ZnO film properties. For the samples deposited on fused silica, the model fitting indicated a thickness of 28.1 nm with a surface roughness of 1.6 nm, and the refractive index values obtained were in agreement with reported literature when using a fixed bandgap energy of 3.15 eV [30], [31]. The refractive index of zirconia ceramic, used as a substrate for the fiber samples, was also successfully determined and matched values from established dispersion relations, confirming its suitability as a base material [32]. The fused silica and ceramic fiber refractive index estimation and ZnO extinction coefficient were reported in **Figure 5a-b**, respectively.

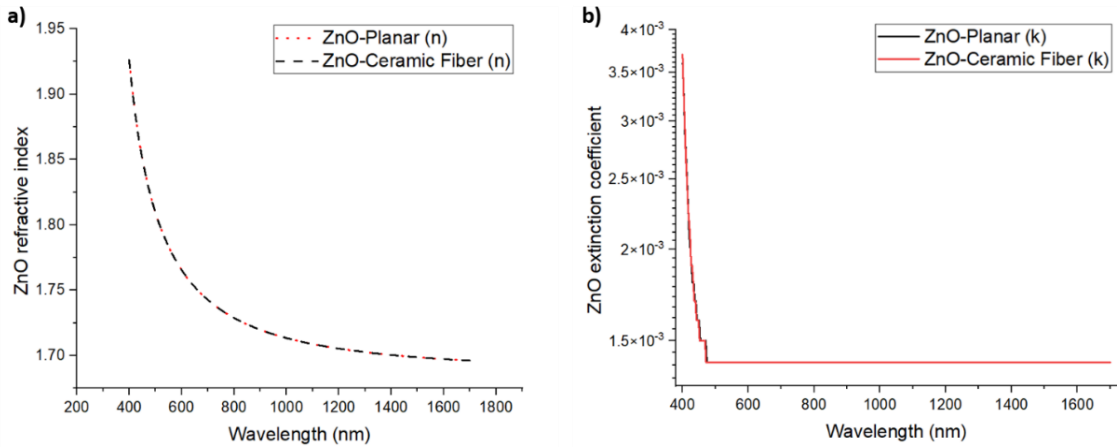


Figure 5. a) Refractive index characterization of fused silica and ceramic fiber, with b) extinction coefficient analysis of ZnO thin films.

In particular, for ZnO films on ceramic fiber tips, ellipsometric fitting revealed a thickness of approximately 32.5 nm and a surface roughness of 1.4 nm, which highlighted slight deviations compared to planar samples, largely due to substrate geometry and material differences. Reflection spectroscopy supported these results by showing interference fringes that confirmed the presence and uniformity of the thin film layers, as reported in **Figure 6**.

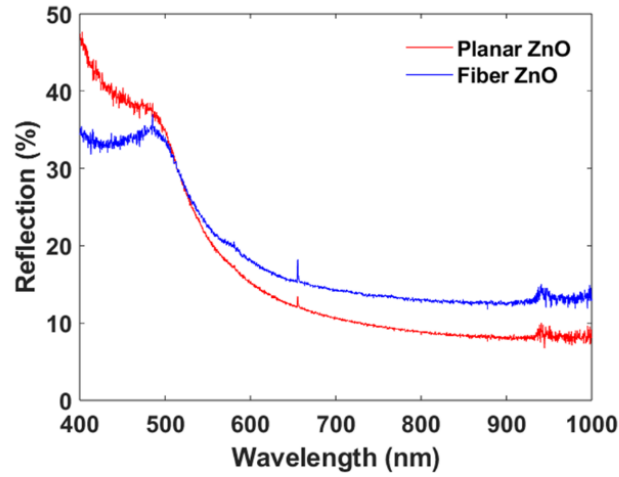


Figure 6. Reflection spectroscopy analysis of ZnO thin film on planar and fiber substrates.

Furthermore, distinct spectral features were observed near the absorption edge of ZnO at around 395 nm, corresponding to its bandgap, which verified the optical quality of the films. Overall, the findings confirmed the reproducibility of the deposition method while illustrating the influence of substrate choice on the final optical properties of the ZnO coatings.

3.2 SEM of SERS optical fiber architectures

SEM analysis of the SERS optical fibers revealed clear morphological variations among the unmasked and masked samples. In the unmasked fiber, ZnO nanopillars grown hydrothermally appeared relatively dense, though their spatial distribution retains a degree of randomness, as shown in **Figure 7**.

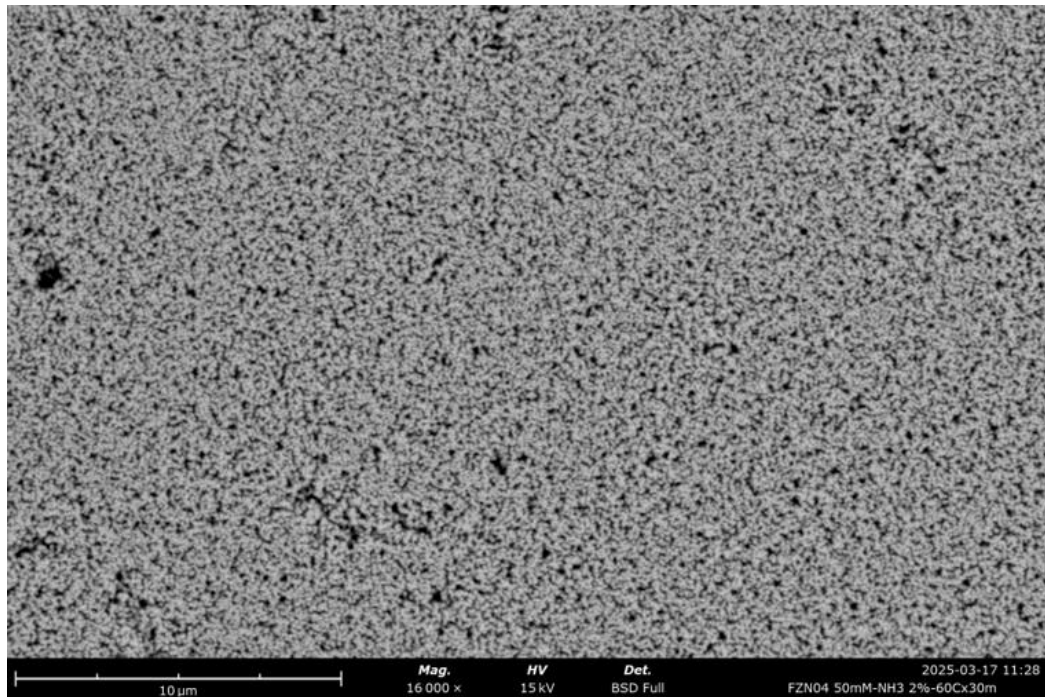


Figure 7. SEM micrograph of the unmasked sample.

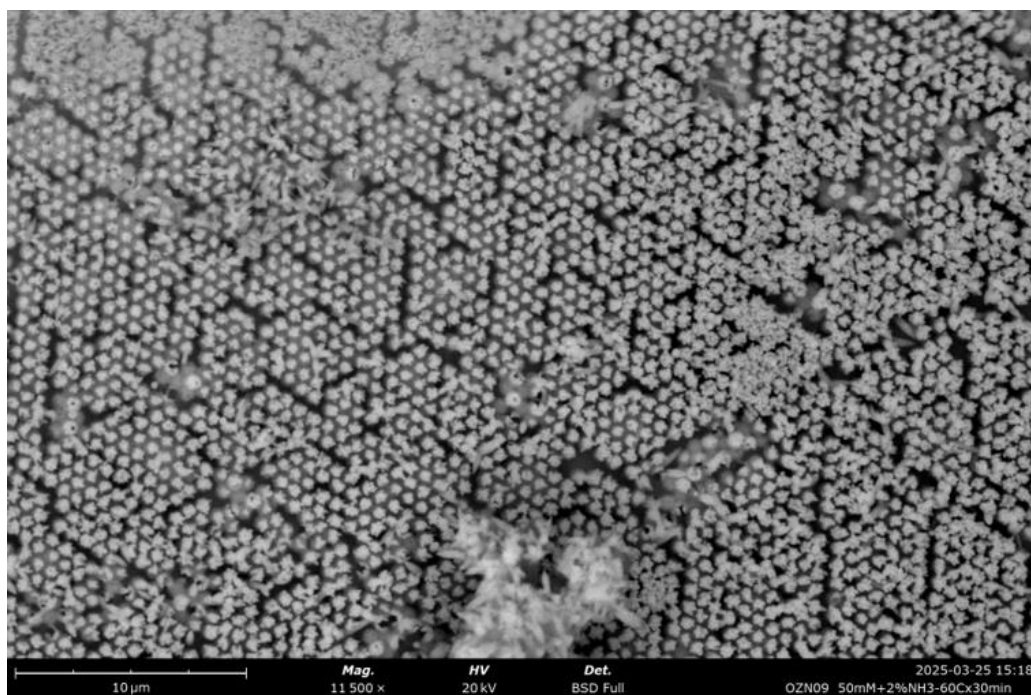


Figure 8. SEM micrograph of a masked fiber sample.

By contrast, masked fibers exhibit a more ordered arrangement of nanopillars, as the masking limits nucleation to defined regions, producing localized arrays of nanostructures with clearer separation, as reported in **Figure 8**. Taken together, these results show that unmasked fibers favor dense but less uniform growth, while masking improves structural order.

3.3 Assessment of SERS optical fiber performance under excitation with two distinct laser wavelengths

SERS characterization of optical fibers integrated with various Au-coated ZnO nanopillar nanostructures and functionalized with BPT molecules, were analyzed in both direct and indirect configurations using 785 and 633 nm excitation wavelengths, offered a comprehensive understanding of their plasmonic behavior. This approach investigated how the interplay between different fabrication steps, nanostructure ordering and excitation wavelength governs the overall enhancement efficiency, spectral response, and reproducibility of the different SERS probes. The variations of these conditions revealed the synergistic effects that determine the plasmonic performance of fiber-based SERS platforms. For the practical deployment of SERS technology, one of the most important challenges lies in reliably controlling the spatial distribution of plasmonic hot spots over large substrate areas, essential to ensure both high sensitivity and reproducibility of Raman signals. The optical behavior of metallic nanostructures is highly dependent on their morphological parameters, such as size and shape, as well as on the physical and chemical properties of their surrounding environment [33]. In this context, a key distinction can be made between ordered nanopillar arrays, produced through masked growth, and disordered nanopillars, resulting from unmasked growth. At 785 nm in the direct configuration, the average SERS spectra of BPT-functionalized SERS fibers are presented in **Figure 9a–b**.

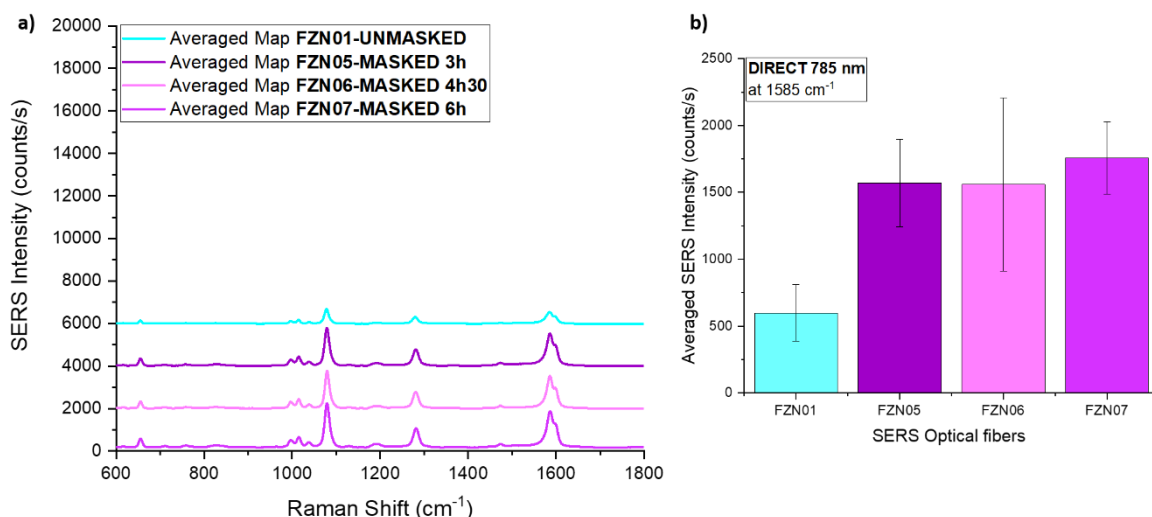


Figure 9. a) Average SERS spectra of BPT obtained from SERS map ($100 \times 100 \mu\text{m}^2$) collected in direct configuration at 785 nm from each SERS-active substrates onto the tip end of unmasked and masked samples. Vertical offsets were applied to the SERS spectra for clearer visualization. b) Statistical distribution of BPT SERS intensities at 1585 cm^{-1} for all fiber types, with error bars representing the standard deviation. Each sample is represented by a distinct color for unambiguous identification.

BPT, an alkane thiol, is used in SERS studies because it forms self-assembled monolayers (SAMs) on metal surfaces. Thiophenol is especially favored for gold or silver due to strong S–Au/S–Ag bonds, well-ordered SAMs, and high SERS activity from both the benzene ring’s Raman cross-section and chemical enhancement [34]. BPT is widely employed as a model molecule for assessing SERS substrates due to its well-defined and reproducible Raman features. In **Figure 9a**, its characteristic bands appear at 1080 cm^{-1} , corresponding to C–H bending vibrations, and at 1281 and 1585 cm^{-1} , associated with ring C=C stretching modes, with some additional minor peaks [35]. These distinct signals make BPT an ideal probe for evaluating both the enhancement efficiency and spectral identification of SERS-active surfaces.

Compared with all other samples, the masked FZN07 SERS fiber exhibited enhanced SERS intensity along with improved signal reproducibility, highlighting its superior structural order relative to other masked samples and the disordered FZN01 fiber. Notably, FZN07 (6 h) showed the strongest SERS response among the masked fibers, outperforming FZN05 (3 h) and FZN06 (4 h 30 min). This trend indicated that extended masked growth promoted optimal ordering of ZnO nanopillars and the formation of stable plasmonic hotspots. This observation was further supported by the statistical distribution of SERS intensities for the Raman band at 1585 cm^{-1} , as shown in **Figure 9b**. These findings suggested that prolonged hydrothermal growth yields ZnO nanopillars with greater structural uniformity which, after gold coating, generate a higher density of well-defined and reproducible plasmonic hotspots, thereby enhancing SERS performance compared to fibers grown for shorter durations.

Moreover, **Table 1** showed the intensities and standard deviations of the characteristic BPT Raman band at 1585 cm^{-1} for each SERS optical fiber measured in direct configuration under excitation wavelengths of 785 nm. The evaluation of both parameters allowed for a clearer differentiation of substrate performance and highlights how structural ordering at the basis of different plasmonic responses.

Table 1. The average intensity values and corresponding standard deviations of the BPT Raman band at 1585 cm^{-1} were reported for each SERS optical fiber in direct configuration using a 785 nm excitation wavelength.

DIRECT 785 NM		
SERS Optical fibers	Average Intensity (counts/s)	Standard Deviation
FZN01-UNMASKED	598.63	212.62
FZN05-MASKED 3h	1569.44	325.76
FZN06-MASKED 4h30	1559.51	647.06
FZN07-MASKED 6h	1756.95	271.10

Additionally, the SERS performance of Au–ZnO nanostructures integrated onto the fiber tip were also evaluated under 633 nm excitation. As shown in **Figure 10a**, the complete set of SERS spectra obtained with this excitation source provides a comprehensive overview of the spectral features across all samples.

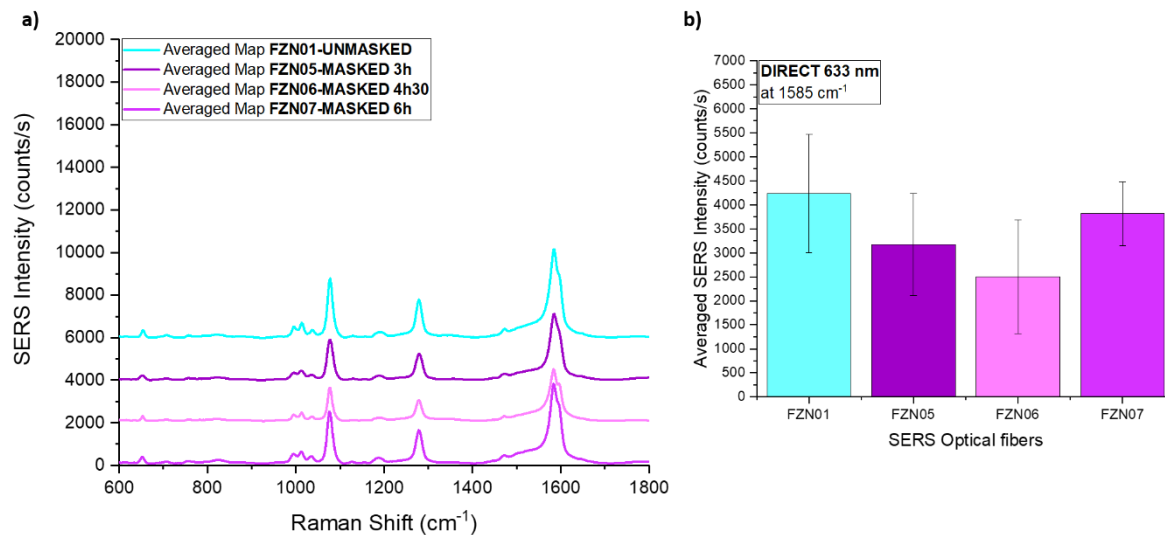


Figure 10. a) Average SERS spectra of BPT obtained from SERS map ($100 \times 100 \mu\text{m}^2$) collected in direct configuration at 633 nm from each SERS-active substrates onto the tip end of all unmasked and masked samples. Vertical offsets were applied to the SERS spectra for clearer visualization. b) Statistical distribution of BPT SERS intensities at 1585 cm^{-1} for all fiber types, with error bars representing the standard deviation. Each sample is represented by a distinct color for unambiguous identification.

At 633 nm excitation, both ordered and disordered arrays generated strong SERS signals, though fabrication differences became more critical in determining performance. Notably, FZN01 exhibited the highest signal intensity among all samples, despite its disordered nanostructure, although its performance was limited by low reproducibility, as reflected by a higher standard deviation. Furthermore, the masked samples followed a trend consistent with the results obtained at 785 nm: FZN07 continued to outperform FZN05 and FZN06, confirming that extended growth time leads to more ordered nanopillar arrays and stronger SERS enhancement. **Figure 10b** illustrated the distribution of peak intensities at 1585 cm^{-1} across all samples, providing a direct comparison of their relative performance. Among all tested samples, FZN07 demonstrated the best overall performance, successfully combining

high SERS enhancement and reproducibility at 633 nm. In **Tables 2** were summarized the mean intensities and standard deviations values of the characteristic BPT Raman band at 1585 cm^{-1} for each SERS optical fiber measured in direct configuration under excitation wavelengths of 633 nm.

Table 2. The average intensity values and corresponding standard deviations of the BPT Raman band at 1585 cm^{-1} were reported for each SERS optical fiber in direct configuration using a 633 nm excitation wavelength.

DIRECT 633 NM		
SERS Optical fibers	Average Intensity (counts/s)	Standard Deviation
FZN01-UNMASKED	4238.54	1231.20
FZN05-MASKED 3h	3173.56	1066.07
FZN06-MASKED 4h30	2503.18	1187.04
FZN07-MASKED 6h	3817.92	664.57

Although the measurements with the two different excitation wavelengths in the configuration were carried out in the same way using identical parameters—except for the power, which was adjusted to prevent damage to the nanostructures—excitation at 633 nm resulted in higher recorded SERS intensities. Evidently this improvement can be explained cause 633 nm is closer to the peak of the plasmon resonance compared to 785 nm. These results underscore the importance of tuning the excitation wavelength to align with the plasmon resonance for maximizing SERS performance [36]. Additionally, because the scattering intensity depends on the inverse fourth power of the excitation wavelength, shorter wavelengths provide stronger signals. Consequently, excitation at 633 nm yields substantially higher scattering intensity compared to excitation at 785 nm [37]. Interestingly, although the same experimental parameters were applied in the direct configuration for both lasers, with powers of 2.08 mW at 785 nm and 1.05 mW at 633 nm, the 633 nm laser—despite operating at lower power—produced stronger SERS signals. Moreover, masked FZN07 fiber sample showed the best reproducibility, as confirmed by analyses with both lasers.

