

Single-Cell Analysis and Classification according to Multiplexed Proteins via Microdroplet-Based Self-Driven Magnetic Surface-Enhanced Raman Spectroscopy Platforms Assisted with Machine Learning Algorithms

Jiaqi Wang, Lili Cong, Wei Shi, Weiqing Xu, and Shuping Xu*



Cite This: *Anal. Chem.* 2023, 95, 11019–11027



Read Online

ACCESS |



Metrics & More

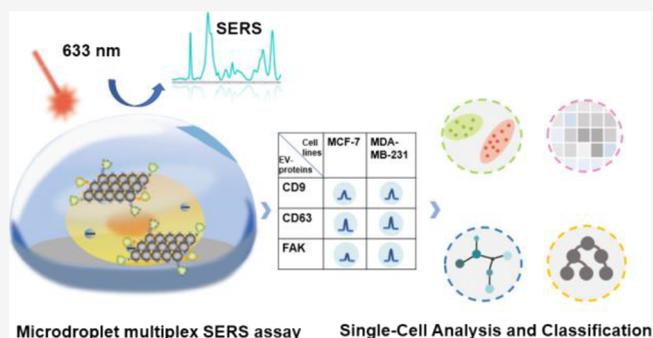


Article Recommendations



Supporting Information

ABSTRACT: A microdroplet-based surface-enhanced Raman spectroscopy (microdroplet SERS) platform was constructed to envelop individual cells in microdroplets, followed by the SERS detection of their extracellular vesicle-proteins (EV-proteins) via the in-drop immunoassays by use of immunomagnetic beads (*i*MBs) and immuno-SERS tags (*i*SERS tags). A unique phenomenon is found that *i*MBs can start a spontaneous reorientation on the probed cell surface based on the electrostatic force-driven interfacial aggregation effect, which leads EV-proteins and *i*SERS tags to be gathered from a liquid phase to a cell membrane interface and significantly improves SERS sensitivity to the single-cell analysis level due to the formation of numbers of SERS hotspots. Three EV-proteins from two breast cancer cell lines were collected and further analyzed by machine learning algorithmic tools, which will be helpful for a deeper understanding of breast cancer subtypes from the view of EV-proteins.



Microdroplet multiplex SERS assay Single-Cell Analysis and Classification

INTRODUCTION

Breast cancer (BC) is the most common female malignancy. Among the four subtypes of BC, triple-negative breast cancer (TNBC) has a substantially lower survival rate than the nonmetastatic subtypes due to its high invasiveness, metastasis susceptibility, poor prognosis, ease of recurrence, and drug resistance.^{1,2} Meanwhile, TNBC lacks a broad spectrum of marker receptors.³ Therefore, finding appropriate biomarkers for early diagnosis and detection of different subtypes is essential for cancer treatment. Related studies have shown that surface proteins of extracellular vesicles (EVs) isolated from the body fluids of BC patients are expected to be biomarkers for prediction, diagnosis, chemotherapy response, and prognosis.⁴ EVs are differentiated from endosomes, plasma membranes, or the endoplasmic reticulum, which are released from intracellular to extracellular, and they play a crucial role by carrying a number of bioactive molecules that act as mediators for intercellular communication.^{5,6} Among them, the regulation of intercellular signaling communication events cannot be achieved without the involvement of integrin membrane proteins. Tetraspanin proteins, also known as transmembrane four superfamily proteins, are widely distributed in EVs, and they are involved in integrin-dependent functions of tumor cells through extracellular structural domains, transmembrane structural domains, and cytoplasmic

structural domains, as well as in the regulation of physiopathological events such as receptor signaling.^{7–10} In refractory BC, focal adhesion kinase (FAK), closely associated with tumorigenesis, progression, and metastasis, plays a vital role in maintaining tumor angiogenesis and becomes a target for cancer therapy. Increased FAK expression is usually associated with metastatic diseases.^{11–13} Thus, analyzing exosomal markers can help reveal the characteristics of different BC subtypes and provide a better strategy for early cancer diagnosis.

