

Electronic Thesis and Dissertation Repository

---

9-30-2021 9:30 AM

## Characterization of Extracellular Vesicles by Surface-Enhanced Raman Spectroscopy

Nina M. Culum, *The University of Western Ontario*

Supervisor: Laguné-Labarthet, François, *The University of Western Ontario*

A thesis submitted in partial fulfillment of the requirements for the Master of Science degree in Chemistry

© Nina M. Culum 2021

Follow this and additional works at: <https://ir.lib.uwo.ca/etd>

 Part of the [Analytical Chemistry Commons](#), and the [Physical Chemistry Commons](#)

---

### Recommended Citation

Culum, Nina M., "Characterization of Extracellular Vesicles by Surface-Enhanced Raman Spectroscopy" (2021). *Electronic Thesis and Dissertation Repository*. 8183.  
<https://ir.lib.uwo.ca/etd/8183>

This Dissertation/Thesis is brought to you for free and open access by Scholarship@Western. It has been accepted for inclusion in Electronic Thesis and Dissertation Repository by an authorized administrator of Scholarship@Western. For more information, please contact [wlsadmin@uwo.ca](mailto:wlsadmin@uwo.ca).

## Abstract

Extracellular vesicles (EVs), which are nanoscale vesicles secreted by cells into biofluids, are of research interest due to their roles in intercellular communication. EVs released from mesenchymal stromal cells (MSCs) have tremendous potential in cell-free regenerative medicine, while EVs released from diseased cells are being studied as biomarkers for minimally invasive and early disease detection. Presented in this thesis are gold nanohole arrays for the capture and sensitive detection of EVs by surface-enhanced Raman spectroscopy (SERS), a plasmonic technique capable of single molecule detection. Herein, we have characterized EVs released from MSCs and ovarian cancer cells, with a focus on cell lines that have been underexplored by SERS in literature. Using a hybrid principal component analysis-machine learning approach, we have demonstrated the platform's potential in classifying EV groups with high (~ 99 %) accuracy, sensitivity, and specificity, which we hope will one day translate to point-of-care detection for disease diagnosis.

## Keywords

Extracellular vesicles, nanohole arrays, surface-enhanced Raman spectroscopy, plasmonics, electron-beam lithography, mesenchymal stromal cells, ovarian cancer, machine learning, biosensing

## Summary for Lay Audience

The fields of rehabilitative and diagnostic medicine are constantly evolving, where the former is seeking safer and more effective ways to repair tissue and organ damage, and the latter is developing methods for rapid, non-invasive, and early disease detection. One area of research with applications in both these fields is extracellular vesicle (EV)-based technology. EVs are a complex group of membrane-bound vesicles released from cells into biofluids including blood, saliva, and urine. EVs are traditionally separated into three subclasses based on attributes such as size, biomolecular cargo, and mechanisms of formation and release. The most interesting subclasses of EVs consist of exosomes and microvesicles since they play roles in intercellular communication. Released from mesenchymal stromal cells (MSCs), EVs have shown immense promise in cell-free regenerative and restorative applications, while EVs released from diseased cells (e.g., cancer cells) are studied for their applications as disease biomarkers. However, the nanoscale size and molecular heterogeneity of EVs pose a significant research problem since sensitive methods are required for their detection.

Surface-enhanced Raman spectroscopy (SERS) is a sensitive, non-destructive plasmonic technique that has shown potential in biosensing applications, including EV detection. The highly sensitive nature of SERS is based on the collective oscillation of free electrons at a nanoscale, metallic surface. Consequently, large electromagnetic fields are produced at the edges of the nanoscale features, and analytes that are confined to these regions experience significant enhancement with respect to their Raman signal intensity. Proposed in this work are gold nanohole arrays for the capture and SERS characterization of EVs. The SERS spectra gathered provide insight into the biochemical composition of EVs, and EVs from both MSCs and ovarian cancer cells are explored. SERS characterization of EVs from the specific cell sources investigated in this thesis have been largely underexplored in literature, and so the work presented here is novel. Statistical analysis is utilized to find patterns in the complex spectra acquired, and machine learning is further implemented to classify EVs from various cell sources. The high (~ 99 %) accuracies, sensitivities, and specificities reported in this thesis demonstrate great promise for translation to clinical testing.

## Co-Authorship Statement

Prof. François Lagugné-Labarhet was the corresponding author on the papers presented in this thesis and was responsible for the supervision of Nina Culum during her studies. This thesis includes materials from two publications, all of which had contributions from co-authors.

Nina Culum was the primary authors of the papers presented in Chapters 3 and 4. She was responsible for the acquisition and processing of most of the experimental work in the papers, as well as the writing and revision for all drafts, including the final manuscripts. Throughout this work, data was additionally acquired and analyzed by co-authors as summarized below:

**Chapter 3:** Culum, N. M.; Cooper, T. T.; Bell, G. I.; Hess, D. A.; Lagugné-Labarhet, F. Characterization of Extracellular Vesicles Derived from Mesenchymal Stromal Cells by Surface Enhanced Raman Spectroscopy. *Anal. Bioanal. Chem.* **2021**, *413*, 5013-5024.

Dr. Tyler T. Cooper was a post-doctoral fellow in Prof. David A. Hess' group at Robarts Research Institute. Dr. Cooper assisted in the conceptualization of the manuscript and was responsible for the cell culturing and isolation of the EVs. Gillian I. Bell is a research technician in Professor Hess' group who was also responsible for collecting EV samples.

**Chapter 4:** Culum, N. M.; Cooper, T. T.; Lajoie, G. A.; Dayarathna, T.; Pasternak, S. H.; Liu, J.; Fu, Y.; Postovit, L.-M.; Lagugné-Labarhet, F. Characterization of Ovarian Cancer-Derived Extracellular Vesicles Derived by Surface Enhanced Raman Spectroscopy. *Analyst* **2021**, Accepted manuscript, DOI: 10.1039/d1an01586a.

Dr. Tyler T. Cooper is a post-doctoral fellow in Prof. Gilles A. Lajoie's group in the Department of Biochemistry at the University of Western Ontario. Dr. Cooper assisted in the conceptualization of the manuscript and was responsible for the proteomic analysis of the EVs. Dr. Thamara Dayarathna is a research associate in Prof. Stephen H. Pasternak's group at Robarts Research Institute who gathered nanoscale flow cytometry data for the manuscript. Dr. Jiahui Liu is a research associate in Prof. Lynne-Marie Postovit's group in the Department of Oncology at the University of Alberta who was responsible for cell culturing and EV isolation. Prof. Postovit is also associated with the Department of Biomedical & Molecular Sciences at Queen's University. Professor Yangxin Fu, in the Department of Oncology at the

University of Alberta, kindly provided primary cell lines. The human immortalized surface epithelial cells were also kindly provided by Dr. David Huntsman at the BC Cancer Agency.

## Dedication

To my grandmother, Gospava Simendić, the most loving and resilient woman I know

## Acknowledgments

First, I would like to thank my supervisor, Prof. François Lagurné-Labarhet. This project would not have been possible without your guidance and expertise. You have provided me with the tools and knowledge to develop this project into what it has become today, and have given me immense opportunities to grow and communicate my skills. You have challenged me to become a more critical thinker, to step out of my comfort zone, and to actively seek out new literature each week. The scientist I have become with your support is unrecognizable from the student I was when I first joined your group. For that, I am deeply grateful.

I must also thank Dr. Danielle McRae, Alex Therien, and Maria Olivia Avilés in the FLL group for their advice and training during my first year at Western, as well as Dr. Joachim Jelken for (often) troubleshooting and repairing instruments. I would also like to acknowledge FLL group members Dr. Leila Mazaheri, Pooneh Farhat, and Sydney Legge. I am very thankful to have been surrounded by such bright researchers and wonderful people. Our daily long lunches and coffee breaks were ended too early by the pandemic, and I will continue to miss them after I graduate.

I would like to deeply thank Tim Goldhawk and Dr. Todd Simpson of the Western Nanofabrication Facility for their training and assistance throughout the nanofabrication process. This project could not have succeeded without your advice and troubleshooting, of which plenty was needed. I am also very grateful to all of our collaborators at the University of Western Ontario, University of Alberta, and Queen's University. Specifically, I must thank Dr. Tyler Cooper for helping structure the projects presented in this thesis and for his expertise on EVs. Thank you for quickly and patiently responding to the many emails I have sent over the years!

I am grateful to my family, for their unconditional love, and my friends (both old and new), for always being there to support me. A special thank you goes out to one of my closest and oldest friends, Maja Lopandić, for graciously proofreading many of my academic works over these past two years, including this thesis. I can't wait to return the favour! Lastly, I must thank my fiancé. You have shown me endless love, encouragement, and patience, and you have helped me through some of the most stressful moments of my academic career.

# Table of Contents

Abstract.....	ii
Summary for Lay Audience.....	iii
Co-Authorship Statement.....	iv
Dedication.....	vi
Acknowledgments.....	vii
Table of Contents.....	viii
List of Tables.....	xii
List of Figures.....	xiv
List of Appendices.....	xx
List of Abbreviations, Symbols, and Nomenclature.....	xxi
Chapter 1.....	1
1 General Introduction.....	1
1.1 EV Background.....	1
1.1.1 Overview of EV Classification.....	2
1.1.2 EVs in Intercellular Communication.....	6
1.2 Cellular Origins.....	7
1.2.1 Mesenchymal Stromal Cell (MSC)-Derived EVs.....	8
1.2.2 Cancer-Derived EVs.....	9
1.3 Methodologies for EV Detection and Characterization.....	10
1.3.1 Common Physical Characterization Methods.....	11
1.3.2 Conventional Biochemical Characterization Methods and their Limitations.....	13
1.3.3 Nanoplasmonic Approaches for Biochemical Characterization.....	17
1.4 Scope of Thesis.....	20
1.5 References.....	20



Chapter 2.....	34
2 Principles of EV Detection and Characterization .....	34
2.1 Principles of Raman Spectroscopy .....	34
2.1.1 Instrumentation for Raman Spectroscopy.....	36
2.1.2 Raman Spectroscopy and Biological Samples.....	37
2.2 Principles of SERS.....	40
2.2.1 Chemical Enhancement .....	41
2.2.2 EM Enhancement.....	42
2.2.3 Materials for SERS .....	44
2.3 Spectral Interpretation.....	45
2.3.1 Principles of PCA .....	45
2.3.2 Machine Learning .....	47
2.4 SERS Probes and Platforms for EV Capture and Characterization .....	52
2.4.1 Nanoparticle Probes for Direct Sensing.....	52
2.4.2 Periodic Arrays of Nanostructures for Direct Sensing .....	59
2.4.3 Probe-Based Immunocomplexes for Indirect Sensing.....	65
2.4.4 Sandwich-Type Immunoassays for Indirect Sensing.....	70
2.5 Summary.....	75
2.6 References.....	75
Chapter 3.....	86
3 Characterization of Extracellular Vesicles Derived from Mesenchymal Stromal Cells by Surface-Enhanced Raman Spectroscopy.....	86
3.1 Introduction.....	86
3.2 Methods.....	88
3.2.1 Nanohole Array Fabrication by EBL.....	88
3.2.2 Absorption Measurements .....	89

3.2.3	Cell Culture.....	89
3.2.4	EV Isolation .....	89
3.2.5	AFM Measurements.....	90
3.2.6	SERS Measurements.....	90
3.2.7	Statistical Analysis and Machine Learning.....	91
3.3	Results and Discussion .....	91
3.3.1	Nanohole Array Fabrication and Characterization .....	91
3.3.2	EV Characterization.....	95
3.3.3	Statistical Analysis and Machine Learning.....	102
3.4	Conclusions.....	105
3.5	References.....	106
Chapter 4	.....	111
4	Characterization of Ovarian Cancer-Derived Extracellular Vesicles by Surface-Enhanced Raman Spectroscopy .....	111
4.1	Introduction.....	111
4.2	Methods.....	113
4.2.1	Cell Culture.....	113
4.2.2	EV Isolation .....	114
4.2.3	EV Protein Extraction and Digestion.....	114
4.2.4	Ultraperformance Liquid Chromatography Coupled to Tandem Mass Spectrometry (UPLC-MS/MS) .....	115
4.2.5	UPLC-MS/MS Data Analysis.....	116
4.2.6	nFC Measurements .....	116
4.2.7	Nanohole Array Fabrication .....	117
4.2.8	SERS Measurements.....	117
4.2.9	Statistical Analysis and Machine Learning.....	117
4.3	Results and Discussion .....	118

4.3.1	Validation of EV Size and Proteomic Cargo .....	118
4.3.2	EV Characterization by SERS .....	122
4.3.3	Spectral Analysis by PCA.....	130
4.3.4	Machine Learning for Ovarian Cancer Diagnosis .....	133
4.4	Conclusion .....	137
4.5	References.....	138
Chapter 5	.....	142
5	Conclusions and Outlook.....	142
5.1	References.....	148
Appendix A:	Confusion Matrix Calculations .....	151
Appendix B:	Copyrights .....	152
Curriculum Vitae	.....	157

## List of Tables

Table 1-1: Characterization of EVs based on their size, density, contents, biogenesis, and mechanism of release. ....	6
Table 3-1: Assignments of common peaks shared among Panc-MS C EV spectra from Figure 3-5. ....	98
Table 3-2: Assignments of common peaks shared among BM-MS C and Panc-MS C EV spectra from Figure 3-6. ....	101
Table 3-3: Comparison of abilities of PCA and various machine learning algorithms to discriminate BM-MS C and Panc-MS C EVs based on SERS spectra. ....	105
Table 4-1: Parameters for Q Exactive™ Plus. ....	115
Table 4-2: Summary of peak assignments of hIOSE EVs, as highlighted in Figure 4-4A. ...	122
Table 4-3: Summary of peak assignments of OV-90 EVs, as highlighted in Figure 4-4B. ...	124
Table 4-4: Summary of peak assignments of OVCAR3 EVs, as highlighted in Figure 4-4B. ....	125
Table 4-5: Summary of peak assignments of EOC6 EVs, as highlighted in Figure 4-4C. ....	127
Table 4-6: Summary of peak assignments of EOC18 EVs, as highlighted in Figure 4-4C. ..	128
Table 4-7: SERS peaks that best differentiate normal vs. cancer groups, as highlighted in Figure 4-7A. ....	133
Table 4-8: SERS peaks that best differentiate low-grade vs. high-grade cancer groups, as highlighted in Figure 4-7B. ....	133
Table 4-9: Comparison of accuracies, sensitivities, and specificities achieved with the six machine learning algorithms shown in Figure 4-8. ....	134
Table 4-10: Comparison of accuracies, sensitivities, and specificities achieved with each group compared in Figure 4-8. ....	137

Table A-1: Example confusion matrix. .... 151

# List of Figures

Figure 1-1: (A) Schematic illustration of the three subclasses of EVs and their biogenesis and release with (B) transmission electron microscope (TEM) images of the three EV types for comparison. TEM images are adapted with permission from reference [24] (copyright 2016 Elsevier B. V.). ..... 4

Figure 1-2: Schematic illustration of the three general mechanisms of intercellular communication via EV exchange between a secreting cell and a recipient cell. .... 7

Figure 1-3: Schematic illustration of some high-resolution microscopic techniques for EV characterization, including (A) TEM, (B) SEM, and (C) AFM. .... 13

Figure 1-4: Schematic illustration of two common immunoassay techniques for EV detection and characterization, including (A) ELISA, which highlights four common detection methods, and (B) LFIA, which highlights sample flow along four main components. .... 16

Figure 1-5: (A) Schematic illustration of an SPP at the interface of a dielectric material and metal surface, and (B) the strength of the resulting electric field, which decays exponentially away from the metal surface. (C) Schematic illustration of the LSPR of a metal nanoparticle induced by light. .... 19

Figure 2-1: Schematic illustration of the Rayleigh and Raman (Stokes and anti-Stokes) scattering processes. .... 35

Figure 2-2: Schematic illustration of a general Raman spectrometer setup, where the green beams and arrows represent the incident photons and the red beams and arrows represent the Raman-scattered photons. .... 37

Figure 2-3: Raman spectral regions associated with proteins, nucleic acids, lipids, and carbohydrates. .... 38

Figure 2-4: Schematic illustration of the process of PCA. (A) The original 5-variable data set can be visually represented by a maximum of 3 axes. The first PC is fitted so that it contains the maximum variance, and the second is positioned perpendicular to it. (B) The 5 calculated PCs can be visualized in a scree plot, and (C) a score plot is created by re-casting the original data on PC1 and PC2 axes. .... 46

Figure 2-5: Schematic illustration of the process of supervised machine learning. .... 48

Figure 2-6: Schematic illustrations of some machine-learning algorithms, highlighting (A) decision trees, (B) feed-forward ANNs, (C) Bayesian networks, (D) kNNs, and (E) SVMs. .... 51

Figure 2-7: A 100 nm AuNP-covered SERS substrate for the analysis of lung cancer-derived EVs. (A) SEM images and an optical image (inset, top left) of the AuNPs assembled over an APTES-covered substrate. (B) Schematic illustration of the acquisition of SERS spectra. (C) Lung cancer cell line-derived EVs and normal lung cell line-derived EVs are used to train a deep learning algorithm for lung cancer diagnosis. (D) The deep learning model is tested with EVs derived from lung cancer patients and healthy individuals, and their similarities to the lung cancer cell-derived EVs are compared. Adapted with permission from reference [113] (copyright 2020 American Chemical Society). .... 54

Figure 2-8: DMAP-coated AuNPs with silver nanoshells are used as SERS probes for the analysis of melanoma-derived EVs. Schematic illustrations of SERS acquisition (A) without and (B) with the silver nanoshell, highlighting the closer proximity of the EV membrane to the SERS-active metal surface when the probes are coated *in situ* with a silver layer. Analysis of melanoma cell line-derived EVs (red cluster) and RBC-derived EVs (green cluster) (C) without and (D) with the silver nanoshell, illustrating higher discriminatory power when the silver nanoshell is implemented. Adapted with permission from reference [117] (copyright 2019 American Chemical Society). .... 57

Figure 2-9: FASERS substrate for the analysis of breast cancer-derived EVs and cell lysates. (A) Schematic illustration of FASERS. Each SAM uniquely provides insight into the biological composition of a given analyte. (B) Schematic illustration (top), top-view SEM image (bottom left), and tilted-view SEM image (bottom right) of the bare FASERS (scale bars = 400 nm).

(C) SERS spectra of breast cancer cell line-derived EVs collected with each FASERS. Adapted with permission from reference [126] (copyright 2020 Yarovsky, I. & Stevens M. M.). ..... 61

Figure 2-10: A beehive-inspired three-dimensional gold-coated TiO<sub>2</sub> MIO structure for the analysis of prostate cancer, lung cancer, liver cancer, and colon cancer derived-EVs. (A) Schematic illustration of the structure fabrication and acquisition of SERS spectra. (B) SEM image of the substrate that has captured 4 EVs, which are circled in red (scale bar = 300 nm). (C) The P-O band intensities of EVs derived from cancer patients are significantly higher on average compared to the P-O band intensities of EVs derived from healthy individuals. Adapted with permission from reference [130] (copyright 2020 American Chemical Society). ..... 64

Figure 2-11: Sandwich-type immunocomplex for the analysis of breast cancer, colorectal cancer, and prostate cancer-derived EVs. (A) Schematic illustration of the fabrication of capture probes and SERS probes, as well as the formation of the immunocomplex and acquisition of SERS spectra. (B) Simultaneous detection of breast cancer, colorectal cancer, and prostate cancer markers demonstrated using patient-derived EVs. Adapted with permission from reference [120] (copyright 2018 Royal Society of Chemistry). ..... 67

Figure 2-12: EV-probe assembly for the SERS analysis of pancreatic cancer-derived EVs. (A) Schematic illustration of the uptake of SERS probes into EVs and the affinity of Fe<sub>3</sub>O<sub>4</sub> core-TiO<sub>2</sub> shell MBs to the EV surface. (B) SERS maps of normal pancreatic cell line-derived EVs (left) and pancreatic cancer cell line-derived EVs (right), highlighting the higher SERS intensity observed when the target is present. (C) Higher characteristic peak intensity is also observed with EVs derived from clinical samples. Adapted with permission from reference [134] (copyright 2021 American Chemical Society). ..... 69

Figure 2-13: Flocked gold nanopillars for the SERS analysis of breast cancer-derived EVs. (A) Schematic illustration of the fabrication and functionalization of the nanopillars. (B) Top-view (left) and tilt-view (right) SEM images of upright nanopillars. (C) Top-view (left) and tilt-view (right) SEM images of head-flocked nanopillars. Adapted with permission from reference [137] (copyright 2019 Wiley-VCH). ..... 72



Figure 2-14: (A) TEM image of a gold octahedral self-assembled monolayer. (B) Schematic illustration of the sensing principle of the structure, in which the hairpin structure P opens upon hybridization with let-7a, forcing the Cy5 molecules (red star) farther away from the gold surface. (B) SERS spectra of Cy5 in the (a) absence and (b) presence of let-7a, highlighting the change its change in intensity when the target is present. Adapted with permission from reference [139] (copyright 2021 American Chemical Society). ..... 75

Figure 3-1: (A) Schematic illustration of nanohole array fabrication by EBL using a negative-tone resist, with example SEM images of 700 nm circular arrays (B) before lift-off and (C) after lift-off (scale bars = 2  $\mu\text{m}$ ). ..... 92

Figure 3-2: SEM images of 1000 nm (A) triangular, (B) square, and (C) circular arrays, and 500 nm (D) triangular, (E) square, and (F) circular arrays (scale bars = 1  $\mu\text{m}$ ). (G) Background-corrected absorption spectra of 500 – 1000 nm triangular arrays and (H) corresponding raw absorption spectra. .... 93

Figure 3-3: (A) SERS map comprised of 100 spectra of BM-MSV EVs over an area of 16 1.0  $\mu\text{m}$ -sized square nanoholes, where high intensity areas (blue) correspond to trapped EVs and low intensity areas (red) correspond to empty holes. (B) SERS map from (A) smoothed and overlaid with an optical image of the array and (C) corresponding spectra of points (1) – (3) from (B), with spectra shifted vertically for clarity. .... 95

Figure 3-4: (A)  $(5 \times 5) \mu\text{m}^2$  AFM scan of Panc-MSV EVs on a bare glass coverslip and (B) the cross-section of a single EV, as indicated by the white line in (A). (C)  $(1 \times 1) \mu\text{m}^2$  AFM scan of an individual BM-MSV EV on a bare glass coverslip. (D) A histogram representing the height distribution of a Panc-MSV sample containing 106 EVs. .... 96

Figure 3-5: Three SERS spectra of Panc-MSV EVs acquired with a 632.8 nm laser with a background spectrum for comparison and common peaks highlighted in blue. .... 97

Figure 3-6: SERS spectra of (A) Panc-MSV EVs and (B) BM-MSV EVs acquired with a 785 nm laser, where individual spectra are represented by lighter lines, average spectra are represented by darker lines, and common spectral peaks are highlighted in yellow. .... 100

Figure 3-7: PCA score plots of Panc-MSC and BM-MSC clusters constructed with (A) the first and second PCs and (B) the first and third PCs, where one data point corresponds to one SERS spectrum. (C) Comparison of final output scores of data given by PCA only (PC1 vs. PC3) and five machine learning algorithms, where one data point corresponds to one SERS spectrum. (D) ROC curves comparing various machine learning algorithms. .... 104

Figure 4-1: Schematic illustration of nFC and UPLC-MS/MS analysis of OvCa EVs. .... 118

Figure 4-2: (A) Representative flow cytometry plots (top) and corresponding size distributions (bottoms) of EV samples from OV-90 (left) and hIOSE (middle) cell lines, as well as conditioned media (right). (B) nFC results revealed a near-linear decrease in the number of particles as the volume of sample analyzed diminished. The horizontal dashed line indicates the mean particle number when an equal volume of background media (PBS) was analyzed. (C) Size distribution of EVs. .... 119

Figure 4-3: (A) Number of MS1 and MS2 scans for duplicate injections of each EV sample. (B) Number of peptides identified by de novo sequencing, which led to confident protein identification with false discovery rate ( $p < 0.01$ ). (C) Venn diagram highlighting the distinct proteomics cargo contained within EVs from various OvCa cell lines, leading to distinct proteomic “fingerprints” as demonstrated by (D) PCA. (E) Overlap of OvCa EV proteomes with Vesiclepedia database filtered for OvCa. (F) Heat map indicating that classical EV markers (e.g., CD9/81/63) were confidently identified in all OvCa EV samples. .... 121

Figure 4-4: Average SERS spectra of (A) ovarian epithelial cell line-derived EVs (control), (B) established OvCa cell line-derived EVs, and (C) primary OvCa cell line-derived EVs. Average SERS spectra of hIOSE, OV-90, OVCAR3, EOC6, and EOC18 are comprised of 207, 123, 106, 166, and 156 spectra, respectively. .... 129

Figure 4-5: PC score plots comparing (A) normal (hIOSE) vs. cancer (OV-90, OVCAR3) EVs, (B) normal vs. high-grade cancer (EOC6) EVs, (C) normal vs. low-grade cancer (EOC18) EVs, and (D) all groups to highlight separation along PC18. In (A) – (C), each point corresponds to one SERS spectrum while the points in (D) correspond to the centroids of each group in (A) – (C). .... 131

Figure 4-6: Average SERS spectra of EVs derived from each cell type (also shown in Figure 4-5) compared with (A) PC18 and (B) PC25 loading spectra. Key peaks (i.e., spectral peaks that best describe each PC) are highlighted with vertical dashed lines. .... 132

Figure 4-7: ROC curves comparing six different machine learning algorithms: SVM, logistic regression, random forest, kNN, Naïve Bayes, and CN2 rule inducer. The upper left-most portion is zoomed in and highlighted in red (right). .... 134

Figure 4-8: Heat maps (left) to visualize output scores of (A) normal *vs.* established cancer cell line EVs, (B) normal *vs.* high-grade cancer EVs, (C) normal *vs.* low-grade cancer EVs, and (D) low-grade *vs.* high-grade cancer EVs, with corresponding confusion matrices (right). In the heat maps, each horizontal bar corresponds to one SERS spectrum. Groups labeled on the left indicate the origin of the sample, while the color of the bar indicates the algorithm's prediction. In (A) – (C), a red or dark orange color (output score > 0.5) corresponds to a cancer prediction and a yellow or light orange color (output score < 0.5) corresponds to a normal prediction. In (D), a red or dark orange color (output score > 0.5) corresponds to a high-grade cancer prediction and a yellow or light orange color (output score < 0.5) corresponds to a low-grade cancer prediction. .... 136

Figure 5-1: (A) Schematic illustration comparing photolithography, with proximity printing shown as an example, and EBL for nanohole array fabrication. (B) Schematic illustration of photolithography by projection printing. .... 145

Figure 5-2: Schematic illustration of antibody functionalization by EDC chemistry. .... 147

## List of Appendices

Appendix A: Confusion Matrix Calculations .....	151
Appendix B: Copyrights .....	152

## List of Abbreviations, Symbols, and Nomenclature

2NAT	2-naphthalenethiol
ABC	ammonium bicarbonate
AFM	atomic force microscope
AGC	automatic gain control
AgNC	silver nanocube
AgNP	silver nanoparticle
ALIX	apoptosis-linked gene 2-interacting protein X
ANN	artificial neural network
Anti-CD9	CD9 antibody
Anti-CD63	CD63 antibody
Anti-EpCAM	epithelial cell adhesion molecule antibody
Anti-MCSP	melanoma chondroitin sulphate proteoglycan antibody
APTES	(3-aminopropyl)triethoxysilane
ATCC	American Type Tissue Collection
AUC	area under the curve
AuNP	gold nanoparticle
AuNR	gold nanorod
AuNS	gold nanostar
BC	bacterial cellulose
BCA	bicinchronic assay
BM-MSC	bone marrow-mesenchymal stromal cell
c	concentration
CCD	charge-coupled device
CD-R	compact disc-recordable
CEA	carcinoembryonic antigen
CHMP4	charged multivesicular body protein 4
circRNA	circular ribonucleic acid
CM	conditioned media
CTAB	cetrimonium bromide
CV	cross validation

Cy3	cyanine dye-3
Cy5	cyanine dye-5
DA	discriminant analysis
DCA	discriminant correspondence analysis
DFA	discriminant function analysis
DLS	dynamic light scattering
DMAP	4-(dimethylamino)pyridine
DNA	deoxyribonucleic acid
DTNB	5,5-dithiobis-(2-nitrobenzoic acid)
DTT	dithiothreitol
DVD-R	digital video disc-recordable
EBL	electron-beam lithography
EDC	1-ethyl-3-(3-dimethylaminopropyl)carbodiimide
EF	enhancement factor
EF <sub>chem</sub>	chemical enhancement factor
EF <sub>EM</sub>	electromagnetic enhancement factor
EGFR	epidermal growth factor receptor
E <sub>i</sub>	energy level i
ELISA	enzyme-linked immunosorbent assay
E <sub>loc</sub>	local electric field
EM	electromagnetic
EOC	epithelial ovarian cancer
EPAC	extracellular vesicle phenotype analyzer chip
EpCAM	epithelial cell adhesion molecule
ESCRT	endosomal sorting complex required for transport
EV	extracellular vesicle
FA	formic acid
FASERS	functionalized array for surface-enhanced Raman spectroscopy
FBS	fetal bovine serum
GC	gas chromatography
GO	graphene oxide
HER2	human epidermal growth factor receptor 2

hIOSE	human immortalized ovarian surface epithelial cell
HISTH2A	histone H2A type 1
HPAEC	primary human pulmonary artery endothelial cells
HPDE	human pancreatic duct epithelial cells
h $\nu$	light
I	intensity
IB	immunoblotting
IGFR	insulin-like growth factor
ILV	intraluminal vesicle
IPA	isopropanol
ISEV	International Society for Extracellular Vesicles
k	Boltzmann constant
kNN	<i>k</i> -nearest neighbors
LC	liquid chromatography
LDA	linear discriminant analysis
let	lethal
LFIA	lateral flow immunoassay
LNA	locked nucleic acid
lncRNA	long non-coding ribonucleic acid
LNGFR	low-affinity nerve growth factor receptor
LOOCV	leave-one-out cross validation
LR	logistic regression
LSPR	localized surface plasmon resonance
MB	magnetic bead
MBA	4-mercaptobenzoic acid
MCAM	melanoma cell adhesion molecule
MCSP	melanoma chondroitin sulphate proteoglycan
MHC-II	major histocompatibility class-II
MIBK	methyl isobutyl ketone
MIO	macroporous inverse opal
miRNA	micro ribonucleic acid
MMC	2,7-mercapto-4-methylcoumarin

MPY	4-mercaptopyridine
mRNA	messenger ribonucleic acid
MS	mass spectrometry
MSC	mesenchymal stromal cell
MVB	multivesicular body
n	number of molecules
N.A.	numerical aperture
NB	Naïve Bayes
nFC	nanoscale flow cytometry
NHA	normal human astrocytes
OvCa	ovarian cancer
Panc-MS	pancreatic-mesenchymal stromal cell
<i>p</i> ATP	4-aminobenzenethiol
PBS	phosphate-buffered saline
PC	principal component
PCA	principal component analysis
PDA	polydopamine
PDAC	pancreatic ductal adenocarcinoma
PLS	partial least squares
qRT-PCR	quantitative reverse transcription-polymerase chain reaction
RBC	red blood cell
RF	random forest
RI	refractive index
RIE	reactive ion etching
RNA	ribonucleic acid
ROC	receiver operating characteristic curve
RT	room temperature
SAM	self-assembled monolayer
SEM	scanning electron microscope
SERS	surface-enhanced Raman spectroscopy
SPP	surface plasmon polariton
SPR	surface plasmon resonance



SPRi	surface plasmon resonance imaging
SVM	support vector machine
T	temperature
T-cell	thymus cell
TEM	transmission electron microscope
TFA	trifluoroacetic acid
TFMBA	2,3,5,6-tetrafluoro-4-mercaptobenzoic acid
TN	triple-negative
TSG101	tumor susceptibility gene 101
TSPAN8	tetraspanin-8
UPLC	ultra-high performance liquid chromatography
UV	ultraviolet
$v_0$	ground vibrational level
$v_1$	first excited vibrational level
vis-NIR	visible-near infrared
VPS4	vascular protein sorting-associated protein 4
$\sigma$	Raman scattering cross-section
$\omega$	frequency

# Chapter 1

## 1 General Introduction

Extracellular vesicles (EVs) are complex, membrane-bound, nanoscale vesicles that are secreted by nearly all cells into human biofluids. EVs play an important role in intercellular signaling and communication by the transfer of proteins and RNA to near and distant cells. There is a strong research interest in EVs since they have high potential as non-invasive disease biomarkers, may additionally offer prognostic information in a variety of diseases, and can be used in therapeutic applications. Discussed in this introductory chapter are the classification of EVs in terms of size, contents, and biogenesis as well as current detection methods and their limitations. Nanoplasmonic techniques for EV characterization, which are used extensively throughout this thesis, are also introduced in this chapter.

### 1.1 EV Background

Interestingly, the concept of minute-sized cell derivatives dates back approximately 150 years to Darwin's pangenesis theory.<sup>[1]</sup> This theory postulated that every cell type in the human body generates small particles called gemmules which communicate to other cell types through the transfer of molecules and may also mediate the maternal-fetal transfer of heritable information. Although this theory was not accepted at the time due to lack of experimental evidence, scientists today can relate the concept of Darwin's gemmules to EVs that indeed carry nucleic acid and protein cargo of their parent cell,<sup>[2,3]</sup> and also mediate maternal-fetal dialogue.<sup>[4]</sup>

Although historically thought of as carriers of cell waste, EVs have long been implicated in cell development,<sup>[5]</sup> intercellular signaling,<sup>[6]</sup> cellular stress responses,<sup>[7]</sup> and cell maturation.<sup>[8]</sup> Early work in EV research in 1969 recognized gene expression via EVs and their functional roles when EVs were found to be associated with epiphyseal cartilage matrix calcification in mice.<sup>[5]</sup> Around the same time, EVs were identified in the periaxonal space within the mouse atrium and a model was proposed for neuronal signaling.<sup>[6]</sup> Half a decade later, early evidence of the association of EVs with mechanisms for cellular stress was uncovered.<sup>[7]</sup> In the 1980s, it was observed sheep reticulocytes shed their transferrin

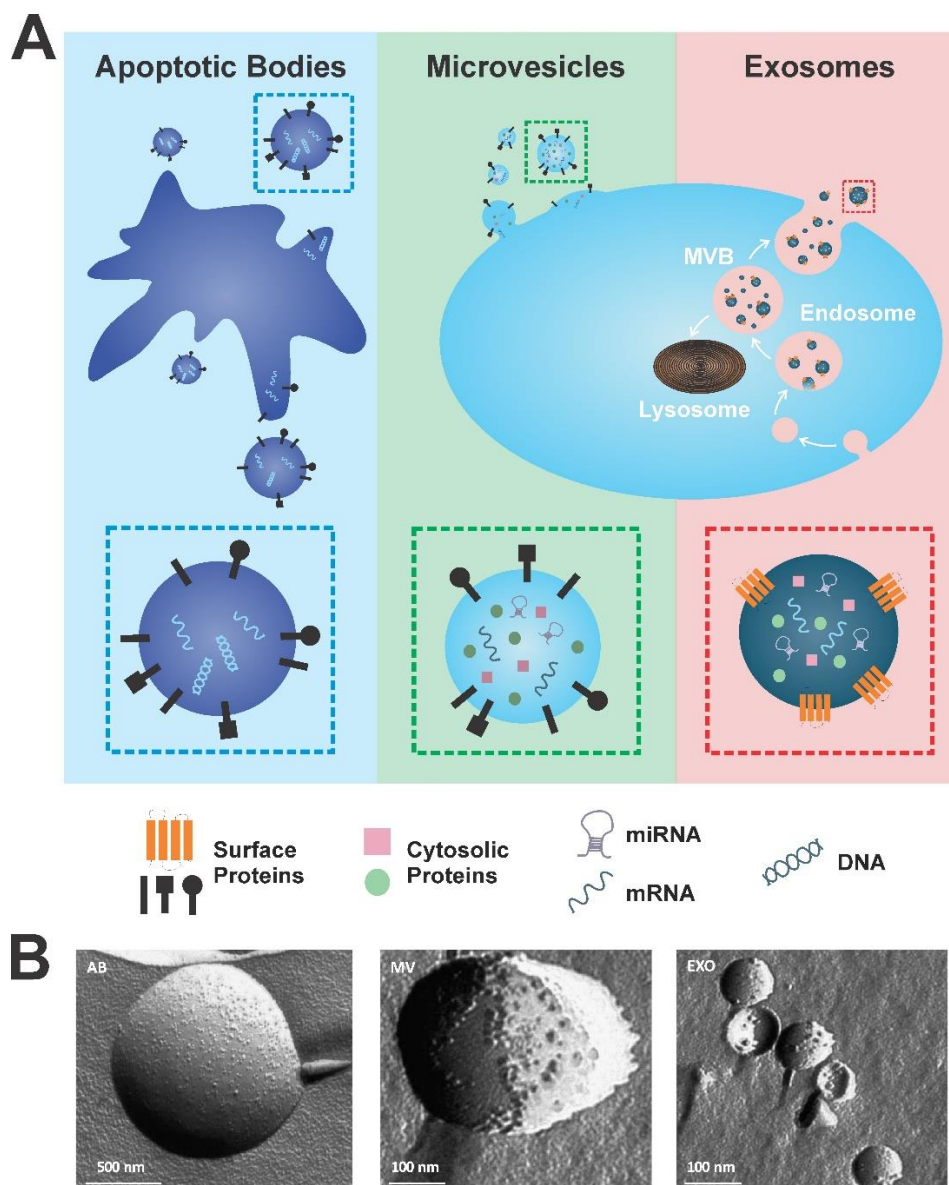
receptors by releasing EVs during maturation and it was theorized that EV secretion was a mechanism for the removal of membrane components no longer need during reticulocyte maturity.<sup>[8,9]</sup> These findings, however, were misinterpreted, leading to the misidentification of EVs as simply a means of the disposal of unwanted cellular components or cellular waste.<sup>[10]</sup>

Nearly three decades after EVs were first identified, serious work in elucidating EV signaling functions began to be published, and EVs were finally established as intercellular communication agents. In 1996, it was observed that B-cells from both humans and mice secrete EVs carrying the major histocompatibility class-II (MHC-II) molecule initiated T-cell responses.<sup>[11]</sup> Later studies further reported EV-associated signaling activities such as enhancing cell proliferation, survival, adhesion, and chemotaxis.<sup>[12]</sup> Multiple studies in 2006 through 2008 elucidated EV nucleic acid cargo and determined that EV-derived RNAs were transferrable to recipient cells,<sup>[13-15]</sup> leading to the exponential increase in EV research over the past decade.

### 1.1.1 Overview of EV Classification

EVs are secreted by nearly all cell types and are found in biofluids including blood, urine, breast milk, cerebrospinal fluid, amniotic fluid, and ascites.<sup>[16]</sup> EVs are generally separated into three subclasses: apoptotic bodies, microvesicles, and exosomes (**Figure 1-1**). Multiple factors are considered when dividing extracellular vesicles into these three subclasses, including size, chemical and biomolecular composition, biogenesis, and mechanism of formation.<sup>[17]</sup> Apoptotic bodies are the largest subclass of EVs in terms of their size distribution, as they typically have a diameter range of 1 – 5  $\mu\text{m}$ .<sup>[18]</sup> Apoptotic bodies are formed, as their name suggests, from cells in the late stages of apoptosis, and are released by directly budding from the cell's plasma membrane.<sup>[17]</sup> They typically contain cytosolic contents including proteins, RNA, and fragmented DNA, as well as cellular organelles.<sup>[19]</sup> Microvesicles, like apoptotic bodies, are also released by directly budding from the plasma membrane.<sup>[20]</sup> However, microvesicles are not released by cells undergoing apoptosis, so they are more interesting to researchers in their roles in intercellular communication. Microvesicles are also smaller than apoptotic bodies as their diameter range is approximately 100 – 1000 nm.<sup>[17]</sup> Microvesicles also contain contents

that are reflective of their parent cell and carry surface markers (e.g., integrins and selectins), cytosolic proteins, lipids, and nucleic acids such as messenger RNAs (mRNAs), microRNAs (miRNAs), circular RNAs (circRNAs) and long noncoding RNAs (lncRNAs).<sup>[21]</sup> Exosomes are the smallest subclass of EVs in terms of size with a diameter range of approximately 30 – 150 nm.<sup>[18]</sup> Exosome cargo and content is similar to microvesicles, albeit with distinct surface markers such as tetraspanins CD9, CD63, and CD81,<sup>[22]</sup> and their main distinction is in their biogenesis. Exosomes do not bud directly from the plasma membrane, but are rather released following the fusion of multivesicular bodies (MVBs) with the plasma membrane.<sup>[23]</sup> MVBs contain intraluminal vesicles (ILVs), which once expelled from the cell, are termed exosomes. Alternatively, ILVs may fuse with lysosomes for the degradation of their content.



**Figure 1-1: (A) Schematic illustration of the three subclasses of EVs and their biogenesis and release with (B) transmission electron microscope (TEM) images of the three EV types for comparison. TEM images are adapted with permission from reference [24] (copyright 2016 Elsevier B. V.).**

ILVs are formed within MVBs by the invagination of late endosomal membranes. During this process, certain proteins are incorporated into the invaginating membrane while cytosolic contents are engulfed within the ILVs. There are two pathways that are proposed for the mechanism of ILV formation: one that is dependent on the endosomal sorting complex required for transport (ESCRT) function, and one that is ESCRT-independent.<sup>[25]</sup>

A complex protein machinery made up of four separate ESCRTs (0 through III) work together to facilitate the formation of MVBs, vesicle budding, and protein cargo-sorting.<sup>[26,27]</sup>

The ESCRT-dependent mechanism is initiated by the recognition and sequestration of ubiquitinated proteins to specific domains of the endosomal membrane via the ubiquitin-binding subunits of ESCRT-0. After interaction with ESCRT-I and ESCRT-II, the total complex combines with ESCRT-III, which is the protein complex that is involved in promoting the budding process. The MVB membrane buds are cleaved to form ILVs and the ESCRT-III complex separates from the membrane using energy supplied by a sorting protein (VPS4). ESCRT components (e.g., TSG101 and CHMP4), ESCRT-associated proteins (e.g., ALIX), and ubiquitinated proteins have been identified in exosomes originating from various cell types, thus providing evidence for the ESCRT-dependent pathway.<sup>[25]</sup>

The ESCRT-independent biogenesis mechanism instead depends on raft-based microdomains for the lateral segregation of cargo within the endosomal membrane and emphasizes the key role of exosomal lipids in exosome biogenesis. Microdomains (i.e., highly enriched sphingomyelinases) can form ceramides via the hydrolytic removal of the phosphocholine moiety.<sup>[25]</sup> Ceramides induce the lateral phase separation and coalescence of microdomains in membranes. Furthermore, the cone-shape structure of ceramides may cause the spontaneous negative curvature of the endosomal membrane, thus promoting domain-induced budding. The ESCRT-independent mechanism is supported by the fact that ILVs can be formed despite silencing key subunits of all ESCRT complexes.<sup>[28,29]</sup>

Researchers warn that although it is useful to conceptualize EV biogenesis and classify EVs by either direct plasma membrane budding or fusion of the plasma membrane with MVBs, this is an oversimplified model that does not entirely encapsulate or explain the complexity of EVs. Furthermore, the distinction between the two pathways is becoming less clear as researchers discover and elucidate pathway interdependencies and cell specialization. For example, the interdependencies of EV biogenesis pathways can lead to EVs that biochemically and biophysically resemble an endosomally-derived EV (i.e., an

exosome) that were produced by budding directly from the plasma membrane.<sup>[30]</sup> This phenomenon indicates that cells can indeed specialize to operate beyond a binary classification of EV biogenesis pathways. Furthermore, while it is conceptually simple to classify EVs based on their size (i.e., exosomes are smaller than microvesicles, which are smaller than apoptotic bodies), there is substantial overlap of size distributions among each group, and there exist variations in the size distributions from author to author. Nevertheless, the generally accepted criteria for EV classification are summarized in **Table 1-1**.

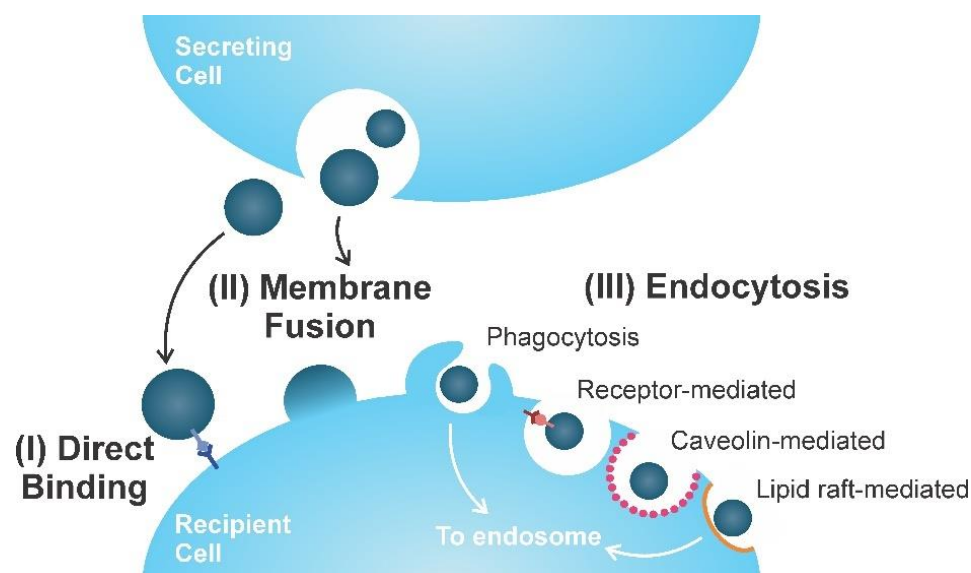
**Table 1-1: Characterization of EVs based on their size, density, contents, biogenesis, and mechanism of release.**

Feature	Exosomes	Microvesicles	Apoptotic Bodies
Size	30 – 150 nm	100 – 1000 nm	1 – 5 $\mu$ m
Content	Proteins (e.g., CD9/63/81, ALIX, TSG101), lipids, RNAs (e.g., mRNAs, miRNAs, lncRNAs), DNAs	Proteins (e.g., CD154), lipids, RNAs (e.g., mRNAs, miRNAs, lncRNAs), DNAs	Cytosolic content (proteins, RNAs, fragmented DNA) and cellular organelles
Mechanism of Release	Release of ILVs from MVBs following the fusion of MVBs with plasma membrane	Direct outward budding from the plasma membrane	Plasma membrane blebbing during apoptosis

### 1.1.2 EVs in Intercellular Communication

EVs, specifically microvesicles and exosomes, play an important role in intercellular signaling and communication either by transmitting information via direct contact or through the transfer of proteins and nucleic acids,<sup>[31]</sup> and can occur by three general mechanisms (**Figure 1-2**). The first mechanism involves the “docking” of secreted EVs onto the membrane of the recipient cell via a ligand-receptor interaction. This interaction

involves the transfer of information between cells rather than the delivery of EV cargo into the recipient cell, and examples include immunomodulation<sup>[11,32]</sup> and morphogen signaling.<sup>[33,34]</sup> The second mechanism involves the direct fusion of secreted EVs with the membrane of the recipient cell, resulting in the direct uptake of proteins and nucleic acids from the donor cell into the cytosol of the recipient cell. For example, the delivery of mRNAs,<sup>[14]</sup> miRNAs,<sup>[35,36]</sup> lncRNAs,<sup>[37]</sup> DNAs,<sup>[38,39]</sup> and oncoproteins<sup>[40,41]</sup> by EVs has been reported. The third mechanism, which also results in the direct uptake of donor cell cargo, involves the ingestion of secreted EVs into the recipient cell via endocytosis.<sup>[42,43]</sup> Endocytosis is a broad term for cellular internalization, and can occur, for example, by phagocytosis (i.e., the secreted EV is “eaten” by the recipient cell), or be receptor-mediated, caveolin-mediated, or lipid raft-mediated.<sup>[44]</sup> For successful intercellular communication to occur, particularly via receptor-mediated pathways, secreted EVs must contain recognition molecules that will direct them to the proper target and be accepted onto or into the recipient cell.



**Figure 1-2: Schematic illustration of the three general mechanisms of intercellular communication via EV exchange between a secreting cell and a recipient cell.**

## 1.2 Cellular Origins

Since EVs play an important role in intercellular communication, there is strong research interest in their therapeutic roles as well as their roles as biomarkers in disease detection,



diagnosis, and prognosis. EVs that are released from mesenchymal stromal cells (MSCs) are considered to be potent cell-free regenerative and curative agents.<sup>[45]</sup> EVs have also been implicated as mediators of many cardiovascular diseases (e.g., heart failure and obesity),<sup>[46]</sup> neurodegenerative diseases (e.g., Alzheimer's and Parkinson's),<sup>[47]</sup> inflammatory diseases (e.g., arthritis),<sup>[48]</sup> and musculoskeletal diseases.<sup>[49]</sup> One of the most prominent research areas of EVs is their role as cancer biomarkers, since they have been associated with many cancers such as ovarian,<sup>[50]</sup> prostate,<sup>[51]</sup> and breast cancer.<sup>[52]</sup>

### 1.2.1 Mesenchymal Stromal Cell (MSC)-Derived EVs

MSCs, which are multipotent stem cells capable of differentiating into many different cell types, serve to maintain and repair tissues *in vivo*.<sup>[53]</sup> Although found in a variety of different sources in the body, MSCs are most commonly isolated from bone marrow, adipose tissue, and umbilical cord blood.<sup>[54]</sup> MSCs that are administered locally or systemically can release a mixture of bioactive molecules, including EVs, which can promote the activation of endogenous repair pathways and lead to anti-inflammatory, anti-apoptotic, and restorative effects.<sup>[55,56]</sup> However, EVs isolated from MSCs can be used as therapeutic agents by themselves, and have shown similar and even superior functional capacities compared to MSCs. There are additionally safety advantages of administering EVs over living cells, since EVs are non-oncogenic, less immunogenic, less toxic, and easier to sterilize, handle, and store than living cells.<sup>[57]</sup>

MSC-derived EVs have demonstrated to be effective therapeutic agents in many conditions and diseases. For example, EVs have been used in the therapy of lung diseases,<sup>[58]</sup> such as pulmonary infections (including COVID-19),<sup>[59,60]</sup> asthma,<sup>[61,62]</sup> acute respiratory distress syndrome,<sup>[63,64]</sup> idiopathic pulmonary fibrosis,<sup>[65,66]</sup> and cystic fibrosis.<sup>[67,68]</sup> In addition to lung disease therapies, MSC-derived EVs have been implicated in tissue engineering and many regenerative therapies.<sup>[69]</sup> In neurological regeneration, EVs have been used following nerve crush injury,<sup>[70-72]</sup> stroke,<sup>[73,74]</sup> traumatic brain injury,<sup>[75]</sup> and spinal cord injury.<sup>[76-78]</sup> MSC-derived EVs have also been used to improve motor function outcomes in multiple sclerosis<sup>[79]</sup> and autoimmune encephalitis,<sup>[80]</sup> as well as to improve cognitive function recovery in Alzheimer's disease.<sup>[81]</sup> In the case of cardiovascular regeneration, EVs have been used to reduce infarct size,<sup>[82]</sup> increase myocardial viability,<sup>[83]</sup> and increase

cardiac repair<sup>[84]</sup> following myocardial ischemia/reperfusion injury. MSC-derived EVs have also been in bone regeneration,<sup>[85-87]</sup> cartilage regeneration,<sup>[88-90]</sup> muscle regeneration,<sup>[91-93]</sup> kidney regeneration,<sup>[94-96]</sup> and liver regeneration.<sup>[97-99]</sup> Lastly, EVs have been used in wound healing therapies including photo-damaged dermal fibroblasts,<sup>[100]</sup> excisional skin wounds,<sup>[101]</sup> ischemic wounds,<sup>[102]</sup> cutaneous wounds,<sup>[103]</sup> and second-degree burns.<sup>[104]</sup>

Although the application of EVs in therapeutic and regenerative medicine is a very interesting area of research, this thesis will focus on the development of methods for the sensitive detection, characterization, and analysis of MSC-derived EVs. Clinical applications of MSC-derived EVs in therapeutics is challenged by the lack of standardized techniques for EV isolation, purification, characterization, and analysis. The topic of the characterization of MSC-derived EVs is explored in detail in Chapter 3.