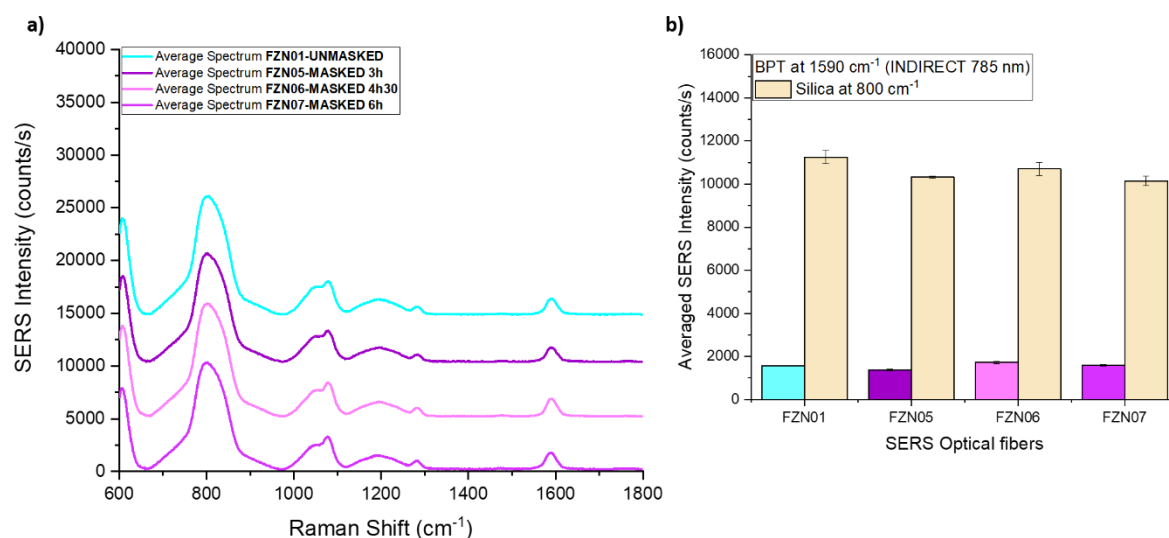


Figure 11. a) Averaged SERS spectra of BPT, obtained from three single-point measurements acquired in the indirect configuration at 785 nm, are shown for each unmasked and masked SERS-active substrate integrated at the fiber tip. Vertical offsets have been applied to the spectra to aid visual clarity. b) Statistical analysis of SERS signal intensities at 1590 cm^{-1} (BPT) and at 800 cm^{-1} (silica reference band) across all fiber types. Error bars represent standard deviations, while distinct colors are used to clearly differentiate between the sample groups.

Subsequently, the SERS performance of all fibers was evaluated in the indirect configuration at 785 nm and 633 nm, in which the excitation light first propagates through the fiber before reaching the SERS substrate and subsequently interacting with the BPT molecules. However, in the indirect configuration some prominent silica Raman bands appeared in SERS spectra, approximatively at 800, 1050 and 1195-1200 cm^{-1} , with the peak at 800 cm^{-1} more intense than the other silica bands [38]. The results of SERS analysis in the remote configuration with 785 nm were reported in **Figure 11a-b**.

At 785 nm in the indirect configuration, fibers with longer masked growth, such as FZN06 and FZN07, generated stronger SERS signals than the shorter-growth FZN05, even under remote conditions. Notably, the masked FZN06 fiber, with its well-ordered architecture, exhibited higher SERS intensity than the disordered unmasked FZN01 fiber (**Figure 11b**, **Table 3**). FZN06 showed the highest overall SERS performance among the masked fibers, emphasizing how structural ordering enhances SERS signals compared to disordered architectures.

Table 3. The average intensity values and corresponding standard deviations of the BPT Raman band at 1590 cm^{-1} were reported for each SERS optical fiber in indirect configuration using a 785 nm excitation wavelength.

INDIRECT 785 NM		
SERS Optical fibers	Average Intensity	Standard Deviation
FZN01-UNMASKED	1563.91	20.31
FZN05-MASKED 3h	1365.97	32.88
FZN06-MASKED 4h30	1720.84	53.57
FZN07-MASKED 6h	1590.80	34.61

In the indirect configuration at 633 nm, the silica background was particularly intense (up to ~20k counts/s), partially masking BPT signals, while at 785 nm it was weaker (~10–12k counts/s), allowing clearer detection despite the lower plasmonic efficiency (**Figure 12a-b**).

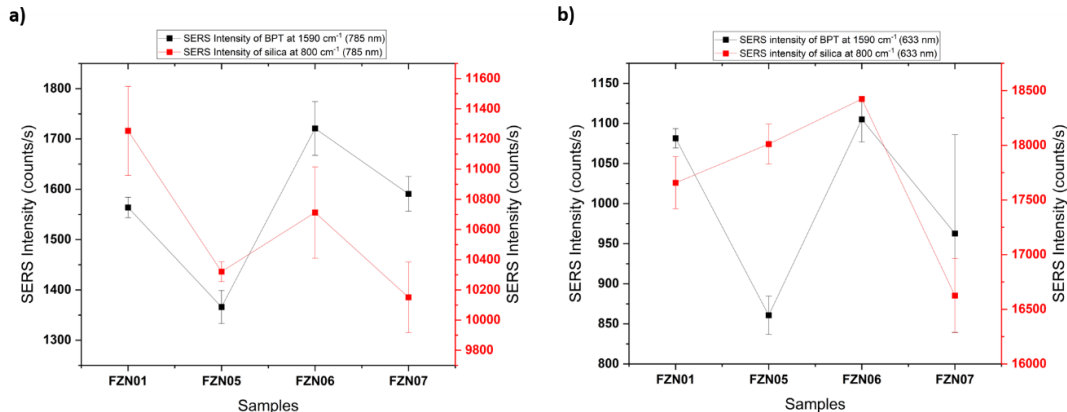


Figure 12. Evaluation of the silica interference on the SERS performance of all SERS fibers. a) Distribution of SERS intensities of BPT and silica at 1590 (black square) and 800 cm^{-1} (red square) at 785 nm and b) 633 nm excitation wavelength. Error bar represent standard deviations.

In particular, it is important to note that in the indirect configuration as well, the same parameters were used to perform the SERS analysis, employing the maximum power for both lasers, although this corresponded to different absolute power values. Furthermore, under a 633 nm laser, the unmasked fiber FZN01 and the masked fiber FZN06 showed comparable performance in terms of SERS intensity, as reported in **Figure 13**. The BPT SERS intensities accompanied by the relative standard deviation across all fiber samples were described in **Table 4**.

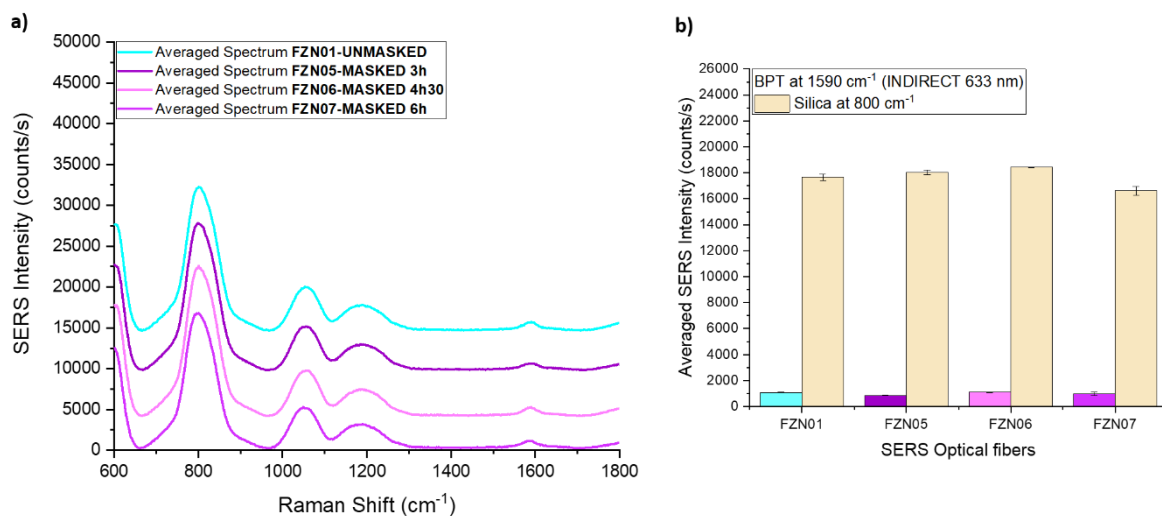


Figure 13. a) Averaged SERS spectra of BPT, obtained from three single-point measurements acquired in the indirect configuration at 633 nm, for each unmasked and masked. Vertical offsets have been applied to the spectra to aid visual clarity. b) Statistical analysis of SERS signal intensities at 1590 cm^{-1} (BPT) and at 800 cm^{-1} (silica reference band) across all fiber types. Error bars represent standard deviations, while distinct colors are used to clearly differentiate between the sample groups.

These results indicated that fiber performance depended critically on both structural design and excitation wavelength. For direct interrogation at 785 nm, FZN07 (ordered, masked for 6 hours) emerged as the fiber with the best SERS performances, benefiting from their optimized structural uniformity and consistent SERS hotspots, which enhance reproducibility across multiple measurements.

Table 4. The average intensity values and corresponding standard deviations of the BPT Raman band at 1590 cm^{-1} were reported for each SERS optical fiber in indirect configuration using a 633 nm excitation wavelength.

INDIRECT 633 NM		
SERS Optical fibers	Average Intensity	Standard Deviation
FZN01-UNMASKED	1081.44	12.15
FZN05-MASKED 3h	860.74	23.97
FZN06-MASKED 4h30	1104.99	28.02
FZN07-MASKED 6h	962.78	123.36

At 633 nm, the masked sample FZN07 outperformed all other samples, demonstrating that the synergy between plasmonic resonance and ordered architecture enhances signal reproducibility. This configuration achieves an optimal balance between high SERS intensity and uniformity across the substrate.

In remote mode, although identical parameters were applied in both configurations, including 100% laser power, this value corresponds to different actual outputs depending on the laser used (30.1 mW for 785 nm and 7.05 mW for 633 nm). Despite operating at a higher power, the 785 nm laser suppressed the silica background and produced stronger SERS signals from BPT, in particular, by employing the masked FZN06 SERS-based fiber.

These findings highlight that optimizing growth duration and excitation wavelength is key to producing reliable, high-performance SERS fiber probes with both strong enhancement and reproducible responses.

3.4 Comparison direct and indirect configuration for SERS optical fibers

Subsequently, the SERS performance of the optical fibers were compared in the direct and indirect configuration under the respective optimized experimental conditions. For the initial direct measurements, a 50× objective, a confocal pinhole of 300 μm aperture were used, applying laser powers of 2.08 mW for 785 nm and 1.05 mW for 633 nm, respectively. In contrast, the indirect measurements employed a 10× objective and a pinhole with a larger 600 μm diameter, along with higher laser powers of 30.1 mW for 785 nm and 7.05 mW for 633 nm to compensate for the longer optical path and ensure sufficient excitation at the SERS substrate.

The comparison of SERS intensities obtained under direct and indirect interrogation modes at 785 nm and 633 nm excitation wavelengths reveals distinct performance trends across the different optical fibers. At 785 nm, both direct and indirect configurations yielded comparable performance although the masked configurations exhibit slight better intensities (**Figure 14a**).

At 633 nm, the difference in SERS intensities at 1585 cm^{-1} between direct and indirect configurations was more marked than at 785 nm (**Figure 14b**). The lower efficiency of the indirect configuration at this wavelength was likely due to a reduced collection efficiency of the optical fiber and the related worse coupling with the microscope objective.

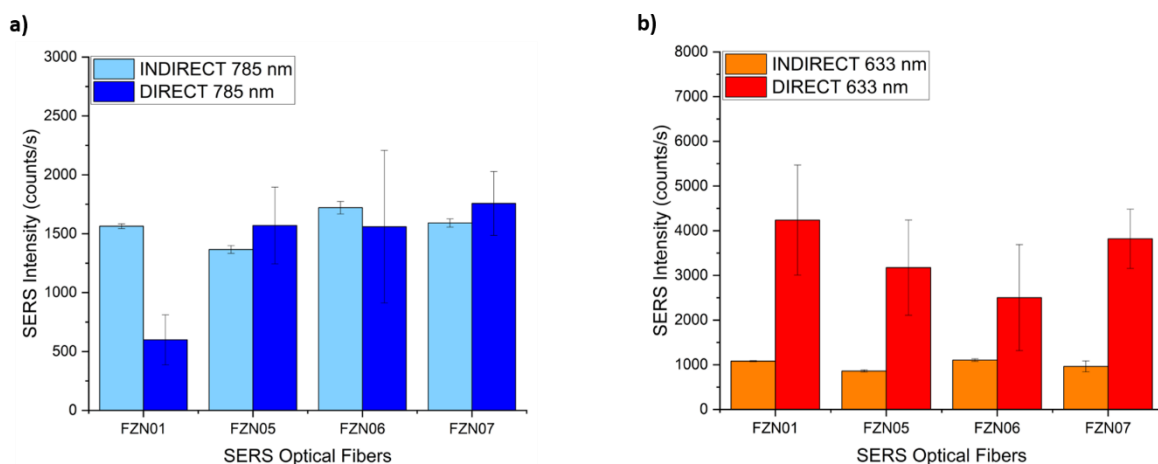


Figure 14. Comparison of BPT SERS intensities for various SERS optical fibers under direct and indirect interrogation. a) Plot of intensities distribution at 1585 cm^{-1} at 785 nm excitation in indirect and direct configuration. b) Plot of intensities distribution at 1585 cm^{-1} in indirect and direct configuration at 633 nm. Error bars represent the standard deviations.

Overall, the use of the optical fiber (indirect configuration) leads to a maximum reduction factor of about 3-4-fold with respect to the SERS-active substrate (direct configuration). Such reduction is substantially congruent with similar works [38] and it can be considered acceptable considering the unrivalled application the optical fiber tip can afford.

4. Conclusions

In this chapter, the systematic investigation of Au-coated ZnO nanopillar arrays integrated onto optical fiber tips has provided a comprehensive understanding of how fabrication strategies, structural ordering and excitation wavelength collectively influence SERS performance. By comparing masked and unmasked hydrothermal growth approaches, the study demonstrated that both morphological control is decisive factor in generating reproducible plasmonic hotspots and stable enhancement.

In particular, masked fibers grown for longer durations exhibited superior ordering of nanopillars (lower standard deviation in the direct configuration intensities), which translated into higher signal reproducibility and intensity at 785 and 633 nm excitation in direct interrogation. Conversely, the unmasked fiber exhibited lower SERS performance than the masked FZN07 fiber in direct mode at 785 nm, and lower reproducibility than the masked FZN06 fiber at 633 nm. The interplay of these effects revealed that ordered arrays maximize reproducibility compared with disordered platform.

The influence of excitation wavelength was also decisive. Direct interrogation at 633 nm consistently produced higher SERS intensities than at 785 nm due to resonance alignment and the λ^{-4} dependence of scattering efficiency, though silica background limited its utility for remote sensing. Nonetheless, indirect interrogation at 785 nm proved more favorable for practical applications, as it minimized interference from silica background signals while still enabling robust enhancement. This highlights the importance of matching the excitation wavelength not only to the plasmon resonance but also to the optical environment of the fiber probe.

Overall, the comparative analysis revealed that no single fiber architecture dominates under all conditions. Instead, optimal performance depends on the sensing modality and experimental configuration. Ordered masked fibers, such as FZN07 with extended growth, delivered the most consistent performance at 785 nm in direct mode in terms of SERS intensity and reproducibility.

Conversely, the unmasked FZN01 fiber exhibited the highest SERS intensity at 633 nm, while the masked FZN07 fiber demonstrated superior reproducibility with comparable signal strength. Under indirect interrogation, FZN06 consistently exhibited the highest SERS intensity at both 785 nm and 633 nm, with performance comparable to the unmasked sample, especially at 633 nm. The SERS optical fiber best performance were summarized in the **Table 5**.

By carefully tuning structural ordering and excitation conditions, it is possible to tailor fiber-based optrodes for different operational requirements.

Table 5. Summary of best-performing SERS optical fiber for each configuration and excitation wavelengths.

Configuration	Laser (nm)	Best-Performing Fibers	Key Features of Performance
Direct	785	FZN07 (Masked, 6 h)	FZN07 shows the highest SERS intensity and reproducibility
Direct	633	FZN07 (Masked, 6 h)	FZN07 provides the best balance between the SERS intensity and reproducibility
Indirect	785	FZN06 (Masked, 4 h 30 min)	FZN06 exhibits the highest BPT-to-silica signal ratio
Indirect	633	FZN01 (Unmasked 3 h)/ FZN06 (Masked, 4 h 30 min)	FZN01 and FZN06 showed comparable results in terms of BPT-to-silica signal ratio

This adaptability represents a crucial step toward the realization of robust, reproducible, and scalable lab-on-fiber SERS devices, suitable for real-world biomedical diagnostics.

Beyond these findings, several critical factors emerge for practical SERS fiber applications. The mechanical stability and adhesion of nanostructures to the fiber surface are essential for long-term use, especially in *in vivo* or flow-through sensing [39]. Fiber geometry and tip functionalization significantly affect light coupling and hotspot accessibility, influencing both sensitivity and uniformity [38]. Background suppression and spectral clarity remain key for accurate analyte detection, suggesting a role for optical filtering or advanced signal processing. Additionally, reproducibility and scalability of fabrication are vital, as small variations in nanostructure morphology or chemical treatment can lead to substantial differences in SERS performance [40]. Then, further improvements will be necessary to enhance the SERS performance of the developed optical fibers, including refining the treatment conditions, and ensuring greater reproducibility and stability of the SERS signal under varying experimental conditions.

Overall, these results emphasize the need to carefully balance nanostructural ordering, chemical treatments, excitation parameters, and practical fiber considerations to optimize SERS-active optical fibers, enabling their effective integration with real-time detection capabilities. By integrating these features, such platforms have the potential to revolutionize biomedical diagnostics, enabling earlier and more reliable detection of disease markers, while also facilitating continuous, minimally invasive monitoring in clinical settings. Their inherent miniaturization—achieved through integration onto the tips of optical fibers—makes them particularly well-suited for *in vivo* applications, where compact, user-friendly devices are essential. Looking ahead, the integration of these technological advances will be expected to accelerate the translation of lab-on-fiber systems from proof-of-concept studies into practical, deployable tools for precision medicine and beyond.

5. References

- [1] R. K. Gangwar *et al.*, “Optical fiber SERS sensors: Unveiling advances, challenges, and applications in a miniaturized technology,” *Coord. Chem. Rev.*, vol. 510, no. April, p. 215861, 2024, doi: 10.1016/j.ccr.2024.215861.