Various techniques have been reported for studying EV-proteins, including thermophoretic aptasensors,¹⁴ nanomixing-enhanced EV subpopulation characterization platforms,¹⁵ surface plasmon resonance (SPR) sensors,¹⁶ electrochemical immunoassay,¹⁷ immunofluorescence, and surface-enhanced Raman spectroscopy (SERS).¹⁶ Most of these studies have been conducted on population cells, and studies based on single-cell diversity and heterogeneity have been little

Received: March 23, 2023

Accepted: June 28, 2023

Published: July 7, 2023



developed. In contrast, biological heterogeneity is widespread and intricate in tumor cells, especially in highly metastatic subtypes. Thus, single-cell analysis is critical in identifying different subpopulations and revealing the interactions of analytes. The microfluidic droplet technology that allows for high-throughput production of picoliter volumes of microdroplets by precisely manipulating the reagent feed volume has become one of the popular single-cell analysis techniques. Each microdroplet is an individual microreactor, and each microchamber is an independent closed system, which makes the microdroplet-based technology available for high-throughput single-cell analysis.¹⁸

Single-cell analysis is always accompanied by a large amount of high-dimensional sample data, which is increasingly limited by manual control-based methods. High-dimensional computational tools not only allow the handling of richer and more complex samples but also enable exploratory analysis, such as exploiting biological component interaction networks and patterns to reveal latent cellular subpopulations, which can be adopted clinically to predict disease progression outcomes. Machine learning-related dimensionality reduction, clustering, and prediction algorithms are the most widely used methods, in which unsupervised learning-based dimensionality reduction or clustering methods can explore and characterize unknown subpopulations for single-cell data from multiple samples that can be analyzed combinatorially or compared, and supervised learning-based prediction class methods rely on input from external variables to return interpretable models for prediction of new samples. These algorithms generate different insights from single-cell data sets.^{19–21}

SERS technology has been widely used in the biomedical field because of its high sensitivity, noninvasive *in situ* analysis, multiplexed detection with single wavelength excitation, and subcellular region analysis.²² SERS hotspots, where the SERS gain is very significant, arise from metallic nanogaps when two metal nanoparticles are close enough. The combination of SERS technology and microdroplet technology enables high-throughput, nondestructive, and long-term monitoring of single cells.²³ In order to analyze biological samples secreted from a single cell by SERS, constructing an elaborate SERS strategy that supports enough SERS hotspots is highly required to meet high demands for single-cell detection sensitivity.²⁴ The developed single-cell SERS strategies are mainly based on magnetic SERS, which can form collective metal nanoaggregates due to the magnetic field driving.²³ Recently, a cell membrane pretreatment strategy was developed by our group for building an immunoassay above the cell membrane for secreted proteins in microdroplets, which achieved the SERS collections of the single-cell secreted vascular endothelial growth factors.²⁵

Here, we constructed a novel microdroplet SERS-based single-cell detection platform and developed a workflow for single-cell heterogeneity analysis and classification of different BC subtypes. In this study, single cells were encapsulated in microdroplets along with capture probes (immunomagnetic beads, *i*MBs) and reporter probes (immuno-SERS tags, *i*SERS tags). In these microdroplets, we observed a unique interfacial oriented aggregation (IOA) effect of *i*MBs, in which the negatively charged cell membranes tend to attract the electrically neutral capture probes (*i*MBs) to accumulate them on the cell membrane surface spontaneously. The EV-proteins and the *i*SERS tags were sequentially bioconjugated above the *i*MBs to realize sandwich assays. The collection of

*i*SERS tags contributes dramatically to the SERS signal amplification due to the SERS hotspot effect.

Nonmetastatic (MCF-7) and metastatic (MDA-MB-231) cell lines were selected to explore the expression differences of three EV-proteins. CD9 and CD63 belong to the tetraspanin protein family, which are located in the exosomal membrane and play essential roles in cell adhesion and metastasis. FAK protein belongs to the exosomal intramembrane protein and is highly expressed in metastatic breast tumor exosomes.^{4,13} Based on this platform, we explored the levels of three EV-proteins related to metastasis in two BC cell lines and further analyzed and classified single-cell SERS spectra using supervised/unsupervised machine learning algorithms. Based on the single-cell data set, the presence of different feature subpopulations was identified, and a method to distinguish different BC subtypes was established. This study helps reveal the existence of potential subtypes mediated by tumor cell heterogeneity, analyze the association between different variable proteins, and establish classification methods for the diagnosis of BC subtypes.

