Title

Towards Minimally Invasive Cancer Detections through Label-free Surface Enhanced Raman Spectroscopy of Individual Small Extracellular Vesicles

Author

Liu, Zirui

Publication Date

2022

Peer reviewed|Thesis/dissertation

UNIVERSITY OF CALIFORNIA

Los Angeles

Towards Minimally Invasive Cancer Detections through Label-free Surface Enhanced Raman Spectroscopy of Individual Small Extracellular Vesicles

A dissertation submitted in partial satisfaction of the requirements for the degree of Doctor of Philosophy in Materials Science and Engineering

by

Zirui Liu

2022

ABSTRACT OF THE DISSERTATION

Towards Minimally Invasive Cancer Detections through

Label-free Surface Enhanced Raman Spectroscopy of

Individual Small Extracellular Vesicles

by

Zirui Liu

Doctor of Philosophy in Materials Science and Engineering

University of California, Los Angeles, 2022

Professor Ya-Hong Xie, Chair

Cancer is one of the leading causes of premature death worldwide. Currently, when diagnosed at advanced stages, the five-year survival rates of cancers remain low even with tremendous amounts of medical resources dedicated to the treatments. To enhance the survival rates and reduce the cancer-caused socioeconomic burdens, population-based cancer screening is desired. Such screening requires the detection techniques to be accurate, patient friendly, and easy to operate at low costs. Over the past decade, cell-released small extracellular vesicles (sEVs, sized ⌀ 30−150 nm) have been attracting increasingly amounts of attention as a source of biomarkers for minimally invasive cancer detections. While existing stably in the bodily fluids, sEVs reflect

their cells of origins, including cancer cells, through the encapsulated cargos such as DNA/RNA and proteins. Therefore, it is possible to detect cancers by investigating the sEVs captured in the bodily fluids. This thesis aims to explore the potential clinical applicability of analyzing individual sEVs by surface-enhanced Raman spectroscopy (SERS) for minimally invasive cancer detections. To objectively examine the SERS spectral features collected from sEVs, customized machine learning algorithm is utilized. Further, the combination between SERS and machine learning is named as "SERS Identification of Molecules" or "SIM".

The thesis work begins with understanding the heterogeneity of sEVs, namely sEV subpopulations, from the single-vesicle level by SIM. Currently, sEVs, even from the same cell of origin, are recognized to be heterogeneous which can be further sub-fractionated. Size discrepancy is one of the most popularly used criteria for separating the vesicular subpopulations. In Chapter 2, SIM was employed to spectrally detect and analyze individual sEVs from cell lines isolated based on their size discrepancies, forming subpopulations. The results suggested that sEVs in different size groups carried different chemical compositions, which could be reflected by the distinguishable SERS spectral features.

Chapter 3 & 4 include the explorations of the clinical applicability of SIM in the sEV-based minimally invasive cancer detections. In the first case, SIM was applied to analyze sEVs derived from the saliva, blood, and tissue samples between gastric cancer (GC) patients and non-GC participants (n = 15 each). The algorithm prediction accuracies were reportedly 90, 85, and 72%. "Leave-a-pair-of-samples out" validation was further performed to test the clinical potential. The area under the curve of each receiver operating characteristic curve was 0.96, 0.91, and 0.65 in tissue, blood, and saliva, respectively. In addition, by comparing the SERS fingerprints of individual vesicles, a possible way of tracing the biogenesis pathways of patient-specific sEVs

from tissue to blood to saliva was provided. The second case involved SIM in analyzing sEVs derived from bronchoalveolar fluid (BAL) for non-small cell lung cancer (NSCLC) detection. BAL samples were collected from NSCLC patients and non-cancer participants (n = 10 each). Analyzing the SERS spectra collected from the BAL-derived sEVs revealed the SERS spectral distinguishability between the patient and the control group in general. In addition, the sEV spectra collected from patients with early and late-stage NSCLC were also distinct. Blind test to examine the clinical usage of SIM was perform (n = 6 each for the machine learning model developing and n = 4 for the testing). Under such setup, the model correctly predicted diagnostic results from all the eight individuals in the testing set. Collectively the results obtained from the two cancer case studies indicate the clinical potential of the minimally invasive cancer detections via SIM analyzing sEVs derived from bodily fluids.

The dissertation of Zirui Liu is approved.

Ioanna Kakoulli

Ximin He

Gal Bitan

Ya-Hong Xie, Committee Chair

University of California, Los Angeles

2022

To my wife, for her unconditional love and support

Table of Contents

List of Figures and Tables

Acknowledgements

The PhD journey at UCLA has been the most challenging but fruitful adventure in my life. Conducting independent research was never easy for me. Luckily, I managed to surmount with the kind help and guidance from many people. First and foremost, I would like to express my sincere gratitude to my advisor, Prof. Ya-Hong Xie, for the guidance and patient instructions throughout my entire PhD career. His erudite knowledge in physics, rigorous attitude towards science, and high requirements on the performance have inserted profound influence, making me a better researcher. Outside the academics, Prof. Xie has been incredibly supportive and caring about me and my family. Health is the top priority as he often says. It is my great honor and pleasure being as a student of him. What I learned from Prof. Xie throughout these years will benefit me at all times.

I would like to thank my collaborators: Dr. David T.W. Wong, Dr. Yong Kim, and the colleagues at UCLA Dentistry; Dr. Tony Jun Huang and his colleagues at Duke University; Dr. Robert Raffai and his colleagues at UCSF; Dr. Huinan Liu and her colleagues at UC Riverside; Sumita T. Jonak, Brian V. Dsouza, J. Alan Schiaffino, and the colleagues at Nurlabs. I really enjoyed working and learning from you. Your talents and expertise greatly benefit the interdisciplinary studies and projects that we have faced together.

Working with the amazing people in Prof. Xie's group has been inspiring and joyful. Tieyi Li, thank you for your work on developing the machine learning algorithm. Jun Liu, thank you for taking care of the substrate fabrications for the entire research group. Siddharth Srivastava, thank you for helping me with the detailed data analyses. In addition to the current colleagues, I received help from some of the former group members as well. Dr. Shan Huang, I still remember the day when we stayed up the whole night trying to secure a research grant for the group. Dr. Zhongbo

Yan, thank you for taking the initiation and laying the foundation for the extracellular vesicle research in our group. Dr. Xinke Yu, thank you for leading the research on the machine learning-based spectral analysis in our group. I felt intriguing working with these great talents. Moreover, throughout the years, we have become close friends.

I would like to show my gratefulness to my family. My wife, Qi Yan, has been staying with me from the very beginning of this journey. My parents, Dr. Muyu Liu and Dr. Hong Li, have not only provided me with financial supports, but also offered suggestions on my academic performances from their professional perspectives as faculty members. Together with others from the family, they are my strongest support. I would not be here without them.

Last but not least, I would like to thank all other people who have helped me through this adventure.

VITA

2015	B.S. in Materials Science and Engineering
	University of California, Los Angeles
	Los Angeles, CA, U.S.A.
2022	Ph.D. Candidate in Materials Science and Engineering
	University of California, Los Angeles
	Los Angeles, CA, U.S.A.

Chapter 1. Introduction

1.1. Motivation of the thesis

Cancer happens where some of the cells grow uncontrollably with the potential of spreading across the entire body. It is notorious as being the second leading cause of death worldwide. In 2020, estimated 19.3 million new cases and 10.0 million deaths occurred. Cancer in general causes medical, social, and economic burdens globally. One key for improving the cancer survival rate lies in the early detections of the incidences. Early detections of cancers depend on reliable screening techniques that are minimally invasive to the patients. In addition, such techniques need to be low cost and rather easy to operate to be applicable to the general public. Clinically available screening techniques currently include but not limited to computed tomography (CT), endoscopy, genomic tests, and liquid biopsies (e.g., blood tests, saliva tests, and urine tests) based on biomarkers. With the development of the field, extracellular vesicles (EVs), especially small extracellular vesicles (sEVs) have attracted increasing amounts of attention as a potential new source of biomarkers for liquid biopsies of cancer.

sEVs are cell released vesicles, ⌀ 30–150 nm, that participate in the circulation. Research has shown that sEVs play important roles in cell-cell communications through cargo transportation such as proteins, DNAs, and RNAs. They are capable of regulating the recipient cells while representing their cells of origins. The stable existences in bodily fluids grant the potentials for them being biomarkers for cancer liquid biopsy. By detecting those sEVs, opportunities exist for non-invasive detections of cancers.

Surface-enhanced Raman spectroscopy (SERS) uses surface plasmon to boost the signal from the inelastic scattering between laser light and the matter, also known as Raman scattering. It has proven to be able to provide the spectroscopic information about the compositions and/or structures of molecules through detecting the collective of Raman active chemical bonds. Because of its high specificity, SERS becomes progressively attractive in bio-sensing including disease detections. Particular for sEVs, with the sensitivity at the single molecule level, SERS obsesses the capability of detecting single vesicles. Works have been shown SERS detecting cancer by distinguishing sEVs from different sources (e.g., normal tissue vs. cancer cells). While developing rapidly, the field is still relatively new comparing to those clinically available techniques for cancer screenings mentioned above. Research interests remain in presenting the potential clinical value of SERS on cancer screening via measuring sEVs. In addition, insights about the biogenesis and the metabolism of the sEVs from the spectroscopic perspectives are yet to be fully revealed.

The main objective of this thesis is to explore the clinical applicability of SERS for minimally invasive cancer detections via analyzing single sEVs. To achieve this, the combination of a gold nanopyramid SERS substrate that our group developed previously, and machine learning was utilized. Through the studies of samples from human donors, both the promising features and the potential areas for improvements are provided. The following of this chapter presents the background introductions of the entire thesis. Chapter 1.2 provides the overview of the cancer caused burdens worldwide and a brief review about different techniques for cancer screenings. Chapter 1.3 gives the introduction of sEVs from the perspectives of structure, biogenesis, isolation, and analysis. Chapter 1.4 contains the introduction of SERS and SERS based bio-sensing. Chapter 1.5 includes the information about the gold nanopyramid substrate and the introduction of machine learning that was involved in this thesis. Chapter 1.6 offers the outline of this dissertation.

1.2. Overview of Cancer Caused Burdens and Cancer Screening

Cancer continues ranking as one of the leading causes of death and imposing challenges for life expectancy improvements across almost every country around the world.[1] In general, the rapid increasing of the incidence and the mortality rate of cancer as a result of aging and population growth reflects a strong tie with the socioeconomic development. According to International Agency for Research on Cancer (IARC), by 2040, the cancer caused burden is estimated to grow to 27.5 million new cases with 16.3 million deaths.[2] Figure 1.1 shows the comparison between the national ranking of cancer as the cause of premature death and the human development index across the countries worldwide, indicating the positive correlation between cancer as one of the top causes of premature death and the degree of socioeconomic growth.[3]

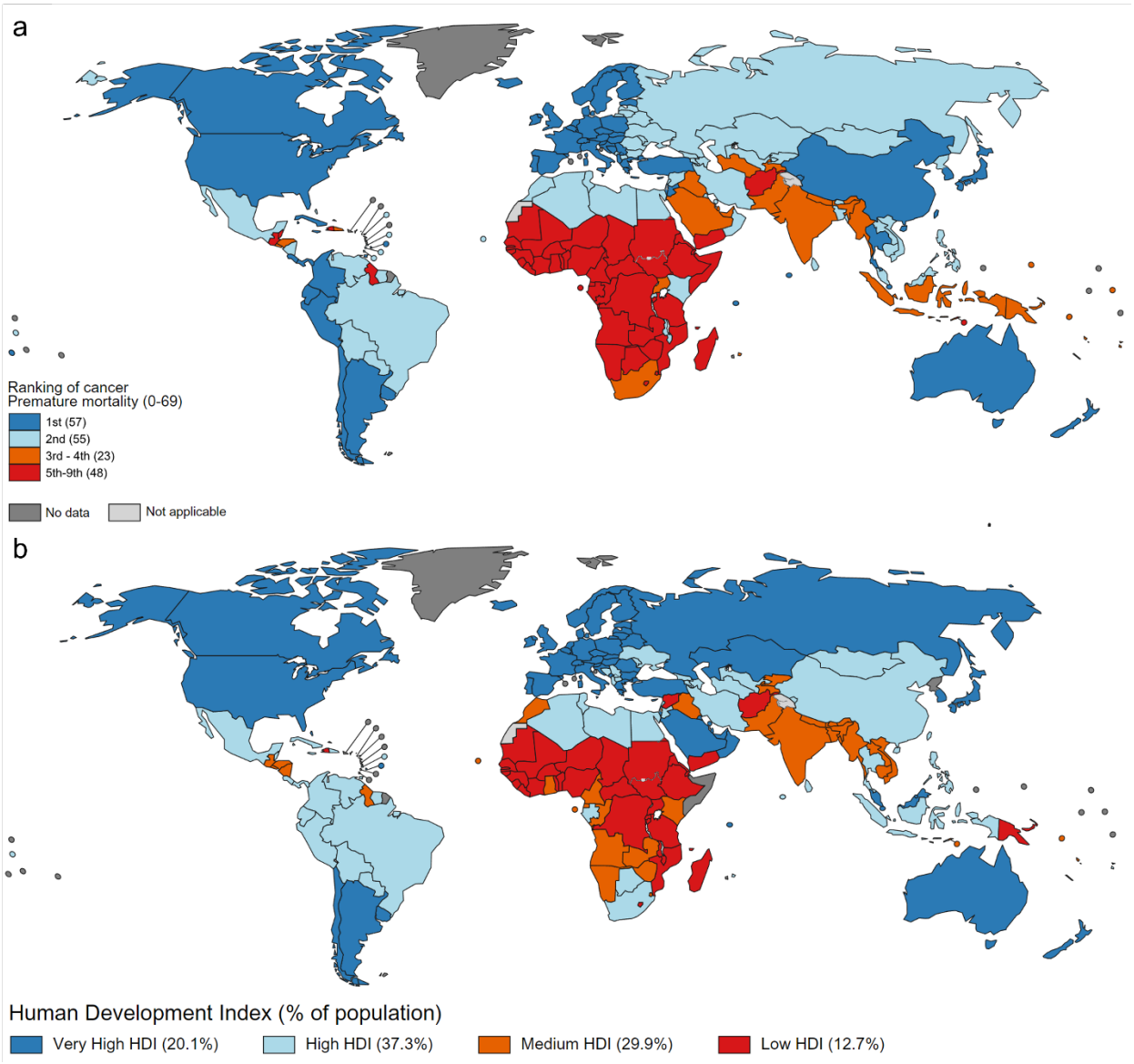

Figure 1.1. Rankings of cancer as a cause of premature death (< age 70) across the countries globally, a); Human Development Index associated with different regions worldwide

Breaking down the total incidence and death cases attributed to cancer in 2020 (shown in Figure 1.2, according to the world health organization, WHO) by income reveals that while high-income population accounts for 40.6% of the cancer incidences, it only accounts for 29% of the total mortality. The situation is worse in the low-and-middle income population which accounts

for 70% of the death cases. In general, cancer causes tremendous social and economic burdens particularly because when diagnosed at the advanced stages, the 5-year survival rates remain low even with large amount of medical resource spent at high costs. One key to relief the burden and increase the patient survival rate of cancer lies in the early detection of the disease. For example, in the cases of stomach cancers, the 5-year survival rate remains low (below 20-25%) if diagnosed at advanced stages. However, for patients diagnosed of GC at early stages, five-year survival rates of 95% or higher have been observed. Considering the cancer caused burden to the entire society, population-based early detection programs are desired. Such early detection of cancer depends on fast and reliable screening methods with reasonable amount of operating costs.

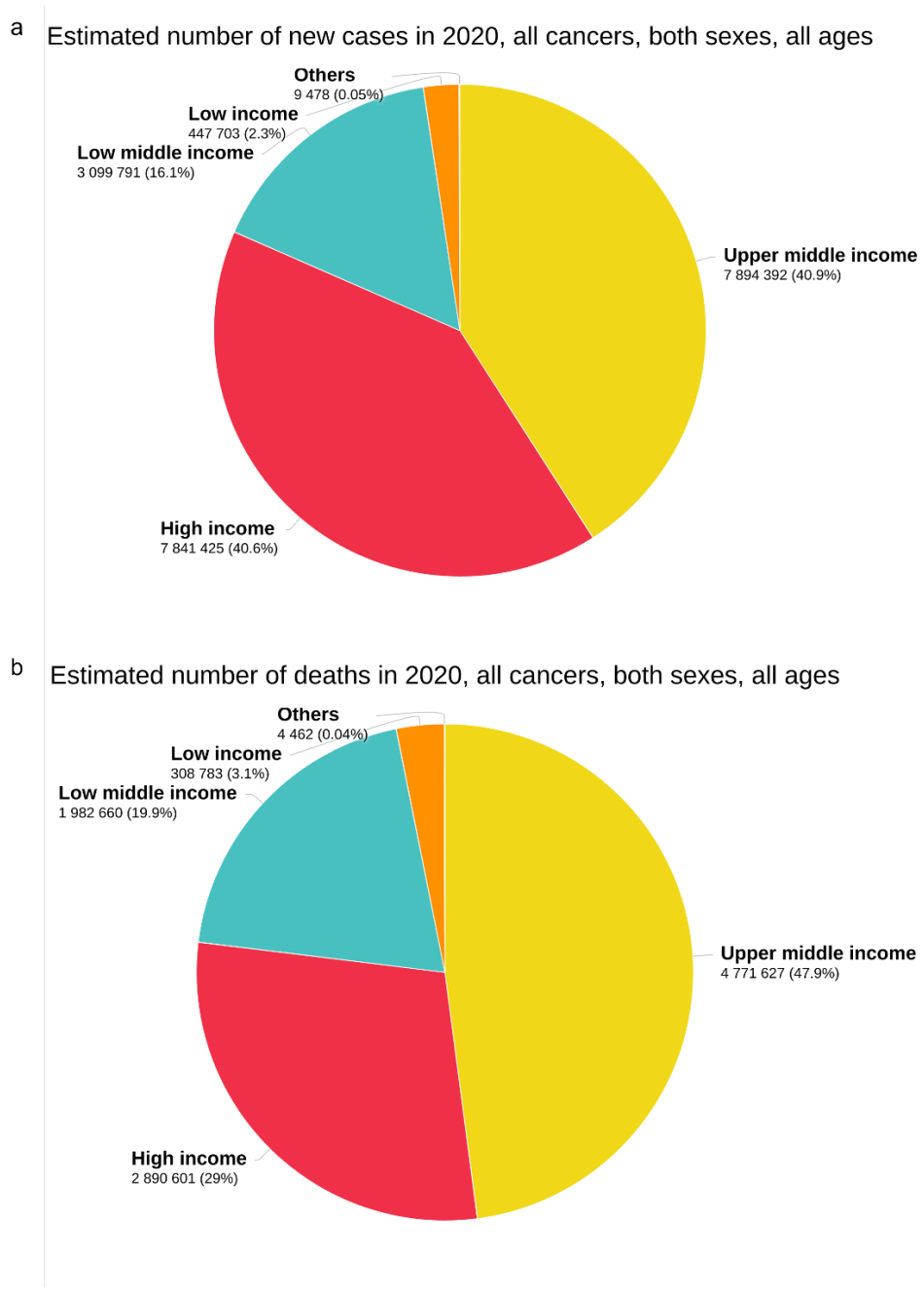

Figure 1.2. Break-down of the total cancer cases according to the income levels, a); Break-down of the total cancer caused death cases according to the income levels

Currently, there are several cancer screening methods available for various types of cancers. For breast cancer, mammography has been utilized to generate the picture of the breast for

screening and it is reported to lower the breast cancer caused deaths among women ages of 40-74, especially those from 50 to 69.[4, 5] For lung cancer, the most lethal type of cancer, low-dose helical CT scan has been shown to reduce the death cases among people at high risks, for example, heavy smokers.[6, 7] For cancers related to the digestive system such as stomach and colorectal cancers, endoscopy together with other techniques such as stool and/or blood tests are used as screening methods which are proven to improve the survival rates of the patients.[8, 9] The techniques mentioned here are popular ones but do not exhaust all the screening methods that currently available since the screening methods are cancer type specific. In addition to the detection methods, one successful example of the primary prevention program of cervical cancer is based on population human papillomavirus (HPV) tests and HPV vaccines.[10, 11] Emerging works also have been focusing on the improvements of the available techniques or developing novel screening methods such as cancer genomic tests and multicancer biomarkers tests in blood.[12-14] Apart from the promising aspects of the current cancer screening methods, challenges, such as delayed treatments caused by false negative results and overdiagnosis/overtreatments caused by false positive results, still remain. Moreover, some of the procedures could be uncomfortable and/or have side effects during the processes. Therefore, research interests on improving the current techniques or developing novel methods for cancer screening have been attracting increasingly amount of attention. In recent years, extracellular vesicles (EVs), especially small extracellular vesicles (sEVs) have been regarded as a potential new type of biomarkers for minimally invasive cancer detections based on liquid biopsy.

1.3. Introduction of small extracellular vesicles (sEVs)

Small extracellular vesicles (sEVs), currently also known as exosomes, are cell released nanoscale vesicles sizing from ~30-150 nm in diameters.[15] The vesicles consist of a lipid bilayer membrane encapsulating the contents such as proteins, DNAs, and RNAs.[15, 16] They play crucial roles in intracellular communications by their biological contents carried, also known as cargos.[15-17] The membrane contains two layers of lipid molecules which serve as the protection of inner cargos during the trafficking. A lipid molecule has a hydrophobic tail and a hydrophilic head. When closely aligned form a bilayer structure, the membrane prevents the substances exchange between the encapsulated contents and the microenvironment while allowing the vesicle to move freely during the circulation. A sample schematic of sEV is shown in Figure 1.3. The reason why sEVs reflect their parental cells lies in their biogenesis. In general, the biogenesis of sEVs starts with the invagination of the cell plasma membrane to incorporate surface proteins and soluble proteins, forming endosomes.[18] Then the endosomes are sorted to load the cargos that are inside the cell.[19] After sorting, the endosomes eventually generate multivesicular endosomes (MVBs) by the invagination of the endosomal membrane which is also referred to as to the double invagination of the plasma membrane.[20] sEVs are released from the cell when the MVBs fuse to the cell membrane. Figure 1.4 exhibits the schematic of the sEV biogenesis. Such process is different from the simple budding of the membrane. The uptake mechanism of sEVs by the recipient cells are yet to be fully established, especially from the vesicular function perspective. But such process could be modeled as fusing with the cell plasma membrane, similar to virus-cell interactions.[21]

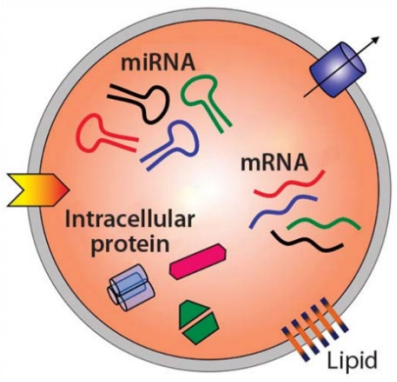

Figure 1.3. Schematic of the structure of a small extracellular vesicle

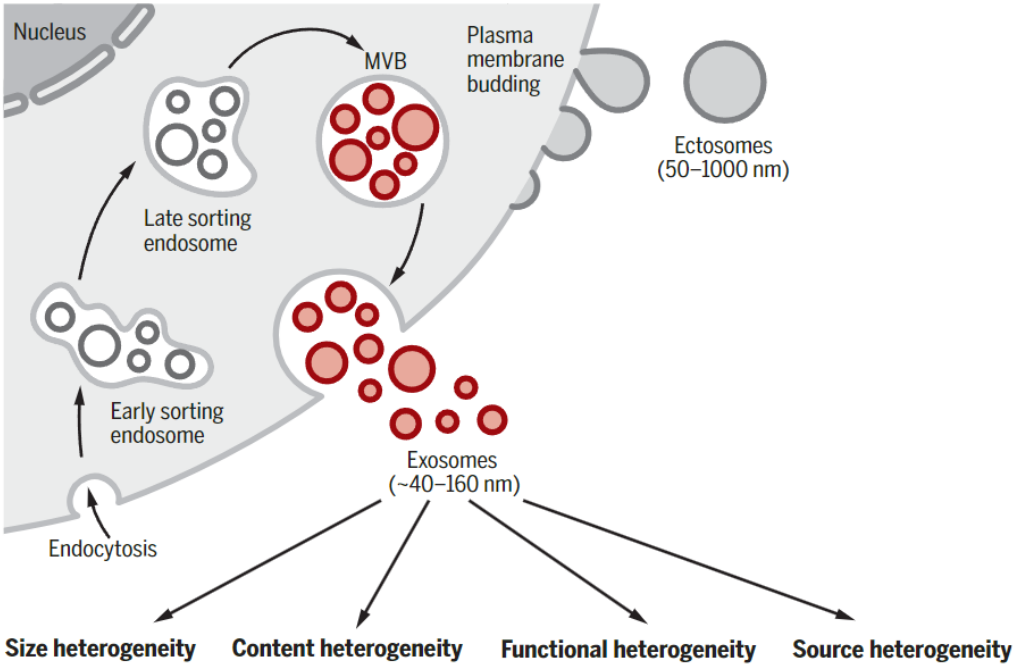

Figure 1.4. Schematic of sEV biogenesis

Based on the biogenesis of sEVs, works have shown that sEVs reflect their parental cells. Such fact grants the potential of sEVs being as biomarkers for disease detections.[22] Particularly for cancer, proteomic analysis has shown that sEVs were released from different cancer cells with unique cargos reflecting the parental cells.[23] More importantly, they are detectable in multiple

types of bodily fluids such as blood, saliva, and urine, making the cancer-derived sEVs potential biomarkers for liquid biopsy based minimally invasive cancer detections. The fact that it is capable of obtaining sEVs in a non-invasive fashion makes monitoring of the conditions possible via the cancer-derived sEVs at different time points. Further studies recognize that sEVs released, including the cancer derived ones, are heterogeneous from the same cell type or even the same individual cell. Such heterogeneity can be defined from multiple aspects, such as size, density, and the biochemical compositions.[24] Figure 1.5 provides an example illustration about the sEV heterogeneity associated with cancers. Such realization leads to the research on the subpopulation of the sEVs for further understanding the group. Currently, it is challenging for the analytical techniques to detect and analyze sEVs directly from the bodily fluids. Isolation and purification of the vesicles are often required following the raw sample harvesting from the patients before the downstream analysis of the contents.

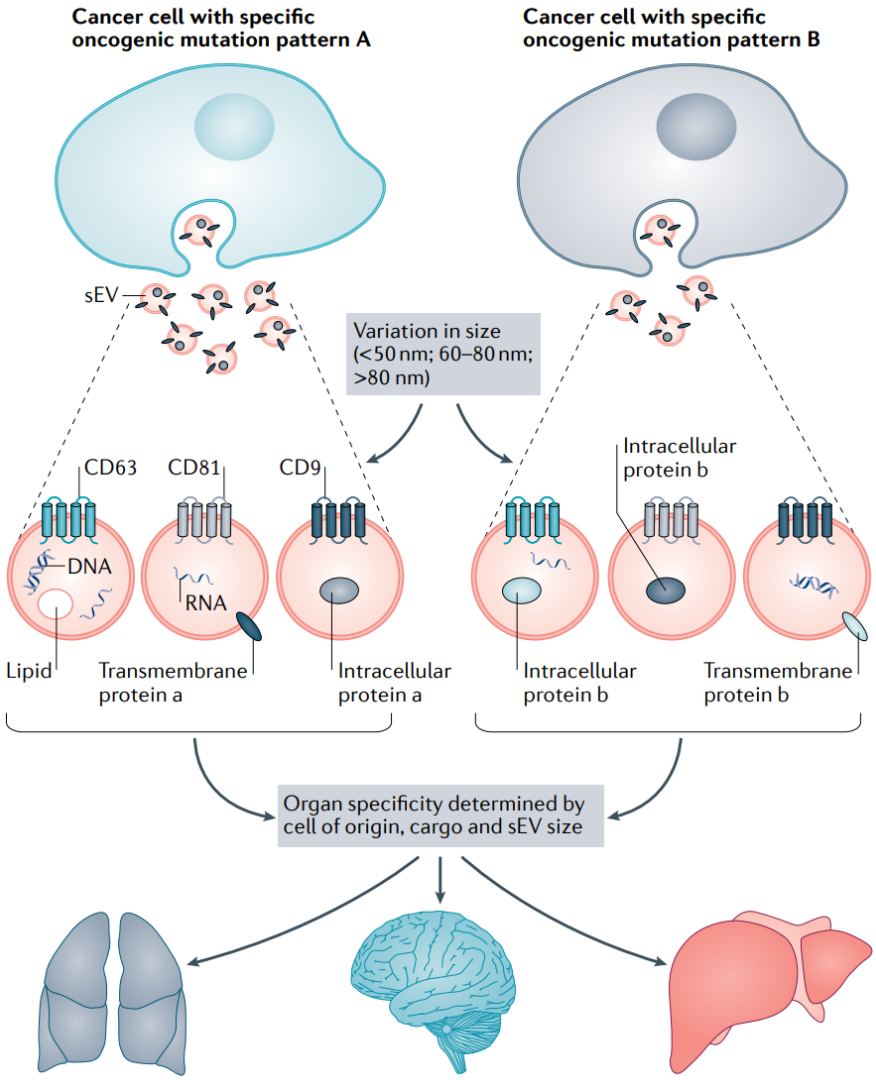

Figure 1.5. Schematic of the sEV heterogeneity associated with cancer

1.4. Introduction of Surface-Enhanced Raman Spectroscopy (SERS)

Surface-enhanced Raman spectroscopy originates from Raman spectroscopy, rooted at Raman scattering.[25] When light interacts with matters, most of the photons are scattered elastically, meaning that the scattered photon has the same energy with the incident photon. On the other hand, a small portion of the photons scatter inelastically, indicating an energy difference between the

incident and the scattered photon. As such, an energy transfer happens between the photon and the matter based on the conservation of energy, leading to the matter either gain or lose energy. In general, such inelastic scattering is called Raman scattering. Specifically, when the matter gains energy during the process, it is called Stokes scattering. If the matter loses energy, it is called Anti-Stokes scattering. The transferred energy is corresponding to the vibrational energy of the substance. The shift in energy offers the information about the vibrational modes of the matter. Given the vibrational energy states are unique to the molecules which depend on the chemical bonding strengths, such information obtained from Raman scattering holds high specificity about the matter. Therefore, Raman spectroscopy is used to provide structural fingerprints of those molecules that are detectable. Figure 1.6 shows the energy diagrams, illustrating different types of light-matter interactions. Despite the promising features of detecting molecules, the chance of Raman scattering is extremely low (roughly 10^{-6}).[26] To boost the signal from the Raman spectroscopy and make the system more sensitive, surface-enhanced Raman spectroscopy (SERS) was developed utilizing the localized surface plasmons resonance (LSPR) generated by the metallic nanostructures.