- [2] P. R. Stoddart and D. J. White, "Optical fibre SERS sensors," *Anal. Bioanal. Chem.*, vol. 394, no. 7, pp. 1761–1774, 2009, doi: 10.1007/s00216-009-2797-6.
- [3] A. Ricciardi *et al.*, "Lab-on-fiber technology: A new vision for chemical and biological sensing," *Analyst*, vol. 140, no. 24, pp. 8068–8079, 2015, doi: 10.1039/c5an01241d.
- [4] M. Pisco and A. Cusano, "Lab-on-fiber technology: A roadmap toward multifunctional plug and play platforms," *Sensors (Switzerland)*, vol. 20, no. 17, pp. 1–20, 2020, doi: 10.3390/s20174705.
- [5] M. Consales, A. Ricciardi, A. Crescitelli, E. Esposito, A. Cutolo, and A. Cusano, "Lab-on-fiber technology: Toward multifunctional optical nanoprobe," *ACS Nano*, vol. 6, no. 4, pp. 3163–3170, 2012, doi: 10.1021/nn204953e.
- [6] P. Vaiano *et al.*, "Lab on Fiber Technology for biological sensing applications," *Laser Photonics Rev.*, vol. 10, no. 6, pp. 922–961, 2016, doi: 10.1002/lpor.201600111.
- [7] B. Du, Y. Xu, L. Zhang, and Y. Zhang, "Plasmonic Functionality of Optical Fiber Tips: Mechanisms, Fabrications, and Applications," *Materials (Basel)*, vol. 16, no. 9, 2023, doi: 10.3390/ma16093596.
- [8] L. Meng *et al.*, "Optical fiber optrodes with silver-coated gold nanocavity ordered arrays for highly sensitive surface enhanced Raman spectrum," *Sensors Actuators B Chem.*, vol. 380, no. August 2022, p. 133314, 2023, doi: 10.1016/j.snb.2023.133314.
- [9] G. Kostovski, P. R. Stoddart, and A. Mitchell, "The optical fiber tip: An inherently light-coupled microscopic platform for micro- and nanotechnologies," *Adv. Mater.*, vol. 26, no. 23, pp. 3798–3820, 2014, doi: 10.1002/adma.201304605.
- [10] R. Huang *et al.*, "High-sensitivity and throughput optical fiber SERS probes based on laser-induced fractional reaction method," *Results Phys.*, vol. 48, no. January, p. 106410, 2023, doi: 10.1016/j.rinp.2023.106410.
- [11] A. Dhawan, M. Gerhold, and T. Vo-Dinh, "Theoretical Simulation and Focused Ion Beam Fabrication of Gold Nanostructures For Surface-Enhanced Raman Scattering (SERS)," *Nanobiotechnology J. Intersect. nanotechnology, Mol. Biol. Biomed. Sci.*, vol. 3, no. 3–4, pp. 164–171, Dec. 2007, doi: 10.1007/s12030-008-9017-x.
- [12] A. Foti *et al.*, "A Shape-Engineered Surface-Enhanced Raman Scattering Optical Fiber Sensor Working from the Visible to the Near-Infrared," *Plasmonics*, vol. 8, no. 1, pp. 13–23, 2013, doi: 10.1007/s11468-012-9371-3.
- [13] X. Ma *et al.*, "Surface-enhanced Raman scattering sensor on an optical fiber probe fabricated with a femtosecond laser," *Sensors*, vol. 10, no. 12, pp. 11064–11071, 2010, doi: 10.3390/s101211064.
- [14] A. Bouvrée *et al.*, "Nanostructured and nanopatterned gold surfaces: Application to the surface-enhanced Raman spectroscopy," *Gold Bull.*, vol. 46, no. 4, pp. 283–290, 2013, doi: 10.1007/s13404-013-0127-4.
- [15] G. Kostovski, D. J. White, A. Mitchell, M. W. Austin, and P. R. Stoddart, "Nanoimprinted optical fibres: Biotemplated nanostructures for SERS sensing," *Biosens. Bioelectron.*, vol. 24, no. 5, pp. 1531–1535, 2009, doi: 10.1016/j.bios.2008.10.016.
- [16] X. Zhang *et al.*, "Remote Chemical Sensing by SERS with Self-Assembly Plasmonic Nanoparticle Arrays on a Fiber," *Front. Phys.*, vol. 9, no. January, pp. 1–6, 2022, doi: 10.3389/fphy.2021.752943.
- [17] Z. Huang *et al.*, "Tapered Optical Fiber Probe Assembled with Plasmonic Nanostructures for Surface-Enhanced Raman Scattering Application," *ACS Appl. Mater. Interfaces*, vol. 7, no. 31, pp. 17247–17254, 2015, doi: 10.1021/acsami.5b04202.
- [18] J. A. Kim, D. J. Wales, A. J. Thompson, and G. Z. Yang, "Fiber-Optic SERS Probes Fabricated Using Two-Photon Polymerization For Rapid Detection of Bacteria," *Adv. Opt. Mater.*, vol. 8, no. 9, pp. 1–12, 2020, doi: 10.1002/adom.201901934.
- [19] Y. Xiong and F. Xu, "Multifunctional integration on optical fiber tips: Challenges and opportunities," *Adv. Photonics*, vol. 2, no. 6, pp. 1–24, 2020, doi: 10.1117/1.AP.2.6.064001.
- [20] J. Liu, B. Zhang, A. Abdullah, S. A. Muhsin, J. Huang, and M. Almasri, "A Fiber optics based surface enhanced Raman spectroscopy sensor for chemical and biological sensing," *Sens. Bio-Sensing Res.*, vol. 46, no. April, p. 100686, 2024, doi: 10.1016/j.sbsr.2024.100686.

- [21] M. Bekeris *et al.*, “Rapid Quantification of Nanosphere Lithography Packing Defects Using Scanning Electron Microscopy Edge Effects,” *Phys. Status Solidi - Rapid Res. Lett.*, vol. 14, no. 11, 2020, doi: 10.1002/pssr.202000328.
- [22] M. Pisco *et al.*, “Nanosphere lithography for optical fiber tip nanoprobe,” *Light Sci. Appl.*, vol. 6, no. 5, pp. 1–12, 2017, doi: 10.1038/lsa.2016.229.
- [23] S. Adesoye, S. Al Abdullah, A. Kumari, G. Pathiraja, K. Nowlin, and K. Dellinger, “Au-Coated ZnO Surface-Enhanced Raman Scattering (SERS) Substrates: Synthesis, Characterization, and Applications in Exosome Detection,” *Chemosensors*, vol. 11, no. 11, 2023, doi: 10.3390/chemosensors11110554.
- [24] A. B. Djurišić and Y. H. Leung, “Optical properties of ZnO nanostructures,” *Small*, vol. 2, no. 8–9, pp. 944–961, 2006, doi: 10.1002/sml.200600134.
- [25] D. K. Sharma, S. Shukla, K. K. Sharma, and V. Kumar, “A review on ZnO: Fundamental properties and applications,” *Mater. Today Proc.*, vol. 49, pp. 3028–3035, 2020, doi: 10.1016/j.matpr.2020.10.238.
- [26] M. Bakry, W. Ismail, M. Abdelfatah, and A. El-Shaer, “Low-cost fabrication methods of ZnO nanorods and their physical and photoelectrochemical properties for optoelectronic applications,” *Sci. Rep.*, vol. 14, no. 1, pp. 1–15, 2024, doi: 10.1038/s41598-024-73352-5.
- [27] M. Z. Hossain, S. M. A. Nayem, M. S. Alam, M. I. Islam, G. Seong, and A. N. Chowdhury, “Hydrothermal ZnO Nanomaterials: Tailored Properties and Infinite Possibilities,” *Nanomaterials*, vol. 15, no. 8, 2025, doi: 10.3390/nano15080609.
- [28] T. T. H. Pham *et al.*, “Ag nanoparticles on ZnO nanoplates as a hybrid SERS-active substrate for trace detection of methylene blue,” *RSC Adv.*, vol. 12, no. 13, pp. 7850–7863, 2022, doi: 10.1039/d2ra00620k.
- [29] A. K. Pal, S. Pagal, K. Prashanth, G. K. Chandra, S. Umapathy, and B. M. D., “Ag/ZnO/Au 3D hybrid structured reusable SERS substrate as highly sensitive platform for DNA detection,” *Sensors Actuators, B Chem.*, vol. 279, no. May 2018, pp. 157–169, 2019, doi: 10.1016/j.snb.2018.09.085.
- [30] W. A. M. Harland G. Tompkins, *Spectroscopic ellipsometry and reflectometry: a user’s guide*. 1999. [Online]. Available: <https://www.wiley.com/Spectroscopic+Ellipsometry+and+Reflectometry%3A+A+User’s+Guide-p-9780471181729#relatedProducts-section>
- [31] O. Aguilar, S. de Castro, M. P. F. Godoy, and M. Rebello Sousa Dias, “Optoelectronic characterization of Zn 1-x Cd x O thin films as an alternative to photonic crystals in organic solar cells,” *Opt. Mater. Express*, vol. 9, no. 9, p. 3638, 2019, doi: 10.1364/ome.9.003638.
- [32] D. L. Wood and K. Nassau, “Refractive index of cubic zirconia stabilized with yttria,” *Appl. Opt.*, vol. 21, no. 16, pp. 2978–2981, Aug. 1982, doi: 10.1364/AO.21.002978.
- [33] Y. T. Kim, J. Schilling, S. L. Schweizer, G. Sauer, and R. B. Wehrspohn, “Au coated PS nanopillars as a highly ordered and reproducible SERS substrate,” *Photonics Nanostructures - Fundam. Appl.*, vol. 25, pp. 65–71, 2017, doi: 10.1016/j.photonics.2017.03.006.
- [34] F. Sun, D. D. Galvan, P. Jain, and Q. Yu, “Multi-functional, thiophenol-based surface chemistry for surface-enhanced Raman spectroscopy,” *Chem. Commun.*, vol. 53, no. 33, pp. 4550–4561, 2017, doi: 10.1039/c7cc01577a.
- [35] S. Managò *et al.*, “Tailoring lab-on-fiber SERS optrodes towards biological targets of different sizes,” *Sensors Actuators, B Chem.*, vol. 339, no. July 2020, 2021, doi: 10.1016/j.snb.2020.129321.
- [36] H. Liu, X. Zhang, T. Zhai, T. Sander, L. Chen, and P. J. Klar, “Centimeter-scale-homogeneous SERS substrates with seven-order global enhancement through thermally controlled plasmonic nanostructures,” *Nanoscale*, vol. 6, no. 10, pp. 5099–5105, 2014, doi: 10.1039/c4nr00161c.
- [37] M. Hardy and H. O. M. Chu, “Laser wavelength selection in Raman spectroscopy,” *Anal.*, vol. 150, no. 10, pp. 1986–2008, 2025, doi: 10.1039/d5an00324e.
- [38] G. Quero *et al.*, “Nanosphere lithography on fiber: Towards engineered lab-on-fiber SERS optrodes,” *Sensors (Switzerland)*, vol. 18, no. 3, pp. 1–15, 2018, doi: 10.3390/s18030680.
- [39] H. Chang, “In vivo surface-enhanced Raman scattering techniques: nanoprobe, instrumentation, and

applications,” *Light Sci. Appl.*, pp. 11–22, 2025, doi: 10.1038/s41377-024-01718-5.

- [40] S. R. Tadi *et al.*, “Recent advances in the design of SERS substrates and sensing systems for (bio)sensing applications: Systems from single cell to single molecule detection,” *F1000Research*, vol. 13, pp. 1–90, 2025, doi: 10.12688/f1000research.149263.3.

Chapter 6: Nanoscale Analysis of Cancer Cell Membranes and Cell Surface Biomarker Using Tip-Enhanced Raman Spectroscopy (TERS)

Abstract

Tip-Enhanced Raman Spectroscopy (TERS) is an advanced optical technique that combines the molecular sensitivity of surface-enhanced Raman scattering (SERS) with the nanoscale spatial resolution of scanning probe microscopy, enabling simultaneous high-resolution biochemical and topographical mapping. This chapter outlines the fundamental principles of TERS, its operating mechanisms, and its application in biomedical field. In the next section, a feasibility study base on the analysis of cancer cell membrane is presented in two phases. First, TERS is utilized in a label-free configuration to characterize the native biochemical composition of the cell membrane, providing insight into its molecular organization. In the second phase, TERS is integrated with functionalized SERS tags to enable targeted detection and identification of specific surface biomarker, overexpressed in hepatocellular carcinoma, namely glypican-3 (GPC-3), demonstrating its capability for selective molecular recognition at the nanoscale. By uniting targeted molecular recognition with localized plasmonic field enhancement, this methodology underscores the remarkable capability of TERS for ultrasensitive molecular detection, establishing it as a highly promising tool for cancer diagnostics and advanced cellular analysis.

1. Introduction

All cells have a cell membrane, a dynamic and fluid structure that function as a barrier between the cytoplasm and the extracellular environment, preserving the distinct chemical conditions inside and outside the cell. Additionally, the cell membrane regulates the transport of molecules, enables communication between cells, and transmits electrical and chemical signals through specialized protein receptors [1]. Membranes display a heterogeneous organization, composed of a complex mixture of lipids, proteins, and carbohydrates. Lipids, with their diverse structures and properties, form the bilayer and contribute to its essential functions. Lipids play a key role in directing protein positioning and signaling, thereby regulating processes such as hormone and neurotransmitter release, protein secretion, cell division, differentiation, and gene expression. They can also serve as ligands or substrates for signaling proteins and enzymes, making interactions between proteins and lipids, as well as between lipids themselves, vital for proper cell function [2]. Cancer cells display modified lipid metabolism, resulting in abnormal membrane composition that promotes accelerated growth. Such alterations in lipid profiles may act as biomarkers of cancer and are linked to increased chemotherapy resistance and impaired cellular signaling [3]. Membrane proteins exist in different forms—integral (embedded in the bilayer), peripheral (associated with the membrane surface), and anchoring (attached indirectly through lipids). These proteins interact with lipids via hydrophobic or ionic forces, creating lateral domains that facilitate the transport of ions and molecules, electron transfer, signal transduction, enzymatic activity, and intercellular communication [4]. Despite

accounting for a substantial fraction of all proteins (25–30%) and being critical to cellular function, detailed structural and functional knowledge of membrane proteins remains limited in comparison to other protein classes, reflecting the challenges associated with studying these complex [5]. Alteration in membrane proteins and their associated signaling pathways play a key role in the onset and progression of cancer, making them promising candidates for diagnostic, prognostic, and therapeutic applications [6]. Moreover, cell surfaces are coated with glycoproteins and glycolipids that mediate cell recognition, development, immune interactions, and pathogen binding. Changes in these carbohydrates can disrupt recognition or promote cancer, making their molecular understanding important for biology and medicine [7].

Membrane structure is highly organized at the nanoscale, with functional regions that operate on millisecond timescales, making the study of membranes particularly challenging. As a result, understanding membrane dynamics and organization requires advanced imaging techniques that provide both high spatial and temporal resolution, ideally under conditions that closely mimic those within living organisms. Revealing the nanoscale organization of cellular biomolecules provides crucial insight into their unresolved details, structure–function relationships, cellular pathways, and receptor interactions that underpin the rational design of targeted therapeutics with optimal binding performance or intracellular internalization. Despite this importance, existing analytical methods remain limited in their ability to perform non-destructive, label-free molecular mapping of biological cells with nanoscale spatial resolution. Among current technologies, super-resolution fluorescence microscopy (SRFM) comes closest to meeting these requirements, as it enables non-destructive visualization of molecular interaction, dynamic cellular processes and cellular structure distributions at nanometer scale [8]. However, SRFM depends on the conjugation of specific fluorophores to the target molecules, a process that can alter the molecules’ native behavior and potentially interfere with their biological function. Moreover, SRFM exclusively captures signal from fluorescently labeled molecules, rendering most unlabeled biomolecules invisible, while inaccuracies in molecular quantification caused by fluorophore blinking constitute an additional major limitation of the technique [9], [10].

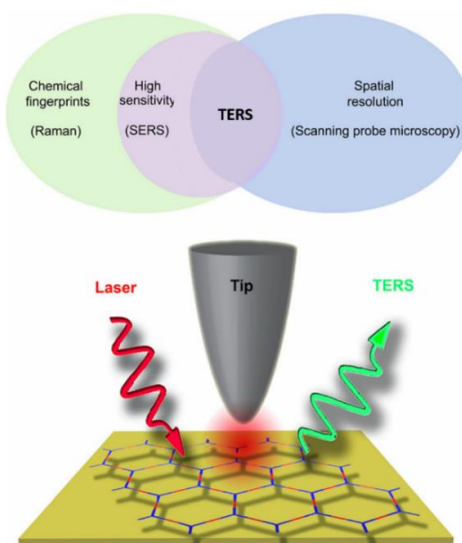


Figure 1. Schematic illustration of the TERS approach, which combines the high sensitivity of SERS—enabled by the metallic tip that amplifies the Raman fingerprints of molecules beneath it—with the high spatial resolution provided by scanning probe microscopy, from Ref. [12].

Scanning probe microscopy techniques, including atomic force microscopy (AFM) and scanning tunneling microscopy (STM), primarily provide topographical and structural information, with no capacity for direct chemical characterization [11]. Despite detailed knowledge of the molecular components of cell membranes, their spatial organization and distribution remain largely poorly understood. To unravel this structural complexity, it is essential to employ a non-invasive approach that can provide molecular-level sensitivity and nanoscale resolution.

Tip-enhanced Raman spectroscopy is an advanced analytical technique that combines the nanoscale spatial precision of scanning probe microscopy, such as AFM, with the strong electromagnetic field enhancement generated by plasmonic nanostructures at the tip apex [13]. This synergistic integration not only amplifies the inherently weak Raman scattering signal but also enables simultaneous acquisition of high-resolution biochemical information and nanoscale topographical mapping of samples, as shown in **Figure 1** [14], [15], [16].

The TERS versatility further enables the analysis of a wide range of materials, making TERS a powerful and adaptable tool across diverse scientific disciplines [17]. In particular, TERS is extensively utilized for material characterization, as reported in **Figure 2**. It enables the investigation of graphene and carbon nanotubes by revealing their structural features, defects, impurities, strain, and doping levels [18], [19]. TERS is also applied to study defects and strain in silicon and other semiconductors [20], [21], analyze polymers [22], develop and refine intensity models [23], [24], and monitor chemical reactions and the formation of self-assembled monolayers [25].

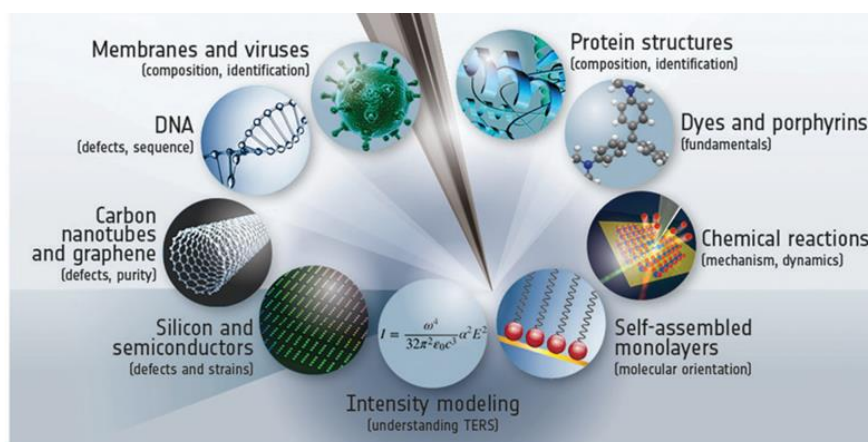


Figure 2. Overview of TERS application, from Ref. [17].

TERS spectroscopy offers remarkable advantages, including its non-invasive approach and the simultaneous acquisition of topographical and chemical information in real time. These features position TERS as a valuable complementary tool in the field of bionanoimaging, with the potential to advance our understanding of cell membrane architecture, molecular composition, and receptor behavior associated with cancer. By enabling biomolecular investigation of cells at nanometer resolution, TERS could significantly contribute to the rational design and optimization of targeted therapeutic drugs. In this regard, TERS is gaining recognition in biomedicine as an innovative technique that combines structural imaging with molecular-level spectroscopy. Unlike conventional optical methods constrained by spatial resolution, TERS enables simultaneous visualization of cellular and

biomolecular ultrastructures while revealing their distinct biochemical signatures through vibrational fingerprints. This dual capability offers an integrated perspective on both morphology and chemistry within complex biological systems, opening new avenues for exploring fundamental biological processes at unprecedented resolution [26]. Beyond basic research, TERS also holds promise for medical diagnostics and therapeutics, providing highly sensitive and non-destructive detection and characterization of biomolecules [27].

Applied to nucleic acids, TERS can differentiate among nucleobases in DNA and RNA and has even demonstrated sub-nanometer resolution sufficient to resolve individual DNA bases—an achievement that opens the door to direct, label-free sequencing [28], [29]. Furthermore, TERS allows simultaneous examination of both the structural and functional characteristics of proteins. This approach has been utilized to study cytochrome c, amyloid fibrils, and insulin fibrils, as well as to differentiate between glycosylated and native forms of RNase [30], [31], [32], [33]. At the cellular level, TERS offers a unique view of the nanoscale organization of complex environments, such as membranes, allowing clear differentiation of microdomains with distinct protein and lipid distributions. In this regard, Böhme *et al.* demonstrated the potential of TERS by analyzing a lipid bilayer model (POPC/DOPS) and correlating the results with measurements on fixed human keratinocytes to determine their biochemical composition. Notably, the spectra revealed strong lipid contributions on the cell surface membrane, occasionally intermixed with protein and oligosaccharide signals [34]. In another work, Richter *et al.* employed hyperspectral TERS mapping with multivariate analysis (N-FINDR algorithm) to examine the membrane regions of human colon-cancer cells (HT29) [35]. Specifically, protein domains were distinguished by their Amide I ($1663\text{--}1637\text{ cm}^{-1}$) and Amide III signatures, whereas lipid regions were identified through phosphate vibrations at 1087 and $1209\text{--}1215\text{ cm}^{-1}$, CH_2/CH_3 chain modes, ester C=O stretching at $1725\text{--}1728\text{ cm}^{-1}$, and a characteristic cholesterol band at $690\text{--}699\text{ cm}^{-1}$. Using these spectral markers, protein and lipid domains were resolved with lateral dimensions down to $10\text{--}20\text{ nm}$.