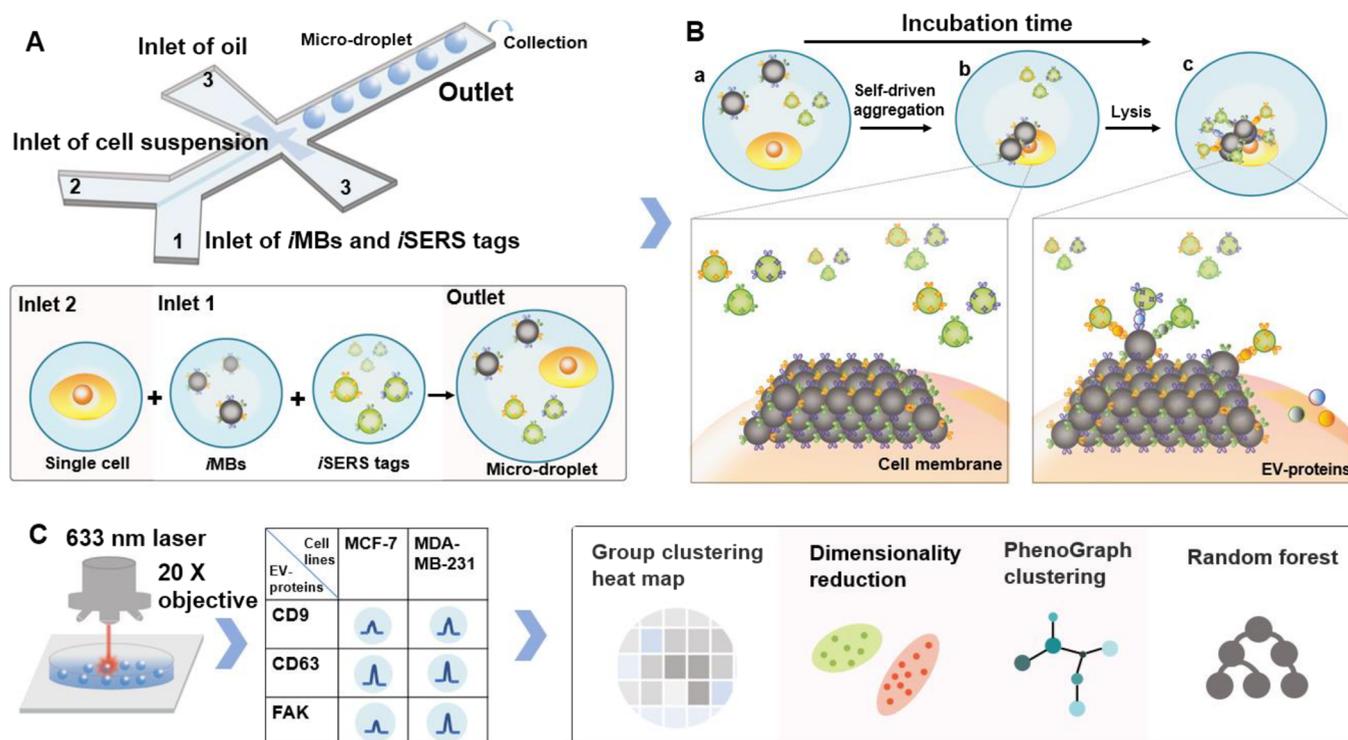
■ MATERIALS AND METHODS

Synthesis of *i*SERS Tags. The preparation process of labeling probes is shown in Figure S1A. Silver nanoparticles were selected as the SERS-active substrate and synthesized by the silver nitrate-sodium citrate reduction.²⁶ AgNPs (3.0 mL) and Tween 20 (2 μ L) were added into glass jars and stirred at room temperature for 30 min. Then, the carboxyl activation reagents, including EDC (1 μ L, 0.05 M) and NHS (1 μ L, 0.05 M), were added to the mixture with stirring for 0.5 h. Two microliters of the probe molecules [4-nitrophenylthiophenol (NTP), 2-mercapto-4-methyl-5-thiazolyl acetic acid (MMTAA), or 3-hydroxythiophenol (HTP), 1.0 mM] and 2 μ L of the monoclonal antibodies (CD63, CD9, or FAK, 0.05 mg/mL) were added to jars containing the activated AgNPs. The mixtures were stirred for 3 h. Finally, BSA (1 μ L, 0.1 mg/mL) was added and kept stirring for 0.5 h to block the unoccupied sites above the Ag nanoparticle surfaces. The mixture was centrifuged at 3310 g for 10 min to remove the supernatant and concentrated to 1.0 mL. Thus, the *i*SERS tags, AgNP@probe@mAb, were obtained.