## 1.2.2 Cancer-Derived EVs

In 2020 alone, approximately 19.3 million new cancer cases were diagnosed globally, and nearly 10.0 million cancer deaths occurred.<sup>[105]</sup> Cancer remains one of the leading causes of premature death worldwide,<sup>[106]</sup> and so there is a need for the development of effective methods for cancer treatment and prevention. During cancer development and progression, a tumor develops for a long time at the subclinical or microscopic level before it metastasizes; to be clinically detectable, a tumor must grow to a size of approximately 1 cm<sup>3</sup>, containing about 10<sup>9</sup> cells.<sup>[107]</sup> Since there is a high probability of prior dissemination at the time of diagnosis, methods for early cancer diagnosis must be developed to combat the global mortality of cancer. Since EVs have a specific profile of miRNAs, proteins, and lipids that mirror the cargo of their parent cells, EVs can be used as biomarkers for cancer detection, and can even be used to monitor cancer progression and drug resistance.<sup>[108,109]</sup> EV-derived miRNAs have been identified as potential biomarkers of many different cancers, including liver,<sup>[110]</sup> lung,<sup>[111,112]</sup> gastrointestinal,<sup>[113]</sup> colorectal,<sup>[114]</sup> pancreatic,<sup>[115]</sup> breast,<sup>[116]</sup> and ovarian cancer,<sup>[117]</sup> as well as melanoma.<sup>[118-120]</sup> Furthermore, it has been observed that cancer cells secrete a higher number of EVs compared to their normal or non-malignant counterparts.<sup>[121]</sup>

In addition to acting as cancer fingerprints, EVs are involved in a wide range of processes involved in cancer progression, including inflammatory responses, angiogenesis, lymphogenesis, cell migration and proliferation, immune suppression, and metastasis.<sup>[121-123]</sup> Cancer-derived EVs have also been shown to influence non-cancer cells to generate a tumor microenvironment conducive for tumor growth by influencing endothelial cells to support neoangiogenesis,<sup>[124]</sup> as well as metastasis by inducing vascular permeability.<sup>[125]</sup> Furthermore, cancer-derived EVs have been shown to not only suppress immune responses,<sup>[126]</sup> but to also modify immune cells towards pro-tumorigenic and pro-metastatic phenotypes through the transfer of functional oncoproteins from cancerous donor cells to recipient cells.<sup>[125,127]</sup> EV transfer of miRNAs as well as oncogenic entities (e.g., mutated proteins, fusion gene mRNAs, and lncRNAs) from cancer cells to neighboring cells within the tumor microenvironment can not only drive malignancy,<sup>[128]</sup> but lead to drug resistance.<sup>[129]</sup>

Clinical translation of EV technology in cancer diagnosis, especially at the early stage, is hindered by the lack of standardization of methods for the isolation, quantification, characterization, and analysis of EVs, particularly from complex biofluids like blood. Another challenge is the fact that the percentage of total blood vesicles that are cancer-derived is unknown,<sup>[130]</sup> even for advanced stage cancers that are associated with the release of a higher number of EVs compared to early stage cancers. Therefore, it is currently unknown if a given blood sample will contain levels of cancer-derived EVs within the range of current technological capabilities for EV detection.<sup>[121]</sup> Although EVs are involved in many facets of cancer biology, the focus of this thesis is the sensitive detection and characterization of EVs for early cancer diagnosis, which is the topic of Chapter 4.

### 1.3 Methodologies for EV Detection and Characterization

EVs are generally characterized by at least two methods. One of these methods is a physical characterization method to confirm that the analytes are within the acceptable size ranges for EVs, as well as for EV enumeration and morphology. Although this is an important study, physical characterization alone is not suitable for diagnostic studies, and so

biochemical characterization is required. Biochemical characterization is used to determine the surface composition of EVs as well as the composition of their cargo.

### 1.3.1 Common Physical Characterization Methods

Commonly used methods for EV size quantification are dynamic light scattering (DLS), nanoparticle tracking analysis (NTA), and flow cytometry. Microscopy techniques can also be used for EV enumeration and size quantification in addition to providing insight into EV morphology and topography. Commonly used microscopy techniques for EV characterization include transmission electron microscopy (TEM), scanning electron microscopy (SEM), and atomic force microscopy (AFM).

Both DLS and NTA measure the size of particles based on their Brownian motion in solution, the basis of which is that lighter particles will diffuse faster, and that speed is relative to particle size.<sup>[131]</sup> In DLS, EVs are illuminated with a laser beam and the light is scattered. The intensity changes of scattered light are recorded as a function of time, thus allowing the Brownian motion of EVs to be observed.<sup>[132]</sup> While DLS can detect particles as small as 10 nm in diameter, the detection of particles in a heterogenous solution is biased towards larger particles since they scatter more light compared to smaller particles.<sup>[132]</sup> Therefore, in a solution containing exosomes and microvesicles, the average EV diameter calculated is more likely to be skewed towards microvesicle size. In NTA, the Brownian motion of individual EVs is tracked and their size and total concentration is calculated based on their mean velocity.<sup>[131]</sup> NTA can detect EVs in the range of approximately 50 – 1000 nm, which nearly encompasses the generally accepted size ranges for exosomes and microvesicles.<sup>[133]</sup> NTA however cannot typically detect particles larger than 1000 nm in diameter since larger particles have limited Brownian motion, and therefore move too slowly.

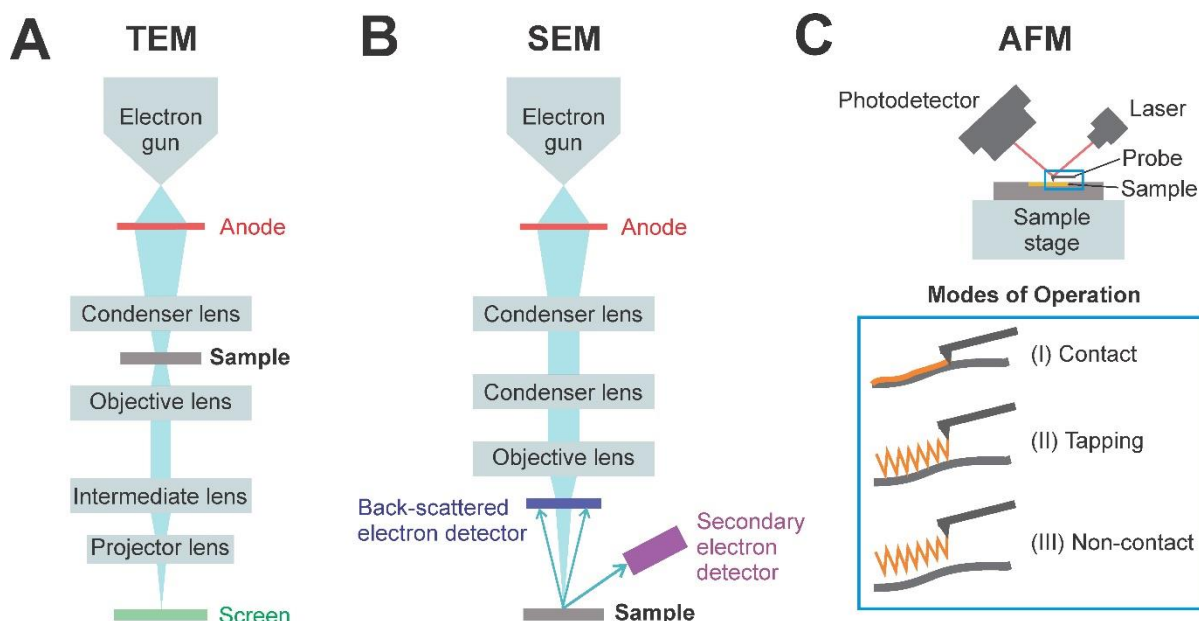
Flow cytometry has been commonly used over the past couple of decades for EV analysis. With a resolution of approximately 300 – 500 nm, depending on the setup, flow cytometry is not as sensitive as DLS or NTA, but it is advantageous in that it offers a higher throughput and multiple markers can be determined simultaneously.<sup>[132]</sup> Flow cytometry operates by directing a laser beam through a stream of sheath fluid containing suspended

EVs, which cause light scattering, similarly to DLS or NTA. At the same time, the EVs are labelled with fluorescent dyes either in their interior or on their surface, thus providing insight into their cellular origins.<sup>[134]</sup> Nanoscale flow cytometry (nFC) has emerged in more recent years, following the same operating principles as conventional flow cytometry, but capable of detecting particles as small as approximately 100 nm.<sup>[131]</sup> Additionally, nFC can discriminate EVs from other nanoparticles or contaminants in the solution.

Electron microscopy techniques like TEM and SEM are attractive for imaging EVs since they offer much higher resolution compared to light microscopy techniques due to the fact that the wavelength of an electron is approximately 100 000-fold smaller than the wavelength of a photon in visible light.<sup>[135]</sup> In TEM, stained, thin sections of a preserved sample are imaged by passing a broad electron beam through the sample above a fluorescent screen or charge coupled device (CCD) (**Figure 1-3A**).<sup>[134]</sup> The resulting two-dimensional image provides insight into the internal structure of the sample. In SEM, a focused beam of electrons is scanned across the sample surface, interacting with atoms on the sample surface (**Figure 1-3B**). Secondary electrons are collected and counted by a detector, producing a three-dimensional image containing information of the sample's surface topography.<sup>[134]</sup> Both TEM and SEM offer sub-nanometer resolution,<sup>[135,136]</sup> but require laborious sample preparation when imaging biological samples like EVs, which must be cryogenically or chemically preserved.<sup>[137]</sup> Freezing the EVs also causes them to shrink, and so volume lost must be accounted for in order to accurately determine EV size.

Like electron microscopy techniques, AFM can offer high resolution images of EVs. Images are generated by quantifying the forces between a probe (cantilever tip with a very low spring constant, usually composed of silicon or silicon nitride) and the sample surface (**Figure 1-3C**).<sup>[138]</sup> The cantilever scans the surface while a piezoelectric crystal raises or lowers the cantilever to maintain constant bending. A laser beam is constantly reflected from the top of the cantilever towards a position-sensitive photodetector, recording the actual position of the cantilever. Since AFM ages are three-dimensional, height information can be obtained simultaneously with lateral information, with resolutions of approximately 0.001 nm and 0.1 – 1.0 nm, respectively.<sup>[138]</sup> The tip can interact with the surface topography of the sample either by directly tracing the sample (contact mode), oscillating

vertically at or below its resonant frequency and gently tapping the sample surface (tapping mode), or oscillating above the sample surface with a small amplitude at a frequency larger than its resonant frequency (non-contact mode).<sup>[139]</sup> An advantage of AFM over TEM and SEM is that minimal sample preparation is required for EV imaging, since scans can be conducted in air or even in fluid, thereby mimicking the native conditions of the EVs.<sup>[137]</sup> However, high resolution scans are very time consuming, especially when large areas are imaged.



**Figure 1-3: Schematic illustration of some high-resolution microscopic techniques for EV characterization, including (A) TEM, (B) SEM, and (C) AFM.**

Although physical characterization of EVs is important for the quantification of their size, these methods alone are not suitable for applications such as disease diagnosis. They are usually used to supplement such studies and confirm the presence of EVs in a given sample, but diagnostic studies require at least one type of biochemical characterization method to determine EV chemical composition.

### 1.3.2 Conventional Biochemical Characterization Methods and their Limitations

Specific proteins can be detected by immunoblotting (IB) techniques that utilize labelled antibodies for protein capture. For EV detection by IB, purified EVs are generally lysed to

release their proteins. Western blotting is commonly used to confirm the presence of EVs in a sample by detecting EV-associated surface proteins like CD9, CD63, and CD81, as well as EV-associated cytosolic proteins like ALIX and TSG101.<sup>[134]</sup> Common protein contaminants like apolipoproteins, albumin, and uromodulin that are often co-isolated with EVs in complex samples (e.g., blood) can also be identified in this manner. However, IB techniques like Western blotting are semi-quantitative, require a large sample volume, and only allow for bulk assays, so individual EVs cannot be analyzed.<sup>[134]</sup> IB methods are typically not used for EV diagnostic applications, but are simple ways to confirm the presence of EVs and fulfil EV characterization guidelines outlined by the International Society of Extracellular Vesicles (ISEV).

Mass spectrometry (MS)-based “omic” analyses such as lipidomics and proteomics have been used for high-throughput, multiplexed EV profiling of lipids and proteins, respectively.<sup>[10]</sup> Lipidomics involves the characterization and quantitation of lipids in biological samples as well as lipid compositional responses of cell and organism levels to stimuli.<sup>[140]</sup> The typical workflow involves EV lysis, lipid extraction, and the separation of lipid species, usually by gas chromatography (GC)-MS or liquid chromatography (LC)-MS.<sup>[140]</sup> Lipidomic analysis of EVs is mainly used to elucidate and understand EV biogenesis and protein packaging pathways, but it has also recently been explored in determining lipid-based biomarkers for disease diagnosis, prognosis, and treatment. For example, prostate cancer cell-derived exosomes have been found to be highly enriched in glycosphingolipids, sphingomyelin, cholesterol, and phosphatidylserine compared to their parent cells, and such lipids could potentially be exploited as cancer biomarkers.<sup>[141]</sup> Similarly to lipidomics, the two main goals of EV proteomic analysis include the determination of disease-associated proteins that can be exploited as diagnostic and prognostic biomarkers,<sup>[142]</sup> and the investigation of the roles of EVs and their proteomes in biological processes and phenomena (e.g., disease progression).<sup>[143]</sup> Proteomic analysis workflow is similar to lipidomic analysis in that EVs are lysed, and proteins are extracted and fractionated (typically by LC-MS).<sup>[27]</sup> MS-based proteomic analysis that utilizes advanced instruments allows for the rapid identification of thousands of proteins.<sup>[140]</sup> However, analysis is complicated by contamination from high-abundant proteins found in EV source media, as well as low EV recovery (and thus protein yield) due to losses at

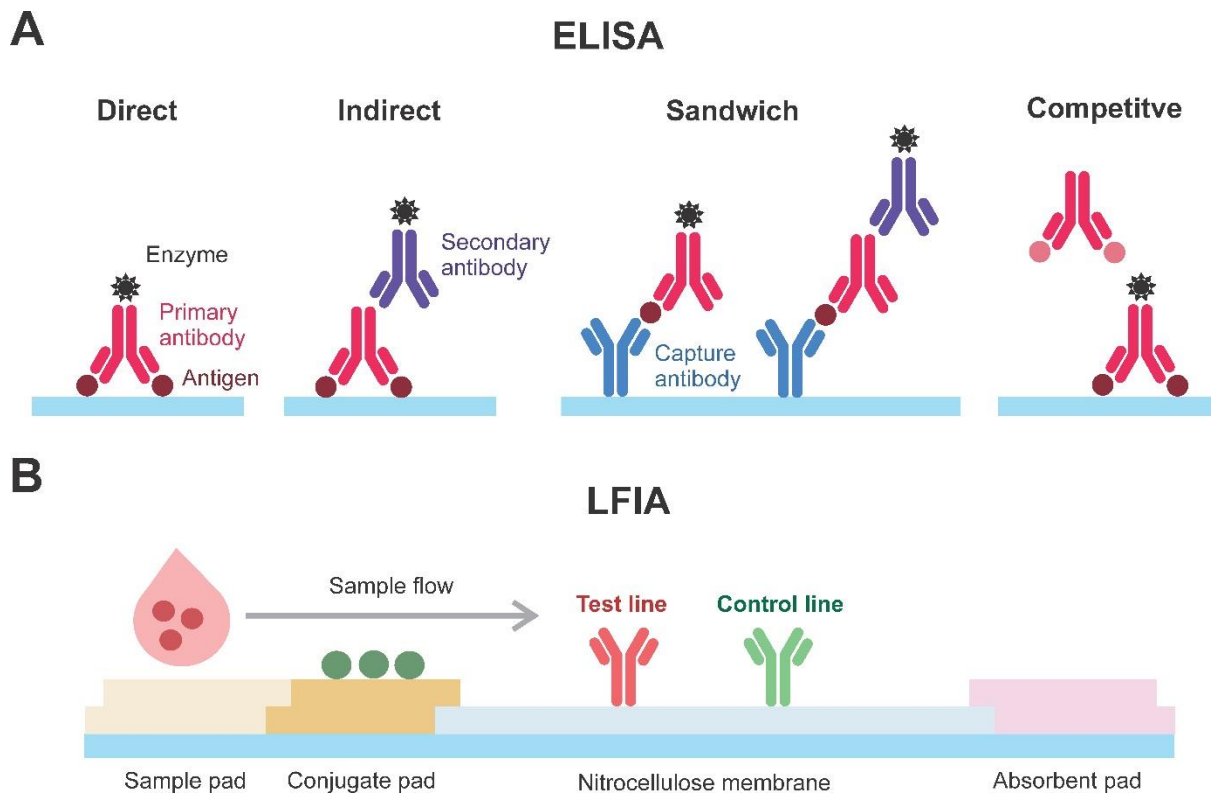
various stages of isolation, sample processing, and fractionation.<sup>[140]</sup> Although perhaps not yet ideal for point-of-care diagnostic applications, the increased depth of EV proteomic profiling is beneficial for the complete characterization of EV proteomes and the understanding of their functions.

Enzyme-linked immunosorbent assay (ELISA) is commonly used and widely available for the detection of an analyte based on the antigen-antibody reaction. In ELISA, an antibody is linked to an enzyme, and a detectable signal like a color change is detected by a spectrophotometer upon the addition of a substrate.<sup>[144]</sup> ELISAs are usually performed on 96-well polystyrene plates, allowing for the analysis of multiple samples per assay, and can be performed by one of four methods: direct, indirect, sandwich, and competitive (**Figure 1-4A**).<sup>[144]</sup> Direct ELISA is the fastest of the four, as the target antigen is immobilized on the well and detected with an antibody conjugated to an enzyme. In contrast, indirect ELISA utilizes a primary antibody for antigen detection and an enzyme-coupled secondary antibody is used to detect the primary antibody. In sandwich ELISA, the capture antibody is coated on the well and the sample is added. Based on the detection antibody, this method can be either direct or indirect. Lastly, in competitive ELISA, the primary antibody is incubated with the sample antigen, and the antigen-antibody complexes are added to a plate containing the same antigen, resulting in a competitive reaction between the sample antigen and the antigen bound to the plate. If more antigen is present in the sample, less primary antibody will be available to bind to the coated antigen. ELISA is a powerful tool for EV analysis since the use of tetraspanins offers high sensitivity for EV detection. For example, EVs derived from ovarian cancer cells can be identified based on their CD24 and EpCAM expression.<sup>[145]</sup>

Lateral-flow immunoassay (LFIA) is another method that can be used for EV detection. The operation principle is similar to ELISA, except the capture antibody or immobilized antigen is bound to a membrane like nitrocellulose instead of a plastic well. Unlike ELISAs, LFIAs can be performed in a single step and in a few minutes, thus avoiding the tedious steps and long incubation times required for ELISA.<sup>[146]</sup> In LFIAs, four components are integrated serially, with small overlap between components allowing for sample flow (**Figure 1-4B**).<sup>[144]</sup> A small sample volume is added to a sample pad, which then migrates



to the conjugate pad onto which the detection reagent has been conjugated. The analyte interacts with the conjugate and both migrate to the membrane onto which the capture reagent has been immobilized. Immunoassay formats in LFIA can vary like in ELISA, and typically sandwich and competitive formats are selected.<sup>[147]</sup> After interaction, excess sample and reagent migrate to the absorbent pad, which is used as storage and waste. In the case of EVs, tetraspanins are commonly used for detection, and the LFIA platform can even be modified into a multiple target assay by utilizing different antibodies on different test lines,<sup>[144]</sup> thereby allowing for the detection of a broad range of EVs based on their surface content.<sup>[148]</sup>



**Figure 1-4: Schematic illustration of two common immunoassay techniques for EV detection and characterization, including (A) ELISA, which highlights four common detection methods, and (B) LFIA, which highlights sample flow along four main components.**

Immunoassay approaches as described above can be attractive methods for EV detection and characterization, but there are several limitations. For example, ELISA involves multiple wash steps that introduce ample opportunity for error, and the process is time-

consuming.<sup>[144]</sup> Microfluidic biosensors are currently being developed to reduce assay time and cost of analysis, possibly rendering traditional ELISA as obsolete in the future.<sup>[149]</sup> Although LFIA is quicker, simpler, and easier to operate than ELISA, its simplicity limits its sensitivity, and thus its performance for point-of-care diagnostics.<sup>[144]</sup> General limitations regarding immunoassays such as high cost, lack of automation, low sensitivity, and low specificity can be overcome by the use of nanoplasmonic techniques.<sup>[150]</sup>

### 1.3.3 Nanoplasmonic Approaches for Biochemical Characterization

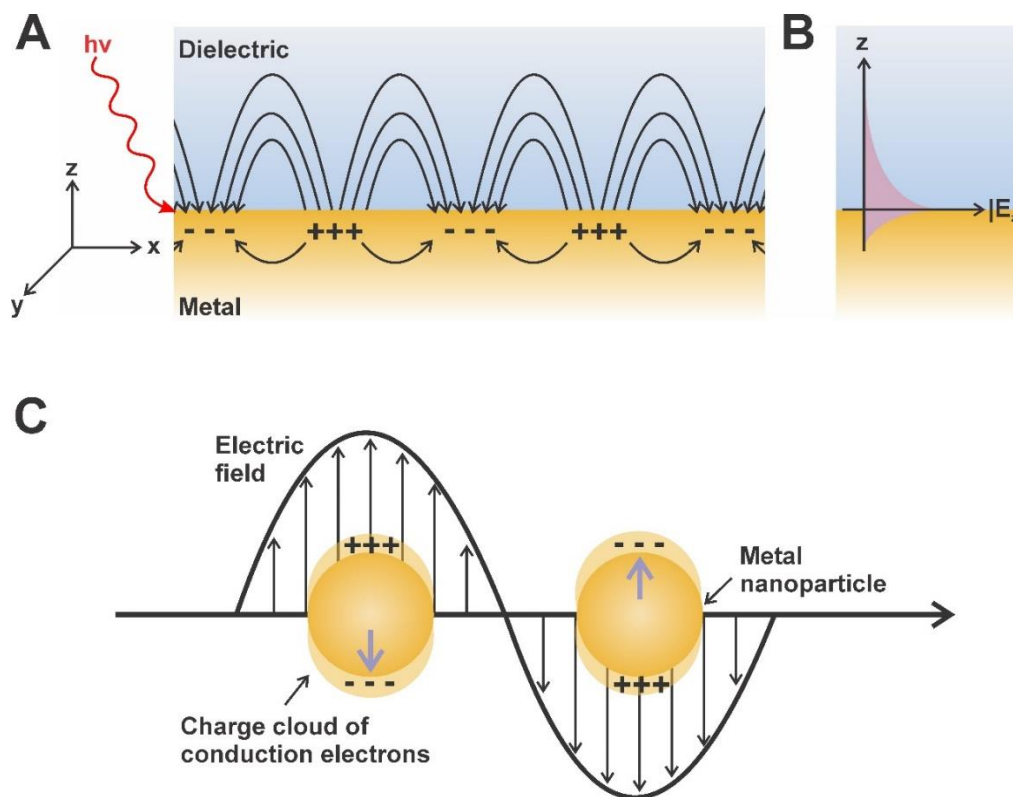
Nanoplasmonics refers to the study and application of light-metal interactions at the nanoscale.<sup>[151,152]</sup> Metals can be viewed as a plasma in which conduction electrons move throughout the bulk material. When incident light interacts with the metal, the photons induce a collective oscillation of free electrons in the conduction band of the metal, producing surface plasmon polaritons (SPPs) that propagate at the metal-dielectric interface (**Figure 1-5A**).<sup>[150]</sup> Nanoplasmonic biosensors are beneficial in that there is potential for label-free quantitative analysis, a high degree of multiplexing, and potential for miniaturization.<sup>[153]</sup> Prominent examples of nanoplasmonic techniques for biosensing include surface plasmon resonance (SPR) sensing, SPR imaging (SPRi), localized surface plasmon resonance (LSPR) sensing, and surface-enhanced Raman spectroscopy (SERS).

SPR occurs when the frequency of incident light matches the oscillation frequency of the surface plasmons. A common SPR experimental setup known as the Kretschmann configuration implements a prism tightly connected to a glass sensor chip coated with a thin film of gold.<sup>[154,155]</sup> For biosensing, the gold film is usually covered with antibodies for the target capture of biological analytes.<sup>[156]</sup> Analytical signals of biomolecules can then be measured by considering the refractive index (RI) of the volume in which the surface plasmons are travelling.<sup>[150]</sup> The resonance condition for a given plasmonic setup is sensitive to changes in RI within a nanometer-sized length from the material. In SPR, this area extends usually 100 – 400 nm away from the metal surface (**Figure 1-5B**).<sup>[156]</sup> Upon the binding of biomolecules to the metal surface (e.g., by antigen-antibody interactions), subtle changes in the RI lead to large observable changes in the resonance condition. Thus, SPR is a very sensitive technique for biosensing. Although miniaturization of SPR devices

for clinical applications is being explored and developed, commercially available setups are still quite bulky as well as costly.<sup>[150]</sup> Furthermore, SPR devices have limited multiplexing capability, often requiring multiple sensor chips and multi-flow channels.<sup>[156]</sup>

As opposed to conventional SPR sensing, SPRi allows for detection and imaging in the entire sensor area, thus providing multiplexed detection. Most SPRi devices utilize the aforementioned Kretschmann configuration but with a CCD camera for the detection of reflected light, allowing for real-time visualization of the chip.<sup>[154,157]</sup> However, since a less compact optical configuration is used with less sensitive detectors, SPRi suffers from reduced sensitivity (about one order of magnitude) compared to SPR sensing.<sup>[156]</sup>

LSPR occurs when the surface plasmon is confined to a nanoparticle surface with dimensions that are far smaller than the wavelength of incident light.<sup>[150]</sup> LSPR is fundamentally similar to SPR as photons of the incident light interact with metallic nanostructures to create a free electron oscillation in the conduction band of the metal (**Figure 1-5C**). In contrast to SPR, the free electron oscillation is non-propagating, and the electric field enhancement observed is confined much closer to the metallic surface (approximately 10 – 30 nm).<sup>[150]</sup> Although LSPR sensing requires nanoscale probes or substrates with nanoscale features, thereby increasing cost and complexity, LSPR offers high surface sensitivity compared to SPR, has negative contribution from the bulk sensitivity and bulk temperature fluctuations, and has high potential for miniaturization.<sup>[156,158]</sup>



**Figure 1-5: (A) Schematic illustration of an SPP at the interface of a dielectric material and metal surface, and (B) the strength of the resulting electric field, which decays exponentially away from the metal surface. (C) Schematic illustration of the LSPR of a metal nanoparticle induced by light.**

The theory of LSPR is also fundamental to SERS, which is the analytical technique of choice for EV characterization in this thesis. Although conventional Raman spectroscopy offers high chemical specificity, is non-destructive, and inert to aqueous background, the recorded signal is very weak (i.e., low sensitivity) and often masked by fluorescent background, particularly in biological samples.<sup>[150]</sup> These drawbacks are overcome by SERS, which exploits LSPR to achieve very high sensitivity (i.e., single molecule detection).<sup>[159]</sup> The theory of Raman spectroscopy and SERS are explained in Chapter 2 of this thesis, and several probes and substrates used in the SERS characterization of EVs are reviewed.

## 1.4 Scope of Thesis

This thesis presents the fabrication of plasmonic gold nanohole arrays for EV capture and subsequent EV spectral characterization and analysis by SERS. The following four chapters are summarized below.

In Chapter 2, the theory of and instrumentation for Raman spectroscopy and SERS are explained in detail. Further reviewed in this chapter are SERS probes and platforms that have been proposed for EV characterization in recent years. Chapter 2 also details the methodology and theory behind several other techniques used throughout this thesis, including electron-beam lithography, principal component analysis (PCA), and machine learning.

In Chapter 3, nanohole arrays fabricated for EV capture and SERS characterization are presented and characterized in detail. Further characterized are EVs isolated from MSCs derived from two sources (bone marrow and pancreatic tissue). The SERS signals from these samples are further analyzed by PCA and machine learning.

In Chapter 4, the nanohole arrays presented in Chapter 3 are used again but are challenged with a larger scale of EV samples, this time derived from ovarian cancer cells. Four ovarian cancer cell lines are explored by SERS here and compared with the SERS signals from a normal ovarian cell line as a control. The SERS signals from these samples are further analyzed by PCA, and machine learning is used to discriminate the cancerous EVs from the normal EVs.

In Chapter 5, the thesis is concluded and summarized. Also discussed in this chapter are suggested ideas for future work.

## 1.5 References

1. Margolis, L.; Sadovsky, Y. *PLoS Biol.* **2019**, *17*, e3000363.
2. Raposo, G.; Stoorvogel, W. *J. Cell Biol.* **2013**, *200*, 373-383.

3. Lawson, C.; Vicencio, J. M.; Yellon, D. M.; Davidson, S. M. *J. Endocrinol.* **2016**, *228*, R57-R71.
4. Ouyang, Y.; Mouillet, J. F.; Coyne, C. B.; Sadovsky, Y. *Placenta* **2014**, *35*, S69-S73.
5. Anderson, H. C. *J. Cell Biol.* **1969**, *41*, 59-72.
6. Grillo, M. A. *J. Cell Biol.* **1970**, *47*, 547-553.
7. Allan, D.; Billah, M. M.; Finean, J. B.; Michell, R. H. *Nature* **1976**, *261*, 58-60.
8. Pan, B.-T.; Johnstone, R. M. *Cell* **1983**, *33*, 967-978.
9. Johnstone, R. M.; Adam, M.; Hammond, J. R.; Orr, L.; Turbide, C. *J. Biol. Chem.* **1987**, *262*, 9412-9420.
10. Veziroglu, E. M.; Mias, G. I. *Front. Genet.* **2020**, *11*, 700.
11. Raposo, G.; Nijman, H. W.; Stoorvogel, W.; Liejendekker, R.; Harding, C. V.; Melief, C. J.; Geuze, H. J. *J. Exp. Med.* **1996**, *183*, 1161-1172.
12. Baj-Krzyworzeka, M.; Majka, M.; Pratico, D.; Ratajczak, J.; Vilaire, G.; Kijowski, J.; Reca, R.; Janowska-Wieczorek, A.; Ratajczak, M. Z. *Exp. Hematol.* **2002**, *30*, 450-459.
13. Ratajczak, J.; Miekus, K.; Kucia, M.; Zhang, J.; Reca, R.; Dvorak, P.; Ratajczak, M. Z. *Leukemia* **2006**, *20*, 847-856.
14. Valadi, H.; Ekström, K.; Bossios, A.; Sjöstrand, M.; Lee, J. J.; Lötvall, J. O. *Exosome-Nat. Cell Biol.* **2007**, *9*, 654-659.
15. Skog, J.; Würdinger, T.; van Rijn, S.; Meijer, D. H.; Gainche, L.; Curry, W. T.; Carter, B. S.; Krichevsky, A. M.; Breakefield, X. O. *Nat. Cell Biol.* **2008**, *10*, 1470-1476.
16. Battistelli, M.; Falcieri, E. *Biology* **2020**, *9*, 21.
17. Stremersch, S.; De Smedt, S. C.; Raemdonck, K. *J. Control. Release* **2016**, *244*, 167-183.

18. Shpacovitch, V.; Hergenröder, R. *Anal. Chim. Acta* **2018**, *1005*, 1-15.
19. Saraste, A.; Pulkki, K. *Cardiovasc. Res.* **2000**, *45*, 528-537.
20. Krämer-Albers, E. M.; Hill, A. F. *Curr. Opin. Neurobiol.* **2016**, *39*, 101-107.
21. Shah, R.; Patel, T.; Freedman, J. E. *N. Engl. J. Med.* **2018**, *379*, 958-966.
22. Chettimada, S.; Lorenz, D. R.; Misra, V.; Dillon, S. T.; Reeves, R. K.; Manickam, C.; Morgello, S.; Kirk, G. D.; Mehta, S. H.; Gabuzda, D. *Sci. Rep.* **2018**, *8*, 7227.
23. Ståhl, A.-I.; Johansson, K.; Mossberg, M.; Kahn, R.; Karpman, D. *Pediatr. Nephrol.* **2019**, *34*, 11-30.
24. Mihály, J.; Deák, R.; Szigyártó, I. C.; Bóta, A.; Beke-Somfai, T.; Varga, Z. *Biochim. Biophys. Acta, Biomembr.* **2017**, *1859*, 459-466.
25. Zhang, Y.; Liu, Y.; Liu, H.; Tang, W. H. *Cell Biosci.* **2019**, *9*, 19.
26. Henne, W. M.; Buchkovich, N. J.; Emr, S. D. *Dev. Cell* **2011**, *21*, 77-91.
27. Mallia, A.; Gianazza, E.; Zoanni, B.; Brioschi, M.; Barbieri, S. S.; Banfi, C. *Diagnostics* **2020**, *10*, 843.
28. Stuffers, S.; Sem Wegner, C.; Stenmark, H.; Brech, A. *Traffic* **2009**, *10*, 925-937.
29. Hessvik, N. P.; Llorente, A. *Cell. Mol. Life Sci.* **2018**, *75*, 193-208.
30. Booth, A. M.; Fang, Y.; Fallon, J. K.; Yang, J.-M.; Hildreth, J. E. K.; Gould, S. J. *J. Cell Biol.* **2006**, *172*, 923-935.
31. Mathieu, M.; Martin-Jaular, L.; Lavieu, G.; Théry, C. *Nat. Cell Biol.* **2019**, *21*, 9-17.
32. Zitvogel, L.; Regnault, A.; Lozier, A.; Wolfers, J.; Flament, C.; Tenza, D.; Ricciardi-Castagnoli, P.; Raposo, G.; Amigorena, S. *Nat. Med.* **1998**, *4*, 594-600.

33. Gross, J. C.; Chaudhary, V.; Bartscherer, K.; Boutros, M. *Nat. Cell Biol.* **2012**, *14*, 1036-1045.
34. Zhang, L.; Wrana, J. L. *Curr. Opin. Genet. Dev.* **2014**, *27*, 14-19.
35. Pegtel, D. M.; Cosmopoulos, K.; Thorley-Lawson, D. A.; van Eijndhoven, M. A. J.; Hopmans, E. S.; Lindenberg, J. L.; de Gruijl, T. D.; Würdinger, T.; Middeldorp, J. M. *Proc. Natl. Acad. Sci. U. S. A.* **2010**, *107*, 6328-6333.
36. Melo, Sonia A.; Sugimoto, H.; O'Connell, Joyce T.; Kato, N.; Villanueva, A.; Vidal, A.; Qiu, L.; Vitkin, E.; Perelman, Lev T.; Melo, Carlos A., et al. *Cancer Cell* **2014**, *26*, 707-721.
37. Qu, L.; Ding, J.; Chen, C.; Wu, Z.-J.; Liu, B.; Gao, Y.; Chen, W.; Liu, F.; Sun, W.; Li, X.-F., et al. *Cancer Cell* **2016**, *29*, 653-668.
38. Cai, J.; Han, Y.; Ren, H.; Chen, C.; He, D.; Zhou, L.; Eisner, G. M.; Asico, L. D.; Jose, P. A.; Zeng, C. *J. Mol. Cell Biol.* **2013**, *5*, 227-238.
39. Sansone, P.; Savini, C.; Kurelac, I.; Chang, Q.; Amato, L. B.; Strillacci, A.; Stepanova, A.; Iommarini, L.; Mastroleo, C.; Daly, L., et al. *Proc. Natl. Acad. Sci. U. S. A.* **2017**, *114*, E9066-E9075.
40. Al-Nedawi, K.; Meehan, B.; Kerbel, R. S.; Allison, A. C.; Rak, J. *Proc. Natl. Acad. Sci. U. S. A.* **2009**, *106*, 3794-3799.
41. Ricklefs, F.; Mineo, M.; Rooj, A. K.; Nakano, I.; Charest, A.; Weissleder, R.; Breakefield, X. O.; Chiocca, E. A.; Godlewski, J.; Bronisz, A. *Cancer Res.* **2016**, *76*, 2876-2881.
42. Montecalvo, A.; Larregina, A. T.; Shufesky, W. J.; Beer Stolz, D.; Sullivan, M. L. G.; Karlsson, J. M.; Baty, C. J.; Gibson, G. A.; Erdos, G.; Wang, Z., et al. *Blood* **2012**, *119*, 756-766.



43. Verweij, F. J.; Revenu, C.; Arras, G.; Dingli, F.; Loew, D.; Pegtel, D. M.; Follain, G.; Allio, G.; Goetz, J. G.; Zimmermann, P., et al. *Dev. Cell* **2019**, *48*, 573-589.
44. Mulcahy, L. A.; Pink, R. C.; Carter, D. R. F. *J. Extracell. Vesicles* **2014**, *3*, 24641.
45. Jing, H.; He, X.; Zheng, J. *Transl. Res.* **2018**, *196*, 1-16.
46. Koenen, R. R.; Aikawa, E. *Front. Cardiovasc. Med.* **2018**, *5*, 133.
47. Coleman, B. M.; Hill, A. F. *Semin. Cell Dev. Biol.* **2015**, *40*, 89-96.
48. Buzas, E. I.; Gyorgy, B.; Nagy, G.; Falus, A.; Gay, S. *Nat. Rev. Rheumatol.* **2014**, *10*, 356-364.
49. Murphy, C.; Withrow, J.; Hunter, M.; Liu, Y.; Tang, Y. L.; Fulzele, S.; Hamrick, M. W. *Mol. Aspects Med.* **2018**, *60*, 123-128.
50. Tang, M. K. S.; Wong, A. S. T. *Cancer Lett.* **2015**, *367*, 26-33.
51. Skotland, T.; Ekroos, K.; Kauhanen, D.; Simolin, H.; Seierstad, T.; Berge, V.; Sandvig, K.; Llorente, A. *Eur. J. Cancer* **2017**, *70*, 122-132.
52. Wang, M.; Ji, S.; Shao, G.; Zhang, J.; Zhao, K.; Wang, Z.; Wu, A. *Clin. Transl. Oncol.* **2018**, *20*, 906-911.
53. Tonk, C. H.; Witzler, M.; Schulze, M.; Tobiasch, E. Mesenchymal Stem Cells. In *Essential Current Concepts in Stem Cell Biology*, Brand-Saberi, B., Ed. Springer: Cham, 2020; pp 21-39.
54. Park, K.-S.; Bandeira, E.; Shelke, G. V.; Lässer, C.; Lötvall, J. *Stem Cell Res. Ther.* **2019**, *10*, 288.
55. Lai, R. C.; Yeo, R. W.; Lim, S. K. *Semin. Cell Dev Biol.* **2015**, *40*, 82-88.
56. Lopes-Pacheco, M.; Robba, C.; Rocco, P. R. M.; Pelosi, P. *Cell Biol. Toxicol.* **2020**, *36*, 83-102.

57. Giebel, B.; Kordelas, L.; Börger, V. *Stem Cell Investig.* **2017**, *4*, 84-84.
58. Abreu, S. C.; Lopes-Pacheco, M.; Weiss, D. J.; Rocco, P. R. M. *Front. Cell Dev. Biol.* **2021**, *9*, 600711.
59. Khatri, M.; Richardson, L. A.; Meulia, T. *Stem Cell Res. Ther.* **2018**, *9*, 17.
60. Lopes-Pacheco, M.; Silva, P. L.; Cruz, F. F.; Battaglini, D.; Robba, C.; Pelosi, P.; Morales, M. M.; Caruso Neves, C.; Rocco, P. R. M. *Front. Physiol.* **2021**, *12*, 593223.
61. Du, Y. M.; Zhuansun, Y. X.; Chen, R.; Lin, L.; Lin, Y.; Li, J. G. *Exp. Cell Res.* **2018**, *363*, 114-120.
62. Aguiar, F. S.; Melo, A. S.; Araújo, A. M. S.; Cardoso, A. P.; de Souza, S. A. L.; Lopes-Pacheco, M.; Cruz, F. F.; Xisto, D. G.; Asensi, K. D.; Faccioli, L., et al. *Stem Cell Res. Ther.* **2020**, *11*, 167.
63. Hu, S.; Park, J.; Liu, A.; Lee, J.; Zhang, X.; Hao, Q.; Lee, J. W. *Stem Cells Transl. Med.* **2018**, *7*, 615-624.
64. Wang, J.; Huang, R.; Xu, Q.; Zheng, G.; Qiu, G.; Ge, M.; Shu, Q.; Xu, J. *Crit. Care Med.* **2020**, *48*, e599-e610.
65. Mansouri, N.; Willis, G. R.; Fernandez-Gonzalez, A.; Reis, M.; Nassiri, S.; Mitsialis, S. A.; Kourembanas, S. *JCI Insight* **2019**, *4*, e128060.
66. Kang, J. H.; Jung, M. Y.; Choudhury, M.; Leof, E. B. *FASEB J.* **2020**, *34*, 2213-2226.
67. Vituret, C.; Gallay, K.; Confort, M. P.; Ftaich, N.; Matei, C. I.; Archer, F.; Ronfort, C.; Mornex, J. F.; Chanson, M.; Di Pietro, A., et al. *Hum. Gene Ther.* **2016**, *27*, 166-183.
68. Zulueta, A.; Colombo, M.; Peli, V.; Falleni, M.; Tosi, D.; Ricciardi, M.; Baisi, A.; Bulfamante, G.; Chiaramonte, R.; Caretti, A. *Cell. Signal.* **2018**, *51*, 110-118.
69. Tsiapalis, D.; O'Driscoll, L. *Cells* **2020**, *9*, 991.

70. Ma, Y.; Ge, S.; Zhang, J.; Zhou, D.; Li, L.; Wang, X.; Su, J. *Int. J. Clin. Exp. Pathol.* **2017**, *10*, 10032-10039.
71. Bucan, V.; Vaslaitis, D.; Peck, C. T.; Strauß, S.; Vogt, P. M.; Radtke, C. *Mol. Neurobiol.* **2019**, *56*, 1812-1824.
72. Ma, Y.; Dong, L.; Zhou, D.; Li, L.; Zhang, W.; Zhen, Y.; Wang, T.; Su, J.; Chen, D.; Mao, C., et al. *J. Cell Mol. Med.* **2019**, *23*, 2822-2835.
73. Xin, H.; Li, Y.; Buller, B.; Katakowski, M.; Zhang, Y.; Wang, X.; Shang, X.; Zhang, Z. G.; Chopp, M. *Stem Cells* **2012**, *30*, 1556-1564.
74. Xin, H.; Li, Y.; Cui, Y.; Yang, J. J.; Zhang, Z. G.; Chopp, M. *J. Cereb. Blood Flow Metab.* **2013**, *33*, 1711-1715.
75. Zhang, Y.; Chopp, M.; Meng, Y.; Katakowski, M.; Xin, H.; Mahmood, A.; Xiong, Y. *J. Neurosurg.* **2015**, *122*, 856-867.
76. Li, D.; Zhang, P.; Yao, X.; Li, H.; Shen, H.; Li, X.; Wu, J.; Lu, X. *Front. Neurosci.* **2018**, *12*, 845.
77. Lu, Y.; Zhou, Y.; Zhang, R.; Wen, L.; Wu, K.; Li, Y.; Yao, Y.; Duan, R.; Jia, Y. *Front. Neurosci.* **2019**, *13*, 209.
78. Li, C.; Jiao, G.; Wu, W.; Wang, H.; Ren, S.; Zhang, L.; Zhou, H.; Liu, H.; Chen, Y. *Cell Transplant.* **2019**, *28*, 1373-1383.
79. Clark, K.; Zhang, S.; Barthe, S.; Kumar, P.; Pivetti, C.; Kreutzberg, N.; Reed, C.; Wang, Y.; Paxton, Z.; Farmer, D., et al. *Cells* **2019**, *8*, 1497.
80. Riazifar, M.; Mohammadi, M. R.; Pone, E. J.; Yeri, A.; Lässer, C.; Segaliny, A. I.; McIntyre, L. L.; Shelke, G. V.; Hutchins, E.; Hamamoto, A., et al. *ACS Nano* **2019**, *13*, 6670-6688.

81. Reza-Zaldivar, E.; Hernández-Sapiéns, M. A.; Gutiérrez-Mercado, Y. K.; Sandoval-Ávila, S.; Gomez-Pinedo, U.; Márquez-Aguirre, A. L.; Vázquez-Méndez, E.; Padilla-Camberos, E.; Canales-Aguirre, A. *Neural Regen. Res.* **2019**, *14*, 1626-1634.
82. Lai, R. C.; Arslan, F.; Lee, M. M.; Sze, N. S.; Choo, A.; Chen, T. S.; Salto-Tellez, M.; Timmers, L.; Lee, C. N.; El Oakley, R. M., et al. *Stem Cell Res.* **2010**, *4*, 214-222.
83. Arslan, F.; Lai, R. C.; Smeets, M. B.; Akeroyd, L.; Choo, A.; Aguur, E. N.; Timmers, L.; van Rijen, H. V.; Doevendans, P. A.; Pasterkamp, G., et al. *Stem Cell Res.* **2013**, *10*, 301-312.
84. Adamiak, M.; Cheng, G.; Bobis-Wozowicz, S.; Zhao, L.; Kedracka-Krok, S.; Samanta, A.; Karnas, E.; Xuan, Y.-T.; Skupien-Rabian, B.; Chen, X., et al. *Circ. Res.* **2018**, *122*, 296-309.
85. Lu, Z.; Chen, Y.; Dunstan, C.; Roohani-Esfahani, S.; Zreiqat, H. *Tissue Eng. Part A* **2017**, *23*, 1212-1220.
86. Diomedede, F.; Gugliandolo, A.; Cardelli, P.; Merciaro, I.; Ettore, V.; Traini, T.; Bedini, R.; Scionti, D.; Bramanti, A.; Nanci, A., et al. *Stem Cell Res. Ther.* **2018**, *9*, 104.
87. Chew, J. R. J.; Chuah, S. J.; Teo, K. Y. W.; Zhang, S.; Lai, R. C.; Fu, J. H.; Lim, L. P.; Lim, S. K.; Toh, W. S. *Acta Biomater.* **2019**, *89*, 252-264.
88. Vonk, L. A.; van Dooremalen, S. F. J.; Liv, N.; Klumperman, J.; Coffey, P. J.; Saris, D. B. F.; Lorenowicz, M. J. *Theranostics* **2018**, *8*, 906-920.
89. Wang, R.; Xu, B.; Xu, H. *Cell Cycle* **2018**, *17*, 2756-2765.
90. Chen, P.; Zheng, L.; Wang, Y.; Tao, M.; Xie, Z.; Xia, C.; Gu, C.; Chen, J.; Qiu, P.; Mei, S., et al. *Theranostics* **2019**, *9*, 2439-2459.
91. Wu, R.; Huang, C.; Wu, Q.; Jia, X.; Liu, M.; Xue, Z.; Qiu, Y.; Niu, X.; Wang, Y. *Stem Cell Res. Ther.* **2019**, *10*, 80.

92. Mitchell, R.; Mellows, B.; Sheard, J.; Antonioli, M.; Kretz, O.; Chambers, D.; Zeuner, M.-T.; Tomkins, J. E.; Denecke, B.; Musante, L., et al. *Stem Cell Res. Ther.* **2019**, *10*, 116.
93. Figliolini, F.; Ranghino, A.; Grange, C.; Cedrino, M.; Tapparo, M.; Cavallari, C.; Rossi, A.; Togliatto, G.; Femminò, S.; Gugliuzza, M. V., et al. *Arterioscler. Thromb. Vasc. Biol.* **2020**, *40*, 239-254.
94. Ranghino, A.; Bruno, S.; Bussolati, B.; Moggio, A.; Dimuccio, V.; Tapparo, M.; Biancone, L.; Gontero, P.; Frea, B.; Camussi, G. *Stem Cell Res. Ther.* **2017**, *8*, 24.
95. Matsukura, T.; Inaba, C.; Weygant, E. A.; Kitamura, D.; Janknecht, R.; Matsumoto, H.; Hyink, D. P.; Kashiwada, S.; Obara, T. *Physiol. Rep.* **2019**, *7*, e14172.
96. Zhou, Y.; Liu, S.; Zhao, M.; Wang, C.; Li, L.; Yuan, Y.; Li, L.; Liao, G.; Bresette, W.; Zhang, J., et al. *J. Control. Release* **2019**, *316*, 93-104.
97. Du, Y.; Li, D.; Han, C.; Wu, H.; Xu, L.; Zhang, M.; Zhang, J.; Chen, X. *Cell. Physiol. Biochem.* **2017**, *43*, 611-625.
98. Yao, J.; Zheng, J.; Cai, J.; Zeng, K.; Zhou, C.; Zhang, J.; Li, S.; Li, H.; Chen, L.; He, L., et al. *FASEB J.* **2019**, *33*, 1695-1710.
99. Mardpour, S.; Ghanian, M. H.; Sadeghi-abandansari, H.; Mardpour, S.; Nazari, A.; Shekari, F.; Baharvand, H. *ACS Appl. Mater. Interfaces* **2019**, *11*, 37421-37433.
100. Choi, J. S.; Cho, W. L.; Choi, Y. J.; Kim, J. D.; Park, H. A.; Kim, S. Y.; Park, J. H.; Jo, D. G.; Cho, Y. W. *J. Extracell. Vesicles* **2019**, *8*, 1565885.
101. da Fonseca Ferreira, A.; da Silva Cunha, P.; Carregal, V. M.; da Silva, I. P.; de Miranda, M. C.; Kunrath-Lima, M.; de Melo, M. I. A.; Faraco, C. C. F.; Barbosa, J. L.; Frezard, F., et al. *Stem Cells Int.* **2017**, *2017*, 9841035.
102. Cooper, D. R.; Wang, C.; Patel, R.; Trujillo, A.; Patel, N. A.; Prather, J.; Gould, L. J.; Wu, M. H. *Adv. Wound Care* **2018**, *7*, 299-308.

103. Pelizzo, G.; Avanzini, M. A.; Cornaglia, A. I.; Silvestri, A. D.; Mantelli, M.; Travaglino, P.; Croce, S.; Romano, P.; Avolio, L.; Iacob, G., et al. *Regen. Med.* **2018**, *13*, 385-394.
104. Zhang, B.; Shi, Y.; Gong, A.; Pan, Z.; Shi, H.; Yang, H.; Fu, H.; Yan, Y.; Zhang, X.; Wang, M., et al. *Stem Cells* **2016**, *34*, 2485-2500.
105. Sung, H.; Ferlay, J.; Siegel, R. L.; Laversanne, M.; Soerjomataram, I.; Jemal, A.; Bray, F. *CA Cancer J. Clin.* **2021**, *71*, 209-249.
106. Bray, F.; Laversanne, M.; Weiderpass, E.; Soerjomataram, I. *Cancer* **2021**, *127*, 3029-3030.
107. Guerrini, L.; Garcia-Rico, E.; O'Loghlen, A.; Giannini, V.; Alvarez-Puebla, R. A. *Cancers* **2021**, *13*, 2179.
108. Urabe, F.; Kosaka, N.; Yoshioka, Y.; Egawa, S.; Ochiya, T. *Clin. Transl. Med.* **2017**, *6*, 45.
109. Vasconcelos, M. H.; Caires, H. R.; Ābols, A.; Xavier, C. P. R.; Linē, A. *Drug Resist. Updat.* **2019**, *47*, 100647.
110. Sugimachi, K.; Matsumura, T.; Hirata, H.; Uchi, R.; Ueda, M.; Ueo, H.; Shinden, Y.; Iguchi, T.; Eguchi, H.; Shirabe, K., et al. *Br. J. Cancer* **2015**, *112*, 532-538.
111. Liu, Q.; Yu, Z.; Yuan, S.; Xie, W.; Li, C.; Hu, Z.; Xiang, Y.; Wu, N.; Wu, L.; Bai, L., et al. *Oncotarget* **2017**, *8*, 13048-13058.
112. Jin, X.; Chen, Y.; Chen, H.; Fei, S.; Chen, D.; Cai, X.; Liu, L.; Lin, B.; Su, H.; Zhao, L., et al. *Clin. Cancer Res.* **2017**, *23*, 5311-5319.
113. Nedaeinia, R.; Manian, M.; Jazayeri, M. H.; Ranjbar, M.; Salehi, R.; Sharifi, M.; Mohaghegh, F.; Goli, M.; Jahednia, S. H.; Avan, A., et al. *Cancer Gene Ther.* **2017**, *24*, 48-56.

114. Ogata-Kawata, H.; Izumiya, M.; Kurioka, D.; Honma, Y.; Yamada, Y.; Furuta, K.; Gunji, T.; Ohta, H.; Okamoto, H.; Sonoda, H., et al. *PLoS One* **2014**, *9*, e92921.
115. Madhavan, B.; Yue, S.; Galli, U.; Rana, S.; Gross, W.; Müller, M.; Giese, N. A.; Kalthoff, H.; Becker, T.; Büchler, M. W., et al. *Int. J. Cancer* **2015**, *136*, 2616-2627.
116. Chen, I. H.; Xue, L.; Hsu, C. C.; Paez, J. S.; Pan, L.; Andaluz, H.; Wendt, M. K.; Iliuk, A. B.; Zhu, J. K.; Tao, W. A. *Proc. Natl. Acad. Sci. U. S. A.* **2017**, *114*, 3175-3180.
117. Zhou, J.; Gong, G.; Tan, H.; Dai, F.; Zhu, X.; Chen, Y.; Wang, J.; Liu, Y.; Chen, P.; Wu, X., et al. *Oncol. Rep.* **2015**, *33*, 2915-2923.
118. Alegre, E.; Sanmamed, M. F.; Rodriguez, C.; Carranza, O.; Martín-Algarra, S.; González, A. *Arch. Pathol. Lab. Med.* **2014**, *138*, 828-832.
119. Ragusa, M.; Barbagallo, C.; Statello, L.; Caltabiano, R.; Russo, A.; Puzzo, L.; Avitabile, T.; Longo, A.; Toro, M. D.; Barbagallo, D., et al. *Cancer Biol. Ther.* **2015**, *16*, 1387-1396.
120. Fleming, N. H.; Zhong, J.; da Silva, I. P.; Vega-Saenz de Miera, E.; Brady, B.; Han, S. W.; Hanniford, D.; Wang, J.; Shapiro, R. L.; Hernando, E., et al. *Cancer* **2015**, *121*, 51-59.
121. Xu, R.; Rai, A.; Chen, M.; Suwakulsiri, W.; Greening, D. W.; Simpson, R. J. *Nat. Rev. Clin. Oncol.* **2018**, *15*, 617-638.
122. Xie, F.; Zhou, X.; Fang, M.; Li, H.; Su, P.; Tu, Y.; Zhang, L.; Zhou, F. *Adv. Sci.* **2019**, *6*, 1901779.
123. Kogure, A.; Yoshioka, Y.; Ochiya, T. *Int. J. Mol. Sci.* **2020**, *21*, 4463.
124. Hsu, Y. L.; Hung, J. Y.; Chang, W. A.; Lin, Y. S.; Pan, Y. C.; Tsai, P. H.; Wu, C. Y.; Kuo, P. L. *Oncogene* **2017**, *36*, 4929-4942.