In a different study, Kumar *et al.* used deuterated stearate isotope labeling in mouse pre-adipocytes, allowing detection of newly synthesized phospholipids via their distinct C–D Raman bands and mapping them using TERS [9]. They achieved non-destructive molecular mapping with $\sim 19\text{ nm}$ resolution, revealing thin, chain-like features likely representing the smooth endoplasmic reticulum, demonstrating the ability of TERS to map cellular molecules at nanoscale. Recently, Mrđenović employed label-free TERS imaging to visualize the molecular distribution in the membranes of pancreatic cancer cells (BxPC-3) [36]. TERS images were generated from the absolute intensities of Raman marker bands, complemented by correlative AFM and confocal Raman imaging. The study revealed nanoscale segregation of phenylalanine-rich and histidine-rich regions, PC lipid- and protein-rich domains, as well as cholesterol-rich areas in the cell membranes, including zones where cholesterol and protein co-localized. Notably, higher-resolution TERS imaging showed that these molecular domains were not chemically pure, containing other biomolecules such as proteins and DNA within a few nanometers, achieving spatial resolution down to $\sim 2.5\text{ nm}$. In another approach, Shibata *et al.* pioneered intracellular TERS by inserting silver nanoparticle–functionalized AFM probes directly into living HeLa cells [37]. This enabled selective detection of DNA, proteins, and lipids in distinct compartments, distinguishing nuclear from cytoplasmic

regions and even monitoring time-dependent changes in glycogen and protein signals. Collectively, these studies demonstrate the versatility of TERS for high-resolution, label-free molecular imaging, ranging from nanoscale mapping of cell surfaces to real-time tracking of biochemical composition and processes in cells.

Furthermore, TERS has been applied to intact mammalian cell membranes to detect membrane receptors, underlining its potential for high-resolution molecular mapping in biological systems. In addition, TERS is also used for probing protein-ligand interactions at the nanoscale with high chemical specificity. Carrier *et al.* showed that individual biotin-labeled gold nanoparticles bound to streptavidin on glass slides produced enhanced Raman signals from both the ligand and receptor, demonstrating the potential of single-particle TERS for studying binding events [38]. In a different study, Xiao *et al.* (2016) used cyclic RGD-conjugated gold nanoparticles (cRGD-GNPs) and multivariate curve resolution (MCR) analysis to distinguish between integrins $\alpha 5 \beta 1$ and $\alpha v \beta 3$ on surfaces and in SW480 colon cancer cell membranes, revealing unique spectral signatures for each receptor and colocalization within ~ 60 nm [39]. Extending this approach, Xiao *et al.* investigated the relationship between spectral variability in SERS and TERS and ligand specificity in affinity-based assays targeting integrin $\alpha v \beta 3$. The study examined interactions with three peptide ligands, both using purified proteins and within the native environment of intact SW620 cancer cell membranes [40]. Their results revealed that greater spectral variance indicated non-specific ligand-receptor interactions. Furthermore, by integrating this approach with single-particle tracking, they distinguished ligands based on their binding affinity, establishing a promising new strategy for probing drug-target specificity at the molecular level. In parallel, Alexander and Schultz applied TERS microscopy to visualize antibody-conjugated nanoparticles directly on intact cellular membranes to study biomolecular interactions in situ [41]. This technique allowed for the detection of individual nanoparticles and the mapping of specific antibody-antigen interactions on cell surfaces, achieving spatial resolution below 50 nm. Importantly, it also provided detailed chemical information about the molecules involved in these interactions, offering a powerful tool for investigating molecular recognition processes within the complex cellular environment. Together, these works illustrate TERS as a powerful tool for investigating the chemical nature and specificity of ligand-receptor interactions directly in cells.

This chapter presents a proof-of-concept study on the use of tip-enhanced Raman spectroscopy for the label-free characterization of the human cell membrane, using the hepatocarcinoma cell line HepG2 as a model system. The approach highlights the capability of TERS to deliver biochemical information at the nanometer scale, offering valuable insights into the molecular composition of cancer cell membranes. Furthermore, the chapter explores the TERS capability to reveal the interaction between SERS tags incubated with HepG2 cells to selectively detect glypican-3 (GPC-3), an important cancer biomarker in hepatocellular carcinoma. The SERS tags were functionalized with the Raman reporter 4-MBA and an anti-GPC-3 antibody as the recognition element, enabling their specific binding to the overexpressed glypican-3 protein on the outer membrane of hepatocellular carcinoma cells. These interactions were investigated using gap-mode TERS, established between the gold-coated tip and the functionalized gold nanoparticles under side-illumination, which generated distinct Raman fingerprints enabling the precise identification of GPC-3. Beyond

confirming the presence of this biomarker, the findings underscore the promise of combining TERS with SERS-based probes as a powerful strategy for biomarker detection, cancer cell characterization, and potential utility in cancer diagnostics.

2. TERS basic principles

Conventional optical microscopy and Raman spectroscopy are fundamentally limited by the diffraction limit of spatial resolution at about 200, nm according to the Abbe criterion [42]. TERS overcomes this limitation by confining and enhancing the Raman signal within a nanoscale hotspot generated at the apex of a plasmonically active metallic tip [17], [43]. This enables TERS imaging with resolutions well below the diffraction limit—typically in the 20 - 30 nm range—where performance is largely determined by the geometry and dimensions of the tip [14]. In TERS, the tip is crucial for both signal enhancement and spatial resolution, its stability affects measurement quality. TERS tip nanostructures can have various geometries, including spheres, nanorods, or nanocones, as shown in **Figure 3** [17], [43]. Gold and silver are commonly used for tips due to their plasmonic properties, typically deposited on AFM cantilevers via evaporation or electrodeposition [12]. Gold is generally preferred for excitation wavelengths between 633 and 785 nm, while silver is more effective for 488 to 568 nm [17].

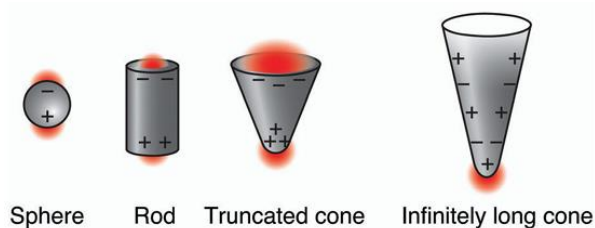


Figure 3. Common metal tip geometries for TERS setups include spherical, rod-shaped, truncated, and tapered tips. From Ref. [17].

Tip-enhanced and Surface-enhanced Raman spectroscopy are both based on electromagnetic and chemical enhancement, but they operate differently. In SERS, multiple plasmonic hot spots are distributed across a laser-illuminated substrate, producing very high enhancement factors and often benefiting from stronger chemical contributions. By contrast, TERS relies on a single metallic tip that scans across the sample surface, with enhancement factors typically lower (10^3 – 10^6) compared to SERS [17], [43]. The enhancement in TERS arises mainly from electromagnetic effects linked to the tip acting as a nanoscale optical antenna [17]. Three main physical effects underlie this amplification. First, localized surface plasmon resonances (LSPR) occur when the tip's conduction electrons collectively oscillate in response to incident light, concentrating the electromagnetic field in the immediate vicinity of the apex. In particular, when a focused laser illuminates the apex of a metallic or metal-coated nanotip, it excites localized surface plasmons at the tip, creating a confined nanolight source that produces evanescent light. This nanoscale light excites Raman scattering from the sample directly under the tip, as shown in **Figure 4a**.

The scattered signal, generated in the near-field, is then directed back to the far-field by the tip and collected using conventional optics and a spectrometer.

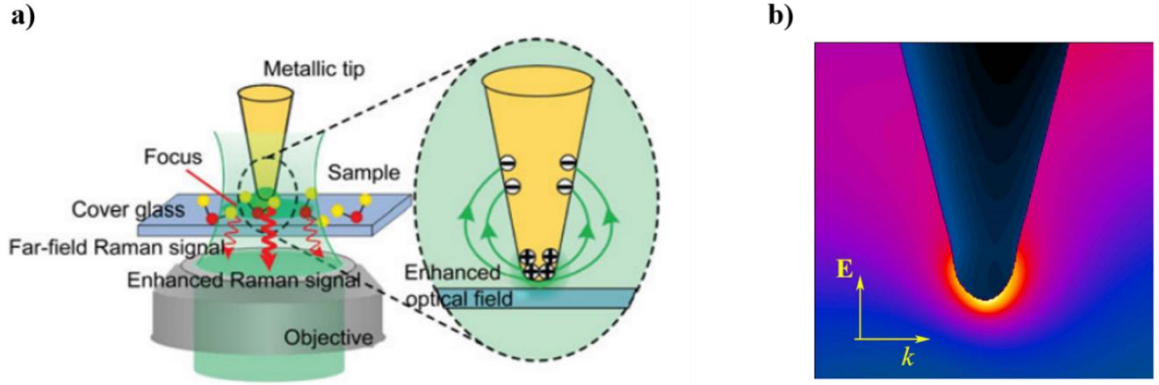


Figure 4. TERS principles. Schematic representation of the enhancement of the optical field at the apex of a sharp metallic tip when illuminated by a laser source in TERS setup (a), from Ref.[45]. Illustration of the lightning rod effect, where a sharp tip concentrates the electric field at its apex, particularly when the incident radiation is polarized along the tip axis, producing a strong local field enhancement (b). E: electric field, k: wave vector of the excitation laser. From Ref. [16].

By scanning the tip across the sample surface, Raman signals from each nanoscale region under the apex are recorded, yielding an image with spatial resolution determined by the size of the tip apex [43]. Furthermore, LSPR strength and efficiency depend on factors such as the material choice, curvature, dielectric environment, and illumination parameters [44]. In TERS, the performance of the system is highly influenced by how light is delivered to the tip—particularly its wavelength, power, and polarization. The polarization of the incoming light plays a crucial role in concentrating the electric field at the tip. The greatest enhancement typically occurs when the light’s electric field is aligned along the tip axis (p-polarization). If the polarization deviates from this alignment, such as having a component perpendicular to the tip, the resulting enhancement is significantly reduced [15]. In this case, the second effect represented by the lightning rod effect further amplifies the local field by gathering charges at the tip’s sharpest regions, independent of the light’s frequency, but only when the incident radiation is parallel to the tip axis as depicted in **Figure 4b** [16]. The third mechanism, antenna resonances, enhances the response when the tip length matches the effective excitation wavelength [12]. As a result, metallic nanostructures like TERS tips can both intensify and confine electromagnetic fields, amplifying Raman scattering from nearby molecules with signal strength scaling approximately with the fourth power of the local field enhancement. Scanning the tip across a surface provides not only enhanced Raman spectra but also simultaneous nanometer-resolved topography and optical imaging, making TERS a uniquely powerful nanoscale characterization tool [16]. In TERS, the metal tip amplifies the local electric field compared to the incoming light. This enhancement for the excitation light can be expressed as [15]:

$$G_{\omega_0} = E_{loc}(\omega_0) / E_0(\omega_0)$$

Where $E_{loc}(\omega_0)$ is the electric field at the tip, $E_0(\omega_0)$ is the incident light electric field without the tip in contact. Similarly, the scattered Raman light is also locally enhanced [15]:

$$G_{\omega_R} = E_{loc}(\omega_R) / E_0(\omega_R)$$

When the laser and Raman wavelengths are close, the tip enhances both fields nearly equally. As a result, the overall electromagnetic enhancement factor for TERS can be approximated as the fourth power of the local field enhancement [15]:

$$EF_{TERS,EM} = (G_{\omega_0})^2 / (G_{\omega_R})^2 \approx (G)^4$$

Maximizing this enhancement requires careful optimization of the tip's material, shape, curvature radius, and the excitation laser wavelength [14].

In particular, a typical TERS setup integrates three essential systems: a scanning probe microscope (SPM), an optical excitation and collection pathway, and a spectrometer. The SPM provides nanoscale surface characterization by positioning and controlling a metallic tip just a few nanometers above the sample surface.

When a laser beam is focused onto the tip apex using the microscope objective, a highly confined electromagnetic field is generated at the tip-sample junction, resulting in a substantial enhancement of the Raman signal from nearby molecules. The scattered photons are then collected by the objective and directed into the spectrometer for spectral analysis. This configuration enables simultaneous acquisition of topographical and chemical information. However, it should be emphasized that the confocal optical pathway of the objective captures not only the enhanced near-field TERS signal but also background contributions from far-field Raman scattering [15]. A typical side-illumination TERS-setup, used in this chapter, is shown in **Figure 5a**.

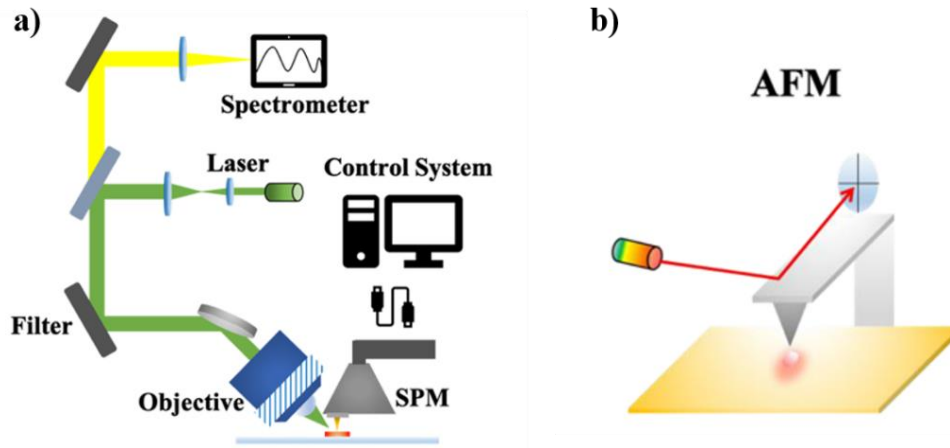


Figure 5. Diagram of the typical TERS setup in side-illumination configuration (a). AFM feedback system in the AFM-TERS system (b). Re-adapted from Ref. [15].

The AFM-TERS system is a specialized form of scanning probe microscopy that uses a sharp metallic tip mounted on a cantilever to scan the sample surface with nanometer precision. The cantilever vibrations caused by tip-molecule interactions are tracked optically using a laser and photodetector, while a feedback loop regulates the tip-sample distance to ensure stable, high-resolution topographic imaging. The AFM feedback system is reported in **Figure 5b**. Moreover, the technique supports multiple operational modes, including contact and tapping; the latter is especially suited for delicate specimens, as the reduced tip-sample interaction minimizes the likelihood of structural damage [46], [47].

In TERS, the choice of optical illumination is a key experimental factor that directly affects signal intensity, spatial resolution, and overall measurement performance. Three main optical geometries can be used: bottom, side, and top illumination, each optimized for specific sample types and measurement goals.

Bottom illumination is typically employed for transparent substrates, using a transmission-mode setup. Here, the laser is focused from below through the substrate onto both the metallic tip apex and the sample region directly beneath it, creating a highly localized interaction zone (**Figure 6a**). This approach benefits from high-numerical-aperture (NA) objectives, which enable strong confinement of the laser energy, minimize background illumination, enhance signal collection, and make the setup compatible with delicate biological samples.

Side illumination involves directing the excitation laser at an oblique angle relative to the sample surface, generally between 45° and 70° , using long-working-distance objectives (**Figure 6b**). Although these objectives have lower NA values (typically 0.28-0.55), which produce a larger focal spot, the reduced focus can be compensated by higher laser power (5-10 mW) and the strong plasmonic enhancement provided by metallic surfaces in the gap between tip and sample. This configuration works well for both transparent and opaque substrates. Top illumination delivers light along the tip axis, perpendicular to the sample surface, providing a more symmetric focal spot and moderate NA (**Figure 6c**). While this geometry allows direct illumination of both tip and sample, it requires precise tip positioning to avoid partial blockage of the beam. Top illumination is versatile and compatible with a variety of sample types but generally offers moderate signal enhancement compared with the other configurations [48].

In addition to the conventional TERS configuration, there is a variant known as gap-mode TERS, in which a sharp metallic tip interacts with a nearby metallic substrate.

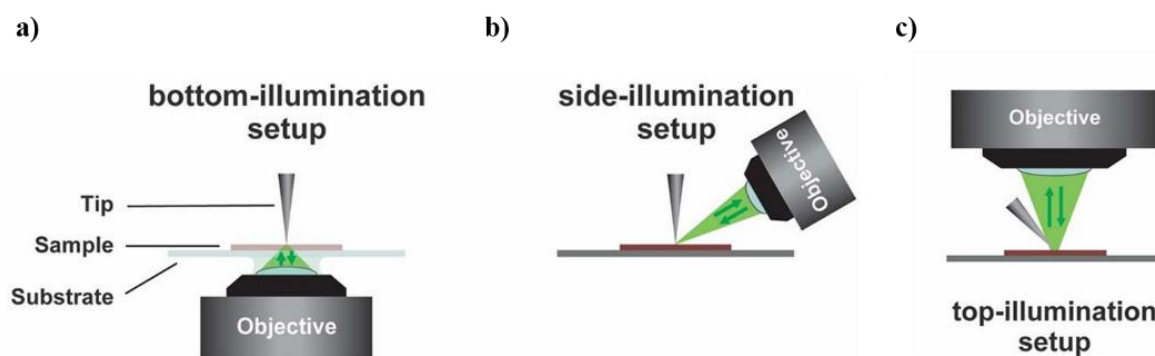


Figure 6. Schematic of the various optical configurations used in tip-enhanced Raman spectroscopy. Re-adapted from Ref [48].

Upon illumination, the localized plasmon of the tip couples with a mirror plasmon induced in the substrate, creating a highly confined electromagnetic hotspot within the nanoscale gap, as shown in **Figure 7**. This interaction can boost Raman signals by up to two orders of magnitude, making single-molecule detection possible. The approach is especially beneficial for biological samples, where the small tip–substrate spacing allows efficient coupling [48], [49].

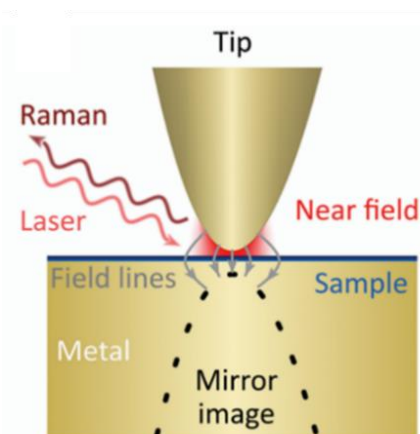


Figure 7. Schematic illustration of gap-mode TERS, Re-adapted from Ref. [49].

TERS offers significant advantages due to its ability to achieve the nanometer-scale resolution, often just a few nanometers, thanks to a key element: a plasmonic metallic tip that amplifies Raman signals through the LSPR phenomenon. This enhancement not only maintains chemical specificity but also greatly increases sensitivity, enabling the detection of single molecules [50]. TERS uniquely provides both chemical and topographical information simultaneously, delivering insights into the molecular composition and the morphology of the sample. As a label-free technique, it allows the study of samples in their native state, and when implemented in an AFM-TERS configuration, it can operate in liquid environments, making it particularly suitable for biological systems. Despite these strengths, TERS faces practical limitations. The reproducibility and stability of tips remain major concerns: silver tips, while highly enhancing, are prone to rapid oxidation. Furthermore, the intense localized laser field can induce local heating, leading to bleaching or chemical damage of sensitive samples [14], [43]. Tip contamination and limited lifetimes are common, especially when probing sticky or rough samples [51], [26]. Far-field background signals often compete with the near-field contribution, reducing image contrast and requiring sophisticated background suppression or illumination schemes [43]. Interpreting TERS spectra also presents challenges, as spectral shifts or distortions compared to conventional Raman or SERS can complicate band assignment and hinder quantitative analysis. Moreover, fluctuations in laser power and instabilities in tip-sample distance control can further complicate TERS measurements [12].

3. Materials and Methods

3.1 Materials and reagents

N-hydroxy succinimide (NHS) 1-ethyl-3-(3-dimethylamino-propyl)carbodiimide (EDC) were obtained from Cytiva (Marlborough, MA, USA). Mouse monoclonal GPC3 antibody (9C2) (Cat. No. NBP2-37483), recombinant GPC-3 (2119-GP) were purchased from Biotechne (Minneapolis, MN, USA). Peroxidase AffiniPure Goat Anti-Mouse IgG secondary antibody was from Jackson ImmunoResearch (Cat. No. 115035003; Cambridge, UK). QuantaRed™ Enhanced Chemifluorescent HRP Substrate kit was obtained from Thermo Fisher Scientific (Waltham, MA, USA). Methanol (MeOH), Phosphate-buffered saline (PBS), borate buffer (BB), 4-mercaptobenzoic acid (4-MBA), Bovine Serum Albumin

(BSA), 4-mercaptobenzoic acid and 2-(N-Morpholino)ethanesulfonic acid (MES) were supplied by Merck-Millipore (Milan, Italy). Gold nanoparticles (40 nm) were from BBI Solutions (Crumlin, United Kingdom).