Synthesis of *i*MBs. The *i*MBs were used as a capture probe in this study. Figure S1B shows the preparation process for the capture probe. Carboxyl paramagnetic nanoparticles (MBs) with a size of about 250 nm were used for preparing *i*MBs, which have a superparamagnetic characteristic and a long suspension time. MBs (diluted to 1 mg/mL) were washed three times with deionized water. Then, EDC and NHS (250 μ L, 0.05 M) were added to the MBs and the mixture was shaken for 0.5 h, followed by deionized water washing twice and magnetic separation. After this, monoclonal antibodies of CD63, CD9, and FAK (13.3 μ L, 0.5 mg/mL, for each) were added to the MBs with 0.8 mL of deionized water for constant volume. The MBs were shaken overnight (12–18 h) and washed twice. Next, BSA (0.8 mL, 10 mg/mL) was added subsequently. After 4 h of shaking and washing twice with deionized water, the *i*MBs were achieved.

Fabrication of Microfluidic Chips and Preparation of Liquid Droplets. The microfluidic chip fabrication process for the microdroplet is presented in Figure S2. The chip was constructed using the soft lithography technique.²⁷ The chip channel pattern was designed using AutoCAD and printed to a high-resolution transparent photomask. Briefly, a 4 inch silicon

Scheme 1. Schematic Design of the Single-Cell Analysis and Classification according to Multiplexed EV-Proteins Using the Microdroplet-Based Self-driven Magnetic SERS Platform Assisted with Machine Learning Algorithmic Tools^a



^a(A) A platform of microdroplet production using a polydimethylsiloxane (PDMS)-based microfluidic chip. Inlet 1: labeling probes (iSERS tags), capturing probes (iMBs), and lysate. Inlet 2: cell suspension. Inlet 3: oil phase. (B) The microdroplet states with the incubation time. (a) A mixture of iSERS tags, iMBs, and a single cell. (b) The iMBs are gathered on the cytomembrane of the target cell. (c) Immunosandwich structures. Reaction process: (1) electrostatic attraction, and (2) lysis. (C) The platform of the single-cell SERS detection in microdroplets with the liquid seal protection, the heterogeneity analysis of single-cell multiplexed proteins data set, and the classification of different cell lines from SERS spectra information, using machine learning algorithms, including group clustering heat map, dimensionality reduction, PhenoGraph clustering, and random forest.

wafer was rinsed with deionized water, ethanol, and acetone, respectively, to clean its surface. The wafer was dried on a 95 °C heating plate for 1 h. Then, the negative photoresist (SU-8) was spun onto the wafer using a spin coater (500 rpm for 10 s, then 1000 rpm for 50 s). The SU-8 layers were then cured on a heating plate at 65 °C for 3 min and at 95 °C for 9 min. The cured SU-8 layers were then exposed to Xenon lamp radiation through a photomask (20 s, 20 mW/cm²) and next baked at 65 °C for 2 min and 95 °C for 7 min. The unexposed SU-8 was removed by soaking in the SU-8 developer for 10 min. The wafer was then cleaned with isopropyl alcohol and dried with filtered nitrogen gas. The silicon wafer with patterns was achieved, which was used as a template for PDMS.

The PDMS base and curing agent were mixed at a ratio of 10:1 w/w. After air bubbles were removed, they were poured onto a wafer template and fully cured for 30 min at 85 °C. After heat curing, the PDMS layer was stripped from the Si wafer. Above the PDMS layer, the entrance and exit holes were made with a 1.2 mm biopsy punch. The PDMS layer was bonded immediately after oxygen plasma treatment (oxygen plasma processor).