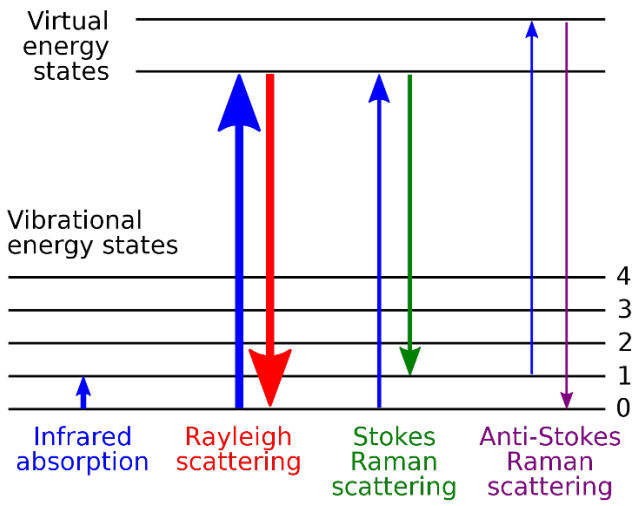

Figure 1.6. Energy diagrams for different interactions between light and matter

LSPR is triggered when the collective oscillation frequency of the valence electrons in a metallic nanoparticle/nanostructure matches the frequency of the incident light.[27] As a result, the electric field near the surface of the nanostructure is enhanced, illustrated in Figure 1.7. The place where the electric field is mostly enhanced is called the "hotspot".[28] It should be noted at the enhancement dissipates quickly away from the hotspot (within few hundred nanometers to the most). SERS overcomes the low signal issue in Raman spectroscopy and pushes the detection limit to even the single molecule level. With the inherent molecular specificity and high detection sensitivity, SERS has been utilized to identify molecules in various areas of research including biomedicines such as disease detections. Particular for EVs and sEVs, emerging works have been published showing the potential of SERS analyzing the vesicles. Penders et al. (2021) developed a Raman trapping technique of single EVs to establish the breast cancer cell-derived EVs as biomarkers.[29] Kruglik et al. (2019) advanced a Raman tweezer to be able to analyze sEVs at 100nm diameters.[30] Shin et al. (2020) combined SERS of sEVs and deep learning for early-stage lung cancer detection with area under the curve (AUC) > 0.9.[31] Dong et al. (2020) reported that SERS

spectral variation of protein phosphorylation inside the vesicles could serve as the indicator for detecting four types of cancers.[32] To prevent vesicle solution druing during the measurements, Rojalin et al. (2020) developed a porous scaffold SERS platform, resulting a clear spectral distinguishability between cancer patients and healthy individuals.[33] In addition to detect the spectral features directly, Banaei et al. (2021) modified SERS substrate for vesicle immobilization allows quantitative analysis via variations of the SERS indicator intensities.[34] Our group previously developed a SERS substrate based on gold nanopyramids which has been shown to have the single molecule sensitivity.[35] SERS measurements of the following works in this thesis is based on such substrate.

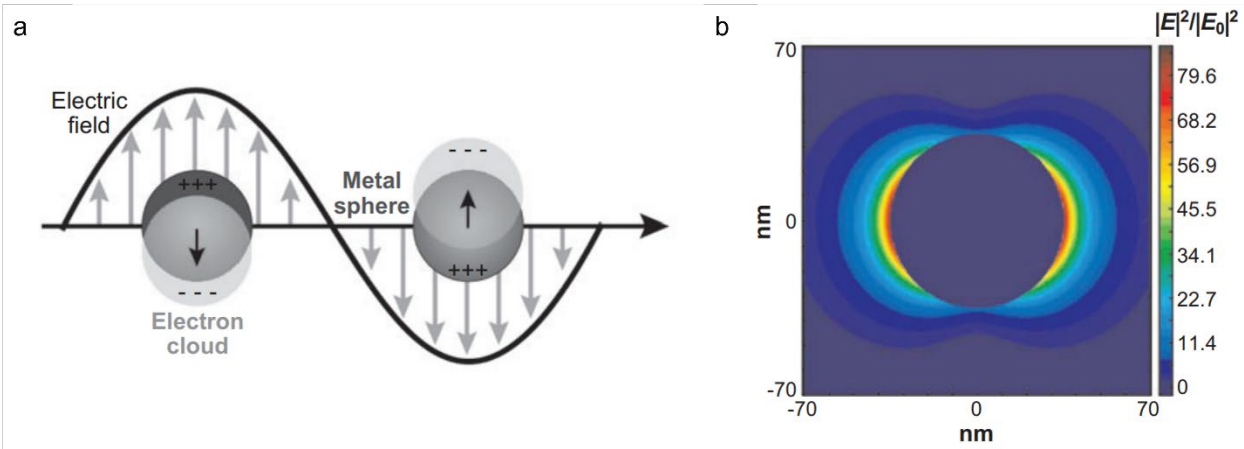

Figure 1.7. Schematic of LSPR, a); Illustration of the electric field enhancements, b)

1.5. Introduction of the SERS Gold Nanopyramids and Machine Learning for Analyzing sEVs

The SERS substrate developed previously by our group consists of an array of gold nanopyramids. Figure 1.8 illustrates the detailed structure of the substrate. The distance between the neighboring nanopyramids is 250nm. Based on the simulation, the hotspots of our gold nanopyramid located at the centers of the ladders with the hotspot size of 150-200 nm. The design

of the substrate makes it a strong candidate for sEV analysis because of two characteristics. First, the neighboring distance of adjacent nanopyramids is large enough to allow vesicles to fall in between, being close to the hotspots without stacking. Second, the hotspot size is comparable to the sizes of the sEVs, making it possible to extract the information from the entire vesicle.

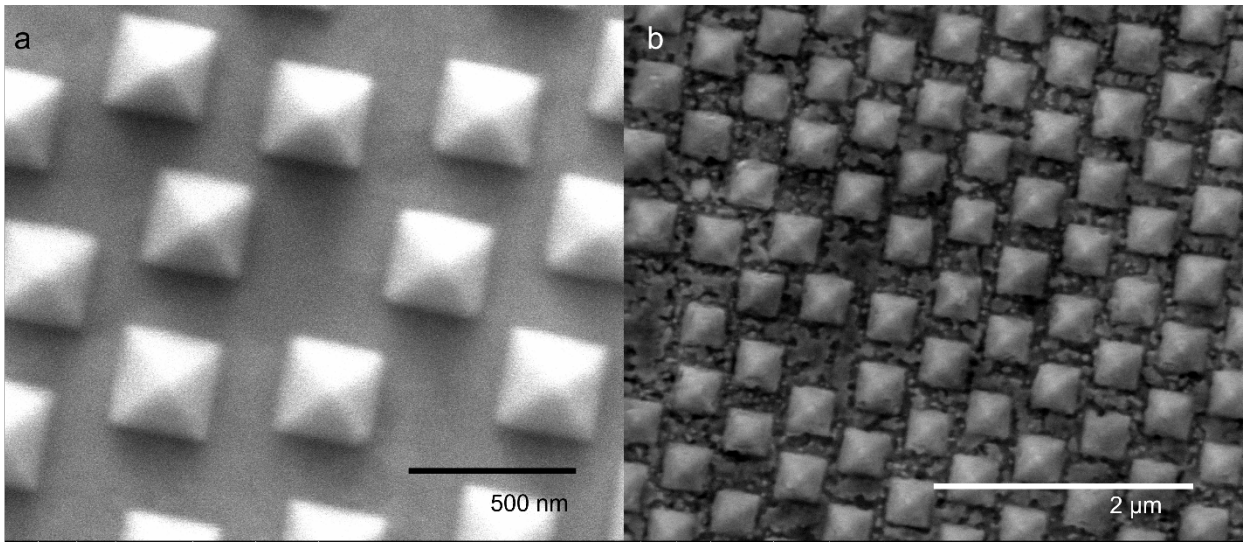

Figure 1.8. SEM images of the gold nanopyramid substrate with 160,000× magnification, a); and 60,000× magnification, b)

The substrate has been applied for sEV studies and demonstrated the single vesicle sensitivity through a correlative study between SERS and SEM.[36] To correlate between SEM visualization and Raman signature measurements, serial-dilution experiments was conducted, in which unaltered sEV preparation or samples diluted 3- or 10-times were utilized. Such study was led by Dr. Zhongbo Yan, a former member of our group, and Dr. Suman Dutta, a former member of Dr. Gal Bitan's group at UCLA David Geffen School of Medicine. Raman spectra were collected across a 10 × 10-pixel area at each concentration and visualized the same hybrid substrates by SEM (Figure 1.9). The Raman mapping results of the three samples showed a density change consistent with the change of the sample concentration. To analyze the spectral data

collected from the sEVs to establish the objective distinguishability for cancer detection, a data analysis mechanism is required. Machine learning has shown tremendous promise in meeting this challenge.

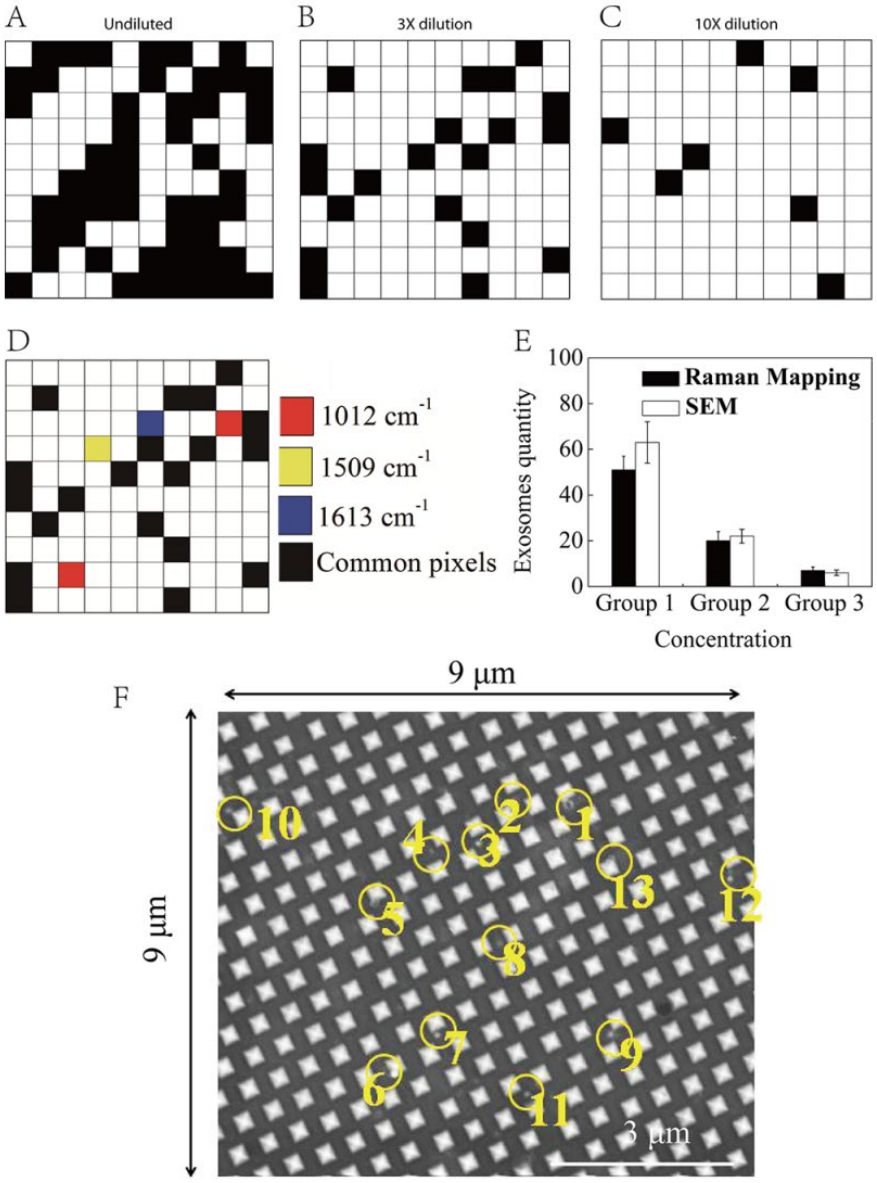

Figure 1.9. SERS and SEM mapping of sEVs adsorbed on the the hybrid substrate. (A−C) Raman mapping of the same undiluted (A), 3-times diluted (B), or 10-times diluted (C) sEV preparation. (D) Pixel assignment for the Raman signature of sEVs. The red, yellow, and blue pixels represent

the presence of 1012, 1509, and 1613 cm^{-1} peaks in the Raman spectrum, respectively. The black pixels are those in which all three peaks were detected. Only black pixels were considered as containing sEVs. (E) Comparison of the sEV density obtained through Raman mapping and SEM at three different exosom concentrations. (F) A representative 9 × 9- μm SEM micrograph of sEVs attached to the graphene-covered surface at 35,000× magnification. The yellow circles mark the presence of sEVs within this region.

Machine learning is a branch of artificial intelligence (AI) that has been attracting increasing attention and has experienced a great acceleration in development pace over the past decade. It has been applied to assist with disease diagnosis and big data analysis in preventative biomedical settings. With respect to SERS data collected from sEVs in the form of spectra, machine learning is an indispensable tool to provide objective determination regarding the similarities and differences among large numbers of Raman spectra collected from sEVs. In addition, machine learning is capable of further grouping the spectra from different individual sEVs based on their spectral features, enabling the study of sEV subpopulations. Combining the gold nanopyramid SERS substrate with machine learning, the entire platform is named as "SERS Identification of Molecules" or "SIM"

1.6. Outline of the Thesis

The following four chapters of this thesis goes from the studies of understanding sEV subpopulations isolated based their size discrepancies to the applications of SIM on two types of

cancers, exploring the potential clinical applicability of the minimally invasive cancer detection based on SERS detecting sEVs in the bodily fluids.

Chapter 2 describes an experimental study of SIM analyzing individual sEVs isolated based on their size discrepancies forming subpopulations. Also, it elaborates the validation the origins of the SERS spectral features identified from sEVs via a comparative analysis between SERS and mass spectrometry (MS). The results suggest that there exists a correlation between sEV size-based subpopulations and their biomolecular compositions from the perspective of individual vesicles. Such findings highlight the possibility that the biogenesis and respective biological functionalities of the various sEV subpopulations, as distinguished by the vesicle-size discrepancies, may be inherently different.

Chapter 3 includes the results from experimental studies of SIM analyzing sEVs from stomach tissue, blood, and saliva from human donors for minimally invasive detection of gastric cancer. It illustrates the potential value of SIM as a platform for gastric cancer population screening based on liquid biopsy of sEVs and discusses the challenges as well. One example could be found at the less than desired accuracy caused by low cancer-derived sEV counts in bodily fluids. In addition, through comparing the SERS fingerprints of individual vesicles, a possible way of tracing the biogenesis pathways of patient-specific sEVs from tissue to blood to saliva was provided.

Chapter 4 illustrates the experimental studies of SIM examining sEVs isolated from bronchoalveolar lavage fluid (BAL) for non-smaller cell lung cancer (NSCLC) detection. The selection of BAL is to harvest sEVs mostly from the lung, in the consideration of the low-cancer-derived-sEV-count challenge mentioned above. It also elaborates a blind test based on the spectra collected from different donors to preliminarily test the applicability of SIM in the clinical usage.

Chapter 5 gives the summary of the thesis by providing the major outcomes and findings in this work together with potential directions for future research that can be built on the presented results.

1.7. Reference

1. Fidler, M. M.; Bray, F.; Soerjomataram, I., The global cancer burden and human development: A review. *Scand J Public Health* **2018,** *46* (1), 27-36.

2. Soerjomataram, I.; Bray, F., Planning for tomorrow: global cancer incidence and the role of prevention 2020-2070. *Nat Rev Clin Oncol* **2021,** *18* (10), 663-672.

3. Sung, H.; Ferlay, J.; Siegel, R. L.; Laversanne, M.; Soerjomataram, I.; Jemal, A.; Bray, F., Global Cancer Statistics 2020: GLOBOCAN Estimates of Incidence and Mortality Worldwide for 36 Cancers in 185 Countries. *CA Cancer J Clin* **2021,** *71* (3), 209-249.

4. Walter, L. C.; Schonberg, M. A., Screening mammography in older women: a review. *JAMA* **2014,** *311* (13), 1336-47.

5. Hodgson, R.; Heywang-Kobrunner, S. H.; Harvey, S. C.; Edwards, M.; Shaikh, J.; Arber, M.; Glanville, J., Systematic review of 3D mammography for breast cancer screening. *Breast* **2016,** *27*, 52-61.

6. Gierada, D. S.; Black, W. C.; Chiles, C.; Pinsky, P. F.; Yankelevitz, D. F., Low-Dose CT Screening for Lung Cancer: Evidence from 2 Decades of Study. *Radiol Imaging Cancer* **2020,** *2* (2), e190058.

7. Kovalchik, S. A.; Tammemagi, M.; Berg, C. D.; Caporaso, N. E.; Riley, T. L.; Korch, M.; Silvestri, G. A.; Chaturvedi, A. K.; Katki, H. A., Targeting of low-dose CT screening according to the risk of lung-cancer death. *N Engl J Med* **2013,** *369* (3), 245-254.

8. Kaminski, M. F.; Thomas-Gibson, S.; Bugajski, M.; Bretthauer, M.; Rees, C. J.; Dekker, E.; Hoff, G.; Jover, R.; Suchanek, S.; Ferlitsch, M.; Anderson, J.; Roesch, T.; Hultcranz, R.; Racz, I.; Kuipers, E. J.; Garborg, K.; East, J. E.; Rupinski, M.; Seip, B.; Bennett, C.; Senore, C.; Minozzi, S.; Bisschops, R.; Domagk, D.; Valori, R.; Spada, C.; Hassan, C.; Dinis-Ribeiro,

M.; Rutter, M. D., Performance measures for lower gastrointestinal endoscopy: a European Society of Gastrointestinal Endoscopy (ESGE) Quality Improvement Initiative. *Endoscopy* **2017,** *49* (4), 378-397.

9. Valdastri, P.; Simi, M.; Webster, R. J., 3rd, Advanced technologies for gastrointestinal endoscopy. *Annu Rev Biomed Eng* **2012,** *14*, 397-429.

10. Roden, R.; Wu, T. C., How will HPV vaccines affect cervical cancer? *Nat Rev Cancer* **2006,** *6* (10), 753-63.

11. Lei, J.; Ploner, A.; Elfstrom, K. M.; Wang, J.; Roth, A.; Fang, F.; Sundstrom, K.; Dillner, J.; Sparen, P., HPV Vaccination and the Risk of Invasive Cervical Cancer. *N Engl J Med* **2020,** *383* (14), 1340-1348.

12. Loeb, S.; Ross, A. E., Genomic testing for localized prostate cancer: where do we go from here? *Curr Opin Urol* **2017,** *27* (5), 495-499.

13. Yanes, T.; Willis, A. M.; Meiser, B.; Tucker, K. M.; Best, M., Psychosocial and behavioral outcomes of genomic testing in cancer: a systematic review. *Eur J Hum Genet* **2019,** *27* (1), 28-35.

14. Kim, Y.; Yeo, I.; Huh, I.; Kim, J.; Han, D.; Jang, J. Y.; Kim, Y., Development and Multiple Validation of the Protein Multi-marker Panel for Diagnosis of Pancreatic Cancer. *Clin Cancer Res* **2021,** *27* (8), 2236-2245.

15. Jeppesen, D. K.; Fenix, A. M.; Franklin, J. L.; Higginbotham, J. N.; Zhang, Q.; Zimmerman, L. J.; Liebler, D. C.; Ping, J.; Liu, Q.; Evans, R.; Fissell, W. H.; Patton, J. G.; Rome, L. H.; Burnette, D. T.; Coffey, R. J., Reassessment of Exosome Composition. *Cell* **2019,** *177* (2), 428-445 e18.

16. Doyle, L. M.; Wang, M. Z., Overview of Extracellular Vesicles, Their Origin, Composition, Purpose, and Methods for Exosome Isolation and Analysis. *Cells* **2019,** *8* (7).

17. Maia, J.; Caja, S.; Strano Moraes, M. C.; Couto, N.; Costa-Silva, B., Exosome-Based Cell-Cell Communication in the Tumor Microenvironment. *Front Cell Dev Biol* **2018,** *6*, 18.

18. Kalluri, R.; LeBleu, V. S., The biology, function, and biomedical applications of exosomes. *Science* **2020,** *367* (6478).

19. Hessvik, N. P.; Llorente, A., Current knowledge on exosome biogenesis and release. *Cell Mol Life Sci* **2018,** *75* (2), 193-208.

20. Jadli, A. S.; Ballasy, N.; Edalat, P.; Patel, V. B., Inside(sight) of tiny communicator: exosome biogenesis, secretion, and uptake. *Mol Cell Biochem* **2020,** *467* (1-2), 77-94.

21. Mathieu, M.; Martin-Jaular, L.; Lavieu, G.; Thery, C., Specificities of secretion and uptake of exosomes and other extracellular vesicles for cell-to-cell communication. *Nat Cell Biol* **2019,** *21* (1), 9-17.

22. Pegtel, D. M.; Gould, S. J., Exosomes. *Annu Rev Biochem* **2019,** *88*, 487-514.

23. Abhange, K.; Makler, A.; Wen, Y.; Ramnauth, N.; Mao, W.; Asghar, W.; Wan, Y., Small extracellular vesicles in cancer. *Bioact Mater* **2021,** *6* (11), 3705-3743.

24. Moller, A.; Lobb, R. J., The evolving translational potential of small extracellular vesicles in cancer. *Nat Rev Cancer* **2020,** *20* (12), 697-709.

25. Kudelski, A., Analytical applications of Raman spectroscopy. *Talanta* **2008,** *76* (1), 1-8.

26. Das, R. S.; Agrawal, Y. K., Raman spectroscopy: Recent advancements, techniques and applications. *Vibrational Spectroscopy* **2011,** *57* (2), 163-176.

27. Hutter, E.; Fendler, J. H., Exploitation of Localized Surface Plasmon Resonance. *Advanced Materials* **2004,** *16* (19), 1685-1706.

28. Stiles, P. L.; Dieringer, J. A.; Shah, N. C.; Van Duyne, R. P., Surface-enhanced Raman spectroscopy. *Annu Rev Anal Chem (Palo Alto Calif)* **2008,** *1*, 601-26.

29. Penders, J.; Nagelkerke, A.; Cunnane, E. M.; Pedersen, S. V.; Pence, I. J.; Coombes, R. C.; Stevens, M. M., Single Particle Automated Raman Trapping Analysis of Breast Cancer Cell-Derived Extracellular Vesicles as Cancer Biomarkers. *ACS Nano* **2021**.

30. Kruglik, S. G.; Royo, F.; Guigner, J. M.; Palomo, L.; Seksek, O.; Turpin, P. Y.; Tatischeff, I.; Falcon-Perez, J. M., Raman tweezers microspectroscopy of circa 100 nm extracellular vesicles. *Nanoscale* **2019,** *11* (4), 1661-1679.

31. Shin, H.; Oh, S.; Hong, S.; Kang, M.; Kang, D.; Ji, Y. G.; Choi, B. H.; Kang, K. W.; Jeong, H.; Park, Y.; Hong, S.; Kim, H. K.; Choi, Y., Early-Stage Lung Cancer Diagnosis by Deep Learning-Based Spectroscopic Analysis of Circulating Exosomes. *ACS Nano* **2020,** *14* (5), 5435-5444.

32. Dong, S.; Wang, Y.; Liu, Z.; Zhang, W.; Yi, K.; Zhang, X.; Zhang, X.; Jiang, C.; Yang, S.; Wang, F.; Xiao, X., Beehive-Inspired Macroporous SERS Probe for Cancer Detection through Capturing and Analyzing Exosomes in Plasma. *ACS Appl Mater Interfaces* **2020,** *12* (4), 5136-5146.

33. Rojalin, T.; Koster, H. J.; Liu, J.; Mizenko, R. R.; Tran, D.; Wachsmann-Hogiu, S.; Carney, R. P., Hybrid Nanoplasmonic Porous Biomaterial Scaffold for Liquid Biopsy Diagnostics Using Extracellular Vesicles. *ACS Sens* **2020,** *5* (9), 2820-2833.

34. Banaei, N.; Moshfegh, J.; Kim, B., Surface enhanced Raman spectroscopy-based immunoassay detection of tumor-derived extracellular vesicles to differentiate pancreatic cancers from chronic pancreatitis. *Journal of Raman Spectroscopy* **2021,** *52* (11), 1810-1819.

35. Wang, P.; Liang, O.; Zhang, W.; Schroeder, T.; Xie, Y. H., Ultra-sensitive graphene-plasmonic hybrid platform for label-free detection. *Adv Mater* **2013,** *25* (35), 4918-24.

36. Yan, Z.; Dutta, S.; Liu, Z.; Yu, X.; Mesgarzadeh, N.; Ji, F.; Bitan, G.; Xie, Y. H., A Label-Free Platform for Identification of Exosomes from Different Sources. *ACS Sens* **2019,** *4* (2), 488-497.

Chapter 2. Understanding sEVs Subpopulations from Single-Vesicle Level

2.1. Introduction

Increasing evidence recognizes that sEVs are heterogeneous even from the same type of cell of origin, leading to the interest in understanding the biological relevance of specific sEV subpopulations.[1-3] Such heterogeneity is defined from both the perspectives of physical properties such as size and density as well as chemical compositions. Particularly related to the sEV based disease detections such as cancer, understanding the vesicular subpopulations facilitates the better sorting and targeting of the subgroup of sEVs with the higher values in diagnosis. Published studies have shown the variations of surface markers and/or specific cargo molecules among the groups of sEV subpopulations. Such investigations of sEV subpopulations have been focusing on the different vesicle groups isolated based on various physical properties.[4-7] Further understanding the heterogeneity of sEVs at the single-vesicle level attracts research interest as it provides an alternative or complementary perspective comparing to the population-based analysis. One question yet to be fully answered is that whether there is a correlation between the physical property based sEV sub-fractionations and their chemical content differences. In order to explore such correlation, it requires an sEV subpopulation isolation technique that is based on the physical properties and a vesicle characterization method that is based on chemical compositions.

Currently, there exist several isolation and sub-fractionation of sEVs techniques that are based on their biophysical properties, e.g., their size or density using techniques such as ultracentrifugation and size exclusion chromatography (SEC). Each of these techniques have their advantages and space for improvements.[8-10] Ultracentrifugation, is the first and remains most

widely used method for the isolation of sEVs.[11, 12] It consists of a sequence of centrifugation steps to selectively deplete free floating protein complexes, cellular debris, organelles and larger microvesicles with the highest centrifugal force of around 100,000 x g used to pellet sEVs. Major limitations to the approach include vesicle loss, prolonged processing time (~5-8hrs), and the frequently required large starting volume of the analytes.[13, 14] In addition, it is challenging for conventional ultracentrifugation to effectively isolate sEV subpopulations as all EVs are collectively isolated in a pellet, a process that has been shown to cause EV aggregations.[15-17] Recent studies have led to improvements with ultracentrifugation including by combining density gradient ultracentrifugation and iso-osmotic gradient ultracentrifugation, methods that have been shown to resolve subpopulations of EVs in a more gentle manner.[18, 19] Other types of isolation techniques have been developed to improve sEV sub-fractionation, including ultrafiltration and size-based chromatography (SEC).[20] Ultrafiltration is frequently used as a size-based sEVs isolation techniques. The working principle is that sEV subpopulations can be isolated based on the differences in their sizes and molecular weights by using a sub-micrometer sized filter membrane.[21, 22] In addition to enabling the isolation of EV subpopulations, ultrafiltration has a relatively low processing time and cost. To overcome the challenge of low vesicle isolation yields and potential bias for the isolated EV subpopulations, ongoing studies based on ultrafiltration aim to improve EV purification by reducing membrane clogging during the isolation process.[23, 24]

SEC is recognized as a method that can gently fractionate an EV sample within a solution of mobile phase running through a porous filtration matrix fixed within a column as the stationary phase, resulting a differential size elution.[25] Using this approach, larger particles elute before smaller ones, ultimately a sized-based separation of sEV subpopulations. In addition to the relatively rapid processing time and low cost, SEC possesses the advantage of preserving the

structure and size of sEVs after their fractionation.[26-30] Therefore, SEC is increasingly regarded as a promising method for sEV subpopulation isolation for downstream biochemical analyses. While it certainly has limitations at the current stage, work has been done to improve the scalability and throughput, overcoming the main challenges in SEC-based sEV isolation.[31]

Other methods aimed at sEV subpopulation isolation are increasingly being developed. Among these the one deserving special mention is acoustofluidics, that represents a combination of microfluidics and acoustics, providing a nanostructure-based microfluidic sEV isolation platform.[32-34] In addition to tremendous effort in improving the isolation of sEV subpopulations, with respect to throughput and size resolutions, emerging works on downstream characterizations based on biochemical content have also been reported. Collectively the field is evolving to explore and further establish the underlying basis for the biogenesis and function of sEV subpopulations.

Approaches commonly used for the biomolecular characterization of sEV cargo content often involves proteomic and genomic techniques. In proteomics studies, the Western blot is still regarded as a gold standard for protein detection and analyses in biological samples and has been widely applied in various sEVs and sEV subpopulation studies. In order to utilize Western blot-based proteomics, protein samples need to first be fragmented and subsequently denatured to prevent their degradation. In addition, the Western blot is inherently a bulk-based detection method. Typical experiments require more than 10^6 sEVs for one analysis.[35] Therefore, distinctive features associated with individual types of sEVs are at the risk of being lost during the procedure. Works are in progress to improve the subpopulation detection via Western blot.[36]

In addition to the Western blot, enzyme linked immunosorbent essay (ELISA) is another popular proteomic-based characterization method for identifying unique markers on the surface of sEVs (e.g., CD63 & CD9). As an antibody-based technique, ELISA provides marker information with

high specificity.[37-39] The process normally requires the determination of the target surface marker prior to the analysis. Reported works show the potential of improving the capability of multicomplex for EV subpopulation analysis together with reducing the non-specific cross-reactvity.[40]

Flow cytometry with fluorescently labelled antibodies has been increasingly attracting interest because it has single vesicle sensitivity and is capable of providing biochemical information with the specific binding between antibody and antigen displayed on sEVs.[41, 42] Studies with next generation cytometers have attempted to improve the limit of detection of conventional systems, which ranges from around 100nm to 300nm, in order to improve the study of vesicles with smaller sizes such as sEVs.[43-45] Vesicle-flow cytometry offers a powerful approach for analyzing vesicle subpopulations via detecting the target surface markers on the vesicle samples. [46, 47] Moreover, flow cytometry is capable of running in a "label-free" format to determine the size and concentration of the EVs.

Genomic studies of EV cargo often make use of unbiased sequencing and polymerase chain reaction (PCR) to analyze their oligonucleotide content.[48] These approaches provide an alternative to surface marker-based characterization techniques with the capability of providing molecular content information for the vesicles.[49, 50] Based on the principles, next generation sequencing was introduced to improve the multicomplex of the sequencing with less operating time. Also, emerging works have been focused on increasing the detection limit to the single vesicle level[51, 52]

This chapter presents the combined use of the SERS gold nanopyramid substrate along with the customized machine learning program to analyze the spectral signatures collected from individual sEVs isolated based on their size. To this end, a well-defined HEK293 cell line and its HRAS transgenic variant cell line were used as parental cells to derive vesicles of interest. The

sEVs were isolated from the conditioned cell culture medium using a combination of cushioned-ultracentrifugation, which served to gently concentrate the sEVs that were next fractionated by size using SEC.[18] To firmly establish this approach, the SERS data with mass spectrometry (MS) data obtained from the same samples of sEVs were compared. Results of the study reveal that the SERS spectral signatures correlate with the biomolecular composition of sEVs. The findings are supported by SERS spectra that were analyzed using linear discriminant analysis (LDA). SERS spectral signatures collected from vesicles in different size groups revealed clearly distinguishable differences with small overlaps, suggesting that some of the sEVs in different size-based groups shared common biomolecular composition. This was achieved despite the non-zero spectral feature variations within each of the subpopulations, likely caused by an inherent biological variability among individual EVs. Collectively, results of the study lay the foundation for future investigations into the relationship between sEV size and their biomolecular composition at the single EV level.

2.2. Experimental Procedures

2.2.1. sEV isolation and nanoparticle tracking analysis (NTA)

sEVs were isolated from conditioned cell culture medium by first performing Cushioned-Ultracentrifugation,[18] followed by Size Exclusion Chromatography using IZON 35 nm qEV columns. Briefly, HEK293 cells carrying a doxycycline-inducible transgene, expressing mNeonGreen or HRAS-mNeonGreen fusion protein, were cultured in Pro293a Chemically Defined Medium (Lonza) supplemented with 1% Penicillin/Streptomycin (Gibco), 1% GlutaMAX (Gibco), and 1% fetal bovine serum (Gibco). Cells were cultured in 225 cm^2 cell culture flask

(Corning) until they reached 90% confluency. The cells were then washed twice with phosphate buffered saline (PBS) (Corning) and cultured in serum-free Pro293a Chemically Defined Medium (Lonza) supplemented with 1% Penicillin/Streptomycin (Gibco), 1% GlutaMAX (Gibco), and 1 ug/mL of Doxycycline (Sigma Aldrich). The conditioned media was collected after 48 hours.