3.2 Cell culture and preparation

HepG2 cells were cultured in DMEM/F12 medium (Thermo Fisher Scientific Inc.) supplemented with 10% fetal bovine serum (Biochrom GmbH) and maintained in a humidified incubator at 37 °C with 5% CO₂. After cultivation, the cells were fixed with 2% paraformaldehyde and stored in PBS at 4 °C until use in TERS analysis, either for label-free TERS investigation of the cell membrane or for GPC-3 targeting via SERS nanotags in combination with TERS. For the latter, fixed HepG2 cells were incubated with SERS nanotags by adding a 600 µL aliquot of functionalized nanoparticle solution onto a calcium fluoride slide within a six-well plate. The incubation proceeded for 2 h at room temperature, after which samples were gently rinsed with 2 mM BB containing 1% (w/v) BSA. The fixed cells were subsequently air-dried prior to SERS and TERS analysis.

3.3 Enzyme-linked immunosorbent assay (ELISA)

The functional activity of the immobilized anti-GPC-3 antibody was validated by enzyme-linked immunosorbent assay (ELISA). Microtiter plates were coated in triplicate with GPC-3 antigen at concentrations of 0, 10, 50, and 500 ng/mL and incubated at 37 °C for 2 h. Wells were then blocked with 3% (w/v) BSA for 1 h, followed by incubation with the antibody-conjugated nanoparticles for 1 h at 37 °C. After washing with PBS, bound antibodies were detected using goat anti-mouse HRP-conjugated secondary antibody, followed by signal development with the QuantaRed™ Enhanced Chemifluorescent HRP Substrate Kit according to the manufacturer's protocol. Fluorescence was measured at 590 nm using an EnSpire® Multimode Plate Reader (PerkinElmer, Italy).

3.4 Functionalization of AuNP-based SERS Tags

Gold nanoparticles (40 nm) were functionalized in a stepwise manner with a Raman reporter molecule, followed by conjugation of anti-GPC-3 antibody. First, 1 mL of colloidal gold suspension was added to 40 µL of 50 mM borate buffer (pH 8.5). Next 14 µL of a 1.0 mM 4-MBA solution in methanol was added, and the mixture was vortexed briefly before incubation at 4 °C for 90 min. Carboxyl groups of the surface-anchored reporter molecules were then activated by treatment with freshly prepared EDC (0.4 M) and NHS (0.1 M) in 0.1 M MES buffer (pH 4.8) for 5 min under stirring. Excess coupling reagents were removed by centrifugation (4000 g, 15 min, 4 °C), and the pellet was resuspended in 0.1 mM MES buffer (pH 4.8). For antibody attachment, 14 µL of anti-GPC-3 antibody (100 µg/mL in MES buffer) was conjugated to the activated nanoparticles and incubated at 4 °C for 90 min. Functionalized AuNPs were then subjected to three centrifugation cycles to remove unconjugated antibodies. To stabilize the conjugates and block residual reactive sites, 100 µL of 10% (w/v) BSA in 2 mM BB (pH 8.5) was added. After a final centrifugation step, the resulting SERS probes were re-suspended in 2 mM BB containing 1% (w/v) BSA and

stored at 4 °C until further use. The entire functionalization workflow was presented in **Figure 8**.

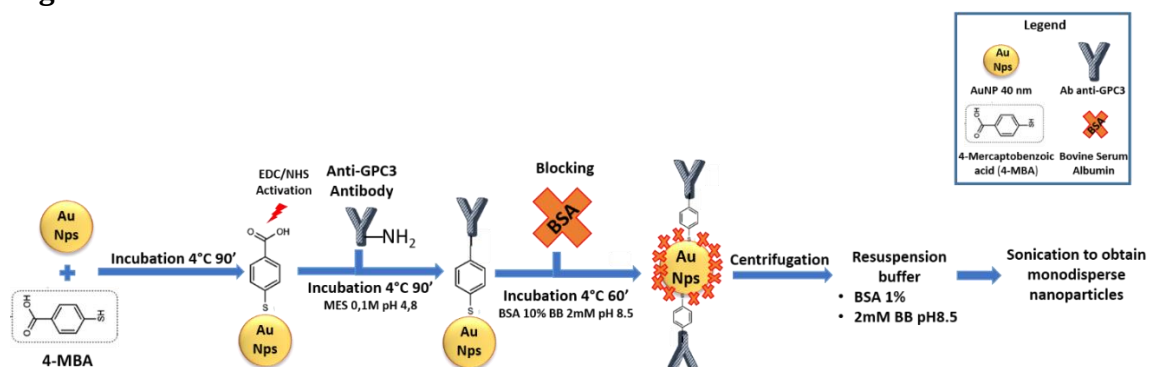


Figure 8. Graphical representation of the gold nanoparticle functionalization protocol for GPC3 detection.

3.5 Characterization of SERS tags

The physicochemical properties of the SERS tags were assessed using dynamic light scattering (DLS) on a Zetasizer Ultra (633 nm laser, 173° scattering angle; Malvern Instruments Ltd., UK,) to determine hydrodynamic diameter, polydispersity index (PDI) and zeta potential. UV–Vis absorption spectra were acquired with a NanoDrop Microvolume Spectrophotometer (Thermo Scientific).

The characterization of SERS tags was performed by SERS analysis using a LabRAM HR Nano system (HORIBA Scientific, France) equipped with a 633 nm laser as the excitation source. A 100× air objective (Olympus, Tokyo, Japan) with a numerical aperture of 0.90 was used. Measurement parameters included an acquisition time of 1 s, a single accumulation, and an ND filter set to 25%. Raman lines were dispersed by a 300 g/mm grating and collected by a CCD camera cooled to -60°C. The confocal hole was set to 100 μm to spatially filter the Raman signal. Spectra were recorded in the range of 400–1800 cm⁻¹. After spike removal, baseline correction was applied using a 4th-degree polynomial (200 points) with Labspec 6 software.

3.6 TERS setup, measurements and data analysis

TERS measurements were performed using a LabRAM HR Nano system (HORIBA Scientific, France), which integrates a scanning probe microscope (OmegaScope) with a confocal Raman micro-spectrometer. AFM maps of untreated cells (40 × 40 μm²) and cells incubated with SERS tags (30 × 30 μm²) were acquired in dynamic (AC) mode with a resolution of 512 × 512 pixels.

TERS analyses were performed with two laser excitations, 532 nm and 633 nm, both incident at 60° relative to the sample plane. In each case, the laser was focused through a 100× objective (NA = 0.7) mounted on a piezo scanner, ensuring accurate positioning and focusing of the Raman spot at the probe tip apex. Raman signals were dispersed with a 300 lines/mm grating and detected by a CCD camera cooled to -60 °C. The TERS probes used were cantilever-based gold AFM-TERS tips (OMNI TERS-SNC-Au App Nano, HORIBA Scientific).

Label-free TERS spectra of HepG2 cell membranes were collected with the 532 nm laser from five single-point measurements on two different cells. The acquisition parameters were: 20 s acquisition time, 3 accumulations, 100 μm hole, and 25% laser power. Additionally, a $5 \times 5 \mu\text{m}^2$ TERS map with 50×50 -pixel resolution was acquired on the second cell under the same conditions, except with a reduced acquisition time of 1 s per pixel to minimize laser-induced damage.

For the detection of GPC-3 on HepG2 cell membranes incubated with SERS tags, TERS measurements were performed with the 633 nm laser under the following conditions: 20 s acquisition time, 3 accumulations, 100 μm hole, and 25% laser power. To verify tip integrity and rule out contamination, Raman spectra of the retracted TERS tip were recorded before and after each experiment. All TERS data were processed with spike removal and baseline correction using a 7th-order polynomial with 200 points.

4. Results and Discussion

4.1 TERS label-free investigation of HepG2 cancer cell membrane

Topographic imaging of untreated HepG2 cells in dynamic AFM mode (AC mode) provided detailed visualization of membrane contours and shapes, offering structural information before TERS spectroscopic analysis. In particular, the Mag AFM image shown in **Figure 9** revealed fine structural and morphological details of the analyzed biological samples. In AFM dynamic imaging, the magnitude signal serves as an error indicator that reflects deviations from the specified set-point [52]. Consequently, the magnitude (or amplitude) image illustrated how the cantilever's oscillation amplitude varies locally relative to the predefined set-point. Bright zones generally correspond to softer cellular components such as the cytoplasm, while darker regions indicate stiffer or more adhesive structures, including the cytoskeleton-dense areas. Smooth variations in the magnitude signal indicate a uniform and regular cell surface, whereas a noisy or irregular magnitude signal suggests a corrugated or morphologically altered surface [52]. Thus, the magnitude image provides a qualitative representation of the spatial variations in tip-sample interaction strength, highlighting mechanical contrasts and structural boundaries within the cell.

In **Figure 9**, the two cells appeared elongated and nuclear region could not be easily distinguished from the cytoplasm. The texture displayed both smoother regions and rougher, elevated areas, which may indicate cellular features such as cytoskeletal structures and membrane roughness. In particular, the AFM analysis revealed heterogeneous cell surfaces, possibly corresponding to different membrane regions such as lipid-enriched domains, protein clusters and membrane protrusions [53], [54]. The nanoscale variations observed across the cell surface emphasize the inherent complexity of the plasma membrane, where subtle topographical differences may modulate molecular organization and govern cellular interactions.

The AFM map was essential for accurately guiding the placement of TERS measurements, enabling investigation of specific regions of interest on the sample. The mag AFM image underscores the importance of integrating structural mapping with chemical analysis to better unravel the complex organization of the cell membrane [52].

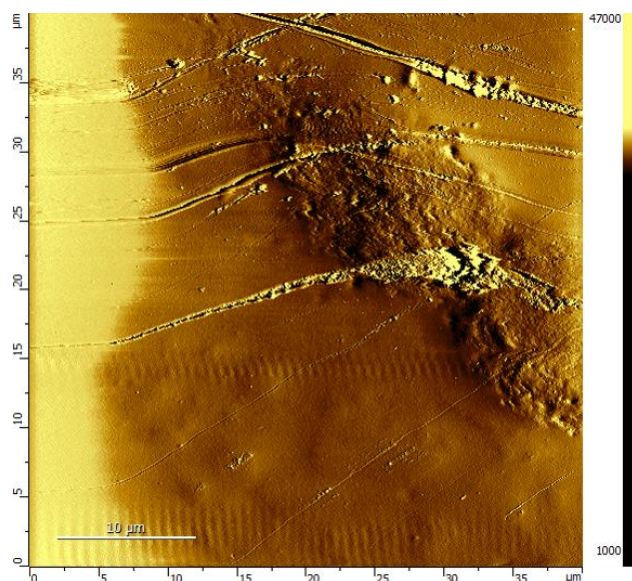


Figure 9. Atomic force microscopy (AFM) magnitude (Mag) image of HepG2 cells ($40 \times 40 \mu\text{m}$). Cells were scanned in air at room temperature. Scale bar: $10 \mu\text{m}$. The vertical color scale, ranging from dark brown to light yellow, represents regions with progressively stronger to weaker interactions between the sample and the TERS tip.

TERS analysis was carried out by acquiring five spectra at different locations on two individual cell using a 532 nm laser, as illustrated in **Figure 10a** and **c**. The cellular TERS spectra were recorded with the tip in direct contact with the cell membrane surface. For comparison, the full set of spectra obtained from both cell samples, along with reference spectra collected outside the cells and with the tip positioned far from the membrane, are presented in **Figure 10b** and **d**. Each spectrum exhibited a distinct molecular fingerprint, highlighting the variability in biochemical composition across different plasma membrane regions. The TERS bands detected in TERS spectra for both samples, along with their corresponding biomolecular assignments, are summarized in **Table 1**. Notably, the TERS analysis of cell 1 and cell 2 revealed both common features and marked differences in spectral band positions and biochemical composition. Both cell 1 and 2 displayed well-defined protein-related signals, such as the phenylalanine band at 1006 cm^{-1} and amide I mode at $1657\text{--}1669 \text{ cm}^{-1}$ within the fingerprint region ($400\text{--}1800 \text{ cm}^{-1}$). In addition, the recurrent band at $1588\text{--}1591 \text{ cm}^{-1}$ across several TERS spectra of cell 1 indicated an enrichment of phenylalanine and hydroxyproline compared with cell 2 in which was present only in spectrum 3. Fatty acid and cholesterol-related bands were detected at $2884\text{--}2887 \text{ cm}^{-1}$ and 2967 cm^{-1} Raman peaks, while cholesterol-associated Raman bands were absent in cell 2. The asymmetric stretching modes of choline were identified in all spectra from both cells. Overall, these findings indicate that cell 1 exhibited a higher contribution of cholesterol and fatty acids within its membrane, along with enhanced signals from the amino acid phenylalanine. In this context, cholesterol-rich membrane domains are increasingly acknowledged as crucial regulators of cancer cell behavior, affecting proliferation, survival, apoptosis, and invasion. In particular cholesterol plays a crucial role in regulating the fluidity of cell membranes, directly influencing cancer cell behavior. When cholesterol levels are high, the membrane becomes more rigid, which contributes to chemotherapy resistance by limiting drug penetration.

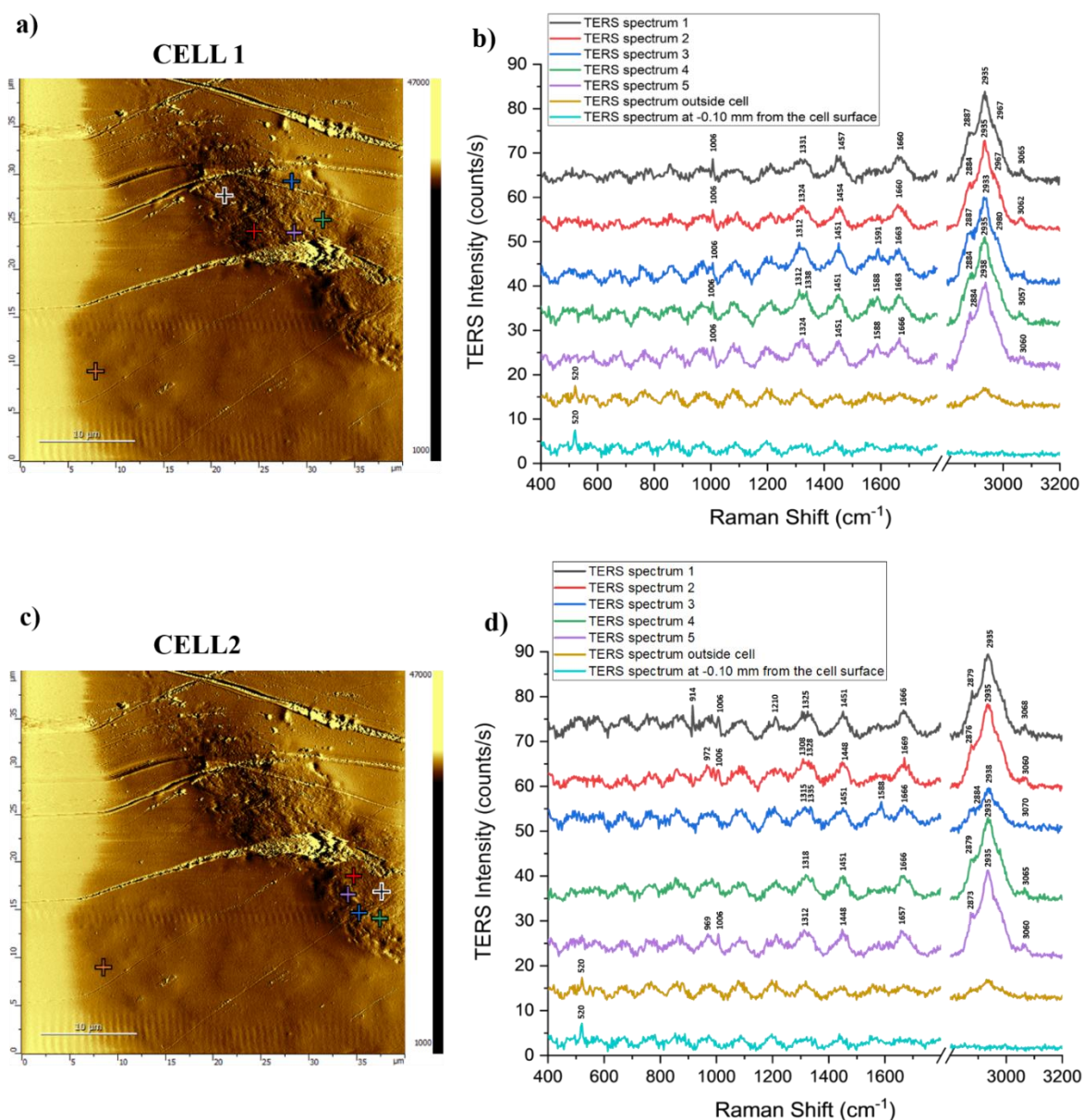


Figure 10. TERS spectra were acquired from five distinct locations on the plasma membrane surface HepG2 cells indicated in the a) and c) Mag AFM images, respectively. AFM Mag Images Scale bar: 10 μm. The vertical color scale, ranging from dark brown to light yellow, represents regions with progressively stronger to weaker interactions between the sample and the TERS tip. The tip positions for cell 1 (a) and cell 2 (c) are marked with colored crosses, each color corresponding to its respective spectrum. The complete sets of spectra for cell 1 (b) and cell 2 (d), together with reference spectra collected outside the cells and at a distance from the membrane, are shown. Characteristic band positions identified in each spectrum are indicated, and the spectra have been vertically offset for clarity.

Table 1. TERS band assignments for TERS spectrum recorded from the two HepG2 cells are reported.

Raman Shift (cm ⁻¹)	Assignment
914	C-C skeletal mode (protein), carboxylate str. [34]
969-972	Lipids [55]
1006	Phe ring br. mode (protein) [56]
1210	Trp, Tyr, Phe C-C ₆ H ₅ str.[57]
1308-1318	CH ₃ CH ₂ twist. of protein/lipid [57]
1324-1328	def. CH (protein), def. CH bend., def. OH (oligosaccharide) [34]
1331-1338	Cα-H def. in Trp (protein) [36]

1448-1454	CH ₂ def. lipid/protein [58]
1588-1591	Phe, Hypro (protein) [57]
1657-1669	Amide I (protein) [57]
2879	CH ₂ asym. str. and CH str. of lipid/protein [57]
2884-2887	str. CH ₃ , lipids, fatty acids [59]
2933-2938	CH str. in-CH ₃ protein/lipid [56]
2967	str. CH ₃ , lipids, fatty acids, cholesterol and its ester [57]
2980	Asym. CH ₃ str. [60]
3060-3070	Asym. str. of CH ₃ in choline [11]

sym.= symmetric, asym. = asymmetric, str.= stretching mode, bend. = bending mode, twist. = twisting mode, def.= deformation mode, ring br.= ring breathing, Hypro= hydroxyproline, Phe = phenylalanine, Pro= proline, Trp = tryptofan, Tyr = tyrosine, Val = valine

Conversely, when cholesterol levels are low, the membrane becomes more fluid, enabling cancer cells to change shape, move more easily, and spread to other tissues. Thus, cholesterol balance determines whether cancer cells are more resistant to treatment or more capable of metastasis [3]. Notably, several studies have reported a high content of phenylalanine and choline derivatives in HepG2 cells [61]. Variations in the TERS spectra acquired at different spots on the two cells indicate a non-uniform (heterogeneous) distribution of biomolecules at the nanoscale on the HepG2 cell surface.

Subsequently, a TERS map ($5 \times 5 \mu\text{m}^2$) of a specific region within cell 2 provided valuable information about the spatial arrangement of biomolecules. After pre-processing, the visualization shown in **Figure 11** indicated that phenylalanine (1000 cm^{-1}) signals appeared as red pixels, while high wavenumber vibrations ($2800 - 3200 \text{ cm}^{-1}$) were shown as green pixel.