125. Peinado, H.; Alečković, M.; Lavotshkin, S.; Matei, I.; Costa-Silva, B.; Moreno-Bueno, G.; Hergueta-Redondo, M.; Williams, C.; García-Santos, G.; Ghajar, C., et al. *Nat. Med.* **2012**, *18*, 883-891.
126. Wieckowski, E. U.; Visus, C.; Szajnik, M.; Szczepanski, M. J.; Storkus, W. J.; Whiteside, T. L. *J. Immunol.* **2009**, *183*, 3720-3730.
127. Liu, C.; Yu, S.; Zinn, K.; Wang, J.; Zhang, L.; Jia, Y.; Kappes, J. C.; Barnes, S.; Kimberly, R. P.; Grizzle, W. E., et al. *J. Immunol.* **2006**, *176*, 1375-1385.
128. Al-Nedawi, K.; Meehan, B.; Micallef, J.; Lhotak, V.; May, L.; Guha, A.; Rak, J. *Nat. Cell Biol.* **2008**, *10*, 619-624.
129. Wei, Y.; Lai, X.; Yu, S.; Chen, S.; Ma, Y.; Zhang, Y.; Li, H.; Zhu, X.; Yao, L.; Zhang, J. *Breast Cancer Res. Treat.* **2014**, *147*, 423-431.
130. Arraud, N.; Linares, R.; Tan, S.; Gounou, C.; Pasquet, J. M.; Mornet, S.; Brisson, A. R. *J. Thromb. Haemost.* **2014**, *12*, 614-627.
131. Carnino, J. M.; Lee, H.; Jin, Y. *Respir. Res.* **2019**, *20*, 240.
132. Szatanek, R.; Baj-Krzyworzeka, M.; Zimoch, J.; Lekka, M.; Siedlar, M.; Baran, J. *Int. J. Mol. Sci.* **2017**, *18*, 1153.
133. Dragovic, R. A.; Gardiner, C.; Brooks, A. S.; Tannetta, D. S.; Ferguson, D. J. P.; Hole, P.; Carr, B.; Redman, C. W. G.; Harris, A. L.; Dobson, P. J., et al. *Nanomedicine* **2011**, *7*, 780-788.
134. Hartjes, T. A.; Mytnyk, S.; Jenster, G. W.; van Steijn, V.; van Royen, M. E. *Bioengineering* **2019**, *6*, 7.
135. Winey, M.; Meehl, J. B.; O'Toole, E. T.; Giddings, T. H., Jr. *Mol. Biol. Cell* **2014**, *25*, 319-323.
136. Golding, C. G.; Lamboo, L. L.; Beniac, D. R.; Booth, T. F. *Sci. Rep.* **2016**, *6*, 26516.



137. Dufrêne, Y. F.; Ando, T.; Garcia, R.; Alsteens, D.; Martinez-Martin, D.; Engel, A.; Gerber, C.; Müller, D. J. *Nature Nanotech.* **2017**, *12*, 295-307.
138. Chatterjee, S.; Gadad, S. S.; Kundu, T. K. *Resonance* **2010**, *15*, 622-642.
139. Sharma, S.; LeClaire, M.; Gimzewski, J. K. *Nanotechnology* **2018**, *29*, 132001.
140. Kreimer, S.; Belov, A. M.; Ghiran, I.; Murthy, S. K.; Frank, D. A.; Ivanov, A. R. *J. Proteome Res.* **2015**, *14*, 2367-2384.
141. Llorente, A.; Skotland, T.; Sylvänne, T.; Kauhanen, D.; Róg, T.; Orłowski, A.; Vattulainen, I.; Ekroos, K.; Sandvig, K. *Biochim. Biophys. Acta* **2013**, *1831*, 1302-1309.
142. Boukouris, S.; Mathivanan, S. *Proteomics Clin. Appl.* **2015**, *9*, 358-367.
143. Simpson, R. J.; Lim, J. W. E.; Moritz, R. L.; Mathivanan, S. Exosomes: Proteomic Insights and Diagnostic Potential. *Expert Rev. Proteom.* **2009**, *6*, 267-283.
144. Serrano-Pertierra, E.; Oliveira-Rodríguez, M.; Matos, M.; Gutiérrez, G.; Moyano, A.; Salvador, M.; Rivas, M.; Blanco-López, M. C. *Biomolecules* **2020**, *10*, 824.
145. Im, H.; Shao, H.; Park, Y. I.; Peterson, V. M.; Castro, C. M.; Weissleder, R.; Lee, H. *Nat. Biotechnol.* **2014**, *32*, 490-495.
146. Eltzov, E.; Guttel, S.; Low Yuen Kei, A.; Sinawang, P. D.; Ionescu, R. E.; Marks, R. S. *Electroanalysis* **2015**, *27*, 2116-2130.
147. O'Farrell, B. *Top. Companion Anim. Med.* **2015**, *30*, 139-147.
148. Oliveira-Rodríguez, M.; Serrano-Pertierra, E.; García, A. C.; López-Martín, S.; Yañez-Mo, M.; Cernuda-Morollón, E.; Blanco-López, M. C. *Biosens. Bioelectron.* **2017**, *87*, 38-45.
149. Zhang, P.; He, M.; Zeng, Y. *Lab Chip* **2016**, *16*, 3033-3042.
150. Rojalín, T.; Phong, B.; Koster, H. J.; Carney, R. P. *Front. Chem.* **2019**, *7*, 279.

151. Jackman, J. A.; Rahim Ferhan, A.; Cho, N.-J. *Chem. Soc. Rev.* **2017**, *46*, 3615-3660.
152. Marquez, S.; Morales-Narváez, E. *Front. Bioeng. Biotechnol.* **2019**, *7*, 69.
153. Lopez, G. A.; Estevez, M.-C.; Soler, M.; Lechuga, L. M. *Nanophotonics* **2016**, *6*, 123-136.
154. Puiu, M.; Bala, C. *Sensors* **2016**, *16*, 870.
155. Michel, D.; Xiao, F.; Alameh, K. *Sens. Actuators B Chem.* **2017**, *246*, 258-261.
156. Mauriz, E.; Dey, P.; Lechuga, L. M. *Analyst* **2019**, *144*, 7105-7129.
157. Zhu, L.; Wang, K.; Cui, J.; Liu, H.; Bu, X.; Ma, H.; Wang, W.; Gong, H.; Lausted, C.; Hood, L., et al. *Anal. Chem.* **2014**, *86*, 8857-8864.
158. Špačková, B.; Šípová-Jungová, H.; Käll, M.; Fritzsche, J.; Langhammer, C. *ACS Sens.* **2021**, *6*, 73-82.
159. Otto, A. *J. Raman. Spectrosc.* **2002**, *33*, 593-598.

## Chapter 2

### 2 Principles of EV Detection and Characterization

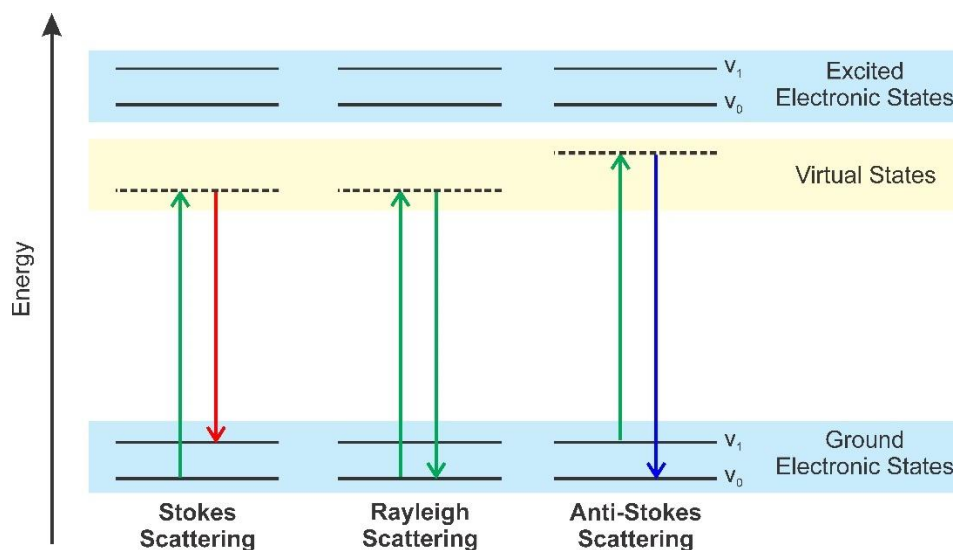
This chapter focuses on the fundamentals of Raman spectroscopy and SERS, the latter of which was the main technique used throughout this thesis for EV characterization. To simplify the analysis and characterization of EVs, PCA is used in this thesis, as well as machine learning for the classification of various EV sources. Therefore, the fundamentals of PCA are also discussed, and an overview of machine learning and its various algorithms is also provided. The performance and fabrication of several SERS probes and substrates used for EV detection in recent years, including both direct and indirect methods, are also reviewed.

#### 2.1 Principles of Raman Spectroscopy

Raman spectroscopy is a technique used to study the vibrational modes of a system based on Raman scattering, a phenomenon that was observed by Raman and Krishnan in 1928.<sup>[1,2]</sup> Photons can be elastically or inelastically scattered from an atom or molecule. When a photon is elastically scattered from an atom or molecule, the scattered photon has an energy that is equal in energy to the incident radiation.<sup>[3]</sup> This process is commonly referred to as Rayleigh scattering. When a photon is inelastically scattered from an atom or molecule, the scattered photon has an energy that is not equal to that of the incident radiation. The inelastic scattering process is referred to as Raman scattering. Raman scattering occurs very infrequently compared to Rayleigh scattering (approximately 1 in  $10^7$  photons).<sup>[4]</sup>

Depending on the initial vibrational state of the atom or molecule, Raman scattered photons can be of lower or higher energy than the incident radiation (**Figure 2-1**).<sup>[5]</sup> A molecule in its ground vibrational state ( $\nu = 0$ ) can absorb a photon of energy  $h\nu_{ex}$  and reemit a photon of energy  $h(\nu_{ex} - \nu_V)$ . When Raman scattered photons are of lower energy than the incident radiation, they are said to be Stokes scattered. A molecule in an excited vibrational state ( $\nu = 1$ ) can absorb a photon of energy  $h\nu_{ex}$  and reemit a photon of energy  $h(\nu_{ex} + \nu_V)$ . When a scattered photon is of higher energy than the incident radiation, it is said to be anti-Stokes scattered. At room temperature, more molecules exist in their ground

vibrational state than an excited vibrational state, and so Stokes Raman scattering occurs more frequently than anti-Stokes Raman scattering.<sup>[6]</sup> In both cases of Raman scattering ( $\pm h\nu_V$ ), the observed wavenumber or Raman shift is a direct measure of the molecule's vibrational energy and specific to chemical bonds in molecules (i.e., “fingerprints”).



**Figure 2-1: Schematic illustration of the Rayleigh and Raman (Stokes and anti-Stokes) scattering processes.**

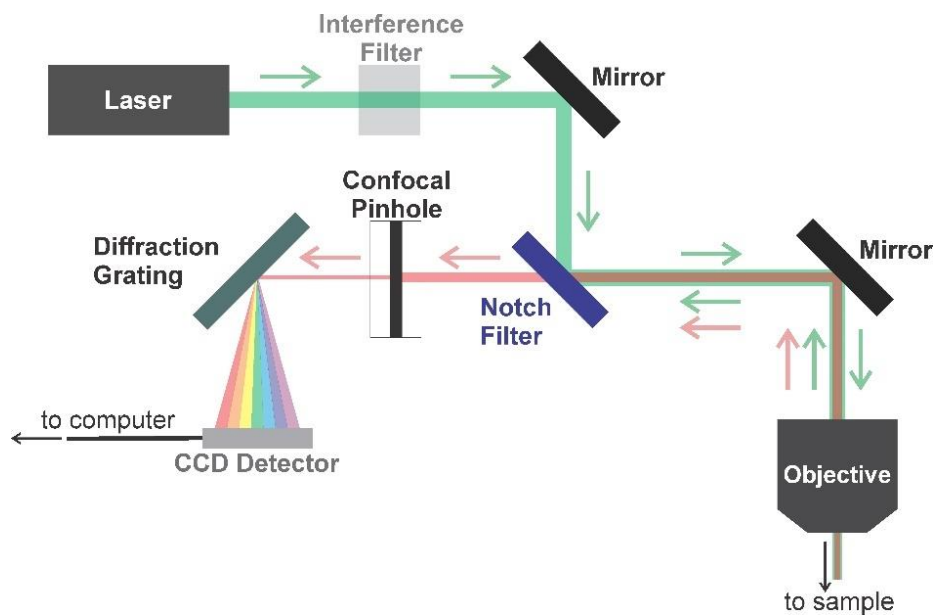
Raman line intensity ( $I_R$ ) can be described by:<sup>[4]</sup>

$$I_R \propto \nu^4 \cdot \sigma \cdot I \cdot e^{-\frac{E_i}{kT}} \cdot c \quad (2-1)$$

In **Equation 2-1**,  $\nu$  and  $I$  are the frequency and intensity of the incident radiation, respectively,  $\sigma$  is the Raman scattering cross-section,  $e^{-\frac{E_i}{kT}}$  is the Boltzmann factor for state  $i$ , and  $c$  is the concentration of the molecule that scattered the radiation. The typical value for Raman scattering cross-section of molecules is between  $10^{-31}$  and  $10^{-26}$   $\text{cm}^{-2}$  molecule<sup>-1</sup>.<sup>[4]</sup> A direct consequence of this low cross-section is one of the main limitations of Raman spectroscopy: low signal intensity.

### 2.1.1 Instrumentation for Raman Spectroscopy

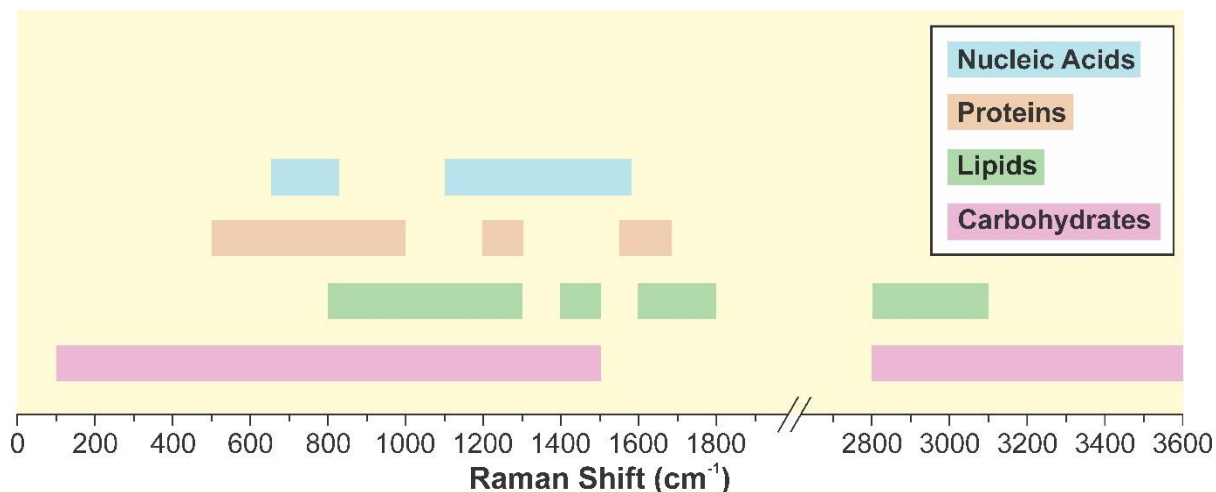
Modern instruments for Raman spectroscopy generally consist of an optical microscope, an excitation laser, light filters, and a sensitive detector (**Figure 2-2**). Light from a laser beam, usually in the green (e.g., 532 nm), red (e.g., 633 nm), or near-infrared (e.g., 785 nm) regions, is passed through an interference filter to maintain a near-zero absorption coefficient for the desired wavelength. Samples are then illuminated with the laser beam focused through an optical microscope objective. High magnification and high numerical aperture (N.A.) objectives provide higher spatial resolution than lower magnification and lower N.A. objectives, which are beneficial when probing biological samples since specific biological architecture can be examined.<sup>[7]</sup> Wavelengths that are close to the laser wavelength, due to Rayleigh scattering, must be optically filtered out since Rayleigh scattering is much more intense compared to Raman scattering.<sup>[8]</sup> Commonly used filters include edge filters, which transmit light above or below the laser wavelength, and notch filters, which allow both Stokes and anti-Stokes measurements by filtering out only the excitation laser wavelength. The remaining unfiltered light is then passed through a confocal pinhole, which allows only Raman scattering through from the focal plane of the microscope, and is dispersed by a diffraction grating onto a detector and recorded. Since Raman scattering intensity is very weak, detectors used in Raman systems must be very sensitive. CCDs are most commonly used for their high quantum efficiencies and low signal-to-noise ratios, especially in comparison to other detectors such as photomultiplier tubes and photodiode arrays.<sup>[9]</sup> CCDs are made up of arrays of pixels that can each collect charge from scattered photons, which is directly proportional to the Raman scattering intensity. The resulting spectrum obtained is a plot of Raman scattering intensity versus Raman shift in wavenumbers ( $\text{cm}^{-1}$ ).



**Figure 2-2: Schematic illustration of a general Raman spectrophotometer setup, where the green beams and arrows represent the incident photons and the red beams and arrows represent the Raman-scattered photons.**

### 2.1.2 Raman Spectroscopy and Biological Samples

Although Raman spectroscopy is hailed for its molecular specificity, Raman analysis of biological samples can be quite complicated since all biological systems are composed of biological macromolecules such as lipids, proteins, nucleic acids, and carbohydrates.<sup>[10]</sup> Vibrations of all modes from these macromolecules will be present in the spectra, and often bands from different origins overlap, confounding the spectra and challenging spectral analysis (**Figure 2-3**). However, there is a wealth of information available to researchers on the classification of Raman spectra of biological systems, and with informed judgement on the researcher's behalf, peak assignments can be made.



**Figure 2-3: Raman spectral regions associated with proteins, nucleic acids, lipids, and carbohydrates.**

Proteins serve many biological functions in cells, including transmitted information and controlling the passage of molecules around the cell membrane. As discussed in Chapter 1, proteins play important roles in EV formation and biogenesis. Raman spectroscopy is an excellent technique to probe the protein structure of biological samples, whose signals are found in the fingerprint region. Frequencies associated with protein backbones can be found at  $1230 - 1300 \text{ cm}^{-1}$  (amide III),<sup>[11,12]</sup>  $1550 \text{ cm}^{-1}$  (amide II),<sup>[13]</sup> and  $1630 - 1680 \text{ cm}^{-1}$  (amide I).<sup>[14-18]</sup> These ranges encompass  $\alpha$ -helices ( $1270 - 1300 \text{ cm}^{-1}$  and  $1650 - 1655 \text{ cm}^{-1}$ ),<sup>[11,18]</sup>  $\beta$ -sheets ( $1230 - 1240 \text{ cm}^{-1}$ ,<sup>[12]</sup>  $1630 - 1635 \text{ cm}^{-1}$ ,<sup>[14,17]</sup> and  $1660 - 1670 \text{ cm}^{-1}$ ),<sup>[15]</sup>  $\beta$ -turns ( $1670 - 1680 \text{ cm}^{-1}$ ),<sup>[16]</sup> and random coils ( $1240 - 1250 \text{ cm}^{-1}$  and  $1670 - 1680 \text{ cm}^{-1}$ ).<sup>[12]</sup> Peaks corresponding to disulfides can be located around  $505 - 550 \text{ cm}^{-1}$ , which includes the *gauche-gauche-gauche* conformation at  $505 - 515 \text{ cm}^{-1}$ , the *gauche-gauche-trans* and *gauche-trans-gauche* conformations at  $520 - 530 \text{ cm}^{-1}$ , and the *trans-gauche-trans* conformation at  $540 - 545 \text{ cm}^{-1}$ .<sup>[19,20]</sup> Lastly, common amino acid side chain peaks can be found at  $643 \text{ cm}^{-1}$ ,  $830 \text{ cm}^{-1}$ , and  $1615 \text{ cm}^{-1}$  (tyrosine),<sup>[21]</sup>  $750 \text{ cm}^{-1}$ ,  $1010 \text{ cm}^{-1}$ , and  $1550 \text{ cm}^{-1}$  (tryptophan),<sup>[22]</sup> and  $1003 \text{ cm}^{-1}$ ,  $1205 \text{ cm}^{-1}$ , and  $1609 \text{ cm}^{-1}$  (phenylalanine).<sup>[23]</sup>

Also found in the fingerprint region are nucleic acids. Many thorough studies have been conducted to elucidate peaks for various nucleic acid moieties including sugars, phosphates, and bases, which provide insight into backbone conformation and base pairing. Regarding the bases, frequencies associated with ring breathing modes can be found at 668

$\text{cm}^{-1}$  (guanine),  $668 \text{ cm}^{-1}$ ,  $746 \text{ cm}^{-1}$ , and  $785 \text{ cm}^{-1}$  (thymine),  $726 \text{ cm}^{-1}$  (adenine),  $781 \text{ cm}^{-1}$  and  $785 \text{ cm}^{-1}$  (cytosine), and  $785 \text{ cm}^{-1}$  (uracil).<sup>[24]</sup> Frequencies of unspecified ring modes can be found at  $1257 \text{ cm}^{-1}$  (cytosine and thymine)<sup>[24]</sup> and  $1336 \text{ cm}^{-1}$  (adenine and guanine),<sup>[25]</sup> while frequencies of ring stretching modes can be found at  $1485 \text{ cm}^{-1}$  (guanine and adenine) and  $1576 \text{ cm}^{-1}$  (adenine and guanine).<sup>[25]</sup> Peaks corresponding to the backbone of A-DNA are located at  $807 \text{ cm}^{-1}$  (O-P-O stretching),  $1099 \text{ cm}^{-1}$  ( $\text{PO}_2^-$  symmetric stretching), and  $1415 \text{ cm}^{-1}$  ( $\text{CH}_2$  deformation), while peaks corresponding to the backbone of B-DNA are located at  $784 \text{ cm}^{-1}$  (O-P-O symmetric stretching),  $830 \text{ cm}^{-1}$  (O-P-O asymmetric stretching),  $1090 \text{ cm}^{-1}$  ( $\text{PO}_2^-$  symmetric stretching), and  $1420 \text{ cm}^{-1}$  ( $\text{CH}_2$  deformation).<sup>[26]</sup> Lastly, peaks belonging to the backbone of Z-DNA are found at  $748 \text{ cm}^{-1}$  (O-P-O symmetric stretching),  $748 \text{ cm}^{-1}$  and  $792 \text{ cm}^{-1}$  (O-P-O asymmetric stretching),  $810 \text{ cm}^{-1}$  (O-P-O asymmetric stretching),  $1095 \text{ cm}^{-1}$  ( $\text{PO}_2^-$  symmetric stretching), and  $1425 \text{ cm}^{-1}$  ( $\text{CH}_2$  deformation).<sup>[26]</sup>

Spectral peaks of lipids and carbohydrates are generally found in the high frequency region, although they can also be present in the fingerprint region. Raman spectroscopy can be used to estimate certain structural properties of lipids such as the degree of unsaturation. Spectral regions associated with lipids are as follows:  $2800 - 3100 \text{ cm}^{-1}$  (CH/CH<sub>2</sub>/CH<sub>3</sub> stretching),  $1600 - 1800 \text{ cm}^{-1}$  (C=C and C=O stretching),  $1400 - 1500 \text{ cm}^{-1}$  (CH<sub>2</sub>/CH<sub>3</sub> bending and scissoring),  $1200 - 1300 \text{ cm}^{-1}$  (CH deformation and CH<sub>2</sub> twisting),  $1050 - 1200 \text{ cm}^{-1}$  (C-C and P-O stretching),  $800 - 1050 \text{ cm}^{-1}$  (CH bending, a skeletal C-O-O mode, and  $\text{N}^+(\text{CH}_3)_3$  asymmetric stretching), and  $500 - 700 \text{ cm}^{-1}$  (CH<sub>2</sub> bending,  $\text{N}^+(\text{CH}_3)_3$  symmetric stretching, C=O-O deformation, and cholesterol ring deformation).<sup>[27]</sup> Specific lipid bands that have been identified include  $1430 \text{ cm}^{-1}$  (CH<sub>2</sub> scissoring),  $1443 \text{ cm}^{-1}$  (CH<sub>2</sub>/CH<sub>3</sub> deformation of lipids and triglycerides), and  $1453 \text{ cm}^{-1}$  (C-H bending),<sup>[28]</sup> as well as  $1749 \text{ cm}^{-1}$  (C=O).<sup>[27]</sup> Bands at  $2846 \text{ cm}^{-1}$  and  $3015 \text{ cm}^{-1}$  have been used to quantify total fatty acid content and total unsaturated fatty acid content, respectively.<sup>[29]</sup> Structural investigations on different types of carbohydrates are very difficult since the basic monomeric unit in polymeric carbohydrate chains is very similar. However, broad spectral regions have been identified as follows:  $3100 - 3600 \text{ cm}^{-1}$  (O-H stretching),  $2800 - 3100 \text{ cm}^{-1}$  (CH/CH<sub>2</sub> stretching),  $1200 - 1500 \text{ cm}^{-1}$  (CH/CH<sub>2</sub> deformation),  $800 - 1200 \text{ cm}^{-1}$  (C-O and C-C stretching), and  $100 - 800 \text{ cm}^{-1}$  (CCO deformation).<sup>[30]</sup>



As mentioned in Section 2.1, the biggest disadvantage of Raman spectroscopy is the low signal intensity caused by small Raman scattering cross-sections associated with molecules (approximately  $10^{-31} - 10^{-26} \text{ cm}^{-2} \text{ molecule}^{-1}$ ).<sup>[4]</sup> Raman spectroscopy of biological molecules can be further complicated by the fact that many biological compounds are fluorescent in nature. The fluorescence scattering cross-section is typically much larger compared to the Raman scattering cross-section (approximately  $10^{-16} \text{ cm}^{-2} \text{ molecule}^{-1}$ ) and a resulting fluorescence background may suppress any underlying Raman spectrum to the point where it is no longer detectable.<sup>[4,7]</sup> Fortunately, many techniques have emerged over the past few decades that are capable of enhancing this weak Raman signal, one of which is SERS.

## 2.2 Principles of SERS

As the name implies, SERS provides the same chemical information as conventional Raman spectroscopy but with a dramatically enhanced signal (i.e., up to  $10^{10}$  enhancement).<sup>[31]</sup> Additionally, minimal sample preparation is generally required, and measurements can easily be multiplexed.<sup>[32]</sup> Since SERS is a surface-sensitive and non-destructive technique capable of single molecule detection,<sup>[33]</sup> it is a promising technique for the characterization of EVs.<sup>[34,35]</sup>

SERS was first observed in 1974 by Fleischmann et al. when an unexpectedly large Raman signal arose from pyridine adsorbed onto a roughened silver electrode.<sup>[36]</sup> Not long after this observation, Albrecht and Creighton hypothesized that the phenomenon was due to the formation of a molecule-metal complex.<sup>[37]</sup> Concurrently, Jeanmaire and van Duyne hypothesized that the observed phenomenon arose from strong electrochemical fields at the surface of the metal.<sup>[38]</sup> Not long after, Moskovits proposed the idea that the collective oscillation of conduction electrons at the nanoscale roughness features on a metal surface was responsible for the intense signal.<sup>[39,40]</sup> Studies since have confirmed that the SERS enhancement arises from two distinct mechanisms based on initial hypothesis by Albrecht and Creighton as well as Jeanmaire and van Duyne: the chemical enhancement mechanism and the electromagnetic (EM) enhancement mechanism.<sup>[41,42]</sup>

The enhancement factor  $EF$  of a substrate can be determined experimentally by:<sup>[43]</sup>

$$EF = \frac{I_{SERS}}{I_{Raman}} \cdot \frac{n_{Raman}}{n_{SERS}} \quad (2-2)$$

In **Equation 2-2**,  $I_{SERS}$  and  $I_{Raman}$  represent the observed intensities with and without a SERS substrate, respectively, of a specific vibrational mode, while  $n_{SERS}$  and  $n_{Raman}$  represent the number of molecules contributing to SERS and normal Raman scattering, respectively. This equation is generally used to estimate the enhancement factor of a substrate when both enhancement mechanisms are present. The main difficulty of accurately estimating  $EF$  is properly defining the number of molecules in a given focal volume that yield a Raman spectrum and a SERS spectrum.

### 2.2.1 Chemical Enhancement

The chemical enhancement mechanism arises from the physio-chemical interactions between a substrate and analyte. For chemical enhancement to be observed, the analyte must either be adsorbed during onto the substrate surface, or a very small (i.e., a few Ångströms) distance away from the substrate.<sup>[32,44]</sup> Therefore, it is considered to be a short-range effect. This mechanism is site-specific, analyte-dependent, and dependent also on the orientation of the adsorbed analytes.<sup>[45]</sup>

The adsorption of a molecule on a substrate can be physical in nature (“physisorption”) or chemical in nature (“chemisorption”). The former is driven by Van der Waals forces, and so the interaction enthalpy is generally around  $-20 \text{ kJ mol}^{-1}$ .<sup>[46]</sup> As a result, the structure of the molecule is modified only slightly. In the latter, a chemical bond is formed between the molecule and the surface, and so stronger perturbation is expected. The interaction enthalpy in this process is around  $-200 \text{ kJ mol}^{-1}$ .<sup>[46]</sup> In both physisorption and chemisorption processes, the molecule’s electronic and geometric structure is changed, albeit to a different extent. Therefore, the Raman cross-sections of its vibrational modes will generally be different with respect to the free molecule, and the chemical enhancement factor  $EF_{chem}$  can thus be defined as:<sup>[32]</sup>

$$EF_{chem} = \frac{\sigma_k^{ads}}{\sigma_k^{free}} \quad (2-3)$$

In **Equation 2-3**,  $\sigma_k^{ads}$  and  $\sigma_k^{free}$  refer to the Raman scattering cross-sections of the  $k$ -th vibrational mode of the adsorbed and free molecule, respectively.

Chemical enhancement can arise from two different mechanisms, including the non-resonant chemical effect and the resonant charge transfer effect.<sup>[47]</sup> In the non-resonant mechanism, the molecular orbitals of the analyte do not lay close enough in energy to the Fermi level of the metal, and thus a new electronic state is not formed. This interaction, however, may induce a geometrical change in the molecule as well as a change in the electronic structure of the molecule, leading to a small change in the Raman shifts and intensity of the vibrational modes.<sup>[32]</sup> Alternatively, chemical enhancement can arise from the resonant charge transfer effect, which increases the probability of a Raman transition by providing a pathway for resonant excitation.<sup>[44]</sup> If Raman scattering is excited with a laser that is in resonance or pre-resonance with the metal-molecule charge transfer state, Raman modes may be strongly enhanced, especially those coupled to the allowed electronic transitions.<sup>[32]</sup> The resonant charge transfer effect can also be transient in nature, in which temporary electron or hole transfer occurs between the metal and the analyte.<sup>[41]</sup>

Out of the two accepted theories for SERS enhancement, chemical enhancement is generally thought to contribute less to the overall enhancement observed compared to electromagnetic enhancement. The approximate contributions of the chemical enhancement mechanism typically range from magnitudes of  $10^2 - 10^4$ , where  $10^2$  enhancement is achieved from atomic scale roughness and  $10^4$  enhancement is achieved through charge transfer resonance.<sup>[48]</sup>

## 2.2.2 EM Enhancement

The EM enhancement mechanism is based on the LSPR processes described in Section 1.3.3. Unlike the chemical enhancement mechanism, EM enhancement is analyte-independent but substrate-dependent. The LSPR of a given substrate, which is the resonant frequency of the conduction electrons at the metal surface, is governed by the size and shape of nanoscale features on the metal substrate as well as the nature of the metal itself.<sup>[44]</sup> When the frequency of impinging light matches the LSPR of the substrate, large EM fields or “hot spots” are generated at the edges of the “roughness” features of the metal. When

analytes are confined within these large EM regions, the resulting Raman intensity is greatly amplified.

The Raman scattering cross-section of a given molecule is dependent on the strength of the local field  $|E_{loc}|$ , and the enhanced local field acts on both incident and scattered rays of light. Therefore, the overall Raman scattering cross-section is enhanced by the enhancement factor  $EF_{EM}$  which can be described as:<sup>[49]</sup>

$$EF_{EM} = \frac{|E_{loc}(\omega_I)|^2}{|E_0|^2} \cdot \frac{|E_{loc}(\omega_R)|^2}{|E_0|^2} \quad (2-4)$$

In **Equation 2-4**,  $E_{loc}$  is the local field,  $\omega_I$  is the frequency of the incident light,  $\omega_R$  is the frequency of Raman-scattered light, and  $E_0$  is the electric field of the incident light. In the case of small Raman shifts, where  $\omega_I \cong \omega_R$ , **Equation 2-2** can be simplified to:<sup>[50]</sup>

$$EF_{EM} = \frac{|E_{loc}(\omega_I)|^4}{|E_0|^4} \quad (2-5)$$

**Equation 2-5** is known as the  $|E|^4$  approximation. Consequently, a modest field enhancement by a factor of 10 will yield an  $EF$  of  $10^4$ . It is important to note that this approximation is more accurate at small Raman shifts if  $E_{loc}(\omega_I)$  is not too sharp near  $\omega_I$ ,<sup>[32]</sup> and that some situations require a more specific approach for quantifying enhancement. Nevertheless, the scaling of the enhancement factor to the fourth power greatly increases the sensitivity of SERS for molecules that are located near or at hot spot areas, allowing for even single molecule detection to be achieved.

The SERS effect arising from EM enhancement is generally strongest when the analyte is confined within 1 – 10 nm from the roughness features, a sensing distance much farther than what is required for chemical enhancement to be observed.<sup>[32]</sup> Therefore, the EM enhancement mechanism can be thought of as the long-range effect. As mentioned above, the EM enhancement mechanism is the dominant mechanism that contributes to the SERS effect, which can reach magnitudes of  $10^8 - 10^{10}$ , where  $10^8$  enhancement is observed when averaged over the substrate, and  $10^{10}$  enhancement is observed in a hot spot area.<sup>[51]</sup>

### 2.2.3 Materials for SERS

SERS substrates are traditionally fabricated with metals whose resonance positions are located in the visible and near-infrared regions, thereby matching the wavelengths of commonly used laser excitation sources (i.e., 532 nm, 633 nm, and 785 nm). Examples of commonly used metals for SERS probes and substrates include gold, silver, copper, and aluminum.<sup>[44]</sup> Gold and silver are the most commonly used metals for SERS applications,<sup>[52,53]</sup> particularly in biological studies, with their own unique advantages over the other. For example, gold has low toxicity and is far more stable in air than silver, which tends to readily oxidize and react with sulfur compounds in the atmosphere.<sup>[54,55]</sup> Copper and aluminum are less chemically stable in the environment as well, as they also form oxide layers in air.<sup>[56,57]</sup> Oxidation of SERS substrates and probes can decrease their plasmonic performances; for example, an oxide layer limits the proximity of an analyte to the SERS-active metal surface, lowering the enhancement observed, and can modify the analyte's affinity towards the metal surface. However, silver, copper, and aluminum substrates are much less expensive than gold substrates, which is not an insignificant factor when considering a platform for various applications, such as point-of-care diagnostics.<sup>[58]</sup>

Less traditional materials for SERS substrates are dielectric<sup>[59]</sup> and semiconductor materials,<sup>[60]</sup> which are being investigated to avoid optical losses<sup>[61]</sup> and unwanted plasmonic heating<sup>[62]</sup> that can occur with metallic substrates. The absorption and dissipation processes in metals leads to the release of heat which can modify an analyte. For example, plasmonic heating has been observed to change the composition of protein corona.<sup>[63]</sup> Non-metallic materials that have been studied in SERS substrates include transparent conductive oxides (e.g., indium tin oxide, aluminum-doped zinc oxide, and gallium-doped zinc oxide),<sup>[61,64]</sup> refractory transition metal nitrides (e.g., titanium nitride and zirconium nitride),<sup>[65,66]</sup> and two-dimensional materials (e.g., graphene).<sup>[67,68]</sup> Although interesting alternatives to conventional metallic probes and substrates, these materials have not been widely adopted in the SERS characterization and analysis of EVs.

## 2.3 Spectral Interpretation

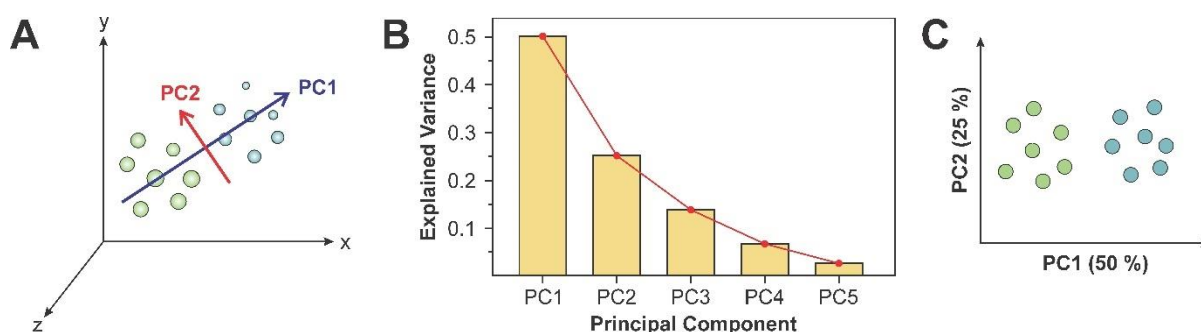
Like Raman spectroscopy, a disadvantage of using SERS to probe complex biological systems and samples such as EVs is that, unlike a small molecule, there is no single unique spectral fingerprint or spectrum that can belong to a single EV. Unlike the standard biochemical detection methods introduced in Section 1.3.2, which can identify specific protein and lipid cargo of EVs, SERS can only be used to identify small molecular moieties not necessarily specific to a particular biological macromolecule. Therefore, SERS spectra are generally interpreted by reducing their dimensionality, determining peaks responsible for the most variance within a data set, and establishing patterns for a set of EVs. Such methods include the commonly used PCA,<sup>[69,70]</sup> which can be used on its own or in series with stronger classification methods such as linear discriminant analysis (LDA)<sup>[71,72]</sup> or machine learning.<sup>[73,74]</sup>

### 2.3.1 Principles of PCA

PCA is a method for dimensionality-reduction that involves transforming a large set of variables into a smaller one that contains most of the information of the original data set.<sup>[75]</sup> The new variables produced by this method are referred to as principal components (PCs) and are constructed as linear combinations of the initial variables.<sup>[76]</sup> PCs are uncorrelated and most of the information of the initial data set is contained in the first PCs. For example, a data set with 5 dimensions will be represented by 5 PCs, where the maximum variance or information is squeezed into the first PC, then next, and so on until 100 % of variance is explained (**Figure 2-4A**). PCA is extremely valuable for the analysis of spectral data since these data sets can have hundreds or thousands of dimensions (i.e., wavenumbers).

Before treating data with PCA, the data must be standardized so that each initial variable contributes equally to the analysis.<sup>[77]</sup> PCA is sensitive to the variances in the initial data set, and variables with larger ranges will dominate over those with smaller ranges, leading to biased results. Covariance matrices are computed to identify correlations among the input data set. The sign of the covariances indicate how the variables are correlated: if the covariance is positive, then the two variables increase or decrease together (i.e., directly correlated), but if the covariance is negative, then one variable increases while the other

decreases (i.e., inversely correlated). Simply put, covariance matrices are tables that summarize the correlations between all possible pairs of variables. PCs can be solved by determining the eigenvectors and eigenvalues for the covariance matrices. For a data set with 5 dimensions, there are 5 variables, and so 5 eigenvectors will be calculated with 5 eigenvalues. The eigenvectors of a covariance matrix are the directions of the axes where there is the most variance (i.e., most information). In other words, the eigenvectors corresponding to the covariance matrix are the PCs. Ranking the eigenvalues from highest to lowest will rank the PCs in order of significance, and the graphical interpretation of this is what is known as a scree plot (**Figure 2-4B**). The percentage of explained variance of each PC is calculated by dividing the eigenvalue of each component by the sum of eigenvalues. Recasting the data along the PC axes generates score plots (**Figure 2-4C**), which can be useful in classification tasks and examining trends in the data.



**Figure 2-4: Schematic illustration of the process of PCA. (A) The original 5-variable data set can be visually represented by a maximum of 3 axes. The first PC is fitted so that it contains the maximum variance, and the second is positioned perpendicular to it. (B) The 5 calculated PCs can be visualized in a scree plot, and (C) a score plot is created by re-casting the original data on PC1 and PC2 axes.**

Naturally, simplifying a data set in this manner will lead to a decrease in accuracy. However, reducing the complexity of a spectral data set makes it easier to visualize and analyze the data, and also reduces the computational demand when further using machine learning or deep learning algorithms. However, PCA alone often fails to distinguish classes or separate them with a high enough accuracy to be used in diagnostic applications. This is due to the fact that PCA is an unsupervised transformation technique, meaning it ignores class labels. Usually, PCA is used in combination with techniques that explicitly attempt

to model the difference between classes of data such as LDA. While PCA determines the relationship between independent variables, the supervised technique LDA determines the relationship between dependent and independent variables, resulting in class separation by finding a lower-dimensional space that has better discriminatory power.<sup>[78]</sup> Another method that can be used to achieve higher classification accuracy than PCA is machine learning.

### 2.3.2 Machine Learning

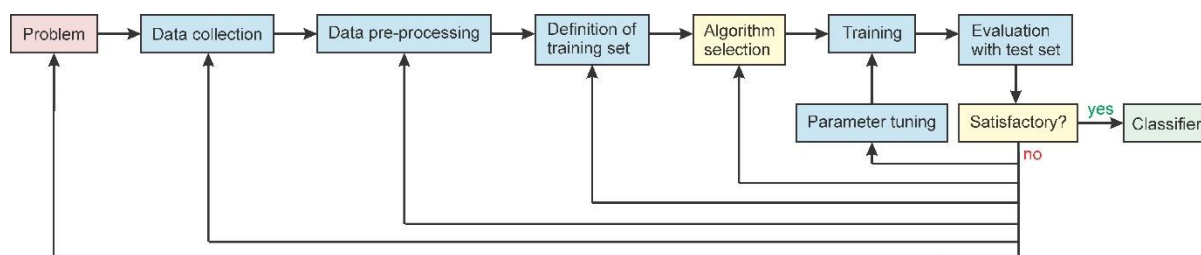
Machine learning is a type of artificial intelligence that is based on the idea that systems, with minimal intervention from humans, can learn from data, identify patterns in data, and make decisions. Although machine learning has been around as early as the 1950s,<sup>[79]</sup> recent developments made in the past couple of decades has led to its widespread use in many applications, including, but not limited to, finance analysis,<sup>[80,81]</sup> fraud detection,<sup>[82,83]</sup> search engines,<sup>[84,85]</sup> marketing,<sup>[86,87]</sup> and medical diagnosis.<sup>[88,89]</sup>

Two broad categories of machine learning are unsupervised machine learning, in which no labels are provided to the system and the algorithm finds hidden patterns in a data set, and supervised machine learning, whereby the system is trained through example inputs and their desired outputs.<sup>[79]</sup> Some of the most common supervised machine learning techniques include regression and forecasting, in which the machine analyzes trends in past and present data to make predictions about the future, and classification, in which the machine learns to draw conclusions from the data set and determines into which categories new observations belong.

Supervised machine learning with the goal of finding a classifier generally consists of the identification of a problem and required data, data pre-processing, selection of a training set, selection of an algorithm, training, and evaluation with a test set (**Figure 2-5**).<sup>[90]</sup> The researcher is, of course, responsible for the collection of data and its pre-processing to reduce noise and outliers. In cases with large data sets, dimensionality reduction (e.g., by PCA) is suggested to reduce computational time and improve computational efficiency. The choice of learning algorithm is a critical step in the machine's classification performance and is assessed by computing prediction or classification accuracy. Classifier



evaluation can be achieved by randomizing the data and splitting it into training and test sets, the ratio of which is defined by the user (e.g., 80 % is used for training and 20 % is used for testing).<sup>[91]</sup> Alternatively, classifiers can be evaluated by cross validation (CV), in which the data set is split into  $n$  mutually exclusive and equally sized subsets. In CV,  $n - 1$  subsets are used as training sets while the remaining set is used as a test group until each subset has been used for testing once, and the error rate of each subset is averaged to estimate the error rate of the classifier.<sup>[90]</sup> Leave-one-out CV (LOOCV) is a special case of CV in which each piece of data is used to train and test the classifier, which is the most accurate estimator of the classifier's error rate, but the most computationally expensive.<sup>[90]</sup> If the classification performance is unsatisfactory, modifications must be made, which can look like fine-tuning training parameters, selecting a different algorithm, reducing data or problem dimensionality, or even collecting a larger data set for training. The best learning method for a set of data can be chosen from a variety of different algorithms, which can be logic-based, perceptron-based, statistical, instance-based, or support-vector machines (SVM).



**Figure 2-5: Schematic illustration of the process of supervised machine learning.**

A common example of a logic-based algorithm is the decision tree, which classifies instances (i.e., rows of data) by sorting them based on features, or input variables, in a sequence of branching statements (**Figure 2-6A**).<sup>[92]</sup> The first split, or root node, considers all features of the training data during the test. Branches represent outcomes of tests, and can either flow into internal nodes, which are tests on the attributes, or leaf nodes, which hold class labels. Decision trees tend to perform better when features are discrete or categorical rather than continuous.<sup>[90]</sup> While decision trees are easy to read and interpret, they can create overly complex models that pose the risk of overfitting data.<sup>[93]</sup> While decision trees can be translated into a set of rules by creating a separate rule for each branch,

rules can also be directly induced from the training data via a rule-based algorithm,<sup>[94]</sup> such as the CN2 algorithm.<sup>[95]</sup> Rule-based algorithms are typically more comprehensible compared to decision trees when learning binary problems since rule-based algorithms learn a set of rules for the positive class only. However, if multiple classes must be learned, decision trees are more efficient than rule-based algorithms since they consider the entire data set one class at a time.<sup>[90]</sup>

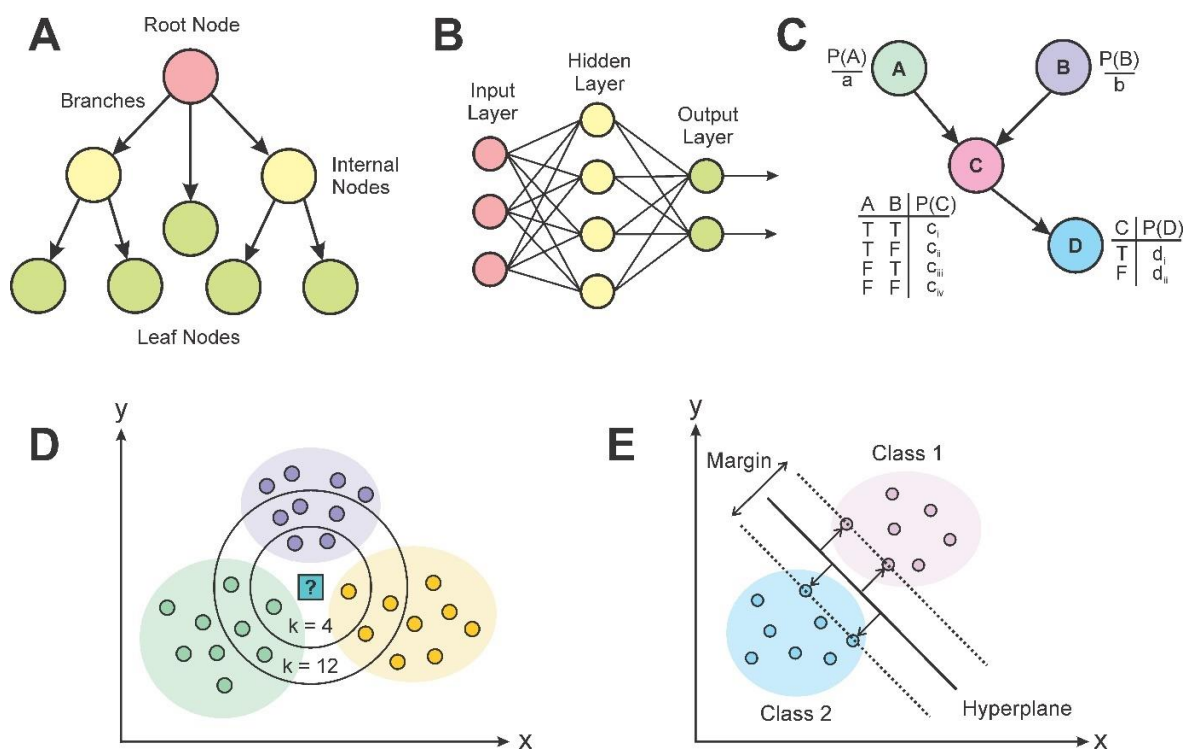
Perceptrons or “artificial neurons” are nodes in artificial neural networks (ANNs) that take one or more input values, run a (typically nonlinear) function on the weighted sum of the inputs, and compute a single output value.<sup>[96]</sup> Feed-forward ANNs allow signals to travel in only one direction (i.e., input to output) (**Figure 2-6B**).<sup>[97]</sup> A hidden layer between the input and output layers contains a function for training. The network is initially trained on a set of paired data to determine input-output mapping, and weights of the connections between neurons are fixed. When a new data set is introduced, the network is used to determine new classifications. Multilayer perceptron-based ANNs can usually provide incremental learning more easily compared to decision trees, but training time is usually much longer and they perform about equally as well.<sup>[90]</sup>

Statistical learning algorithms, in contrast to neural networks, have an explicit underlying probability model, meaning they calculate a probability that an instance belongs in each class, rather than a simple classification.<sup>[90]</sup> Examples of statistical learning algorithms include LDA, discriminant correspondence analysis (DCA), Bayesian networks, and logistic regression. LDA, as mentioned in Section 2.3.1, finds the linear combination of features which best separates classes.<sup>[98]</sup> LDA is used when observations are continuous quantities, while DCA is an equivalent technique used for discrete or categorical data.<sup>[90]</sup> Bayesian networks are graphical models for probability relationships among a set of variables and, unlike decision trees or neural networks, can consider prior information about a given problem in terms of structural relationships among its features (**Figure 2-6C**).<sup>[99]</sup> However, since Bayesian network classifiers try to construct a very large network, it is not a suitable method for examining data sets with many features.<sup>[100]</sup> Logistic regression is often used in binary classification problems and implements a sigmoid function to convert the raw prediction of a linear model into a value between 0 and 1.<sup>[101]</sup>

This output value can be interpreted as either the probability that the instance belongs to the positive class in the binary classification problem or as a value to be compared against a classification threshold (i.e., a criterion separating positive and negative classes).

Instance-based learning algorithms are specific types of statistical methods. Since instance-based learning algorithms delay the generalization process until classification is performed, they are known as lazy-learning algorithms.<sup>[102]</sup> Such algorithms require less computational time during the training phase compared to eager learning methods (e.g., decision trees, neural networks, and Bayesian networks), but require more computational time during classification. One of the most common instance-based learning algorithms is *k*-Nearest Neighbor (kNN), which is based on the idea that instances within a data set will generally exist in close proximity to other instances that have similar qualities or properties (**Figure 2-6D**).<sup>[103]</sup> Using labelled instances, the value of the label of an unclassified instance can be determined by observing the class of its nearest neighbors, and the algorithm locates its *k* nearest instances and determines its class by identifying the single most frequent class label. Although kNN can be very powerful, they can become very computationally expensive since the only to choose *k* (i.e., number of nearest instances to consider) is through cross validation.<sup>[90]</sup>

Compared to other machine learning algorithms, SVM has emerged quite recently. SVMs are based on “margins” surrounding hyperplanes that separate data classes (**Figure 2-6E**).<sup>[104]</sup> Support vectors refer to the data points closest to the hyperplane that influence its optimal position and orientation. By maximizing the margin, the largest possible distance is created between the separating hyperplane and the instances on either side of it, which increases the confidence in the projections and reduces the expected generalization error. Since the number of support vectors selected by the SVM is usually small, the number of features in a data set does not affect the model complexity of an SVM, making SVMs suitable to deal with learning tasks in which the number of features is large compared to the number of training instances.<sup>[90]</sup>



**Figure 2-6: Schematic illustrations of some machine-learning algorithms, highlighting (A) decision trees, (B) feed-forward ANNs, (C) Bayesian networks, (D) kNNs, and (E) SVMs.**

The discriminatory power of classification models is commonly compared using accuracy, sensitivity, specificity, and receiver operator characteristic (ROC) curves.<sup>[101]</sup> Sensitivity, which can also be referred to as recall, is a measure of the true positive rate, while specificity is a measure of the true negative rate. Plotting  $(1 - \text{specificity})$  versus sensitivity at varying values of classification threshold generates an ROC curve, which is a simple graphical tool for displaying the accuracy of a diagnostic test.<sup>[105]</sup> The area under the curve (AUC) of the ROC curve is therefore a summary of the diagnostic accuracy. An AUC value of 0.5 corresponds to random chance, while an AUC value of 1.0 corresponds to perfect accuracy. While an AUC greater than 0.5 is generally good, researchers aim for as close to 1.0 as possible, which is especially important in medical applications. If the AUC is less than 0.5, the test is worse than random chance.<sup>[106]</sup>

There is no single learning algorithm that can uniformly outperform other algorithms across all data sets. The simplest approach to determine the most accurate algorithm for a

particular data set is to estimate the accuracies of several algorithms and select the algorithm that is most accurate. In Chapters 3 and 4, several algorithms are tested and compared to determine the best algorithms for classifying MSC-derived EVs and cancer-derived EVs.