For sEV isolation, the conditioned media was first centrifuged at 400 x g for 10 min at 4°C to pellet debris and dead cells, then centrifuged at 2000 x g for 20 min at 4°C to pellet large vesicles and leftover debris. The supernatant was filtered (0.2 μm) and centrifuged on 2 mL of a 60% iodixanol cushion (Stem Cell Technologies) at 100,000 x g for 3 hours (Type 50.2 Ti, Beckman Coulter). The resulting concentrated cushion was extracted with an extra 1 mL and further purified by loading onto a 35 nm qEV column (IZON), with PBS as the diluting buffer. Afterwards, 13 distinct fractions were collected. Particles in fractions 7, 8, and 9 were subjected to size and concentration measurement by Nanosight LM14 (Malvern Instruments, Westborough, MA) performed using a 488 nm detection wavelength. The analysis settings were optimized and standardized for each sample. Samples were diluted in either 1:100 or 1:200 PBS and measured in triplicates. The detection threshold was set at 3, and 3 videos (1 min long each) were captured to give the mean, mode, median, and estimated concentration for each sample. Data were analyzed using the NTA 3.3 software.

2.2.2. Immunoblotting

Each fraction of the IZON SEC purified sEVs (37.5 μL) was mixed with 12.5 μL of 4x Laemmli buffer (Bio-Rad) and heated at 95°C for 5 minutes. Samples were then loaded on a 4-20% SDS-PAGE gel and transferred onto a PVDF membrane (Bio-Rad). The membranes were blocked with 5% non-fat milk dissolved in PBS for one hour, and then were probed with the following primary antibodies: anti-CD81 (1:500, Santa Cruz Biotechnology) and anti-CD63 (1:500, BD

Biosciences). After 4 washed in PBS containing 0.1% Tween (PBST), the membranes were incubated with HRP-conjugated secondary antibodies: anti-Mouse IgG-HRP (1:1000, Santa Cruz Biotechnology) for 1hr and washed with PBST. Signals were visualized after incubation with Amersham ECL Prime substrate and imaged using an ImageQuant LAS 4000.

2.2.3. Transmission electron microscopy (TEM)

sEV morphology was assessed by loading 5 μL of a sample onto a glow-discharged 300 mesh Formvar-coated copper grid. The particles were left to settle for 2 minutes, and then wicked of excess moisture with filter paper. The grids were then washed 3 times with 1% uranyl acetate (UA), after which the grid was left to rest on a drop of 1% UA for 1 minute. Excess UA was wicked off with filter paper. Grids were imaged at 120kV using a Tecnai 12 Transmission Electron Microscope (FEI) at the EM-Lab at the University of California Berkeley.

2.2.4. Scanning electron microscopy (SEM)

SEM used in this study was Nova 230. The working distance was ~5.0mm. The acceleration voltage was 10 kilovolts. The images were taken at the magnification between 45,000× and 55,000×. The electron detector used was TLD (through the lens) detector to obtain the signal from the secondary electrons.

2.2.5. SERS substrate fabrication

A single layer of self-assembled polystyrene (PS) balls (⌀ 500 nm) was generated on a surface of DI water using the Langmuir–Blodgett patterning. As shown in Figure 2.1, the layer was then transferred to a 4” (001) silicon wafer with a layer of 50 nm SiO_2 deposited on top. The PS balls were later removed using chloroform after a deposition of 50nm Cr. The exposed SiO_2 were etched using reactive ion etching to selectively expose Si. Next, the exposed silicon was

etched using KOH. Inverted nanopyramids with sidewalls at 57.5-degree angles were created because of different etching rates along the [001] and [111] directions of silicon. The model was finished by removing the residual Cr and SiO_2 using 48% HF solution. Then, 200 nm of gold film was deposited onto the pitted surface by electron beam deposition and bonded to a carrier wafer using epoxy before lifting off.

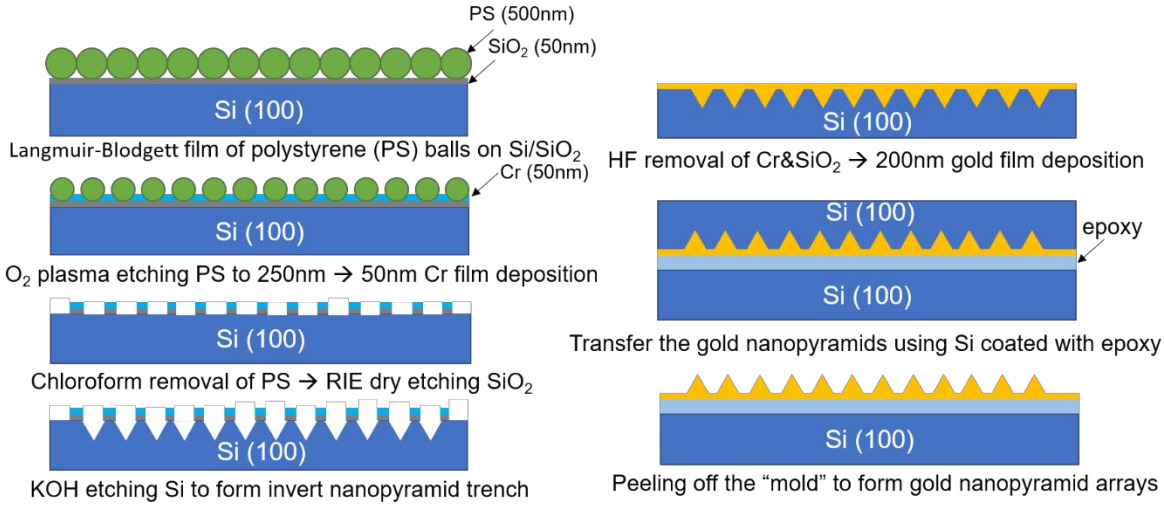

Figure 2.10. Schematic of SERS gold-pyramid substrate fabrication

2.2.6. Raman spectroscopy

First, 5 µL of each sEV sample solution was dropped onto the SERS substrate using a micropipette and dried. Raman measurements were performed using the Reinshaw inVia Raman spectrometer at room temperature. The laser excitation wavelength was 785 nm. The power used was 5 mW. 500x optical microscope was used for focusing the laser beam, resulting a ~1 µm laser spot diameter. Before measuring sEVs, the system was calibrated using the 520 cm^{-1} peak of silicon. The exposure time was 0.5 s to avoid sample overheating. For collecting SERS spectra from multiple sEVs, a Raman mapping measurement was performed over a 1.2 mm × 1.2 mm square with respect to the center of each sample droplet. One spectrum was collected from each of

the data spots and the step width was 5 μm to avoid collecting multiple spectra from one data spot, thus over-fitting in spectral analysis.

2.2.7. MS-based proteomics analysis

HEK293+HRAS sEV samples were first lysed using 5% SDS. Protein solutions were denatured with 10 mM DTT for 15 mins at 37 °C and alkylated with 50 mM iodoacetamide in the dark for 15 mins at room temperature. Afterward, the sample was added a final concentration of 2.5% phosphoric acid and then six volumes of binding buffer (90% methanol; 100 mM triethylammonium bicarbonate, TEAB; pH 7.1). After mixing, the protein solution was loaded to an S-Trap filter (ProtiFi), spun at 10000g for 1 min and then the filter was washed with 150 μL of binding buffer for 3 times. Finally, 1 μg of Lys-C and sequencing-grade trypsin and 20 μL of digestion buffer (50 mM TEAB) were added into the filter and the sample was digested at 37 °C for 16 h. To elute the peptides, 40 μL of 50 mM TEAB, 40 μL of 0.2% formic acid in H_2O, and 40 μL of 80% acetonitrile in H_2O were added sequentially. The peptide solutions were pooled and quantified BCA protein assay (Thermo Fisher Scientific). The peptides were dried with SpeedVac and stored at −80 °C until LC-MS/MS analysis.

The sEV peptides were reconstituted in 12 μL of 0.1% TFA with 2% ACN containing 0.01% DDM to reach a final concentration of 0.1 μg/μL, and 5 μL of the resulting sample was analyzed by LC-MS/MS using an Orbitrap Fusion Lumos Mass Spectrometer (Thermo Scientific) connected to a nanoACQUITY UPLC system (Waters) (buffer A: 0.1% FA with 3% ACN and buffer B: 0.1% FA in 90% ACN). Peptides were separated on an analytical column (75 μm i.d. × 20 cm) packed using 1.9-μm ReproSil C18 and with a column heater set at 48 °C, using an LC gradient (buffer A: 0.1% FA with 3% ACN and buffer B: 0.1% FA in 90% ACN): 2-6% buffer B in 1 min, 6-30% buffer B in 84 min, 30-60% buffer B in 9 min, 60-90% buffer B in 1 min, and

finally 90% buffer B for 5 min at 200 nL/min. Data were acquired in a data-dependent acquisition mode and the peptides were isolated using a quadrupole system (the isolation window was 0.7). Ionized peptides with a mass range of 350-1650 m/z were scanned at 60,000 resolutions with maximum injection time (IT) of 50 ms and 100% automatic gain control (AGC) target (4E5). Precursor ions with intensities > 1E4 were selected for fragmentation by higher-energy collisional dissociation (HCD) at 30% collision energy and scanned in an orbitrap with a 100% AGC (5E4) and an IT of 300 ms.

The raw MS/MS data were processed with MSFragger via FragPipe.[53, 54] The MS/MS spectra were searched against a human UniProt database (fasta file dated July 31, 2021, with 34,386 sequences which contain 17,193 decoys) and (initial) fragment mass tolerances were set to 20 ppm. A peptide search was performed with full tryptic digestion (Trypsin) and allowed a maximum of two missed cleavages. Carbamidomethyl (C) was set as a fixed modification; acetylation (protein N-term), and oxidation (M) were set as variable modifications. For match-between-run (MBR) analysis, 10 ppm m/z tolerance, 1.5 mins RT tolerance and 0.05 MBR ions FDR were used for analysis. The final reports were then generated (peptide-spectrum match (PSM), ion, peptide, and protein) and filtered at each level (1% protein FDR plus 1% PSM/ion/peptide-level FDR). The intensities of each protein/peptide were extracted from FragPipe outputs.

2.2.8. SERS spectral data analysis

For spectral analysis, algorithms were adopted from the open-source libraries of Numpy, Pandas, Sklearn, and Scipy. To generate the simulated spectra based on MS, the relative abundances of top 90 proteins were obtained using MS. Using web-scraping of the Uniprot online database, the amino acid sequence of each of the proteins was obtained. Combining MS data, the relative abundance of the amino acids was obtained. Further, SERS spectra of each of the 20 amino

acids (purchased from Sigma Aldrich) were measured and averaged 75 spectra with the best signal-to-noise ratios were obtained. All the Raman spectra were obtained under the same settings; hence the signal-to-noise ratios were used as an indicator of SERS activity. Combining the signal-to-noise ratios and the MS derived relative abundance, 20 independent coefficients for the 20 amino acids were obtained respectively. These coefficients were used to obtain a linear combination of the 20 SERS spectra, which was the final simulated SERS spectrum. To generate the fitted spectrum, machine learning, specifically Adaboost, was used. The dataset included 20 average SERS spectrum of amino acids as data instances, and 1117 points of Raman shift as the features. The fitting coefficients from Adaboost were then compared with the normalized relative-abundance ratios obtained from MS. Since the data has 1117 dimensions corresponding to the 1117 data points on the Raman shift axis, to visualize the clusters of spectra, Linear Discriminant Analysis (LDA) was used for the dimensionality reduction and supervised clustering. The 1117 dimensions were reduced to 1 dimension in the case of a binary cluster system (for e.g., HEK293 vs HEK293+HRAS), and 2 dimensions for (for e.g., IZON SEC fractions 7, 8, 9). The LDA model was trained on the following dataset – 1117 Raman shifts as features, and each SERS spectrum as a data instance. The trained model was used to transform the existing data set to a reduced dimension dataset for the purpose of visualization. The transformed data was plotted in 2 dimensions, and the results are presented in two dimensional plots.

2.3. Results and Discussion

2.3.1. Sample characterizations

sEVs were first isolated from conditioned cell culture mediums of HEK293 and HEK293+HRAS cells using cushioned-ultracentrifugation to concentrate the EVs. Subsequently, sEVs were sub-fractionated based on their size using SEC with qEV Izon columns. Purified sEV

resided mainly within three fractions: 7, 8 and 9 (F7, F8, F9 in short). Figure 2.2 exhibits the NTA characterizations of the purified vesicles. Each of the fractions displayed a different particle size distribution. It verified that the vesicles fractionated by SEC from the respective conditioned medium of HEK293 and HEK293+HRAS cells, eluted in different fractions indicating size differences in the sEV subpopulations. The TEM images of the vesicles are shown in Figure 2.3a. The lipid bilayer of the particles can be clearly observed, and the average sizes fell within the range of sEVs. In addition, western blot analyses confirmed the presence of CD81 positive sEVs derived from the cultured HEK293 and HEK293+HRAS cells respectively, shown in Figure 2.3b&c. IZON SEC fraction 7, 8, and 9 for sEVs produced by either cell type demonstrated robust CD81 immunoreactivity and were thus selected for further study.

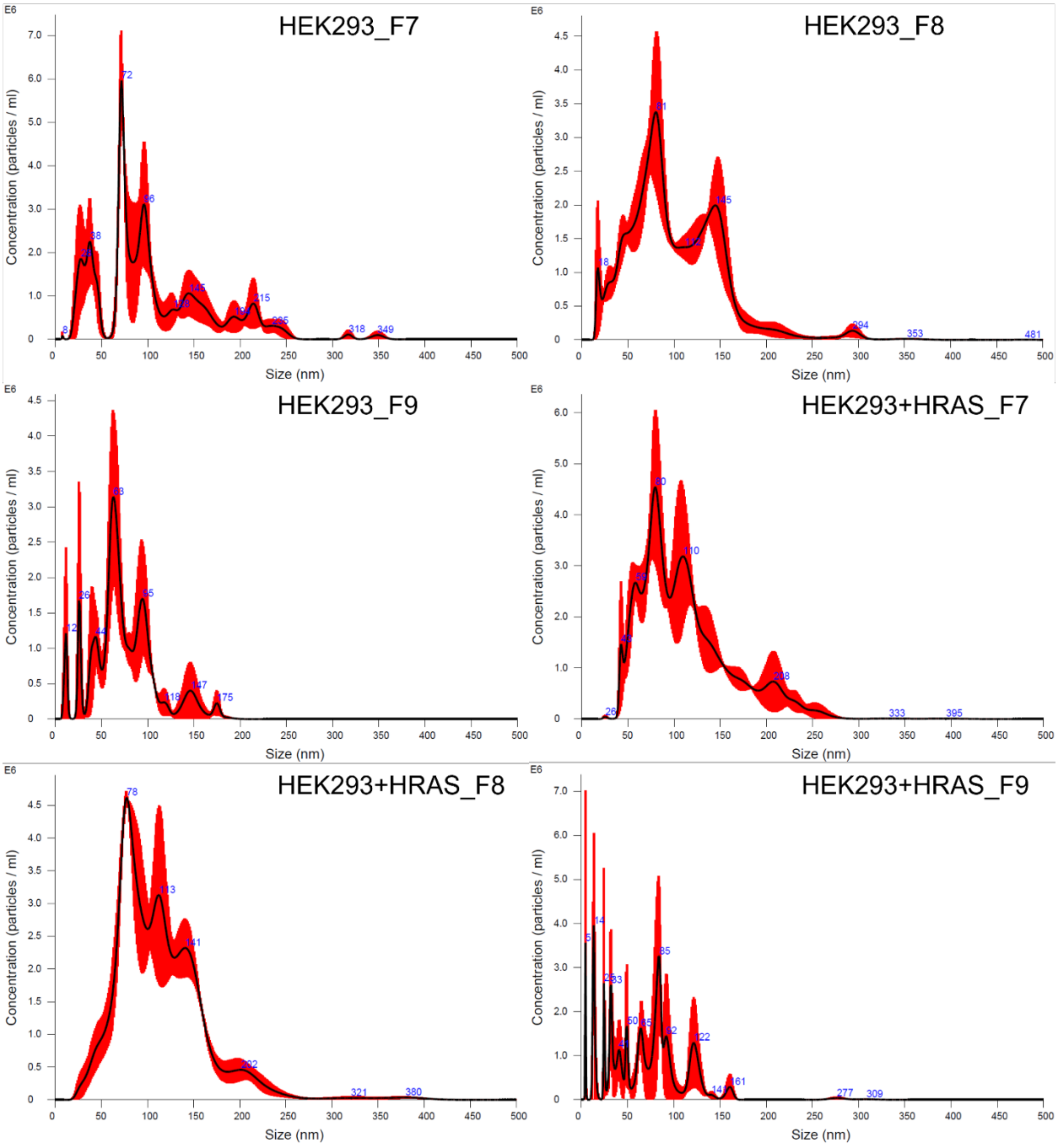

Figure 2.11. NTA characterizations of the isolated sEVs

Data in Figures 2.3d&e illustrates representative SEM images of the SERS substrates before (2d) and after (2e) the introduction of sEV samples. The nano-pyramids are clearly visible in images taken for both conditions. Findings shown in Figure 2.3e show the interaction between

the substrate and the sEVs. Furthermore, dendrites visualized on top of the nano-pyramids were attributed to phosphate buffered saline (PBS) crystals that had formed during the drying process of the sample solution. On the top right portion of Figure 2.3e, is an example of PBS crystals covering the nano-pyramids. The charging effect observed on the SERS substrate caused by PBS crystallization prevented the acquisition of super resolution SEM images of the structures.

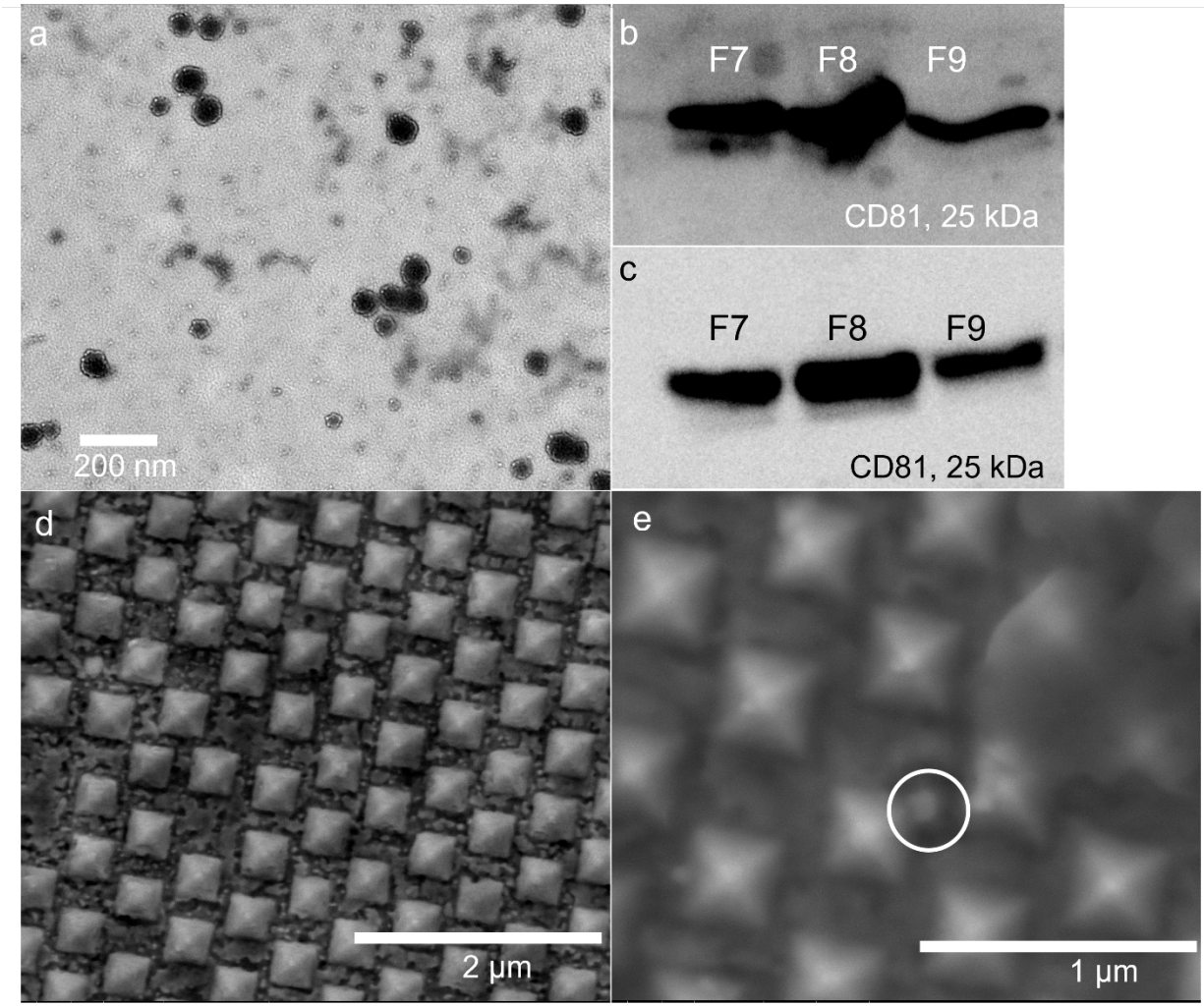

Figure 2.12. Overview of the vesicle samples and the SERS substrate; a) Cryo-EM image of the sEV; b) Western blot of HEK293 derived vesicles; c) Western blot of HEK293+HRAS derived

vesicles; d) SEM image of the SERS substrate; e) SEM image of the SERS substrate after sample introduction

2.3.2. Validating the spectral origin of SERS via a comparison with Mass Spectrometry (MS)

The use of Raman/SERS is recognized as a powerful tool in determining the compositional "fingerprint" of biological substances. Therefore, based on the physical principles of Raman/SERS, it can be inferred that SERS spectra from the collective of Raman-active chemical bonds inside sEVs could serve to resolve the biomolecular composition of individual vesicles. The working hypothesis was tested by correlating the sEV SERS spectral data to their proteomic analysis by classical mass spectrometry (MS). The primary sEV samples used for this application were from cultured HEK293+HRAS fraction 8. This fraction was chosen as it contained the most abundant amount of sEVs as detected by NTA. For SERS mappings, the substrate did not render background peaks and spectra were collected from the each of the SERS mappings by filtering out the spectra with pure noise. It should be noted that the Raman laser spot size during the measurements was larger than the vesicle diameters and that vesicles were invisible due to the instrumental limitation. There is a non-zero probability that occasional ones could be derived from more than one vesicle. Nonetheless, the data indicated that statistically the measurements were based on single vesicles. SERS mappings covered 1.2mm × 1.2mm area which was representatively large considering the size of the sample droplet. The 5 μm separation between the two nearest data spots was to avoid collecting multiple spectra from one vesicle sample which caused data over fitting issue. The spectra collected was observed to be distributed randomly rather than closely concentrated. Also, the spot checks via SEM also indicated that vesicles did not show tendency of self-clustering.

Following the collection of mass spectrometry (MS) data, which revealed the amino acid frequency in the sEVs (Table 2.1), it was further verified how the data correlated to the natural

abundance of the 20 amino acids within proteins that commonly occur in human tissues. Subsequently, the SERS spectra for all 20 amino acids were measured and presented in Figure 2.4. Each presented spectra file consists of the average of 75 spectra collected from the corresponding amino acid sample. Given the fact that different substances display uniquely different Raman interaction cross-sections, the relative SERS responses of all 20 amino acids by comparing the signal-to-noise ratios (SNR) of the SERS spectra were determined next, shown in the middle column of Table 2.1. In making sure to consider the relative natural amino acid abundance and relative SERS activities, a simulated spectrum which mirrors the results of the proteomics analysis of the sEVs was generated. In addition, a fitted spectrum of an averaged spectrum measured from 51 different sEVs derived from fraction 8 sEVs produced by HEK293+HRAS cells was generated, solely considering the relative SERS activities of the 20 amino acids measured. Figure 2.5a displays a comparison between the simulated spectrum derived from MS data and averaged measurements of SERS spectrum derived from the vesicles. The qualitative matching rate for the data was 89% from the perspective of their peak locations. As shown in Figure 2.5b, the reported qualitative matching rate corresponded to 89% of the fitted spectrum and measured spectrum. Similarly, the matching rate increased to 94% when comparing the simulated and fitted spectrum, as shown in Figure 2.5c. By comparing the "assigned" relative amino acid abundances derived from the fitted spectrum to the experimentally MS-derived abundances of amino acids, the averaged deviation corresponded to 9% (Figure 2.5d). When comparing results derived from SERS and proteomics analyses, a high consistency was observed, demonstrating that SERS spectra collected from individual sEVs faithfully reflected the biomolecular composition of the sEVs.

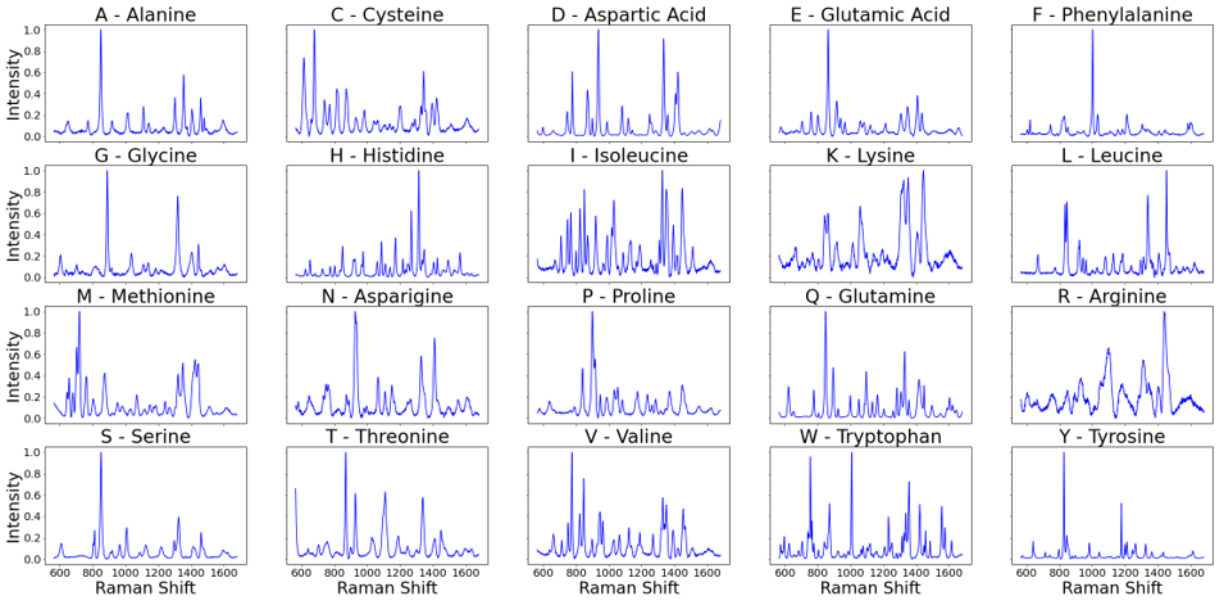

Figure 2.13. Spectra of the 20 amino acids, each averaged from 75 individual spectra

It is noted that the SERS spectral quality from the vesicle measurements were not as high comparing to that of the amino acids. This could be interpreted from the perspective that laser beam diffracted by PBS crystals during the SERS measurements. PBS was introduced in the vesicle sample solution to avoid osmotic discrepancies inside and outside of the vesicles, preventing bursting. Crystals were formed during drying that laying on top of the substrate (suggested by the SEM imaging). Amino acids were suspended in DI water followed immediately by drying and measurements. Works are in progress for reducing the influence of PBS crystals during the measurements. Nonetheless, the data suggested that the spectral features could be extracted from the SERS spectra of the vesicles for comparison and further analyses. There existed the mismatches in the comparative data, especially those derived from peak intensities recorded from simulated and fitted-measured comparisons. Such mismatches could legitimately be attributed to the presence of lipids and/or other elements (e.g., RNA & DNA) present in the vesicles which are not expected to be located in the simulated spectrum only from amino acids.

Furthermore, it is widely recognized that biomolecular substances can display different responses (interaction cross-sections) when tested by SERS and MS. Such non-linear fluctuations could in part explain the observed mismatches between results recorded by the use of the two methods, especially in the peak intensities. Nonetheless, the comparisons helped validate the utility of SERS spectra in determining the biomolecular compositions of sEVs with less than 10% deviation between the measured and the "assigned" amino acid abundance. The findings therefore help to pave the way for future studies aimed at further resolving the link between sEV sizes and their biomolecular composition.