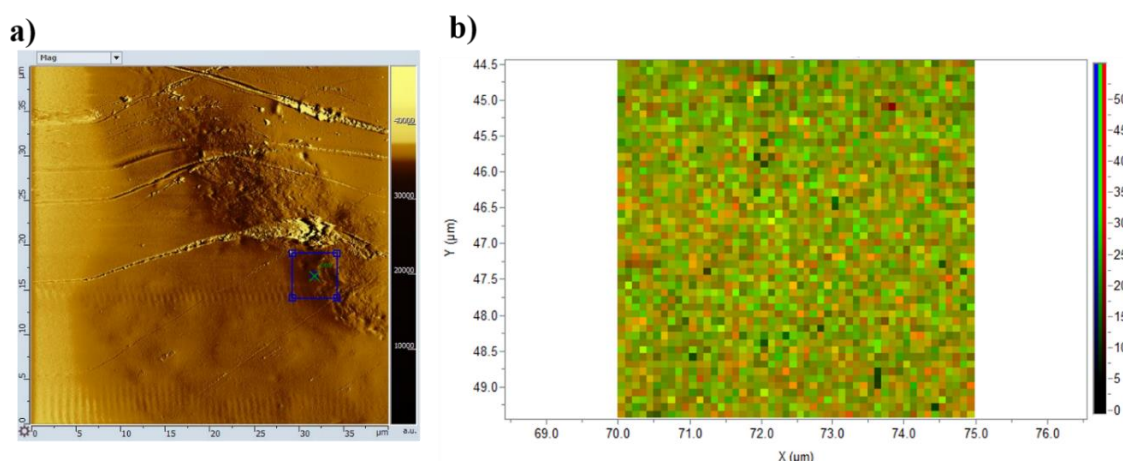


Figure 11. AFM-TERS map of HepG2 Cell 2. a) AFM Mag image of HepG2 cell 2. Scale bar: $10 \mu\text{m}$. The vertical color scale, ranging from dark brown to light yellow, represents regions with progressively stronger to weaker interactions between the sample and the TERS tip. The blue square outlines the $5 \times 5 \mu\text{m}$ region of the cell selected for scanning. b) The corresponding pre-processed TERS map provides simultaneous spatial and biochemical information. The TERS image is generated by mapping the intensity of the Raman band of phenylalanine at 1000 cm^{-1} and the high-wavenumber region ($2800\text{--}3200 \text{ cm}^{-1}$), represented by red and green pixels, respectively. These spectral features correspond to proteins in the case of the phenylalanine band, and to lipids along with protein contributions in the high-wavenumber region. This combined mapping highlights the heterogeneous organization of biomolecules across the cell membrane.

The resulting map demonstrated that proteins and lipids—the primary components of cellular plasma membranes—exhibit heterogeneous distribution patterns even within a

single cell. Future research will focus on acquiring nanoscale cell membrane maps to identify specialized microdomains containing concentrated levels of particular biomolecules. These microdomains may play crucial roles in cellular signaling processes and membrane structural organization.

Furthermore, applying TERS microscopy to cellular membranes is particularly challenging due to the fragile nature of lipid–lipid and protein–lipid interactions, which can be easily disrupted by external influences. While label-free TERS offers great potential for investigating membrane organization, the complex, heterogeneous, and highly dynamic nature of these structures complicates data interpretation. Moreover, because cells extend over tens of micrometers, accurately locating and probing a specific region of the membrane with nanometer precision remains a considerable technical challenge [41].

4.2 TERS for GPC-3 recognition on the cell membrane of HepG2

Label-free TERS provides important information about the biochemical composition of the cell membrane; however, its ability to distinguish specific protein cancer biomarkers is limited. Without advanced computational methods, such as machine learning for spectral deconvolution, reliably assigning a given signal to an individual biomarker becomes challenging [39]. The main issue arises from the inherent complexity of the membrane, which contains not only proteins but also lipids and carbohydrates, all contributing to overlapping and often indistinguishable spectral signatures. To address these limitations, a hybrid strategy combining TERS with functionalized SERS tags was adopted.

In this approach, SERS tags carrying strong and highly specific Raman reporters were conjugated to ligands or antibodies directed against the target cancer biomarker, in this case represented by glypican-3 (GPC-3). GPC-3, a cell-surface glycoprotein, with heparan sulfate glycosaminoglycan chains, has attracted considerable interest for its involvement in hepatocellular carcinoma, particularly in driving tumor progression and metastasis. Structurally, GPC-3 consists of a 60–70 kDa core protein anchored to the plasma membrane via a glycosylphosphatidylinositol (GPI) linkage. Its carboxyl-terminal region is further modified by the attachment of heparan sulfate chains, which are critical for its functional interactions with signaling molecules [62]. While GPC-3 is absent in healthy and benign liver tissues, it becomes notably reactivated in cancerous tissue. This upregulation in HCC correlates with increased tumor invasiveness and poorer patient prognosis, manifesting as shortened overall survival and increased recurrence rates, underscoring its significance as both a diagnostic biomarker and therapeutic target [63].

The targeted functionalization adopted enhanced spectral specificity and signal intensity, enabling precise recognition and localization of GPC-3 on the cell surface, even within the complex molecular environment of the plasma membrane. By coupling the high spatial resolution of TERS with the molecular specificity of SERS tags, this method provides a powerful platform for probing cancer-associated biomarkers with improved accuracy and reliability.

4.2.1 SERS tags profiling

As an initial step, an ELISA assay was performed to verify the specificity and selectivity of the anti-GPC-3 antibody before its conjugation with the Raman reporter in the SERS tags.

As shown in **Figure 12**, the experiment demonstrated a clearly dose-dependent response, which was the expected outcome when validating a detection system. In the absence of GPC-3 protein (0 ng/mL), the system registered only minimal background signal. This is crucial, as it indicates low non-specific binding and minimal assay noise. Introducing a concentration of just 10 ng/mL of GPC-3 resulted in a clear and immediate increase in the RFU signal. This initial jump confirmed the high sensitivity of the assay: the system can effectively detect the protein even at low concentrations. As the GPC-3 concentration increased, the fluorescence signal continued to rise, establishing a clear quantitative relationship: i) moving from 10 to 50 ng/mL, the increase persisted, though it was slightly less steep, suggesting that the specificity of the binding is maintained; ii) at the maximum tested concentration of 500 ng/mL, the peak signal was achieved. This confirms that the assay functions across a wide dynamic range of GPC-3 concentrations. In conclusion, the ELISA confirms the specific recognition of GPC-3 by the detection system. This makes the antibody a reliable bio-recognition element for quantifying GPC-3 protein also in unknown biological samples, potentially supporting its use in diagnostic applications.

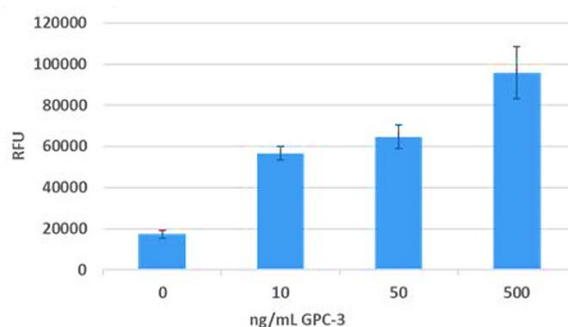


Figure 12. Enzyme-linked immunosorbent assay (ELISA) results obtained for GPC-3 concentrations of 0, 10, 50, and 500 ng/mL. Data are expressed as relative fluorescence units (RFU) on the y-axis, while GPC-3 concentration is plotted on the x-axis. The negative control (0 ng/mL) corresponds to phosphate-buffered saline (PBS).

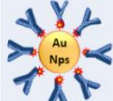
Subsequently, gold nanoparticles (40 nm) were functionalized with 4-MBA, which presents reactive carboxyl (-COOH) groups on the nanoparticle surface. These groups were then used to covalently conjugate anti-GPC3 antibodies through an activation step with 1-ethyl-3-carbodiimide hydrochloride (EDC) and N-hydroxysuccinimide (NHS), facilitating the formation of stable amide bonds between the nanoparticle carboxyl groups and the antibody amine groups. Following removal of unbound antibodies, the nanoparticles were passivated with BSA and sonicated, yielding monodisperse particles with enhanced target recognition efficiency.

The SERS tags are developed through the sequential functionalization of 40 nm spherical gold nanoparticles with the Raman reporter molecule, 4-MBA, followed by anti-GPC-3 antibodies to for selective recognition of the protein biomarker in the cell membrane. A blocking step was then performed to cover any exposed gold. To evaluate the functionalization protocol and verify the stability of the functionalized gold nanoparticles, the hydrodynamic radius, polydispersity index (PDI) and surface charge (Z pot) were measured by DLS and compared to those of naked nanoparticles. DLS analyzes assessed the size of the gold nanoparticles (~40 nm) before surface modification, their stability (PDI of 0.13), and a strong negative charge (**Table 2**). After the functionalization protocol, which

involved covalent binding of the Raman reporter, antibody and passivation of the remaining gold moieties, the particles had a diameter of 126 nm and a surface charge of -26.1 mV, while maintaining a reasonable PDI (0.26), suggesting monodisperse functionalized gold nanoparticles (**Table 2**).

These findings were consistent with the successful covalent conjugation of anti-GPC-3 antibodies. Fu *et al.* reported that nanoprobe uptake depends on both particle size and the fixation method, with paraformaldehyde fixation creating a stronger barrier that limits penetration even for 14 nm particles in cancer cells [64]. Based on this, it can be concluded that the functionalized SERS probes remained confined to the membrane of HepG2 cells due to their size.

Table 2. Dynamic light scattering analysis of naked and functionalized nanoparticles.

Sample	Size (d nm)	PDI	Z pot (mV)
 Naked 40 nm	41.26 ± 0.05	0.13 ± 0.01	-39.70 ± 1.02
 Post functionalization (AuNPs + 4-MBA + Ab anti-GPC-3 + BSA)	126.00 ± 1.79	0.26 ± 0.02	-26.10 ± 0.79

Before performing TERS measurements, the presence and stability of 4-MBA attached to gold nanoparticles were evaluated by Raman measurement. The SERS spectra of the tags are characterized by two intense bands at 1075 and 1585 cm^{-1} , corresponding to vibrations of the aromatic ring of the molecule (**Figure 13**) [65], [66].

Additional vibrational signals were observed at 509–520 cm^{-1} , which can be attributed to the recognition element, specifically the anti-GPC-3 antibody. The band in the 509–520 cm^{-1} region is consistent with vibrational modes of disulfide (S–S) bonds commonly present in IgG antibodies, reflecting the protein's tertiary structure and stability [67], [68]. Furthermore, the peak at 1178 cm^{-1} may arise from tryptophan residues within the antibody structure or could correspond to a minor vibrational mode of the 4-MBA reporter [41], [66].

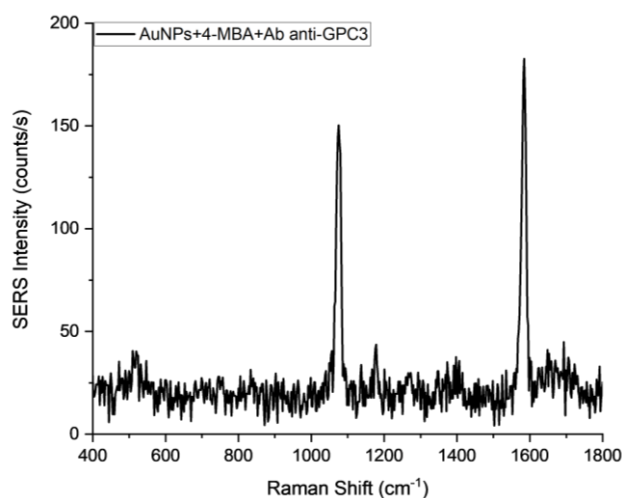


Figure 13. Representative SERS spectrum of gold nanoparticles functionalized with 4-MBA and anti-GPC3 antibodies, recorded with a 633 nm excitation laser prior to cell incubation.

These characterizations confirmed that the SERS tags can perform two main functions: impart immune specificity for GPC-3 and produce an intense Raman spectrum when irradiated with an adequate excitation source.

To investigate targeted GPC-3 detection, HepG2 cells were incubated for 2 hours at room temperature with gold nanoparticles functionalized with 4-MBA and anti-GPC-3 antibodies, followed by two washing steps to remove nonspecific interactions. As shown in **Figure 14a**, the cell image after incubation with the SERS tags revealed pronounced changes. In the bright-field images, HepG2 cells incubated with functionalized gold nanoparticles displayed distinct dark spots, indicating the surface binding of the SERS tags [69], [70]. The presence of gold nanoparticles enhanced optical signals through plasmonic effects, explaining the increased fluorescence observed in treated cells. The intense color of gold nanoparticles resulted from their strong light scattering at the surface plasmon resonance wavelength, which varied with particle size and shape. Huang et al. reported that with an SPR positioned near 540 nm on the cell monolayer, the particles produced a pronounced green-to-yellow scattering signal [71]. Additionally, BSA incubation was found to affect the visualization of the cells, producing discernible differences in their appearance. A SERS point measurement was performed out to validate the effective recognition of GPC-3 expressed on the membrane of HepG2 cells by the functionalized SERS tags before conducting the TERS analysis. The obtained SERS spectra of 4-MBA (**Figure 14b**) exhibited characteristic vibrational fingerprints, with strong C–C ring-breathing modes at 1080 and 1587 cm^{-1} and a weaker C–H bending mode at 1181 cm^{-1} , consistent with previously reported Raman features of 4-MBA [66].

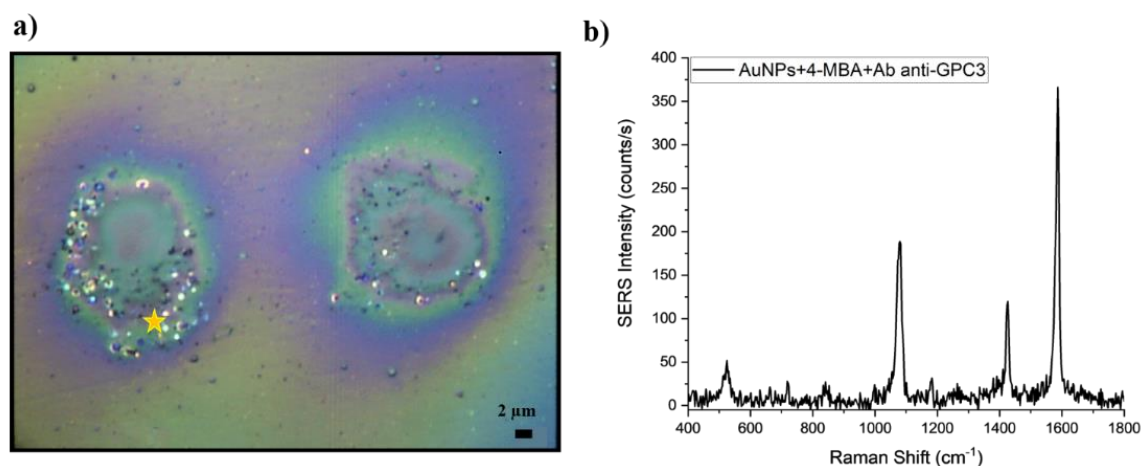


Figure 14. Bright-field micrograph of HepG2 cells after incubation with gold nanoparticles functionalized with the Raman reporter 4-MBA, an anti-GPC3 antibody for specific targeting, and bovine serum albumin to stabilize the plasmonic nanostructures and prevent nonspecific interactions (a). Image acquired at 100 $\times$ magnification; scale bar: 2 μm . Representative SERS spectrum of SERS tags incubated with HepG2 cells (b). The yellow star indicates the specific point where the spectrum was recorded.

These bands confirm the successful indirect detection of GPC-3 on the HepG2 cell surface. Notably, the antibody-associated band at 524 cm^{-1} was observed with higher intensity in this case compared to the SERS spectrum of free SERS tags not incubated with cells (**Figure 14**), which could suggest a strong interaction between the antibody and the GPC-3 cancer biomarker. The enhancement of this band provides further evidence that the immuno-

recognition process occurred effectively, enabling the selective anchoring of the SERS tags onto the HepG2 cell membrane. This result confirms both the molecular specificity of the functionalized probes and the reliability of using SERS as a pre-validation step before TERS analysis.

4.2.2 Combined TERS and SERS for GPC-3 recognition

The strongest SERS enhancements occurred at the junctions between nanoparticles, and coupling SERS nanotags with a TERS tip provided a powerful approach to amplify Raman signals associated with antigen–antibody interactions, enabling the nanometric localization of GPC-3 on the cancer cell membrane.

As a first step, a topographical image of a HepG2 cell incubated with SERS tags was recorded, as shown in **Figure 15a**. The Mag AFM map revealed that the nuclear region is poorly differentiated from the surrounding cytoplasm. Interestingly, the interaction of the SERS tags with the cell surface induced notable structural modifications at the level of the plasma membrane.

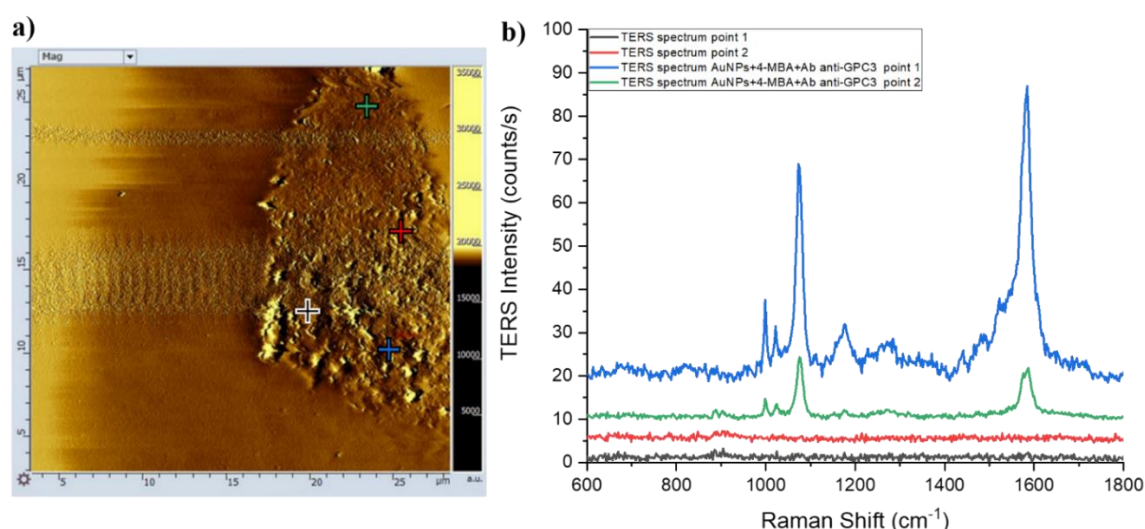


Figure 15. AFM magnitude image of fixed HepG2 cells is shown in (a). Scale bar: 10 μm. The vertical color scale, ranging from dark brown to light yellow, represents regions with progressively stronger to weaker interactions between the sample and the TERS tip. TERS spectra were collected at the specific positions marked with colored crosses on the AFM map in (b).

Specifically, numerous nanoscale depressions and pore-like structures became evident across the membrane, reflecting localized perturbations of the lipid bilayer and cytoskeletal reorganization in response to nanoparticle binding. These structural rearrangements were consistent with strong adhesion and local clustering of the SERS tags at the cell surface, leading to significant nanoscale modifications in the membrane architecture [52], [72].

In **Figure 15a**, the specific positions where TERS spectra were collected are indicated with different colors, allowing direct correlation between spatially resolved topography and molecular detection. The black and red TERS spectra correspond to regions of the membrane where no GPC-3 expression was detected, as evidenced by the absence of the characteristic Raman bands at 1080 cm⁻¹ and 1580 cm⁻¹. Conversely, the green and blue TERS spectra clearly exhibit the fingerprint signals of SERS tags, confirming the successful binding of anti-GPC-3-functionalized SERS nanoprobes at these membrane locations. When the gold

TERS tip comes into contact with gold SERS tags bound to GPC-3, the detected signal rises dramatically due to the interaction of coupled plasmonic nanostructures. This amplification is driven by two main effects. The first is the intense SERS enhancement that occurs within the nanometer-scale gaps between plasmonic structures, where electromagnetic fields are most strongly confined. The second is a plasmon coupling effect: while isolated 40 nm gold nanoparticles display a plasmon resonance at 533 nm, the close proximity of neighboring nanostructures allows their plasmons to couple, shifting the resonance toward longer wavelengths. This spectral shift aligns the resonance more closely with the 633 nm laser excitation, thereby amplifying the Raman response of molecules near the nanoparticles [38]. As a result, the Raman scattering from the SERS tags, which indirectly reveals the interaction between the GPC-3 protein and the anti-GPC-3 antibody, is greatly amplified, as evidenced by the pronounced green and, in particular, blue TERS spectra in **Figure 15b**. Notably, the blue TERS spectrum shows a markedly higher signal intensity compared to all other positions, indicative of a localized region with an elevated density of GPC-3 receptors. This observation points to a non-uniform distribution of GPC-3 across the plasma membrane, with distinct micro-domains enriched in surface cancer biomarker expression. Such spatial heterogeneity may be biologically significant, as it could influence nanoparticle binding efficiency, cell–nanoparticle interactions, and potentially downstream signaling events. Importantly, the spatial correspondence between AFM topography and TERS data provides compelling evidence of the specificity of the targeting strategy. Areas exhibiting pronounced membrane depressions in AFM coincide with regions showing strong SERS tag signals, reinforcing the interpretation that structural perturbations are directly linked to the localized accumulation of functionalized nanoparticles. These findings collectively demonstrate the feasibility of tip-enhanced Raman spectroscopy for nanoscale biomarker detection in cancer cells. The combination of antibody-mediated targeting and the localized electromagnetic enhancement of the TERS setup enables the selective identification of biomolecules with nanometric resolution. The ability to unambiguously detect GPC-3 expression on the surface of HepG2 cells underscores the potential of this approach as a powerful analytical tool for cancer diagnostics, offering both ultra-sensitivity and spatial precision beyond the reach of conventional techniques.