## 2.4 SERS Probes and Platforms for EV Capture and Characterization

SERS probes and platforms can generally be grouped into two methods: direct or “label-free” detection and indirect detection. Direct detection methods allow for direct probing of an analyte without a label or “tag”, since the analyte’s SERS spectrum is recorded.<sup>[107,108]</sup> Indirect detection methods rely on a SERS tag for analyte detection. SERS tags consist of an efficient Raman reporter, which is a molecule with a large Raman-scattering cross-section and characteristic peak or spectrum, and are engineered to selectively bind to the analyte. The analyte is thus detected through the Raman reporter, whose signal intensity is proportional to the analyte concentration.<sup>[109,110]</sup> Indirect detection can be used when analytes are immersed in complex matrices (e.g., plasma) that can generate an interfering signal with the analyte. Although indirect detection often requires more complex fabrication and analysis compared to direct detection, the available sensitivity and specificity are oftentimes much higher. The main disadvantage of indirect detection is that, although readout is greatly simplified for medical personnel who are not experts in Raman spectroscopy, rich biological information is lost when true analyte signals are not tracked. SERS probes and platforms have been developed for EV analysis using both direct detection and indirect detection methods; the former is reviewed in Sections 2.4.1 and 2.4.2 while the latter is reviewed in Sections 2.4.3 and 2.4.4. Although both are reviewed to discuss recent advances and shortcomings of current studies, only direct SERS sensing is used in Chapters 3 and 4.

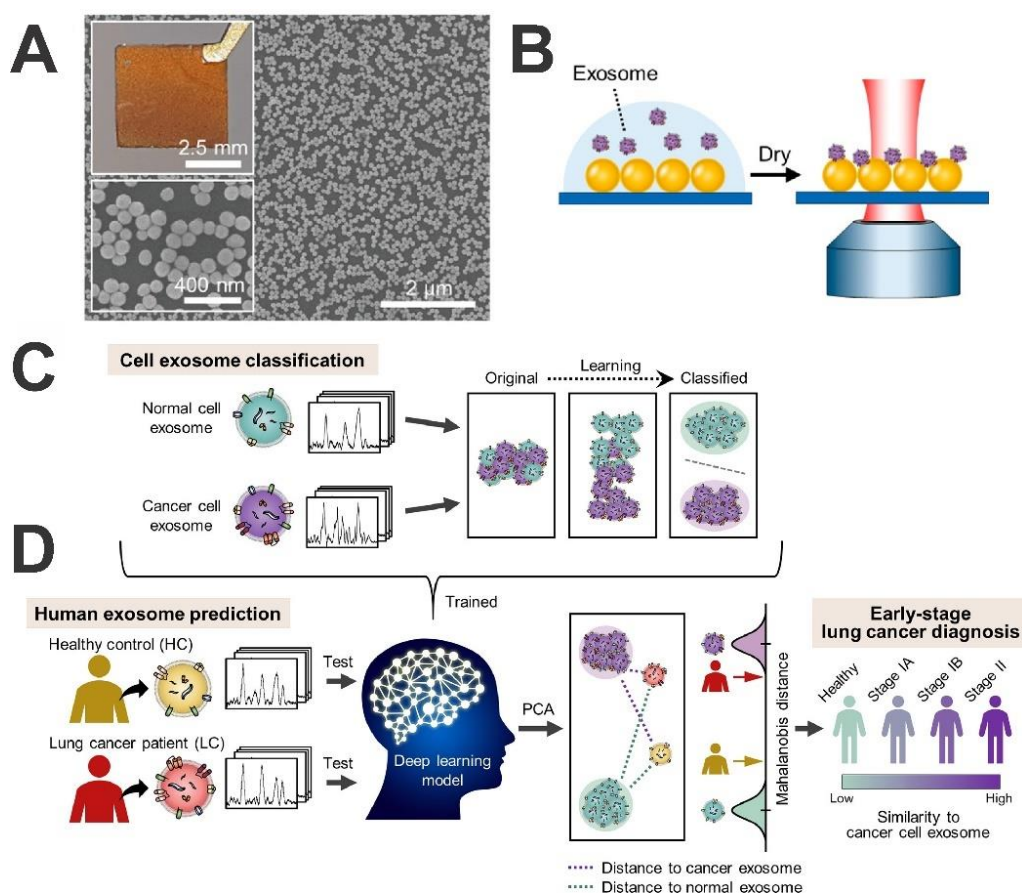
### 2.4.1 Nanoparticle Probes for Direct Sensing

Nanoparticle probes, either cast on a substrate or in suspension, are commonly used for SERS due to their simplicity. Simple gold nanoparticles (AuNPs) have been used as SERS probes in the characterization of EVs derived from lung cancer cells.<sup>[111-113]</sup> AuNPs with a

diameter of approximately 80 nm were implemented by Park et al. to probe EVs analyzed from two lung cancer cell lines, H1299 and H522, as well as from alveolar cells as a normal control.<sup>[111]</sup> These AuNPs were dried on coverglasses, and SERS measurements were performed on dried EVs overtop the substrate with an excitation wavelength of 785 nm. Using PCA, cancerous EVs and normal EVs could be discriminated with 95.3 % sensitivity and 97.3 % specificity, although accuracy was not reported. The authors further attempted to use PCA score plots to predict the outcomes of clinical samples consisting of 2 lung cancer patients and 2 healthy individuals. However, the platform failed at this stage, as the clinical sample scores did not fall into the 95 % confidence ellipses of their respective group. One year later, Shin et al. built on this work, instead clustering AuNPs on a (3-aminopropyl)triethoxysilane (APTES)-coated substrate.<sup>[112]</sup> In contrast to the previous work, the authors instead performed SERS measurements in a liquid state to prevent signal deformation by EV damage or salt formation. The AuNPs were also coated with cysteamine, which has been used to induce the electrostatic adsorption of EVs via interactions between the cationic amino groups of cysteamine and anionic EV surfaces. Cancer cell lines PC9 and H1299 were investigated with the normal lung cell line HPAEC as a control. Although cancerous and noncancerous groups could be distinguished by PCA, the authors did not report sensitivity or specificity for comparison with the previous study. Additionally, despite the proposed SERS effect, the raw spectra still contained background fluorescence, thus requiring treatment prior to analysis.

Recently, Shin et al. implemented the same substrate idea, but greatly improved analysis with a sophisticated deep learning algorithm.<sup>[113]</sup> AuNPs with 100 nm diameter were coated onto an APTES-covered substrate (**Figure 2-7A**), and used to analyze dried EVs from lung cancer cell lines (A549, H460, H1299, H1763, and PC9) and from normal alveolar epithelial cells HPAEC (**Figure 2-7B**). The deep learning algorithm trained with 80 % of the cell-derived EV data could predict the validation set with 94.8 % accuracy, and thus the entire cell-derived data set was used as a training set to predict the outcomes of clinical plasma-derived EVs (**Figure 2-7C**). The clinical data set was comprised of 20 healthy individuals and 43 lung adenocarcinoma patients. The lung cancer patients were further identified by their disease stage, where 22 patients had stage IA cancer, 16 patients had stage IB cancer, and 5 patients had stage II cancer. PCA was employed to determine the

similarity between the clinical samples and the cancer cell line-derived EVs, which was quantified by calculating the Mahalanobis distances between PC score clusters (**Figure 2-7D**). On average, cancer patients displayed 1.5 times more similarity to the cancer cell-derived EVs compared to the healthy individuals, and the cancer patients could be discriminated from the healthy individuals with 84 % sensitivity and 85 % specificity. Furthermore, the authors established the potential of the device for early cancer screening by discriminating the stage IA patients from the healthy individuals with 73 % sensitivity and 85 % specificity. However, like the previous study, the raw spectra required significant preprocessing prior to analysis due to high background fluorescence, and the calculated sensitivity and specificity remains low for clinical translation.



**Figure 2-7:** A 100 nm AuNP-covered SERS substrate for the analysis of lung cancer-derived EVs. (A) SEM images and an optical image (inset, top left) of the AuNPs assembled over an APTES-covered substrate. (B) Schematic illustration of the acquisition of SERS spectra. (C) Lung cancer cell line-derived EVs and normal lung

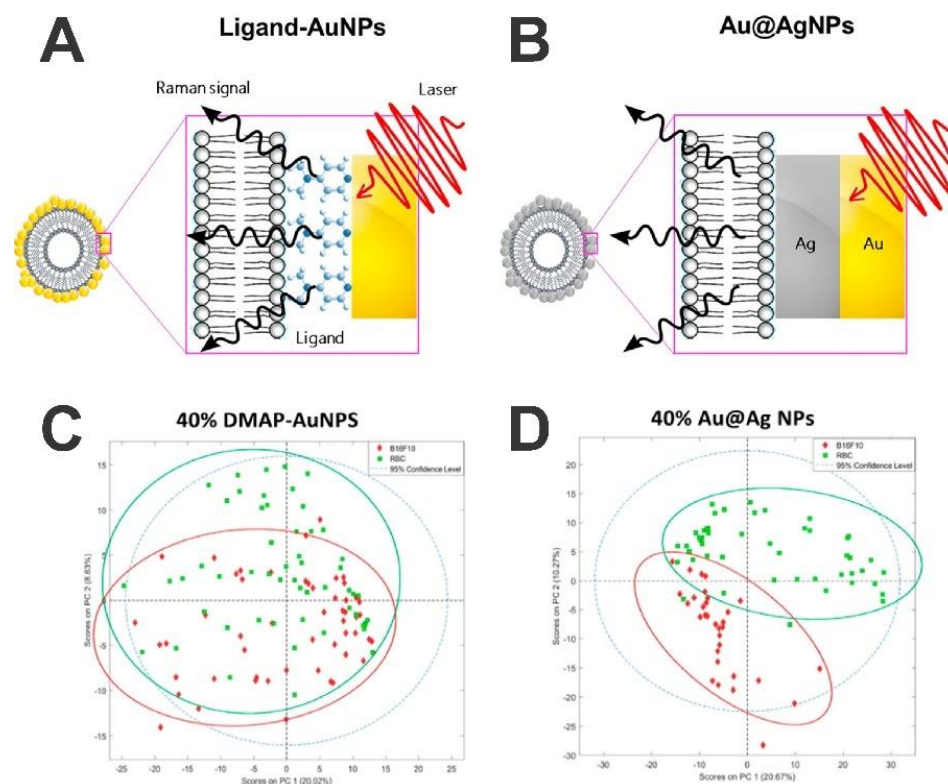
**cell line-derived EVs are used to train a deep learning algorithm for lung cancer diagnosis. (D) The deep learning model is tested with EVs derived from lung cancer patients and healthy individuals, and their similarities to the lung cancer cell-derived EVs are compared. Adapted with permission from reference [113] (copyright 2020 American Chemical Society).**

Simple nanoparticles have additionally been used as SERS probes to characterize EVs derived from pancreatic and cervical cancer cells.<sup>[114,115]</sup> AuNPs with 10 nm diameter were utilized by Carmichael et al. to analyze EVs from two pancreatic cancer cell lines, CD18/HPAF and MiaPaCa, and a normal pancreatic cell line, HPDE, and SERS measurements were performed with a 785 nm excitation wavelength.<sup>[114]</sup> With principal component analysis-discriminant function analysis (PCA-DFA), cancer EVs could be discriminated from normal EVs with 90.0 % accuracy, 90.6 % sensitivity, and 97.1 % specificity. The probes were further tested with clinical samples from patient sera, examining 10 healthy individuals and 10 early-stage pancreatic cancer patients. However, the discriminatory capabilities were not as strong as with the cell line models, with characterization efficiencies ranging from 20 – 87 % for healthy individuals and 30 – 90 % for cancer patients. Another research group utilized citrate-reduced silver nanoparticles (AgNPs) with 200 nm diameter to analyze cervical cancer EVs.<sup>[115]</sup> Such nanoparticles were able to surround individual EVs (approximately 2 – 5 AgNPs per vesicle), and displayed a resonance position around 434 nm. Spectra of EVs from two cervical cancer cell lines, HeLa and Atg<sup>5-/-</sup>, and one noncancerous cell line, HEK293, were collected with a 633 nm excitation wavelength and analyzed by PCA. Although the cluster of normal EVs was separate from the HeLa-derived EVs, the normal EVs overlapped with the Atg<sup>5-/-</sup>-derived EVs, and accuracy, sensitivity, and specificity were not reported. Furthermore, each spectrum was heavily dominated by background peaks resulting from the total exosome isolation reagent used for EV isolation.

Recognizing the importance of single vesicle characterization for diagnostic applications, Stremersch et al. implemented 4-(dimethylamino)pyridine (DMAP)-coated AuNPs approximately 10 nm in diameter.<sup>[116]</sup> DMAP-coated AuNPs were able to form irregularly-shaped nanoshells around single EVs based on the electrostatic absorption of cationic



DMAP-AuNPs onto anionic EV surfaces, and display an LSPR band around 522 nm. The authors analyzed EVs derived from melanoma cell line B16F16 and red blood cell (RBC)-derived vesicles, and SERS spectra were obtained with a 785 nm excitation wavelength. Using partial least squares-discriminant analysis (PLS-DA), melanoma EVs could be discriminated from RBC-derived EVs with 88.0 % sensitivity and 95.4 % specificity, while accuracy was not reported. Although the authors attempted to characterize both spectra, EV-related peaks were generally poorly resolved, and each spectrum was additionally dominated by presumed DMAP molecules. In addition to having a high Raman scattering cross-section, DMAP molecules are closest to the AuNP surface, and therefore at the region of highest enhancement. To address the problem of DMAP-dominated spectra, Fraire et al. developed *in situ* silver-coated AuNPs.<sup>[117]</sup> *In situ*-coating of silver allowed the EV membrane to be in direct contact with a metal surface (**Figure 2-8A,B**), as opposed to DMAP, and also shifted the LSPR to 490 nm, with an additional band appearing at 418 nm. The same EV sources were analyzed again, and with PLS-DA, melanoma EVs could be discriminated from RBC-derived EVs with 91.7 % sensitivity and 96.9 % specificity, an improvement compared to the use of bare DMAP-coated AuNP probes (**Figure 2-8C,D**). Although DMAP peaks no longer dominated each spectrum, there remained substantial noise in both spectra, and both studies may have benefited from using a laser excitation wavelength better matched to the LSPR of the probes, such as a 532 nm laser.



**Figure 2-8: DMAP-coated AuNPs with silver nanoshells are used as SERS probes for the analysis of melanoma-derived EVs. Schematic illustrations of SERS acquisition (A) without and (B) with the silver nanoshell, highlighting the closer proximity of the EV membrane to the SERS-active metal surface when the probes are coated *in situ* with a silver layer. Analysis of melanoma cell line-derived EVs (red cluster) and RBC-derived EVs (green cluster) (C) without and (D) with the silver nanoshell, illustrating higher discriminatory power when the silver nanoshell is implemented. Adapted with permission from reference [117] (copyright 2019 American Chemical Society).**

Magnetic beads (MBs) functionalized for EV capture are commonly used in the SERS analysis of EVs. With the application of an external magnet, EVs can be concentrated under a laser spot, further enriching signal intensity.<sup>[118-121]</sup> Li et al. have reported gold nanodot-covered MBs for the analysis of breast cancer EVs.<sup>[118]</sup> Superparamagnetic nanoparticles were modified to introduce thiol groups for the anchorage of gold nanodots and were further functionalized with anti-CD9 for EV capture. Such probes had a diameter of 360 nm and displayed an LSPR band around 540 nm. EVs from two breast cancer cell lines, MCF-7 and MDA-MB-231, were chosen due to the cell lines' differences in metastatic

potential. With PCA, the two groups were able to be discriminated with 100 % sensitivity and specificity. The authors further tested the platform with clinical serum samples from 14 breast cancer patients and 6 healthy individuals, and were able to discriminate these groups with 91.67 % sensitivity and 100 % specificity. Although the calculated sensitivities and specificities appear extremely high, it is important to note that a very limited sample size was used, and the reported SERS spectra across the groups were virtually identical. The spectra appear not to have been standardized prior to PCA, in which case the PC score plots are based on variations in signal intensity rather than variations in Raman shift. Variations in Raman signal intensity cannot be interpreted without knowing analyte concentrations.

An unconventional substrate consisting of AgNPs grown *in situ* into bacterial cellulose (BC) from commercial *nata de coco* was described by Ferreira et al.<sup>[122]</sup> The motivation behind this work was to develop a cost-efficient SERS substrate for breast cancer diagnosis, since silver is less expensive than the more commonly used gold, and *nata de coco* is a low-cost and mass-produced source of BC. The AgNPs grown into the BC membrane were approximately 92 nm in diameter, providing an LSPR band around 390 – 460 nm. EVs were analyzed from the breast cancer cell line MDA-MB-231, with nontumorous breast epithelial cells MCF-10A used as a control. A laser with an excitation wavelength of 532 nm was used to probe 20 cancerous and 18 normal samples, and PCA was used to discriminate the groups. No overlap was observed between the 95 % confidence ellipses of the two groups. Furthermore, 2 test samples (i.e., 1 cancer-derived sample and 1 normal sample) were used as proof-of-concept to demonstrate diagnostic potential, and each fell into their respective cluster. However, limited sample sets were used in this study.

Another novel nanoparticle-based substrate was proposed by Pramanik et al., who implemented a mixed-dimensional heterostructure platform consisting of two-dimensional graphene oxide and gold nanostars (GO-AuNS) for the analysis of breast cancer EVs.<sup>[123]</sup> The SERS effect is influenced by both the GO and the AuNSs, the former of which contributes via the chemical enhancement mechanism, whereas the latter contributes via the stronger electromagnetic enhancement mechanism. A strong LSPR band around 580

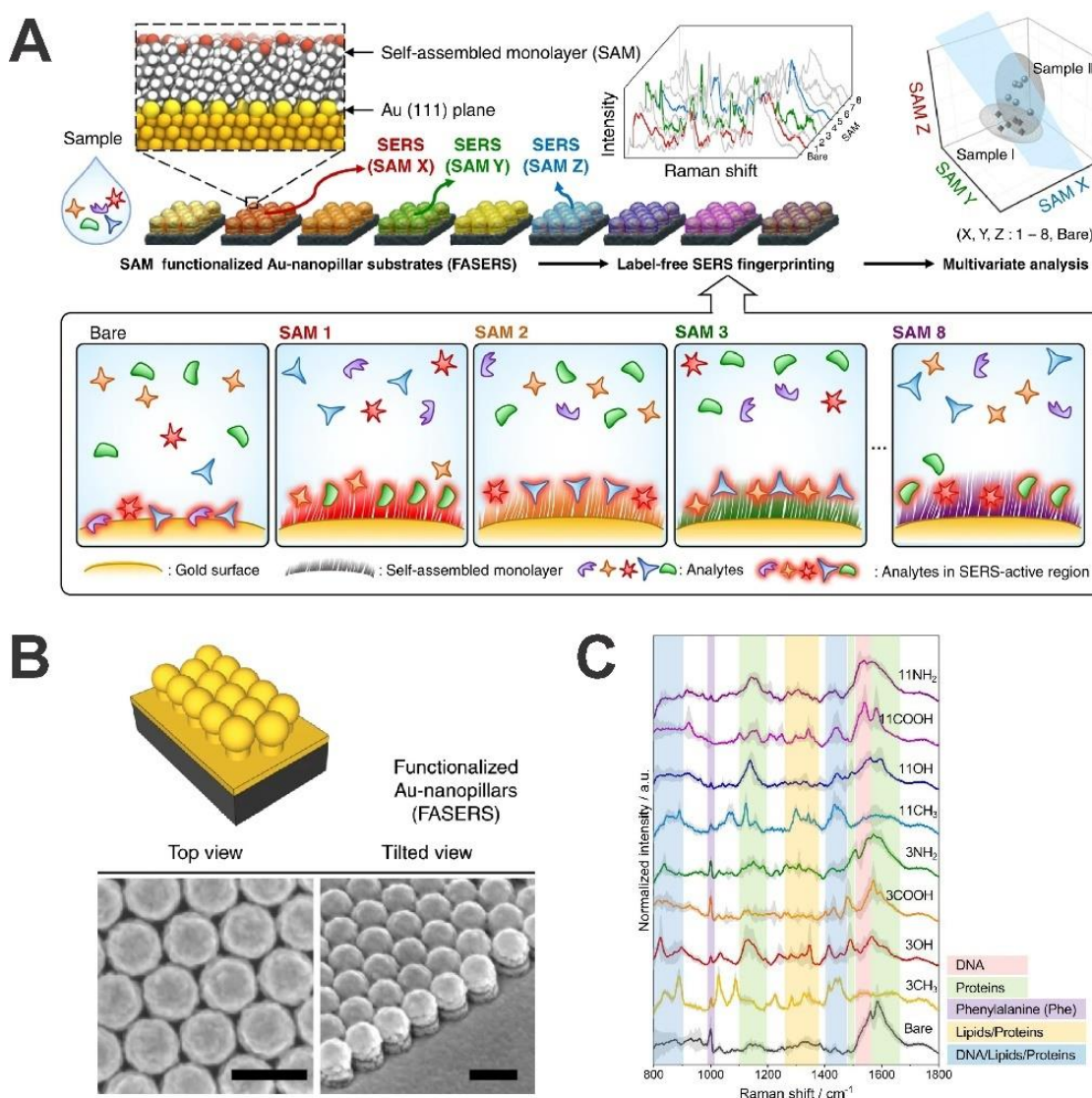
nm is observed, as well as a broader band around 720 nm, and so the authors selected a 670 nm excitation laser. EVs from breast cancer cell lines MDA-MB-231 and SKBR3 were characterized by SERS. Although the spectra were well-characterized, a normal control group was not utilized, and so conclusions cannot be drawn to the platform's diagnostic potential.

#### 2.4.2 Periodic Arrays of Nanostructures for Direct Sensing

Nanopillars are common, simple SERS substrates for EV analysis.<sup>[124-126]</sup> Early work involving pillars for SERS analysis of EVs was published by Tirinato et al., who fabricated super hydrophobic structures (silicon micropillars, approximately 10 nm in diameter) by standard optical lithography techniques and reactive ion etching (RIE).<sup>[124]</sup> The pillars were further covered with silver nanoaggregates. EVs from a colon cancer cell line (HCT116) and from healthy colon cells (CCD-841-CoN) were analyzed by SERS with a 514 nm excitation wavelength. However, the SERS spectra of both EV types appeared nearly identical, with variations limited to differences in intensity rather than peak positions. In addition to limited characterization, the diagnostic potential of the platform was not assessed. Many years later, Sivashanmugan et al. proposed a gold nanorod (AuNR) array covered with silver nanocubes (AgNCs) on top of the rods for the analysis of lung cancer EVs.<sup>[125]</sup> AuNRs, with diameters of approximately 126 nm and heights of approximately 400 nm, were fabricated by a focused ion beam method. AgNCs with edge lengths of approximately 65 nm were able to self-assemble overtop the pillars. The resulting gap mode substrate provided three LSPR modes around 529 nm, 752 nm, and 881 nm. EVs from three lung cancer cell lines (HCC827, H1975, and PC9) and three normal lung cell lines (L929, BEAS-20, and NL-20) were analyzed by SERS with a selected excitation wavelength of 785 nm. The normal EVs exhibited strong protein, nucleic acid, and lipid signals, while the cancer-derived EVs exhibited only strong protein signals. Additionally, the authors found that the cancer-derived EVs displayed a higher diversity in peaks compared to the normal EVs, which was confirmed by Western blotting, and potentially the result of the highly dysregulated feature of cancer cells. Although the authors describe the potential of this platform in diagnostic applications, diagnostic criteria such as

accuracy, sensitivity, and specificity were not assessed. Furthermore, the SERS spectra of all six EV types were extremely noisy, with peaks not well-resolved.

A novel nanopillar array was proposed by Kim *et al.* and named Functionalized Array for SERS (FASERS).<sup>[126]</sup> Gold nanopillars with 300 nm diameter and 500 nm height were fabricated by colloidal lithography and plasma etching and were functionalized with 8 different self-assembled monolayers (SAMs) (**Figure 2-9A,B**). The SAMs varied in functional group (i.e., alkyl, hydroxyl, carboxyl, and amine) as well as in carbon chain length (i.e., 3-carbon and 11-carbon). SAMs were selected for EV capture as an alternative to an immunoaffinity-based approach, so that compositional diversity of biological samples are not lost. When examined with small test molecules, the SERS signal was observed to substantially diminish with longer-chain SAMs, owing to the fact that the molecule is positioned farther away from the enhancement region of the nanopillars. Carboxyl and amine groups were also generally able to provide more favorable interactions with test molecules, leading to a higher observed SERS signal compared to the other SAMs. The platform was tested with EVs from the breast cancer cell line MDA-MB-231 using an excitation wavelength of 785 nm (**Figure 2-9C**). However, since EV analysis was not the focus of the paper and rather for proof-of-concept, multivariate analysis was not reported. Instead, the authors tested the platforms with more complex cell lysates from the breast cancer cell line Hs578T and from the normal cell line Hs578Bst. With PCA-LDA, the lowest obtained accuracy was 41.7 %, obtained with the 11 carbon-length COOH SAM. The highest accuracy obtained was with the 3 carbon-length NH<sub>2</sub> SAM at 91.7 %, with a corresponding sensitivity and specificity of 83.3 % and 100 %, respectively. Although an interesting platform design, the diagnostic potential is not as high as other SERS platforms that use EVs for analysis.



**Figure 2-9: FASERS substrate for the analysis of breast cancer-derived EVs and cell lysates. (A) Schematic illustration of FASERS. Each SAM uniquely provides insight into the biological composition of a given analyte. (B) Schematic illustration (top), top-view SEM image (bottom left), and tilted-view SEM image (bottom right) of the bare FASERS (scale bars = 400 nm). (C) SERS spectra of breast cancer cell line-derived EVs collected with each FASERS. Adapted with permission from reference [126] (copyright 2020 Yarovsky, I. & Stevens M. M.).**

A periodic array of gold pyramidal nanostructures overlaid with single-layer graphene for the analysis of lung cancer-derived EVs was reported by Yan et al.<sup>[127]</sup> Such nanopyramids have a base edge length of approximately 200 nm and 57.7 ° sidewall angle, as well as a

center-to-center distance of approximately 400 nm. The graphene layer serves a dual purpose: to provide a biocompatible and protective layer over the metal surface (e.g., prevent oxidation) since it is chemically inert, and to gauge local EM field intensity. Gauging the EM field intensity by normalizing the biological SERS signal with the graphene peaks allows for the direct correlation of the biological analyte peaks with the amount of analyte present in the sample, as opposed to a convolution existing between the EM field intensity and the amount of analyte. EVs from two lung cancer cell lines, HCC827 and H1975, were analyzed by SERS with a 785 nm excitation wavelength, and human and bovine serum were used as controls. Using PCA, all four sample sources could be separated with < 5 % overlap among groups and a sensitivity > 84. %. Interestingly, the highest degree of overlap was not between the two lung cancer cell groups, but between the H1975-derived EVs and the bovine serum. However, the sensitivity achieved is not competitive with other SERS substrates and methodologies for cancer diagnosis, and accuracy and specificity were not assessed.

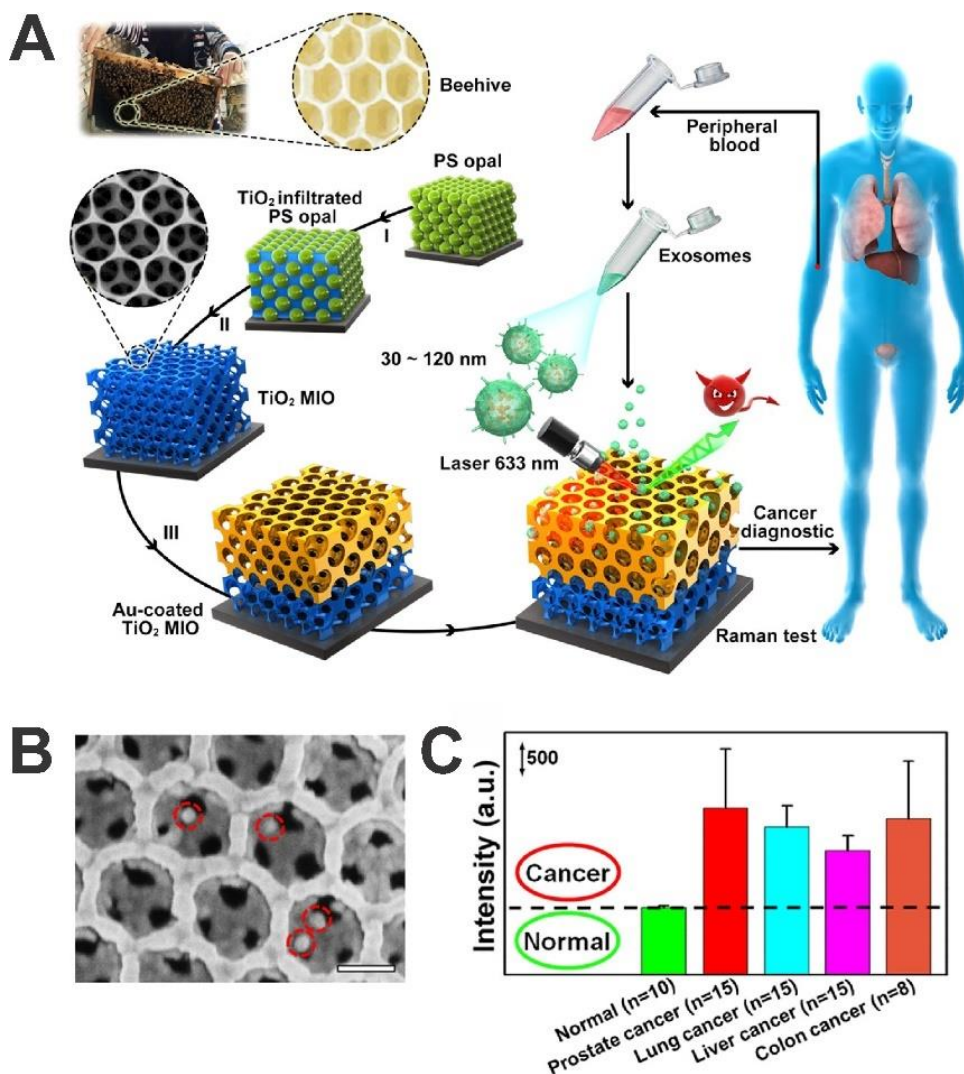
Recently, a microfluidic device based on silver nanobowtie-shaped antennae was reported by Jalali et al.<sup>[128]</sup> These antennae were approximately 150 nm in length, with gaps ranging from 0 – 100 nm, and provided three LSPR modes at 410 nm, 460 nm, and 580 nm. The microfluidic device was fabricated by a top-down standard lithography method while a bottom-up method based on the self-assembly of a polystyrene monolayer on silicon was used to fabricate the antennae. EVs derived from two human glioma cell lines, U373 and U87, were analyzed by SERS using an excitation wavelength of 532 nm, along with synthetic liposomes and normal glial cells (NHA) as controls. Such groups could be distinguished by PCA, as their 95 % confidence ellipses displayed low overlap. Additionally, the U87-derived EVs could be discriminated from the U373-derived EVs, the latter of which shared more similarity with the normal EVs. However, a limited sample size was studied, and clinical samples have not yet been analyzed. Furthermore, the antennae were not uniformly distributed over the substrate surface, meaning the EM field intensity is not homogenous across the substrate. Since the authors utilized silicon for substrate fabrication, a strong silicon peak around  $520\text{ cm}^{-1}$  was present in the spectra, and so the spectra required normalization.

An unconventional, grooved SERS substrate for EV analysis was proposed by Avella-Oliver et al.<sup>[129]</sup> These structured polycarbonate substrates were obtained from standard recordable disks (CD-R and DVD-R) and coated with silver for SERS. The goal of this study was to develop an extremely cost-effective device by reusing disks and by utilizing silver instead of gold. The groove pitches for the DVD-based and CD-based substrates were approximately 1.4  $\mu\text{m}$  and 0.7  $\mu\text{m}$ , respectively. The resulting LSPRs were located at 650 nm for the DVD-based substrates and at 780 nm for the CD-based substrates. EVs from lung cancer cell line A549 were analyzed by SERS, with excitation wavelengths of 633 nm (for the DVD-based substrates) and 785 nm (CD-based substrates). Although the SERS spectra were published, the authors did not characterize the spectral peaks. The paper was positioned as proof-of-concept, focusing more on the novelty of the substrate rather than the diagnostic potential, so normal controls were not analyzed as opposed to most studies.

A novel beehive-inspired three-dimensional gold-coated  $\text{TiO}_2$  macroporous inverse opal (MIO) structure was introduced by Dong et al. (**Figure 2-10A**).<sup>[130]</sup> Advantages of such a substrate include ideal trapping conditions for separating EVs from biofluids (**Figure 2-10B**) and the enhancement of Raman signal by both LSPR and by the slow light effect (i.e., the ability to trap and scatter light multiple times). With a pore diameter of approximately 290 nm, the resulting resonance band was located around 630 nm. EVs from prostate cancer cell line LNCaP and normal prostate cell line RWPE-1 were analyzed by SERS using an excitation wavelength of 633 nm. Since protein phosphorylation is one of the most fundamental mechanisms for the regulation and control of protein activities and functions within cells, the authors monitored the degree of protein phosphorylation by quantifying the spectral peak at  $1087\text{ cm}^{-1}$ , corresponding to the P-O bond. LNCaP-derived EVs displayed a P-O peak approximately 3.0 times higher than RWPE-1-derived EVs, indicating a higher degree of protein phosphorylation in the cancer cells. The authors also analyzed EVs derived from patient samples, assessing not only prostate cancer patients, but lung, liver, and colon cancer patients as well. EVs were derived from 15 prostate cancer patients, 15 lung cancer patients, 15 liver cancer patients, 8 colon cancer patients, and 10 healthy individuals. Again, a higher degree of protein phosphorylation was observed in the cancer-derived EVs, with the P-O peak 2 – 2.5 times more intense than that from the normal



control group (**Figure 2-10C**) and was stronger in over 93 % of cancer cases compared to the healthy individuals. However, the patient data sample size was limited, and the authors did not report accuracy, sensitivity, and specificity based on their method.



**Figure 2-10:** A beehive-inspired three-dimensional gold-coated TiO<sub>2</sub> MIO structure for the analysis of prostate cancer, lung cancer, liver cancer, and colon cancer derived-EVs. (A) Schematic illustration of the structure fabrication and acquisition of SERS spectra. (B) SEM image of the substrate that has captured 4 EVs, which are circled in red (scale bar = 300 nm). (C) The P-O band intensities of EVs derived from cancer patients are significantly higher on average compared to the P-O band intensities of EVs derived from healthy individuals. Adapted with permission from reference [130] (copyright 2020 American Chemical Society).

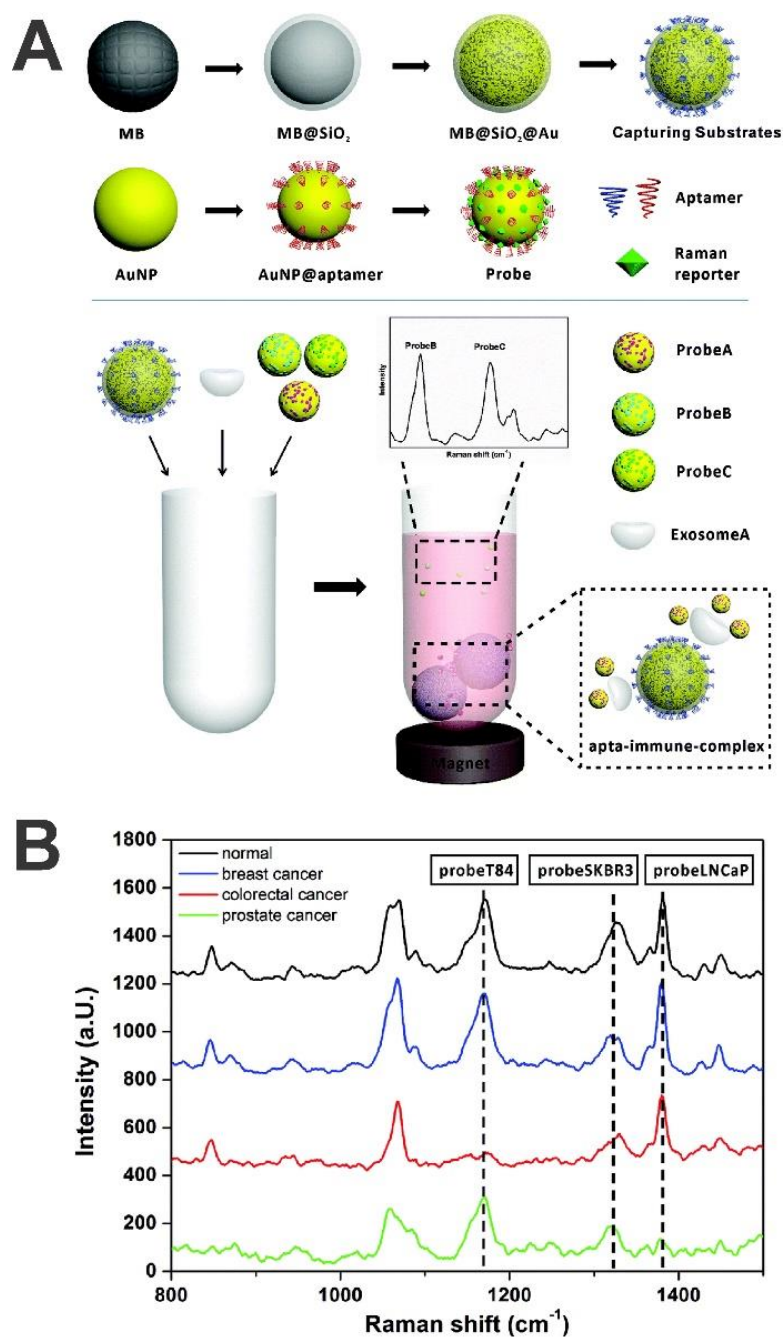
Other nanoarrays consisting of cavities for EV capture have been reported, such as silver nanobowls fabricated by soft lithography<sup>[131]</sup> and nanoholes fabricated by electron-beam lithography.<sup>[132,133]</sup> Since these platforms were used in the investigation of ovarian cancer-derived cells in literature and for this thesis, respectively, they are reviewed in detail in Chapter 4.

### 2.4.3 Probe-Based Immunocomplexes for Indirect Sensing

One sandwich-type immunocomplex detailed by Zong et al. consists of a magnetic capture probe and a gold core-silver shell nanorod (Au@AgNR) as a SERS tag.<sup>[119]</sup> Prior to functionalization with capture antibody anti-CD63, the Fe<sub>3</sub>O<sub>4</sub> MBs are coated with a silica shell to provide more available surface area for the antibodies as well as to protect the MBs from degradation. The Au@AgNRs are attached to Raman reporter 5,5-dithiobis-(2-nitrobenzoic acid) (DTNB) via Ag-S bonds and further coated with a silica shell to protect the DTNB from disturbance from the outer environment. For binding with EVs, the DTNB-tagged Au@AgNRs are further functionalized with rabbit anti-human HER2 antibodies, which specifically targets a biomarker in breast cancer cells, and the final probe offers LSPR positions around 430 nm, 500 – 550 nm (shoulder peak), and 690 nm. EVs from a breast cancer cell line (SKBR3) and normal lung fibroblasts (MRC5) were used as models for cancerous and normal cells, respectively, and probed with a 632.8 nm laser. The characteristic peak for DTNB is located at 1327 cm<sup>-1</sup>, which corresponds to the nitro group symmetric stretching, and was found to decrease with decreasing EV concentration, while the controls were associated with low intensity, and the detection limit was estimated to be 1200 EVs. In fact, the DTNB signal intensity from the cancer-derived EVs was approximately 2.5 times stronger than the normal EVs, and even 3.8 times stronger when cell numbers were normalized, since cancerous cells tend to secrete more EVs than their normal counterparts.

Another sandwich-type immunocomplex was proposed by Wang et al., which consists of MBs covered by a silica layer and gold shell and modified by aptamers of CD63 via Au-S bonds for EV capture (**Figure 2-11A**).<sup>[120]</sup> Three different SERS probes based on AuNPs were implemented with varying aptamers, each targeting a specific cancer type, and Raman reporters for multiplexed detection. For the detection of breast cancer EVs derived from

the SKBR3 cell line, AuNPs were modified with the aptamer of HER2 (i.e., breast cancer biomarker) and Raman reporter DTNB, as in the previous paper reviewed. For the detection of colorectal cancer EVs derived from the T84 cell line, AuNPs were modified with the aptamer of CEA, which binds to a carcinoembryonic antigen overexpressed in T84 cells, and Raman reporter 2,7-mercapto-4-methylcoumarin (MMC). MMC has a characteristic peak located at  $1170\text{ cm}^{-1}$ , corresponding to the triangular and symmetric benzene ring breathing deformations including the in-plane deformations of the C(O)-O group. Lastly, for the detection of prostate cancer EVs derived from the LNCaP cell line, AuNPs were modified with the aptamer of PSMA, which binds to a prostate-specific membrane antigen, and Raman reporter 2-naphthalenethiol (2NAT), which has a characteristic peak at  $1378\text{ cm}^{-1}$  corresponding to ring stretching. Each EV type was analyzed by SERS using a  $632.8\text{ nm}$  excitation wavelength, and the detection limits for the SKBR3 EVs, T84 EVs, and LNCaP EVs were estimated to be  $32\text{ EVs }\mu\text{L}^{-1}$ ,  $73\text{ EVs }\mu\text{L}^{-1}$ , and  $203\text{ EVs }\mu\text{L}^{-1}$ , respectively. In this case, characteristic peaks were found to decrease with increasing EV concentration. The performance of the model was further validated with clinical samples (3 cancer patients and 1 healthy individual) (**Figure 2-11B**), but accuracy was not reported, and limited sample sizes were used in both cell line and clinical cases.

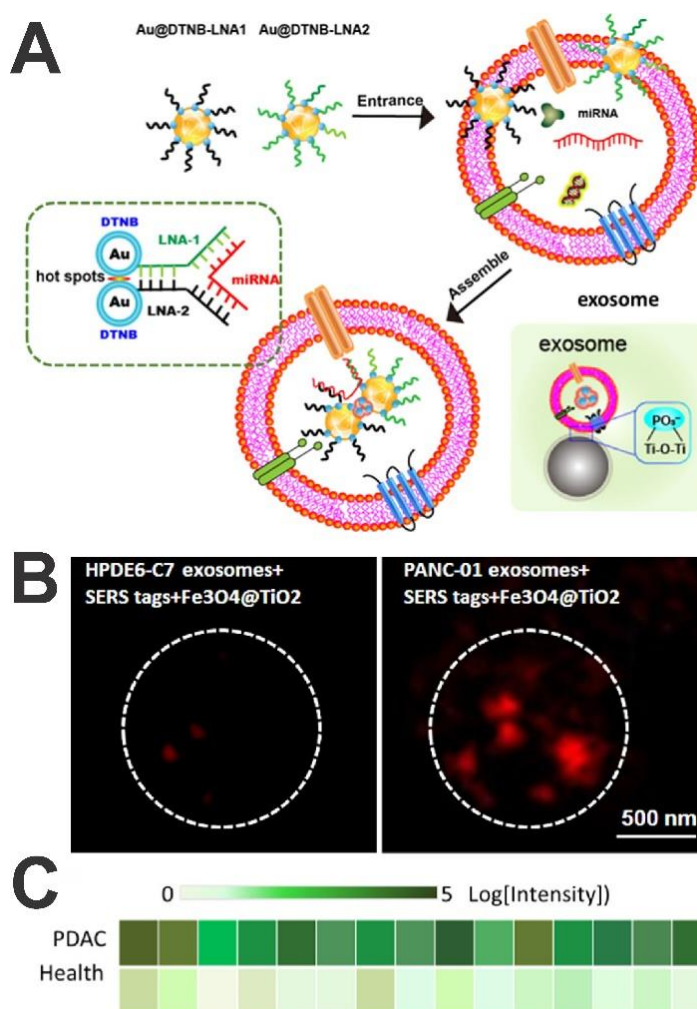


**Figure 2-11: Sandwich-type immunocomplex for the analysis of breast cancer, colorectal cancer, and prostate cancer-derived EVs. (A) Schematic illustration of the fabrication of capture probes and SERS probes, as well as the formation of the immunocomplex and acquisition of SERS spectra. (B) Simultaneous detection of breast cancer, colorectal cancer, and prostate cancer markers demonstrated using patient-derived EVs. Adapted with permission from reference [120] (copyright 2018 Royal Society of Chemistry).**

A sandwich-type immunocomplex that targets EVs via bivalent cholesterol-labelled DNA anchors has been proposed by Tian et al.<sup>[121]</sup> Herein, the capture probes consist of MBs functionalized with anti-CD9, and DNA-labelled SERS probes consist of AuNSs modified with Raman reporter 4-mercaptobenzoic acid (MBA) and covered with a gold nanoshell. The SERS probe allows for monitoring of EVs via the characteristic  $1078\text{ cm}^{-1}$  peak and offers an LSPR around 615 nm. The probes were tested with EVs derived from liver cancer cell line HepG2. The authors observed the characteristic MBA peak only in the presence of EVs, since the immunocomplex cannot form without them. The detection limit was estimated to be  $27\text{ particles }\mu\text{L}^{-1}$ , far more sensitive than commercially available methods and even other SERS techniques. Human serum samples from 3 liver cancer patients and 3 healthy individuals were also used to validate this method, and EVs from these samples were quantified using a calibration equation established by the cell line model. The EV amounts were further validated by the standardized method qNano, and the results were in agreement. Although a lower EV count was identified in the healthy patients compared to the cancer patients, a diagnostic threshold was not established, and the diagnostic potential of the biosensor was therefore not quantified.

An EV-probe assembly was detailed by Jiang et al., in which DTNB-modified AuNPs were functionalized with locked nucleic acids (LNAs) targeting miRNA-10b.<sup>[134]</sup> Rather than targeting EVs via surface interactions, the LNA-modified SERS tags are transported into EVs by incubation (**Figure 2-12A**), which the authors confirmed with TEM imaging and fluorescence measurements. These SERS probes also provided LSPR modes around 520 nm and 625 nm.  $\text{Fe}_3\text{O}_4$  core- $\text{TiO}_2$  shell MBs were used for EV capture, which can concentrate EVs with an efficiency of 96.5 % in 10 minutes with the application of an external magnet due to the high specificity binding between  $\text{TiO}_2$  and the phosphate groups on EV membranes. EVs from pancreatic ductal adenocarcinoma (PDAC) cells (PANC-01) and normal pancreatic cells (HPDE6-E7) were probed under a 785 nm excitation wavelength laser. The characteristic DTNB peak was found to be 3-fold higher in the cancer-derived EVs compared to the normal EVs (**Figure 2-12B**). The authors demonstrated high specificity by distinguishing the target miRNA-10b from the single base mismatched miRNA-10a and estimated the detection limit to be 0.21 fM. Since miRNA-10b has also been reported to be overexpressed in colorectal cancer cell EVs, the authors

further validated miRNA-10b detection with EVs derived from the colorectal cancer cell line HCT116 and the normal colonic epithelial cell line CCD841. The authors went on to monitor miRNA-10b levels in sera of 15 PDAC patients and 15 individuals and found that the characteristic peak intensity in the case of the cancer patients was 3-fold higher compared to the healthy individuals (**Figure 2-12C**). While the authors assessed the diagnostic capabilities of the assembly with an ROC curve and found the AUC to be quite high at 0.996, accuracy, sensitivity, and specificity were not reported.



**Figure 2-12: EV-probe assembly for the SERS analysis of pancreatic cancer-derived EVs. (A) Schematic illustration of the uptake of SERS probes into EVs and the affinity of Fe<sub>3</sub>O<sub>4</sub> core-TiO<sub>2</sub> shell MBs to the EV surface. (B) SERS maps of normal pancreatic cell line-derived EVs (left) and pancreatic cancer cell line-derived EVs (right), highlighting the higher SERS intensity observed when the target is present. (C)**

**Higher characteristic peak intensity is also observed with EVs derived from clinical samples. Adapted with permission from reference [134] (copyright 2021 American Chemical Society).**

#### 2.4.4 Sandwich-Type Immunoassays for Indirect Sensing

One sandwich-type immunoassay reported by Li et al. consists of a glass slide modified by the self-polymerizing polydopamine (PDA) and antibodies (e.g., anti-MIF) for EV capture.<sup>[135]</sup> PDA provides a biocompatible, uniform layer into which antibodies can be homogeneously encapsulated with high efficiency. The SERS probes utilized were multilayered, consisting of an 18 nm gold core, silver shell, Raman reporter 4-aminobenzenethiol (*p*ATP), PDA, and antibodies, and provided LSPR modes around 400 nm and 500 nm. The Raman reporter has a characteristic peak around 1072 cm<sup>-1</sup>, corresponding to the benzene ring breathing vibration, and the PDA layer served to protect the silver layer from oxidation. EVs from the pancreatic cancer cell line PANC-01 and normal cell line HPDE6-C7 were analyzed by SERS using a 785 nm excitation wavelength, and the detection limit was estimated to be 544 particles  $\mu\text{L}^{-1}$ . The biosensor was further tested with EVs derived from patient serum samples (71 PDAC patients and 32 healthy individuals) and was able to distinguish the serum of cancer patients from that of healthy donors, but with low sensitivity (62.5 %) and low specificity (76.2 %). The authors noted that the biosensor was even able to distinguish 37 patients with higher stage (P3) stage tumors from 4 patients with lower stage (P1-2) tumours. Although the sensitivity was high in this instance (95.7 %), the specificity was again found to be low (53.4 %). When the authors attempted to distinguish 10 patients with metastatic cancer from 61 patients with nonmetastatic cancer, low sensitivity and specificity were again achieved, at 53.3 % and 73.9 %, respectively. Additionally, the laser excitation wavelength at 785 nm was not well matched to the LSPR bands of the biosensor, which poses the question of whether the SERS effect was indeed observed.