Table 2.1. Coefficients assignments of the 20 amino acids for generating the simulated spectrum

Amino Acid	Amino Acid Abundance Ratio based on Mass-Spec	Relative SERS Activity	Assigned Coefficients for Simulations
A - Alanine	7.88	0.39	3.05
C - Cysteine	2.34	0.29	0.68
D - Aspartic Acid	6.5	1	6.5
E - Glutamic Acid	9.48	0.56	5.26
F - Phenylalanine	4	0.66	2.64
G - Glycine	14.82	0.44	6.48
H - Histidine	3.48	0.63	2.2
I - Isoleucine	5.69	0.18	1.04
K - Lysine	6.58	0.15	0.98
L - Leucine	9.83	0.21	2.05
M - Methionine	2.07	0.46	0.95
N - Asparigine	4.87	0.31	1.52
P - Proline	4.67	0.49	2.3
Q - Glutamine	7	0.85	5.94
R - Arginine	7.28	0.13	0.95
S - Serine	15.01	0.83	12.53
T - Threonine	7.3	0.44	3.23
V - Valine	7.15	0.28	1.99
W - Tryptophan	1	0.35	0.35
Y - Tyrosine	3.93	0.83	3.26

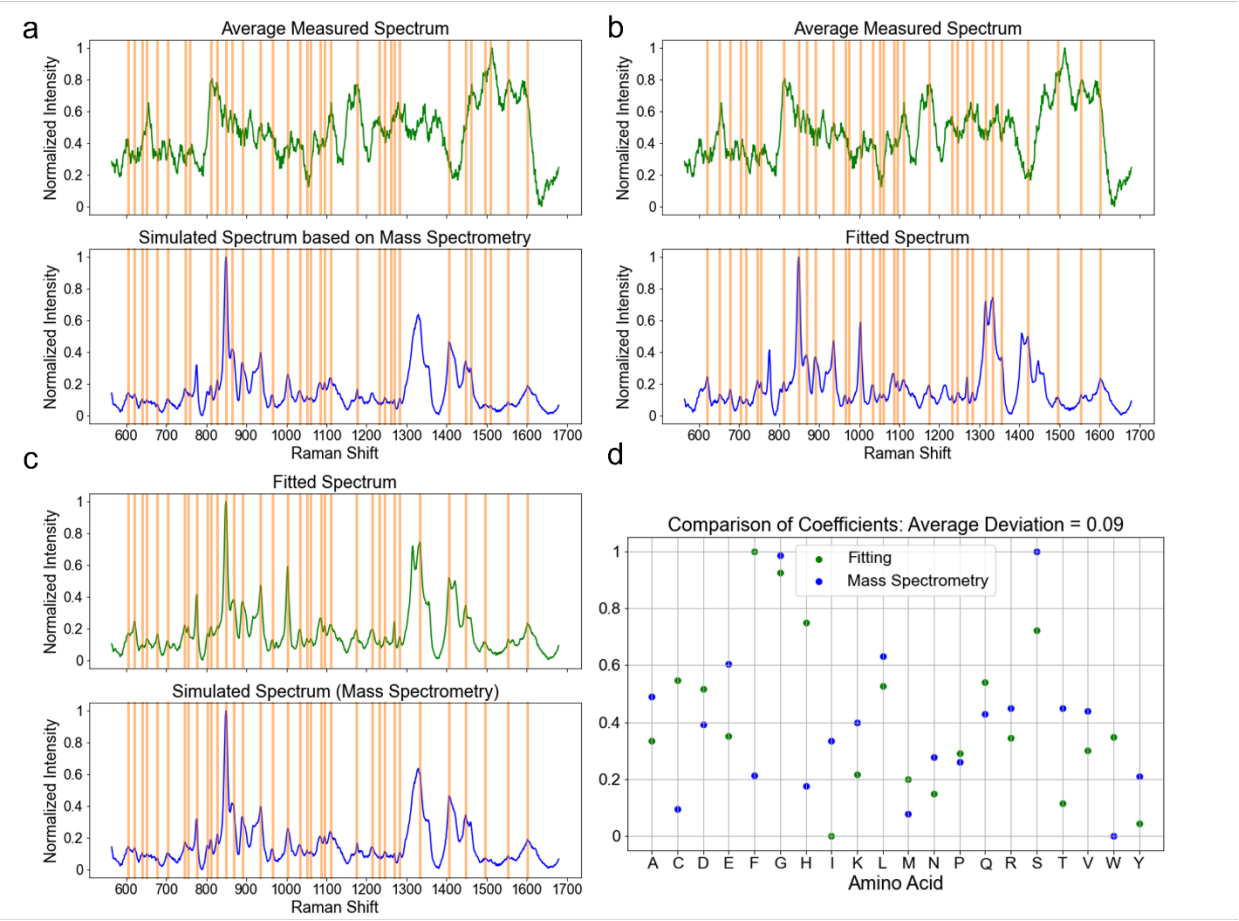

Figure 2.14. Comparison between the SERS data and MS data; a) peak location comparison between the averaged spectrum from the spectra measured from 68 vesicles (HEK293+HRAS F8) and the simulated spectrum; b) peak location comparison between the fitted spectrum and averaged measured spectrum; c) peak location comparison between the fitted spectrum and the simulated spectrum; d) comparison of the coefficients assigned for the 20 amino acids between the fitted and simulated spectra

2.3.3. SERS Analyzing Individual sEVs Isolated Based on Size Discrepancies

Data was collected using sEVs that had been isolated into three different sized groups (fraction 7-9) based on their elution prolife from the SEC column shown in Figure 2.2b&c. The

number of spectra collected by SERS from HEK293-derived fractions 7, 8, and 9 consisted of 59, 115, and 47 respectively. Similarly, the number of HEK293+HRAS derived vesicles interrogated by SERS consisted of 65, 68, 31 data spots respectively. The assessment of biochemical contents in vesicles was carried out by comparing the spectral overlaps among the SERS spectra collected from individual sEVs. Results were projected onto a two-dimensional map using the LDA algorithm, as shown in Figure 2.6. Each dot displayed in the plot represents a SERS spectrum recorded from an individual vesicle. The spectral variations between the two types of vesicles were measured by their Euclidean distances. In this case, the spectral difference was revealed by the magnitude of the distance between the data dots. As shown in Figure 2.6a, the LDA map comparing the SERS spectra collected from HEK293-derived sEVs exhibited some level of internal spectral difference within each of the size groups. These data reveal the magnitude of differences in the biochemical content among individual vesicles. Importantly, when comparing the LDA maps between vesicles of the three size groups, no significant overlap was detected, suggesting that the external spectral differences among the three groups were larger than the spectral differences among individual sEVs within each group. The data demonstrate that size-based fractionations of a bulk preparation of HEK293 cell-derived EVs resulted in the purification of homogeneous classes of sEVs with a defined and non-overlapping biomolecular composition as determined by SERS analysis. Furthermore, a similar pattern emerged when analyzing the SERS spectral features of HEK293+HRAS cell-derived vesicles using the LDA algorithm, shown in Figure 2.6b. In general, based on the LDA maps, less than 16% of the vesicle spectra were found "common" between different size-distribution groups.

Interestingly, the data demonstrate that sEV groups with distinct size distributions can be differentiated based on their biomolecular composition through their SERS spectral signature. The

data suggesting the existences of “common” vesicles present across different size groups likely derive from limitations associated with SEC to fully resolve sEVs exclusively according to distinct sizes. Indeed, the data demonstrate the existence of vesicles, which are similarly sized, that elute across different fractions eluting from the SEC column. These observations highlight limitations associated with the use of current SEC methodologies for the isolation of distinct sEV subpopulations. Given the situation, the single-vesicle SERS platform offers an opportunity to further characterize the variation in biomolecular composition among individual sEVs within a subpopulation of vesicles isolated within a SEC sub-fraction. The methodology thus addresses a gap in the field of sEV biology that is limited to bulk and/or specific marker-based analysis.

Lastly, to evaluate the distribution of individual sEVs within the comparison of HEK293- and HEK293+HRAS-derived vesicles, an LDA map by pooling the SERS spectra contributed by all three size groups of both cell types was produced, shown in Figure 2.6c. This approach revealed an overlap between the two clusters, suggesting that the HRAS transgene expressed in HEK293+HRAS cells caused the releases of vesicles with both similar and different biomolecular composition compared to those released by HEK293 cells. By retrieving the spectral sources of the dots produced from the HEK293+HRAS sEV group that had LD scores within the range of those produced from the HEK293 sEV group, a comparison noting the distribution of the individual vesicles (Figure 2.6d) was generated. Using this approach, it was noted that the overlapping rates increased from 6.2% to 12.9% across fractions F7 to F9 in the HRAS group. This finding suggests that the biological perturbations caused HRAS expression in HEK293 cells results in altering the biomolecular composition of a defined population of sEVs released by the cells. However, a substantial portion of the sEVs shared similar biomolecular compositions with the ones released by HEK293 control cells. Interestingly, when considering that SERS spectral

signatures are rooted in the biomolecular composition of the vesicles, data demonstrate that sEVs in larger size groups produced by HRAS-cells shared less biomolecular similarity with vesicles of similar size released by the control HEK293 cells than those sEVs with smaller sizes.

Comparing to other published works of analyzing EVs using SERS including but not limited to the ones mentioned previously[55-60], this study focused on the subpopulations of the sEVs isolated based on size distributions. Also, the study attempted to provide experimental data indications that the origins of the SERS spectral features are from the biomolecules of the vesicles by linking the results between SERS and MS. Collectively, results of this study support the value of the single-vesicle based SERS platform to explore the biomolecular composition of individual sEVs with different sizes and/or released by different parental cells. Currently, such information is yet to be established and recognized by the community. The fact that our platform is capable of distinguishing vesicles released by two types of cells with only one gene alternation underscores the potential of this system for disease screening, diagnosis and monitoring. A limitation of this single vesicle analysis system includes the low throughput capacity for sEV scanning by SERS. Current work in our laboratory aims to improve both the areal density available to the vesicles when introduced to the substrate and the efficiency of the SERS measurement process. Nonetheless, the results illustrate the value of the SERS platform as a sensitive sEV detecting technique that can differentiate nano-sized vesicles based on their biomolecular composition at the single vesicle level. The approach overcomes the current challenges associated with population averaging in the study of bulk EV biology. The technology could therefore not only benefit sEV-based disease diagnosis/screening, but also facilitate further studies investigating the biogenesis and activity of sub-fractionation of vesicles released by the same and/or different parental cells.

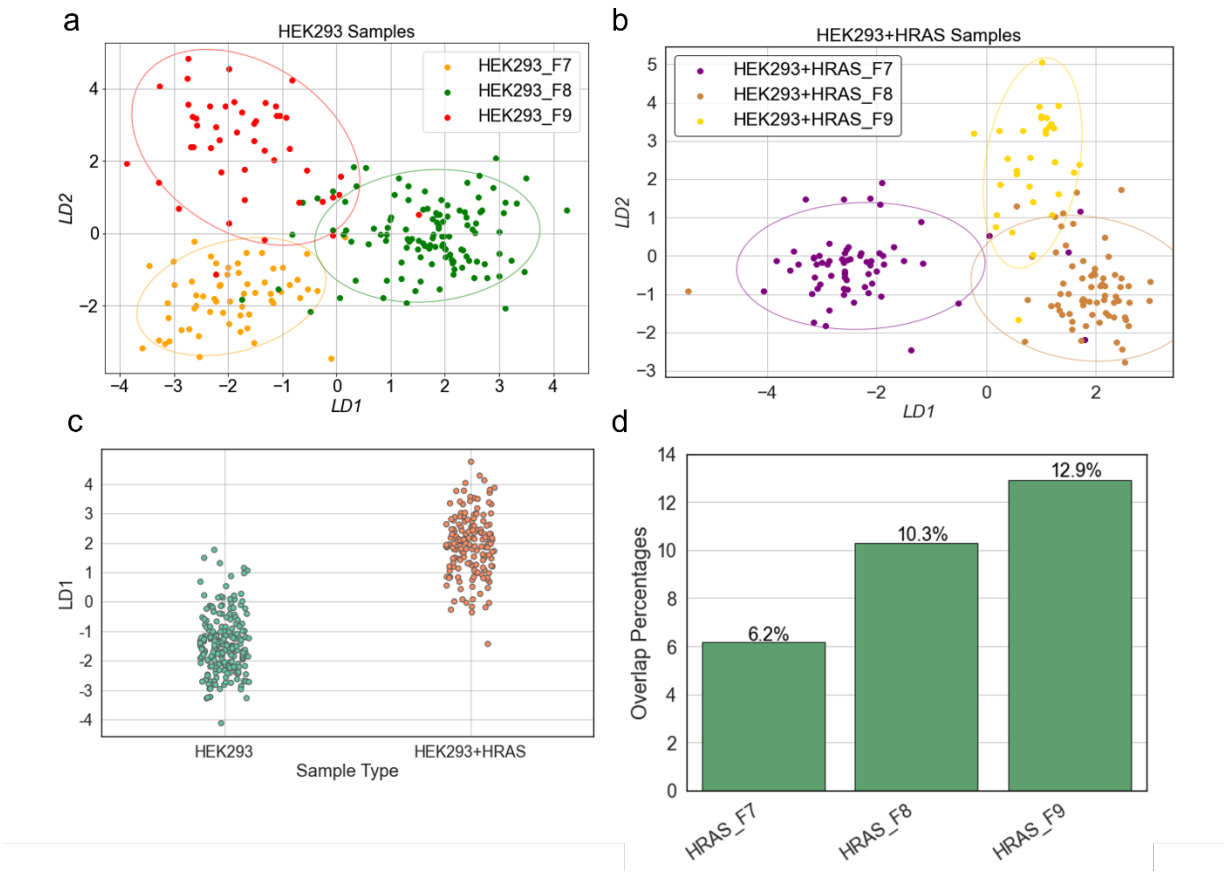

Figure 2.15. LDA analysis results of SERS signatures obtained from sEVs of different size groups; a) LDA of the spectral signatures obtained from vesicles of HEK293 fractions 7, 8, 9 respectively; b) LDA of the spectral signatures obtained from vesicles of HEK293+HRAS fractions 7, 8, 9 respectively; c) Polling the SERS signature from the vesicles of the three size fractions from the HEK293 and HEK293+HRAS respectively; d) The SERS signature overlapping rate of each fraction from the HRAS group

2.4. Conclusion

Findings from this study validate the use of SERS for investigating the biomolecular composition of small extracellular vesicles (sEVs) produced by cultured cells. In using this approach, the work attempted to uncover a strong correlation between the sEV size and their biomolecular composition. Specifically, 16% or less vesicles in each of the individual size groups displayed an overlapping biomolecular composition with sEVs of different size groups. The <10% deviation of data derived from sEV analysis by SERS and mass spectrometry supports the robustness of SERS as a method to fingerprint the biomolecular composition of sEVs. Future studies using SERS will allow to further understand the biogenesis, diversity and functional consequences of vesicles released by cultured cells and those in more complex biofluids.

2.5. Reference

1. Willms, E.; Cabanas, C.; Mager, I.; Wood, M. J. A.; Vader, P., Extracellular Vesicle Heterogeneity: Subpopulations, Isolation Techniques, and Diverse Functions in Cancer Progression. *Front Immunol* **2018,** *9*, 738.

2. Smith, Z. J.; Lee, C.; Rojalin, T.; Carney, R. P.; Hazari, S.; Knudson, A.; Lam, K.; Saari, H.; Ibanez, E. L.; Viitala, T.; Laaksonen, T.; Yliperttula, M.; Wachsmann-Hogiu, S., Single exosome study reveals subpopulations distributed among cell lines with variability related to membrane content. *J Extracell Vesicles* **2015,** *4*, 28533.

3. Willms, E.; Johansson, H. J.; Mager, I.; Lee, Y.; Blomberg, K. E.; Sadik, M.; Alaarg, A.; Smith, C. I.; Lehtio, J.; El Andaloussi, S.; Wood, M. J.; Vader, P., Cells release subpopulations of exosomes with distinct molecular and biological properties. *Sci Rep* **2016,** *6*, 22519.

4. Choi, D.; Montermini, L.; Jeong, H.; Sharma, S.; Meehan, B.; Rak, J., Mapping Subpopulations of Cancer Cell-Derived Extracellular Vesicles and Particles by Nano-Flow Cytometry. *ACS Nano* **2019,** *13* (9), 10499-10511.

5. Singh, K.; Nalabotala, R.; Koo, K. M.; Bose, S.; Nayak, R.; Shiddiky, M. J. A., Separation of distinct exosome subpopulations: isolation and characterization approaches and their associated challenges. *Analyst* **2021,** *146* (12), 3731-3749.

6. Lee, S. S.; Won, J. H.; Lim, G. J.; Han, J.; Lee, J. Y.; Cho, K. O.; Bae, Y. K., A novel population of extracellular vesicles smaller than exosomes promotes cell proliferation. *Cell Commun Signal* **2019,** *17* (1), 95.

7. Panagopoulou, M. S.; Wark, A. W.; Birch, D. J. S.; Gregory, C. D., Phenotypic analysis of extracellular vesicles: a review on the applications of fluorescence. *J Extracell Vesicles* **2020,** *9* (1), 1710020.

8. Crescitelli, R.; Lasser, C.; Lotvall, J., Isolation and characterization of extracellular vesicle subpopulations from tissues. *Nat Protoc* **2021,** *16* (3), 1548-1580.

9. Wang, W.; Luo, J.; Wang, S., Recent Progress in Isolation and Detection of Extracellular Vesicles for Cancer Diagnostics. *Adv Healthc Mater* **2018,** *7* (20), e1800484.

10. Konoshenko, M. Y.; Lekchnov, E. A.; Vlassov, A. V.; Laktionov, P. P., Isolation of Extracellular Vesicles: General Methodologies and Latest Trends. *Biomed Res Int* **2018,** *2018*, 8545347.

11. Johnstone, R. M.; Adam, M.; Hammond, J. R.; Orr, L.; Turbide, C., Vesicle formation during reticulocyte maturation. Association of plasma membrane activities with released vesicles (exosomes). *Journal of Biological Chemistry* **1987,** *262* (19), 9412-9420.

12. Thery, C.; Witwer, K. W.; Aikawa, E.; Alcaraz, M. J.; Anderson, J. D.; Andriantsitohaina, R.; Antoniou, A.; Arab, T.; Archer, F.; Atkin-Smith, G. K.; Ayre, D. C.; Bach, J. M.; Bachurski, D.; Baharvand, H.; Balaj, L.; Baldacchino, S.; Bauer, N. N.; Baxter, A. A.; Bebawy, M.; Beckham, C.; Bedina Zavec, A.; Benmoussa, A.; Berardi, A. C.; Bergese, P.; Bielska, E.; Blenkiron, C.; Bobis-Wozowicz, S.; Boilard, E.; Boireau, W.; Bongiovanni, A.; Borras, F. E.; Bosch, S.; Boulanger, C. M.; Breakefield, X.; Breglio, A. M.; Brennan, M. A.; Brigstock, D. R.; Brisson, A.; Broekman, M. L.; Bromberg, J. F.; Bryl-Gorecka, P.; Buch, S.; Buck, A. H.; Burger, D.; Busatto, S.; Buschmann, D.; Bussolati, B.; Buzas, E. I.; Byrd, J. B.; Camussi, G.; Carter, D. R.; Caruso, S.; Chamley, L. W.; Chang, Y. T.; Chen, C.; Chen, S.;

Cheng, L.; Chin, A. R.; Clayton, A.; Clerici, S. P.; Cocks, A.; Cocucci, E.; Coffey, R. J.; Cordeiro-da-Silva, A.; Couch, Y.; Coumans, F. A.; Coyle, B.; Crescitelli, R.; Criado, M. F.; D'Souza-Schorey, C.; Das, S.; Datta Chaudhuri, A.; de Candia, P.; De Santana, E. F.; De Wever, O.; Del Portillo, H. A.; Demaret, T.; Deville, S.; Devitt, A.; Dhondt, B.; Di Vizio, D.; Dieterich, L. C.; Dolo, V.; Dominguez Rubio, A. P.; Dominici, M.; Dourado, M. R.; Driedonks, T. A.; Duarte, F. V.; Duncan, H. M.; Eichenberger, R. M.; Ekstrom, K.; El Andaloussi, S.; Elie-Caille, C.; Erdbrugger, U.; Falcon-Perez, J. M.; Fatima, F.; Fish, J. E.; Flores-Bellver, M.; Forsonits, A.; Frelet-Barrand, A.; Fricke, F.; Fuhrmann, G.; Gabrielsson, S.; Gamez-Valero, A.; Gardiner, C.; Gartner, K.; Gaudin, R.; Gho, Y. S.; Giebel, B.; Gilbert, C.; Gimona, M.; Giusti, I.; Goberdhan, D. C.; Gorgens, A.; Gorski, S. M.; Greening, D. W.; Gross, J. C.; Gualerzi, A.; Gupta, G. N.; Gustafson, D.; Handberg, A.; Haraszti, R. A.; Harrison, P.; Hegyesi, H.; Hendrix, A.; Hill, A. F.; Hochberg, F. H.; Hoffmann, K. F.; Holder, B.; Holthofer, H.; Hosseinkhani, B.; Hu, G.; Huang, Y.; Huber, V.; Hunt, S.; Ibrahim, A. G.; Ikezu, T.; Inal, J. M.; Isin, M.; Ivanova, A.; Jackson, H. K.; Jacobsen, S.; Jay, S. M.; Jayachandran, M.; Jenster, G.; Jiang, L.; Johnson, S. M.; Jones, J. C.; Jong, A.; Jovanovic-Talisman, T.; Jung, S.; Kalluri, R.; Kano, S. I.; Kaur, S.; Kawamura, Y.; Keller, E. T.; Khamari, D.; Khomyakova, E.; Khvorova, A.; Kierulf, P.; Kim, K. P.; Kislinger, T.; Klingeborn, M.; Klinke, D. J., 2nd; Kornek, M.; Kosanovic, M. M.; Kovacs, A. F.; Kramer-Albers, E. M.; Krasemann, S.; Krause, M.; Kurochkin, I. V.; Kusuma, G. D.; Kuypers, S.; Laitinen, S.; Langevin, S. M.; Languino, L. R.; Lannigan, J.; Lasser, C.; Laurent, L. C.; Lavieu, G.; Lazaro-Ibanez, E.; Le Lay, S.; Lee, M. S.; Lee, Y. X. F.; Lemos, D. S.; Lenassi, M.; Leszczynska, A.; Li, I. T.; Liao, K.; Libregts, S. F.; Ligeti, E.; Lim, R.; Lim, S. K.; Line, A.; Linnemannstons, K.; Llorente, A.; Lombard, C. A.; Lorenowicz, M. J.; Lorincz, A. M.; Lotvall, J.; Lovett, J.; Lowry, M. C.; Loyer, X.; Lu, Q.; Lukomska, B.; Lunavat,

T. R.; Maas, S. L.; Malhi, H.; Marcilla, A.; Mariani, J.; Mariscal, J.; Martens-Uzunova, E. S.; Martin-Jaular, L.; Martinez, M. C.; Martins, V. R.; Mathieu, M.; Mathivanan, S.; Maugeri, M.; McGinnis, L. K.; McVey, M. J.; Meckes, D. G., Jr.; Meehan, K. L.; Mertens, I.; Minciacchi, V. R.; Moller, A.; Moller Jorgensen, M.; Morales-Kastresana, A.; Morhayim, J.; Mullier, F.; Muraca, M.; Musante, L.; Mussack, V.; Muth, D. C.; Myburgh, K. H.; Najrana, T.; Nawaz, M.; Nazarenko, I.; Nejsum, P.; Neri, C.; Neri, T.; Nieuwland, R.; Nimrichter, L.; Nolan, J. P.; Nolte-'t Hoen, E. N.; Noren Hooten, N.; O'Driscoll, L.; O'Grady, T.; O'Loghlen, A.; Ochiya, T.; Olivier, M.; Ortiz, A.; Ortiz, L. A.; Osteikoetxea, X.; Ostergaard, O.; Ostrowski, M.; Park, J.; Pegtel, D. M.; Peinado, H.; Perut, F.; Pfaffl, M. W.; Phinney, D. G.; Pieters, B. C.; Pink, R. C.; Pisetsky, D. S.; Pogge von Strandmann, E.; Polakovicova, I.; Poon, I. K.; Powell, B. H.; Prada, I.; Pulliam, L.; Quesenberry, P.; Radeghieri, A.; Raffai, R. L.; Raimondo, S.; Rak, J.; Ramirez, M. I.; Raposo, G.; Rayyan, M. S.; Regev-Rudzki, N.; Ricklefs, F. L.; Robbins, P. D.; Roberts, D. D.; Rodrigues, S. C.; Rohde, E.; Rome, S.; Rouschop, K. M.; Rughetti, A.; Russell, A. E.; Saa, P.; Sahoo, S.; Salas-Huenuleo, E.; Sanchez, C.; Saugstad, J. A.; Saul, M. J.; Schiffelers, R. M.; Schneider, R.; Schoyen, T. H.; Scott, A.; Shahaj, E.; Sharma, S.; Shatnyeva, O.; Shekari, F.; Shelke, G. V.; Shetty, A. K.; Shiba, K.; Siljander, P. R.; Silva, A. M.; Skowronek, A.; Snyder, O. L., 2nd; Soares, R. P.; Sodar, B. W.; Soekmadji, C.; Sotillo, J.; Stahl, P. D.; Stoorvogel, W.; Stott, S. L.; Strasser, E. F.; Swift, S.; Tahara, H.; Tewari, M.; Timms, K.; Tiwari, S.; Tixeira, R.; Tkach, M.; Toh, W. S.; Tomasini, R.; Torrecilhas, A. C.; Tosar, J. P.; Toxavidis, V.; Urbanelli, L.; Vader, P.; van Balkom, B. W.; van der Grein, S. G.; Van Deun, J.; van Herwijnen, M. J.; Van Keuren-Jensen, K.; van Niel, G.; van Royen, M. E.; van Wijnen, A. J.; Vasconcelos, M. H.; Vechetti, I. J., Jr.; Veit, T. D.; Vella, L. J.; Velot, E.; Verweij, F. J.; Vestad, B.; Vinas, J. L.; Visnovitz, T.; Vukman, K. V.; Wahlgren, J.; Watson, D. C.; Wauben,

M. H.; Weaver, A.; Webber, J. P.; Weber, V.; Wehman, A. M.; Weiss, D. J.; Welsh, J. A.; Wendt, S.; Wheelock, A. M.; Wiener, Z.; Witte, L.; Wolfram, J.; Xagorari, A.; Xander, P.; Xu, J.; Yan, X.; Yanez-Mo, M.; Yin, H.; Yuana, Y.; Zappulli, V.; Zarubova, J.; Zekas, V.; Zhang, J. Y.; Zhao, Z.; Zheng, L.; Zheutlin, A. R.; Zickler, A. M.; Zimmermann, P.; Zivkovic, A. M.; Zocco, D.; Zuba-Surma, E. K., Minimal information for studies of extracellular vesicles 2018 (MISEV2018): a position statement of the International Society for Extracellular Vesicles and update of the MISEV2014 guidelines. *J Extracell Vesicles* **2018,** *7* (1), 1535750.

13. Doyle, L. M.; Wang, M. Z., Overview of Extracellular Vesicles, Their Origin, Composition, Purpose, and Methods for Exosome Isolation and Analysis. *Cells* **2019,** *8* (7).

14. Ludwig, N.; Whiteside, T. L.; Reichert, T. E., Challenges in Exosome Isolation and Analysis in Health and Disease. *Int J Mol Sci* **2019,** *20* (19).

15. Zhang, M.; Jin, K.; Gao, L.; Zhang, Z.; Li, F.; Zhou, F.; Zhang, L., Methods and Technologies for Exosome Isolation and Characterization. *Small Methods* **2018,** *2* (9).

16. Yang, D.; Zhang, W.; Zhang, H.; Zhang, F.; Chen, L.; Ma, L.; Larcher, L. M.; Chen, S.; Liu, N.; Zhao, Q.; Tran, P. H. L.; Chen, C.; Veedu, R. N.; Wang, T., Progress, opportunity, and perspective on exosome isolation - efforts for efficient exosome-based theranostics. *Theranostics* **2020,** *10* (8), 3684-3707.

17. Linares, R.; Tan, S.; Gounou, C.; Arraud, N.; Brisson, A. R., High-speed centrifugation induces aggregation of extracellular vesicles. *J Extracell Vesicles* **2015,** *4*, 29509.

18. Duong, P.; Chung, A.; Bouchareychas, L.; Raffai, R. L., Cushioned-Density Gradient Ultracentrifugation (C-DGUC) improves the isolation efficiency of extracellular vesicles. *PLoS One* **2019,** *14* (4), e0215324.

19. Chen, J.; Li, P.; Zhang, T.; Xu, Z.; Huang, X.; Wang, R.; Du, L., Review on Strategies and Technologies for Exosome Isolation and Purification. *Front Bioeng Biotechnol* **2021,** *9*, 811971.

20. Van Deun, J.; Mestdagh, P.; Sormunen, R.; Cocquyt, V.; Vermaelen, K.; Vandesompele, J.; Bracke, M.; De Wever, O.; Hendrix, A., The impact of disparate isolation methods for extracellular vesicles on downstream RNA profiling. *J Extracell Vesicles* **2014,** *3*.

21. Lobb, R. J.; Becker, M.; Wen, S. W.; Wong, C. S.; Wiegmans, A. P.; Leimgruber, A.; Moller, A., Optimized exosome isolation protocol for cell culture supernatant and human plasma. *J Extracell Vesicles* **2015,** *4*, 27031.

22. Witwer, K. W.; Buzas, E. I.; Bemis, L. T.; Bora, A.; Lasser, C.; Lotvall, J.; Nolte-'t Hoen, E. N.; Piper, M. G.; Sivaraman, S.; Skog, J.; Thery, C.; Wauben, M. H.; Hochberg, F., Standardization of sample collection, isolation and analysis methods in extracellular vesicle research. *J Extracell Vesicles* **2013,** *2*.

23. Cho, S.; Yang, H. C.; Rhee, W. J., Development and comparative analysis of human urine exosome isolation strategies. *Process Biochemistry* **2020,** *88*, 197-203.

24. Xu, R.; Greening, D. W.; Zhu, H. J.; Takahashi, N.; Simpson, R. J., Extracellular vesicle isolation and characterization: toward clinical application. *J Clin Invest* **2016,** *126* (4), 1152-62.

25. Monguio-Tortajada, M.; Galvez-Monton, C.; Bayes-Genis, A.; Roura, S.; Borras, F. E., Extracellular vesicle isolation methods: rising impact of size-exclusion chromatography. *Cell Mol Life Sci* **2019,** *76* (12), 2369-2382.

26. Yang, Y.; Wang, Y.; Wei, S.; Zhou, C.; Yu, J.; Wang, G.; Wang, W.; Zhao, L., Extracellular vesicles isolated by size-exclusion chromatography present suitability for RNomics analysis in plasma. *J Transl Med* **2021,** *19* (1), 104.

27. Davis, C. N.; Phillips, H.; Tomes, J. J.; Swain, M. T.; Wilkinson, T. J.; Brophy, P. M.; Morphew, R. M., The importance of extracellular vesicle purification for downstream analysis: A comparison of differential centrifugation and size exclusion chromatography for helminth pathogens. *PLoS Negl Trop Dis* **2019,** *13* (2), e0007191.

28. Nordin, J. Z.; Lee, Y.; Vader, P.; Mager, I.; Johansson, H. J.; Heusermann, W.; Wiklander, O. P.; Hallbrink, M.; Seow, Y.; Bultema, J. J.; Gilthorpe, J.; Davies, T.; Fairchild, P. J.; Gabrielsson, S.; Meisner-Kober, N. C.; Lehtio, J.; Smith, C. I.; Wood, M. J.; El Andaloussi, S., Ultrafiltration with size-exclusion liquid chromatography for high yield isolation of extracellular vesicles preserving intact biophysical and functional properties. *Nanomedicine* **2015,** *11* (4), 879-83.

29. Xiang, X.; Guan, F.; Jiao, F.; Li, H.; Zhang, W.; Zhang, Y.; Qin, W., A new urinary exosome enrichment method by a combination of ultrafiltration and TiO2 nanoparticles. *Anal Methods* **2021,** *13* (13), 1591-1600.

30. Mol, E. A.; Goumans, M. J.; Doevendans, P. A.; Sluijter, J. P. G.; Vader, P., Higher functionality of extracellular vesicles isolated using size-exclusion chromatography compared to ultracentrifugation. *Nanomedicine* **2017,** *13* (6), 2061-2065.

31. Sidhom, K.; Obi, P. O.; Saleem, A., A Review of Exosomal Isolation Methods: Is Size Exclusion Chromatography the Best Option? *Int J Mol Sci* **2020,** *21* (18).

32. Liga, A.; Vliegenthart, A. D.; Oosthuyzen, W.; Dear, J. W.; Kersaudy-Kerhoas, M., Exosome isolation: a microfluidic road-map. *Lab Chip* **2015,** *15* (11), 2388-94.

33. Yeo, J. C.; Kenry; Zhao, Z.; Zhang, P.; Wang, Z.; Lim, C. T., Label-free extraction of extracellular vesicles using centrifugal microfluidics. *Biomicrofluidics* **2018,** *12* (2), 024103.

34. Gardiner, C.; Di Vizio, D.; Sahoo, S.; Thery, C.; Witwer, K. W.; Wauben, M.; Hill, A. F., Techniques used for the isolation and characterization of extracellular vesicles: results of a worldwide survey. *J Extracell Vesicles* **2016,** *5*, 32945.

35. MacPhee, D. J., Methodological considerations for improving Western blot analysis. *J Pharmacol Toxicol Methods* **2010,** *61* (2), 171-7.

36. Allelein, S.; Medina-Perez, P.; Lopes, A. L. H.; Rau, S.; Hause, G.; Kolsch, A.; Kuhlmeier, D., Potential and challenges of specifically isolating extracellular vesicles from heterogeneous populations. *Sci Rep* **2021,** *11* (1), 11585.

37. Yoshioka, Y.; Kosaka, N.; Konishi, Y.; Ohta, H.; Okamoto, H.; Sonoda, H.; Nonaka, R.; Yamamoto, H.; Ishii, H.; Mori, M.; Furuta, K.; Nakajima, T.; Hayashi, H.; Sugisaki, H.; Higashimoto, H.; Kato, T.; Takeshita, F.; Ochiya, T., Ultra-sensitive liquid biopsy of circulating extracellular vesicles using ExoScreen. *Nat Commun* **2014,** *5*, 3591.

38. Dahiya, B.; Khan, A.; Mor, P.; Kamra, E.; Singh, N.; Gupta, K. B.; Sheoran, A.; Sreenivas, V.; Mehta, P. K., Detection of Mycobacterium tuberculosis lipoarabinomannan and CFP-10 (Rv3874) from urinary extracellular vesicles of tuberculosis patients by immuno-PCR. *Pathog Dis* **2019,** *77* (5).