5. Conclusions

In this chapter, the comprehensive investigation of tip-enhanced Raman spectroscopy for the analysis of HepG2 hepatocellular carcinoma cells highlights the significant potential of this technique in biomedical research. Label-free TERS enabled visualization of heterogeneous distributions of proteins and lipids at the single cell level, providing high-resolution, chemical specificity, and spatially resolved insights unmatched by other label-free imaging methods. This approach revealed variations in molecular composition and organization, including contributions from cholesterol, fatty acids, and amino acids such as phenylalanine, offering a broad survey of native biochemical diversity. To overcome limitations in selectively detecting cancer biomarkers, a hybrid strategy combining TERS with SERS tags was employed, enabling targeted identification of glypican-3, a key hepatocellular carcinoma marker. Functionalized SERS tags conjugated with 4-MBA Raman reporter and anti-GPC-3 antibody bound specifically to overexpressed GPC-3 on the cell membrane, and

gap-mode TERS amplification allowed precise molecular identification. This integration of antibody-mediated targeting with localized electromagnetic enhancement demonstrated both molecular specificity and nanometric spatial resolution, providing a focused examination of targeted biomolecular interactions alongside the broader biochemical landscape. Looking forward, TERS is poised to become a core technology in personalized medicine, where understanding the precise molecular environment of diseased cells is critical for tailoring treatment strategies. As technical barriers are progressively addressed, integration into automated, high-throughput platforms may further expand its impact. With its combination of ultrasensitive detection, chemical specificity, nanometric spatial resolution, and compatibility with molecular targeting, TERS has the potential to redefine the limits of cell biology and medical diagnostics.

6. References

- [1] Q. Zhang, S. Li, Y. Yang, Y. Shan, and H. Wang, "Studying structure and functions of cell membranes by single molecule biophysical techniques," *Biophys. Reports*, vol. 7, no. 5, pp. 384–398, 2021, doi: 10.52601/bpr.2021.210018.
- [2] P. V. Escribá *et al.*, "Membranes: A meeting point for lipids, proteins and therapies: Translational Medicine," *J. Cell. Mol. Med.*, vol. 12, no. 3, pp. 829–875, 2008, doi: 10.1111/j.1582-4934.2008.00281.x.
- [3] W. Szlasa, I. Zendran, A. Zalesińska, M. Tarek, and J. Kulbacka, "Lipid composition of the cancer cell membrane," *J. Bioenerg. Biomembr.*, vol. 52, no. 5, pp. 321–342, 2020, doi: 10.1007/s10863-020-09846-4.
- [4] O. Ali and A. Szabó, "Review of Eukaryote Cellular Membrane Lipid Composition, with Special Attention to the Fatty Acids," *Int. J. Mol. Sci.*, vol. 24, no. 21, 2023, doi: 10.3390/ijms242115693.
- [5] M. Jelokhani-Niaraki, "Membrane Proteins: Structure, Function and Motion," *Int. J. Mol. Sci.*, vol. 24, no. 1, 2023, doi: 10.3390/ijms24010468.
- [6] C. Y. Lin *et al.*, "Membrane protein-regulated networks across human cancers," *Nat. Commun.*, vol. 10, no. 1, 2019, doi: 10.1038/s41467-019-10920-8.
- [7] J. Chen *et al.*, "Systemic localization of seven major types of carbohydrates on cell membranes by dSTORM imaging," *Sci. Rep.*, vol. 6, no. March, pp. 1–10, 2016, doi: 10.1038/srep30247.
- [8] B. Huang, M. Bates, and X. Zhuang, "Super-resolution fluorescence microscopy," *Annu. Rev. Biochem.*, vol. 78, pp. 993–1016, 2009, doi: 10.1146/annurev.biochem.77.061906.092014.
- [9] N. Kumar, M. M. Drozd, H. Jiang, D. M. Santos, and D. J. Vaux, "Nanoscale mapping of newly-synthesised phospholipid molecules in a biological cell using tip-enhanced Raman spectroscopy," *Chem. Commun.*, vol. 53, no. 16, pp. 2451–2454, 2017, doi: 10.1039/c6cc10226c.
- [10] M. B. Stone, S. A. Shelby, and S. L. Veatch, "Super-Resolution Microscopy: Shedding Light on the Cellular Plasma Membrane," *Chem. Rev.*, vol. 117, no. 11, pp. 7457–7477, Jun. 2017, doi: 10.1021/acs.chemrev.6b00716.
- [11] D. Mrđenović *et al.*, "Regioselective Tip-Enhanced Raman Spectroscopy of Lipid Membranes with Sub-Nanometer Axial Resolution," *Nano Lett.*, vol. 23, no. 9, pp. 3939–3946, 2023, doi: 10.1021/acs.nanolett.3c00689.
- [12] F. Shao and R. Zenobi, "Tip-enhanced Raman spectroscopy: principles, practice, and applications to nanospectroscopic imaging of 2D materials," *Anal. Bioanal. Chem.*, vol. 411, no. 1, pp. 37–61, 2019, doi: 10.1007/s00216-018-1392-0.
- [13] R. Treffer *et al.*, "Advances in TERS (tip-enhanced Raman scattering) for biochemical applications," *Biochem. Soc. Trans.*, vol. 40, no. 4, pp. 609–614, 2012, doi: 10.1042/BST20120033.
- [14] Z. Zhang, S. Sheng, R. Wang, and M. Sun, "Tip-Enhanced Raman Spectroscopy," *Anal. Chem.*, vol. 88, no. 19, pp. 9328–9346, 2016, doi: 10.1021/acs.analchem.6b02093.
- [15] Y. Cao and M. Sun, "Tip-enhanced Raman spectroscopy," *Rev. Phys.*, vol. 8, p. 100067, 2022, doi: 10.1016/j.revip.2022.100067.
- [16] B. S. Yeo, J. Stadler, T. Schmid, R. Zenobi, and W. Zhang, "Tip-enhanced Raman Spectroscopy - Its status, challenges and future directions," *Chem. Phys. Lett.*, vol. 472, no. 1–3, pp. 1–13, 2009, doi: 10.1016/j.cplett.2009.02.023.

- [17] T. Deckert-Gaudig, A. Taguchi, S. Kawata, and V. Deckert, "Tip-enhanced Raman spectroscopy-from early developments to recent advances," *Chem. Soc. Rev.*, vol. 46, no. 13, pp. 4077–4110, 2017, doi: 10.1039/c7cs00209b.
- [18] A. Foti, S. Venkatesan, B. Lebental, G. Zucchi, and R. Ossikovski, "Comparing Commercial Metal-Coated AFM Tips and Home-Made Bulk Gold Tips for Tip-Enhanced Raman Spectroscopy of Polymer Functionalized Multiwalled Carbon Nanotubes," *Nanomaterials*, vol. 12, no. 3, 2022, doi: 10.3390/nano12030451.
- [19] R. Beams, "Tip-enhanced Raman scattering of graphene," *J. Raman Spectrosc.*, vol. 49, no. 1, pp. 157–167, 2018, doi: <https://doi.org/10.1002/jrs.5211>.
- [20] D. Mehtani *et al.*, "Chapter 6 - Scanning nano-Raman spectroscopy of silicon and other semiconducting materials," in *Tip Enhancement*, S. Kawata and V. M. Shalae, Eds. Amsterdam: Elsevier, 2007, pp. 177–203. doi: <https://doi.org/10.1016/B978-044452058-6/50007-4>.
- [21] C. Mancini *et al.*, "High-Resolution and Surface-Sensitive Tip-Enhanced Raman Spectroscopy Characterization of Strained-Silicon Devices through Cleanroom-Compatible Plasmonic Probes," *Adv. Mater. Interfaces*, vol. 12, no. 10, p. 2400876, 2025, doi: <https://doi.org/10.1002/admi.202400876>.
- [22] X. Yan, H. Sato, and Y. Ozaki, "Chapter 4 - Raman and tip-enhanced Raman scattering spectroscopy studies of polymer nanocomposites," in *Spectroscopy of Polymer Nanocomposites*, S. Thomas, D. Rouxel, and D. Ponnamm, Eds. William Andrew Publishing, 2016, pp. 88–111. doi: <https://doi.org/10.1016/B978-0-323-40183-8.00004-5>.
- [23] Y. Litman, F. P. Bonafé, A. Akkoush, H. Appel, and M. Rossi, "First-Principles Simulations of Tip Enhanced Raman Scattering Reveal Active Role of Substrate on High-Resolution Images," *J. Phys. Chem. Lett.*, vol. 14, no. 30, pp. 6850–6859, Aug. 2023, doi: 10.1021/acs.jpclett.3c01216.
- [24] M. Fadel, L. Henrard, and V. Meunier, "Calculation of tip-enhanced Raman spectra of carbon nanostructures," *Carbon Trends*, vol. 12, p. 100286, 2023, doi: <https://doi.org/10.1016/j.cartre.2023.100286>.
- [25] S. Bienz, C. Xu, Y. Xia, A. Dutta, R. Zenobi, and N. Kumar, "Advancements in Nanoscale Chemical Analysis using Tip-Enhanced Raman Spectroscopy," *Chimia (Aarau)*, vol. 79, no. 1–2, pp. 52–59, 2025, doi: 10.2533/chimia.2025.52.
- [26] S. Bonhommeau, G. S. Cooney, and Y. Huang, "Nanoscale chemical characterization of biomolecules using tip-enhanced Raman spectroscopy," *Chem. Soc. Rev.*, vol. 51, no. 7, pp. 2416–2430, 2022, doi: 10.1039/d1cs01039e.
- [27] W. Wang and V. Deckert, "Tip-Enhanced Raman Spectroscopy," in *Raman Spectroscopy: Advances and Applications*, D. K. Singh, A. Kumar Mishra, and A. Materny, Eds. Singapore: Springer Nature Singapore, 2024, pp. 233–289. doi: 10.1007/978-981-97-1703-3_11.
- [28] Z. He *et al.*, "Tip-Enhanced Raman Imaging of Single-Stranded DNA with Single Base Resolution," *J. Am. Chem. Soc.*, vol. 141, no. 2, pp. 753–757, Jan. 2019, doi: 10.1021/jacs.8b11506.
- [29] Z. He *et al.*, "Resolving the Sequence of RNA Strands by Tip-Enhanced Raman Spectroscopy," *ACS Photonics*, vol. 8, no. 2, pp. 424–430, Feb. 2021, doi: 10.1021/acsphotonics.0c01486.
- [30] D. Talaga, A. Bremner, T. Buffeteau, R. A. L. Vallée, S. Lecomte, and S. Bonhommeau, "Total Internal Reflection Tip-Enhanced Raman Spectroscopy of Cytochrome c," *J. Phys. Chem. Lett.*, vol. 11, no. 10, pp. 3835–3840, May 2020, doi: 10.1021/acs.jpclett.0c00579.
- [31] V. Sereda and I. K. Lednev, "Two Mechanisms of Tip Enhancement of Raman Scattering by Protein Aggregates," *Appl. Spectrosc.*, vol. 71, no. 1, pp. 118–128, 2017, doi: 10.1177/0003702816651890.
- [32] D. P. Cowcher, T. Deckert-Gaudig, V. L. Brewster, L. Ashton, V. Deckert, and R. Goodacre, "Detection of Protein Glycosylation Using Tip-Enhanced Raman Scattering," *Anal. Chem.*, vol. 88, no. 4, pp. 2105–2112, 2016, doi: 10.1021/acs.analchem.5b03535.
- [33] M. Tabatabaei, F. A. Caetano, F. Pashee, S. S. G. Ferguson, and F. Lagugné-Labarthe, "Tip-enhanced Raman spectroscopy of amyloid β at neuronal spines," *Analyst*, vol. 142, no. 23, pp. 4415–4421, 2017, doi: 10.1039/C7AN00744B.
- [34] R. Böhme, M. Richter, D. Cialla, P. Rösch, V. Deckert, and J. Popp, "Towards a specific characterisation of components on a cell surface – combined TERS-investigations of lipids and human cells," *J. Raman Spectrosc.*, vol. 40, no. 10, pp. 1452–1457, 2009, doi: 10.1002/jrs.2433.
- [35] M. Richter, M. Hedegaard, T. Deckert-Gaudig, P. Lampen, and V. Deckert, "Laterally resolved and direct spectroscopic evidence of nanometer-sized lipid and protein domains on a single cell," *Small*, vol. 7, no. 2, pp. 209–214, 2011, doi: 10.1002/sml.201001503.
- [36] D. Mrđenović, W. Ge, N. Kumar, and R. Zenobi, "Nanoscale Chemical Imaging of Human Cell Membranes Using

- Tip-Enhanced Raman Spectroscopy,” *Angew. Chemie - Int. Ed.*, vol. 61, no. 43, pp. 1–20, 2022, doi: 10.1002/anie.202210288.
- [37] T. Shibata *et al.*, “Photocatalytic nanofabrication and intracellular raman imaging of living cells with functionalized AFM probes,” *Micromachines*, vol. 11, no. 5, 2020, doi: 10.3390/MI11050495.
 - [38] S. L. Carrier, C. M. Kownacki, and Z. D. Schultz, “Protein-ligand binding investigated by a single nanoparticle TERS approach,” *Chem. Commun.*, vol. 47, no. 7, pp. 2065–2067, 2011, doi: 10.1039/c0cc05059h.
 - [39] L. Xiao, H. Wang, and Z. D. Schultz, “Selective Detection of RGD-Integrin Binding in Cancer Cells Using Tip Enhanced Raman Scattering Microscopy,” *Anal. Chem.*, vol. 88, no. 12, pp. 6547–6553, 2016, doi: 10.1021/acs.analchem.6b01344.
 - [40] L. Xiao, K. A. Bailey, H. Wang, and Z. D. Schultz, “Probing Membrane Receptor-Ligand Specificity with Surface- and Tip- Enhanced Raman Scattering,” *Anal. Chem.*, vol. 89, no. 17, pp. 9091–9099, 2017, doi: 10.1021/acs.analchem.7b01796.
 - [41] K. D. Alexander and Z. D. Schultz, “Tip-enhanced Raman detection of antibody conjugated nanoparticles on cellular membranes,” *Anal. Chem.*, vol. 84, no. 17, pp. 7408–7414, Sep. 2012, doi: 10.1021/ac301739k.
 - [42] E. McLeod and A. Ozcan, “Nano-imaging enabled via self-assembly,” *Nano Today*, vol. 9, no. 5, pp. 560–573, 2014, doi: <https://doi.org/10.1016/j.nantod.2014.08.005>.
 - [43] P. Verma, “Tip-Enhanced Raman Spectroscopy: Technique and Recent Advances,” *Chem. Rev.*, vol. 117, no. 9, pp. 6447–6466, 2017, doi: 10.1021/acs.chemrev.6b00821.
 - [44] T. A. Yano and S. Kawata, “Tip-Enhanced Raman Spectroscopy (TERS) for Nanoscale Imaging and Analysis,” *Wiley Blackwell 6*, vol. 9781118359, pp. 139–161, 2014, doi: 10.1002/9781118703601.ch7.
 - [45] M. Zhang, R. Wang, X. Wu, and J. Wang, “Principle, system, and applications of tip-enhanced Raman spectroscopy,” *Sci. China Physics, Mech. Astron.*, vol. 55, no. 8, pp. 1335–1344, 2012, doi: 10.1007/s11433-012-4769-7.
 - [46] L. Gao, H. Zhao, T. Li, P. Huo, D. Chen, and B. Liu, “Atomic force microscopy based tip-enhanced Raman spectroscopy in biology,” *Int. J. Mol. Sci.*, vol. 19, no. 4, 2018, doi: 10.3390/ijms19041193.
 - [47] K. C. Chang, Y. W. Chiang, C. H. Yang, and J. W. Liou, “Atomic force microscopy in biology and biomedicine,” *Tzu Chi Med. J.*, vol. 24, no. 4, pp. 162–169, 2012, doi: 10.1016/j.tcmj.2012.08.002.
 - [48] J. Stadler, T. Schmid, and R. Zenobi, “Developments in and practical guidelines for tip-enhanced Raman spectroscopy,” *Nanoscale*, vol. 4, no. 6, pp. 1856–1870, 2012, doi: 10.1039/c1nr11143d.
 - [49] G. Kolhatkar, J. Plathier, and A. Ruediger, “Nanoscale investigation of materials, chemical reactions, and biological systems by tip enhanced Raman spectroscopy-a review,” *J. Mater. Chem. C*, vol. 6, no. 6, pp. 1307–1319, 2018, doi: 10.1039/c7tc05688e.
 - [50] N. Jiang, D. Kurouski, E. A. Pozzi, N. Chiang, M. C. Hersam, and R. P. Van Duyne, “Tip-enhanced Raman spectroscopy: From concepts to practical applications,” *Chem. Phys. Lett.*, vol. 659, pp. 16–24, 2016, doi: 10.1016/j.cplett.2016.06.035.
 - [51] B. Prats-Mateu and N. Gierlinger, “Tip in-light on: Advantages, challenges, and applications of combining AFM and Raman microscopy on biological samples,” *Microsc. Res. Tech.*, vol. 80, no. 1, pp. 30–40, 2017, doi: 10.1002/jemt.22744.
 - [52] M. Miletic *et al.*, “Combined Raman and AFM detection of changes in HeLa cervical cancer cells induced by CeO₂nanoparticles-molecular and morphological perspectives,” *Analyst*, vol. 145, no. 11, pp. 3983–3995, 2020, doi: 10.1039/c9an02518a.
 - [53] H. Y. Wang, D. Bharti, and I. Levental, “Membrane Heterogeneity Beyond the Plasma Membrane,” *Front. Cell Dev. Biol.*, vol. 8, no. November, 2020, doi: 10.3389/fcell.2020.580814.
 - [54] V. D. Gordon, T. J. O’Halloran, and O. Shindell, “Membrane adhesion and the formation of heterogeneities: biology, biophysics, and biotechnology,” *Phys. Chem. Chem. Phys.*, vol. 17, no. 24, pp. 15522–15533, Jun. 2015, doi: 10.1039/c4cp05876c.
 - [55] T. Tolstik *et al.*, “Discrimination and classification of liver cancer cells and proliferation states by raman spectroscopic imaging,” *Analyst*, vol. 139, no. 22, pp. 6036–6043, 2014, doi: 10.1039/c4an00211c.
 - [56] E. Szafraniec *et al.*, “Raman spectroscopy-based insight into lipid droplets presence and contents in liver sinusoidal endothelial cells and hepatocytes,” *J. Biophotonics*, vol. 12, no. 4, pp. 1–11, 2019, doi: 10.1002/jbio.201800290.
 - [57] Z. Movasaghi, S. Rehman, and I. U. Rehman, “Raman spectroscopy of biological tissues,” *Appl. Spectrosc. Rev.*,

vol. 42, no. 5, pp. 493–541, 2007, doi: 10.1080/05704920701551530.