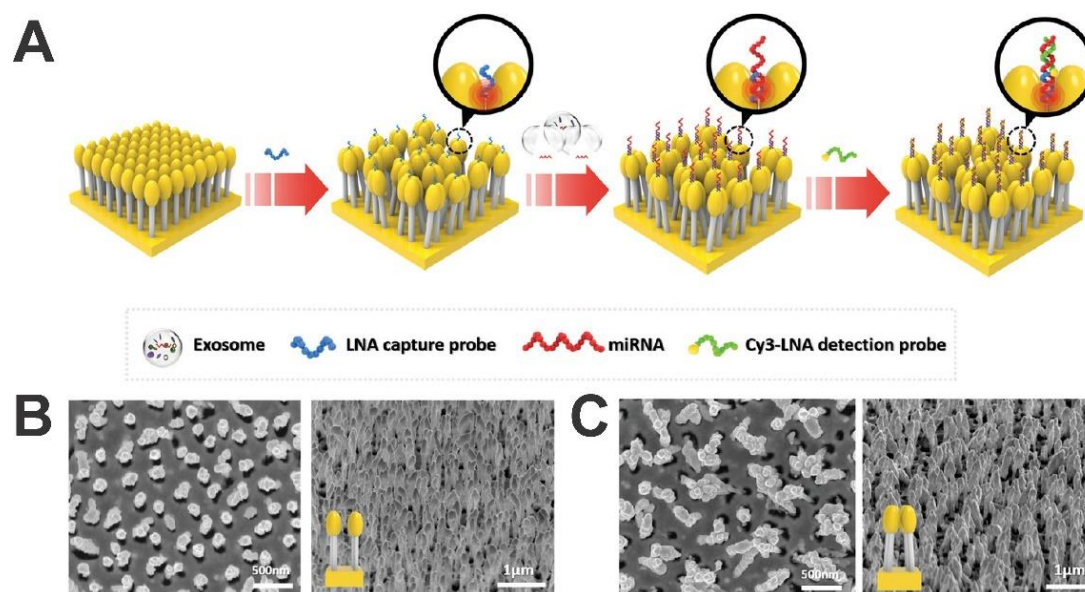
Another sandwich-type immunoassay was proposed by Kwizera et al. consisting of antibody (anti-CD63) arrays on a gold-coated slide for specific EV capture.<sup>[136]</sup> The SERS probes consisted of AuNRs of 35 nm length and 12 nm width capped with cetrimonium bromide (CTAB) and coated with an organic dye, QSY<sup>TM</sup> 21 carboxylic acid-succinimidyl

ester (QSY21), as a Raman reporter, providing an LSPR band around 720 nm. Cationic CTAB was chosen to allow for the electrostatic adsorption of the SERS probes onto the anionic lipid EV bilayer, and to provide a hydrophobic pocket into which the hydrophobic Raman reporter could be loaded. QSY21 was chosen as the Raman reporter for its nonfluorescent properties and fingerprinting signal at  $1497\text{ cm}^{-1}$ . EVs from three breast cancer cell lines (MDA-MB-231, MDA-MB-468, and SKBR3) were analyzed with the normal breast cell line MCF12A used as a control. The EVs were probed with a 785 nm excitation wavelength laser, and the detection limit was estimated to be  $2000\text{ EVs }\mu\text{L}^{-1}$ , approximately 500 times lower than the estimated concentration of EVs in plasma. MDA-MB-231 was used as the model cell line to determine protein expression in breast cancer EVs by examining levels of the epithelial marker EpCAM, breast cancer markers CD44, HER2, EGFR, and IGFR, and exosome markers CD9, CD63, and CD81. MDA-MB-231-derived EVs displayed a high expression of CD44, and low expression of EpCAM and other breast cancer markers, consistent with literature and validated with the gold standard ELISA. MDA-MB-468-derived EVs also displayed high EpCAM expression and moderate HER2 expression, while SKBR3-derived EVs displayed high EpCAM and HER2 expressions. The control EVs were also positive for EpCAM, but with a much lower level compared to the cancer-derived EVs. All EVs also displayed high exosome marker expressions. The authors further evaluated the biosensor with plasma samples from 10 breast cancer patients (2 stage I patients, 2 stage II patients, and 6 stage III patients) and 5 healthy individuals. The cancer-derived EVs showed much higher expressions of EpCAM and HER2 compared to the healthy controls, whereas the differences in CD44 among all patients were less noticeable. Both cancer-derived EVs and normal EVs additionally displayed strong exosomal marker expressions. However, accuracy, sensitivity, and specificity were not reported, and a limited data set was evaluated. The authors also noted that EV isolation from plasma was required prior to analysis.

An immunoassay targeting miRNAs via specific LNA probes was proposed by Lee et al.<sup>[137]</sup> Flocked gold nanopillars with 200 nm diameter and 800 nm height were used as a SERS substrate, fabricated by the RIE of silicon followed by electron beam evaporation of gold (**Figure 2-13A**). The nanopillars are termed flocked since the pillars lean towards each other due to a capillary force developed during solvent evaporation, improving the



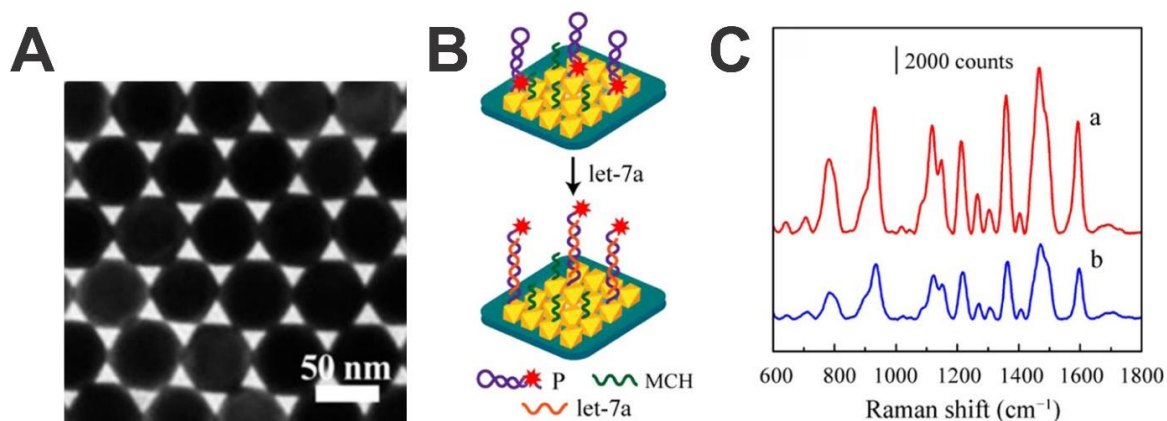
coupling of localized surface plasmons by reducing the gap between nanopillars (approximately 100 – 200 nm) (**Figure 2-13B,C**). Each LNA probe used targets miRNAs closely related to breast cancer (miRNA-21, miRNA-222, miRNA-200c). The probes were labelled with Cy3 dye, providing a characteristic peak at  $1150\text{ cm}^{-1}$ . EVs from luminal subtype (MCF7, BT474), HER2<sup>+</sup> subtype (SKBR3, AU565), and TN subtype (MDA-MB-231, HCC1143) breast cancer cell lines were analyzed by SERS using a 785 nm excitation wavelength. The probes were able to discriminate single base-mismatched miRNAs, and the detection limit was estimated to be 1 aM. By monitoring SERS signal intensity, the authors found that miRNA-21 was significantly higher in the luminal and TN subtypes compared to the HER2<sup>+</sup> subtype, and miRNA-222 and miRNA-200c were also clearly distinguishable. These findings were also validated with the standard quantitative reverse transcription-polymerase chain reaction (qRT-PCR). While the authors proposed a highly selective and sensitive platform for breast cancer diagnosis, the study lacked a normal control group, so the diagnostic capability could not be assessed.



**Figure 2-13: Flocked gold nanopillars for the SERS analysis of breast cancer-derived EVs. (A) Schematic illustration of the fabrication and functionalization of the nanopillars. (B) Top-view (left) and tilt-view (right) SEM images of upright nanopillars. (C) Top-view (left) and tilt-view (right) SEM images of head-flocked nanopillars. Adapted with permission from reference [137] (copyright 2019 Wiley-VCH).**

A microfluidic sandwich-type immunoassay device named the EV phenotype analyzer chip (EPAC) was introduced by Wang et al.<sup>[138]</sup> Capture substrates consisted of a series of asymmetric gold electrodes fabricated by photolithography, which allowed for nanoscopic lateral fluid flow, and were aligned with microfluidic channels fabricated by soft lithography. The substrate or “chip” was functionalized with capture antibodies anti-CD63 and anti-MSCP, although a lower detection limit could be achieved with the anti-MSCP-functionalized EPAC. Four SERS tags based on 60 nm AuNPs were implemented, each targeting a single biomarker. Raman reporters MBA and DTNB, described above, were used to target markers MCSP and ErbB3, respectively. Raman reporters 4-mercaptopyridine (MPY) and 2,3,5,6-tetrafluoro-MBA (TFMBA), which have characteristic peaks at  $1000\text{ cm}^{-1}$  and  $1375\text{ cm}^{-1}$ , respectively, targeted markers LNGFR and MCAM. EVs from melanoma cell line SK-MEL-28 were analyzed under a 632 nm excitation wavelength laser. SK-MEL-28 cells are expected to have high expressions of MCSP and MCAM and low expressions of ErbB3 and LNGFR. As a negative control, EVs from breast cancer cell line MCF7 were utilized, which are expected to have low expressions of all four markers. Indeed, SK-MEL-28 EVs demonstrated quantifiable signal while negligible signal was observed with MCF7-derived EVs. The response of EVs from SK-MEL-28 along with EVs derived from two other BRAF mutation cell lines (LM-MEL-33 and LM-MEL-64) when treated with a BRAF inhibitor drug was also monitored. Two control groups were also used, including LM-MEL-64 EVs without treatment and NRAS mutation cell line LM-MEL-35, which is not expected to respond to this type of drug treatment. After 30 days of treatment, EVs from all three BRAF mutation cell lines showed obvious phenotypic changes (quantified by LDA), while the negative control EVs displayed no phenotypic changes. The authors further challenged the biosensor with patient plasma-derived EVs (11 melanoma patients and 12 healthy individuals), and found that based on MCSP levels, melanoma patients could be differentiated from healthy controls. These findings were also validated with the standard ELISA by monitoring ErbB3 levels. Phenotypic changes of these EVs in response to drug treatment was also monitored, but in some cases, changes were not observed by SERS despite radiological imaging confirming disease progression. The authors therefore highlighted the need to further study and evaluate the choice and number of biomarkers implemented in the EPAC.

Probing DNAs immobilized onto gold octahedra with an edge length of 50 nm have been introduced by Kang et al. for the analysis of miRNA-7a in breast cancer EVs.<sup>[139]</sup> Gold octahedral NPs are able to self-assemble into a hexagonal arrangement in liquid and were transferred to a solid (silicon) substrate (**Figure 2-14A**), providing a uniform electric field distribution and an LSPR mode around 666 nm. The SERS tag consisted of a Cy5 dye-labelled oligonucleotide probe was designed into a hairpin structure and immobilized on the array via the Au-S bond. Cy5 has prominent spectral peaks at 938  $\text{cm}^{-1}$  (C-H deformation), 1120  $\text{cm}^{-1}$  (C-H in-plane bending), 1230  $\text{cm}^{-1}$  (C-N stretching), 1359  $\text{cm}^{-1}$  (methine chain deformation), 1467  $\text{cm}^{-1}$  (C=C ring stretching), and 1597  $\text{cm}^{-1}$  (C=N stretching). Additionally, Cy5 has an absorption centered around 660 nm, overlapping with the resonance of the array and leading to coresonance. When the let-7a target is present, the probe hybridizes with let-7a and opens its hairpin structure, forcing Cy5 away from the SERS-active surface and resulting in a signal intensity decrease across all Cy5 peaks (**Figure 2-14B,C**). The detection limit was estimated to be 5.3 aM, and the authors also demonstrated specificity by introducing other miRNAs to the biosensor such as miRNA-21, miRNA-375, let-7fi (4 base mismatch), let-7i (2 base mismatch), and let-7f (1 base mismatch). EVs from breast cancer cell line MCF7 and normal breast cell line MCF-10A were analyzed using an excitation wavelength of 633 nm, and the expression of let-7a in MCF7-derived EVs was found to be 3.27-fold higher compared to MCF-10A-derived EVs. The authors also monitored the response of let-7a levels to treatment with the chemotherapeutic drug 5-fluorouracil for 48 hours and found a higher intensity change in treated EVs compared to non-treated EVs, which was validated with qRT-PCR. However, at this stage, the authors have not challenged their biosensor with clinical samples.



**Figure 2-14: (A) TEM image of a gold octahedral self-assembled monolayer. (B) Schematic illustration of the sensing principle of the structure, in which the hairpin structure P opens upon hybridization with let-7a, forcing the Cy5 molecules (red star) farther away from the gold surface. (C) SERS spectra of Cy5 in the (a) absence and (b) presence of let-7a, highlighting the change in intensity when the target is present. Adapted with permission from reference [139] (copyright 2021 American Chemical Society).**

## 2.5 Summary

The focus of this chapter was to provide detailed background information of techniques utilized throughout this thesis. The principles of Raman spectroscopy and SERS were introduced in this chapter. Raman spectroscopy and SERS were then explained in the context of biological systems, reviewing expected peaks arising from biological macromolecules and explaining how spectra are typically used and analyzed in biological applications. PCA was introduced and explained in detail, since it is the dimensionality-reduction method used to process SERS spectra in this thesis. Since machine learning was used in series with PCA for EV classification in this thesis, the process of machine learning and various algorithms that can be used were also reviewed. Finally, various SERS probes and platforms that have been reported in literature were thoroughly reviewed, highlighting both recent advances in the field and limitations of current methods.

## 2.6 References

1. Raman, C. V.; Krishnan, K. S. *Nature* **1928**, *121*, 501-502.

2. Raman, C. V. A New Radiation. *Indian J. Phys.* **1928**, *2*, 387-398.
3. Atkins, P.; de Paula, J. Materials 1: Macromolecules and Aggregates. In *Physical Chemistry*, 8th ed.; W. H. Freeman and Company: New York, 2006; pp 652-696.
4. Sur, U. K. *Resonance* **2010**, *15*, 154-164.
5. Atkins, P.; de Paula, J. Molecular Spectroscopy 1: Rotational and Vibrational Spectra. In *Physical Chemistry*, 8th ed.; W. H. Freeman and Company: New York, 2006; pp 430-480.
6. Mishchenko, E. G. *Phys. Rev. B* **1999**, *59*, 14892-14895.
7. Butler, H. J.; Ashton, L.; Bird, B.; Cinque, G.; Curtis, K.; Dorney, J.; Esmonde-White, K.; Fullwood, N. J.; Gardner, B.; Martin-Hirsch, P. L., et al. *Nat. Protoc.* **2016**, *11*, 664-687.
8. Pitt, G.; Batchelder, D.; Bennett, R.; Bormett, R.; Hayward, I.; Smith, B. J.; Williams, K.; Yang, Y.; Baldwin, K.; Webster, S. *IEE P-Sci. Meas. Tech.* **2005**, *152*, 241-318.
9. Cooper, J. B.; Aust, J.; Stellman, C.; Chike, K.; Myrick, M. L.; Schwartz, R.; Longmire, M. *Spectrochim. Acta A Mol. Biomol. Spectrosc.* **1994**, *50*, 567-575.
10. Kuhar, N.; Sil, S.; Verma, T.; Umopathy, S. *RSC Adv.* **2018**, *8*, 25888-25908.
11. Bandekar, J. *Biochim. Biophys. Acta, Protein Struct. Mol. Enzymol.* **1992**, *1120*, 123-143.
12. Cai, S.; Singh, B. R. *Biophys. Chem.* **1999**, *80*, 7-20.
13. Rygula, A.; Majzner, K.; Marzec, K. M.; Kaczor, A.; Pilarczyk, M.; Baranska, M. *J. Raman Spectrosc.* **2013**, *44*, 1061-1076.
14. Williams, R. W.; Dunker, A. K. *J. Mol. Biol.* **1981**, *152*, 783-813.
15. Hamodrakas, S. J.; Asher, S. A.; Mazur, G. D.; Regier, J. C.; Kafatos, F. C. *Biochim. Biophys. Acta, Protein Struct. Mol. Enzymol.* **1982**, *703*, 216-222.

16. Maiti, N. C.; Apetri, M. M.; Zagorski, M. G.; Carey, P. R.; Anderson, V. E. *J. Am. Chem. Soc.* **2004**, *126*, 2399-2408.
17. Tuma, R. *J. Raman Spectrosc.* **2005**, *36*, 307-319.
18. Mangialardo, S.; Piccirilli, F.; Perucchi, A.; Dore, P.; Postorino, P. *J. Raman Spectrosc.* **2012**, *43*, 692-700.
19. Dong, A.; Huang, P.; Caughey, W. S. *Biochemistry* **1990**, *29*, 3303-3308.
20. Nakamura, K.; Era, S.; Ozaki, Y.; Sogami, M.; Hayashi, T.; Murakami, M. *FEBS Lett.* **1997**, *417*, 375-378.
21. Hernández, B.; Coïc, Y.-M.; Pflüger, F.; Kruglik, S. G.; Ghomi, M. *J. Raman Spectrosc.* **2016**, *47*, 210-220.
22. Takeuchi, H. *Biopolymers* **2003**, *72*, 305-317.
23. Hernández, B.; Pflüger, F.; Kruglik, S. G.; Ghomi, M. *J. Raman Spectrosc.* **2013**, *44*, 827-833.
24. Myintzu Hlaing, M.; Wood, B.; McNaughton, D.; Ying, D.; Augustin, M. A. *J. Biophotonics.* **2017**, *10*, 589-597.
25. Benevides, J. M.; Overman, S. A.; Thomas Jr, G. J. *J. Raman Spectrosc.* **2005**, *36*, 279-299.
26. Benevides, J. M.; Thomas, G. J., Jr. *Nucleic Acids Res.* **1983**, *11*, 5747-5761.
27. Czamara, K.; Majzner, K.; Pacia, M. Z.; Kochan, K.; Kaczor, A.; Baranska, M. *J. Raman Spectrosc.* **2015**, *46*, 4-20.
28. Singh, S. P.; Krishna, C. M. *Anal. Methods* **2014**, *6*, 8613-8620.
29. Münchberg, U.; Wagner, L.; Rohrer, C.; Voigt, K.; Rösch, P.; Jahreis, G.; Popp, J. *Anal. Bioanal. Chem.* **2015**, *407*, 3303-3311.

30. Wiercigroch, E.; Szafraniec, E.; Czamara, K.; Pacia, M. Z.; Majzner, K.; Kochan, K.; Kaczor, A.; Baranska, M.; Malek, K. *Spectrochim. Acta A Mol. Biomol. Spectrosc.* **2017**, *185*, 317-335.
31. Fang, Y.; Seong, N.-H.; Dlott, D. D. *Science* **2008**, *321*, 388-392.
32. Pilot, R.; Signorini, R.; Durante, C.; Orian, L.; Bhamidipati, M.; Fabris, L. *Biosensors* **2019**, *9*, 57.
33. Etchegoin, P. G.; Le Ru, E. C. *Phys. Chem. Chem. Phys.* **2008**, *10*, 6079-6089.
34. Ferhan, A. R.; Jackman, J. A.; Park, J. H.; Cho, N. J.; Kim, D. H. *Adv. Drug Deliv. Rev.* **2018**, *125*, 48-77.
35. Chin, L. K.; Son, T.; Hong, J.-S.; Liu, A.-Q.; Skog, J.; Castro, C. M.; Weissleder, R.; Lee, H.; Im, H. *ACS Nano* **2020**, *14*, 14528-14548.
36. Fleischmann, M.; Hendra, P. J.; McQuillan, A. J. *Chem. Phys. Lett.* **1974**, *26*, 163-166.
37. Albrecht, M. G.; Creighton, J. A. *J. Am. Chem. Soc.* **1977**, *99*, 5215-5217.
38. Jeanmaire, D. L.; Van Duyne, R. P. *J. Electroanal. Chem. Interf. Electrochem.* **1977**, *84*, 1-20.
39. Moskovits, M. *J. Chem. Phys.* **1978**, *69*, 4159-4161.
40. Moskovits, M. *Rev. Mod. Phys.* **1985**, *57*, 783-826.
41. Otto, A. *J. Raman Spectrosc.* **2005**, *36*, 497-509.
42. Lombardi, J. R.; Birke, R. L. *Acc. Chem. Res.* **2009**, *42*, 734-742.
43. Le Ru, E. C.; Blackie, E.; Meyer, M.; Etchegoin, P. G. *J. Phys. Chem. C* **2007**, *111*, 13794-13803.
44. Haynes, C. L.; McFarland, A. D.; Van Duyne, R. P. *Anal. Chem.* **2005**, *77*, 338 A-346 A.

45. Bantz, K. C.; Meyer, A. F.; Wittenberg, N. J.; Im, H.; Kurtuluş, Ö.; Lee, S. H.; Lindquist, N. C.; Oh, S.-H.; Haynes, C. L. *Phys. Chem. Chem. Phys.* **2011**, *13*, 11551-11567.
46. Atkins, P.; de Paula, J. Processes at Solid Surfaces. In *Physical Chemistry*, 8th ed.; W. H. Freeman and Company: New York, 2006; pp 909-958.
47. Jensen, L.; Aikens, C. M.; Schatz, G. C. *Chem. Soc. Rev.* **2008**, *37*, 1061-1073.
48. Champion, A.; Kambhampati, P. *Chem. Soc. Rev.* **1998**, *27*, 241-250.
49. Mayer, K. M.; Hafner, J. H. *Chem. Rev.* **2011**, *111*, 3828-3857.
50. Le Ru, E. C.; Etchegoin, P. G. *Chem. Phys. Lett.* **2006**, *423*, 63-66.
51. Ding, S.-Y.; You, E.-M.; Tian, Z.-Q.; Moskovits, M. *Chem. Soc. Rev.* **2017**, *46*, 4042-4076.
52. Liu, M.; Wang, Z.; Zong, S.; Zhang, R.; Zhu, D.; Xu, S.; Wang, C.; Cui, Y. *Anal. Bioanal. Chem.* **2013**, *405*, 6131-6136.
53. Quester, K.; Avalos-Borja, M.; Vilchis-Nestor, A. R.; Camacho-López, M. A.; Castro-Longoria, E. *PLoS ONE* **2013**, *8*, e77486.
54. Kunzmann, A.; Andersson, B.; Thurnherr, T.; Krug, H.; Scheynius, A.; Fadeel, B. *Biochim. Biophys. Acta, Gen. Subj.* **2011**, *1810*, 361-373.
55. Bondarenko, O.; Juganson, K.; Ivask, A.; Kasemets, K.; Mortimer, M.; Kahru, A. *Arch. Toxicol.* **2013**, *87*, 1181-1200.
56. Chan, G. H.; Zhao, J.; Hicks, E. M.; Schatz, G. C.; Van Duyne, R. P. *Nano Lett.* **2007**, *7*, 1947-1952.
57. Gérard, D.; Gray, S. K. *J. Phys. D Appl. Phys* **2014**, *48*, 184001.
58. Markin, A. V.; Markina, N. E.; Popp, J.; Cialla-May, D. *Trends Anal. Chem.* **2018**, *108*, 247-259.



59. Alessandri, I.; Lombardi, J. R. *Chem. Rev.* **2016**, *116*, 14921-14981.
60. Lombardi, J. R.; Birke, R. L. *J. Phys. Chem. C* **2014**, *118*, 11120-11130.
61. West, P. R.; Ishii, S.; Naik, G. V.; Emani, N. K.; Shalaev, V. M.; Boltasseva, A. *Laser Photonics Rev.* **2010**, *4*, 795-808.
62. Naik, G. V.; Shalaev, V. M.; Boltasseva, A. *Adv. Mater.* **2013**, *25*, 3264-3294.
63. Mahmoudi, M.; Lohse, S. E.; Murphy, C. J.; Fathizadeh, A.; Montazeri, A.; Suslick, K. *S. Nano Lett.* **2014**, *14*, 6-12.
64. Wang, Z.; Chen, C.; Wu, K.; Chong, H.; Ye, H. *Phys. Status Solidi A* **2019**, *216*, 1700794.
65. Patsalas, P.; Kalfagiannis, N.; Kassavetis, S. *Materials* **2015**, *8*, 3128-3154.
66. Juneja, S.; Shishodia, M. S. *Opt. Commun.* **2019**, *433*, 89-96.
67. Ling, X.; Xie, L.; Fang, Y.; Xu, H.; Zhang, H.; Kong, J.; Dresselhaus, M. S.; Zhang, J.; Liu, Z. *Nano Lett.* **2010**, *10*, 553-561.
68. Silver, A.; Kitadai, H.; Liu, H.; Granzier-Nakajima, T.; Terrones, M.; Ling, X.; Huang, S. *Nanomaterials* **2019**, *9*, 516.
69. He, X.; Liu, Y.; Huang, S.; Liu, Y.; Pu, X.; Xu, T. *RSC Adv.* **2018**, *8*, 23348-23352.
70. Beattie, J. R.; Esmonde-White, F. W. L. *Appl. Spectrosc.* **2021**, *75*, 361-375.
71. Liu, W.; Sun, Z.; Chen, J.; Jing, C. *J. Spectrosc.* **2016**, *2016*, 1603609.
72. Song, D.; Yu, F.; Chen, S.; Chen, Y.; He, Q.; Zhang, Z.; Zhang, J.; Wang, S. *Biomed. Opt. Express* **2020**, *11*, 1061-1072.
73. Howley, T.; Madden, M. G.; O'Connell, M.-L.; Ryder, A. G. The Effect of Principal *Knowl. Based Syst.* **2006**, *19*, 363-370.

74. Das, B.; Manohara, K. K.; Mahajan, G. R.; Sahoo, R. N. *Spectrochim. Acta A Mol. Biomol. Spectrosc.* **2020**, *229*, 117983.
75. Ricciardi, C.; Valente, A. S.; Edmund, K.; Cantoni, V.; Green, R.; Fiorillo, A.; Picone, I.; Santini, S.; Cesarelli, M. *Health Informatics J.* **2020**, *26*, 2181-2192.
76. Abdi, H.; Williams, L. J. *Wiley Interdiscip. Rev. Comput. Stat.* **2010**, *2*, 433-459.
77. Jolliffe, I. T.; Cadima, J. *Philos. Trans. Royal Soc. A* **2016**, *374*, 20150202.
78. Boedeker, P.; Kearns, N. T. *Adv. Meth. Pract. Psychol. Sci.* **2019**, *2*, 250-263.
79. Bell, J. What Is Machine Learning? In *Machine Learning: Hands-on for Developers and Technical Professionals*, 2nd ed.; Andresen, J., Ed. Wiley & Sons, Inc.: Indianapolis, 2020; pp 1-14.
80. Rundo, F.; Trenta, F.; di Stallo, A. L.; Battiato, S. *Appl. Sci.* **2019**, *9*, 5574.
81. Warin, T.; Stojkov, A. *J. Risk Financial Manag.* **2021**, *14*, 302.
82. Hajek, P.; Henriques, R. *Knowl. Based Syst.* **2017**, *128*, 139-152.
83. Carcillo, F.; Dal Pozzolo, A.; Le Borgne, Y.-A.; Caelen, O.; Mazzer, Y.; Bontempi, G. *Inf. Fusion* **2018**, *41*, 182-194.
84. Nagarajan, G.; Thyagarajan, K. K. *Procedia Eng.* **2012**, *38*, 2164-2171.
85. Matošević, G.; Dobša, J.; Mladenčić, D. *Future Internet* **2021**, *13*, 9.
86. Brei, V. A. *Found. Trends Mark.* **2020**, *14*, 173-236.
87. Hair, J. F.; Sarstedt, M. *J. Mark. Theory Pract.* **2021**, *29*, 65-77.
88. Erickson, B. J.; Korfiatis, P.; Akkus, Z.; Kline, T. L. *RadioGraphics* **2017**, *37*, 505-515.
89. Richens, J. G.; Lee, C. M.; Johri, S. *Nat. Commun.* **2020**, *11*, 3923.

90. Kotsiantis, S. *Informatica* **2007**, *31*, 249-268.
91. Vabalas, A.; Gowen, E.; Poliakoff, E.; Casson, A. J. *PLoS ONE* **2019**, *14*, e0224365.
92. Murthy, S. K. *Data Min. Knowl. Discov.* **1998**, *2*, 345-389.
93. Bell, J. Working with Decision Trees. In *Machine Learning: Hands-on for Developers and Technical Professionals*, 2nd ed.; Andresen, J., Ed. Wiley & Sons, Inc.: Indianapolis, 2020; pp 81-102.
94. Fürnkranz, J. *Artif. Intell. Rev.* **1999**, *13*, 3-54.
95. Clark, P.; Niblett, T. *Mach. Learn.* **1989**, *3*, 261-283.
96. Bell, J. Artificial Neural Networks. In *Machine Learning: Hands-on for Developers and Technical Professionals*, 2nd ed.; Andresen, J., Ed. Wiley & Sons, Inc.: Indianapolis, 2020; pp 165-196.
97. Svozil, D.; Kvasnicka, V.; Pospichal, J. *Chemometr. Intell. Lab Syst.* **1997**, *39*, 43-62.
98. Friedman, J. H. *J. Am. Stat. Assoc.* **1989**, *84*, 165-175.
99. Heckerman, D. A Tutorial on Learning with Bayesian Networks. In *Innovations in Bayesian Networks: Theory and Applications*, 1st ed.; Holmes, D. E.; Jain, L. C., Eds. Springer: Berlin, Heidelberg, 2008; pp 33-82.
100. Cheng, J.; Greiner, R.; Kelly, J.; Bell, D.; Liu, W. *Artif. Intell.* **2002**, *137*, 43-90.
101. Dreiseitl, S.; Ohno-Machado, L. *J. Biomed. Inform.* **2002**, *35*, 352-359.
102. Aha, D. W.; Kibler, D.; Albert, M. K. *Mach. Learn.* **1991**, *6*, 37-66.
103. Zhang, Z. *Ann. Transl. Med.* **2016**, *4*, 218-218.

104. Bell, J. Support Vector Machines. In *Machine Learning: Hands-on for Developers and Technical Professionals*, 2nd ed.; Andresen, J., Ed. Wiley & Sons, Inc.: Indianapolis, 2020; pp 143-164.
105. Zou, K. H.; O'Malley, A. J.; Mauri, L. *Circulation* **2007**, *115*, 654-657.
106. Hanley, J. A.; McNeil, B. J. *Radiology* **1982**, *143*, 29-36.
107. Bonifacio, A.; Cervo, S.; Sergo, V. *Anal. Bioanal. Chem.* **2015**, *407*, 8265-8277.
108. Garcia-Rico, E.; Alvarez-Puebla, R. A.; Guerrini, L. *Chem. Soc. Rev.* **2018**, *47*, 4909-4923.
109. Fabris, L. *J. Opt.* **2015**, *17*, 114002.
110. Fabris, L. *ChemNanoMat* **2016**, *2*, 249-258.
111. Park, J.; Hwang, M.; Choi, B.; Jeong, H.; Jung, J.-H.; Kim, H. K.; Hong, S.; Park, J.-H.; Choi, Y. *Anal. Chem.* **2017**, *89*, 6695-6701.
112. Shin, H.; Jeong, H.; Park, J.; Hong, S.; Choi, Y. *ACS Sens.* **2018**, *3*, 2637-2643.
113. Shin, H.; Oh, S.; Hong, S.; Kang, M.; Kang, D.; Ji, Y.-G.; Choi, B. H.; Kang, K.-W.; Jeong, H.; Park, Y., et al. *ACS Nano* **2020**, *14*, 5435-5444.
114. Carmicheal, J.; Hayashi, C.; Huang, X.; Liu, L.; Lu, Y.; Krasnoslobodtsev, A.; Lushnikov, A.; Kshirsagar, P. G.; Patel, A.; Jain, M., et al. *Nanomedicine* **2019**, *16*, 88-96.
115. Chalapathi, D.; Padmanabhan, S.; Manjithaya, R.; Narayana, C. *J. Phys. Chem. B* **2020**, *124*, 10952-10960.
116. Stremersch, S.; Marro, M.; Pinchasik, B.-E.; Baatsen, P.; Hendrix, A.; De Smedt, S. C.; Loza-Alvarez, P.; Skirtach, A. G.; Raemdonck, K.; Braeckmans, K. *Small* **2016**, *12*, 3292-3301.

117. Fraire, J. C.; Stremersch, S.; Bouckaert, D.; Monteyne, T.; De Beer, T.; Wuytens, P.; De Rycke, R.; Skirtach, A. G.; Raemdonck, K.; De Smedt, S., et al. *ACS Appl. Mater. Interfaces* **2019**, *11*, 39424-39435.
118. Li, G.; Zhu, N.; Zhou, J.; Kang, K.; Zhou, X.; Ying, B.; Yi, Q.; Wu, Y. *J. Mater. Chem. B* **2021**, *9*, 2709-2716.
119. Zong, S.; Wang, L.; Chen, C.; Lu, J.; Zhu, D.; Zhang, Y.; Wang, Z.; Cui, Y. *Anal. Methods* **2016**, *8*, 5001-5008.
120. Wang, Z.; Zong, S.; Wang, Y.; Li, N.; Li, L.; Lu, J.; Wang, Z.; Chen, B.; Cui, Y. *Nanoscale* **2018**, *10*, 9053-9062.
121. Tian, Y.-F.; Ning, C.-F.; He, F.; Yin, B.-C.; Ye, B.-C. *Analyst* **2018**, *143*, 4915-4922.
122. Ferreira, N.; Marques, A.; Águas, H.; Bandarenka, H.; Martins, R.; Bodo, C.; Costa-Silva, B.; Fortunato, E. *ACS Sens.* **2019**, *4*, 2073-2083.
123. Pramanik, A.; Mayer, J.; Patibandla, S.; Gates, K.; Gao, Y.; Davis, D.; Seshadri, R.; Ray, P. C. *ACS Omega* **2020**, *5*, 16602-16611.
124. Tirinato, L.; Gentile, F.; Di Mascolo, D.; Coluccio, M. L.; Das, G.; Liberale, C.; Pullano, S. A.; Perozziello, G.; Francardi, M.; Accardo, A., et al. *Microelectron. Eng.* **2012**, *97*, 337-340.
125. Sivashanmugan, K.; Huang, W.-L.; Lin, C.-H.; Liao, J.-D.; Lin, C.-C.; Su, W.-C.; Wen, T.-C. *J. Taiwan Inst. Chem. Eng.* **2017**, *80*, 149-155.
126. Kim, N.; Thomas, M. R.; Bergholt, M. S.; Pence, I. J.; Seong, H.; Charchar, P.; Todorova, N.; Nagelkerke, A.; Belessiotis-Richards, A.; Payne, D. J., et al. *Nat. Commun.* **2020**, *11*, 207.
127. Yan, Z.; Dutta, S.; Liu, Z.; Yu, X.; Mesgarzadeh, N.; Ji, F.; Bitan, G.; Xie, Y.-H. *ACS Sens.* **2019**, *4*, 488-497.

128. Jalali, M.; Isaac Hosseini, I.; AbdelFatah, T.; Montermini, L.; Wachsmann Hogiu, S.; Rak, J.; Mahshid, S. *Lab Chip* **2021**, *21*, 855-866.
129. Avella-Oliver, M.; Puchades, R.; Wachsmann-Hogiu, S.; Maquieira, A. *Sens. Actuators B Chem.* **2017**, *252*, 657-662.
130. Dong, S.; Wang, Y.; Liu, Z.; Zhang, W.; Yi, K.; Zhang, X.; Zhang, X.; Jiang, C.; Yang, S.; Wang, F., et al. *ACS Appl. Mater. Interfaces* **2020**, *12*, 5136-5146.
131. Lee, C.; Carney, R. P.; Hazari, S.; Smith, Z. J.; Knudson, A.; Robertson, C. S.; Lam, K. S.; Wachsmann-Hogiu, S. *Nanoscale* **2015**, *7*, 9290-9297.
132. Kaufman, L.; Cooper, T.; Wallace, G.; Hawke, D.; Betts, D.; Hess, D.; Lagugné-Labarthe, F. Trapping and SERS Identification of Extracellular Vesicles Using Nanohole Arrays. In *Proc. SPIE 10894, Plasmonics in Biology and Medicine XVI*, 108940B: 2019.
133. Culum, N. M.; Cooper, T. T.; Bell, G. I.; Hess, D. A.; Lagugné-Labarthe, F. *Anal. Bioanal. Chem.* **2021**, *413*, 5013-5024.
134. Jiang, S.; Li, Q.; Wang, C.; Pang, Y.; Sun, Z.; Xiao, R. *ACS Sens.* **2021**, *6*, 852-862.
135. Li, T.-D.; Zhang, R.; Chen, H.; Huang, Z.-P.; Ye, X.; Wang, H.; Deng, A.-M.; Kong, J.-L. *Chem. Sci.* **2018**, *9*, 5372-5382.
136. Kwizera, E. A.; O'Connor, R.; Vinduska, V.; Williams, M.; Butch, E. R.; Snyder, S. E.; Chen, X.; Huang, X. *Theranostics* **2018**, *8*, 2722-2738.
137. Lee, J. U.; Kim, W. H.; Lee, H. S.; Park, K. H.; Sim, S. J. *Small* **2019**, *15*, 1804968.
138. Wang, J.; Wuethrich, A.; Sina, A. A. I.; Lane, R. E.; Lin, L. L.; Wang, Y.; Cebon, J.; Behren, A.; Trau, M. *Sci. Adv.* **2020**, *6*, eaax3223.
139. Kang, T.; Zhu, J.; Luo, X.; Jia, W.; Wu, P.; Cai, C. *Anal. Chem.* **2021**, *93*, 2519-2526.

## Chapter 3

### 3 Characterization of Extracellular Vesicles Derived from Mesenchymal Stromal Cells by Surface-Enhanced Raman Spectroscopy

A version of this work has been published in the journal *Analytical and Bioanalytical Chemistry*: Culum, N. M.; Cooper, T. T.; Bell, G. I.; Hess, D. A.; Laguné-Labarthe, F. Characterization of Extracellular Vesicles Derived from Mesenchymal Stromal Cells by Surface Enhanced Raman Spectroscopy. *Anal. Bioanal. Chem.* **2021**, *413*, 5013-5024. This chapter has been reproduced with permission from Springer Nature.

In this chapter, EVs released by MSCs derived from bone marrow (BM-MSC) and pancreatic tissue (Panc-MSC) are characterized by SERS. The SERS platform fabricated by electron-beam lithography consists of gold nanohole arrays of varying size (100 – 1000 nm) and shape (triangles, circles, and squares). While BM-MSC EVs have been previously investigated by conventional Raman spectroscopy, they have not to date been characterized by SERS. Panc-MSC EVs have indeed been characterized by SERS in the past, although this study looked into a larger sample size. Discussed in this chapter are the main compositional differences between BM- and Panc-MSC EVs determined by PCA. Machine learning was implemented in order to discriminate the two groups from each other and demonstrate the classification power of the platform.

#### 3.1 Introduction

An overview of MSC-derived EVs and their applications was provided in Section 1.2.1. Briefly, MSCs have been shown to induce cellular changes in nearby cells through the release of chemical messengers, known as paracrine signaling, particularly via their secreted EVs.<sup>[1]</sup> EVs released from MSCs are potent cell-free regenerative and restorative agents that are effective in neural,<sup>[2-4]</sup> myocardial,<sup>[5-7]</sup> hepatic,<sup>[8,9]</sup> renal,<sup>[10-12]</sup> cutaneous,<sup>[13-15]</sup> skeletal,<sup>[16,17]</sup> cartilage,<sup>[18,19]</sup> and muscular regeneration.<sup>[20,21]</sup> In particular, BM-MSCs have been widely studied due to the regenerative potential of their secreted EVs. For example, BM-MSC EVs have been shown to reduce neuroinflammation in traumatic brain injuries,<sup>[22]</sup> promote survival of retinal ganglion cells and the regeneration of their axons,<sup>[23]</sup>

suppress inflammation response in acute myocardial infarction,<sup>[24]</sup> and promote the proliferation of cisplatin-damaged proximal tubular epithelial cells.<sup>[25]</sup> However, to be used in therapeutic and regenerative practices, sensitive and reproducible characterization protocols must be established. The characterization of EVs is challenged by their nanoscale size (30 – 150 nm for exosomes and 100 – 1000 nm for microvesicles) and heterogeneity in terms of size range, morphology, molecular composition, and biogenic mechanisms.<sup>[26-28]</sup>

While SERS has been extensively used in the characterization of cancer-derived EVs, its application in the characterization of MSC EVs has been largely underexplored. Although BM-MSC EVs have been characterized by Raman spectroscopy, none to date to our knowledge have been characterized using SERS.<sup>[29,30]</sup> Gualerzi et al. conducted a Raman spectroscopy study comparing EVs isolated from BM-MSCs, adipose tissue MSCs, as well as EVs released from dermal fibroblasts.<sup>[29]</sup> The authors reported decent discriminatory power using PCA-LDA with 93.7 % accuracy, 88.6 % sensitivity, and 95.1 % specificity. However, the Raman spectra reported in this work were of low resolution and quality, and required baseline correction to remove a fluorescence background, and were bulk measurements of EVs, as opposed to single or near-single EV detection. The subsequent study published compared Raman spectra of BM-MSC EVs and EVs released from human liver stem cells.<sup>[30]</sup> Although the main purpose of this study was to assess EV purity from various isolation protocols (i.e., ultracentrifugation and size-exclusion chromatography) by PCA-LDA, the authors reported that both EV sources could be discriminated from non-EV cell fractions with 97 % accuracy. The main limitation of this study again was the low-quality Raman spectra reported due to the fluorescent nature of biological samples.

In this proof-of-concept study, we have investigated EVs derived from BM-MSCs as well as Panc-MSCs by SERS. Previous preliminary work reported from our group has demonstrated the feasibility of using these platforms for SERS characterization of Panc-MSC EVs, but a small sample size was reported.<sup>[31]</sup> Herein, we have built on these concepts and further explored the capacity of these nanohole arrays to trap, detect, and differentiate EVs from these two sources. We have fabricated plasmonically active gold nanohole arrays of varying size (100 – 1000 nm) and shape (circles, squares, and triangles) by electron-



beam lithography (EBL) that are capable of EV trapping and signal enhancement for SERS. Reported in this chapter are spectral fingerprints associated with both EV sources. The main compositional differences between Panc-MSC and BM-MSC EVs were determined by PCA, and machine learning was further employed to differentiate the two groups with high accuracy, sensitivity, and specificity.

## 3.2 Methods

### 3.2.1 Nanohole Array Fabrication by EBL

Nanohole arrays were fabricated using the protocol established by Kaufman et al.<sup>[31]</sup> A negative-tone resist, ma-N 2405 (Microchem), was spin-coated onto reactive O<sub>2</sub>-cleaned glass coverslips at 3000 rpm for 45 seconds, corresponding to a thickness of approximately  $500 \pm 50$  nm, then baked at 90 °C for 90 seconds. AquaSAVE™ conductive polymer (Sigma-Aldrich) was then spin coated at 1000 rpm for 45 seconds and baked at 90 °C for 45 seconds. EBL and SEM imaging were performed using a LEO 1530 scanning electron microscope (Zeiss) with a 30.0 kV voltage, 10.0 μm aperture, and 30.0 – 50.0 pA current. Arrays of varying shape (square, circle, and triangle) and size (0.1 – 1.0 μm in 0.1 μm increments, 1.0 μm width between holes) were written using ELPHY Quantum software (Raith Nanofabrication). All patches measured  $50 \times 50 \mu\text{m}^2$ . Following beam exposure, substrates were soaked in DI water to remove the conductive layer. Samples were developed in MF-319 (MicroChem) for 40 seconds, soaked in DI water, and air dried to avoid collapsing the nanopillars. Samples were subjected to an O<sub>2</sub> plasma descum process (Trion Technology) for 60 seconds to remove residual resist surrounding the nanopillars. A 3 nm adhesion layer of titanium was then deposited onto the samples followed by 30 nm of gold by electron beam evaporation (Angstrom Engineering). For lift-off, samples were exposed to Remover-PG (MicroChem) and heated to 80 °C for 2 hours. Remover-PG was removed from the samples by soaking in a 1:3 solution of methyl isobutyl ketone (MIBK) and isopropanol (IPA). Samples were then rinsed with IPA and dried under N<sub>2</sub>. For final cleaning, samples were immersed in Nano-Strip® (Cyantek) and heated to 80 °C for 30 minutes, then DI water for 15 minutes, and dried under N<sub>2</sub>. Finally, samples were again subjected to O<sub>2</sub> plasma for 5 minutes to remove any remaining resist from the holes.

### 3.2.2 Absorption Measurements

Vis-NIR spectra of gold nanohole arrays were obtained with a homebuilt setup consisting of an HL-2000 halogen lamp (Ocean Optics), which covers a spectral range of about 400 – 1000 nm, coupled to an inverted optical microscope by a 100  $\mu\text{m}$  optical fiber. The source beam was first expanded by a 10  $\times$  objective (N.A. = 0.25), recollimated using 20  $\times$  objective (N.A. = 0.40), and finally collected by a 20  $\times$  objective (N.A. = 0.50). The resulting spot sizes were approximately 50  $\mu\text{m}$  in diameter, covering the surface of a single array. Scattered light was then analyzed with a USB 4000-Vis-NIR-ES spectrometer (Ocean Optics).

### 3.2.3 Cell Culture

Human bone marrow aspirates were obtained from healthy donors with informed consent from the London Health Sciences Centre, Western University (London, ON) following protocol REB#12934. BM-MSCs were established and cultured in AmnioMax-C100<sup>TM</sup> media with AmnioMax<sup>TM</sup> C100 supplement (Life Technologies) as previously described by Sherman et al.<sup>[32]</sup> Ricordi-chamber isolated human islets were obtained through the Integrated Islet Distribution Program (USA) for the establishment of Panc-MSCs as previously described by Cooper et al.<sup>[33]</sup> 200 islet equivalents were plated in RPMI 1640 + 10% fetal bovine serum (FBS) for up to 7 days. Between 5 – 7 days, adherent fibroblast-like cells were separated from non-adherent islets by media aspiration followed by trypsinization and filtration using a 40  $\mu\text{m}$  cell strainer. Single cell suspensions were subsequently reseeded on tissue culture plastic at 4000 cells/cm<sup>2</sup> and expanded in Amniomax-C100<sup>TM</sup> with AmnioMax<sup>TM</sup> E100 supplement (Life Technologies).

### 3.2.4 EV Isolation

EVs were isolated by ultrafiltration as previously described by Cooper et al.<sup>[34]</sup> Conditioned media (CM) was generated by culturing BM-MSC and Panc-MSC to ~80% confluency, rinsed 3 times with pre-warmed phosphate buffer solution (PBS), and switched to basal AmnioMax<sup>TM</sup> C100 media (Life Technologies) without supplement. Media was collected after 24 hours of cell culture. Cell debris were removed by centrifugation for 10 minutes at 600  $\times$  g. Cell-free CM was concentrated by centrifugation in 100 kDa centrifuge filter units

for 20 minutes at  $2800 \times g$ . A 20 mL solution was concentrated in a single unit, requiring two centrifugations, producing a final volume of 120  $\mu\text{L}$ . After the second centrifugation, 10 mL of 0.22  $\mu\text{m}$ -filtered PBS was used to wash out residual phenols, proteins, and salts. EV samples were collected and placed into Eppendorf tubes and stored at  $-20\text{ }^{\circ}\text{C}$  for up to one month.

### 3.2.5 AFM Measurements

Purified EV solutions were diluted (1:20) in Milli-Q water. 10  $\mu\text{L}$  of the dilute EV samples were drop-casted onto chemically cleaned glass coverslips ( $22\text{ mm} \times 22\text{ mm} \times 0.15\text{ mm}$ ) and dried overnight in a biological safety cabinet. Scans were obtained using a BioScope Catalyst atomic force microscope (Bruker). NCLR-50 Silicon probes (Nanoworld) with a force constant of 48 N/m and a resonance frequency of 190 kHz were employed under tapping mode. Height images were recorded at  $256 \times 256$  pixels and a scan rate of 0.50 Hz. Imaging processing was subsequently performed using Gwyddion software.

### 3.2.6 SERS Measurements

Concentrated EV samples were diluted 1:20 with Milli-Q water. 10 – 20  $\mu\text{L}$  of dilute EV samples were drop-cast onto nanohole arrays. EV-water solution was removed from the array using cohesive properties allowed by a Kimwipe absorbent paper (Kimberly-Clark Inc.). The edge of the absorbent paper was placed on the corner of the solution droplet, allowing solution removal via capillary action. This capillary flow also induces EVs to locate and stay in the nanoholes. Lastly, EV solutions were allowed to dry for 15 – 30 minutes prior to SERS measurements. SERS spectra presented in **Figure 3-5** were acquired with a LabRAM HR spectrometer (Horiba Scientific) using a 632.8 nm excitation laser source, 600 grooves/mm grating,  $100 \times$  objective (N.A. = 0.9), and 200  $\mu\text{m}$  pinhole. Laser power was set to 2.5 mW with an acquisition time of 60 seconds per spectrum. SERS spectra presented in **Figure 3-3** and **Figure 3-6** were extracted from SERS maps that were acquired with an XploRA<sup>TM</sup> PLUS spectrometer (Horiba Scientific) using a 785 nm excitation laser source, 600 grooves/mm grating,  $100 \times$  objective (N.A. = 0.9), and 100  $\mu\text{m}$  pinhole. Laser power was set to 5 mW with an acquisition time of 4 seconds per spectrum.

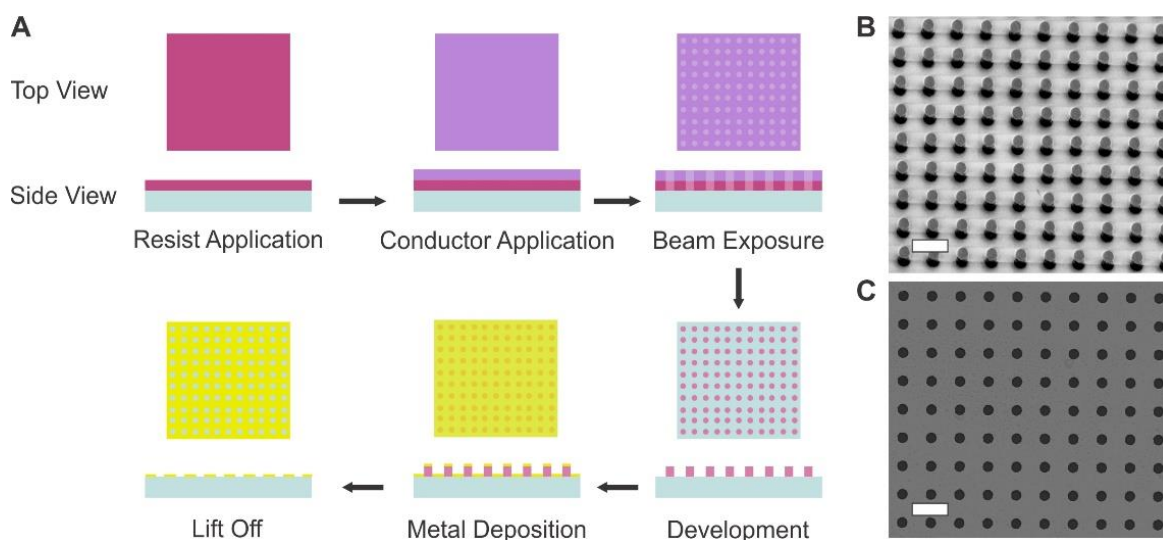
### 3.2.7 Statistical Analysis and Machine Learning

All spectra were normalized prior to PCA. The first 19 PCs were selected to explain 98% of variance among spectra. After score plots were constructed, 95% confidence ellipses were fitted around each cluster type. The first 19 PCs were then used as input data for five different machine learning algorithms: logistic regression, support vector machine, random forest, Naïve Bayes, and CN2 rule induction. In each machine learning case, models were tested using leave-one-out cross validation. PCA and machine learning were performed using Orange software (version 3.27.1).

## 3.3 Results and Discussion

### 3.3.1 Nanohole Array Fabrication and Characterization

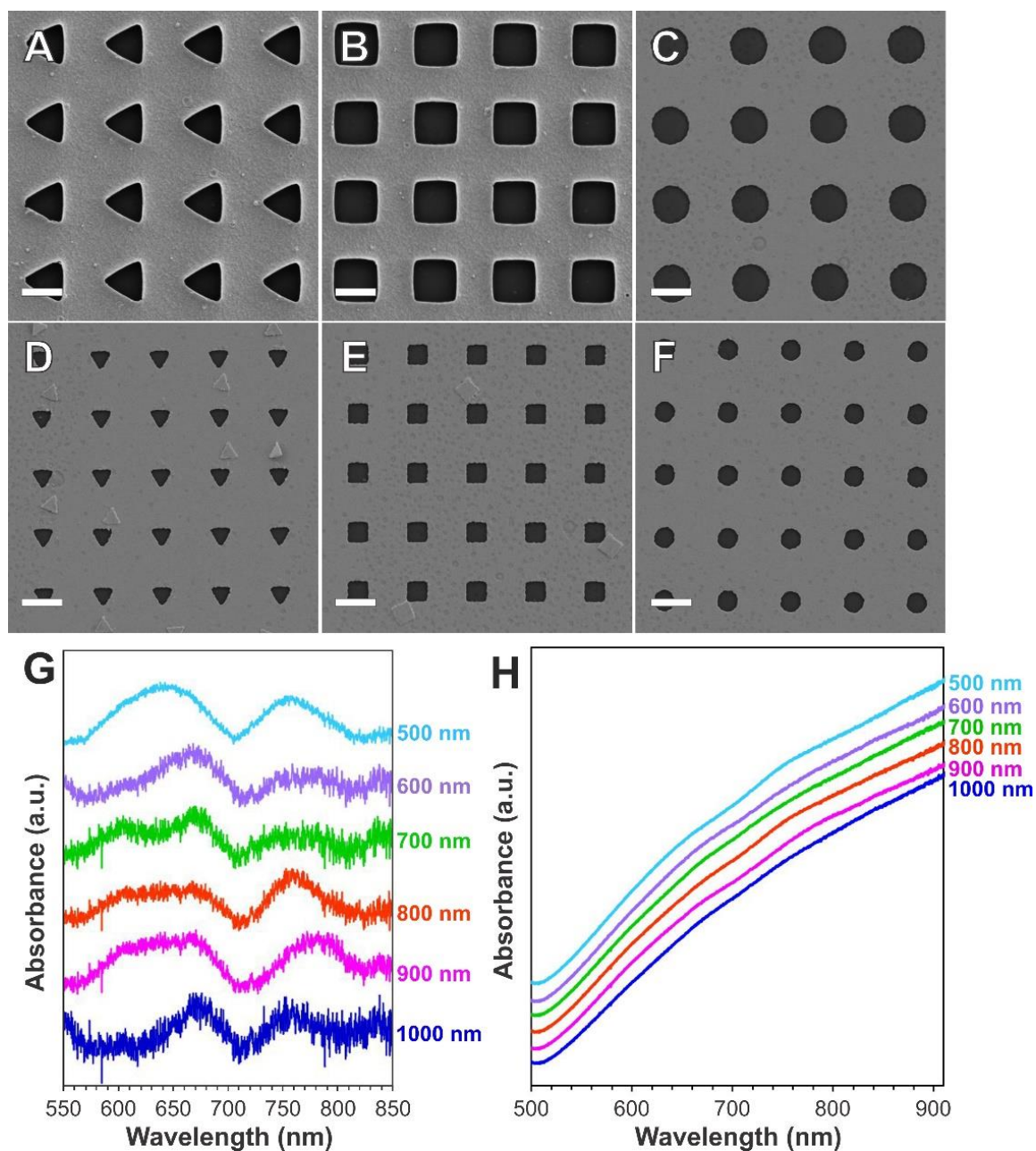
EBL is a nanofabrication technique used to create nanostructures with 20 nm spatial resolution. Nanostructures are fabricated by scanning a focused beam of electrons from an SEM onto an electron-sensitive photoresist, which undergoes chemical changes in exposed areas. The EBL nanofabrication process is illustrated in **Figure 3-1A**. Some substrates, such as the glass coverslips used here, additionally require the application of a conductive layer on top of the resist to prevent charging on the substrate surface during the inscription of the pattern, which minimizes the loss of resolution when the substrate is exposed to the electron beam. Following exposure, the substrate is developed in a chemical bath to remove some of the resist and reveal the desired pattern. For the purpose of fabricating nanohole arrays, a negative-tone resist is desired, and ma-N 2405 was used. Negative-tone resists undergo cross-linking in exposed areas, and non-exposed regions are removed during development, thus producing the reverse or “negative” image of the final pattern. For nanohole arrays, this “negative” image translates to nanopillar arrays (**Figure 3-1B**). Since the objective of the work is to use the nanohole arrays in SERS sensing, metals must be deposited onto the developed substrate for the propagation of plasmons. Gold was selected for these experiments due to its greater stability in air compared to other common SERS-active metals such as silver and copper. Finally, the metallic substrate is placed into a chemical bath to remove any remaining resist and reveal the final nanoholes in a process called lift-off (**Figure 3-1C**).