39. Sunkara, V.; Kim, C. J.; Park, J.; Woo, H. K.; Kim, D.; Ha, H. K.; Kim, M. H.; Son, Y.; Kim, J. R.; Cho, Y. K., Fully Automated, Label-Free Isolation of Extracellular Vesicles from Whole Blood for Cancer Diagnosis and Monitoring. *Theranostics* **2019,** *9* (7), 1851-1863.

40. Serrano-Pertierra, E.; Oliveira-Rodriguez, M.; Matos, M.; Gutierrez, G.; Moyano, A.; Salvador, M.; Rivas, M.; Blanco-Lopez, M. C., Extracellular Vesicles: Current Analytical Techniques for Detection and Quantification. *Biomolecules* **2020,** *10* (6).

41. Pospichalova, V.; Svoboda, J.; Dave, Z.; Kotrbova, A.; Kaiser, K.; Klemova, D.; Ilkovics, L.; Hampl, A.; Crha, I.; Jandakova, E.; Minar, L.; Weinberger, V.; Bryja, V., Simplified protocol for flow cytometry analysis of fluorescently labeled exosomes and microvesicles using dedicated flow cytometer. *J Extracell Vesicles* **2015,** *4*, 25530.

42. van der Pol, E.; van Gemert, M. J.; Sturk, A.; Nieuwland, R.; van Leeuwen, T. G., Single vs. swarm detection of microparticles and exosomes by flow cytometry. *J Thromb Haemost* **2012,** *10* (5), 919-30.

43. Andronico, L. A.; Jiang, Y.; Jung, S. R.; Fujimoto, B. S.; Vojtech, L.; Chiu, D. T., Sizing Extracellular Vesicles Using Membrane Dyes and a Single Molecule-Sensitive Flow Analyzer. *Anal Chem* **2021,** *93* (14), 5897-5905.

44. Tian, Y.; Ma, L.; Gong, M.; Su, G.; Zhu, S.; Zhang, W.; Wang, S.; Li, Z.; Chen, C.; Li, L.; Wu, L.; Yan, X., Protein Profiling and Sizing of Extracellular Vesicles from Colorectal Cancer Patients via Flow Cytometry. *ACS Nano* **2018,** *12* (1), 671-680.

45. Theodoraki, M. N.; Hong, C. S.; Donnenberg, V. S.; Donnenberg, A. D.; Whiteside, T. L., Evaluation of Exosome Proteins by on-Bead Flow Cytometry. *Cytometry A* **2021,** *99* (4), 372-381.

46. van der Pol, E.; Coumans, F.; Varga, Z.; Krumrey, M.; Nieuwland, R., Innovation in detection of microparticles and exosomes. *J Thromb Haemost* **2013,** *11 Suppl 1*, 36-45.

47. Morales-Kastresana, A.; Telford, B.; Musich, T. A.; McKinnon, K.; Clayborne, C.; Braig, Z.; Rosner, A.; Demberg, T.; Watson, D. C.; Karpova, T. S.; Freeman, G. J.; DeKruyff, R. H.; Pavlakis, G. N.; Terabe, M.; Robert-Guroff, M.; Berzofsky, J. A.; Jones, J. C., Labeling Extracellular Vesicles for Nanoscale Flow Cytometry. *Sci Rep* **2017,** *7* (1), 1878.

48. Jia, S.; Zocco, D.; Samuels, M. L.; Chou, M. F.; Chammas, R.; Skog, J.; Zarovni, N.; Momen-Heravi, F.; Kuo, W. P., Emerging technologies in extracellular vesicle-based molecular diagnostics. *Expert Rev Mol Diagn* **2014,** *14* (3), 307-21.

49. Moldovan, L.; Batte, K.; Wang, Y.; Wisler, J.; Piper, M., Analyzing the circulating microRNAs in exosomes/extracellular vesicles from serum or plasma by qRT-PCR. *Methods Mol Biol* **2013,** *1024*, 129-45.

50. Burns, G.; Brooks, K.; Wildung, M.; Navakanitworakul, R.; Christenson, L. K.; Spencer, T. E., Extracellular vesicles in luminal fluid of the ovine uterus. *PLoS One* **2014,** *9* (3), e90913.

51. Bai, Y.; Qu, Y.; Wu, Z.; Ren, Y.; Cheng, Z.; Lu, Y.; Hu, J.; Lou, J.; Zhao, J.; Chen, C.; Mao, H., Absolute quantification and analysis of extracellular vesicle lncRNAs from the peripheral blood of patients with lung cancer based on multi-colour fluorescence chip-based digital PCR. *Biosens Bioelectron* **2019,** *142*, 111523.

52. Crossland, R. E.; Norden, J.; Bibby, L. A.; Davis, J.; Dickinson, A. M., Evaluation of optimal extracellular vesicle small RNA isolation and qRT-PCR normalisation for serum and urine. *J Immunol Methods* **2016,** *429*, 39-49.

53. Kong, A. T.; Leprevost, F. V.; Avtonomov, D. M.; Mellacheruvu, D.; Nesvizhskii, A. I., MSFragger: ultrafast and comprehensive peptide identification in mass spectrometry-based proteomics. *Nat Methods* **2017,** *14* (5), 513-520.

54. Teo, G. C.; Polasky, D. A.; Yu, F.; Nesvizhskii, A. I., Fast Deisotoping Algorithm and Its Implementation in the MSFragger Search Engine. *J Proteome Res* **2021,** *20* (1), 498-505.

55. Penders, J.; Nagelkerke, A.; Cunnane, E. M.; Pedersen, S. V.; Pence, I. J.; Coombes, R. C.; Stevens, M. M., Single Particle Automated Raman Trapping Analysis of Breast Cancer Cell-Derived Extracellular Vesicles as Cancer Biomarkers. *ACS Nano* **2021**.

56. Kruglik, S. G.; Royo, F.; Guigner, J. M.; Palomo, L.; Seksek, O.; Turpin, P. Y.; Tatischeff, I.; Falcon-Perez, J. M., Raman tweezers microspectroscopy of circa 100 nm extracellular vesicles. *Nanoscale* **2019,** *11* (4), 1661-1679.

57. Shin, H.; Oh, S.; Hong, S.; Kang, M.; Kang, D.; Ji, Y. G.; Choi, B. H.; Kang, K. W.; Jeong, H.; Park, Y.; Hong, S.; Kim, H. K.; Choi, Y., Early-Stage Lung Cancer Diagnosis by Deep Learning-Based Spectroscopic Analysis of Circulating Exosomes. *ACS Nano* **2020,** *14* (5), 5435-5444.

58. Dong, S.; Wang, Y.; Liu, Z.; Zhang, W.; Yi, K.; Zhang, X.; Zhang, X.; Jiang, C.; Yang, S.; Wang, F.; Xiao, X., Beehive-Inspired Macroporous SERS Probe for Cancer Detection through Capturing and Analyzing Exosomes in Plasma. *ACS Appl Mater Interfaces* **2020,** *12* (4), 5136-5146.

59. Rojalin, T.; Koster, H. J.; Liu, J.; Mizenko, R. R.; Tran, D.; Wachsmann-Hogiu, S.; Carney, R. P., Hybrid Nanoplasmonic Porous Biomaterial Scaffold for Liquid Biopsy Diagnostics Using Extracellular Vesicles. *ACS Sens* **2020,** *5* (9), 2820-2833.

60. Banaei, N.; Moshfegh, J.; Kim, B., Surface enhanced Raman spectroscopy - based immunoassay detection of tumor - derived extracellular vesicles to differentiate pancreatic cancers from chronic pancreatitis. *Journal of Raman Spectroscopy* **2021,** *52* (11), 1810-1819.

Chapter 3. SIM on sEVs for Gastric Cancer Detection

3.1. Introduction

Gastric cancer (GC) is the fifth-most popular type of malignant tumor and the fourth-most deadly worldwide with over one million new cases, leading to 768,793 deaths in 2020.[1] Majority of the gastric cancers (~90%) are adenocarcinoma which happen at the mucosa layer of the stomach. Although the occurrence and mortality of GC have been on the decline, the five-year survival rate continues to be low.[2] However, for patients diagnosed of GC at early stages, five-year survival rates of 95% or higher have been observed[3, 4] demonstrating the overwhelming importance of early diagnosis and population screening. Early GC diagnosis requires reliable, cheap, and easy-to-operate screening methods that are yet to be available.[5] Currently, screening methods such as barium-meal gastric photofluorography and upper endoscopy followed by biopsy are able to support good population GC screening programs in some countries like South Korea. However such techniques can be costly and time-consuming.[6, 7] These procedures have also been shown to be associated with false negative rates, risks related with the rather invasive procedures, and other side effects.[8-10] Recently extracellular vesicles, especially sEVs, have become potential sources of biomarkers for cancer detections with easy access and minimal invasiveness.[11-14]

sEVs play crucial roles in cell-to-cell communications via the encapsulated cargos which also reflect their parental cells.[15-18] The stable existences in bodily fluids grant the potentials for them being biomarkers for cancer liquid biopsy.[19-21] By detecting those sEVs, opportunities exist for non-invasive cancer detection.[22, 23] For cancer patients, there co-exist sEVs from both normal and cancerous cells, each with their own characteristic biochemical cargo contents, forming different subtypes.[24, 25] One key challenge facing the vesicular liquid biopsy is a technique capable

of examining individual sEVs thereby distinguish the sub-populations that belongs uniquely to abnormal cells [26-28]

This chapter describes experimental studies on exploring the clinical applicability of SIM for minimally invasive GC detections via detecting sEVs. SERS gold nanopyramid platform was applied to spectrally measure vesicles. Instead of using disease-specific SERS tags or focusing on particular spectral features comparisons, the biochemical compositions of the collective Raman active bonds were extracted from individual vesicles directly in the form of SERS spectra. The study aims to examine if these SERS spectra can be used as the sEV "fingerprints" for GC detections. For machine learning, an algorithm was customized to help correct the mislabeling issue in the clinical samples by sub-fractioning the vesicles measured. The results from SIM analysis of cell-lines derived sEVs illustrated the existences of vesicles that were common to both the GC and the normal stomach tissues in addition to the characteristic ones. For clinical samples, vesicles were isolated from the tissue, blood, and saliva of donors from the GC patient and the non-GC control group by an acoustofluidic platform (AFS) developed previously by our collaborators at the Duke University laboratory.[29] The unique capabilities of AFS including high efficiency and low processing time enable a better vesicle recovery rate and quality.[29] The accuracies in identify GC versus control were 90%, 85%, and 72% in tissue, blood, and saliva respectively. "Leave-a-pair-of-samples out" analysis was performed to mimic the potential clinic applications of the platform. The result showed ROCs with AUCs being 0.96, 0.91, and 0.65 respectively in tissue, blood, and saliva cases respectively. Additionally, 9 patients' unique sEV types were found existing across all three sample environments, opening a possibility for tracing the biogenesis of the GC patient-specific sEVs. The methodology involved in this study is amenable for non-invasive detection of diseases other than GC with further validation.

3.2. Experimental Procedures

3.2.1. Cell cultures

Three cell-lines, AGS (ATCC, CRL-1739), NCI-N87 (ATCC, CRL-5822), Hs 738.St/Int (ATCC, CRL-7869), were used in this study. In this study, all the fetal bovine serum (FBS; ATCC, 30-2020) were pre-treated to remove the sEVs. To prepare the complete culture medium for these three different cell lines, F-12K Medium (ATCC, 30-2004), RPMI-1640 Medium (ATCC, 30-2001), and Dulbecco's Modified Eagle's Medium (ATCC, 30-2002) were supplemented with 10% of pre-treated FBS and 1% of penicillin-streptomycin (P/S; ATCC, 30-2300) for AGS, NCI-N87, and Hs 738.St/Int, respectively.

All the cell lines were thawed in a 37 ℃ water bath for 1 min and then transferred into a centrifuge tube containing 9.0 mL complete culture medium and spun at 125 x g in a centrifuge (Rotor F-35-6-30, 5430, Eppendorf, Germany) for 5 minutes to remove the dead cells. The cell pellets were resuspended with the complete culture medium and dispensed into a 75 cm^2 culture flask (T-75 flask; MSPP-90076, VWR, USA). Cells were cultured at 37 ℃ in a humidified incubator (MCO-19AIC, SANYO, USA) with 5% CO_2 and 95% air. Media were collected and changed every 2 - 3 days.

The cells were harvested when the flask reaches over 90% confluency and stored in liquid nitrogen vapor. The medium was aspirated from the cell culture flask and collected for the study. The flasks were rinsed with 10 mL Dulbecco's Phosphate Buffered Saline (D-PBS; ATCC, 30-2200), 2 -3 mL of Trypsin-EDTA solution (ATCC, 30-2101) were added to remove all traces of serum that contains trypsin inhibitor for 5 minutes. Then 6 - 8 mL of complete growth mediums were used to resuspend the cells, the cell suspension was transferred to a centrifuge tube and spun at 125 x g for 5 minutes. The cell pallets were collected and added into a cryovial (5000-1020,

Thermo Fisher Scientific, USA) contains complete culture medium with 5% (v/v) Dimethylsulfoxide (DMSO; ATCC, 4-X) after aspirating out the supernatant medium. The cryovial was placed into a CoolCell (43200, Corning, USA) to freeze down to -80°C. After 2-3 days, store the cells in a liqiud nitrogen tank at vapor phase.

3.2.2. Tissue, plasma, and saliva samples

For each donor, tissue, plasma, and saliva samples were collected respectively at the Samsung Medical Center in Korea. Tissue samples were collected from surgical resection during operation (GC patients) or tissue biopsy during endoscopic examination (non-GC control individuals). The collected tissues were stored at -80°C until use Plasma samples were collected using EDTA (Ethylenediaminetetraacetic acid) tubes following the conventional clinical practice and stored at -80°C until use.[30, 31] Unstimulated whole saliva collected was performed as described previously.[32] From each subject, 5 mL whole saliva was collected and centrifuged at 2,600×g for 15 min at 4°C. SUPERase-In RNase inhibitor was added to the supernatant at 20 U/mL to stabilize salivary exRNA. The cell-free saliva supernatants were stored at -80°C until use.

3.2.3. SERS substrate fabrication

The fabrication process of the SERS substrate used in this study follows the same pattern as the one described in section 2.2.5 of this thesis.

3.2.4. Ultracentrifugation

Cell culture supernatants were first centrifuged at 300 g at 4 °C for 10 min and then at 2,000 g at 4°C for 15 min to remove contaminating cells and apoptotic bodies respectively. The supernatants were then further centrifuged at 12,000 g at 4 °C for 45 min to remove cell debris. The clear supernatant was then filtered using 0.22-μm pore filters, followed by ultracentrifugation

at 110,000 g at 4 °C for 70 mins. The resulting pellets were re-suspended in pre-chilled PBS and again ultra-centrifuged at 110,000 g and 4 °C for 70 min. The final pellet of sEVs was re-suspended, in 50–100 μL.

3.2.5. Acoustofluidic platform (AFS) sEV isolations

The detailed protocol of sEV isolations through AFS in this study was developed by our collaborators, Prof. Tony Huang's group at Duke University, and had been described in the previous publications.[29, 34]

3.2.6. Nanoparticle tracking analysis (NTA)

The sample and PBS (Thermofisher, USA) sheath flow were independently controlled by a syringe pump (neMESYS, CETONI GmbH, Germany). Powered by a variable DC power supply (TP1505D, Tekpower, USA), a Peltier cooling system (TEC1-12730, Hebei IT, China) was used for avoiding excessive heat generation from the SAW device during sEV separation. An upright microscope (BX51WI, Olympus, Japan) combined with a CCD camera (CoolSNAP HQ2, Photometrics, USA) combined with a CCD camera (CoolSNAP HQ2, Photometrics, USA) were used for monitoring separation process. The sEV separation SAW device was powered by a function generator (E4422B, Agilent, USA) and an amplifier (100A250A, Amplifier Research, USA). After separation, collected samples were analyzed by a Nanoparticle Tracking Analysis (NTA, Nanosight LM10, Malvern, England) system for getting the size distribution data.

3.2.7. Transmission electron microscopy (TEM)

TEM validation follows our previous procedure.[34] 4% paraformaldehyde (Sigma-Aldrich, St. Louis, MO) was used for fixing the isolated samples. 100 μL droplet of fixed sample was covered by a 300-meshcopper grid supportfilm (Electron Microscopy Sciences, Hatfield, PA) for absorption. Grids were washed by distilled water and then stained by uranyleacetate solution

(Electron Microscopy Sciences). Grids were washed by distilled water again and dried in room temperature. An electron microscope (FEI, Hillsboro, OR) was used for observation.

3.2.8. Scanning electron microscopy (SEM)

SEM was used to characterize the SERS substrate. Imaging was performed using Nova 230 with accelerating voltage 10 kV. The detector used was in "through the lens" mode to detect secondary electrons, and the images were magnified at ×50,000 to ×55,000.

3.2.9. Raman spectroscopy

Before the Raman test, 5 μL of each sEV sample solution was deposited on the SERS substrate and dried. Raman measurements were performed using the Reinshaw inVia Raman spectrometer at room temperature. The laser excitation wavelength was 785 nm. The power used was 5 mW. Before measuring sEVs, the system was calibrated using the 520 cm^{-1} peak of silicon. Mappings performed at step width 5 μm to collect vesicle spectra across the sample droplet. The exposure time was 0.2 s to avoid sample overheating. For generating SERS intensity maps, a fine mapping was performed at step width 0.1 μm to collect characteristic spectra from the sEV sample after a vesicle was spotted. Again, the exposure time was 0.2 s to avoid sample overheating.

3.2.10. Machine learning analysis

Approximately 50 to 70 different sEVs are obtained for each sample to produce spectra which have 1023 Raman shifts in the range from 553 to 1581 cm^{-1} (biological information rich region). Preprocessing steps are applied to alleviate the spectral signature fluctuations caused by sample variations, SERS platform heterogeneity and instrument fluctuation. Particularly, fluorescence background subtraction and noise reduction are performed by batch processing based on asymmetric least square fitting and Savitzky-Golay filtering, followed by min-max normalization that proportionally compresses the original intensity range to [0, 1]. No initial

feature selections or dimension reduction was performed prior to classification. To reveal the spectral differences among the three cell line groups, linear discriminant analysis (LDA) is used to reduce the dimensionality for visualization. For machine learning model development, predictive model establishment by supervised learning or classification is the core for the proposed technology. It requires appropriate complexity of the classifier to prevent both underfitting and overfitting for the purpose of generalizing the characteristic signature effectively. Conventional but powerful algorithm support vector machine (SVM) for classification tasks was used. Unsupervised learning or clustering analysis is performed by Hierarchical clustering analysis with customized distance metrics, it investigates the intrinsic similarities among the analytes SERS signature and serves as an auxiliary to classification. Randomly 80% of spectra from each of the groups (patient vs. control) were selected for the model training and the rest 20% were left out for cross-validations. 20 rounds of cross-validations were performed, with each round running independent to avoid overfitting. The prediction accuracy is the ratio between the number of correct predictions and the total number of predictions, following the equation eq.1. All the analyses are realized with Python.

$$\text{Accuracy} = \frac{\text{True positive} + \text{True negative}}{\text{True positive} + \text{True negative} + \text{False positive} + \text{False negative}} \text{(eq.3.1)}$$

3.3. Results and Discussion

3.3.1. Experimental flow of SIM measuring sEVs for minimally invasive GC detections

The schematic of the experimental process flow here is shown in Figure 3.1. Briefly, the experimental process can be described as three categories, sEV isolation followed by SERS spectra collections and data analysis. To begin with, vesicles were extracted from three types of samples

(tissue, blood, and saliva) from human donors by AFS. After isolations, the sEV analytes were dispersed onto SERS substrate for measurements. Each sample droplet was 5 µL. After collecting SERS fingerprints, the machine learning algorithm (SVM) was employed to establish the distinguishability. Randomly selected 80% of the spectral data from each of the groups were used as the training set for building-up the machine learning model and the rest of 20% of the spectra in each group were left-out for testing the model's predictability on cancer/control. The prediction results obtained from the testing phase were then compared with the true sample identification to calculate the accuracy. "Leave-a-pair-of-samples out" validation was performed to test the clinical applicability. Additionally, the SERS fingerprints from the patients' unique sEVs were extracted from tissue, blood, and saliva respectively followed by a cross-comparison for studying the possibility of tracing the vesicles through their SERS signatures.

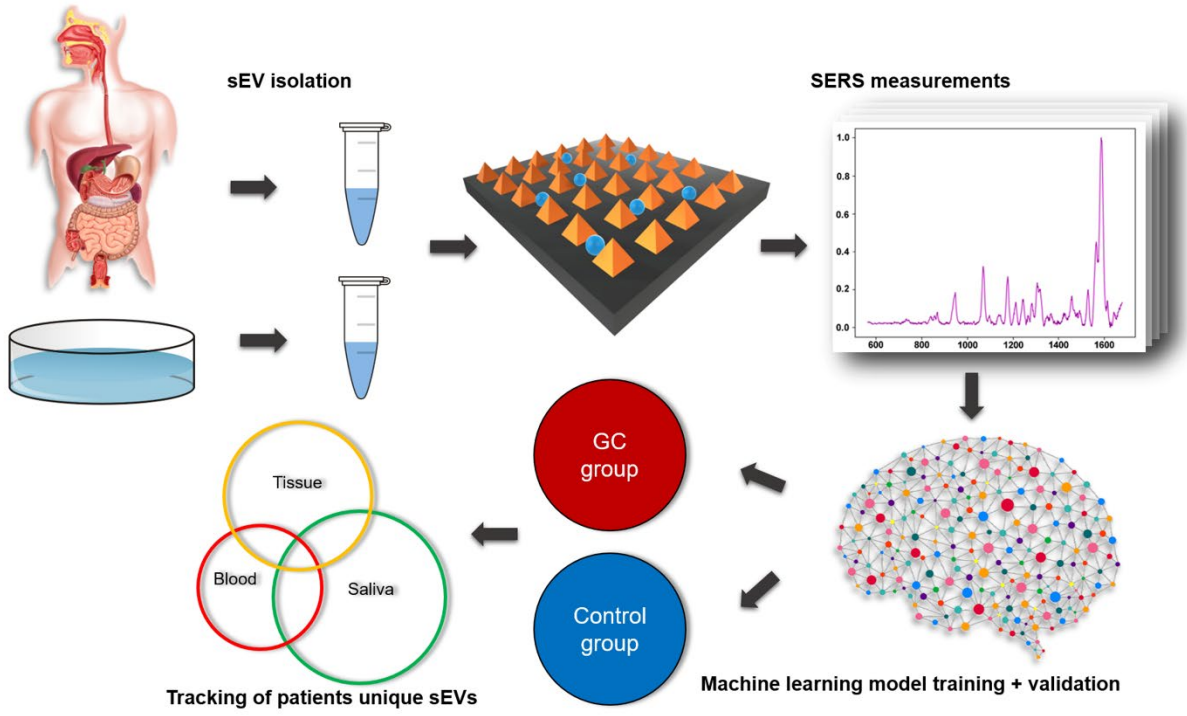

Figure 3.16. Schematic of SERS and machine learning for analyzing sEVs isolated from human samples

3.3.2. Sample characterizations

Figure 3.2a shows the SEM image of the SERS gold nanopyramid substrate. The isolated sEV samples were characterized using NTA and TEM for verifications. Figure 3.2b shows the interaction between the sEV samples and the SERS substrate after the sample droplet had been introduced, suggesting that intact vesicles lying in between the nano-pyramids. Figure 3.2c exhibits the TEM image of the vesicle samples where the sizes fall into the category of sEVs and Figure 3.2d shows the NTA result of the size distribution of the vesicle samples. The substrate itself was Raman inactive, and SERS spectra were collected from the areas with Raman spectral responses on the gold nano-pyramid array. Since the sEVs ⌀ 30–150 nm, it was impossible to observe these vesicles directly under an optical microscope equipped on a conventional Raman spectroscopy system. To verify, SERS mappings with a super-fine grid on areas were performed at areas with Raman signals. The step width for each point in the mapping was set to be 100 nm, the lower limit of the instrument. Since the Raman laser is a Gaussian Beam, a heat mapping based on the particular peak intensity fluctuations across a small area could be generated using such fine grid mapping. Three Raman intensity heat maps were plotted with respect to the Raman peak positions representing phospholipids (1270 cm^{-1}), nucleic acid (1341 cm^{-1}), and protein (1123 cm^{-1}). As shown in Figure 3.2e, the three heat maps were spherical, the shape of a vesicle, with the comparable sizes, suggesting the co-existing of the three substances essential to an sEV. The three SERS intensity maps in Figure 3.2e show the raw shape of a single sEV. But their sizes are larger than the actual vesicle. This is a result of convolution between a tightly focused laser beam of ~ 1 um diameter and a much smaller (~100 nm) hotspot (the source of the Raman signal).[33, 35] The resulting diameter of the heatmap is limited by the larger of the two, in this case that of the excitation laser beam for Raman excitation of 1 um. Nonetheless, the three SERS intensity maps

show the shape of an sEV with comparable sizes consistently, suggesting the co-existing of the three substances, improving the rigor of the result. SERS mapping shows that the average spacing between sEVs is larger than 5 μm, much larger than the size of sEVs and the laser spot size. It needs to be pointed out that because of the instrumentational limitation, specifically the Raman laser spot size being larger than the vesicular diameters, it is challenging to precisely mapping out the morphology of an individual sEV. However, SERS signal intensity is quadratically dependent on the local electromagnetic field intensity, making the signal from one single plasmonic hotpot dominant.[36, 37] When utilizing our SERS platform, the hotspot size matches the size range of sEVs (~100nm).[33, 35, 38] Such feature allows the SERS signal from single sEVs to be collected one at a time. It should be noted that while the majority of the SERS spectra were derived from single sEVs, there is a non-zero probability that occasional ones could be derived from more than one vesicle.

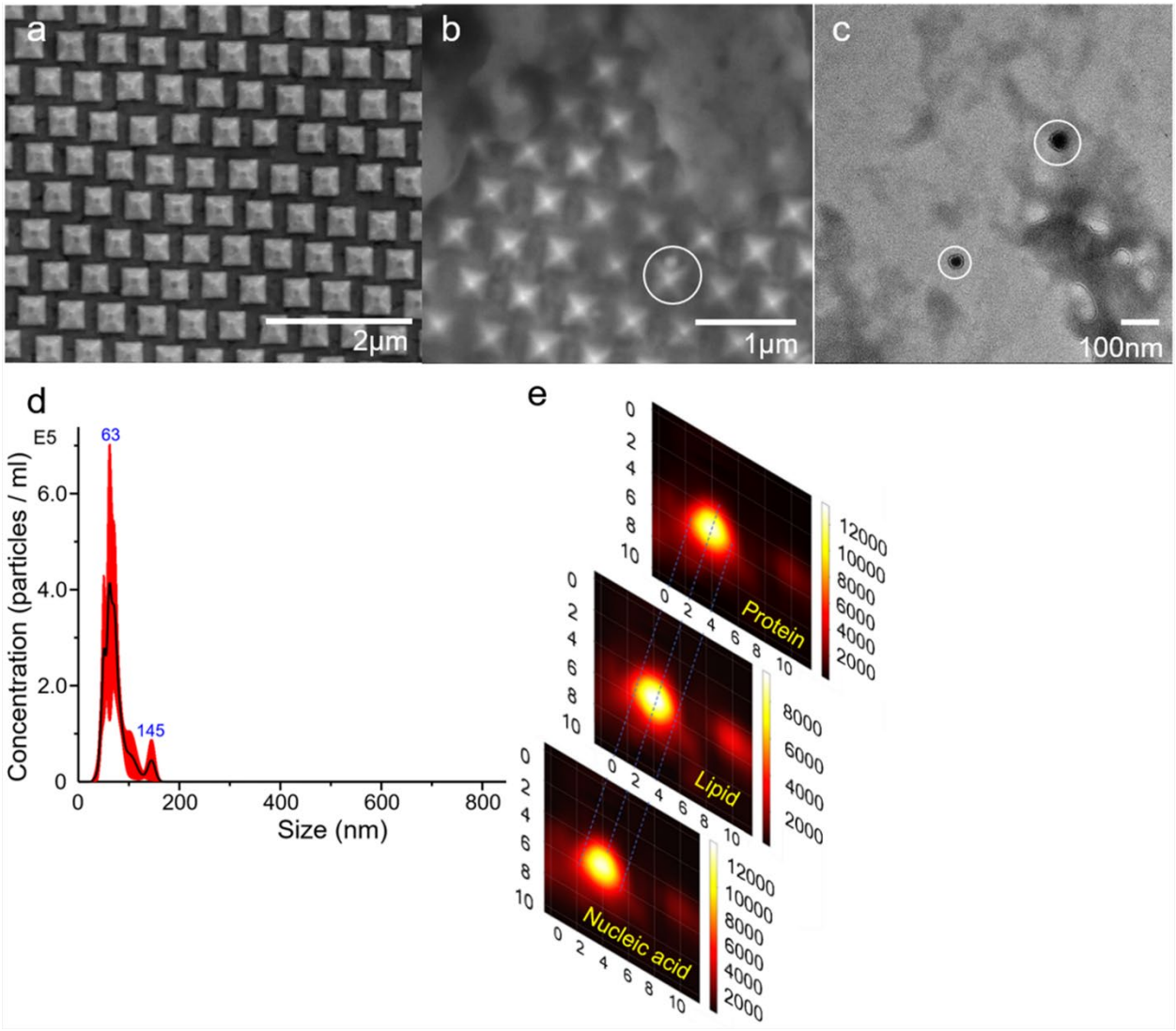

Figure 3.17. a) SEM image of the SERS gold nanopyramids platform; b) SEM image of the SERS substrate after sample solution introduction. c) TEM image of isolated sEVs suspended in PBS; d) NTA result of the isolated vesicles; e) SERS intensity maps generated with respect to nucleic acid, lipid, and protein respectively from the same data spot

3.3.3. SIM analyzing sEVs from cell-lines

Before working on the patient samples, cell-lines derived sEVs by ultracentrifugation were used to establish the capability of SIM to detect and analyze single sEVs. More importantly,

analyzing vesicles from cell-lines provides the information of single sEVs heterogeneities given the purest forms of parental cells. Such knowledge laid the foundation and set the expectations for the further studies with clinical samples. Two gastric cancer cell-lines (CRL-1739 & CRL-5822) and one normal stomach tissue cell-line (CRL-7869) were involved. sEVs from each of the cell-lines were isolated from the culture medium using ultracentrifugation before dropped onto the gold nano-pyramid substrate. Within each sample, SERS measured vesicles one at a time, generating one spectrum per vesicle. For CRL-1739, CRL-5822, and CRL-7869 derived sEVs, 115, 106, and 86 vesicles were measured by SERS respectively. After spectra collection, linear discriminant analysis (LDA) was applied firstly to study the distinguishability of SIM on the three groups. Figure 3.3a showed that the SERS spectra from three group could be distinguished, with the variety of the sEVs within each of the sample groups revealed by SIM. Inherently, LDA is a technique used to separate different groups of data. In order to directly and objectively compare the SERS fingerprints of individual sEVs for determining the common and characteristic vesicles among the three groups, a machine learning-based (SVM) SERS spectral feature comparison mechanism was introduced. Figure 3.3b exhibited the results of comparing the SERS fingerprints of individual vesicles across the three cell-line groups. Internal spectral variations existed within each of the sEV types, but the spectral differences across different vesicle types outweighed the internal spectral variations as measured by the Euclidean distance between nearest neighbors. The results of analyzing three cell-lines illustrated the existences of sEV populations common to both gastric cancer tissue and the normal stomach tissue released groups even in the cell-line forms. Such observation further indicated the existences of sEV populations common to both GC/non-GC groups in the clinical bodily fluids. When analyzing single vesicles, it is necessary to sort out the common populations before an accurate detection could possibly be made.