- [58] K. Kochan, E. Kus, A. Filipek, K. Szafrńska, S. Chlopicki, and M. Baranska, “Label-free spectroscopic characterization of live liver sinusoidal endothelial cells (LSECs) isolated from the murine liver,” *Analyst*, vol. 142, no. 8, pp. 1308–1319, 2017, doi: 10.1039/c6an02063a.
- [59] Z. Movasaghi, R. Shazza, and I. U. and Rehman, “Raman Spectroscopy of Biological Tissues,” *Appl. Spectrosc. Rev.*, vol. 42, no. 5, pp. 493–541, Sep. 2007, doi: 10.1080/05704920701551530.
- [60] D. W. Shipp, F. Sinjab, and I. Notingher, “Raman spectroscopy: techniques and applications in the life sciences,” *Adv. Opt. Photonics*, vol. 9, no. 2, p. 315, 2017, doi: 10.1364/aop.9.000315.
- [61] V. A. Arzumanyan, O. I. Kiseleva, and E. V. Poverennaya, “The Curious Case of the HepG2 Cell Line: 40 Years of Expertise,” *Int. J. Mol. Sci.*, vol. 22, no. 23, Dec. 2021, doi: 10.3390/ijms222313135.
- [62] M. Guo, H. Zhang, J. Zheng, and Y. Liu, “Glypican-3: A new target for diagnosis and treatment of hepatocellular carcinoma,” *J. Cancer*, vol. 11, no. 8, pp. 2008–2021, 2020, doi: 10.7150/jca.39972.
- [63] A. R. Devan *et al.*, “The role of glypican-3 in hepatocellular carcinoma: Insights into diagnosis and therapeutic potential,” *Eur. J. Med. Res.*, vol. 29, no. 1, p. 490, 2024, doi: 10.1186/s40001-024-02073-2.
- [64] K. Fu *et al.*, “Size-Dependent Penetration of Gold Nanoparticles into Fixed Cells,” *ACS Omega*, vol. 6, no. 5, pp. 3791–3799, 2021, doi: 10.1021/acsomega.0c05458.
- [65] S. Managò *et al.*, “Tailoring lab-on-fiber SERS optrodes towards biological targets of different sizes,” *Sensors Actuators, B Chem.*, vol. 339, no. July 2020, 2021, doi: 10.1016/j.snb.2020.129321.
- [66] S. Zhu, C. Fan, J. Wang, J. He, E. Liang, and M. Chao, “Realization of high sensitive SERS substrates with one-pot fabrication of Ag-Fe₃O₄ nanocomposites,” *J. Colloid Interface Sci.*, vol. 438, no. January, pp. 116–121, 2015, doi: 10.1016/j.jcis.2014.09.015.
- [67] I. Ettah and L. Ashton, “Engaging with Raman Spectroscopy to Investigate Antibody Aggregation,” *Antibodies*, vol. 7, no. 3, p. 24, 2018, doi: 10.3390/antib7030024.
- [68] Z. Wang *et al.*, “Raman spectroscopy for monitoring free sulfhydryl formation during monoclonal antibody manufacturing,” *J. Pharm. Biomed. Anal.*, vol. 252, no. July 2024, p. 116530, 2025, doi: 10.1016/j.jpba.2024.116530.
- [69] A. JyothiKumar, M. Kailasnath, J. Simon, and S. Mathew, “BSA stabilized gold nanoparticles: Synthesis and characterization,” *Mater. Today Proc.*, vol. 9, pp. 111–115, 2019, doi: 10.1016/j.matpr.2019.02.044.
- [70] L. Dong *et al.*, “Cytotoxicity of BSA-Stabilized Gold Nanoclusters: In Vitro and In Vivo Study,” *Small*, vol. 11, no. 21, pp. 2571–2581, 2015, doi: <https://doi.org/10.1002/sml.201403481>.
- [71] X. Huang and M. A. El-Sayed, “Gold nanoparticles: Optical properties and implementations in cancer diagnosis and photothermal therapy,” *J. Adv. Res.*, vol. 1, no. 1, pp. 13–28, 2010, doi: 10.1016/j.jare.2010.02.002.
- [72] C. Lara-Cruz, J. E. Jiménez-Salazar, E. Ramón-Gallegos, P. Damian-Matsumura, and N. Batina, “Increasing roughness of the human breast cancer cell membrane through incorporation of gold nanoparticles,” *Int. J. Nanomedicine*, vol. 11, pp. 5149–5161, 2016, doi: 10.2147/IJN.S108768.

Chapter 7: Conclusions

These chapters highlight the pioneering role of Raman-based spectroscopic techniques in advancing cancer diagnostics, offering capabilities that move decisively beyond the limitations of conventional methodologies. Among these, Raman spectroscopy, Surface-enhanced Raman spectroscopy, and Tip-enhanced Raman spectroscopy have emerged as highly versatile tools capable of revealing the cancer-related biochemical and molecular information with remarkable accuracy and sensitivity.

Raman spectroscopy, in particular, provides a non-invasive and label-free approach for probing the biochemical composition of biological samples. By capturing the unique vibrational fingerprints of molecules, it enables the identification of even subtle biochemical alterations. These spectral signatures provide critical insights into the molecular landscape of cancer, revealing changes in major cellular constituents that underlie tumor initiation, progression, and heterogeneity. SERS combines the unique optical properties and the enhancing power of plasmonic nanostructures with Raman spectroscopy, amplifying the intrinsically weak Raman signals by several orders of magnitude. This enhancement enables the detection of biomolecules at ultra-low concentrations, meeting the sensitivity required for clinical applications. TERS advances this concept further by combining the amplification capability of SERS with the spatial precision of scanning probe microscopy. Through this integration, TERS achieves nanoscale resolution and ultrasensitive molecular detection, providing an unprecedented opportunity to investigate tumor at the nanoscale.

Specifically, Chapters 2 and 3 focus on the synergistic integration of Raman spectroscopy with machine learning, showcasing its application in two different case studies: hepatocellular carcinoma (HCC) and non-small cell lung cancer, respectively. In the case of HCC, Raman spectroscopy was employed to capture the biochemical fingerprints of cancerous versus healthy cells, while machine-learning algorithms facilitated accurate discrimination and classification, underscoring the potential of this approach in cancer diagnostics. A key strength of the Raman analysis of HCC-derived primary cells was the innovative use of uncultured primary cells, rather than immortalized cell lines. Although cell lines serve as liver cancer models, their value in translational to clinical settings is limited. Over time, the accumulation of mutations causes them to lose key cellular functions and fail to capture the intrinsic heterogeneity of tumors [1]. In contrast, the Raman analysis of primary tumor and non-tumor cells presented in Chapter 2 enabled, for the first time, the identification of distinct biochemical and spectral profiles of uncultured primary samples. This provides a more accurate representation of the patient's condition and highlights the pronounced nucleic acid signals within the nuclei of cancer cells, which emerged as a key molecular marker of HCC [2]. On one hand, this allowed the unique discrimination between malignant and healthy cells—an essential capability for assessing tumor margins, where ensuring the complete absence of cancer cells is critical. On the other hand, the use of uncultured primary cancer cells more accurately reflects the biological diversity of real tumors, thereby greatly enhancing diagnostic relevance. Nevertheless, while Raman spectra provide abundant of molecular information, their interpretation remains challenging. The complexity arises from several factors, including the overlap of vibrational bands from different biomolecules, the intrinsic biological variability among cell samples and the high

dimensionality of the data. Together, these factors can obscure subtle diagnostic features, making advanced analytical and computational approaches essential for extracting meaningful insights. Therefore, this challenge required the indispensable role of artificial intelligence and machine learning, due to their capability for dimensionality reduction and classification. Accordingly, the performance of LDA- and CNN-LSTM-based models was systematically assessed in distinguishing tumor cells across three experimental settings: (i) a sample composed exclusively of tumor cells and (ii–iii) two mixed samples containing different ratios of tumor to non-tumor cells. At the single-spectrum level, the LDA-based models achieved classification accuracies of nearly 90%, whereas the best-performing CNN-LSTM models reached 93%. The tests were conducted in a completely blind manner to ensure objectivity and robustness. The overall performance demonstrated that the models were capable of identifying tumor cell spectra with consistently high accuracy, thereby confirming their reliability and potential applicability in detecting tumor cells even within heterogeneous and complex biological environments. Future studies will need to involve a larger patient cohort to improve the generalizability of machine learning models and to develop more robust and comprehensive databases, which are essential for capturing inter-patient variability, refining predictive performance, and facilitating the translation of these approaches into clinical practice. However, the findings obtained not only confirm the potential of AI-assisted Raman spectroscopy as a highly sensitive and specific diagnostic tool but also underscore its promise as a minimally invasive approach with potential for clinical translation in advanced hepatocellular carcinoma diagnostics.

In Chapter 3, the effectiveness of combination of Raman spectroscopy and machine learning as a powerful approach for molecular-level analysis was also demonstrated in non-small cell lung cancer. Early detection of circulating tumor cells poses a significant challenge for clinicians, but it plays a vital role in the prevention of tumor dissemination, reduction of intensive treatments, and significantly improvement of patient outcomes. One of the greatest issue in detecting CTCs is their extreme rarity within the bloodstream. Notably, the number of CTCs is extremely low in relation to the abundance of blood cells, which makes their identification and isolation highly complex [3]. This imbalance not only creates technical difficulties for detection methods but also demands highly sensitive and specific analytical approaches to reliably distinguish tumor cells from the surrounding background of normal blood components. The Raman analysis identified clear and distinctive molecular signatures between the cell types: NSCLC cells were characterized by elevated lipid levels, indicative of altered lipid metabolism and enhanced membrane biosynthesis, whereas leukocytes exhibited higher concentrations of nucleic acids and proteins, consistent with their unique functional and structural composition. By capturing these biochemical differences, Raman spectroscopy provides a sensitive, label-free method for molecular-level characterization of cells with fundamentally different origins and functions, enabling the unique identification of the spectral profiles of both NSCLC cell lines and healthy cells [4], [5]. An important innovation of this study was the cultivation of NSCLC cells—epithelial cells that typically grow under adherent conditions—in suspension. This approach effectively recreated the non-adherent state of CTCs in the bloodstream, providing a more physiologically relevant model. By simulating the dynamics that tumor cells experience during circulation, this strategy not only strengthens the biological validity of the experimental system but also

offers valuable insights into the molecular adaptations and survival mechanisms that support metastatic progression. This methodological advancement enhances the relevance of the model, enabling more accurate study of tumor cell dynamics under conditions that closely mimic *in vivo* environments. In addition, this approach offered valuable insights into the metabolic and molecular differences among NSCLC cell lines with distinct mutational profiles, thereby highlighting their unique metabolic landscapes. Notably, A549 cells were predominantly enriched in proteins relative to the other cancer cell lines, consistent with their less invasive phenotype and likely reflecting their origin as a primary, rather than metastatic, cell line. Specifically, H1299 and H1915 showed higher nucleic acid content, reflecting greater proliferative and metastatic potential, while H1650 was lipid-dominant, in line with its EGFR E19del mutation. Among EGFR mutations, the exon 19 deletion (E19del) is the most common (prevalence of 44%) and it is an activating mutation which led to the constitutive activation of the intracellular tyrosine kinase domain of EGFR [6], [7]. EGFR is a critical regulator of lipid metabolism, promoting lipogenesis by activating the PI3K/AKT and MAPK pathways. Through mTOR, this signaling facilitates the nuclear translocation of SREBP, a transcription factor that upregulates genes controlling fatty acid and cholesterol synthesis. Together with cholesterol, these lipids strengthen EGFR signaling by influencing membrane composition and dynamics [8]. Overall, A549 is biosynthetically active, H1299 and H1915 are aggressive and nucleic acid-rich, and H1650 displays altered lipid metabolism, demonstrating how Raman spectroscopy can capture tumor heterogeneity and invasive potential allowed for rapid and precise identification and discrimination of these cells. Furthermore, the use of PCA-LDA model alongside Raman spectroscopy, achieved an impressive 96.40% classification accuracy. Future research will focus on extending the analysis to real CTC samples, with the goal of applying the Raman-PCA-LDA platform to clinical cases and thereby further enhancing its potential to support lung cancer diagnosis. By enabling fast, high-throughput, and accurate profiling of tumor cells, the proposed approach shows promise for early detection of circulating NSCLC cells, real-time monitoring of disease progression, and development of personalized therapies. Overall, it highlights the potential of Raman-based liquid biopsy technologies as powerful tools for fundamental cancer research and future clinical applications, translating molecular-level insights into more precise and patient-focused care.

After validating the integration of Raman spectroscopy with machine learning as a diagnostic tool for cancer, the next step was to design more versatile and sensitive platforms capable of amplifying the intrinsically weak Raman signal, enabling potential applications in SERS. In this context, SERS combined with optical fibers offers a powerful strategy to achieve high sensitivity and *in situ* detection. Conventional SERS sensors are often limited by their rigidity and inefficient light management. In the Chapter 5, these challenges were addressed by employing optical fibers, which offer compact dimensions, low cost, and the ability to enable real-time, remote measurements—features that are particularly valuable for biomedical applications [9]. Specifically, the study investigated the development and performance of Au-coated zinc oxide nanopillars integrated onto optical fiber tips, tested in two configurations (direct and indirect) and with both 633 nm and 785 nm lasers. The research focused on lab-on-tip configurations, where SERS-active platforms are integrated directly onto the fiber end facet. The fabrication involved hydrothermal synthesis of ZnO

nanopillars, exploring both ordered (masked) and disordered (unmasked) growth approaches.

The SERS performance of the investigated optical fiber probes strongly depends on both the interrogation configuration and the excitation wavelength. In the direct configuration, the masked FZN07 fiber consistently outperformed the other samples at both 785 nm and 633 nm, exhibiting the highest SERS intensity and reproducibility. Notably, at 785 nm, ordered masked fiber with extended growth durations (e.g., FZN07 with 6 hours) consistently showed the highest intensities of the BPT Raman band at 1585 cm^{-1} and reproducibility compared to all other masked and unmasked samples. Furthermore, at 633 nm, FZN07 provided the optimal balance between signal strength and reproducibility, highlighting the beneficial role of masking and ordered nanostructuring in enhancing and stabilizing the SERS response. With 633 nm excitation in direct mode, generally higher SERS intensities were observed across all samples, attributed to better alignment with plasmon resonance and the inverse fourth-power dependence of scattering intensity on wavelength. In contrast, under indirect interrogation, the masked FZN06 fiber demonstrated superior performance at 785 nm, achieving the highest BPT-to-silica signal ratio and thus the most effective suppression of the silica background. At 633 nm, the unmasked FZN01 and masked FZN06 fibers showed comparable BPT-to-silica ratios, indicating that both configurations can provide similar signal quality under these conditions. In particular, in the indirect configuration, 785 nm excitation was generally preferred due to reduced silica background interference, making the masked FZN06 fiber well suited for remote sensing applications.

These findings underscore the versatility of Au-coated ZnO nanostructured fibers as SERS platforms and the crucial need to balance structural ordering and excitation parameters for optimal performance. The optimal performance is context-dependent, varying with the experimental configuration. This work significantly contributes to developing more robust, deployable SERS-active optical fiber sensors for further real-time and minimally invasive cancer diagnostics. Subsequently, the focus then shifted to tip-enhanced Raman spectroscopy for nanoscale cellular analysis. In Chapter 6, a proof-of-concept study applied TERS for the first time to label-free investigation of HepG2 liver cancer cell membranes, enabling direct characterization of biomolecules of the cell surface and capturing subtle variations in lipid and protein organization. Specifically, TERS analysis highlighted biochemical heterogeneity between individual cells, with the first cell exhibiting elevated levels of cholesterol and fatty acids as well as pronounced phenylalanine signals compared to the second cell. These observations highlight the role of cholesterol-enriched membrane domains in cancer progression and align with previous findings that HepG2 cells contain elevated phenylalanine and choline derivatives, suggesting potential therapeutic targets [10], [11]. Combining TERS with SERS tags allowed selective detection of specific biomarkers, such as Glypican-3, a heparan sulfate proteoglycan overexpressed in HCC that promotes hepatocellular carcinoma development and progression [12]. This approach enables precise identification of GPC-3 distribution on the cell membrane while maintaining nanoscale spatial resolution. By integrating label-free molecular profiling with targeted biomarker detection, this dual strategy provides a comprehensive view of both the biochemical composition and receptor organization of liver cancer cell membranes. Such insights are

critical for advancing our understanding of tumor biology and for the development of precision diagnostic and therapeutic strategies.

Despite significant advancements, challenges persist across all spectroscopic modalities. For conventional Raman spectroscopy, addressing cellular heterogeneity, expanding spectral datasets, and extending studies to larger patient cohorts for HCC or to real circulating tumor cells in non-small cell lung cancer, will be essential for future applications. These improvements will facilitate the development of more robust AI algorithms for Raman data processing, enhancing classification accuracy and predictive performance.

SERS, particularly when integrated with optical fibers, offers unique opportunities for sensitive, remote, and minimally invasive detection, but it still faces significant challenges. One major limitation is the reproducibility of the SERS-active substrates on fiber tips, which can lead to variability in signal intensity. To address this, ongoing efforts focus on optimizing the chemical fabrication of ZnO nanopillar nanostructures, aiming to produce highly ordered architectures with uniform plasmonic coatings. Such improvements are expected to enhance both the strength and consistency of SERS signals while maintaining the flexibility and compactness inherent to optical fiber platforms.

TERS, while offering unparalleled spatial resolution, still faces challenges including tip reproducibility, potential laser-induced sample damage from local heating, interference from far-field background signals, and the complexity of spectral interpretation. Future efforts will focus on performing TERS mapping over larger membrane areas rather than single-point measurements and optimizing mapping parameters to achieve high-resolution, reproducible characterization of cell membranes. The integration of these advanced spectroscopic techniques represents a transformative approach to cancer diagnostics. Together, these advances promise a new era in cancer diagnostics—one that combines label-free molecular insight, targeted biomarker detection, and nanoscale resolution to facilitate earlier diagnosis, more precise disease characterization, and personalized therapeutic strategies. Continued refinement of these approaches and translation to clinical settings could significantly improve patient outcomes and guide precision oncology in the coming years.

7.1 References

- [1] M. Richter, O. Piwocka, M. Musielak, I. Piotrowski, W. M. Suchorska, and T. Trzeciak, “From Donor to the Lab: A Fascinating Journey of Primary Cell Lines,” *Front. Cell Dev. Biol.*, vol. 9, no. July, pp. 1–11, 2021, doi: 10.3389/fcell.2021.711381.
- [2] C. Esposito *et al.*, “Assessment of Primary Human Liver Cancer Cells by Artificial Intelligence-Assisted Raman Spectroscopy,” *Cells*, vol. 12, no. 22, Nov. 2023, doi: 10.3390/cells12222645.
- [3] S. Sharma *et al.*, “Circulating tumor cell isolation, culture, and downstream molecular analysis,” *Biotechnol. Adv.*, vol. 36, no. 4, pp. 1063–1078, 2018, doi: 10.1016/j.biotechadv.2018.03.007.
- [4] K. D. Sutherland and A. Berns, “Cell of origin of lung cancer,” *Mol. Oncol.*, vol. 4, no. 5, pp. 397–403, 2010, doi: 10.1016/j.molonc.2010.05.002.
- [5] G. Litwack, “Chapter 21 - Blood and Lymphatic System,” in *Human Biochemistry*, G. Litwack, Ed. Boston: Academic Press, 2018, pp. 681–708. doi: <https://doi.org/10.1016/B978-0-12-383864-3.00021-1>.
- [6] L. T. Huang, S. L. Zhang, C. B. Han, and J. T. Ma, “Impact of EGFR exon 19 deletion subtypes on clinical outcomes in EGFR-TKI-Treated advanced non-small-cell lung cancer,” *Lung Cancer*, vol. 166, no. October 2021, pp. 9–16, 2022, doi: 10.1016/j.lungcan.2022.01.014.

- [7] C. wei Xu *et al.*, “Molecular Characteristics and Clinical Outcomes of EGFR Exon 19 C-Helix Deletion in Non–Small Cell Lung Cancer and Response to EGFR TKIs,” *Transl. Oncol.*, vol. 13, no. 9, p. 100791, 2020, doi: 10.1016/j.tranon.2020.100791.
- [8] K. Eltayeb, S. La Monica, M. Tiseo, R. Alfieri, and C. Fumarola, “Reprogramming of Lipid Metabolism in Lung Cancer: An Overview with Focus on EGFR-Mutated Non-Small Cell Lung Cancer,” *Cells*, vol. 11, no. 3, 2022, doi: 10.3390/cells11030413.
- [9] A. Ricciardi *et al.*, “Lab-on-fiber technology: A new vision for chemical and biological sensing,” *Analyst*, vol. 140, no. 24, pp. 8068–8079, 2015, doi: 10.1039/c5an01241d.
- [10] M. Maja and D. Tyteca, “Alteration of cholesterol distribution at the plasma membrane of cancer cells: From evidence to pathophysiological implication and promising therapy strategy,” *Front. Physiol.*, vol. 13, no. November, pp. 1–26, 2022, doi: 10.3389/fphys.2022.999883.
- [11] V. A. Arzumanyan, O. I. Kiseleva, and E. V. Poverennaya, “The Curious Case of the HepG2 Cell Line: 40 Years of Expertise,” *Int. J. Mol. Sci.*, vol. 22, no. 23, Dec. 2021, doi: 10.3390/ijms222313135.
- [12] L. Wang, M. Yao, L. H. Pan, Q. Qian, and D. F. Yao, “Glypican-3 is a biomarker and a therapeutic target of hepatocellular carcinoma,” *Hepatobiliary Pancreat. Dis. Int.*, vol. 14, no. 4, pp. 361–366, 2015, doi: 10.1016/S1499-3872(15)60396-4.