**Figure 3-1: (A) Schematic illustration of nanohole array fabrication by EBL using a negative-tone resist, with example SEM images of 700 nm circular arrays (B) before lift-off and (C) after lift-off (scale bars = 2  $\mu\text{m}$ ).**

Gold nanohole arrays of varying size (100 – 1000 nm) and shape were fabricated by EBL, imaged by SEM, and characterized by vis-NIR spectroscopy. Shapes explored for these arrays consisted of triangles (**Figure 3-2A,D**), squares (**Figure 3-2B,E**), and circles (**Figure 3-2C,F**). Fallen nanopillars or nanocaps are visible on the smaller-sized arrays (**Figure 3-2D-F**) and not on the larger-sized arrays (**Figure 3-2A-C**) since negative resists become increasingly difficult to remove as hole size decreases. Nevertheless, smaller arrays are still suitable for EV capture if the nanocaps do not collapse directly into the holes. Absorption measurements were subsequently performed to determine the position of their plasmonic resonances, as shown on the triangular arrays (**Figure 3-2G,H**). Two resonance wavelengths are observed: one around 650 – 690 nm, and another around 750 – 780 nm. Similar results were observed with the circular and square arrays, and is consistent with the results previously reported by our group.<sup>[31]</sup> When the excitation wavelength matches their resonance wavelengths (i.e., 633 and 785 nm), the conduction electrons at the metal surface of the arrays are driven to collective oscillation at a frequency referred to as the LSPR.<sup>[35]</sup> The higher energy band around 650 – 690 nm is generally referred to as the quadrupolar resonance, while the lower energy band around 750 – 780 nm is referred

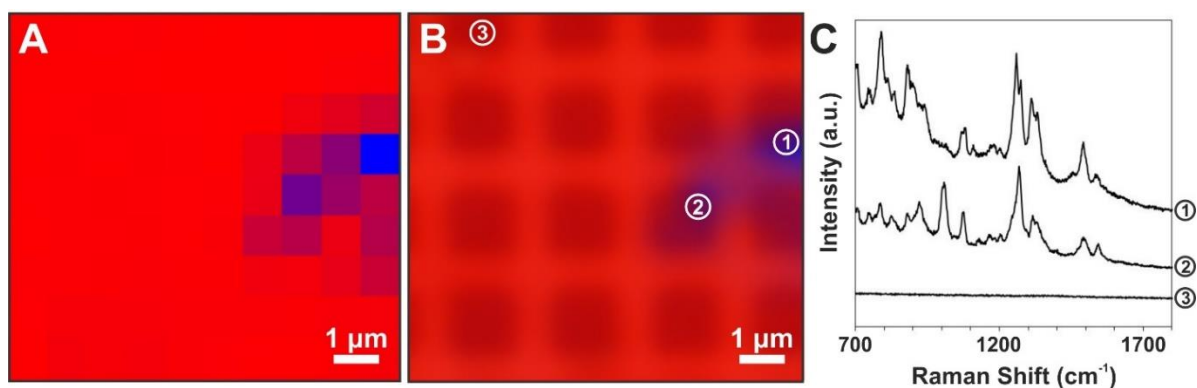
to as the dipolar resonance. Consequently, large enhancements of the local electromagnetic fields of radiation are confined at the vicinity of the nanoholes. Since the EVs are smaller in diameter compared to the hole sizes, we expect EVs to be captured by the nanoholes and their Raman signals to be enhanced as a result.



**Figure 3-2:** SEM images of 1000 nm (A) triangular, (B) square, and (C) circular arrays, and 500 nm (D) triangular, (E) square, and (F) circular arrays (scale bars = 1

**μm). (G) Background-corrected absorption spectra of 500 – 1000 nm triangular arrays and (H) corresponding raw absorption spectra.**

The trapping capabilities of these nanohole arrays have been previously published by our group, both by polystyrene beads as proof-of-concept, as well as with EVs themselves.<sup>[31]</sup> Trapping of the EVs is enabled by the flow of the EV-containing solution and the size match between the EVs and the nanohole cusps. However, trapping of small EVs cannot be observed optically due to the diffraction limited spatial resolution of our optical measurement. Therefore, blind SERS mapping over large areas of the nanohole arrays is necessary to reveal which holes are filled with one or more EVs (areas with signal) and which holes are empty (areas without signal). SERS mapping experiments showed that approximately 12 % of the holes were occupied by one or more EVs (**Figure 3-3**). For SERS of EVs, the circular arrays were the least preferential due to a lower throughput from the EBL process compared to the triangular and square arrays. This lower throughput could potentially be due to the fact that the corners of the triangles and squares provided more anchoring points for the pillars with three and four points, respectively. Conversely, the highest throughput of arrays was achieved with the square arrays, allowing more opportunity for the square arrays to capture EVs. Therefore, square arrays were generally preferred for SERS experiments. By utilizing these SERS platforms in conjunction with lasers of excitation wavelengths that match their LSPRs, we are able to study samples with inherently weak Raman signals without the need to increase laser power or accumulation time, which is likely to burn the samples, or the need to use a higher energy laser wavelength (i.e., green laser), which is likely to induce high background fluorescence.<sup>[36]</sup>

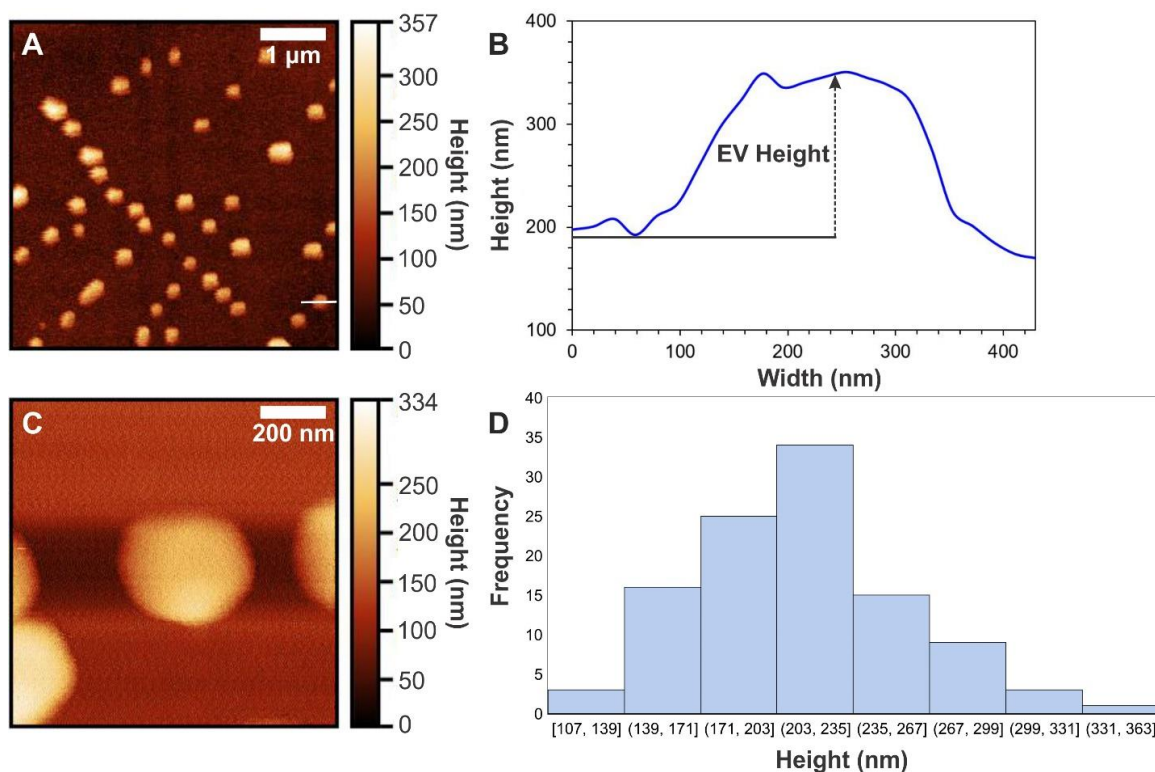


**Figure 3-3:** (A) SERS map comprised of 100 spectra of BM-MSC EVs over an area of 16 1.0  $\mu\text{m}$ -sized square nanoholes, where high intensity areas (blue) correspond to trapped EVs and low intensity areas (red) correspond to empty holes. (B) SERS map from (A) smoothed and overlaid with an optical image of the array and (C) corresponding spectra of points (1) – (3) from (B), with spectra shifted vertically for clarity.

### 3.3.2 EV Characterization

AFM was employed for EV imaging and height quantification since it causes minimal deformations to the soft EV surface when scanning in tapping mode. For AFM imaging, dilute solutions of EVs were drop-casted on clean glass coverslips and allowed to dry. AFM scans of EVs isolated from both Panc-MSC (**Figure 3-4A**) and BM-MSC (**Figure 3-4C**) revealed small, quasi-spherical objects on the substrate surface. A cross-section of one of these features is shown in **Figure 3-4B**. The average height of the adhered EVs from the Panc-MSC samples measured over 106 individual EVs was  $210 \pm 40$  nm, with the size distribution ranging from 110 – 330 nm (**Figure 3-4D**). Similarly, the average height of the adhered EVs from the BM-MSC samples was  $190 \pm 50$  nm, with a size distribution of 50 – 300 nm. Recalling that microvesicles range in diameter from 100 – 1000 nm, these distributions are well within the accepted EV size range.

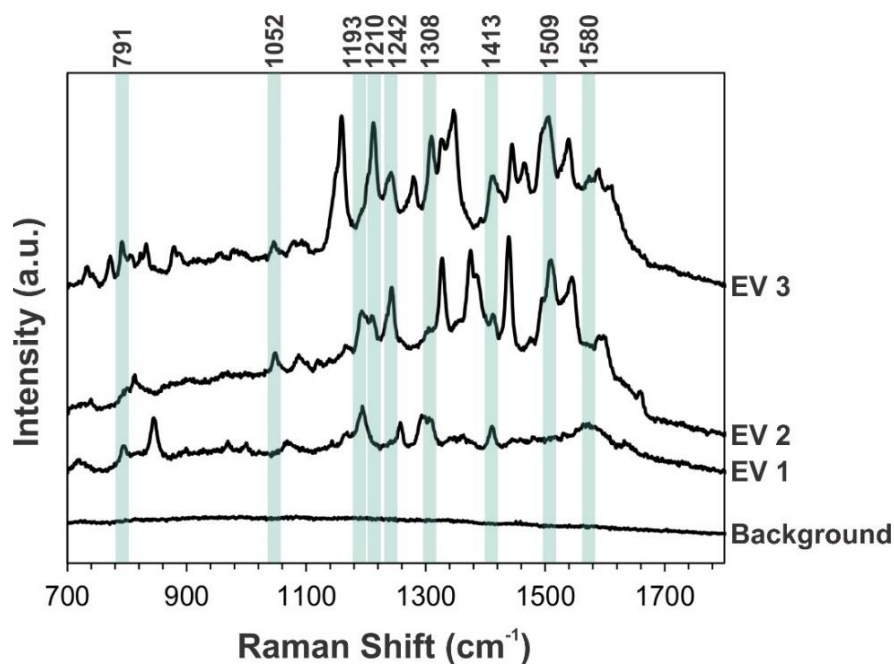




**Figure 3-4:** (A)  $(5 \times 5) \mu\text{m}^2$  AFM scan of Panc-MSC EVs on a bare glass coverslip and (B) the cross-section of a single EV, as indicated by the white line in (A). (C)  $(1 \times 1) \mu\text{m}^2$  AFM scan of an individual BM-MSC EV on a bare glass coverslip. (D) A histogram representing the height distribution of a Panc-MSC sample containing 106 EVs.

Initial SERS characterization of EVs from the Panc-MSC sample was conducted with an excitation wavelength set at 632.8 nm, which utilizes the quadrupolar resonance of the nanohole arrays. Spectra were acquired by focusing a 632.8 nm laser with a  $100\times$  objective (N.A. = 0.90) on holes presumably containing EVs, and a background spectrum was recorded by focusing the same laser off the array on flat gold. The resulting spectra of three individual EVs reveal an abundance of peaks in the  $700 - 1800 \text{ cm}^{-1}$  fingerprint region, as expected since this is a significant Raman spectral window for biological samples, whereas none are visible in the background spectrum (**Figure 3-5**). Since the LSPR decays exponentially away from the platform surface, the effective sensing zone of the plasmonic arrays is confined within the first 10 – 20 nm away from the metal surface.<sup>[37]</sup> Since plasma membranes are typically about 5 nm thick, we expect to detect not only SERS signals

corresponding to the surface content of the EVs (i.e., surface proteins and lipids), but also the SERS signals of their cargo (i.e., proteins and genetic material). The analysis of the collected spectra showed that some Raman modes are common among the three EVs, which are summarized in **Table 3-1**. Protein peaks are observed at  $1052\text{ cm}^{-1}$  and  $1242\text{ cm}^{-1}$  which can be assigned to C-O/C-N stretching and amide III, respectively. Additionally, amino acid peaks are present at  $1210\text{ cm}^{-1}$ , which is found in tyrosine and phenylalanine, and at  $1580\text{ cm}^{-1}$ , which is found in phenylalanine. Nucleic acid peaks are present at  $791\text{ cm}^{-1}$  and  $1509\text{ cm}^{-1}$ , corresponding to pyrimidines and adenine/cytosine, respectively. The peak present at  $1308\text{ cm}^{-1}$  can be attributed to the  $\text{CH}_2/\text{CH}_3$  twisting, bending, or wagging in lipids or collagen. Although these peaks are shared in a couple of spectra, there is still variety in the spectra in terms of peak positions and intensities, which can be attributed to the molecular heterogeneity of EVs.



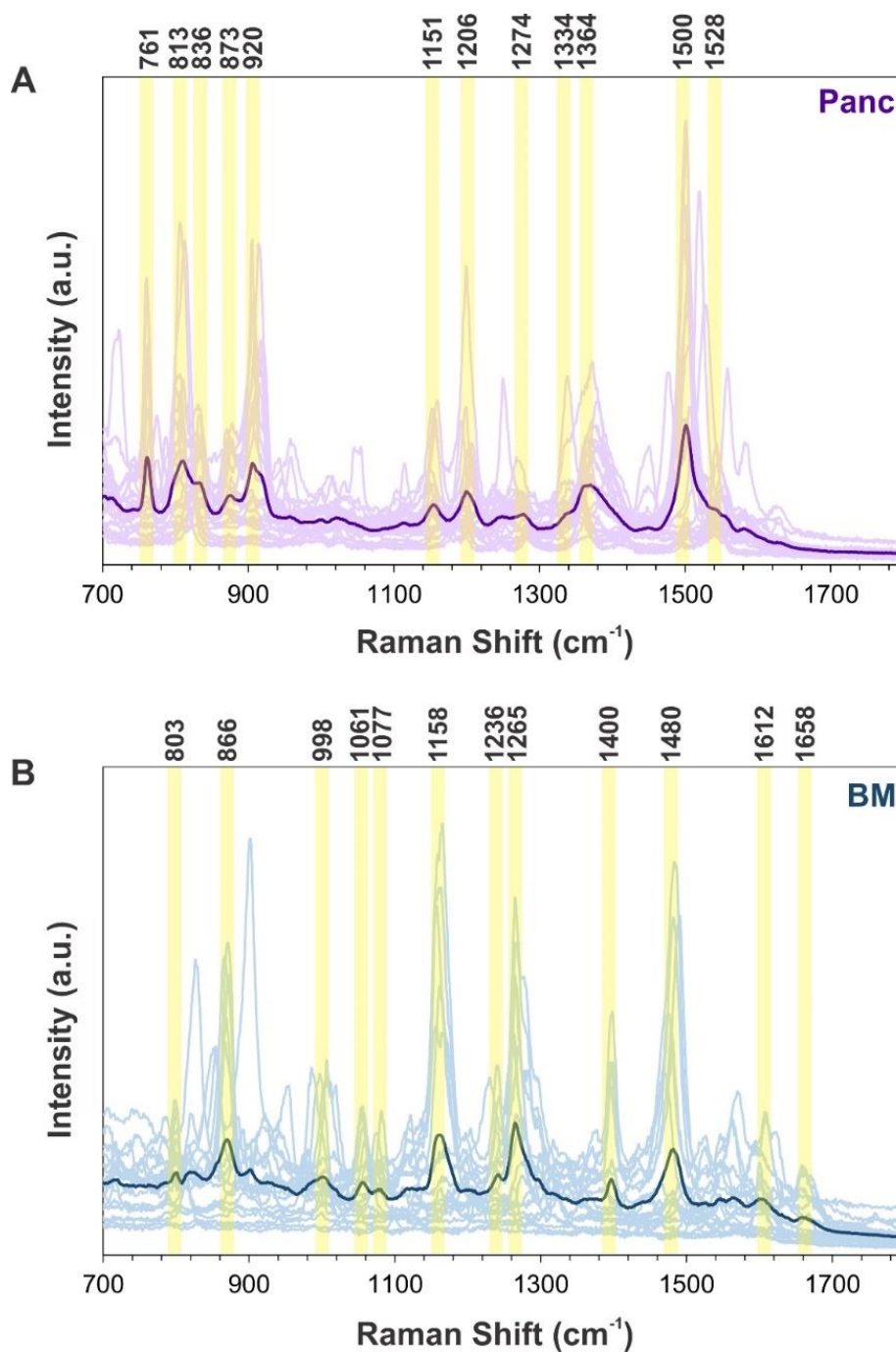
**Figure 3-5: Three SERS spectra of Panc-MSC EVs acquired with a 632.8 nm laser with a background spectrum for comparison and common peaks highlighted in blue.**

**Table 3-1: Assignments of common peaks shared among Panc-MSV EV spectra from Figure 3-5.**

Peak (cm <sup>-1</sup> )	Presumed Origin	Ref. Peak (cm <sup>-1</sup> )	Ref.
791	Pyrimidine	791	[38]
1052	C-O/C-N stretch in proteins	1053	[38]
1193	Ring stretch, CH bend, CH <sub>2</sub> twist	1194	[39]
1210	C-C <sub>6</sub> H <sub>5</sub> stretch in tyrosine and phenylalanine	1210	[38]
1242	Amide III	1243	[38,40]
1308	CH <sub>2</sub> /CH <sub>3</sub> twisting, bending, and wagging in lipids or collagen	1308	[38]
1413	Ring stretch	1412	[38]
1509	Ring breathing mode of adenine and cytosine	1510	[38]
1580	C-C stretch and C=C bend of phenylalanine	1580	[38]

While we were able to collect spectra for the Panc-MSV EVs at 632.8 nm, the spectral acquisition for the BM-MSV EVs was unsuccessful at the same wavelength due to the presence of a fluorescence background. To mitigate the background fluorescence, we switched to a 785 nm laser, thus exploiting the dipolar resonance of the nanohole arrays. Data sets for Panc-MSV and BM-MSV EVs comprised of 25 and 19 SERS spectra, respectively, were obtained with a 785 nm excitation laser source. Since EVs of this size are not visible with an optical microscope, we located EVs on the nanohole arrays by mapping areas approximately (10 × 10) μm<sup>2</sup> in size. Similarly to the Panc-MSV results obtained with the 632.8 nm laser, the spectra corresponding to Panc-MSV (**Figure 3-6A**) and BM-MSV (**Figure 3-6B**) EVs vary considerably, but common peaks within each data set can still be identified. These common peaks are summarized in **Table 3-2**. Regarding the SERS spectra of the Panc-MSV EVs (**Figure 3-6A**), protein peaks are identifiable at 813 cm<sup>-1</sup> and 1151 cm<sup>-1</sup>, corresponding to C-C and C-N stretching, respectively. Additionally, the peak at 1274 cm<sup>-1</sup> belongs to amide III. Many amino acid peaks are

located at  $761\text{ cm}^{-1}$  (tryptophan),  $873\text{ cm}^{-1}$  (hydroxyproline, tryptophan),  $1206\text{ cm}^{-1}$  (hydroxyproline, tyrosine), and  $1364\text{ cm}^{-1}$  (tryptophan). Lastly, the peak at  $1334\text{ cm}^{-1}$  can be attributed to  $\text{CH}_2/\text{CH}_3$  twisting, bending, or wagging in collagen or nucleic acids. Interestingly, there are far less peaks attributable to nucleic acids for the Panc-MSV EV spectra obtained with the 785 nm laser compared to those obtained with the 632.8 nm laser, and far more attributable to proteins and amino acids. The spectral differences between Panc-MSV EVs acquired with the 632.8 nm and 785 nm lasers could be explained by EV rupturing due to differences in laser energy, since the 632.8 nm excitation wavelength used to collect the spectra in **Figure 3-5** is higher in energy compared to the 785 nm excitation wavelength used to collect the spectra in **Figure 3-6A**. Additionally, the spectra presented in **Figure 3-5** were gathered with longer acquisition times than the spectra presented in **Figure 3-6A**, at 60 and 4 seconds, respectively. Although not specific to EVs, targeted laser-induced cell lysis, or the breakdown of the cell membrane, has been demonstrated, and a similar process could be unfolding here.<sup>[41,42]</sup> Since genetic material carried in EVs is contained by a membrane typically decorated with surface proteins, we might expect to see more protein signals compared to nucleic acid signals when the membrane is intact. Regarding the SERS spectra of the BM-MSV EV (**Figure 3-6B**), many protein peaks are also visible at  $866\text{ cm}^{-1}$ ,  $1158\text{ cm}^{-1}$ ,  $1236\text{ cm}^{-1}$ ,  $1265\text{ cm}^{-1}$ , and  $1658\text{ cm}^{-1}$ . A few nucleic acid peaks can be additionally seen at  $803\text{ cm}^{-1}$  (uracil),  $1480\text{ cm}^{-1}$  (guanine and adenine), and  $1612\text{ cm}^{-1}$  (cytosine). A lipid peak can also be found at  $1077\text{ cm}^{-1}$ , corresponding to a C-C/C-O stretch.



**Figure 3-6: SERS spectra of (A) Panc-MSC EVs and (B) BM-MSC EVs acquired with a 785 nm laser, where individual spectra are represented by lighter lines, average spectra are represented by darker lines, and common spectral peaks are highlighted in yellow.**

**Table 3-2: Assignments of common peaks shared among BM-MS and Panc-MS EV spectra from Figure 3-6.**

Peak (cm <sup>-1</sup> )	BM	Panc	Presumed Origin	Ref. Peak (cm <sup>-1</sup> )	Ref.
761		x	Ring breathing in tryptophan	760	[38]
803	x		Ring breathing mode in uracil	802	[38]
813		x	C-C stretch in collagen	813	[38]
836		x	Deformative vibrations of amine groups	838	[38]
866	x		C-C stretch in collagen	868	[39]
873		x	Hydroxyproline, tryptophan	873	[38]
920		x	C-C stretch of proline ring/glucose/lactic acid; collagen assignment	920	[38]
998	x		C-O in ribose, C-C	996	[29,38]
1061	x		C-C in-plane bending; C-N stretching	1053	[38]
1077	x		C-C or C-O stretch in lipids	1078	[43]
1151		x	C-N stretch in proteins	1152	[38,43]
1158	x		C-C/C-N stretch in proteins	1158	[38]
1206		x	Hydroxyproline, tyrosine (collagen assignment)	1206	[38]
1236	x		Amide III	1235	[38]
1265	x		Amide III (collagen assignment)	1265	[43]
1274		x	Amide III	1275	[38]
1334		x	CH <sub>2</sub> CH <sub>3</sub> twisting and wagging in collagen and nucleic acids	1335	[38,43]
1364		x	Tryptophan	1365	[38]
1400	x		N-H in-plane deformation	1400	[38]
1480	x		Ring breathing mode in guanine and adenine	1485	[38]

Peak (cm <sup>-1</sup> )	BM	Panc	Presumed Origin	Ref. Peak (cm <sup>-1</sup> )	Ref.
1500		x	N-H bending	1506	[38]
1528		x	In-plane vibrations of -C=C-	1528	[40]
1612	x		Cytosine (NH <sub>2</sub> )	1610	[38]
1658	x		Amide I ( $\alpha$ -helix)	1658	[38]

### 3.3.3 Statistical Analysis and Machine Learning

PCA was employed to reduce the dimensionality of the spectral data sets and to determine the main compositional differences between the Panc-MS and BM-MS EVs. With respect to the spectra presented in **Figure 3-6**, the first principal component (PC1) that explains 50.7 % of variance corresponds to a protein peak at 1236 cm<sup>-1</sup>, which belongs to amide III. The second principal component (PC2) that explains 15.5 % of variance corresponds to an amino acid peak at 761 cm<sup>-1</sup>, which corresponds to tryptophan. However, plotting PC1 versus PC2 reveals a large overlap in data between the two groups, limiting the ability to separate the two EV types (**Figure 3-7A**). We additionally plotted PC1 versus the third principal component (PC3), which accounts for 7.7 % of variance and corresponds to a peak at 1528 cm<sup>-1</sup> (in-plane -C=C- vibrations). In the second score plot, both data sets are clustered more tightly compared to the first score plot, as evidenced by the smaller confidence ellipses surrounding the former compared to the latter (**Figure 3-7B**). However, there was still a great amount of overlap between the two ellipses, and we could only classify the two EV groups with 82 % accuracy, 74 % sensitivity, and 84 % specificity.

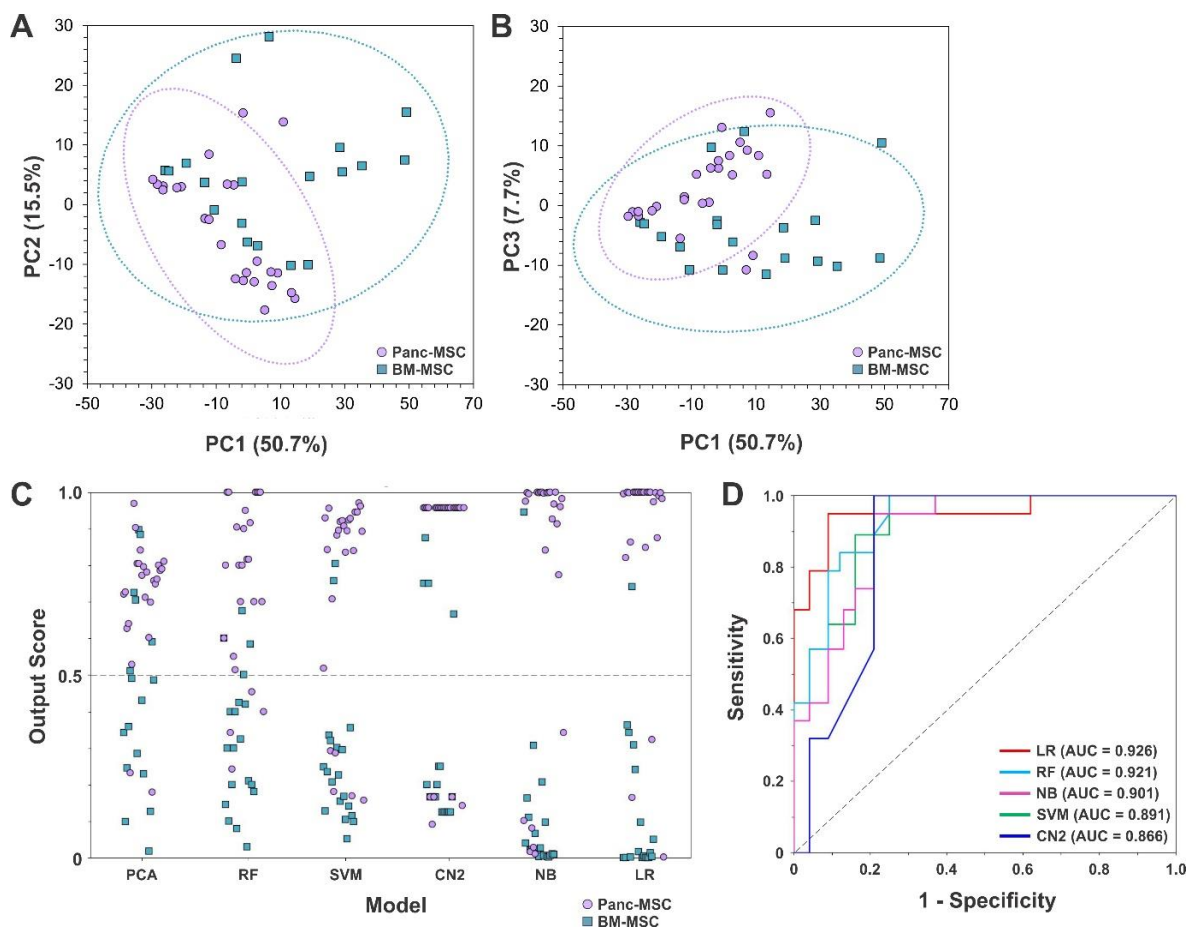
It is not uncommon for PCA to perform poorly in terms of classification tasks since PCA ignores class labels while attempting to maximize variance, and often PCA is used in conjunction with other classification techniques to improve class separability.<sup>[44]</sup> To mitigate this problem, we used the PC scores as classifiers in various machine learning algorithms. Machine learning is a powerful and automatized technique that can be used to make predictions about data without being explicitly programmed for these tasks. Treating spectral data sets with PCA is favored when exploring classification techniques by machine learning since smaller data sets are less computationally demanding for machine learning algorithms, and therefore faster. Furthermore, feeding raw spectral data to a machine learning algorithm can lead to overfitting due to the high dimensionality of the data.<sup>[45]</sup>

Machine learning algorithms in conjunction with the PCA data obtained were thus explored to increase class separation between the BM- and Panc-MSV EV spectral data as well as classification accuracy.

Five different machine learning algorithms were employed: random forest (RF), SVM, Naïve Bayes (NB), CN2 induction (CN2), and logistic regression (LR). RF is a technique that includes an ensemble of decision trees, in which data are modeled in hierarchical structures by a series of if/else statements.<sup>[46,47]</sup> SVM creates an optimal separating line for the classification of all the input data into different classes, while NB is a statistical method that computes the probability of an input's relevance to a pre-defined class.<sup>[47]</sup> In rule induction systems such as CN2, rules are created that fit the example cases, and solutions are found by linking rules to known facts (i.e., the data set).<sup>[48]</sup> Lastly, LR calculates the probability of class membership based on the sigmoid or logistic function.<sup>[49]</sup> Each model was tested using LOOCV to minimize bias that could occur when training with a small sample size. In the LOOCV procedure, one spectrum is held as a test sample while the remaining 43 spectra are used to train the model, until each spectrum has been used as a test sample once.

By visualizing the output scores, we observed that RF could not significantly separate the two groups of EVs, while NB and LR performed the best in terms of class separation (**Figure 3-7C**). To further assess the models, we plotted ROC curves and determined AUCS. While all five models had high AUC values, NB, RF, and LR outperformed CN2 and SVM with AUC values of 0.901, 0.921, and 0.926, respectively, compared to 0.866 and 0.891 (**Figure 3-7D**). Since AUC values are quite close together, it is necessary to also compare the models in terms of classification accuracy, sensitivity, and specificity (**Table 3-3**). CN2 had an unfavorable accuracy of 80 %, which was even lower than the accuracy achieved using PCA only. SVM, RF, and NB performed marginally better than PCA only with accuracies of 84 % each. LR was able to achieve a high accuracy of 89 %, as well as a high sensitivity and specificity of 89 % and 88 %, respectively. While all five machine learning algorithms performed well in terms of differentiating the Panc- and BM-MSV EVs, LR is the most favorable approach for this data set, considering the high AUC, accuracy, sensitivity, and specificity achieved compared to the four other algorithms.





**Figure 3-7: PCA score plots of Panc-MSC and BM-MSC clusters constructed with (A) the first and second PCs and (B) the first and third PCs, where one data point corresponds to one SERS spectrum. (C) Comparison of final output scores of data given by PCA only (PC1 vs. PC3) and five machine learning algorithms, where one data point corresponds to one SERS spectrum. (D) ROC curves comparing various machine learning algorithms.**

**Table 3-3: Comparison of abilities of PCA and various machine learning algorithms to discriminate BM-MSc and Panc-MSc EVs based on SERS spectra.**

Model	Sensitivity	Specificity	Accuracy
PCA Only (PC1 vs. PC3)	74 %	84 %	82 %
Logistic Regression	89 %	88 %	89 %
Support Vector Machine	89 %	80 %	84 %
Random Forest	84 %	84 %	84 %
Naïve Bayes	95 %	76 %	84 %
CN2 Rule Induction	79 %	80 %	80 %

In terms of differentiating MSC-derived EVs from different sources, our model works comparatively well with respect to other published studies, such as Gualerzi et al., who achieved a classification accuracy of 93.7 %.<sup>[29]</sup> Although this accuracy is higher than that achieved by our PCA-LR model, it is important to note that the spectra collected by Gualerzi et al. had to undergo significant treatment prior to data analysis to remove autofluorescence and background induced by their 532 nm laser. The use of a SERS platform and a lower energy excitation wavelength (i.e., 785 nm) usually removes the need for significant data treatment for better peak resolution, as demonstrated by the spectra we obtained in this study (**Figure 3-6**) and highly simplifies the data analysis.

### 3.4 Conclusions

Plasmonically active nanohole arrays were used to trap single EVs isolated from Panc- and BM-MScs, which were subsequently analyzed by SERS. Although the nanohole arrays are plasmonically active in the red and near-infrared wavelength regions, we determined that the near-infrared (785 nm) laser was the most suitable for probing these biological samples. By irradiating both Panc-MSc and BM-MSc EVs with the 785 nm laser, we found that the SERS spectra for both groups contained predominantly protein peaks, as we would expect to find on EV membrane surfaces, as well as some nucleic acid peaks. PCA was employed to determine the main compositional differences between these two EV sources. We determined that the peaks responsible for the most variance were located at 1236  $\text{cm}^{-1}$ , corresponding to amide III, 761  $\text{cm}^{-1}$ , belonging to tryptophan, and 1528  $\text{cm}^{-1}$ ,

corresponding to in-plane -C=C- vibrations. PC scores were then used as simple classifiers in training machine learning algorithms to separate the Panc- and BM-MSC EVs. Using simpler classifiers instead of the entire spectral data lowers the computational demand and time required to complete this classification task. With a logistic regression machine learning algorithm, we were able to distinguish between the two EV types with 89 % accuracy, 89 % sensitivity, and 88 % specificity. In future work, we would like to challenge these platforms with cancer-derived EVs to explore their feasibility as a tool in disease detection and diagnosis.

### 3.5 References

1. Jing, H.; He, X.; Zheng, J. *Transl. Res.* **2018**, *196*, 1-16.
2. Xin, H.; Li, Y.; Buller, B.; Katakowski, M.; Zhang, Y.; Wang, X.; Shang, X.; Zhang, Z. G.; Chopp, M. *Stem Cells* **2012**, *30*, 1556-1564.
3. Takeda, Y. S.; Xu, Q. *PLoS ONE* **2015**, *10*, e0135111.
4. El Bassit, G.; Patel, R. S.; Carter, G.; Shibu, V.; Patel, A. A.; Song, S.; Murr, M.; Cooper, D. R.; Bickford, P. C.; Patel, N. A. *Endocrinology* **2017**, *158*, 183-195.
5. Ibrahim, Ahmed G.-E.; Cheng, K.; Marbán, E. *Stem Cell Rep.* **2014**, *2*, 606-619.
6. Zhao, Y.; Sun, X.; Cao, W.; Ma, J.; Sun, L.; Qian, H.; Zhu, W.; Xu, W. *Stem Cells Int.* **2015**, *2015*, 761643.
7. Agarwal, U.; George, A.; Bhutani, S.; Ghosh-Choudhary, S.; Maxwell, J. T.; Brown, M. E.; Mehta, Y.; Platt, M. O.; Liang, Y.; Sahoo, S., et al. *Circ. Res.* **2017**, *120*, 701-712.
8. Tan, C. Y.; Lai, R. C.; Wong, W.; Dan, Y. Y.; Lim, S. K.; Ho, H. K. *Stem Cell Res. Ther.* **2014**, *5*, 76.
9. Nojima, H.; Freeman, C. M.; Schuster, R. M.; Japtok, L.; Kleuser, B.; Edwards, M. J.; Gulbins, E.; Lentsch, A. B. *J. Hepatol.* **2016**, *64*, 60-68.

10. Borges, F. T.; Melo, S. A.; Özdemir, B. C.; Kato, N.; Revuelta, I.; Miller, C. A.; Gattone li, V. H.; LeBleu, V. S.; Kalluri, R. *J. Am. Soc. Nephrol.* **2013**, *24*, 385-392.
11. Zhou, Y.; Xu, H.; Xu, W.; Wang, B.; Wu, H.; Tao, Y.; Zhang, B.; Wang, M.; Mao, F.; Yan, Y., et al. *Stem Cell Res. Ther.* **2013**, *4*, 34.
12. Jiang, Z. Z.; Liu, Y. M.; Niu, X.; Yin, J. Y.; Hu, B.; Guo, S. C.; Fan, Y.; Wang, Y.; Wang, N. S. *Stem Cell Res. Ther.* **2016**, *7*, 24.
13. Zhang, J.; Chen, C.; Hu, B.; Niu, X.; Liu, X.; Zhang, G.; Zhang, C.; Li, Q.; Wang, Y. *Int. J. Biol. Sci.* **2016**, *12*, 1472-1487.
14. Li, X.; Liu, L.; Yang, J.; Yu, Y.; Chai, J.; Wang, L.; Ma, L.; Yin, H. *EBioMedicine* **2016**, *8*, 72-82.
15. Zhao, B.; Zhang, Y.; Han, S.; Zhang, W.; Zhou, Q.; Guan, H.; Liu, J.; Shi, J.; Su, L.; Hu, D. *J. Mol. Histol.* **2017**, *48*, 121-132.
16. Furuta, T.; Miyaki, S.; Ishitobi, H.; Ogura, T.; Kato, Y.; Kamei, N.; Miyado, K.; Higashi, Y.; Ochi, M. *Stem Cells Transl. Med.* **2016**, *5*, 1620-1630.
17. Qi, X.; Zhang, J.; Yuan, H.; Xu, Z.; Li, Q.; Niu, X.; Hu, B.; Wang, Y.; Li, X. *Int. J. Biol. Sci.* **2016**, *12*, 836-849.
18. Zhang, S.; Chu, W. C.; Lai, R. C.; Lim, S. K.; Hui, J. H. P.; Toh, W. S. *Osteoarthr. Cartil.* **2016**, *24*, 2135-2140.
19. Zhu, Y.; Wang, Y.; Zhao, B.; Niu, X.; Hu, B.; Li, Q.; Zhang, J.; Ding, J.; Chen, Y.; Wang, Y. *Stem Cell Res. Ther.* **2017**, *8*, 64.
20. Nakamura, Y.; Miyaki, S.; Ishitobi, H.; Matsuyama, S.; Nakasa, T.; Kamei, N.; Akimoto, T.; Higashi, Y.; Ochi, M. *FEBS Lett.* **2015**, *589*, 1257-1265.
21. Choi, J. S.; Yoon, H. I.; Lee, K. S.; Choi, Y. C.; Yang, S. H.; Kim, I.-S.; Cho, Y. W. *J. Control. Release* **2016**, *222*, 107-115.

22. Zhang, Y.; Chopp, M.; Zhang, Z. G.; Katakowski, M.; Xin, H.; Qu, C.; Ali, M.; Mahmood, A.; Xiong, Y. *Neurochem. Int.* **2017**, *111*, 69-81.
23. Mead, B.; Tomarev, S. *Stem Cells Transl. Med.* **2017**, *6*, 1273-1285.
24. Teng, X.; Chen, L.; Chen, W.; Yang, J.; Yang, Z.; Shen, Z. *Cell. Physiol. Biochem.* **2015**, *37*, 2415-2424.
25. Tomasoni, S.; Longaretti, L.; Rota, C.; Morigi, M.; Conti, S.; Gotti, E.; Capelli, C.; Introna, M.; Remuzzi, G.; Benigni, A. *Stem Cells Dev.* **2013**, *22*, 772-780.
26. Shpacovitch, V.; Hergenröder, R. *Anal. Chim. Acta* **2018**, *1005*, 1-15.
27. van Niel, G.; D'Angelo, G.; Raposo, G. *Nat. Rev. Mol. Cell Biol.* **2018**, *19*, 213-228.
28. Möller, A.; Lobb, R. J. *Nat. Rev. Cancer* **2020**, *20*, 697-709.
29. Gualerzi, A.; Niada, S.; Giannasi, C.; Picciolini, S.; Morasso, C.; Vanna, R.; Rossella, V.; Masserini, M.; Bedoni, M.; Ciceri, F., et al. *Sci. Rep.* **2017**, *7*, 9820.
30. Gualerzi, A.; Kooijmans, S. A. A.; Niada, S.; Picciolini, S.; Brini, A. T.; Camussi, G.; Bedoni, M. *J. Extracell. Vesicles* **2019**, *8*, 1568780.
31. Kaufman, L.; Cooper, T.; Wallace, G.; Hawke, D.; Betts, D.; Hess, D.; Lagugné-Labarthe, F., Trapping and SERS Identification of Extracellular Vesicles Using Nanohole Arrays. In *Proc. SPIE 10894*, Plasmonics in Biology and Medicine XVI, 108940B: 2019.
32. Sherman, S. E.; Kuljanin, M.; Cooper, T. T.; Putman, D. M.; Lajoie, G. A.; Hess, D. A. *Stem Cells* **2017**, *35*, 1542-1553.
33. Cooper, T. T.; Sherman, S. E.; Bell, G. I.; Ma, J.; Kuljanin, M.; Jose, S. E.; Lajoie, G. A.; Hess, D. A. *Stem Cells* **2020**, *38*, 666-682.
34. Cooper, T. T.; Sherman, S. E.; Bell, G. I.; Dayarathna, T.; McRae, D. M.; Ma, J.; Lagugné-Labarthe, F.; Pasternak, S. H.; Lajoie, G. A.; Hess, D. A. *Stem Cells Dev.* **2021**, *30*, 247-264.

35. Haynes, C. L.; McFarland, A. D.; Van Duyne, R. P. *Anal. Chem.* **2005**, *77*, 338 A-346 A.
36. Sur, U. K. *Resonance* **2010**, *15*, 154-164.
37. Guo, L.; Jackman, J. A.; Yang, H.-H.; Chen, P.; Cho, N.-J.; Kim, D.-H. *Nano Today* **2015**, *10*, 213-239.
38. Talari, A. C. S.; Movasaghi, Z.; Rehman, S.; Rehman, I. U. *Appl. Spectrosc. Rev.* **2015**, *50*, 46-111.
39. Rehman, I. U.; Movasaghi, Z.; Rehman, S. FTIR and Raman Characteristic Peak Frequencies in Biological Studies. In *Vibrational Spectroscopy for Tissue Analysis*, CRC Press: Boca Raton, 2012; Vol. 1, pp 213-294.
40. Stremersch, S.; Marro, M.; Pinchasik, B.-E.; Baatsen, P.; Hendrix, A.; De Smedt, S. C.; Loza-Alvarez, P.; Skirtach, A. G.; Raemdonck, K.; Braeckmans, K. *Small* **2016**, *12*, 3292-3301.
41. Rau, K. R.; Guerra, A.; Vogel, A.; Venugopalan, V. *Appl. Phys. Lett.* **2004**, *84*, 2940-2942.
42. Gazor, M.; Talesh, S. S. A.; kavianpour, A.; Khatami, M.; Javidanbardan, A.; Hosseini, S. N. *Biotechnol. Bioproc. E.* **2018**, *23*, 49-54.
43. Huang, Z.; McWilliams, A.; Lui, H.; McLean, D. I.; Lam, S.; Zeng, H. *Int. J. Cancer* **2003**, *107*, 1047-1052.
44. Shin, H.; Oh, S.; Hong, S.; Kang, M.; Kang, D.; Ji, Y.-G.; Choi, B. H.; Kang, K.-W.; Jeong, H.; Park, Y., et al. *ACS Nano* **2020**, *14*, 5435-5444.
45. Howley, T.; Madden, M. G.; O'Connell, M.-L.; Ryder, A. G., The Effect of Principal Component Analysis on Machine Learning Accuracy with High Dimensional Spectral Data. In *Proc. AI-2005, Applications and Innovations in Intelligent Systems XII: 2005*; pp 209-222.

46. Qi, Y., Random Forest for Bioinformatics. In *Ensemble Machine Learning: Methods and Applications*, Zhang, C.; Ma, Y., Eds. Springer US: Boston, MA, 2012; pp 307-323.
47. Binkhonain, M.; Zhao, L. *Expert Syst. Appl.* **2019**, *1*, 100001.
48. Swe, S. M.; Sett, K. M. *Int. J. Trend Sci. Res. Dev.* **2019**, *3*, 1581-1584.
49. Dreiseitl, S.; Ohno-Machado, L. *J. Biomed. Inform.* **2002**, *35*, 352-359.

## Chapter 4

### 4 Characterization of Ovarian Cancer-Derived Extracellular Vesicles by Surface-Enhanced Raman Spectroscopy

A version of this work has been published in the journal *Analyst*: Culum, N. M.; Cooper, T. T.; Lajoie, G. A.; Dayarathna, T.; Pasternak, S. H.; Liu, J.; Fu, Y.; Postovit, L.-M.; Lagugné-Labarthet, F. Characterization of Ovarian Cancer-Derived Extracellular Vesicles Derived by Surface Enhanced Raman Spectroscopy. *Analyst* **2021**, Accepted manuscript, DOI: 10.1039/d1an01586a. This chapter has been reproduced with permission from the Royal Society of Chemistry.

Ovarian cancer is the most lethal gynecological malignancy, owing to the fact that most cases are diagnosed at a late stage. To improve prognosis and reduce mortality, we must develop methods for the early diagnosis of ovarian cancer. A step towards early and non-invasive cancer diagnosis is through the utilization of EVs, which can be thought of as nanoscale cancer biomarkers. In this chapter, the gold nanohole arrays presented in Chapter 3 are again used for the capture of EVs, this time from ovarian cancer (OvCa) cells, and subsequent characterization by SERS. Investigated in this chapter are EVs isolated from two established OvCa cell lines (OV-90 and OVCAR3), two primary OvCa cell lines (EOC6 and EOC18), and one human immortalized ovarian surface epithelial cell line (hIOSE) by SERS. The main compositional differences among these groups are identified by PCA and machine learning is further employed to discriminate the groups with high accuracy, sensitivity, and specificity.

#### 4.1 Introduction

In 2021, ovarian cancer is estimated to be the fifth most lethal cancer in the United States and the most lethal gynecological malignancy with 13 770 projected deaths.<sup>[1]</sup> The 5-year survival rate for women diagnosed with ovarian cancer is 47 %.<sup>[2,3]</sup> However, ovarian cancer usually presents at a late stage when the 5-year relative survival rate is 27 %, and few cases are diagnosed when the tumor is localized with a 5-year relative survival rate of 92 %.<sup>[4]</sup> The high mortality rate of ovarian cancer is due to the diagnostic delay arising



from the lack of early disease warning signs as well as the lack of early ovarian cancer screening options.<sup>[3]</sup> There is a need for the development of non-invasive approaches for the diagnosis of ovarian cancer, particularly at the early stage, which can be achieved by the characterization of EVs and the analysis of their biomarkers.

Cancer-derived EVs and their functions were introduced in Section 1.2.2. Briefly, EVs are involved in physio-pathological activities including cancer progression by mediating crosstalk between tumor and stromal cells.<sup>[5,6]</sup> Cancer cells have additionally been associated with an increase in EV production in comparison to normal cells, which could be related to the specific conditions of the tumor microenvironment.<sup>[5]</sup> Since EVs carry complex biological information from their parent cells, they are particularly interesting in cancer research as they can be exploited as cancer biomarkers.<sup>[7,8]</sup> SERS, as utilized in Chapter 3, is a promising technique for EV detection and characterization.<sup>[9]</sup> SERS analysis of cancer-derived EVs has been greatly explored, particularly in breast,<sup>[10-16]</sup> lung,<sup>[17-23]</sup> and pancreatic cancer.<sup>[24-26]</sup> Some studies have explored the SERS characterization of OvCa-derived EVs,<sup>[27-30]</sup> but analysis has generally been less thorough compared to the other cancer-derived EVs listed.

Early work in SERS characterization of OvCa EVs dates back to 2014, in which simple AuNPs were utilized to analyze EVs from cell line A2780 grown in normoxic (normal O<sub>2</sub>) and hypoxic (1 % O<sub>2</sub>) conditions.<sup>[27]</sup> However, with principal component analysis (PCA) followed by discriminant function analysis, the two EV types could be only be differentiated with 57.1 % sensitivity and 53.8 % specificity, and accuracy was not reported. One year later, a silver film-coated plasmonic nanobowl substrate fabricated by soft lithography on flexible polydimethylsiloxane was proposed for SERS analysis of EVs isolated from the SKOV-3 cell line.<sup>[29]</sup> The authors did not evaluate the diagnostic potential of the substrate since the purpose of the study was to compare EV purity from different isolation methods. In 2017, a more selective approach using a thiolated peptide ligand for the capture of EVs from SKOV-3 cells was reported.<sup>[28]</sup> The ligand was also bound to silver NPs by a thiol-metallic bond for SERS analysis. While the authors were able to demonstrate the targeted detection of OvCa EVs, they did not perform statistical analysis to evaluate the diagnostic potential of the probe, and so accuracy, sensitivity, and

specificity were not reported. Most recently in 2020, a simple plasmonic scaffold consisting of a cysteamine-treated microscale biosilicate substrate embedded with AgNPs was introduced for SERS analysis of EVs.<sup>[30]</sup> The authors investigated OvCa EVs derived from SKOV-3 cells as well as from patients with ovarian and endometrial cancer. With PCA-LDA, cancer could be diagnosed with 99.4 % accuracy, 100 % sensitivity, and 99.2 % specificity. However, the authors warned that these numbers must be interpreted with caution given the small sample size used in analysis that could lead to biased results.

As evidenced by these four studies, most OvCa EVs analyzed by SERS are derived from the SKOV-3 cell line. For this work, we sought to characterize EVs derived from OVCAR3 and OV-90 cells, which are well-established model systems for epithelial ovarian adenocarcinomas. Given that most (90 %) malignant ovarian tumors are epithelial in origin, and of these cancers, 70 % present as high-grade serous and < 5 % present as low-grade serous,<sup>[4]</sup> we also chose to characterize EVs from high-grade serous (EOC6) and low-grade serous (EOC18) primary cell lines. These four cell lines were compared to EVs derived from a non-malignant human immortalized ovarian surface epithelial (hIOSE) cell line as a control. As a SERS substrate, we again employed gold nanohole arrays fabricated by EBL that we have extensively reported and characterized in past work.<sup>[31][32]</sup> Herein, we report SERS spectra of the five aforementioned cell lines and were able to discriminate their spectral signals by PCA and logistic regression with extremely high accuracy, sensitivity, and specificity of approximately 99 % each.

## 4.2 Methods

### 4.2.1 Cell Culture

OV-90 (ATCC® CRL-11732) and NIH:OVCAR3 (ATCC® HTB-161) were obtained from the ATCC. Human immortalized surface epithelial cells hIOSE (OSE364) were obtained from the Canadian Ovarian Tissue Bank at the BC Cancer Agency. Primary cell lines EOC6 and EOC18 were isolated from the ascites of patients with high-grade and low-grade serous ovarian cancer, respectively. All cell lines, except OVCAR3, were maintained in M199+MCDB105 supplemented with 5 – 15% FBS. NIH:OVCAR3 cells were cultured in RPMI-1640 supplemented with 20 % FBS and 5 µg/mL insulin. Media was exchanged

with serum-free media for 20 – 30 hours to generate conditioned media (CM) for EV purification. All work involving the use of patient samples (cell lines, plasma, and ascites) was approved by the Health Research Ethics Board of the Alberta Cancer Committee.