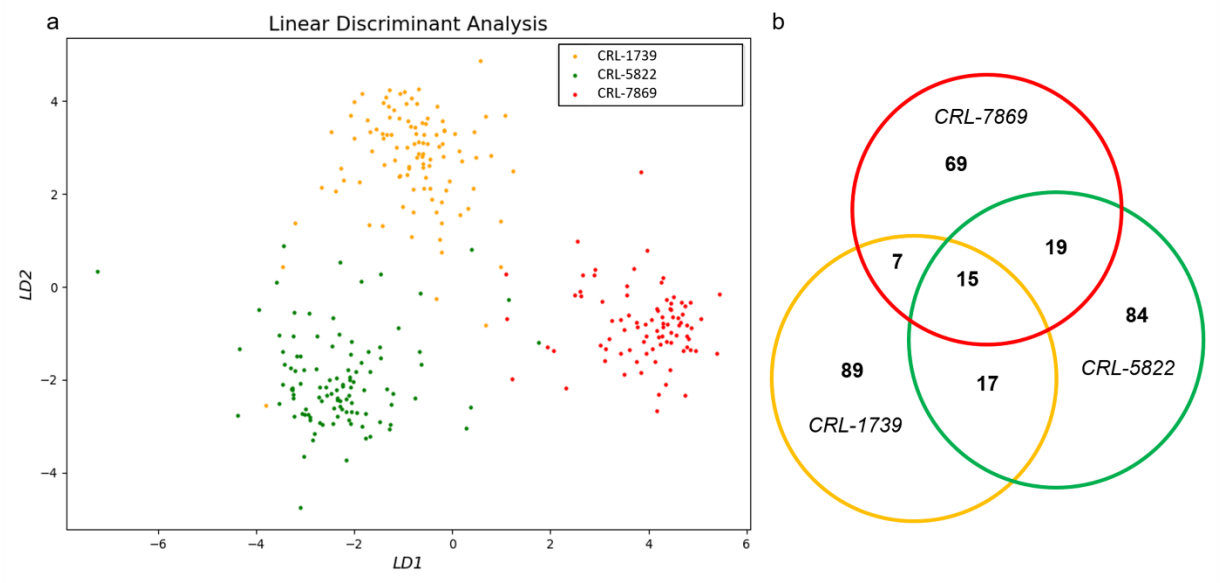

Figure 3.18. a) LDA result distinguishing the SERS spectra of the sEVs derived from cell-lines as three groups; b) Statistical results of SERS signatures comparisons among individual vesicles

3.3.4. SIM Analyzing Clinical Samples for GC Detection

Next, SIM platform was applied to analyzing the tissue, blood, and saliva derived sEVs from gastric cancer patients and non-gastric cancer controls (n = 15 for each of the groups). To address the low vesicle concentration issue, AFS that developed previously at the Duke University laboratory was employed for better vesicle recoveries during the isolation from the human samples to avoid the well-known vesicle loss during the ultracentrifugation isolation.[34] On average, 60 different data spots were measured by SERS for each of the sample droplets. Compared to cell-lines as the sEV extracting sources, two key factors were expected that could raise the complexities when studying vesicles from the human bodies. First, samples from human bodies, especially the bodily fluids, have lower vesicle concentrations comparing to cell culture media. Second, the populations of the vesicles that are common to both patient and the control group are expected to be more in the bodily fluids because the normal cells inside a patient's body secret the same/similar

types of vesicles as the ones inside a non-cancer control person, leading to the mislabeling problem in machine learning and further to damaging the detection accuracy. The LDA results of the clinical samples are shown in Figure 3.4. The y-axes are LD1 scores and the Gaussian like curves record the spectra distributions according to the LD1 scores. Spectral data from sEV samples in saliva and blood were observed to show overlapping between the GC and the non-GC groups from LD1 equals -2 to 2. On the other hand, such overlapping was hardly observed in the tissue sEV samples. Such results indicated that some of the vesicles in saliva and blood shared similar/common SERS spectral features between GC and non-GC, further inferring the existences of normal sEVs inside patients' bloods and saliva. Such observations verified our hypothesis, given that the biochemical compositions of the vesicles could reflect their parental cells. In addition, the LDA results suggested that the mislabeling issue was inevitable, thus requiring relabeling before the SVM classification model training, especially in blood and saliva samples.

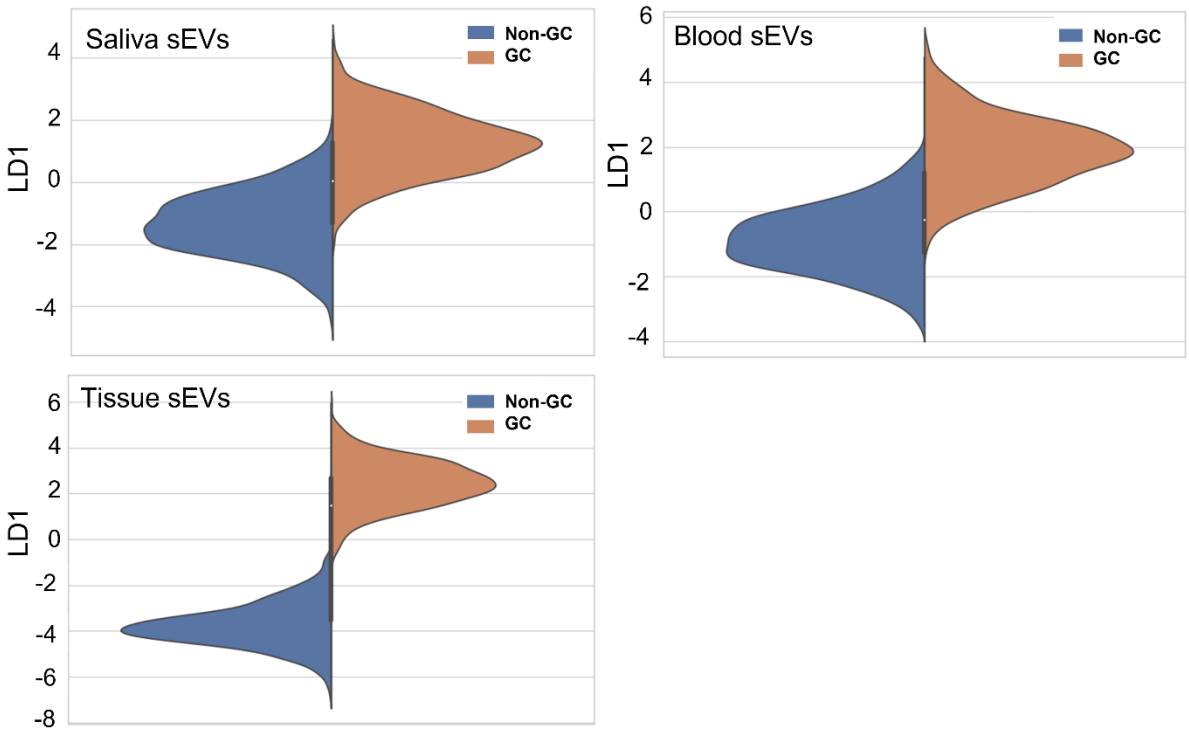

Figure 3.19. LDA results comparing the SERS spectra of sEVs in tissue, blood, and saliva respectively

To address the mislabeling problem in machine learning, a relabeling process was involved before training through sub-fractioning of individual sEVs from patients based on their biochemical compositions reflected on the SERS fingerprints. As shown in Figure 3.5, first the SERS fingerprints of individual sEVs between the patient and the control group were compared, extracting the sEV types that uniquely existed in the patient group (PP). Such process was done by cross-comparing the spectral features of the individual spectra between the GC and the non-GC group. The spectra in the GC group that shared common spectral features with the ones in the non-GC group were regarded as the normal sEVs inside the patients' bodies, given the vesicles could reflect their parental cells. They were further separated from those spectra found only in the GC group (PP). Then, only the SERS signatures from the PP group were labeled as "gastric cancer"

and others were labeled as “control” for the machine learning model training. Such process was referred as “relabeling” and helped correcting the mislabeling issue mentioned previously.

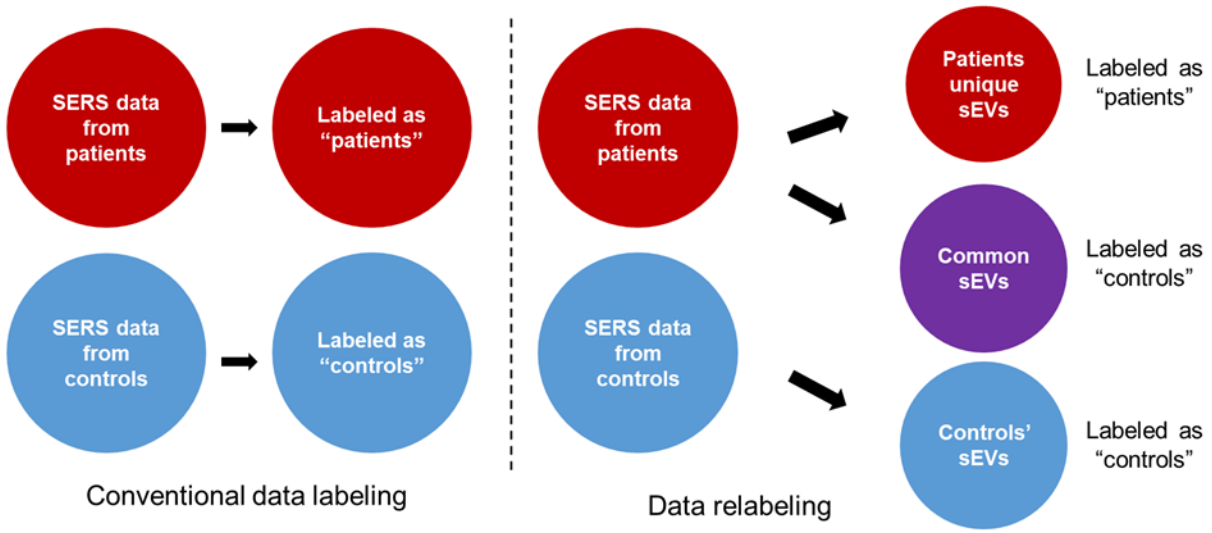

Figure 3.20. Schematics of conventional data labeling in machine learning and the data relabeling process used in this study

Cross-validation of randomly selected 20% of the SERS spectra of sEVs from each of the two groups was done to study the detection accuracy. The selected 20% was intentionally left-out during the training for avoiding the information leak during the training, improving the rigor. For analyzing the SERS fingerprints collected from sEVs of tissue, blood, and saliva, both non-relabeling and relabeling methods were involved and the detection accuracies are exhibited in Table 3.1. Each of the results was an average of 20 rounds of cross-validations ($\sigma^2 \leq 0.000541$). In general, tissue derived sEVs held the highest accuracy of distinguishing followed by the blood derived and saliva derived vesicles. With relabeling, the detection accuracies were improved for sEVs collected from blood and saliva but no significant changed regarding the samples from tissue. For the machine learning model training and cross-validation, the training data set and the validation data set selected have no overlap. Such procedure avoids the pitfall of information leak,

i.e., the machine learning program has been given some clue about the identity of the samples used for cross-validation leading to a falsely high accuracy of detection. The relabeling help correcting the mislabeling issue due to the inevitable existence of normal vesicles in the samples of the cancer patients. To further explore the potential of our analysis in the clinical applications, "leave-a-pair-of-samples out" validations using the relabeled spectra were performed. In such validation, SERS spectra from a randomly selected GC patient and a randomly selected control individual were intentionally excluded from the training set and used as the test set. The random pairing and "leaving-out" continued but would not include samples that had been selected before. One round ended when every sample had been "left-out" once. To test statistical fluctuations, ten rounds for each of the tissue, blood, and saliva samples were performed. Figure 3.6 shows the resulted ROCs. The averaged AUCs were reportedly to be 0.96, 0.91, and 0.65 in tissue, blood, and saliva respectively. It needs to be pointed out that "leave-a-pair-of-samples out" cannot completely substitute the actual clinical blind test. But, to our understanding, such validation offers the chance to test the platform's clinical applicability by providing a scenario closer to the blind test than normal cross-validation. The result suggests that apart from tissue samples, blood holds a greater potential than saliva as a source for sEV based gastric cancer liquid biopsy though involving the minimal invasiveness of puncture.

Table 3.1. SERS spectral distinguishability considering sEVs from different sources

	Tissue sEVs		**Blood sEVs**		**Saliva sEVs**	
	non-relabeld	relabeled	non-relabeld	relabeld	non-relabeled	relabeld
Accuracy	90%	89%	72%	85%	58%	72%

The GC/non-GC distinction results show that there existed improvements in detection accuracies with relabeling in saliva and blood samples but no significant change in the tissue samples. This observation could be explained as the reflection of the different relative population of the GC-specific sEVs. The sEV populations derived from the patients' tissues contain the highest concentration of the patient-characteristic vesicles. Such concentration drops when those sEVs circulate in the bodily fluids where they are joined by vesicles released by other (non-GC) organs. It should be pointed out that such single sEVs detection mechanism has its fundamental limitation on the throughput which is inversely proportional to the required number of sEVs to be examined per patient sample in order for making accurate diagnosis. Nonetheless, this study reveals the feasibility of non-invasive gastric cancer detection/screening through analyzing the composition information of the collective Raman active bonds inside single sEVs isolated from blood and saliva using SIM. SERS measurements provided molecular fingerprints from single vesicles and machine learning algorithm offered the ability to distinguish among the various sub-fractions of the sEVs with objectivity and rigor. Despite the complexity of the microenvironments inside human bodies, the SIM platform has been shown to be capable of reaching the detection accuracy of 90%, 85%, and 72.0% with the reportedly AUCs of 0.96, 0.91, and 0.65 in the "leave-a-pair-of-samples out" validations in tissue, blood, and saliva respectively. The difference in detection accuracy makes intuitive sense considering the difference in the relative concentration of GC-specific sEVs in these three sources. Needless to say, blood and saliva are the more non-invasive sources of sEVs thus with higher promise for clinical practice. Effort is in progress to further optimize the machine learning algorithm with the goal of improving the detection accuracy.

With the development of the field, emerging works related to the innovations and the applications of SERS substrates for cancer detections via extracellular vesicles have been

published. Shin et al. (2020) reported a combination of SERS and deep learning for early stage lung cancer detection via sEVs derived in blood with AUC > 0.9.[39] To achieve such detection, spectral features from the test set were classified based on the degrees of similarities to the training set.[39] In our present study, comparable AUC in "leave-a-pair-of-samples out" validation with the blood sample was observed, but such score dropped with the saliva samples. From the clinical and practical perspectives, the observation of detection accuracy discrepancies in blood and saliva could serve as an indication for biofluid selections in the liquid biopsy based gastric cancer detections. Dong et al. (2020) focused on one specific vesicular contents and reported that the variations of the SERS signal intensity of protein phosphorylation inside sEV between control and patients could serve as the indicator for detecting prostate, liver, lung, and colon cancers.[40] Carmicheal et al. (2019) utilized SERS gold nanoparticles and machine learning for pancreatic cancer detection.[41] sEVs from serum were measured by SERS from 20 (n = 10 for from cancer/control) individuals and machine learning algorithm prediction results indicated the diagnostic potential and the bio-variability dragging the effectiveness due to the diverse origins of the serum vesicles.[41] Rojalin et al. (2020) reported a porous scaffold SERS platform that could prevent vesicles from drying during the SERS measurements with PCA indicating clear separations of the SERS spectra of extracellular vesicles (sEV and larger ones) collected from two control individuals, two ovarian cancer patients, and two endometrial cancer patients.[42] In addition, the dragging of detection accuracy by trypsinization of clinical vesicles was illustrated.[42] Apart from analyzing the spectral features directly, modified SERS substrate for vesicle immobilization allows quantitative analysis of the captured sEV populations between the cancer and the control group through the comparison of intensities of the SERS indicator as demonstrated by Banaei et al. (2021).[43] In this presented study, the single-vesicle based analysis enables identification and

discrimination among different subtypes of sEVs within the same sample group. The relabeling process for enhancing the detection accuracy cannot be done without the signatures from each of the individual vesicles. In addition, the identification of single vesicle offers the feasibility of tracing the trafficking pathways of the patients' sEVs from the cancer tissues to the bodily fluids based on their SERS fingerprints. It opens a new possibility for the further understanding of the sEV biogenesis, specifically the pathway of sEV trafficking from the point of secretion to entering body fluids.

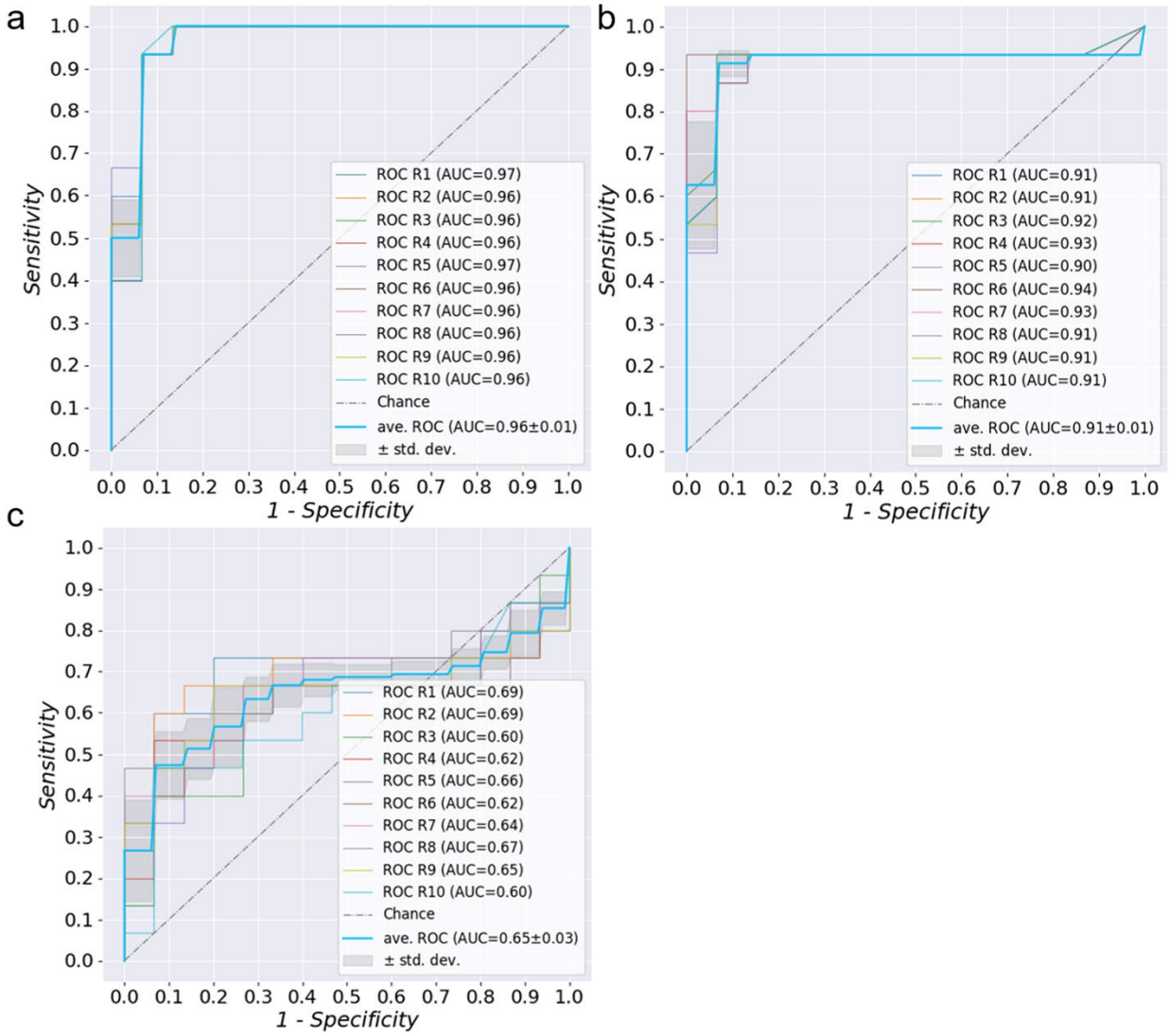

Figure 3.21. ROC curves of the "leave-a-pair-of-samples out" validations for tissue sEVs a), blood sEVs b), and saliva sEVs c)

3.3.5. Tracking sEVs uniquely belonged to the patient group

After illustrating the gastric cancer detectability of SIM through analyzing the SERS signatures of single sEVs, vesicles that uniquely belonged to the patients (PP) in tissue, blood, and saliva attracted interests for further analysis. The goal was to study the feasibility of using SIM to track single sEVs inside the body based on their SERS fingerprints, given the published works

shown different vesicle phenotypes existed in different parts of the human body.[44, 45] Shown in Figure 3.7a, after extracted the PP group from tissue, blood, and saliva respectively as mentioned (PT, PB, and PS), the SERS spectral comparison was introduced to identify the common vesicles existed across tissue, blood, and saliva. 9 vesicle types were identified, existing across all three conditions. Figure 3.7b exhibited the superimposed SERS spectra together with the averaged spectrum for each of the sEV type identified. Within the 9 sEV types, the result of studying the source of each individual vesicles was shown in Figure 3.7c, suggesting the population of patient unique sEVs dropping from tissue to blood/saliva which was consistent with the current understanding about cancerous sEVs circulation. The results of tracking vesicles uniquely belonged to the patient group through their SERS signatures opened a new possibility for tracing vesicle circulations from the tissue of origin to the bodily fluids. In addition, such tracking studies shed light on the impact of the sample of origin on the diagnostic accuracy and practicality: tissue being the most invasive source but contains the highest concentration of disease specific sEVs, whereas saliva being the least invasive but contains less of the disease specific vesicles.

The identification of 9 types of the patients' unique sEVs existed across tissue, blood, and saliva illustrates the feasibility of using SIM for tracking patients' vesicles. But is was not observed that one or more of the nine identified vesicle types existed across all patients. The 9 types of sEVs being not shared among all cancer patients could possibly be due to the small sample size, currently limited by SIM throughput. Using the commercially available Raman spectrometer (In Via by Renishaw) designed for research instead of medical laboratories, the routinely achievable throughput is approximately 15-20 vesicles per hour. It should be pointed out that this throughput is by no means intrinsic to SERS analysis of biological samples. Several areas of the spectrometer could be automated to conceivably improve the Raman mapping throughput by orders of

magnitude. With significantly increased sample size per patient/healthy control, the probability of mislabeling (identifying an sEV type to be cancer simply because such a vesicle happened to be absent from the limited size of healthy control samples) will decrease significantly. Nonetheless, our study has shown a promising pathway for the development of an evidence-based procedure factoring in clinical considerations. With necessary clinical trials for validating, the methodology involved in this study is amenable for non-invasive detection of diseases other than GC and further understanding and tracing of the biogenesis pathway of the sEVs.

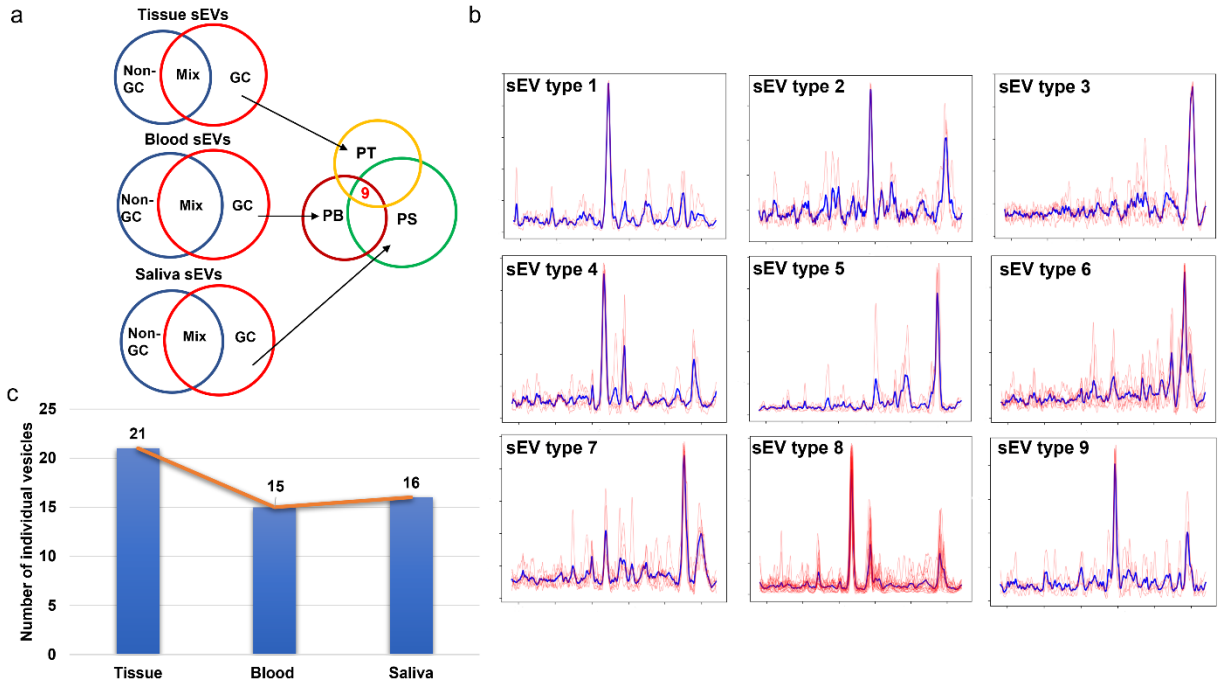

Figure 3.22. a) Schematic of tracking patients' unique sEVs; b) Superimposed SERS spectra (red) and the corresponding average spectrum (blue) of the patients' unique sEVs existed across tissue, blood, and saliva. Horizontal-axis: Raman shift (ranging from 553 to 1581 cm^{-1}). Vertical-axis: Normalized intensity 0-1; c) Distribution of all the individual sEVs of the 9 types presented in b)

3.3.6. Effort in improving SIM throughput through automation of the SERS measurements

Based on the challenges recognized in the presented study, the throughput plays an important role in the applications of SIM in clinical studies to measure greater number cancer-derived vesicles, in the reasonable amount of working time, for enhancing the accuracy and the robustness of pre-clinical and clinical studies. To increase the throughput of SIM, a script to automate the SERS measurements was involved. The detailed developments of the script is led by Tieyi Li of our group. Here, it is worth defining the throughput specially of being the absolute number of vesicles measured within a given amount of time (e.g., a day). With the increasing number of vesicles scanned, the chance of measuring higher number of cancer-derived sEVs also increases. Other methods such increasing the chances of "hitting" cancer-derived vesicles are a separate topic which grants different studies. Discussion onwards carries the effort of increasing the absolute number of vesicles scanned within a given period of time. The automation process is designed to be capable of running the SERS measurements automatically without the need for the attendance from the operator in the middle of the measurements. Two benefits are attributed to the automation. First, comparing the manual setups for SERS measurements, the script performs faster which increases the efficiency of measurements. Second, by running automatically, it opens the possibility of overnight running to improving the machine utilization rate without causing the extra burdens to the operator. A preliminary test to compare the throughputs between taking the measurements manually and applying the automation process was done. The result shown in Figure 3.8 suggests an almost 9× boost in the number of vesicles measured within a day. Further works include optimizing the automation script for enhancing the robustness during the running as the current proof-of-concept version suffers from random systematic errors occasionally during the running.

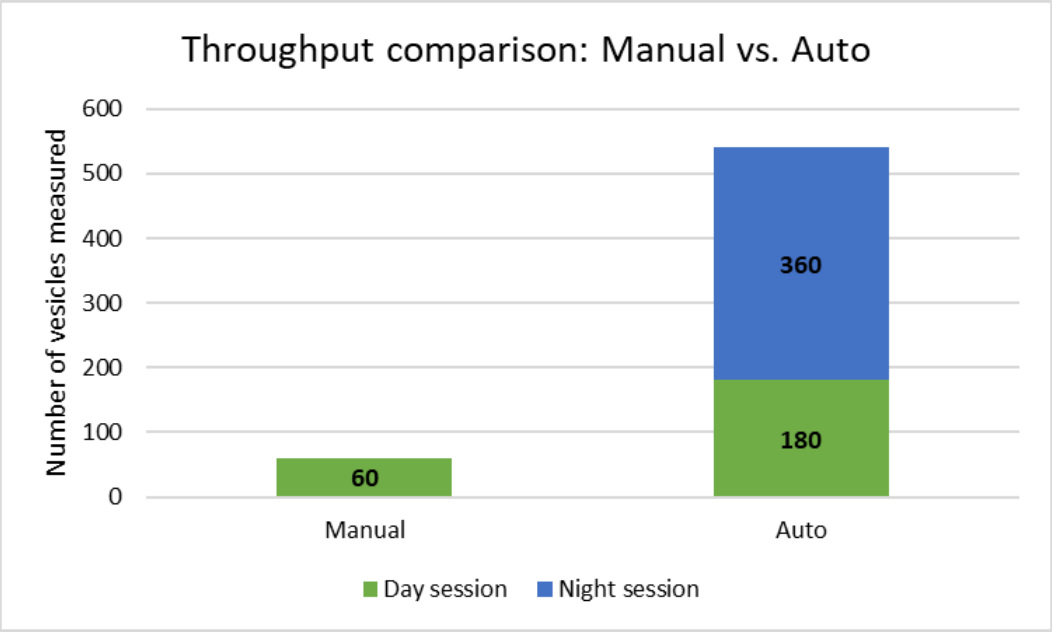

Figure 3.23. Throughput comparison between manual SERS session and automatic SERS session

3.4. Conclusion

In this study, gold nanopyramid platform was applied for SERS measurements of sEVs, exploring the feasibility of non-invasive gastric cancer detection. It demonstrates the feasibility of non-invasive gastric cancer detection/screening through analyzing single sEVs isolated from blood and saliva by SIM, a combination of single vesicle SERS and machine learning. The data obtained from vesicles derived from tissues served as the references for the possible tracking of patient unique vesicles. The distinguishing accuracy of sEVs between gastric cancer patients and non-gastric cancer controls is 90%, 85%, and 72% with the AUC in the "leave-a-pair-of-samples out" validation to be 0.96, 0.91, 0.65 in tissue, blood, and saliva respectively. 9 vesicle types were identified, existing across all three conditions. The methodology developed in this study has the potential to be applied for the detection of other cancers using individual sEVs with further studies for verification.

3.5. Reference

1. Sung, H.; Ferlay, J.; Siegel, R. L.; Laversanne, M.; Soerjomataram, I.; Jemal, A.; Bray, F., Global Cancer Statistics 2020: GLOBOCAN Estimates of Incidence and Mortality Worldwide for 36 Cancers in 185 Countries. *CA Cancer J Clin* **2021,** *71* (3), 209-249.

2. Van Cutsem, E.; Sagaert, X.; Topal, B.; Haustermans, K.; Prenen, H., Gastric cancer. *The Lancet* **2016,** *388* (10060), 2654-2664.

3. Song, Z.; Wu, Y.; Yang, J.; Yang, D.; Fang, X., Progress in the treatment of advanced gastric cancer. *Tumour Biol* **2017,** *39* (7), 1010428317714626.

4. Rawla, P.; Barsouk, A., Epidemiology of gastric cancer: global trends, risk factors and prevention. *Prz Gastroenterol* **2019,** *14* (1), 26-38.

5. Necula, L.; Matei, L.; Dragu, D.; Neagu, A. I.; Mambet, C.; Nedeianu, S.; Bleotu, C.; Diaconu, C. C.; Chivu-Economescu, M., Recent advances in gastric cancer early diagnosis. *World J Gastroenterol* **2019,** *25* (17), 2029-2044.

6. Johnston, F. M.; Beckman, M., Updates on Management of Gastric Cancer. *Curr Oncol Rep* **2019,** *21* (8), 67.

7. Thrift, A. P.; El-Serag, H. B., Burden of Gastric Cancer. *Clin Gastroenterol Hepatol* **2020,** *18* (3), 534-542.

8. Choi, K. S.; Jun, J. K.; Park, E. C.; Park, S.; Jung, K. W.; Han, M. A.; Choi, I. J.; Lee, H. Y., Performance of different gastric cancer screening methods in Korea: a population-based study. *PLoS One* **2012,** *7* (11), e50041.

9. Hamashima, C.; Okamoto, M.; Shabana, M.; Osaki, Y.; Kishimoto, T., Sensitivity of endoscopic screening for gastric cancer by the incidence method. *Int J Cancer* **2013,** *133* (3), 653-9.

10. Miyashiro, I.; Hiratsuka, M.; Sasako, M.; Sano, T.; Mizusawa, J.; Nakamura, K.; Nashimoto, A.; Tsuburaya, A.; Fukushima, N.; Gastric Cancer Surgical Study Group in the Japan Clinical Oncology, G., High false-negative proportion of intraoperative histological examination as a serious problem for clinical application of sentinel node biopsy for early gastric cancer: final results of the Japan Clinical Oncology Group multicenter trial JCOG0302. *Gastric Cancer* **2014,** *17* (2), 316-23.

11. Fu, M.; Gu, J.; Jiang, P.; Qian, H.; Xu, W.; Zhang, X., Exosomes in gastric cancer: roles, mechanisms, and applications. *Mol Cancer* **2019,** *18* (1), 41.

12. Huang, T.; Song, C.; Zheng, L.; Xia, L.; Li, Y.; Zhou, Y., The roles of extracellular vesicles in gastric cancer development, microenvironment, anti-cancer drug resistance, and therapy. *Mol Cancer* **2019,** *18* (1), 62.

13. Kahroba, H.; Hejazi, M. S.; Samadi, N., Exosomes: from carcinogenesis and metastasis to diagnosis and treatment of gastric cancer. *Cell Mol Life Sci* **2019,** *76* (9), 1747-1758.

14. Li, F.; Yoshizawa, J. M.; Kim, K. M.; Kanjanapangka, J.; Grogan, T. R.; Wang, X.; Elashoff, D. E.; Ishikawa, S.; Chia, D.; Liao, W.; Akin, D.; Yan, X.; Lee, M. S.; Choi, R.; Kim, S. M.; Kang, S. Y.; Bae, J. M.; Sohn, T. S.; Lee, J. H.; Choi, M. G.; Min, B. H.; Lee, J. H.; Kim, J. J.; Kim, Y.; Kim, S.; Wong, D. T. W., Discovery and Validation of Salivary Extracellular RNA Biomarkers for Noninvasive Detection of Gastric Cancer. *Clin Chem* **2018,** *64* (10), 1513-1521.

15. Raimondo, F.; Morosi, L.; Chinello, C.; Magni, F.; Pitto, M., Advances in membranous vesicle and exosome proteomics improving biological understanding and biomarker discovery. *Proteomics* **2011,** *11* (4), 709-20.

16. Hannafon, B. N.; Ding, W. Q., Intercellular communication by exosome-derived microRNAs in cancer. *Int J Mol Sci* **2013,** *14* (7), 14240-69.

17. Zhang, J.; Li, S.; Li, L.; Li, M.; Guo, C.; Yao, J.; Mi, S., Exosome and exosomal microRNA: trafficking, sorting, and function. *Genomics Proteomics Bioinformatics* **2015,** *13* (1), 17-24.

18. Hessvik, N. P.; Llorente, A., Current knowledge on exosome biogenesis and release. *Cell Mol Life Sci* **2018,** *75* (2), 193-208.

19. Moller, A.; Lobb, R. J., The evolving translational potential of small extracellular vesicles in cancer. *Nat Rev Cancer* **2020,** *20* (12), 697-709.

20. Srivastava, A.; Amreddy, N.; Pareek, V.; Chinnappan, M.; Ahmed, R.; Mehta, M.; Razaq, M.; Munshi, A.; Ramesh, R., Progress in extracellular vesicle biology and their application in cancer medicine. *Wiley Interdiscip Rev Nanomed Nanobiotechnol* **2020,** *12* (4), e1621.

21. Pan, S.; Zhang, Y.; Natalia, A.; Lim, C. Z. J.; Ho, N. R. Y.; Chowbay, B.; Loh, T. P.; Tam, J. K. C.; Shao, H., Extracellular vesicle drug occupancy enables real-time monitoring of targeted cancer therapy. *Nat Nanotechnol* **2021,** *16* (6), 734-742.

22. Hu, T.; Wolfram, J.; Srivastava, S., Extracellular Vesicles in Cancer Detection: Hopes and Hypes. *Trends Cancer* **2021,** *7* (2), 122-133.

23. Lane, R. E.; Korbie, D.; Hill, M. M.; Trau, M., Extracellular vesicles as circulating cancer biomarkers: opportunities and challenges. *Clin Transl Med* **2018,** *7* (1), 14.

24. Buschmann, D.; Kirchner, B.; Hermann, S.; Marte, M.; Wurmser, C.; Brandes, F.; Kotschote, S.; Bonin, M.; Steinlein, O. K.; Pfaffl, M. W.; Schelling, G.; Reithmair, M., Evaluation of serum extracellular vesicle isolation methods for profiling miRNAs by next-generation sequencing. *J Extracell Vesicles* **2018,** *7* (1), 1481321.

25. Crescitelli, R.; Lasser, C.; Lotvall, J., Isolation and characterization of extracellular vesicle subpopulations from tissues. *Nat Protoc* **2021,** *16* (3), 1548-1580.

26. Willms, E.; Cabanas, C.; Mager, I.; Wood, M. J. A.; Vader, P., Extracellular Vesicle Heterogeneity: Subpopulations, Isolation Techniques, and Diverse Functions in Cancer Progression. *Front Immunol* **2018,** *9*, 738.

27. Thietart, S.; Rautou, P. E., Extracellular vesicles as biomarkers in liver diseases: A clinician's point of view. *J Hepatol* **2020,** *73* (6), 1507-1525.

28. Huang, G.; Lin, G.; Zhu, Y.; Duan, W.; Jin, D., Emerging technologies for profiling extracellular vesicle heterogeneity. *Lab Chip* **2020,** *20* (14), 2423-2437.

29. Wu, M.; Ouyang, Y.; Wang, Z.; Zhang, R.; Huang, P. H.; Chen, C.; Li, H.; Li, P.; Quinn, D.; Dao, M.; Suresh, S.; Sadovsky, Y.; Huang, T. J., Isolation of exosomes from whole blood by integrating acoustics and microfluidics. *Proc Natl Acad Sci U S A* **2017,** *114* (40), 10584-10589.

30. McDonald, T. J.; Perry, M. H.; Peake, R. W.; Pullan, N. J.; O'Connor, J.; Shields, B. M.; Knight, B. A.; Hattersley, A. T., EDTA improves stability of whole blood C-peptide and insulin to over 24 hours at room temperature. *PLoS One* **2012,** *7* (7), e42084.

31. Barra, G. B.; Santa Rita, T. H.; de Almeida Vasques, J.; Chianca, C. F.; Nery, L. F.; Santana Soares Costa, S., EDTA-mediated inhibition of DNases protects circulating cell-free DNA from ex vivo degradation in blood samples. *Clin Biochem* **2015,** *48* (15), 976-81.

32. Navazesh, M., Methods for collecting saliva. *Annals of the New York Academy of Sciences* **1993,** *694(1)*, 72-77.

33. Wang, P.; Liang, O.; Zhang, W.; Schroeder, T.; Xie, Y. H., Ultra-sensitive graphene-plasmonic hybrid platform for label-free detection. *Adv Mater* **2013,** *25* (35), 4918-24.

34. Wang, Z.; Li, F.; Rufo, J.; Chen, C.; Yang, S.; Li, L.; Zhang, J.; Cheng, J.; Kim, Y.; Wu, M.; Abemayor, E.; Tu, M.; Chia, D.; Spruce, R.; Batis, N.; Mehanna, H.; Wong, D. T. W.; Huang, T. J., Acoustofluidic Salivary Exosome Isolation: A Liquid Biopsy Compatible Approach for Human Papillomavirus-Associated Oropharyngeal Cancer Detection. *J Mol Diagn* **2020,** *22* (1), 50-59.

35. Wang, P.; Xia, M.; Liang, O.; Sun, K.; Cipriano, A. F.; Schroeder, T.; Liu, H.; Xie, Y. H., Label-Free SERS Selective Detection of Dopamine and Serotonin Using Graphene-Au Nanopyramid Heterostructure. *Anal Chem* **2015,** *87* (20), 10255-61.

36. Moskovits, M., Surface-enhanced spectroscopy. *Reviews of Modern Physics* **1985,** *57* (3), 783-826.

37. Sharma, B.; Frontiera, R. R.; Henry, A.-I.; Ringe, E.; Van Duyne, R. P., SERS: Materials, applications, and the future. *Materials Today* **2012,** *15* (1-2), 16-25.

38. Yan, Z.; Dutta, S.; Liu, Z.; Yu, X.; Mesgarzadeh, N.; Ji, F.; Bitan, G.; Xie, Y. H., A Label-Free Platform for Identification of Exosomes from Different Sources. *ACS Sens* **2019,** *4* (2), 488-497.

39. Shin, H.; Oh, S.; Hong, S.; Kang, M.; Kang, D.; Ji, Y. G.; Choi, B. H.; Kang, K. W.; Jeong, H.; Park, Y.; Hong, S.; Kim, H. K.; Choi, Y., Early-Stage Lung Cancer Diagnosis by Deep Learning-Based Spectroscopic Analysis of Circulating Exosomes. *ACS Nano* **2020,** *14* (5), 5435-5444.

40. Dong, S.; Wang, Y.; Liu, Z.; Zhang, W.; Yi, K.; Zhang, X.; Zhang, X.; Jiang, C.; Yang, S.; Wang, F.; Xiao, X., Beehive-Inspired Macroporous SERS Probe for Cancer Detection through Capturing and Analyzing Exosomes in Plasma. *ACS Appl Mater Interfaces* **2020,** *12* (4), 5136-5146.

41. Carmicheal, J.; Hayashi, C.; Huang, X.; Liu, L.; Lu, Y.; Krasnoslobodtsev, A.; Lushnikov, A.; Kshirsagar, P. G.; Patel, A.; Jain, M.; Lyubchenko, Y. L.; Lu, Y.; Batra, S. K.; Kaur, S., Label-free characterization of exosome via surface enhanced Raman spectroscopy for the early detection of pancreatic cancer. *Nanomedicine* **2019,** *16*, 88-96.

42. Rojalin, T.; Koster, H. J.; Liu, J.; Mizenko, R. R.; Tran, D.; Wachsmann-Hogiu, S.; Carney, R. P., Hybrid Nanoplasmonic Porous Biomaterial Scaffold for Liquid Biopsy Diagnostics Using Extracellular Vesicles. *ACS Sens* **2020,** *5* (9), 2820-2833.

43. Banaei, N.; Moshfegh, J.; Kim, B., Surface enhanced Raman spectroscopy-based immunoassay detection of tumor-derived extracellular vesicles to differentiate pancreatic cancers from chronic pancreatitis. *Journal of Raman Spectroscopy* **2021,** *52* (11), 1810-1819.

44. Ferguson, S. W.; Nguyen, J., Exosomes as therapeutics: The implications of molecular composition and exosomal heterogeneity. *J Control Release* **2016,** *228*, 179-190.

45. Wen, S. W.; Lima, L. G.; Lobb, R. J.; Norris, E. L.; Hastie, M. L.; Krumeich, S.; Moller, A., Breast Cancer-Derived Exosomes Reflect the Cell-of-Origin Phenotype. *Proteomics* **2019,** *19* (8), e1800180.

Chapter 4. SIM on sEVs for Non-Small Cell Lung Cancer Detection

4.1. Introduction

Lung cancer is the second most common and the most lethal type of cancer worldwide according to American Cancer Society. In 2020, 2.2 million new cases occurred with almost 1.8 million new deaths. A simple but sensitive liquid biopsy platform has the potential to make cancer screening more accessible to a wider variety of patients.[1, 2] This is a great unmet clinical need for early detection diagnostics, particularly for lung cancer.[3, 4] Lung cancer is the single deadliest cancer regardless of gender or ethnicity; however, when caught in early stages, survival rates greatly improve.[5] The current 5-year survival rate of lung cancer is about 22%, which is significantly lower than other leading cancers: colorectal cancer–65%; breast cancer–92%; and prostate cancer–98%.[6, 7] American Cancer Society also lists non-small cell lung cancer (NSCLC) as accounting for almost 85% of all lung cancers. Lung cancer is mostly asymptotic in its early stages, which generally means that when symptoms present and a low-dose CT scan (LDCT) is ordered, the Quality Adjusted Life Expectancy (QALE) is 3.6 years.[8] Using the Surveillance, Epidemiology, and End Results (SEER)-Medicare database and drug costs, the societal burden of treating a Stage IV patient is at least twice as much as a Stage I lung cancer patient. Private insurance costs are extrapolated to mirror similar proportionality.[9, 10] From a purely healthcare economics perspective, this results in higher insurance premiums across the risk pool. But, more importantly, from a humanitarian perspective, the treatment options become dismal as the disease progresses, with most late-stage diagnoses resulting in a poor prognosis. sEV detection via liquid biopsy has the potential to transform the diagnostic and treatment paradigms for lung cancer by making lung cancer screening more accessible to unscreened high-risk populations.[11, 12]

The current standard of care for screening lung cancer is the LDCT scan. However, 93% of the U.S. Preventive Services Task Force's recommended 15 million high-risk populations are not being screened, even before the COVID-19 pandemic paused routine screenings.[13, 14] The National Lung Screening Trial demonstrated a 20% decrease in mortality by using LDCT as a screening modality.[14] However, less than 7% of the at-risk population actually receive the LDCT according to the United States Preventive Services Task Force and the American Lung Association. Simpler screening could overcome resource limitations, reduce the need for annual radiation exposure associated with even a LDCT (approximately 20x conventional X-Ray), detect lung cancer earlier and improve patient outcomes.

With the well-known limitations of imaging technologies in cancer diagnostics, alternative non-invasive biomarkers are needed.[15, 16] Clinical partners have remarked that the problem is not the presence of nodules (which can be detected via a static snapshot with LDCT); the problem is nodules which are increasing in size over time (which require more innovative detection technologies). Longitudinal monitoring via liquid biopsy complements imaging capabilities and can also track minimal residual disease.[17, 18] The LDCT is a complex protocol requiring extensive equipment capital expenditure, nurses, and trained radiologists to interpret the results. Going global, screening becomes even more impractical and inaccessible in undeveloped regions, with lung cancer still claiming almost 1.8 million lives every year.

This chapter includes the experimental studies of SIM spectrally measuring the sEVs collected from bronchoalveolar lavage (BAL) for exploring the feasibility of NSCLC detection via SERS measurements of BAL sEVs. BAL derived sEVs have been used for cargo/content-based analysis (e.g., protein and RNA/DNA expressions discrepancies) between the NSCLC patients and non-cancer control individuals in the format of bulk study.[19-22] Those works indicate that the BAL

derived sEVs hold the potential as the source for NSCLC detection but little has been done from the single-vesicle perspective. Although expected to have low sEV concentration due to its nature of procuring,[23, 24] BAL is directly from the lungs so there is less likelihood of vesicles from other organs based on the current understanding of the sEV biogenesis.[25-27] These purer, lung-derived sEVs allowed us to focus on precise sEV populations with less contamination. The assay successfully distinguished NSCLC from controls after a machine learning (ML) algorithm processed the SERS spectral output. To elucidate the distinguishability between the cancer group and the non-cancer control group, both cross-validation and blind tests were performed. The results further demonstrate the sensitivity of the SERS platform in detecting sEVs in a low vesicle concentration environment. As a pilot study, this work presents that the SERS spectral features collected from the sEVs in the bronchial washes hold the potential for NSCLC detection. At the same time, the methods used in this study are feasible for analyzing sEVs in other biofluids (e.g., saliva, urine, or blood) for disease detection.

4.2. Experimental Procedures

4.2.1. Demographics of BAL sample donors

The participant ages ranged from 20 to 80 and included an equal ratio of males to females with 20% identifying as Hispanic. To complete the study in a timely manner, smoking history was kept as a variable and still obtained almost half smokers v. nonsmokers. It was ensured that the disease samples (n = 10) were representative of diagnoses across all NSCLC stages, from Stage I to Stage IV. The control set (n = 10) was not NSCLC diagnosed.

4.2.2. sEV isolation

Bronchoalveolar lavage fluid samples were stored in the freezer (-20 °C) upon arrival. The isolation process involved size-exclusion chromatography (SEC) using columns purchased from IZON Science. The samples were thawed at room temperature before isolation. A sterile syringe filter (Millex-GP Filter, 0.22 μm) was used to remove the remaining tissue or precipitates in the samples before loading samples into an IZON column (qEV10/35nm) for sEV separation. Sample load volume was set up at 10 mL and phosphate-buffered saline (1x PBS) was used as the medium buffer in the equilibration, elution, and flush process. For detailed information, please refer to the vendor's manual. Isolated sEVs were stored in vials and kept in the -20 °C freezer for future use.

4.2.3. 4.2.3 SERS substrate fabrication

SERS platforms were fabricated according to the same method as previously described.[28] First, a single layer of self-assembled polystyrene (PS) balls (⌀ 500 nm) was generated on a surface of DI water using the Langmuir–Blodgett patterning. The layer was then transferred to a 4" (001) silicon dioxide wafer with a top oxide layer thickness of 50nm. Next, a layer of 50nm of Cr was deposited using electron beam deposition followed by the removal of PS balls using chloroform. The exposed SiO2 were etched using reactive ion etching to selectively expose Si. Next, the exposed silicon was etched using KOH. Inverted nanopyramids with sidewalls at 57.5-degree angles were created because of different etching rates along the [001] and [111] directions of silicon. The model was finished by removing the residual Cr and SiO2 using 48% HF solution. Then, 200 nm of gold film was deposited onto the pitted surface by electron beam deposition and bonded to a carrier wafer using epoxy before lifting off.

4.2.4. Raman spectroscopy

After sEV isolation, 5 μL of each sEV sample solution was transferred onto the SERS substrate using a micropipette. A Reinshaw inVia Raman spectrometer was used for measurements

in this study. All measurements were performed at room temperature. The laser excitation wavelength was 785 nm. The power used was 5 mW. Before usage, the system was first calibrated using the 520 cm-1 peak of silicon. The exposure time was 0.2 s to avoid sample overheating. For collecting SERS spectra from multiple sEVs, a Raman mapping measurement was performed over a 1.2 mm × 1.2 mm square with respect to the center of each sample droplet. The step width was 5 μm to avoid double collecting.

4.2.5. Scanning electron microscopy (SEM)

SEM used in this study was Nova 230. The acceleration voltage was 10 kilovolts. The working distance was ~5.0mm. The images were taken at the magnification between 45,000× and 55,000× (see Figure 2). The electron detector used was TLD (through the lens detector) to obtain the signal from the secondary electrons.

4.2.6. Transmission electron microscopy (TEM)

The TEM operations utilized in this study was assisted by Dr. Wong Hoi Hui at the Electron Imaging Center for Nanomachines, California NanoSystems Institute

4.2.7. SERS Spectral Analysis

On average, 30 different sEVs are obtained for each sample to produce spectra which have 1023 Raman shifts in the range from 553 to 1581 cm-1 (biological information rich region). Preprocessing steps are applied to alleviate the spectral signature fluctuations caused by sample variations, SERS platform heterogeneity and instrument fluctuation. Particularly, fluorescence background subtraction and noise reduction are performed by batch processing based on asymmetric least square fitting and Savitzky-Golay filtering, followed by min-max normalization that proportionally compresses the original intensity range to [0, 1]. To reveal the spectral differences among the three cell line groups, standard linear discriminant analysis (LDA) is used

to reduce the dimensionality for better visualization on a 2-dimensional plot. For machine learning model development, predictive model establishment by supervised learning or classification is the core for the proposed technology. It requires appropriate complexity of the classifier to prevent both underfitting and overfitting for the purpose of generalizing the characteristic signature effectively. Support vector machine (SVM) was used for classification tasks. Unsupervised learning or clustering analysis is performed by Hierarchical clustering analysis with customized distance metrics, it investigates the intrinsic similarities among the analytes SERS signature and serves as an auxiliary to classification. Repeated leave samples out cross validations are then applied to optimize the model settings, followed by blind tests for evaluating the clinical applicability.

4.3. Results and Discussion

4.3.1. Experimental flow

The experimental flow of this study is shown in Figure 4.1. A twenty-participant UCLA Institutional Review Board (IRB) approved study was performed that leveraged the availability of excess bronchoalveolar lavage (BAL) fluid in samples taken from patients suspected of lung cancer. sEVs were then isolated using size-exclusion chromatography (SEC). The isolated sEV solutions were transferred onto the SERS gold nanopyramid substrate before measurements. 5 μL of each of the sample droplets was used for SERS measurements. SERS spectra from individual sEVs were then collected. On average, spectra from 30 individual sEVs were collected from each of the samples. Following confirmed clinical diagnosis, the SERS spectra from the collected sEVs were used to train a customized algorithm to recognize spectral “fingerprints” associated with patients diagnosed as having non-small cell lung cancer/adenocarcinoma (hereafter, NSCLC) or

patients identified as not having NSCLC. An intermediate researcher then provided additional samples to the data scientist without revealing the clinical diagnosis for use as a test set.

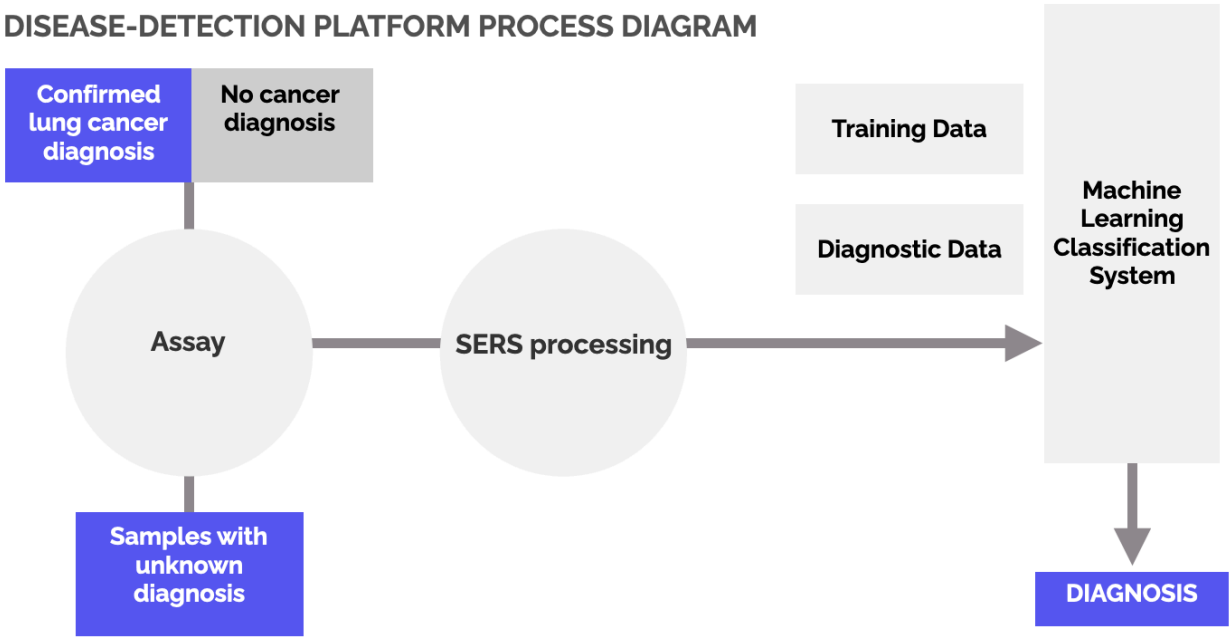

Figure 4.24. Experimental flow of single-vesicle SERS analysis of sEVs from BAL

4.3.2. Sample characterizations

Figure 4.2 exhibits an overview of the sEV samples and the SERS gold nanopyramid substrate used in this study. The transmission electron microscope (TEM) image of the vesicle is shown in Figure 4.2a. The image was taken in the Cryo-EM mode which clearly resolved the lipid membrane of the vesicle. In addition, the image validated the vesicle sizes fell in the range of sEVs. The scanning electron microscope (SEM) images of the SERS substrate before and after sample introduction have been shown in the Figure 4.2b and Figure 4.2c respectively. Figure 4.2b shows the gold nanopyramids array and Figure 4.2c indicates that the vesicles lie in between individual nano-pyramids. The lower quality of resolution in Figure 4.2c comparing to Figure 4.2b

could be attributed to the charging effect caused by the phosphate buffer solution (PBS) crystals lying on top of the substrate.

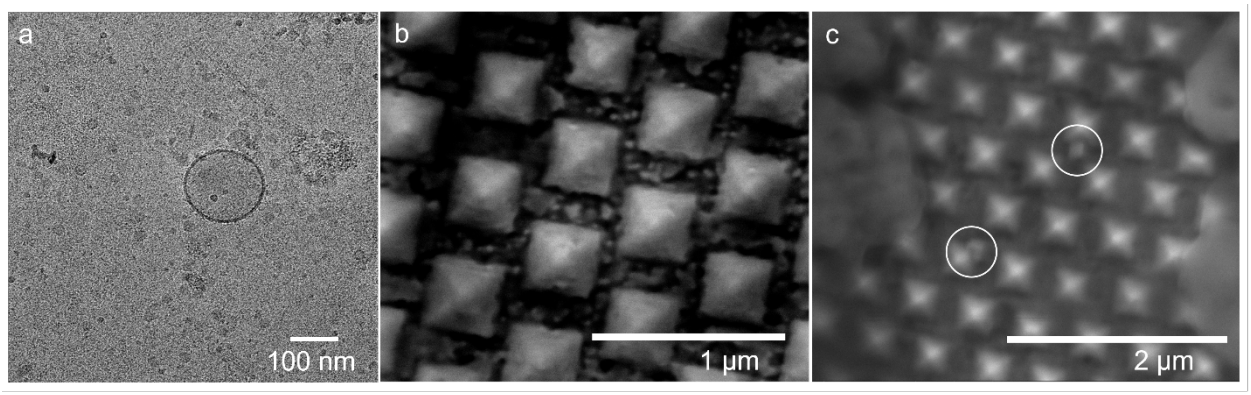

Figure 4.25. a) cryo-EM image of sEVs; b) SEM image of the SERS gold nanopyramid substrate; c) SEM image of the SERS gold nanopyramid substrate after sample introduction

4.3.3. Linear discriminant analysis (LDA) on SERS signatures

LDA was used to reveal the distinguishability between the spectra collected in the NSCLC patient group and the non-NSCLC control group. Such analysis renders the spectral distinguishability before projecting onto a 2D map for better visualization. A 2D rendering is necessary given the complexity of SERS output. Each SERS spectrum is treated as a point in the 1100-dimensional space corresponding to the 1100 wavenumbers of each SERS spectrum (1 sEV = 1100 vectors per spectrum). Next, these "points" are grouped into clusters based on their Euclidian distance from one another. Figure 4.3a exhibits the LDA analysis plot on the SERS spectra collected from sEVs between the cancer and the control group. Each plotted dot represented a single SERS spectrum collected from a single sEV. The spectral differences were measured by the Euclidian distances that separated the dots. Cumulatively, the dots self-clustered according to their origin, indicating spectral distinguishability. Zooming in to the cancer cluster, the spectra collected in NSCLC patients with early and late-stage cancer were compared. The LDA

plot shown in Figure 4.3b also elucidated the spectral distinguishability between the early and late-stage patient groups. Collectively the LDA graphs illustrate the potential of using SERS spectra of individual sEVs derived from BAL for detecting NSCLC. Obtained results indicate the spectral distinguishability between both the control and cancer groups, and also between the early and late-stage NSCLC groups.

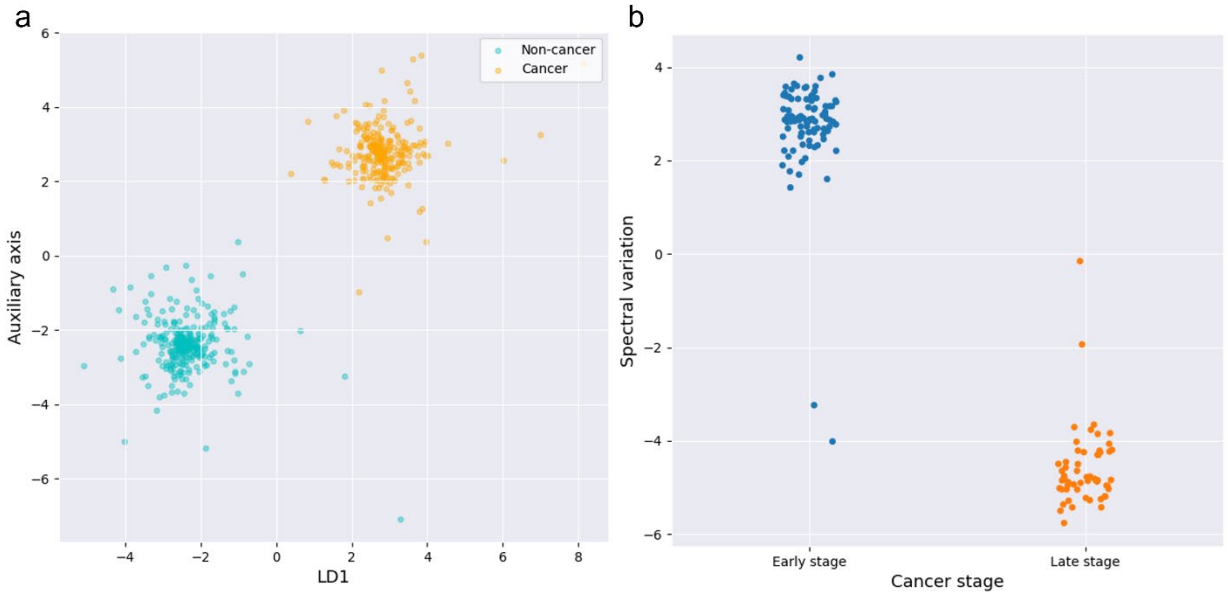

Figure 4.26. a) LDA analysis results of SERS spectra collected from sEVs between NSCLC patients and controls; b) between early stage and late stage NSCLC patients

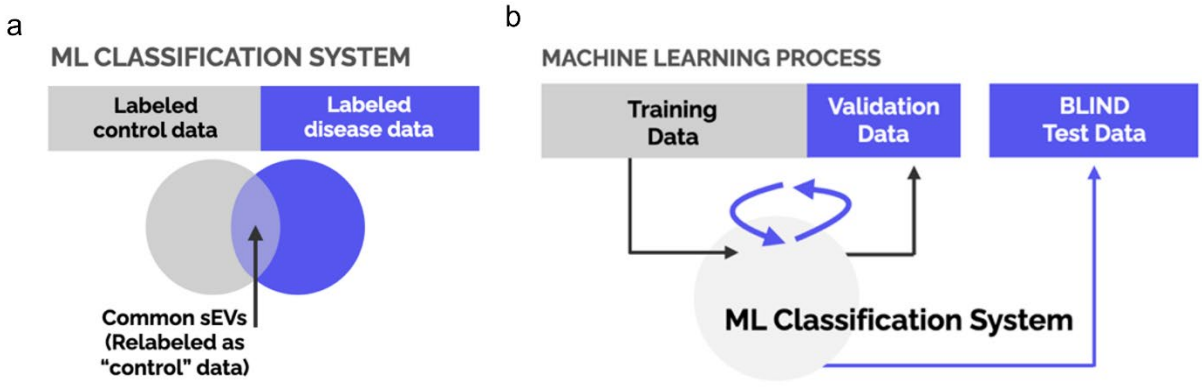

Figure 4.27. a) Schematic of SERS spectra relabeling; b) SVM model training and blind testing

4.3.4. Model training and blind test using support vector machine (SVM)

To explore the clinical applicability of the SIM platform, a blind test was performed. SVM was used to provide the machine learning (ML) model training and testing. SERS spectra collected from 6 donors in each of the patient and the control group (total n = 12) were used as the training set and the SERS spectra collected from the rest of the 8 individuals (4 patients and 4 controls) were left out as the test set. Here the operator for SVM analysis of the SERS spectra was kept "blind" to the diagnosis results until testing concluded. Figure 4.4 shows the schematics of model training and testing.