#### 4.2.2 EV Isolation

OvCa CM samples were first centrifuged at  $200 - 300 \times g$  at  $4\text{ }^{\circ}\text{C}$  to pellet cells. Supernatants (except CM) were diluted 1:10 in PBS and centrifuged at  $3000 \times g$  for 20 minutes at  $4\text{ }^{\circ}\text{C}$  to remove cell debris. To remove large membrane fragments, supernatants were spun at  $10\ 000 \times g$  for an additional 20 minutes at  $4\text{ }^{\circ}\text{C}$ . Lastly, supernatants were ultracentrifuged at  $120\ 000 - 140\ 000 \times g$  (SW-28 rotor) for 2 hours at  $4\text{ }^{\circ}\text{C}$  to pellet EVs on an Optima L-100 XP ultracentrifuge (Beckman Coulter). The supernatant was removed and EVs were resuspended in  $100 - 300\ \mu\text{L}$  of PBS and stored at  $-80\text{ }^{\circ}\text{C}$  until further use.

#### 4.2.3 EV Protein Extraction and Digestion

To prepare CM for proteomic analysis,  $\sim 30\ \mu\text{g}$  of EVs were lyophilized to dryness and reconstituted in 8 M urea, 50 mM ammonium bicarbonate (ABC), 10 mM dithiothreitol (DTT), and 2 % SDS lysis buffer. Proteins were sonicated at  $10 \times 0.5\ \text{s}$  pulses (Level 1) with a probe sonicator (Fisher Scientific, Waltham, MA), reduced in 10 mM DTT for 30 minutes at room temperature (RT), alkylated in 100 mM iodoacetamide for 30 minutes at RT in the dark, and precipitated in chloroform/methanol.<sup>[33]</sup> On-pellet in-solution protein digestion was performed in  $100\ \mu\text{L}$  of 50 mM ABC (pH 8) by adding 1/50 Trypsin/Lys-C (Promega) to digest EV proteins. EV proteins were incubated at  $37\text{ }^{\circ}\text{C}$  overnight ( $\sim 20\ \text{h}$ ) in a ThermoMixer C (Eppendorf) at 900 rpm before acidifying to pH 3 – 4 with 10 % formic acid (FA). Salts and detergents were removed from peptide samples using C18 stagetips made in-house. Briefly, 10 layers were stacked into  $200\ \mu\text{L}$  pipette tips and rinsed with ice-cold methanol. Stagetips were conditioned with solution A (80/20/0.1 %; acetonitrile (ACN)/water/trifluoroacetic acid (TFA)), followed by solution B (5/95/0.1 %; ACN/water/TFA) prior to loading  $\sim 20\ \mu\text{g}$  of peptides resuspended in solution B. Duplicate washes were performed with solution B prior to elution of peptides using solution C (80/20/0.1 %; ACN/water/FA) and final elution using a 50/50 mixture of ACN/0.1 % FA. Peptides were centrifuged at  $45\text{ }^{\circ}\text{C}$  under vacuum and resuspended in 0.1 % FA prior to

quantification by the bicinchonic acid assay (BCA) and injection into the mass spectrometer.

#### 4.2.4 Ultraperformance Liquid Chromatography Coupled to Tandem Mass Spectrometry (UPLC-MS/MS)

Peptides were analyzed using an ACQUITY UPLC M-class system (Waters) connected to a Q Exactive™ Plus mass spectrometer (Thermo Scientific) using a nonlinear gradient. Buffer A consisted of water/0.1 % FA and buffer B consisted of ACN/0.1 % FA. Peptides (~ 1 µg estimated by BCA) were initially loaded onto an ACQUITY UPLC M-Class Symmetry C18 Trap Column (100 Å, 5 µm, 180 µm × 20 mm) and trapped for 5 minutes at a flow rate of 5 µL/min at 99 % A/1 % B. Peptides were separated on an ACQUITY UPLC M-Class Peptide BEH C18 Column (130 Å, 1.7 µm, 75 µm × 250 mm) operating at a flow rate of 300 nL/min at 35 °C using a non-linear gradient consisting of 1 – 10 % B, 10 – 20 % B, 20 – 30 % B, 30 – 40 % B, 40 – 50 % B, 60 – 70 % B, and 80 – 90 % B for 10 minute intervals before cyclic washing between 5 – 95 % B, equaling 140 minute gradient total. Settings for data acquisition on the Q Exactive™ Plus are outlined in **Table 4-1**.

**Table 4-1: Parameters for Q Exactive™ Plus.**

Parameter	Parameter Setting
Orbitrap Resolution (MS1)	$7 \times 10^4$
Mass Range	400 – 1500 m/z
MS1 Injection Time	200 ms
MS1 Automatic Gain Control (AGC) Target	$3 \times 10^6$ ions/cycle
Lock Mass	445.120025 m/z
MS2 Detection	Fourier Transform
MS2 Resolution	$1.75 \times 10^4$
MS2 AGC Target	$2 \times 10^5$ ions/cycle
MS2 Injection Time	50 ms
Loop Count	12
Isolation Width	1.2 m/z

Parameter	Parameter Setting
Isolation Offset	0.5 m/z
MS2 Activation	Higher-energy C-trap dissociation
Normalized Collision Energy	25 %
Dynamic Exclusion	12
Minimum AGC Target	$2 \times 10^3$ ions/cycle
MS2 Intensity Threshold	$8 \times 10^4$
Exclusion Duration	30 s
Charge Exclusion	Unassigned, 1, 8, > 8
Polarity	Positive

#### 4.2.5 UPLC-MS/MS Data Analysis

MS raw files were searched with de novo peptide sequencing software PEAKS (version 10.5) using the Human Uniprot database (reviewed only, updated November 2020). Missed cleavages were set to 5 and I = L. Cysteine carbamidomethylation was set as a fixed modification. Oxidation (M), N-terminal acetylation (protein), and deamidation (NQ) were set as variable modifications (maximum number of modifications per peptide = 7) and all other settings were left as default. Precursor mass deviation was left at 20 ppm and 4.5 ppm for first and main search, respectively. Fragment mass deviation was left at 20 ppm. Protein and peptide false discovery rate was set to 0.01 (1 %) and the decoy database was set to revert. Proteomic data analysis and visualization was performed in a Python 3+ environment, albeit comparison to Vesiclepedia databases was performed using the open-source FunRich (version 3.1.3) software.

#### 4.2.6 nFC Measurements

OvCa EVs were analyzed for the number of microparticles/mL and EV size distribution by nanoscale flow cytometry (nFC). Serial injections (2, 5, 10, or 20 mL) of each concentrate were diluted to 300 mL with 0.22 mm-filtered PBS within low-attachment 96-well plates at RT. EVs were enumerated in duplicate on the Apogee A60 nanoscale flow cytometer with autosampler, capable of EV resolution between 150 and 1000 nm.<sup>[34]</sup> 130  $\mu$ L of diluted CM was injected and analyzed at 1.5  $\mu$ L/min for 1 minute. The size of secreted

microparticles was estimated using silica beads ranging 110 – 1300 nm using properties of large-angle light scatter and small-angle light scatter. Silica beads provide a refractive index ( $k = 1.42$ ) that is closer to cells ( $k = 1.35 - 1.39$ ) than commonly used polystyrene beads ( $k = 1.59$ ).

#### 4.2.7 Nanohole Array Fabrication

The nanohole arrays presented in this chapter were fabricated by EBL following the same protocol outlined in Section 3.2.1, established by Kaufman et al.<sup>[31]</sup>

#### 4.2.8 SERS Measurements

Concentrated EV samples were diluted 1:20 with Milli-Q water. 10 – 20  $\mu\text{L}$  of dilute EV samples were drop-cast onto nanohole arrays and subsequently removed using cohesive properties allowed by a Kimwipe absorbent paper (Kimberly-Clark Inc.). The edge of the absorbent paper was placed on the corner of the solution droplet, allowing for solution removal via capillary action. EV solutions were allowed to dry for 15 – 30 minutes prior to SERS measurements. EVs were located in nanoholes by SERS mapping, and spectra were extracted from these maps. SERS spectra were acquired with an XploRA<sup>TM</sup> PLUS spectrometer (Horiba Scientific) using a 785 nm excitation laser source, 600 grooves/mm grating, 100 $\times$  objective (N.A. = 0.9), and 100  $\mu\text{m}$  pinhole. Laser power was set to 5 mW with an acquisition time of 4 seconds per spectrum.

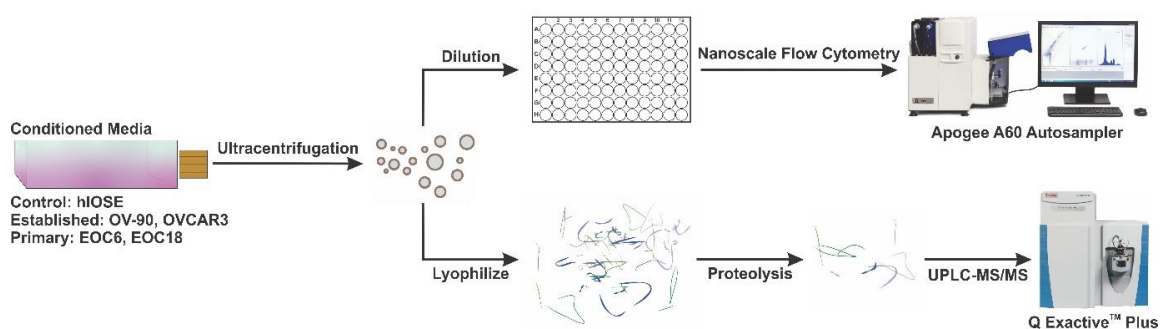
#### 4.2.9 Statistical Analysis and Machine Learning

PCA and machine learning were completed with Orange software (version 3.28.1). All SERS spectra were standardized to  $\mu = 0$ ,  $\sigma^2 = 1$  prior to PCA, and outliers were removed by the Local Outlier Factor method with contamination set to 10 %. Retained spectra were analyzed by PCA and the first 25 PCs were selected to explain 97 % of variance among spectra. The first 25 PCs were used as input data in a logistic regression-based machine learning algorithm. Models were trained and tested using 5-fold CV.

## 4.3 Results and Discussion

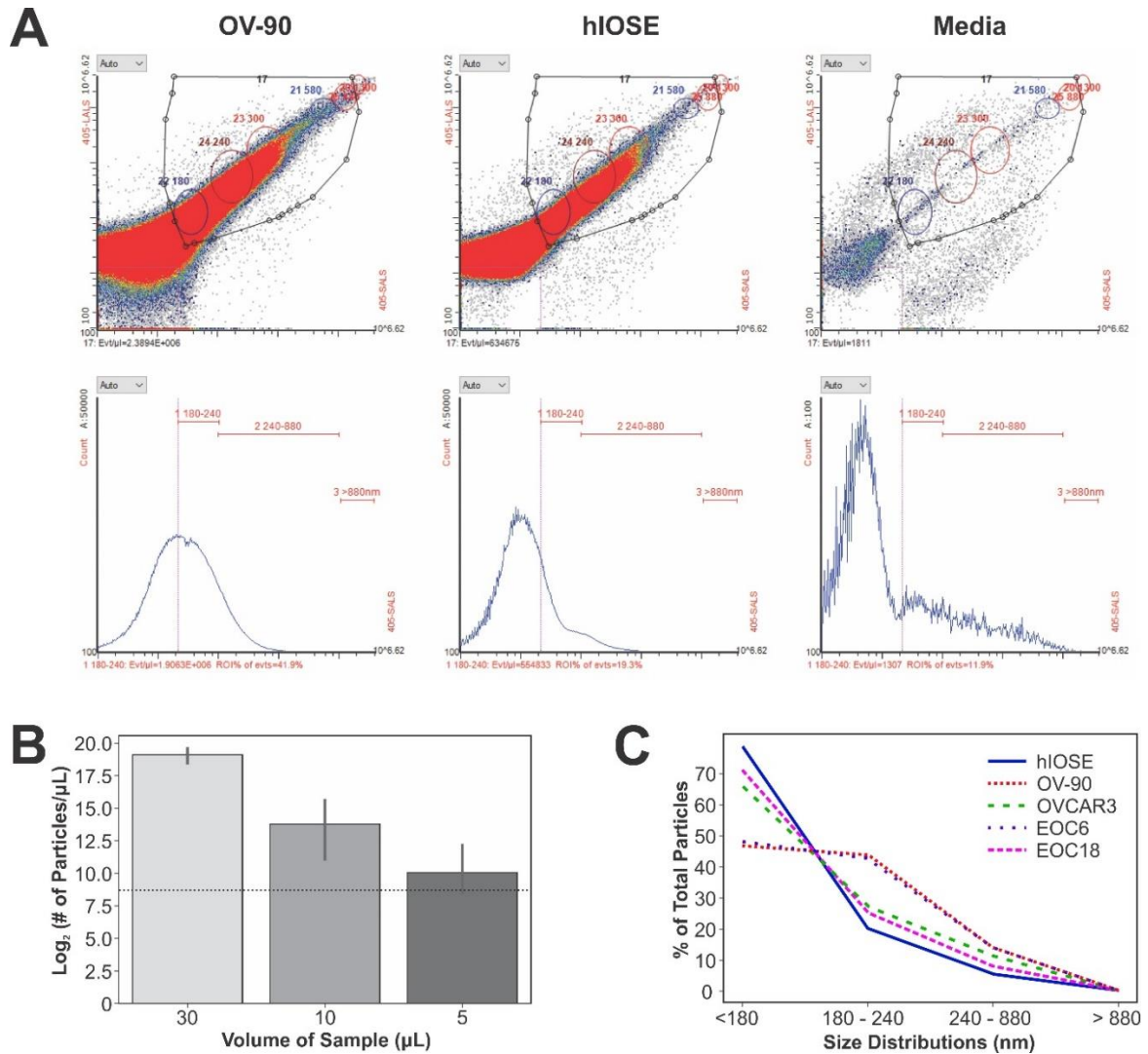
### 4.3.1 Validation of EV Size and Proteomic Cargo

Size quantification of EVs by nFC and proteomic analysis by UPLC-MS/MS was conducted to validate the enrichment of EVs by ultracentrifugation (UC) from CM generated by four OvCa cell lines and one normal control cell line (**Figure 4-1**). Amongst several methods available for EV purification,<sup>[35-38]</sup> this study considered UC as an optimal balance between EV yield and purity. UC provides a heterogenous mixture of EVs which includes both microvesicles and exosomes. Therefore, the use of the term EV throughout this study is in accordance with the ISEV.<sup>[39]</sup>



**Figure 4-1: Schematic illustration of nFC and UPLC-MS/MS analysis of OvCa EVs.**

The size distribution of EVs was verified by nFC (**Figure 4-2A**), estimated by small-angle light scatter measurements. In support of previous studies,<sup>[34]</sup> we demonstrate a near-linear detection of particles as the volume of sample analyzed was diminished, indicating acceptable EV purity (**Figure 4-2B**). The majority of EVs detected were estimated to be less than 240 nm in diameter (**Figure 4-2C**), which is consistent with the size of microvesicles and exosomes.<sup>[40]</sup> However, the resolution of exosomes less than 100 nm in diameter from background noise was unattainable based on the properties of the cytometer used in this study, but subsequent proteomic analysis was able to verify the presence of EV markers in each OvCa sample.

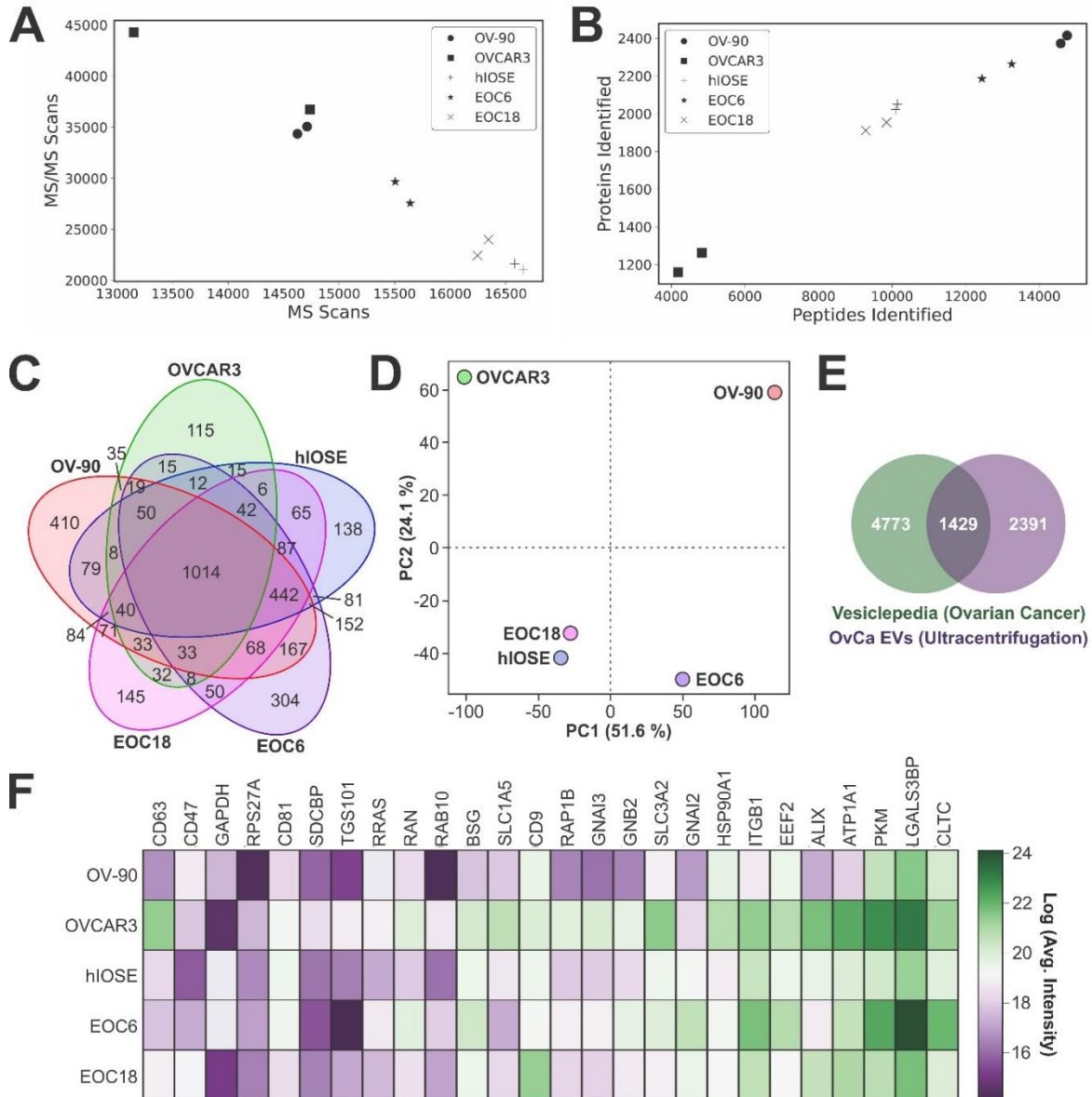


**Figure 4-2: (A) Representative flow cytometry plots (top) and corresponding size distributions (bottoms) of EV samples from OV-90 (left) and hIOSE (middle) cell lines, as well as conditioned media (right). (B) nFC results revealed a near-linear decrease in the number of particles as the volume of sample analyzed diminished. The horizontal dashed line indicates the mean particle number when an equal volume of background media (PBS) was analyzed. (C) Size distribution of EVs.**

UPLC-MS/MS is a powerful tool to detect and quantify proteomic cargo within EVs (**Figure 4-3A-B**). In total, we identified over 3000 proteins in EVs generated by all five cell lines. However, unique proteomic cargo was detected among each. Specifically, 1014 peptides were shared by EVs from all five cell lines, whereas 138, 410, 115, 304, and 145



peptides were unique to hIOSE, OV-90, OVCAR3, EOC6, and EOC18, respectively (**Figure 4-3C**). Distinct proteomic fingerprints were identified by PCA, where PC1 and PC2 scores correspond to tetraspanin-8 (TSPAN8) and histone H2A type 1 (HIST1H2A), respectively (**Figure 4-3D**). EVs isolated from OVCAR3, OV-90, and EOC6 each demonstrate a distinguishing proteome compared to EVs isolated from hIOSE and EOC18. A high PC1 score indicates the presence of TSPAN8 in EVs isolated from OV-90 and EOC6, while a high PC2 score indicates the presence of HIST1H2A in EVs isolated from OVCAR-3 and OV-90. EVs isolated from hIOSE and EOC18 are both associated with low PC1 and PC2 scores, indicating a low presence of TSPAN8 and HIST1H2A in these EVs. Furthermore, EVs from hIOSE and EOC18 are clustered closely to one another on the score plot, indicating that there is a high similarity between normal EVs and low-grade OvCa EVs, as expected. OvCa proteomes identified here were compared with those available in the Vesiclepedia database filtered for OvCa cell lines (**Figure 4-3E**). Of 6202 proteins with OvCa EV proteomes in this database, 1429 were shared with EV proteomes identified within our samples. Classical EV markers (e.g., CD9/81/63) were also confidently detected in EVs isolated from all five cell lines in addition to 22 core exosome proteins recently identified by UPLC-MS/MS (**Figure 4-3F**).<sup>[41]</sup> Collectively, our nFC and proteomic data support UC as a suitable method for the enrichment of EVs for downstream SERS analysis.



**Figure 4-3:** (A) Number of MS1 and MS2 scans for duplicate injections of each EV sample. (B) Number of peptides identified by de novo sequencing, which led to confident protein identification with false discovery rate ( $p < 0.01$ ). (C) Venn diagram highlighting the distinct proteomics cargo contained within EVs from various OvCa cell lines, leading to distinct proteomic “fingerprints” as demonstrated by (D) PCA. (E) Overlap of OvCa EV proteomes with Vesiclepedia database filtered for OvCa. (F) Heat map indicating that classical EV markers (e.g., CD9/81/63) were confidently identified in all OvCa EV samples.

### 4.3.2 EV Characterization by SERS

Gold nanohole arrays of different shapes (squares, triangles) and sizes (500 nm, 1000 nm) were fabricated by EBL as plasmonic substrates for the capture and SERS analysis of normal and OvCa-derived EVs. SEM images of these arrays and the characterization of their LSPRs can be found in Chapter 3. EVs isolated from the hIOSE cell line serve as a control group (**Figure 4-4A**). The majority of peaks found in hIOSE EVs are attributed to proteins and amino acids, which include the peaks at 755  $\text{cm}^{-1}$  (tryptophan), 818  $\text{cm}^{-1}$  (C-C stretching in collagen), 935  $\text{cm}^{-1}$  (proline, valine, protein backbone), 1029  $\text{cm}^{-1}$  (phenylalanine), 1303  $\text{cm}^{-1}$  (collagen, amide III), and 1545  $\text{cm}^{-1}$  (amide II).<sup>[42,43]</sup> Several nucleic acid peaks are located at 724  $\text{cm}^{-1}$  (adenine), 787  $\text{cm}^{-1}$  (cytosine, uracil, thymine), 1185  $\text{cm}^{-1}$  (cytosine, guanine, adenine), 1356  $\text{cm}^{-1}$  (guanine), and 1482  $\text{cm}^{-1}$  (guanine, adenine).<sup>[42,43]</sup> Lipid and carbohydrate peaks can be found at 1254  $\text{cm}^{-1}$  and 865  $\text{cm}^{-1}$ , respectively.<sup>[42,43]</sup> A complete list of SERS peaks of hIOSE EVs are summarized in **Table 4-2**.

**Table 4-2: Summary of peak assignments of hIOSE EVs, as highlighted in Figure 4-4A.**

Peak ( $\text{cm}^{-1}$ )	Presumed Origin	Ref. Peak ( $\text{cm}^{-1}$ )	Ref.
630	Glycerol	630	[42,43]
692	Ring deformation	686	[43]
724	Ring breathing mode of adenine	725	[42,43]
755	Symmetric breathing of tryptophan	755	[43]
787	Ring breathing mode of cytosine, uracil, thymine	786	[42,43]
818	C-C stretching in collagen	817	[42,43]
837	Deformative vibrations of amine groups	838	[42,43]
865	C-C stretching or C-O-C skeletal mode in carbohydrates	868	[42,43]
935	C-C stretching mode of proline, valine, and protein backbone ( $\alpha$ -helix); glycogen	935	[42,43]
961	Unassigned in protein assignments	963	[42,43]

Peak (cm <sup>-1</sup> )	Presumed Origin	Ref. Peak (cm <sup>-1</sup> )	Ref.
1029	Phenylalanine of collagen	1030	[43]
1090	Symmetric phosphate stretching vibrations	1090	[42,43]
1185	Cytosine, guanine, adenine	1180 – 1184	[42,43]
1254	Lipids	1255	[42,43]
1303	CH <sub>2</sub> /CH <sub>3</sub> twisting, wagging, or bending mode of lipid/collagen; amide III	1302	[42,43]
1356	Guanine	1355, 1357	[42,43]
1404	C-H deformation	1404	[42]
1455	Deoxyribose; CH <sub>2</sub> scissoring of proteins and lipids	1455	[43]
1482	Ring breathing mode of guanine, adenine	1485	[42,43]
1545	Amide II	1544	[42]

EVs from two established OvCa cell lines, OV-90 and OVCAR3, serve as cancer models (**Figure 4-4B**). SERS spectra of EVs from both OV-90 and OVCAR3 are dominated by protein and amino acid peaks. For OV-90 EVs, these peaks are located at 755 cm<sup>-1</sup> (tryptophan), 908 cm<sup>-1</sup> (tyrosine), 1008 cm<sup>-1</sup> (phenylalanine), 1036 cm<sup>-1</sup> (phenylalanine), 1151 cm<sup>-1</sup> (C-N stretch), 1274 cm<sup>-1</sup> (amide III), 1335 cm<sup>-1</sup> (collagen, amide III), and 1533 cm<sup>-1</sup> (amide II).<sup>[42,43]</sup> For OVCAR3 EVs, the protein and amino acid peaks are found at 741 cm<sup>-1</sup> (tryptophan), 818 cm<sup>-1</sup> (collagen), 935 cm<sup>-1</sup> (proline, valine, protein backbone), 956 cm<sup>-1</sup> (CH<sub>3</sub> stretching), 1055 cm<sup>-1</sup> (C-O, C-N stretching), 1176 cm<sup>-1</sup> (tyrosine), 1197 cm<sup>-1</sup> (tryptophan), 1226 cm<sup>-1</sup> (amide III), and 1584 cm<sup>-1</sup> (phenylalanine).<sup>[42,43]</sup> Second most abundant for OV-90 and OVCAR3 EVs are nucleic acid peaks. For OV-90 EVs, these nucleic acid peaks are located at 675 cm<sup>-1</sup> (guanine), 782 cm<sup>-1</sup> (thymine, cytosine, uracil), 822 cm<sup>-1</sup> (phosphodiester), 1186 cm<sup>-1</sup> (cytosine, guanine, adenine), and 1201 cm<sup>-1</sup> (nucleic acids and phosphates).<sup>[42,43]</sup> For OVCAR3 EVs, the nucleic acid peaks are located at 675 cm<sup>-1</sup> (guanine), 724 cm<sup>-1</sup> (adenine), 1176 cm<sup>-1</sup> (cytosine, guanine), 1376 cm<sup>-1</sup> (adenine), and 1483 cm<sup>-1</sup> (guanine, adenine).<sup>[42,43]</sup> Carbohydrate peaks in the OV-90 SERS spectrum are found at 943 cm<sup>-1</sup>, 1118 cm<sup>-1</sup>, and 1370 cm<sup>-1</sup>, and in the OVCAR3 SERS spectrum are

found at 848  $\text{cm}^{-1}$ , 929  $\text{cm}^{-1}$ , and 994  $\text{cm}^{-1}$ .<sup>[42,43]</sup> Lastly, OV-90 lipid peaks are located at 714  $\text{cm}^{-1}$ , 719  $\text{cm}^{-1}$ , and 1467  $\text{cm}^{-1}$ , while an OVCAR3 lipid peak is found at 1299  $\text{cm}^{-1}$ .<sup>[42,43]</sup> The complete list of spectral peaks for OV-90 and OVCAR3 EVs are summarized in **Table 4-3** and **Table 4-4**, respectively.

**Table 4-3: Summary of peak assignments of OV-90 EVs, as highlighted in Figure 4-4B.**

Peak ( $\text{cm}^{-1}$ )	Presumed Origin	Ref. Peak ( $\text{cm}^{-1}$ )	Ref.
632	Glycerol	630 *	[42,43]
675	Ring breathing mode in guanine	678	[42,43]
714	C-N (membrane phospholipid head); adenine	717	[43]
719	C-N (membrane phospholipid head), symmetric stretch vibration of choline group $\text{N}^+(\text{CH}_3)_3$ ; nucleotide	719	[42,43]
755	Symmetric breathing of tryptophan	755 *	[43]
782	Ring breathing in thymine, cytosine, uracil	782	[42,43]
822	Phosphodiester	822	[42,43]
908	Tyrosine	906	[42]
943	Skeletal modes in polysaccharides	941	[42,43]
1008	Phenylalanine	1008	[42,43]
1036	C-H in-plane bending mode of phenylalanine	1036	[43]
1118	Glucose	1117	[42,43]
1151	C-N stretching in proteins	1152	[42,43]
1186	Cytosine, guanine, adenine	1180 – 1184 *	[42,43]
1201	Nucleic acids and phosphates; aromatic C-O and C-N	1200	[42,43]
1220	C=N=C stretching	1220	[42,43]
1274	Amide III	1275	[42,43]

Peak (cm <sup>-1</sup> )	Presumed Origin	Ref. Peak (cm <sup>-1</sup> )	Ref.
1335	CH <sub>2</sub> /CH <sub>3</sub> twisting and wagging in collagen and nucleic acids; C-N stretching in amide III	1335	[42,43]
1370	Saccharide band	1370	[42,43]
1403	C-H deformation	1404 *	[42]
1467	Lipids	1465	[42,43]
1533	Amide II	1542	[42]

\* Also present in control (hIOSE) spectrum.

**Table 4-4: Summary of peak assignments of OVCAR3 EVs, as highlighted in Figure 4-4B.**

Peak (cm <sup>-1</sup> )	Presumed Origin	Ref. Peak (cm <sup>-1</sup> )	Ref.
675	Ring breathing mode in guanine	678	[42,43]
687	Ring deformation	686 *	[43]
724	Ring breathing mode of adenine	725 *	[42,43]
741	DNA, tryptophan	742	[42]
818	C-C stretching in collagen	817 *	[42,43]
848	C-O-C skeletal mode in carbohydrates	847	[42,43]
929	Carbohydrates	931	[42,43]
935	C-C stretching mode of proline, valine, and protein backbone ( $\alpha$ -helix); glycogen	935 *	[42,43]
956	CH <sub>3</sub> stretching in proteins ( $\alpha$ -helix)	951	[42,43]
994	C-O ribose, C-C	996	[42,43]
1055	C-O stretching, C-N stretching in proteins	1053	[42,43]
1176	C-H bending in tyrosine	1176	[42]
1197	Tryptophan ring breathing	1199	[42]
1226	Amide III ( $\beta$ -sheet)	1224	[42,43]
1299	Acyl chains, fatty acids	1298	[42,43]
1324	CH <sub>2</sub> /CH <sub>3</sub> wagging mode in collagen and purine bases	1324	[42,43]

Peak (cm <sup>-1</sup> )	Presumed Origin	Ref. Peak (cm <sup>-1</sup> )	Ref.
1346	Adenine and guanine; C-H deformation of proteins	1344	[43]
1376	Ring breathing mode of adenine	1376	[43]
1455	Deoxyribose; CH <sub>2</sub> scissoring of proteins and lipids	1455 *	[43]
1483	Ring breathing mode of guanine, adenine	1485 *	[42,43]
1529	-C=C- in-plane vibrations	1525	[42,43]
1584	C=C bending mode of phenylalanine	1583	[42,43]

\* Also present in control (hIOSE) spectrum.

EVs from two primary OvCa cell lines, EOC6 and EOC18, were also probed as models of high-grade and low-grade serous cancer, respectively (**Figure 4-4C**). Like the EVs from the three aforementioned cell lines, the SERS spectra of EOC6 and EOC18 EVs are dominated by protein and amino acid peaks. For EOC6 EVs, these peaks are located at 639 cm<sup>-1</sup> (tyrosine), 939 cm<sup>-1</sup> (C-C skeletal stretching), 1003 cm<sup>-1</sup> (phenylalanine), 1159 cm<sup>-1</sup> (C-C/C-N stretching), 1162 cm<sup>-1</sup> (tyrosine), 1225 cm<sup>-1</sup> (amide III), 1265 cm<sup>-1</sup> (collagen, phenylalanine), 1439 cm<sup>-1</sup> (collagen), and 1558 cm<sup>-1</sup> (tryptophan, tyrosine, amide II).<sup>[42,43]</sup> For EOC18 EVs, protein and amino acid peaks are found at 756 cm<sup>-1</sup> (tryptophan), 852 cm<sup>-1</sup> (proline, hydroxyproline, tyrosine), 935 cm<sup>-1</sup> (proline, valine, protein backbone), 987 cm<sup>-1</sup> (phenylalanine), 1032 cm<sup>-1</sup> (phenylalanine, proline), 1209 cm<sup>-1</sup> (tryptophan, phenylalanine), 1248 cm<sup>-1</sup> (amide III), 1338 cm<sup>-1</sup> (amide III), and 1362 cm<sup>-1</sup> (tryptophan).<sup>[42,43]</sup> However, fewer nucleic acid peaks are located compared to OV-90 and OVCAR3 EVs, which are found at 797 cm<sup>-1</sup> (uracil) for EOC6 EVs and 677 cm<sup>-1</sup> (guanine) and 1577 cm<sup>-1</sup> (guanine, adenine) for EOC18 EVs.<sup>[42,43]</sup> Lipid peaks for EOC6 EVs are located at 736 cm<sup>-1</sup>, 775 cm<sup>-1</sup>, and 1367 cm<sup>-1</sup>, and in EOC18 EVs are found at 1166 cm<sup>-1</sup>, 1300 cm<sup>-1</sup>, 1466 cm<sup>-1</sup>.<sup>[42,43]</sup> Lastly, carbohydrate peaks of EOC6 EVs are present at 845 cm<sup>-1</sup> and 1023 cm<sup>-1</sup>, while one carbohydrate peak at 1425 cm<sup>-1</sup> is present in the SERS spectra of the EOC18 EVs.<sup>[42,43]</sup> The complete list of spectral peaks of the EOC6 and EOC18 EVs are summarized in **Table 4-5** and **Table 4-6**, respectively.

**Table 4-5: Summary of peak assignments of EOC6 EVs, as highlighted in Figure 4-4C.**

Peak (cm <sup>-1</sup> )	Presumed Origin	Ref. Peak (cm <sup>-1</sup> )	Ref.
639	Tyrosine ring breathing	639	[42]
694	Ring deformation	686 *	[43]
736	Phosphatidylserine	733	[42,43]
775	Phosphatidylinositol	776	[42,43]
797	Ring breathing mode in uracil	802	[42,43]
845	C-O-C skeletal mode in carbohydrates	847	[42,43]
886	Ring deformation and symmetric C-N-C stretching	886	[43]
939	C-C skeletal stretching in proteins	939	[43]
1003	C-C skeletal mode, phenylalanine	1003	[42,43]
1023	Glycogen	1023	[42,43]
1096	Phosphodioxy group (PO <sub>2</sub> <sup>-</sup> in nucleic acids)	1096	[42,43]
1159	C-C/C-N stretching in proteins	1158	[42]
1162	Tyrosine	1163	[42]
1225	Amide III (β-sheet)	1224	[42,43]
1265	Amide III of collagen; C-C <sub>6</sub> H <sub>5</sub> stretching in phenylalanine	1265	[43]
1332	C-C stretching in phenyls, C-O stretching, C-H in-plane bending	1332	[42,43]
1367	CH <sub>3</sub> stretching in phospholipids	1367	[42,43]
1404	C-H deformation	1404 *	[42]
1439	CH <sub>2</sub> /CH <sub>3</sub> deformation in collagen	1439	[42,43]
1520	-C=C- in-plane vibrations	1525	[42,43]
1558	Tryptophan, tyrosine, amide II	1558	[42,43]

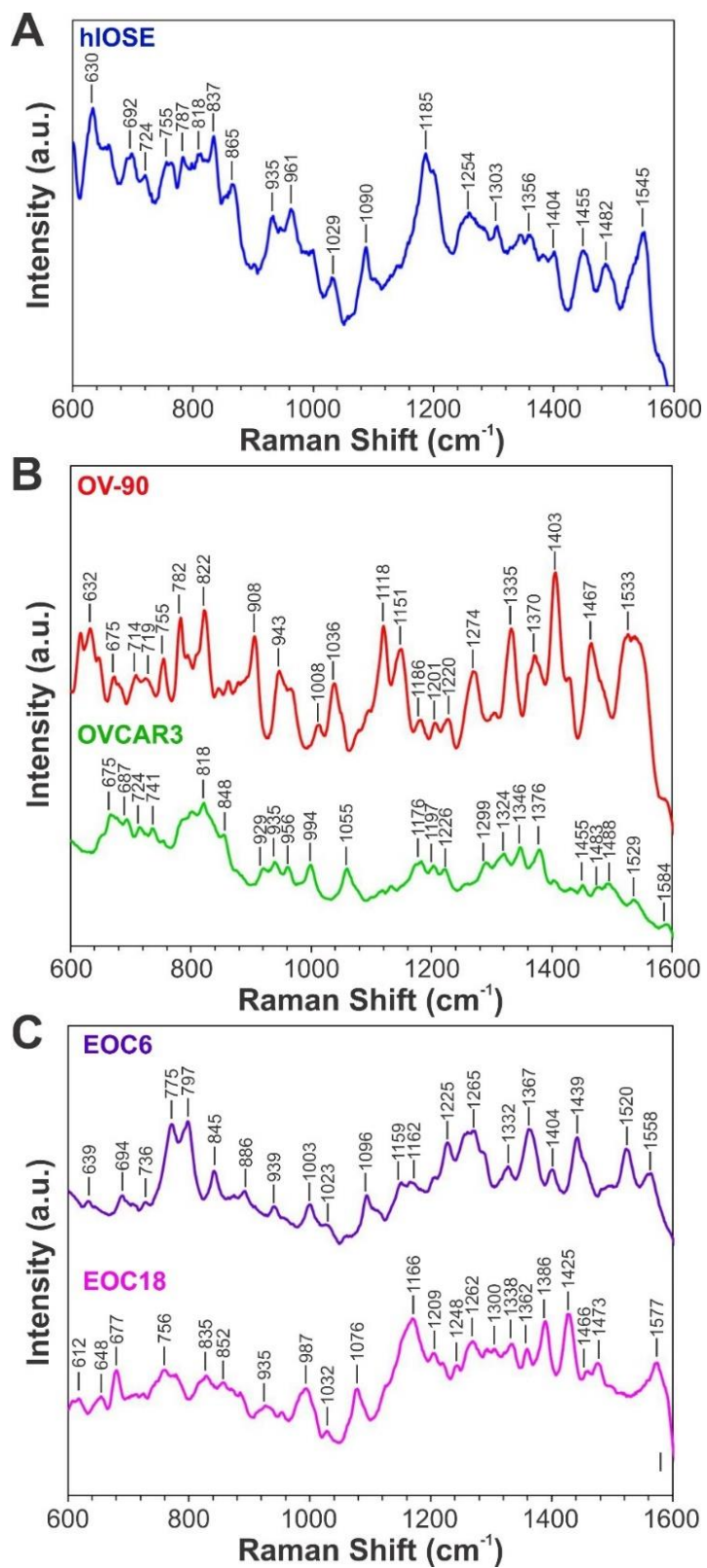
\* Also present in control (hIOSE) spectrum.



**Table 4-6: Summary of peak assignments of EOC18 EVs, as highlighted in Figure 4-4C.**

Peak (cm <sup>-1</sup> )	Presumed Origin	Ref. Peak (cm <sup>-1</sup> )	Ref.
612	Cholesterol ester	614	[42,43]
648	Ring, cyclic deformation	649	[43]
677	Ring breathing in guanine	678	[42,43]
756	Symmetric breathing of tryptophan	755 *	[43]
835	Deformative vibrations of amine groups	838 *	[42,43]
852	Proline, hydroxyproline, tyrosine	852	[42,43]
935	C-C stretching mode of proline, valine, and protein backbone ( $\alpha$ -helix); glycogen	935 *	[42,43]
987	Phenylalanine	991	[42]
1032	CH <sub>2</sub> /CH <sub>3</sub> bending modes of phenylalanine and proline of collagen, phospholipids	1032	[42,43]
1076	Symmetric stretching of PO <sub>4</sub> <sup>3-</sup>	1076	[42]
1166	Lipids	1168	[42,43]
1209	C-C <sub>6</sub> H <sub>5</sub> stretching mode in tryptophan and phenylalanine	1209	[42,43]
1248	Amide III	1248	[42]
1262	Ring breathing mode in thymine, adenine; =C-H bending in proteins	1263	[42,43]
1300	CH <sub>2</sub> twisting in lipids, fatty acids	1300	[42,43]
1338	Amide III	1338	[43]
1362	Tryptophan	1360	[42,43]
1386	CH <sub>3</sub> band	1386	[42,43]
1425	Deoxyribose	1424	[42,43]
1466	Lipids	1465	[42,43]
1473	C=N stretching	1470	[42,43]
1577	Guanine, adenine	1578	[42]

\* Also present in control (hIOSE) spectrum.

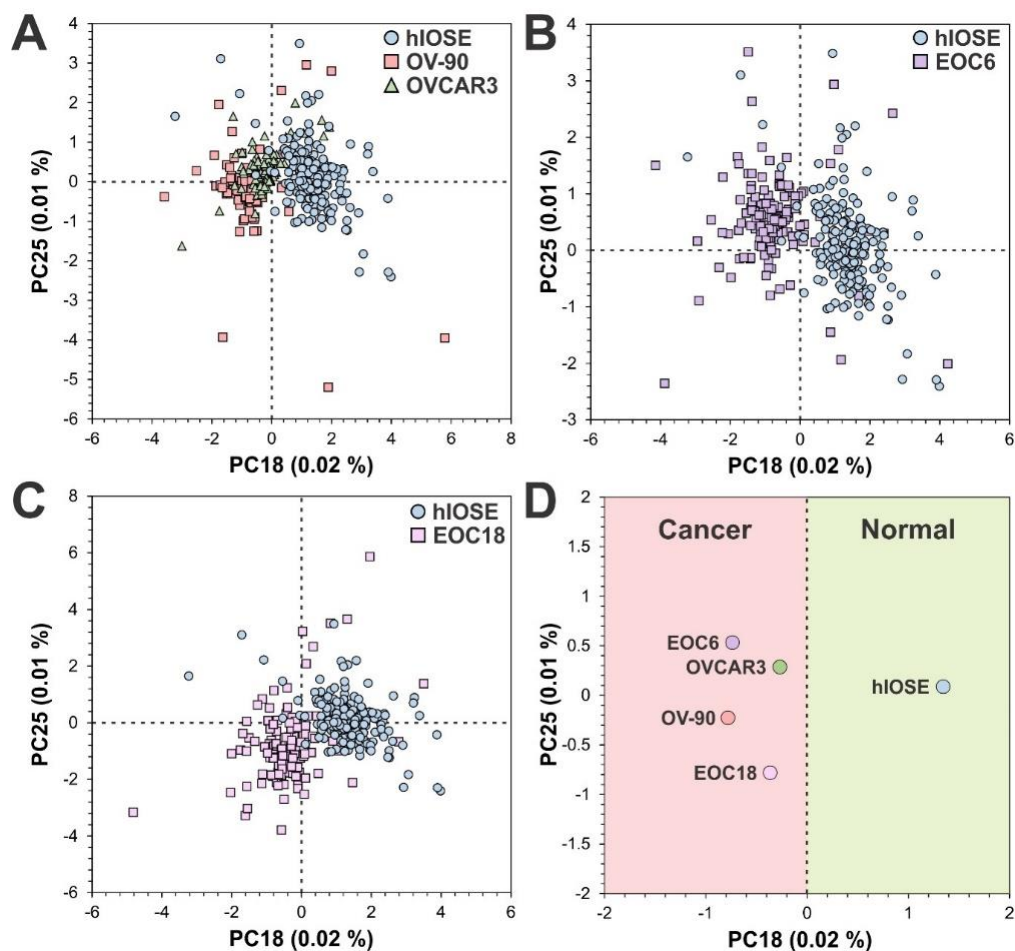


**Figure 4-4: Average SERS spectra of (A) ovarian epithelial cell line-derived EVs (control), (B) established OvCa cell line-derived EVs, and (C) primary OvCa cell line-**

**derived EVs. Average SERS spectra of hIOSE, OV-90, OVCAR3, EOC6, and EOC18 are comprised of 207, 123, 106, 166, and 156 spectra, respectively.**

### 4.3.3 Spectral Analysis by PCA

Often, the first two PCs are used to compare different classes of samples as they contain most of the information of the original data set. However, sometimes the first two PCs alone are not enough to differentiate classes of samples.<sup>[44]</sup> In our case, the PC18 and PC25 scores best separated the EV types, although they account for a very small percentage of the original data set (**Figure 4-5A-C**). These PCs were determined by the informative projections feature in Orange software, which ranks attribute pairs by classification accuracy. PC18 best differentiates the normal group from the cancer groups, where the latter tends to be associated with more negative PC18 scores. (**Figure 4-5D**). PC25 interestingly best separates the cancer groups in terms of the severity of cancer, where higher-grade cancers (e.g., EOC6, OVCAR3, OV-90) tend to be associated with more positive PC25 scores.



**Figure 4-5: PC score plots comparing (A) normal (hIOSE) vs. cancer (OV-90, OVCAR3) EVs, (B) normal vs. high-grade cancer (EOC6) EVs, (C) normal vs. low-grade cancer (EOC18) EVs, and (D) all groups to highlight separation along PC18. In (A) – (C), each point corresponds to one SERS spectrum while the points in (D) correspond to the centroids of each group in (A) – (C).**

PC loading spectra of PC18 and PC25 scores were compared to the average spectrum of each EV type presented in **Figure 4-4** to interpret which spectral peaks are responsible for the most variance in the data set (**Figure 4-6**). PC18 is best described by protein assignments at  $728\text{ cm}^{-1}$  (tryptophan) and  $1237\text{ cm}^{-1}$  (amide III), nucleic acid assignments at  $782\text{ cm}^{-1}$  (thymine, cytosine, uracil),  $1180\text{ cm}^{-1}$  (cytosine, guanine), and  $1483\text{ cm}^{-1}$  (guanine, adenine), and a carbohydrate peak at  $942\text{ cm}^{-1}$  (polysaccharides) (**Figure 4-6A**, **Table 4-7**). The more positive PC18 peaks at  $782\text{ cm}^{-1}$ ,  $1181\text{ cm}^{-1}$ , and  $1483\text{ cm}^{-1}$  tend to be more associated with the normal EVs than the OvCa EVs (i.e., all three peaks are found



**Table 4-7: SERS peaks that best differentiate normal vs. cancer groups, as highlighted in Figure 4-7A.**

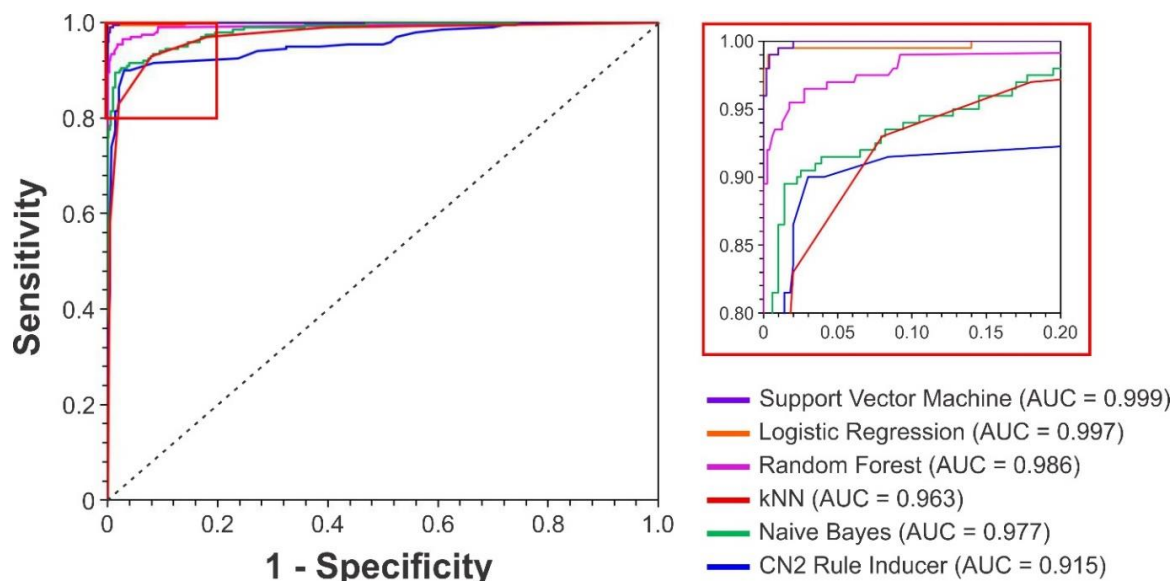
Peak (cm <sup>-1</sup> )	Presumed Origin	Ref. Peak (cm <sup>-1</sup> )	Ref.
728	Ring breathing of tryptophan	728	[42]
782	Thymine, cytosine, uracil ring breathing modes	782	[42,43]
942	Skeletal modes of polysaccharides	941	[42,43]
1181	Cytosine, guanine	1180	[42,43]
1237	Amide III, CH <sub>2</sub> wagging vibrations from glycine backbone and proline side chains	1237	[42,43]
1483	Guanine, adenine ring breathing modes	1485	[42,43]

**Table 4-8: SERS peaks that best differentiate low-grade vs. high-grade cancer groups, as highlighted in Figure 4-7B.**

Peak (cm <sup>-1</sup> )	Presumed Origin	Ref. Peak (cm <sup>-1</sup> )	Ref.
743	DNA, tryptophan	742	[42]
832	Asymmetric O-P-O stretching, tyrosine	831	[42,43]
940	Skeletal modes of polysaccharides	941	[42,43]
973	CH <sub>3</sub> , CCH vibrations in proteins	973	[42,43]
1060	C-C in-plane bending, C-N stretching, ceramide	1061	[42]
1170	C-H in-plane bending mode of tyrosine	1170	[42,43]

#### 4.3.4 Machine Learning for Ovarian Cancer Diagnosis

A logistic regression-based machine learning algorithm was used to classify normal and OvCa EVs based on the first 25 PCs calculated. Several algorithms were tested and compared based on the AUC values of their ROC curves (**Figure 4-7**). Logistic regression was chosen for the high AUC obtained as well as extremely high accuracy, precision, and recall, although some algorithms like SVM worked almost as well (**Table 4-9**). The algorithm was trained and tested with 5-fold CV to reduce potential overfitting.



**Figure 4-7: ROC curves comparing six different machine learning algorithms: SVM, logistic regression, random forest, kNN, Naïve Bayes, and CN2 rule inducer. The upper left-most portion is zoomed in and highlighted in red (right).**

**Table 4-9: Comparison of accuracies, sensitivities, and specificities achieved with the six machine learning algorithms shown in Figure 4-8.**

Model	Accuracy	Precision	Recall
Logistic Regression	98.6 %	98.6 %	98.6 %
SVM	97.3 %	97.4 %	97.3 %
Random Forest	91.0 %	91.1 %	91.0 %
Naïve Bayes	86.5 %	88.4 %	86.5 %
kNN	86.5 %	87.1 %	86.5 %
CN2 Rule Inducer	77.4 %	77.4 %	77.4 %

Four groups of data were compared: (1) normal (hIOSE) EVs vs. established OvCa cell line (OV-90, OVCAR3) EVs (**Figure 4-8A**), (2) normal EVs vs. high-grade OvCa (EOC6) EVs (**Figure 4-8B**), (3) normal EVs vs. low-grade OvCa (EOC18) EVs (**Figure 4-8C**), and (4) low-grade vs. high-grade OvCa EVs (**Figure 4-8D**). Heat maps shown in **Figure 4-8** were created to visualize the machine learning output scores of each sample. A vast majority of OvCa EVs were successfully classified as cancerous, as indicated by the red

and dark orange bars in **Figure 4-8A-C**. Similarly, a majority of the EOC6 EVs were successfully classified as high-grade, also indicated by the red and dark orange bars in **Figure 4-8D**.

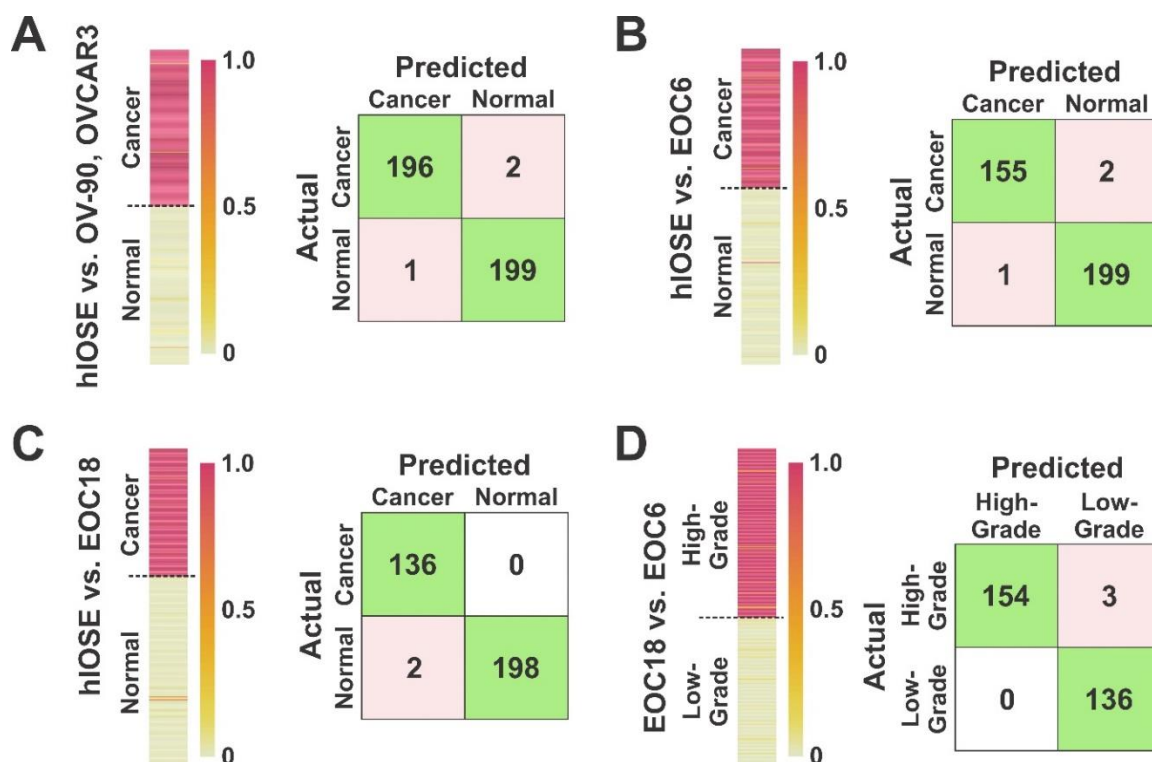
EVs isolated from hIOSE were compared to EVs isolated from OV-90, OVCAR3, and EOC6 to determine the platform's capability of correctly diagnosing OvCa. When the established cell line EVs were used as training sets for the cancer group, 196 out of 198 OvCa EV samples were correctly identified as cancerous, while 199 out of 200 normal EV samples were correctly identified as non-cancerous (**Figure 4-8A**), corresponding to 99.2 % classification accuracy, 99.0 % sensitivity, and 99.5 % specificity. When the EOC6 EVs were used as a training set for a cancer group, 155 out of 157 OvCa samples were correctly identified as cancerous, while 199 out of 200 normal EVs were again correctly classified as non-cancerous (**Figure 4-8B**). The resulting classification accuracy was identical to the normal vs. established OvCa cell line model at 99.2 %. The sensitivity dipped slightly to 98.7 %, but the specificity remained the same at 99.5 %.

EVs isolated from hIOSE were compared to EVs isolated from EOC18 to determine the platform's capability of correctly diagnosing low-grade OvCa, which can be more of a challenge compared to high-grade OvCa since low-grade OvCa cells tend to more closely resemble normal cells. With the EOC18 EVs as a training set for the cancer group, all 136 OvCa samples were correctly identified as cancerous, while 198 out of 200 normal EVs were correctly classified as non-cancerous (**Figure 4-8C**). The 99.4 % accuracy achieved here is comparable to the previous two models, suggesting we can successfully classify both low-grade and high-grade OvCa samples with extremely high accuracy. Furthermore, we were also able to achieve high sensitivity and specificity at 100 % and 99.0 %, respectively.

EVs isolated from EOC6 and EOC18 were compared to determine the platform's capability of predicting whether an OvCa sample is high-grade or low-grade, since the grade of OvCa can help predict the prognosis of the disease as well as how the cancer may respond to treatment. 154 out of 157 EOC6 samples were correctly identified as high-grade, while all



136 EOC18 samples were correctly identified as low-grade, corresponding to 99.0 % accuracy, 98.1 % sensitivity, and a 100 % specificity (Figure 4-8D).



**Figure 4-8: Heat maps (left) to visualize output scores of (A) normal vs. established cancer cell line EVs, (B) normal vs. high-grade cancer EVs, (C) normal vs. low-grade cancer EVs, and (D) low-grade vs. high-grade cancer EVs, with corresponding confusion matrices (right). In the heat maps, each horizontal bar corresponds to one SERS spectrum. Groups labeled on the left indicate the origin of the sample, while the color of the bar indicates the algorithm's prediction. In (A) – (C), a red or dark orange color (output score > 0.5) corresponds to a cancer prediction and a yellow or light orange color (output score < 0.5) corresponds to a normal prediction. In (D), a red or dark orange color (output score > 0.5) corresponds to a high-grade cancer prediction and a yellow or light orange color (output score < 0.5) corresponds to a low-grade cancer prediction.**

The accuracies, sensitivities, and specificities achieved with each of the four models compared in Figure 4-8 is summarized in Table 4-10. The high accuracies presented here indicate that this platform and methodology is not only promising in terms of ovarian

cancer diagnosis, but also promising in determining the grade of disease. Since EVs are found in bodily fluids such as urine, plasma, and saliva, this approach has a strong potential for non-invasive ovarian cancer diagnosis.

**Table 4-10: Comparison of accuracies, sensitivities, and specificities achieved with each group compared in Figure 4-8.**

Model	Accuracy	Sensitivity	Specificity
hIOSE vs. OV-90, OVCAR3	99.2 %	99.0 %	99.5 %
hIOSE vs. EOC6	99.2 %	98.7 %	99.5 %
hIOSE vs. EOC18	99.4 %	100 %	99.0 %
EOC18 vs. EOC6	99.0 %	98.1 %	100 %

#### 4.4 Conclusion

In conclusion, we have presented plasmonic gold nanohole arrays for the capture of single EVs which we subsequently characterized by SERS. To our knowledge, this is the first time EVs from these five cell lines (OV-90, OVCAR3, EOC6, EOC18, and hIOSE) have been characterized by SERS. We then were able to determine their main compositional differences by PCA. We found that normal EVs could generally be differentiated from OvCa EVs by the presence of peaks at  $782\text{ cm}^{-1}$  (thymine, cytosine, uracil ring breathing modes),  $1181\text{ cm}^{-1}$  (cytosine, guanine), and  $1483\text{ cm}^{-1}$  (guanine, adenine ring breathing modes). Interestingly, high-grade cancer could also be predicted based on the presence of peaks at  $940\text{ cm}^{-1}$  (polysaccharides),  $1060\text{ cm}^{-1}$  (C-C in-plane bending, C-N stretching, ceramide), and  $1170\text{ cm}^{-1}$  (tyrosine).

The PC scores calculated were then used as classifiers in a logistic regression-based machine learning algorithm, which was able to differentiate normal EVs from the established OvCa cell line EVs with 99.2 % accuracy, 99.0 % sensitivity, and 99.5 % specificity. Normal EVs could also be discriminated from the high-grade primary cell line EVs with 99.2 % accuracy, 98.7 % sensitivity, and 99.5 % specificity. Interestingly, we were also able to differentiate the normal EVs from the low-grade primary cell line EVs with 99.4 % accuracy, 100 % sensitivity, and 99.0 % specificity. Lastly, the low-grade and high-grade primary cell line EVs could also be discriminated with 99.0 % accuracy, 98.1

% sensitivity, and 100 % specificity. The diagnostic potential presented here is a great step towards early, non-invasive, facile, and rapid ovarian cancer diagnosis. However, at this stage in the study, the results are proof-of-concept; until clinical samples are obtained from ovarian cancer patients and from healthy individuals, we cannot determine precisely how the model will perform with new test data. Therefore, the next step in this study is to challenge the platforms with EVs from clinical samples to determine the feasibility of using this methodology in a clinical setting.

## 4.5 References

1. Siegel, R. L.; Miller, K. D.; Fuchs, H. E.; Jemal, A. *CA Cancer J. Clin.* **2021**, *71*, 7-33.
2. Peres, L. C.; Cushing-Haugen, K. L.; Köbel, M.; Harris, H. R.; Berchuck, A.; Rossing, M. A.; Schildkraut, J. M.; Doherty, J. A. *J. Natl. Cancer Inst.* **2019**, *111*, 60-68.
3. Lheureux, S.; Braunstein, M.; Oza, A. M. *CA Cancer J. Clin.* **2019**, *69*, 280-304.
4. Reid, B. M.; Permeth, J. B.; Sellers, T. A. *Cancer Biol. Med.* **2017**, *14*, 9-32.
5. Lucidi, A.; Buca, D.; Ronsini, C.; Tinari, S.; Bologna, G.; Buca, D.; Leombroni, M.; Liberati, M.; D'Antonio, F.; Scambia, G., et al. *Int. J. Mol. Sci.* **2020**, *21*, 8762.
6. An, T.; Qin, S.; Xu, Y.; Tang, Y.; Huang, Y.; Situ, B.; Inal, J. M.; Zheng, L. *J. Extracell. Vesicles* **2015**, *4*, 27522.
7. Fujita, Y.; Yoshioka, Y.; Ochiya, T. *Cancer Sci.* **2016**, *107*, 385-390.
8. Kinoshita, T.; Yip, K. W.; Spence, T.; Liu, F. F. *J. Hum. Genet.* **2017**, *62*, 67-74.
9. Rojalín, T.; Phong, B.; Koster, H. J.; Carney, R. P. *Front. Chem.* **2019**, *7*, 279.
10. Li, G.; Zhu, N.; Zhou, J.; Kang, K.; Zhou, X.; Ying, B.; Yi, Q.; Wu, Y. *J. Mater. Chem. B* **2021**, *9*, 2709-2716.
11. Zong, S.; Wang, L.; Chen, C.; Lu, J.; Zhu, D.; Zhang, Y.; Wang, Z.; Cui, Y. *Anal. Methods* **2016**, *8*, 5001-5008.

12. Kwizera, E. A.; O'Connor, R.; Vinduska, V.; Williams, M.; Butch, E. R.; Snyder, S. E.; Chen, X.; Huang, X. *Theranostics* **2018**, *8*, 2722-2738.
13. Lee, J. U.; Kim, W. H.; Lee, H. S.; Park, K. H.; Sim, S. J. *Small* **2019**, *15*, 1804968.
14. Kim, N.; Thomas, M. R.; Bergholt, M. S.; Pence, I. J.; Seong, H.; Charchar, P.; Todorova, N.; Nagelkerke, A.; Belessiotis-Richards, A.; Payne, D. J., et al. *Nat. Commun.* **2020**, *11*, 207.
15. Kang, T.; Zhu, J.; Luo, X.; Jia, W.; Wu, P.; Cai, C. *Anal. Chem.* **2021**, *93*, 2519-2526.
16. Ferreira, N.; Marques, A.; Águas, H.; Bandarenka, H.; Martins, R.; Bodo, C.; Costa-Silva, B.; Fortunato, E. *ACS Sens.* **2019**, *4*, 2073-2083.
17. Park, J.; Hwang, M.; Choi, B.; Jeong, H.; Jung, J.-H.; Kim, H. K.; Hong, S.; Park, J.-H.; Choi, Y. *Anal. Chem.* **2017**, *89*, 6695-6701.
18. Shin, H.; Jeong, H.; Park, J.; Hong, S.; Choi, Y. *ACS Sens.* **2018**, *3*, 2637-2643.
19. Shin, H.; Oh, S.; Hong, S.; Kang, M.; Kang, D.; Ji, Y.-G.; Choi, B. H.; Kang, K.-W.; Jeong, H.; Park, Y., et al. *ACS Nano* **2020**, *14*, 5435-5444.
20. Sivashanmugan, K.; Huang, W.-L.; Lin, C.-H.; Liao, J.-D.; Lin, C.-C.; Su, W.-C.; Wen, T.-C. *J. Taiwan Inst. Chem. Eng.* **2017**, *80*, 149-155.
21. Yan, Z.; Dutta, S.; Liu, Z.; Yu, X.; Mesgarzadeh, N.; Ji, F.; Bitan, G.; Xie, Y.-H. *ACS Sens.* **2019**, *4*, 488-497.
22. Dong, S.; Wang, Y.; Liu, Z.; Zhang, W.; Yi, K.; Zhang, X.; Zhang, X.; Jiang, C.; Yang, S.; Wang, F., et al. *ACS Appl. Mater. Interfaces* **2020**, *12*, 5136-5146.
23. Avella-Oliver, M.; Puchades, R.; Wachsmann-Hogiu, S.; Maquieira, A. *Sens. Actuators B Chem.* **2017**, *252*, 657-662.
24. Li, T.-D.; Zhang, R.; Chen, H.; Huang, Z.-P.; Ye, X.; Wang, H.; Deng, A.-M.; Kong, J.-L. *Chem. Sci.* **2018**, *9*, 5372-5382.

25. Carmicheal, J.; Hayashi, C.; Huang, X.; Liu, L.; Lu, Y.; Krasnoslobodtsev, A.; Lushnikov, A.; Kshirsagar, P. G.; Patel, A.; Jain, M., et al. *Nanomedicine* **2019**, *16*, 88-96.
26. Jiang, S.; Li, Q.; Wang, C.; Pang, Y.; Sun, Z.; Xiao, R. *ACS Sens.* **2021**, *6*, 852-862.
27. Kerr, L. T.; Gubbins, L.; Gorzel, K. W.; Sharma, S.; Kell, M.; McCann, A.; Hennelly, B. M., Raman Spectroscopy and SERS Analysis of Ovarian Tumour Derived Exosomes (tEXs): A Preliminary Study. In *Proc. SPIE 9129*, Biophotonics: Photonic Solutions for Better Health Care IV, 91292Q: 2014.
28. Lee, C.; Carney, R.; Lam, K.; Chan, J. W. *J. Raman Spectrosc.* **2017**, *48*, 1771-1776.
29. Lee, C.; Carney, R. P.; Hazari, S.; Smith, Z. J.; Knudson, A.; Robertson, C. S.; Lam, K. S.; Wachsmann-Hogiu, S. *Nanoscale* **2015**, *7*, 9290-9297.
30. Rojalin, T.; Koster, H. J.; Liu, J.; Mizenko, R. R.; Tran, D.; Wachsmann-Hogiu, S.; Carney, R. P. *ACS Sens.* **2020**, *5*, 2820-2833.
31. Kaufman, L.; Cooper, T.; Wallace, G.; Hawke, D.; Betts, D.; Hess, D.; Lagugné-Labarthe, F., Trapping and SERS Identification of Extracellular Vesicles Using Nanohole Arrays. In *Proc. SPIE 10894*, Plasmonics in Biology and Medicine XVI, 108940B: 2019.
32. Culum, N. M.; Cooper, T. T.; Bell, G. I.; Hess, D. A.; Lagugné-Labarthe, F. *Anal. Bioanal. Chem.* **2021**, *413*, 5013-5024.
33. Wessel, D.; Flügge, U. I. *Anal. Biochem.* **1984**, *138*, 141-143.
34. Gomes, J.; Lucien, F.; Cooper, T. T.; Kim, Y.; Williams, K. C.; Liao, X.; Kaufman, L.; Lagugné-Labarthe, F.; Kenyon, O.; Boysen, J., et al. *Thromb. Haemost.* **2018**, *118*, 1612-1624.
35. Raposo, G.; Stoorvogel, W. *J. Cell Biol.* **2013**, *200*, 373-383.
36. Royo, F.; Théry, C.; Falcón-Pérez, J. M.; Nieuwland, R.; Witwer, K. W. *Cells* **2020**, *9*, 1955.

37. Bobrie, A.; Colombo, M.; Krumeich, S.; Raposo, G.; Théry, C. *J. Extracell. Vesicles*. **2012**, *1*, 18397.
38. Cooper, T. T.; Sherman, S. E.; Bell, G. I.; Dayarathna, T.; McRae, D. M.; Ma, J.; Lagugné-Labarthe, F.; Pasternak, S. H.; Lajoie, G. A.; Hess, D. A. *Stem Cells Dev.* **2021**, *30*, 247-264.
39. Théry, C.; Witwer, K. W.; Aikawa, E.; Alcaraz, M. J.; Anderson, J. D.; Andriantsitohaina, R.; Antoniou, A.; Arab, T.; Archer, F.; Atkin-Smith, G. K., et al. *J. Extracell. Vesicles* **2018**, *7*, 1535750.
40. Choi, D. S.; Kim, D. K.; Kim, Y. K.; Gho, Y. S. *Mass Spectrom. Rev.* **2015**, *34*, 474-490.
41. Kugeratski, F. G.; Hodge, K.; Lilla, S.; McAndrews, K. M.; Zhou, X.; Hwang, R. F.; Zanivan, S.; Kalluri, R. *Nat. Cell Biol.* **2021**, *23*, 631-641.
42. Talari, A. C. S.; Movasaghi, Z.; Rehman, S.; Rehman, I. U. *Appl. Spectrosc. Rev.* **2015**, *50*, 46-111.
43. Rehman, I. U.; Movasaghi, Z.; Rehman, S. FTIR and Raman Characteristic Peak Frequencies in Biological Studies. In *Vibrational Spectroscopy for Tissue Analysis*, CRC Press: Boca Raton, 2012; Vol. 1, pp 213-294.
44. Bukva, M.; Dobra, G.; Gomez-Perez, J.; Koos, K.; Harmati, M.; Gyukity-Sebestyen, E.; Biro, T.; Jenei, A.; Kormondi, S.; Horvath, P., et al. *Cancers* **2021**, *13*, 1407.

## Chapter 5

### 5 Conclusions and Outlook

As highlighted in the introduction of this thesis, EVs can generally be split into two groups based on their cellular origins and applications: as cell-free therapeutic agents in regenerative medicine, and as biomarkers for various diseases, including cancer. The main research goal of this work was to fabricate SERS substrates for EV capture and characterization, and EVs from each of the aforementioned groups were studied.

The first chapter of this thesis provided background information of EVs, including their functions and importance, and established the research interest of EVs in both therapeutic and diagnostic applications. Several standard methods for EV characterization were reviewed in this chapter, encompassing both physical (e.g., Brownian motion-based and microscopic techniques) and biochemical methods (e.g., MS-based and immunoassay techniques). An overview of nanoplasmonic approaches was also introduced, including SPR sensing, LSPR sensing, and SERS. Since SERS was the main technique implemented for EV characterization in this thesis, it was discussed further in Chapter 2.

The second chapter of this thesis delved into the fundamentals of Raman spectroscopy and provided background information of the types of spectral fingerprints that could be expected for biological systems arising from biological macromolecules (i.e., proteins, nucleic acids, lipids, and carbohydrates). The main limitation of Raman spectroscopy, low sensitivity, was highlighted, and the need for SERS analysis was presented. In addition to providing the physical background behind SERS sensing, several SERS studies of EVs conducted in recent years were also reviewed. Special interest was taken in the types of probes and substrates used (e.g., nanoparticle-based, nanorod-based, etc.), the type of detection method (i.e., direct or indirect), and their performance in cancer diagnostics. Also explained in detail in this chapter was PCA for its use in EV analysis in Chapters 3 and 4, and an overview of machine learning and its various techniques was given.

The third chapter of this thesis was focused on the characterization of EVs derived from BM- and Panc-MSCs, which are used in therapeutic applications (i.e., regenerative

medicine). The SERS platform consisted of gold nanohole arrays of various size (100 – 1000 nm) and shape (triangles, circles, squares) and were fabricated by EBL. The nanohole arrays provided LSPR modes around 650 – 690 nm and 750 – 780 nm. For SERS analysis, the arrays were best matched with a 785 nm excitation wavelength laser. SERS spectra for each EV group were reported and characterized. PCA was used initially to elucidate trends in the EV groups and to explore the discriminatory power of the platform. While PCA was able to determine the peaks responsible for most variance in the data set, it did not perform well in classification tasks. Thus, various machine learning methods were explored, using the PCs as inputs, to classify the BM- and Panc-MSV EVs. The optimal machine learning algorithm determined for this data set was logistic regression, which could classify the EVs with 89 % accuracy, 89 % sensitivity, and 88 % specificity.

The fourth chapter of this thesis explored ovarian cancer diagnosis via SERS characterization of cancer-derived EVs. The gold nanohole arrays presented and characterized in Chapter 3 were used again for EV trapping and SERS analysis. EVs derived from two established OvCa cell lines (OV-90 and OVCAR3) and two primary OvCa cell lines (EOC6 and EOC18) were investigated. As a normal control, EVs derived from one human immortalized ovarian surface epithelial cell line (hIOSE) were used. PCA was used again to provide insight to the compositional differences among the EVs, and as expected, failed at classification tasks as observed in Chapter 3. However, the machine learning algorithms tested in this chapter performed much better at EV classification using this data set than that from Chapter 3. The logistic regression-based algorithm again performed the best out of the various models tested. The model was used to classify four sets of data, which contained (1) normal and established cell line-derived EVs, (2) normal and high-grade cancer (EOC6) EVs, (3) normal and low-grade cancer (EOC18) EVs, and (4) high- and low-grade cancer EVs. The accuracies, sensitivities, and specificities achieved with such models ranged from 99.0 – 99.4 %, 98.1 – 100 %, and 99.0 – 100 %, respectively. Therefore, there is great potential in using this methodology for ovarian cancer diagnosis in a clinical setting.

While EVs were successfully characterized using nanohole array substrates, there are many areas of this project that could be expanded on and further investigated. Of course, before

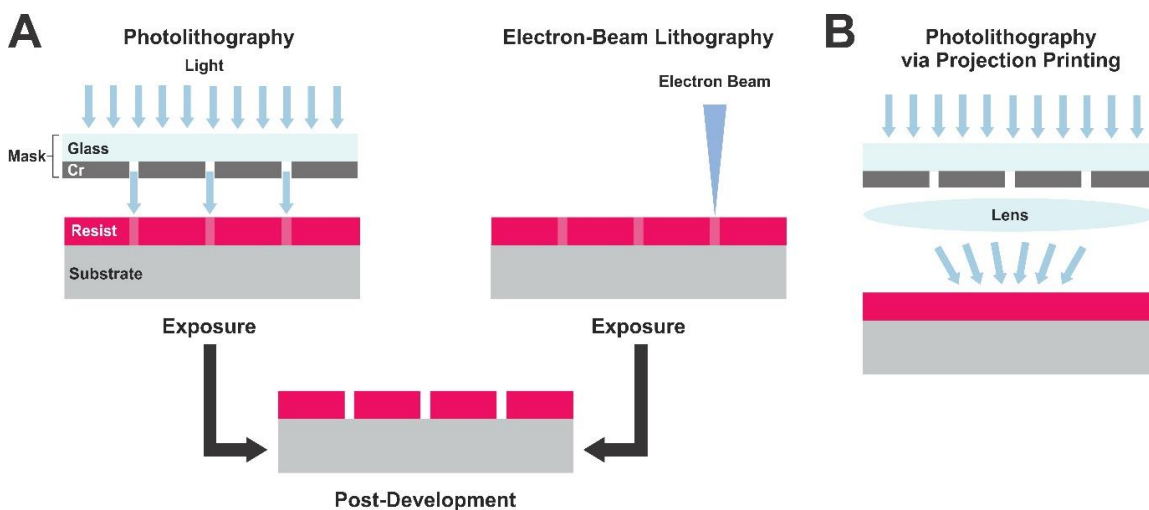


such a methodology could be used in a clinical setting, the platforms must be tested with clinical samples from cancer patients and healthy individuals. Samples from such patients would be gathered in a non- or minimally-invasive fashion (e.g., blood, saliva, urine, etc.). However, since these arrays are unfunctionalized, they are unable to specifically capture EVs from cancerous sources. Therefore, at the moment, analysis would require prior isolation of EVs from the samples rather than direct detection. There are also areas of fabrication that could be further developed and optimized, ranging from the nanofabrication technique, to device functionalization, to implementation in a microfluidic device.

Nanofabrication by EBL is advantageous in that patterns are easily customizable, which is especially useful in research to determine optimal device dimensions and conditions. However, EBL is a very time-consuming and low-throughput process compared to other nanofabrication techniques.<sup>[1]</sup> Fortunately, once device parameters are optimized, EBL can be used to fabricate a mask that can be used in higher-throughput photolithography, which has been applied for nanohole array fabrication in literature.<sup>[2,3]</sup> The photolithography mask, which contains a master pattern, consists of opaque features (e.g., chrome-based) on a transparent substrate (e.g., quartz).<sup>[4]</sup> The principles of photolithography and EBL are very similar, but the former usually uses ultraviolet (UV) light in the 193 – 436 nm range to induce changes in the resist while the latter patterns the resist with an electron beam (**Figure 5-1A**).<sup>[1]</sup> Since each feature is exposed at the same time during photolithography, patterns can be written much faster than in EBL, in which features are written one at a time.

Photolithography can be achieved by contact, proximity, or projection printing. While the most commonly used photolithographic methods are contact and proximity printing for highest throughput, they offer only moderate resolution with a minimum feature size around 2 – 3  $\mu\text{m}$ .<sup>[5]</sup> For EV trapping and sensing applications where smaller features are required, projection printing is best-suited, which allows for resolution as small as 37 nm while still maintaining high throughput.<sup>[6]</sup> Unlike contact and proximity printing for photolithography, projection printing requires a sophisticated optical lens system to project a deep-UV pattern from an excimer laser, usually of 193 nm or 248 nm wavelength (**Figure 5-1B**).<sup>[7]</sup> For comparison, a wafer with a diameter of approximately 7.62 cm can be

completely exposed in less than 1 minute with photolithography,<sup>[5]</sup> while a  $50\ \mu\text{m} \times 50\ \mu\text{m}$  patch of nanohole arrays made for this thesis required 1 – 2 minutes of exposure time with EBL. Photolithography would allow for the rapid and mass production of nanohole arrays, which would be ideal for clinical translation of the methodology, and warrants exploration.

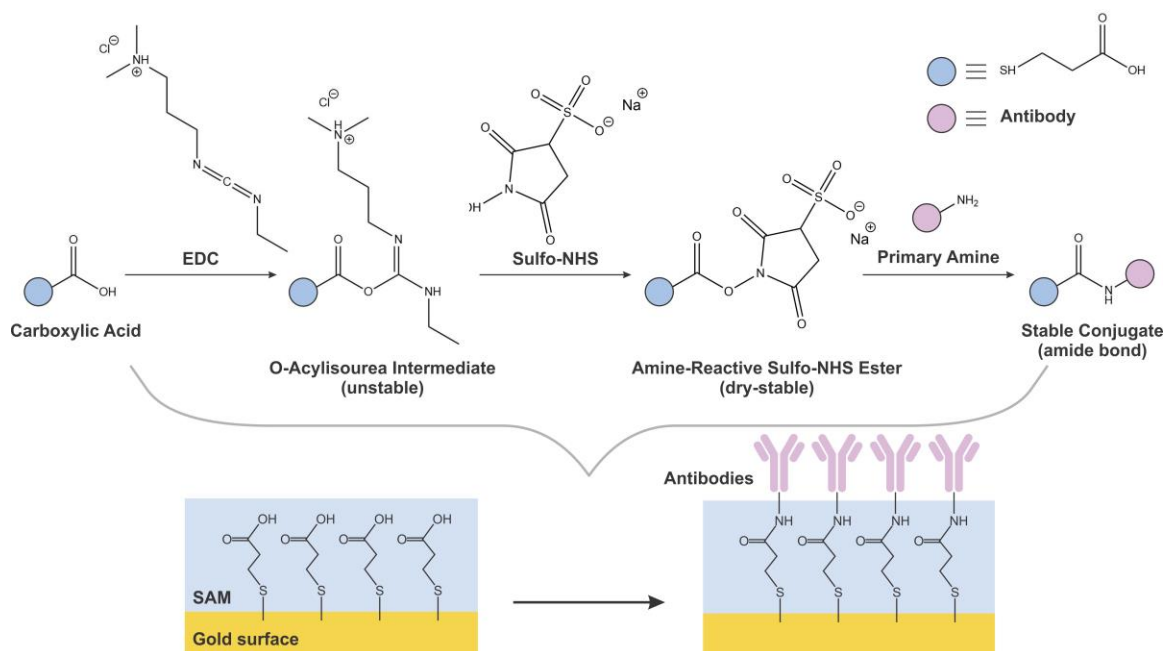


**Figure 5-1: (A) Schematic illustration comparing photolithography, with proximity printing shown as an example, and EBL for nanohole array fabrication. (B) Schematic illustration of photolithography by projection printing.**

Unfunctionalized arrays are advantageous in that rich, heterogeneous biological spectra can be investigated, but a clinical diagnostic setting may benefit from a more specific and efficient capture technique. SERS probes and platforms have been functionalized with cysteamine in literature, which allows for more efficient EV trapping via electrostatic interactions between the cationic cysteamine and anionic EV membrane.<sup>[8,9]</sup> Cysteamine is the simplest aminothiols and a degradation product of cysteine, which is inexpensive and widely commercially available.<sup>[10]</sup> Cysteamine is able to anchor itself to gold surfaces through its terminal thiol group. In an acidic solution (pH = 6.5), the terminal amine group of cysteamine is expected to be protonated, thus allowing for electrostatic interactions with EV surfaces as aforementioned.<sup>[10]</sup>

While cysteamine functionalization would allow for greater adhesion of EVs to the platform, it is not specific to cancer-derived EVs. Therefore, another avenue that could be explored for platform functionalization is immunocapture with antibodies. EpCAM is a

protein that has been identified as a diagnostic and prognostic biomarker in many carcinomas,<sup>[11]</sup> including ovarian cancer.<sup>[12]</sup> EpCAM has an overexpression rate of 55 % in mucinous ovarian cancer and a 76 % overexpression rate in serous and endometrioid ovarian cancers.<sup>[13]</sup> The nanohole arrays could thus be functionalized with anti-EpCAM for the more specific capture of ovarian-cancer derived EVs, and could be achieved by 1-ethyl-3-(3-dimethylaminopropyl)carbodiimide (EDC) crosslinker chemistry (**Figure 5-2**).<sup>[14]</sup> EDC crosslinking begins with activating carboxyl groups for the direct reaction with the primary amines of the antibody. Since this carboxyl-containing molecule should be immobilized onto the gold nanohole arrays, the acid should also contain a terminal thiol (e.g., 3-mercaptopropionic acid). When EDC reacts with carboxylic acid groups, an unstable intermediate (O-acylisourea) that can be displaced by nucleophilic attack from the primary amines is formed.<sup>[15]</sup> However, since it is unstable in aqueous solutions, it is recommended to introduce *N*-hydroxysuccinimide (NHS) or its water-soluble analog (sulfo-NHS). The sulfo-NHS group of this intermediate can then rapidly leave following the addition of amine nucleophiles, thus conjugating anti-EpCAM to the platform via a stable amide bond.<sup>[15]</sup> Anti-EpCAM has even been used in the immunocapture of breast cancer-<sup>[16]</sup> and prostate cancer-derived EVs.<sup>[17]</sup> The platform could thus benefit from more specific ovarian cancer-derived EV capture using anti-EpCAM, and it could also be tested with EVs from other cancer sources.



**Figure 5-2: Schematic illustration of antibody functionalization by EDC chemistry.**

Finally, once optimal fabrication and functionalization conditions are identified, the platform can be embedded in a microfluidic device. A microfluidic device would allow for the probing of EVs more similar to their native state (i.e., in liquid), rather than dry on the platform, which can introduce unwanted salt formation. The analyte flow can be controlled by a dielectric<sup>[18]</sup> or acoustic force.<sup>[19]</sup> Since EVs can be isolated and captured in the microfluidic chamber by both size-based<sup>[20-22]</sup> and immunoaffinity-based methods,<sup>[23-25]</sup> the nanohole arrays can be integrated either unfunctionalized, as was presented in Chapters 3 and 4, or functionalized, as proposed in this chapter. Microfluidics is of interest in the precision medical field because it utilizes a small number of samples,<sup>[26]</sup> allows samples to be recycled,<sup>[27]</sup> and can be used in various analysis methods, including SERS.<sup>[28]</sup>

The goal of this work is to develop a diagnostic device for the minimally-invasive, rapid, and early diagnosis of cancer. The preliminary work presented in this thesis displayed the potential of using gold nanohole arrays for cancer detection by SERS, but further tailoring of the device and testing with clinical samples is required before the device can be used on its own in real diagnosis applications. We hope that the work presented in this thesis will one day translate to point-of-care cancer detection and diagnosis.

## 5.1 References

1. Pimpin, A.; Srituravanich, W. *Eng. J.* **2011**, *16*, 37-56.
2. Debbrecht, B.; McElhiney, M.; Carey, V.; Cullen, C.; Mirotznik, M. S.; DeLacy, B. G. *Optics Express* **2017**, *25*, 24501-24511.
3. Baek, S. H.; Lee, S.; Bae, J.-H.; Hong, C.-W.; Park, M.-J.; Park, H.; Baek, M.-C.; Nam, S.-W. *Mater. Res. Express* **2020**, *7*, 035008.
4. Bandelier, P.; Charley, A.-L.; Lagrange, A. Photolithography. In *Lithography: Main Techniques*, 1st ed.; Landis, S., Ed. ISTE Ltd.: London, 2011; pp 1-40.
5. Pease, F.; Chou, S. *Proc. IEEE* **2008**, *96*, 248-270.
6. Gates, B. D.; Xu, Q.; Stewart, M.; Ryan, D.; Willson, C. G.; Whitesides, G. M. *Chem. Rev.* **2005**, *105*, 1171-1196.
7. Besacier, M.; Constancias, C.; Robic, J.-Y. Extreme Ultraviolet Lithography. In *Lithography: Main Techniques*, 1st ed.; Landis, S., Ed. ISTE Ltd.: London, 2011; pp 41-100.
8. Shin, H.; Jeong, H.; Park, J.; Hong, S.; Choi, Y. *ACS Sens.* **2018**, *3*, 2637-2643.
9. Rojalin, T.; Koster, H. J.; Liu, J.; Mizenko, R. R.; Tran, D.; Wachsmann-Hogiu, S.; Carney, R. P. *ACS Sens.* **2020**, *5*, 2820-2833.
10. Ma, Y.; Jiang, L.; Mei, Y.; Song, R.; Tian, D.; Huang, H. *Analyst* **2013**, *138*, 5338-5343.
11. Went, P. T. H.; Lugli, A.; Meier, S.; Bundi, M.; Mirlacher, M.; Sauter, G.; Dirnhofer, S. *Hum. Pathol.* **2004**, *35*, 122-128.
12. Vorobyeva, A.; Konovalova, E.; Xu, T.; Schulga, A.; Altai, M.; Garousi, J.; Rinne, S. S.; Orlova, A.; Tolmachev, V.; Deyev, S. *Int. J. Mol. Sci.* **2020**, *21*, 3310.

13. Spizzo, G.; Fong, D.; Wurm, M.; Ensinger, C.; Obrist, P.; Hofer, C.; Mazzoleni, G.; Gastl, G.; Went, P. *J. Clin. Pathol.* **2011**, *64*, 415-420.
14. Wong, S. S.; Jameson, D. M. Monofunctional and Zero-Length Cross-Linking Reagents. In *Chemistry of Protein and Nucleic Acid Cross-Linking and Conjugation*, 2nd ed.; CRC Press: Boca Raton, 2012; pp 265-295.
15. Hermanson, G. T. Zero-Length Crosslinkers. In *Bioconjugation Techniques*, 2nd ed.; Elsevier: London, 2008; pp 215-233.
16. Kwizera, E. A.; O'Connor, R.; Vinduska, V.; Williams, M.; Butch, E. R.; Snyder, S. E.; Chen, X.; Huang, X. *Theranostics* **2018**, *8*, 2722-2738.
17. Beekman, P.; Enciso-Martinez, A.; Rho, H. S.; Pujari, S. P.; Lenferink, A.; Zuilhof, H.; Terstappen, L. W. M. M.; Otto, C.; Le Gac, S. *Lab Chip* **2019**, *19*, 2526-2536.
18. Barik, A.; Cherukulappurath, S.; Wittenberg, N. J.; Johnson, T. W.; Oh, S. H. *Anal. Chem.* **2016**, *88*, 1704-1710.
19. Wu, M.; Ouyang, Y.; Wang, Z.; Zhang, R.; Huang, P. H.; Chen, C.; Li, H.; Li, P.; Quinn, D.; Dao, M., et al. *Proc. Natl. Acad. Sci. U. S. A.* **2017**, *114*, 10584-10589.
20. Rho, J.; Chung, J.; Im, H.; Liong, M.; Shao, H.; Castro, C. M.; Weissleder, R.; Lee, H. *ACS Nano* **2013**, *7*, 11227-11233.
21. Lee, K.; Shao, H.; Weissleder, R.; Lee, H. *ACS Nano* **2015**, *9*, 2321-2327.
22. Wunsch, B. H.; Smith, J. T.; Gifford, S. M.; Wang, C.; Brink, M.; Bruce, R. L.; Austin, R. H.; Stolovitzky, G.; Astier, Y. *Nat. Nanotechnol.* **2016**, *11*, 936-940.
23. Kanwar, S. S.; Dunlay, C. J.; Simeone, D. M.; Nagrath, S. *Lab Chip* **2014**, *14*, 1891-900.
24. Zhao, Z.; Yang, Y.; Zeng, Y.; He, M. *Lab Chip* **2016**, *16*, 489-496.
25. Zhang, P.; He, M.; Zeng, Y. *Lab Chip* **2016**, *16*, 3033-3042.

26. Shin, H.; Seo, D.; Choi, Y. *Molecules* **2020**, *25*, 5209.
27. Iliescu, F. S.; Vrtačnik, D.; Neuzil, P.; Iliescu, C. *Micromachines* **2019**, *10*, 392.
28. Han, B.; Zhang, Y.-L.; Zhu, L.; Chen, X.-H.; Ma, Z.-C.; Zhang, X.-L.; Wang, J.-N.; Wang, W.; Liu, Y.-Q.; Chen, Q.-D., et al. *Sens. Actuators B Chem.* **2018**, *270*, 500-507.

## Appendix A: Confusion Matrix Calculations

A confusion matrix is a representation of the classification performance of a given algorithm (**Table A-1**).<sup>[1]</sup> There are four possible outcomes given a classifier and an instance. If a positive instance is correctly classified as positive, it is counted as a true positive (*TP*) outcome, but if it is misclassified as negative, it is counted as a false negative (*FN*). If a negative instance is correctly classified as negative, it is counted as a true negative (*TN*), but if it is misclassified as positive, it is counted as a false positive (*FP*).

**Table A-1: Example confusion matrix.**

		Predicted	
		Positive	Negative
Actual	Positive	<i>TP</i>	<i>FN</i>
	Negative	<i>FP</i>	<i>TN</i>

Sensitivity, which can also be referred to as recall, is a quantification of the true positive rate and is calculated by:<sup>[2]</sup>

$$sensitivity = \frac{TP}{TP+FN} \quad (\text{A-1})$$

Conversely, specificity is a quantification of the true negative rate and is calculated by:<sup>[2]</sup>

$$specificity = \frac{TN}{TN+FP} \quad (\text{A-2})$$

Precision is a quantification of the positive predictive value of an algorithm and is calculated by:<sup>[1]</sup>

$$precision = \frac{TP}{TP+FP} \quad (\text{A-3})$$

Lastly, accuracy is the proportion of all the correct predictions among the total population and is calculated by:<sup>[2]</sup>

$$accuracy = \frac{TP+TN}{TP+FN+FP+TN} \quad (\text{A-4})$$

## References

1. Fawcett, T. *Pattern Recognit. Lett.* **2006**, 27, 861-874.
2. Metz, C. E. *Semin. Nucl. Med.* **1978**, 8, 283-298.



## Appendix B: Copyrights



### Thank you for your order!

Dear Miss. Nina Culum,

Thank you for placing your order through Copyright Clearance Center's RightsLink® service.

#### Order Summary

Licensee: Miss. Nina Culum  
 Order Date: Aug 2, 2021  
 Order Number: 5120831136560  
 Publication: Biochimica et Biophysica Acta (BBA) - Biomembranes  
 Characterization of extracellular vesicles by IR spectroscopy:  
 Title: Fast and simple classification based on amide and CH stretching vibrations  
 Type of Use: reuse in a thesis/dissertation  
 Order Total: 0.00 CAD

### Thank you for your order!

Dear Miss. Nina Culum,

Thank you for placing your order through Copyright Clearance Center's RightsLink® service.

#### Order Summary

Licensee: Miss. Nina Culum  
 Order Date: Aug 2, 2021  
 Order Number: 5120840267346  
 Publication: Analytical and Bioanalytical Chemistry  
 Characterization of extracellular vesicles derived from  
 Title: mesenchymal stromal cells by surface-enhanced Raman spectroscopy  
 Type of Use: Thesis/Dissertation  
 Order Total: 0.00 USD

### Early-Stage Lung Cancer Diagnosis by Deep Learning-Based Spectroscopic Analysis of Circulating Exosomes



Author: Hyunku Shin, Seunghyun Oh, Soonwoo Hong, et al

Publication: ACS Nano

Publisher: American Chemical Society

Date: May 1, 2020

Copyright © 2020, American Chemical Society

#### PERMISSION/LICENSE IS GRANTED FOR YOUR ORDER AT NO CHARGE

This type of permission/license, instead of the standard Terms and Conditions, is sent to you because no fee is being charged for your order. Please note the following:

- Permission is granted for your request in both print and electronic formats, and translations.
- If figures and/or tables were requested, they may be adapted or used in part.
- Please print this page for your records and send a copy of it to your publisher/graduate school.
- Appropriate credit for the requested material should be given as follows: "Reprinted (adapted) with permission from {COMPLETE REFERENCE CITATION}. Copyright {YEAR} American Chemical Society." Insert appropriate information in place of the capitalized words.
- One-time permission is granted only for the use specified in your RightsLink request. No additional uses are granted (such as derivative works or other editions). For any uses, please submit a new request.

If credit is given to another source for the material you requested from RightsLink, permission must be obtained from that source.

[BACK](#)

[CLOSE WINDOW](#)

### Improved Label-Free Identification of Individual Exosome-like Vesicles with Au@Ag Nanoparticles as SERS Substrate



Author: Juan C. Fraire, Stephan Stremersch, Davinia Bouckaert, et al

Publication: Applied Materials

Publisher: American Chemical Society

Date: Oct 1, 2019

Copyright © 2019, American Chemical Society

#### PERMISSION/LICENSE IS GRANTED FOR YOUR ORDER AT NO CHARGE

This type of permission/license, instead of the standard Terms and Conditions, is sent to you because no fee is being charged for your order. Please note the following:

- Permission is granted for your request in both print and electronic formats, and translations.
- If figures and/or tables were requested, they may be adapted or used in part.
- Please print this page for your records and send a copy of it to your publisher/graduate school.
- Appropriate credit for the requested material should be given as follows: "Reprinted (adapted) with permission from {COMPLETE REFERENCE CITATION}. Copyright {YEAR} American Chemical Society." Insert appropriate information in place of the capitalized words.
- One-time permission is granted only for the use specified in your RightsLink request. No additional uses are granted (such as derivative works or other editions). For any uses, please submit a new request.

If credit is given to another source for the material you requested from RightsLink, permission must be obtained from that source.

[BACK](#)

[CLOSE WINDOW](#)



### Surface enhanced Raman scattering artificial nose for high dimensionality fingerprinting

**Author:** Nayoung Kim et al

**Publication:** Nature Communications

**Publisher:** Springer Nature

**Date:** Jan 10, 2020

*Copyright © 2020, The Author(s)*

#### Creative Commons

This is an open access article distributed under the terms of the [Creative Commons CC BY](#) license, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

You are not required to obtain permission to reuse this article.

To request permission for a type of use not listed, please contact [Springer Nature](#)



### Beehive-Inspired Macroporous SERS Probe for Cancer Detection through Capturing and Analyzing Exosomes in Plasma

**Author:** Shilian Dong, Yuhui Wang, Zhengqi Liu, et al

**Publication:** Applied Materials

**Publisher:** American Chemical Society

**Date:** Jan 1, 2020

*Copyright © 2020, American Chemical Society*

#### PERMISSION/LICENSE IS GRANTED FOR YOUR ORDER AT NO CHARGE

This type of permission/license, instead of the standard Terms and Conditions, is sent to you because no fee is being charged for your order. Please note the following:

- Permission is granted for your request in both print and electronic formats, and translations.
- If figures and/or tables were requested, they may be adapted or used in part.
- Please print this page for your records and send a copy of it to your publisher/graduate school.
- Appropriate credit for the requested material should be given as follows: "Reprinted (adapted) with permission from {COMPLETE REFERENCE CITATION}, Copyright {YEAR} American Chemical Society." Insert appropriate information in place of the capitalized words.
- One-time permission is granted only for the use specified in your RightsLink request. No additional uses are granted (such as derivative works or other editions). For any uses, please submit a new request.

If credit is given to another source for the material you requested from RightsLink, permission must be obtained from that source.

[BACK](#)

[CLOSE WINDOW](#)

**Marketplace™**

Dear Nina Culum,

Thank you for placing your order on [Marketplace™](#).

**Order Summary:**  
 Order date: 22 Aug 2021  
 Order number: 1142350  
 No. of items: 1  
 Order total: 0.00 CAD

**Billing Summary:**  
 Payment method: Invoice  
 An invoice will be generated and emailed within 24 hours.

To view your order details, click the following link, sign in, and search for your order:  
[Manage Account](#).


How was your experience? [Click here to give us feedback](#)

Please do not reply to this message.

To speak with a Customer Service Representative, call +1-855-239-3415 toll free or +1-978-646-2600 (24 hours a day), or email your questions and comments to [support@copyright.com](mailto:support@copyright.com).

Sincerely,  
 The CCC Marketplace Team

Tel: 1-855-239-3415 / +1-978-646-2600  
[support@copyright.com](mailto:support@copyright.com)  
[Manage Account](#)





**In Situ Exosomal MicroRNA Determination by Target-Triggered SERS and Fe<sub>3</sub>O<sub>4</sub>@TiO<sub>2</sub>-Based Exosome Accumulation**

Author: Shuqin Jiang, Qing Li, Chongwen Wang, et al

Publication: ACS Sensors

Publisher: American Chemical Society

Date: Mar 1, 2021

*Copyright © 2021, American Chemical Society*

#### PERMISSION/LICENSE IS GRANTED FOR YOUR ORDER AT NO CHARGE

This type of permission/license, instead of the standard Terms and Conditions, is sent to you because no fee is being charged for your order. Please note the following:

- Permission is granted for your request in both print and electronic formats, and translations.
- If figures and/or tables were requested, they may be adapted or used in part.
- Please print this page for your records and send a copy of it to your publisher/graduate school.
- Appropriate credit for the requested material should be given as follows: "Reprinted (adapted) with permission from {COMPLETE REFERENCE CITATION}. Copyright {YEAR} American Chemical Society." Insert appropriate information in place of the capitalized words.
- One-time permission is granted only for the use specified in your RightsLink request. No additional uses are granted (such as derivative works or other editions). For any uses, please submit a new request.

If credit is given to another source for the material you requested from RightsLink, permission must be obtained from that source.

[BACK](#)

[CLOSE WINDOW](#)



## Thank you for your order!

Dear Miss. Nina Culum,

Thank you for placing your order through Copyright Clearance Center's RightsLink® service.

### Order Summary

Licensee: Miss. Nina Culum  
 Order Date: Aug 22, 2021  
 Order Number: 5134521511891  
 Publication: Small  
 Title: Quantitative and Specific Detection of Exosomal miRNAs for Accurate Diagnosis of Breast Cancer Using a Surface-Enhanced Raman Scattering Sensor Based on Plasmonic Head-Flocked Gold Nanopillars  
 Type of Use: Dissertation/Thesis  
 Order Total: 0.00 CAD

View or print complete [details](#) of your order and the publisher's terms and conditions.

Sincerely,

Copyright Clearance Center



### Controlled Self-Assembly of a Close-Packed Gold Octahedra Array for SERS Sensing Exosomal MicroRNAs

Author: Tuli Kang, Jingtian Zhu, Xiaojun Luo, et al

Publication: Analytical Chemistry

Publisher: American Chemical Society

Date: Feb 1, 2021

*Copyright © 2021, American Chemical Society*

### PERMISSION/LICENSE IS GRANTED FOR YOUR ORDER AT NO CHARGE

This type of permission/license, instead of the standard Terms and Conditions, is sent to you because no fee is being charged for your order. Please note the following:

- Permission is granted for your request in both print and electronic formats, and translations.
- If figures and/or tables were requested, they may be adapted or used in part.
- Please print this page for your records and send a copy of it to your publisher/graduate school.
- Appropriate credit for the requested material should be given as follows: "Reprinted (adapted) with permission from {COMPLETE REFERENCE CITATION}. Copyright {YEAR} American Chemical Society." Insert appropriate information in place of the capitalized words.
- One-time permission is granted only for the use specified in your RightsLink request. No additional uses are granted (such as derivative works or other editions). For any uses, please submit a new request.

If credit is given to another source for the material you requested from RightsLink, permission must be obtained from that source.

[BACK](#)

[CLOSE WINDOW](#)

## Curriculum Vitae

**Name:** Nina M. Culum

**Post-secondary Education and Degrees:** University of Western Ontario  
London, Ontario, Canada  
2019-2021 M.Sc., Chemistry

University of Waterloo  
Waterloo, Ontario, Canada  
2015-2019 B.Sc., Honors Chemistry

**Honors and Awards:** Centre for Advanced Materials and Biomaterials Research  
(CAMBR) Day Poster Award  
University of Western Ontario  
2021

Queen Elizabeth II Graduate Scholarship in Science and  
Technology (QEII-GSST)  
University of Western Ontario  
2020-2021

Christian Sivertz Scholarship in Physical Chemistry  
University of Western Ontario  
2019

Dean's Honors List  
University of Waterloo  
2019

President's Scholarship  
University of Waterloo  
2015

**Related Work Experience:** Teaching Assistant  
University of Western Ontario  
2019-2020

Teaching Assistant  
University of Waterloo  
2018-2019

## Publications:

3. **Ćulum, N. M.**; Cooper, T. T.; Lajoie, G. A.; Dayarathna, T.; Pasternak, S. H.; Liu, J.; Fu, Y.; Postovit, L.-M.; Lagugné-Labarthet, F. Characterization of Ovarian Cancer-Derived Extracellular Vesicles by Surface Enhanced Raman Spectroscopy. *Analyst* **2021**, Accepted manuscript, DOI: 10.1039/d1an01586a.
2. **Ćulum, N. M.**; Cooper, T. T.; Bell, G. I.; Hess, D. A.; Lagugné-Labarthet, F. Characterization of Extracellular Vesicles Derived from Mesenchymal Stromal Cells by Surface-Enhanced Raman Spectroscopy. *Anal. Bioanal. Chem.* **2021**, *413*, 5013-5024. **Cover article.**
1. Therien, D. A. B.; **Ćulum, N. M.**; McRae, D. M.; Mazaheri, L.; Lagugné-Labarthet, F. Single-Beam Inscription of Plasmon-Induced Surface Gratings. *Opt. Mater.* **2021**, *112*, 110775.

## Conference Contributions:

### Oral

1. **Ćulum, N. M.**; Bell, G. I.; Cooper, T. T.; Hess, D. A.; Lagugné-Labarthet, F. (2020) Fabrication of Plasmonic Substrates by Electron-Beam Lithography for Extracellular Vesicle Detection and Characterisation. 4<sup>th</sup> Annual Graduate Student Research Symposium, London, Ontario, Canada (Virtual).

### Poster

4. **Ćulum, N. M.**; Bell, G. I.; Cooper, T. T.; Hess, D. A.; Lagugné-Labarthet, F. (2021) Surface-Enhanced Raman Spectroscopic Characterization of Extracellular Vesicles Derived from Mesenchymal Stromal Cells. 104<sup>th</sup> Canadian Chemistry Conference and Exhibition, Montréal, Quebec, Canada (Virtual).
3. **Ćulum, N. M.**; Lui, J.; Fu, Y.; Cooper, T. T.; Postovit, L.-M.; Lagugné-Labarthet, F. (2021) Plasmonic Nanohole Arrays for the Diagnosis of Ovarian Cancer by Surface-Enhanced Raman Spectroscopy. Centre for Advanced Materials and Biomaterials Research Day, London, Ontario, Canada (Virtual).
2. **Ćulum, N. M.**; Lui, J.; Fu, Y.; Cooper, T. T.; Postovit, L.-M.; Lagugné-Labarthet, F. (2021) Characterization of Ovarian Cancer-Derived Extracellular Vesicles by Surface-Enhanced Raman Spectroscopy. 2<sup>nd</sup> Annual Microscopical Society of Canada Symposium, Hamilton, Ontario, Canada (Virtual).
1. **Ćulum, N. M.**; Bell, G. I.; Cooper, T. T.; Hess, D. A.; Lagugné-Labarthet, F. (2020) Investigating Extracellular Vesicles Derived from Mesenchymal Stromal Cells by Surface-Enhanced Raman Spectroscopy. 47<sup>th</sup> Annual Conference of Federation of Analytical Chemistry and Spectroscopy Studies, Reno, Nevada, USA (Virtual).