That cancer cells secrete cancer-cell specific sEVs leads to a natural conclusion that the anticipated composition of sEV populations isolated from biofluids will include disease specific sEVs. However, the disease population is known to be mixed with normal sEV populations in the biofluids. Given this co-existence of healthy and unhealthy sEV populations, ML training based on SERS spectra collected from individual sEVs would cause a mislabeling issue that impedes model accuracy. To overcome this, the spectral features collected from sEVs was relabeled based on their spectral similarities as shown in Figure 4.4a. The relabeling process was through a spectral feature comparison between the spectra from NSCLC patients and the controls. The "common" spectra, indicating the existences of the normal sEVs, in the patient group are relabeled as control. It should be pointed out that the limited control data size makes it challenging to parse all the normal spectra in the patient group. However, the principles behind this process help partially correct the mislabeling issue and significantly improve model accuracy.

Figure 4.4b shows the ML schematic for model training and blind testing based on relabeled data. The classification model was built on the training set of 12 individuals, 6 from each group. Then, the model was applied to predict every donor in the test set whether the person was

likely to be a NSCLC patient or a control. The predicted result was then compared with the actual clinical diagnosis. The SVM model correctly predicted all the diagnosis results of the 8 individuals in the test set based on SERS spectra collected from the BAL derived sEVs. As a proof-of-concept, obtained results suggest the clinical applicability of the SERS platform to analyze sEVs for NSCLC detection. This ML platform would benefit greatly from samples with a higher sEV count (the recent bronchoalveolar lavage study had an average sEV count of 30 with a median of 22) on the vesicle concentrations of the sample biofluids. Future studies involve improving the throughput on both sample sizes and sEV count per sample. Nonetheless, this study clearly illustrates the potential of using label-free single-vesicle SERS to detect and predict NSCLC based on sEV subpopulations. The methodology involved in this work has the potential for disease detection other than NSCLC with specific validations.

4.4. Conclusion

BAL-based liquid biopsy is not a true "noninvasive" approach with certain invasiveness, but BAL has the advantage of sEV purity with most sEVs likely secreted from lung cells as opposed to sEVs from other organs. The reason that BAL was used in this work is to conclusively establish the biomarker value of lung-derived sEVs thereby laying the foundation for subsequent studies aimed at noninvasive biofluids such as blood, saliva, and urine. The fact that the platform is capable of detecting sEVs from BAL with its inherent low vesicle concentration suggests an acceptable detection limit considering the vesicle counts obtained. This work successfully demonstrated that anomalous sEVs are detectable using the platform's nanotechnology protocols with clinical samples from lung cancer patients, and not detectable in the control group. This BAL pilot study allowed us to intentionally stress-test the platform and still achieve positive results

which bodes well for further studies using biofluids with known higher vesicle concentrations. The inherent non-destructive biochemical analysis of single vesicles via SERS makes this biosensing platform powerful when carefully combined with the appropriate ML algorithms.

4.5. Reference

1. Pantel, K.; Alix-Panabieres, C., Real-time liquid biopsy in cancer patients: fact or fiction? *Cancer Res* **2013,** *73* (21), 6384-8.

2. Siravegna, G.; Marsoni, S.; Siena, S.; Bardelli, A., Integrating liquid biopsies into the management of cancer. *Nat Rev Clin Oncol* **2017,** *14* (9), 531-548.

3. Rolfo, C.; Mack, P. C.; Scagliotti, G. V.; Baas, P.; Barlesi, F.; Bivona, T. G.; Herbst, R. S.; Mok, T. S.; Peled, N.; Pirker, R.; Raez, L. E.; Reck, M.; Riess, J. W.; Sequist, L. V.; Shepherd, F. A.; Sholl, L. M.; Tan, D. S. W.; Wakelee, H. A.; Wistuba, II; Wynes, M. W.; Carbone, D. P.; Hirsch, F. R.; Gandara, D. R., Liquid Biopsy for Advanced Non-Small Cell Lung Cancer (NSCLC): A Statement Paper from the IASLC. *J Thorac Oncol* **2018,** *13* (9), 1248-1268.

4. Kwapisz, D., The first liquid biopsy test approved. Is it a new era of mutation testing for non-small cell lung cancer? *Ann Transl Med* **2017,** *5* (3), 46.

5. Sung, H.; Ferlay, J.; Siegel, R. L.; Laversanne, M.; Soerjomataram, I.; Jemal, A.; Bray, F., Global Cancer Statistics 2020: GLOBOCAN Estimates of Incidence and Mortality Worldwide for 36 Cancers in 185 Countries. *CA Cancer J Clin* **2021,** *71* (3), 209-249.

6. Barta, J. A.; Powell, C. A.; Wisnivesky, J. P., Global Epidemiology of Lung Cancer. *Ann Glob Health* **2019,** *85* (1).

7. Ko, E. C.; Raben, D.; Formenti, S. C., The Integration of Radiotherapy with Immunotherapy for the Treatment of Non-Small Cell Lung Cancer. *Clin Cancer Res* **2018,** *24* (23), 5792-5806.

8. Black, W. C., Computed tomography screening for lung cancer in the National Lung Screening Trial: a cost-effectiveness analysis. *J Thorac Imaging* **2015,** *30* (2), 79-87.

9. Enstone, A.; Greaney, M.; Povsic, M.; Wyn, R.; Penrod, J. R.; Yuan, Y., The Economic Burden of Small Cell Lung Cancer: A Systematic Review of the Literature. *Pharmacoecon Open* **2018,** *2* (2), 125-139.

10. Andreas, S.; Chouaid, C.; Danson, S.; Siakpere, O.; Benjamin, L.; Ehness, R.; Dramard-Goasdoue, M. H.; Barth, J.; Hoffmann, H.; Potter, V.; Barlesi, F.; Chirila, C.; Hollis, K.; Sweeney, C.; Price, M.; Wolowacz, S.; Kaye, J. A.; Kontoudis, I., Economic burden of resected (stage IB-IIIA) non-small cell lung cancer in France, Germany and the United Kingdom: A retrospective observational study (LuCaBIS). *Lung Cancer* **2018,** *124*, 298-309.

11. Zhou, B.; Xu, K.; Zheng, X.; Chen, T.; Wang, J.; Song, Y.; Shao, Y.; Zheng, S., Application of exosomes as liquid biopsy in clinical diagnosis. *Signal Transduct Target Ther* **2020,** *5* (1), 144.

12. Salehi, M.; Sharifi, M., Exosomal miRNAs as novel cancer biomarkers: Challenges and opportunities. *J Cell Physiol* **2018,** *233* (9), 6370-6380.

13. Ardila, D.; Kiraly, A. P.; Bharadwaj, S.; Choi, B.; Reicher, J. J.; Peng, L.; Tse, D.; Etemadi, M.; Ye, W.; Corrado, G.; Naidich, D. P.; Shetty, S., End-to-end lung cancer screening with three-dimensional deep learning on low-dose chest computed tomography. *Nat Med* **2019,** *25* (6), 954-961.

14. Tanoue, L. T.; Tanner, N. T.; Gould, M. K.; Silvestri, G. A., Lung cancer screening. *Am J Respir Crit Care Med* **2015,** *191* (1), 19-33.

15. Fass, L., Imaging and cancer: a review. *Mol Oncol* **2008,** *2* (2), 115-52.

16. Liang, W.; Zhao, Y.; Huang, W.; Gao, Y.; Xu, W.; Tao, J.; Yang, M.; Li, L.; Ping, W.; Shen, H.; Fu, X.; Chen, Z.; Laird, P. W.; Cai, X.; Fan, J. B.; He, J., Non-invasive diagnosis

of early-stage lung cancer using high-throughput targeted DNA methylation sequencing of circulating tumor DNA (ctDNA). *Theranostics* **2019,** *9* (7), 2056-2070.

17. Martins, I.; Ribeiro, I. P.; Jorge, J.; Goncalves, A. C.; Sarmento-Ribeiro, A. B.; Melo, J. B.; Carreira, I. M., Liquid Biopsies: Applications for Cancer Diagnosis and Monitoring. *Genes (Basel)* **2021,** *12* (3).

18. Chen, X.; Gole, J.; Gore, A.; He, Q.; Lu, M.; Min, J.; Yuan, Z.; Yang, X.; Jiang, Y.; Zhang, T.; Suo, C.; Li, X.; Cheng, L.; Zhang, Z.; Niu, H.; Li, Z.; Xie, Z.; Shi, H.; Zhang, X.; Fan, M.; Wang, X.; Yang, Y.; Dang, J.; McConnell, C.; Zhang, J.; Wang, J.; Yu, S.; Ye, W.; Gao, Y.; Zhang, K.; Liu, R.; Jin, L., Non-invasive early detection of cancer four years before conventional diagnosis using a blood test. *Nat Commun* **2020,** *11* (1), 3475.

19. Zhang, Y.; Liu, Z.; Li, S.; Wang, M.; Dai, D.; Jing, H.; Liu, L., Upregulation of E-cadherin in bronchoalveolar lavage fluid-derived exosomes in patients with lung cancer. *Thorac Cancer* **2020,** *11* (1), 41-47.

20. Yang, Y.; Ji, P.; Wang, X.; Zhou, H.; Wu, J.; Quan, W.; Shang, A.; Sun, J.; Gu, C.; Firrman, J.; Xiao, W.; Sun, Z.; Li, D., Bronchoalveolar Lavage Fluid-Derived Exosomes: A Novel Role Contributing to Lung Cancer Growth. *Front Oncol* **2019,** *9*, 197.

21. Rodriguez, M.; Silva, J.; Lopez-Alfonso, A.; Lopez-Muniz, M. B.; Pena, C.; Dominguez, G.; Garcia, J. M.; Lopez-Gonzalez, A.; Mendez, M.; Provencio, M.; Garcia, V.; Bonilla, F., Different exosome cargo from plasma/bronchoalveolar lavage in non-small-cell lung cancer. *Genes Chromosomes Cancer* **2014,** *53* (9), 713-24.

22. Kim, J. E.; Eom, J. S.; Kim, W. Y.; Jo, E. J.; Mok, J.; Lee, K.; Kim, K. U.; Park, H. K.; Lee, M. K.; Kim, M. H., Diagnostic value of microRNAs derived from exosomes in

bronchoalveolar lavage fluid of early-stage lung adenocarcinoma: A pilot study. *Thorac Cancer* **2018,** *9* (8), 911-915.

23. Carvalho, A. S.; Moraes, M. C. S.; Hyun Na, C.; Fierro-Monti, I.; Henriques, A.; Zahedi, S.; Bodo, C.; Tranfield, E. M.; Sousa, A. L.; Farinho, A.; Rodrigues, L. V.; Pinto, P.; Barbara, C.; Mota, L.; Abreu, T. T.; Semedo, J.; Seixas, S.; Kumar, P.; Costa-Silva, B.; Pandey, A.; Matthiesen, R., Is the Proteome of Bronchoalveolar Lavage Extracellular Vesicles a Marker of Advanced Lung Cancer? *Cancers (Basel)* **2020,** *12* (11).

24. Johnsen, K. B.; Gudbergsson, J. M.; Andresen, T. L.; Simonsen, J. B., What is the blood concentration of extracellular vesicles? Implications for the use of extracellular vesicles as blood-borne biomarkers of cancer. *Biochim Biophys Acta Rev Cancer* **2019,** *1871* (1), 109-116.

25. Farooqi, A. A.; Desai, N. N.; Qureshi, M. Z.; Librelotto, D. R. N.; Gasparri, M. L.; Bishayee, A.; Nabavi, S. M.; Curti, V.; Daglia, M., Exosome biogenesis, bioactivities and functions as new delivery systems of natural compounds. *Biotechnol Adv* **2018,** *36* (1), 328-334.

26. Hessvik, N. P.; Llorente, A., Current knowledge on exosome biogenesis and release. *Cell Mol Life Sci* **2018,** *75* (2), 193-208.

27. Zhang, Y.; Liu, Y.; Liu, H.; Tang, W. H., Exosomes: biogenesis, biologic function and clinical potential. *Cell Biosci* **2019,** *9*, 19.

28. Wang, P.; Liang, O.; Zhang, W.; Schroeder, T.; Xie, Y. H., Ultra-sensitive graphene-plasmonic hybrid platform for label-free detection. *Adv Mater* **2013,** *25* (35), 4918-24.

Chapter 5. Summary and Future Works

5.1. Summary

The presented dissertation describes the works towards the minimally invasive cancer detections through analyzing individual sEVs in bodily fluids using SERS. Utilizing SIM, which combines single vesicle SERS and machine learning, this thesis goes from analyzing sEVs and their subpopulations to the applications of SIM exploring the potential clinical applicability for minimally invasive cancer detections via detecting sEVs in bodily fluids. The following of this section gives a summary of the presented works.

Chapter 2 includes the experimental studies of SIM examining individual sEVs isolated based on their size discrepancies. The biomolecular composition of each vesicle examined was reflected by its corresponding SERS spectral features (biomolecular "fingerprints") with their roots in the composition of their collective Raman-active bonds. Origins of the SERS spectral features were validated through a comparative analysis between SERS and mass spectrometry (MS). SERS fingerprinting of individual vesicles was effective to overcome the challenge posed by EV population averaging, opening the possibility of analyzing the variations in biomolecular composition between the vesicles of similar and/or different sizes. The obtained data suggested that each of the size-based subpopulation of sEVs contained particles with predominantly similar SERS spectral features. Over 84% of the vesicles residing within in a particular group that were clearly distinguishable from that of the other EV sub-populations despite some features of spectral variations within each sub-population. The results indicate that there exists a correlation between sEV size-based subpopulations and their biomolecular composition from the perspective of

individual vesicles. They further shed light on the possibility that the sEVs with different sizes might inherently contain different cargoes. Such observations potentially contribute to the future developments of sEV sorting and sub-fractionation based on their biochemical compositions.

Chapter 3 illustrates the application of SIM for minimally invasive gastric cancer detection via spectrally analyzing sEVs in different bodily fluids. Based on the results described in Chapter 2, the working hypothesis here is that the composition information of the collective Raman active bonds inside individual sEVs of human donors obtained by SERS hold the potential for non-invasive gastric cancer detections. The complexity when working with samples from the human donors comes from the fact that there co-exist sEVs from both the normal tissues and cancer cells inside the patients' bodily fluids such as blood and saliva. At the same time, currently it is challenging for the isolation methods to separate the two groups. Such co-existence causes "mislabeling" during the machine learning analysis which would reduce the distinguishing accuracy. Here, a machine learning was customized to help overcome or at least partially correct such issue, boosting the detection accuracy. The spectra from the patients were relabeled before machine training and classifier developments based on their spectral similarities to those from the control individuals. The spectra from patients that shared high similarities to the ones from the control group were relabeled as "control" while the unique ones from the patient groups retained their labels of "cancer". sEVs from the tissue, blood, and saliva of GC patients and non-GC participants were collected (n = 15 each) and analyzed. The algorithm prediction accuracies were reportedly 90%, 85%, and 72%. "Leave-a-pair-of-samples out" validation was further performed to test the clinical potential. The area under the curve (AUC) of each the receiver operating characteristic (ROC) curve was 0.96, 0.91, and 0.65 in tissue, blood, and saliva respectively. In addition, through comparing the SERS fingerprints of individual vesicles, a possible way of tracing

the biogenesis pathways of patient-specific sEVs from tissue to blood to saliva was provided based on searching for the patient-unique sEVs spectra existing across tissue, blood and saliva. Through the experiments, areas for further improvements were also proposed. First, for single-vesicle SERS, the efficiency of the measurements needed to be further improved from the technical perspective. Raman machine automation offers a potential solution with further optimizations as the process could run unattended, offering the opportunity of overnight running. Secondly, the chance of hitting cancer-derived sEVs need to be improved to increase the detection accuracy especially when working with the bodily fluids as the source of sEV extraction.

Chapter 4 contains the experimental studies of SIM on non-small cell lung cancer detection (NSCLC) via analyzing sEVs from bronchoalveolar lavage (BAL). The reason for choosing BAL as the source of vesicle isolation is that BAL is directly from lung with the expectation that most of the vesicles come from normal lung tissue or NSCLC depending on the situation of the donor. Such biofluid contains less population of sEVs from other parts of the body which presents less of the challenge included in Chapter 3 even though the collecting of such biofluid is not completely in the fashion of non-invasive detections. LDA plots showed the spectral distinguishability between the sEVs collected from the cancer patients and non-cancer controls. In additions, the spectra collected from the vesicles from early stage and late-stage patients were also revealed to be distinguishable. As a-proof-of-concept-study such results indicate the potential of SIM for precision cancer detections with the need for further validations. Moreover, a blind test blind based on the collected spectra was perform the explore the potential clinical applicability of SIM. Such blind test was conducted in the circumstance that the data analysis operator was kept from the diagnostic results of the donors. Also, the spectra from the “test” set were completely kept from the machine learning training process to mimic the real clinical test. Such process was to test

whether the model developed with the limited training sample size had the potential to be applied for testing unknown samples, a natural question in the development of a cancer screening method. In this pilot study, the machine learning model correctly predicted the diagnostic results of all the samples in the "test" group. Given the limited sample size of n = 10 in each of the cancer and non-cancer group, further validations with increasing number of donors are needed despite the promising results. In addition, BAL is known to have low concentrations of sEVs due to its nature of collecting. Therefore, concentrating of vesicles onto the gold nanopyramid is also desired prior to the SERS measurements and calls for the need of further technical developments.

5.2. Directions of Future Works

5.2.1. Improvement of the effective throughput of SIM detecting sEVs

As a single-vesicle based technology, SIM has the potential to reveal the subtle compositional difference among individual vesicles, providing a complementary perspective to the population/bulk detection methods. However, there is often a tradeoff between detection sensitivity and throughput. One of the major challenges with any single-vesicle based technology is the throughput and SIM is not an exception. Without decent throughput, it is almost impossible to further validate the indications obtained from the current results with large number of patient samples (e.g., > 50 or even > 100 individuals) to further push SIM towards clinical applications.

Raman measurement automation presents a feasible aspect for such improvement from increasing the number of vesicles detected per sample and enabling the measurement processes to run 24/7. However, current implementations of the Raman automation mainly come from the in-house design of a script to mimic the maneuvers of the OS of the Raman system which might be

vulnerable to unexpected systematic errors. Future works could potentially involve the collaborations with the Raman system developer to establish a built-in function of the auto runs that would be robust to operate without the risk of being interrupted by systematic errors.

With the effort on Raman measurement automation, it is possible to increase the absolute number of sEVs measured per sample by at least 10x (e.g., ~500 vesicles measured with the help from overnight runs). However, given the fact that the concentration of the target sEV varies across different types of bodily fluids, even such throughput boost thanks to the automatic measurements might still not be enough to catch even one target sEV from cancer. For example, if the population of the cancer-derived sEVs in certain bodily fluids accounts for less than 0.2% of the total vesicular population, it might not be enough just to scan 500 vesicles per sample. Therefore, though boosting the absolute number of measured vesicles is an important aspect, it is just one side of the story. It is crucial to increase the chance of "hitting" the interested cancer-derived sEVs during the measurements. If successful, not only the throughput of SIM would be significantly boosted, but it also pushes the applicability of SIM for the detection of cancer at the earliest stage where the cancer tissues just start growing and only small population of cancer sEVs are released.

One of the potential methods to achieve such goal is through surface functionalization that attaches antibodies to the surface of the gold nanopyramid substrates. Those antibodies need to be selective to serve the purpose of immobilizing the vesicles of interest and others could be washed off. It should be noted that such process is not designed to capture sEVs from cancers as currently there is not a well-established cancer biomarker for sEVs. The process serves to reduce the number of sEVs from other parts of the body. Taking gastric cancer detection as an example, the goal through functionalization is to be able to focus on sEVs only from stomach, shown in Figure 5.1.

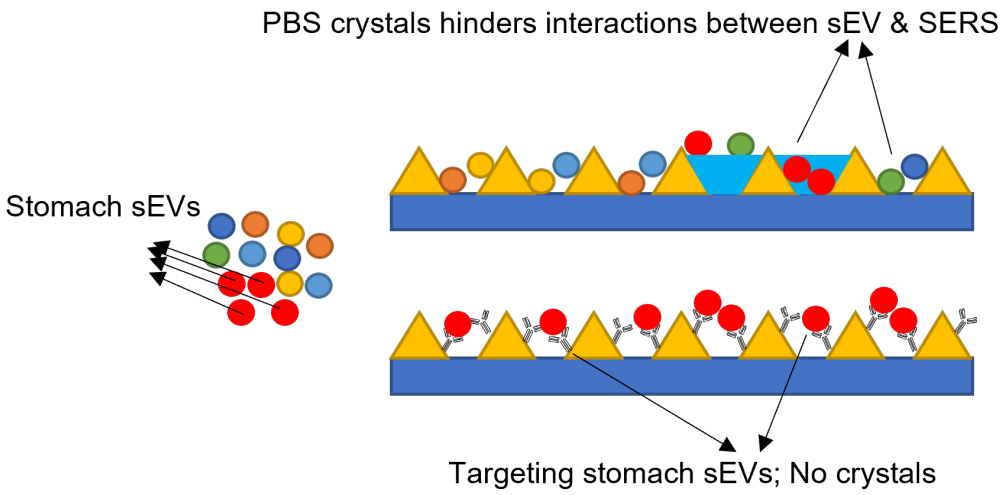

Figure 5.28. Schematic of targeting the sEVs of interest for improving the effective throughput of SIM by surface functionalization of the gold nanopyramid substrate

Vesicles from other parts of the body will be washed off. In this way, most of the vesicles analyze by SIM will be coming from stomach, given the non-zero probability of residual staying after the wash. To achieve this, the selection of antibody is the key. Based on the latest effort of searching, anti-CLDN 18 and especially its isoform anti-CLDN18.2 are the two candidates for the gastric cancer detection case. CLDN18 protein is a cell surface protein that has a significant higher expression level across the stomach cells comparing to other types of cells. Given the understanding of the sEV biogenesis, it is expected that sEVs from the stomach cells also carry higher levels of CLDN18 protein. Future works involve validating this expectation. CLDN18.2 has an expression strictly confined to differentiated epithelial cells of the gastric mucosa where ~90% of the gastric cancers happen. A preliminary experiment has done using anti-CLDN18 to functionalize the gold nanopyramid substrate followed by applying the sEVs extracted from saliva. Here samples come from 4 different donors in each of cancer and non-cancer group. ~200 vesicles were measured for each of the samples, benefited from the Raman automation. As shown in Figure

5.2 the AUC of the ROC is 0.93, comparing to the results presented in Chapter 3 of 0.65. The results show preliminary success but require further validations and optimization as future works. First, validation is needed using increasing number of samples. Second, the protocol used in the functionalization mentioned above was adopted from the one published by ThermoFisher Scientific without any optimization. Since the protocol of the surface functionalization plays an equally important role as antibody selection, future works could involve the protocol optimization with the step-by-step characterizations.

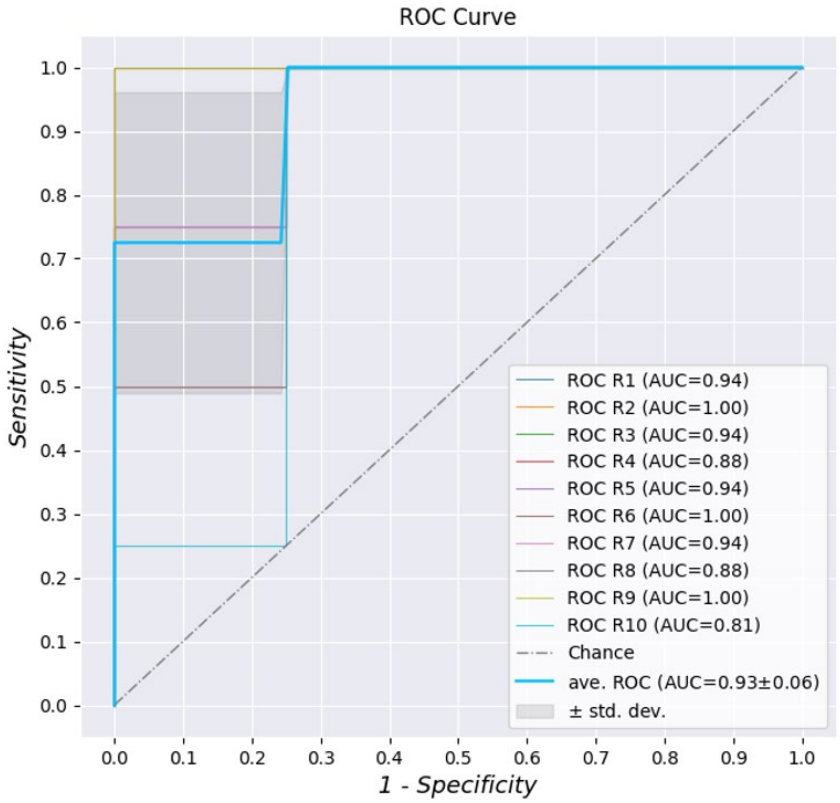

Figure 5.29. ROC curve of distinguishing sEVs from cancer patients and non-cancer individuals (n = 4 for each group). Gold nanopyramid functionalized by anti-CLDN18

5.2.2. Exploring the metabolism and/or function differences among sEV subpopulations

As the results obtained by SIM indicate the possibility that the sEV subpopulations isolated by size discrepancies carry distinctively various cargos, a natural question to ask is whether such variation will lead to functional or metabolic differences.[1, 2] The rationale of such question is also supported by a realization that the fraction of sEVs that carry potentially functional RNAs is at only percent level.[3-5] Data from SIM echoes such observation. As shown in Figure 5.3 by locating the unique peaks of the 4 types of nucleobases in each of the spectra collected from sEVs of 5 different cell-lines, the fraction of sEVs carrying all the 4 types, the prerequisite of RNAs, is only at the percent level in every cell-line.

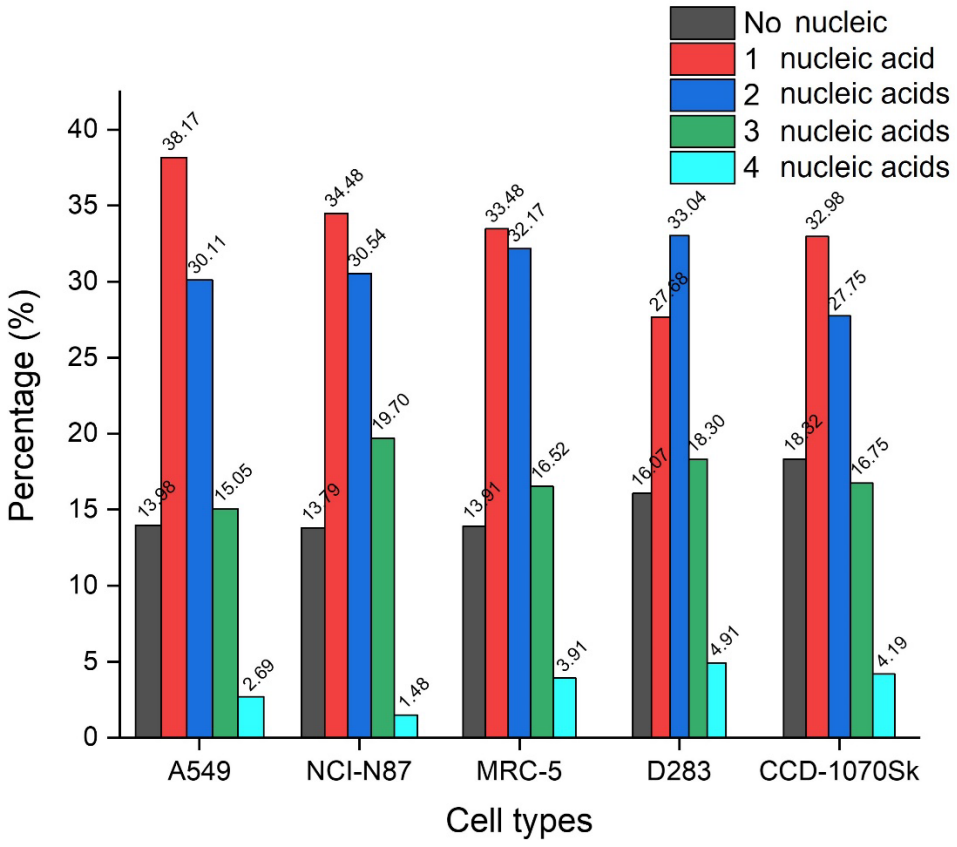

Figure 5.30. Distribution of sEVs containing various types of nucleic acids from five different cell-lines

It is reasonable to question whether those vesicles containing functional RNAs function the same or differently as others, thus worth further investigating the functional differences among different sEV subpopulations. To establish the correlation between various sEV subpopulations and their functions, SIM involves the monitoring the sEVs released by the recipient cells, at the single vesicle level, before and after the uptakes of sEVs subpopulations with different sizes. Such work could contribute to the further establishment of the knowledge base about sEV cargo deliveries. One of the fields that could potentially be benefited from the work is sEV based medicine delivery for disease treatments since it might serve as a guidance for choosing the best group of sEVs for the cargo transit.

5.3. Reference

1. Zhang, H.; Freitas, D.; Kim, H. S.; Fabijanic, K.; Li, Z.; Chen, H.; Mark, M. T.; Molina, H.; Martin, A. B.; Bojmar, L.; Fang, J.; Rampersaud, S.; Hoshino, A.; Matei, I.; Kenific, C. M.; Nakajima, M.; Mutvei, A. P.; Sansone, P.; Buehring, W.; Wang, H.; Jimenez, J. P.; Cohen-Gould, L.; Paknejad, N.; Brendel, M.; Manova-Todorova, K.; Magalhaes, A.; Ferreira, J. A.; Osorio, H.; Silva, A. M.; Massey, A.; Cubillos-Ruiz, J. R.; Galletti, G.; Giannakakou, P.; Cuervo, A. M.; Blenis, J.; Schwartz, R.; Brady, M. S.; Peinado, H.; Bromberg, J.; Matsui, H.; Reis, C. A.; Lyden, D., Identification of distinct nanoparticles and subsets of extracellular vesicles by asymmetric flow field-flow fractionation. *Nat Cell Biol* **2018,** *20* (3), 332-343.

2. Willms, E.; Johansson, H. J.; Mager, I.; Lee, Y.; Blomberg, K. E.; Sadik, M.; Alaarg, A.; Smith, C. I.; Lehtio, J.; El Andaloussi, S.; Wood, M. J.; Vader, P., Cells release subpopulations of exosomes with distinct molecular and biological properties. *Sci Rep* **2016,** *6*, 22519.

3. Ferguson, S. W.; Nguyen, J., Exosomes as therapeutics: The implications of molecular composition and exosomal heterogeneity. *J Control Release* **2016,** *228*, 179-190.

4. Wei, Z.; Batagov, A. O.; Schinelli, S.; Wang, J.; Wang, Y.; El Fatimy, R.; Rabinovsky, R.; Balaj, L.; Chen, C. C.; Hochberg, F.; Carter, B.; Breakefield, X. O.; Krichevsky, A. M., Coding and noncoding landscape of extracellular RNA released by human glioma stem cells. *Nat Commun* **2017,** *8* (1), 1145.

5. Chevillet, J. R.; Kang, Q.; Ruf, I. K.; Briggs, H. A.; Vojtech, L. N.; Hughes, S. M.; Cheng, H. H.; Arroyo, J. D.; Meredith, E. K.; Gallichotte, E. N.; Pogosova-Agadjanyan, E. L.; Morrissey, C.; Stirewalt, D. L.; Hladik, F.; Yu, E. Y.; Higano, C. S.; Tewari, M., Quantitative

and stoichiometric analysis of the microRNA content of exosomes. *Proc Natl Acad Sci U S A* **2014,** *111* (41), 14888-93.