As shown in Figure S3, the width of the obtained chip is 50 μm. The chip has two water phases: one is cell suspension with gradient solution and the other is the mixture of capturing probes and labeling probes (v/v: 1:1) with lysate (10 μL). The oil phase consisted of hexadecane and surfactants. Two water phases met and mixed at the fork and flowed in a laminar way

to the intersection. At the crossing, the water and oil phases in the microchannel formed continuous and stable droplets by extrusion and shearing. The solutions were connected to a chip by a syringe, steel needle, and capillary tube, while the injection pump controlled the injection rate. The oil phase flow rate is 850 μL/h, and for both water phases, the flow rate is 50 μL/h. The iSERS tags (9 × 10⁻¹¹ M), iMBs (0.5 mg/mL), and cells (10⁷ cells/mL) with a cell lysis solution were encapsulated in microdroplets by our prepared microfluidic chip.

The probability that the droplet contains only one cell follows the Poisson distribution. As shown in Figure S4, more than half of the droplets contain no cells, and more than 30% of the droplets contain one cell.

Single-Cell EV-Protein SERS Measurements. SERS measurements were performed using a HORIBA ARAMIS confocal Raman spectrometer with a 633 nm laser and a 20× objective lens. The output laser power was 8 mW. A grating scale of 1800 groove/mm was chosen. The SERS spectra were obtained with an integration time of 400 ms.

The SERS spectrum of iSERS tags in DMEM was first recorded and the mineral oil used for preparing microdroplets was assessed.

After 1 h of coculture of iSERS tags, iMBs, and cells, EV-proteins were produced, and they were captured by iMBs. By applying a magnet, the in-drop iMBs could form a line, which is a similar phenomenon to the report of Griffiths and co-workers.²⁸ Interestingly, we observed that the iMBs sponta-

neously collected on the cytomembrane under the electrostatic attraction without applying a magnet. We further performed the in-drop SERS measurement on the *i*MB gathered cell surfaces, which are also the locations of EV-proteins and *i*SERS tags according to immune identification. For comparison, we collected the SERS from the in-drop *i*MB lines applied by a magnetic bar. Differences in SERS signal intensity between the *i*MB lines or the IOA aggregates were compared. The mineral oil liquid sealing was applied to prevent the droplets from drying and breaking, making the immunoassay robust.

In addition, 20 SERS spectra of the single-cell droplet with and without the cell lysis buffer were recorded. The mean spectra were obtained to compare the effect of cell lysis buffer.

Spectral Treatments. Spectral pretreatments were performed to obtain the single-cell data sets (Figure S5). First, batch correction of baseline processing was applied to the SERS spectra. Second, the characteristic peaks of the *i*SERS tags labeling three EV-proteins as the three variable features were extracted into row 1 of the single-cell data set. Characteristic peaks from three Raman reporters were chosen for their labeling proteins: 993 cm^{-1} for FAK, 1288 cm^{-1} for CD9, and 1332 cm^{-1} for CD63. Third, the spectral intensities corresponding to the three characteristic peaks were extracted and normalized, and the normalized values were filled into the corresponding positions in the blank of single-cell data sets. Normalization aims to eliminate the difference in Raman peak intensity and other errors caused by different Raman cross-sections of the three probe molecules.

Several machine learning algorithmic tools were used: the group clustering heat map uses color differences to represent the similarities and differences between different samples. Usually, the result of hierarchical clustering was added. Dimensionality reduction, including principal component analysis (PCA) and *t*-distributed stochastic neighborhood embedding (*t*-SNE), was used. PCA reduces the dimensionality of a data set by recombining variables into principal components (PCs). Biological variables are usually characterized by highly complex relationships, such as polynomials. Nonlinear dimensionality reduction algorithms can capture more information than PCA.^{21,29} *t*-SNE captures nonlinear relationships of local structures and places similar results in the graph to perform different clusters. In the *t*-SNE result plot, the degree of aggregation between different cell lines and the expression of some variables in different cells could be observed. The PhenoGraph algorithm, developed specifically for the study of cancer cell heterogeneity, was developed using a variant of Louvain clustering. The plurality of cells with similarity to the single-cell data set should be clustered into a class.³⁰ Finally, to establish a model that can classify the two types of cell lines, the random forest classification method was used to train the data and test the model to verify the classification accuracy. Data were processed using RStudio.

RESULTS AND DISCUSSION

Workflow of Single-Cell Analysis of Multiplexed EV-Proteins. We performed microfluidic droplet technology to produce water-in-oil microdroplets that envelop single cells to realize single-cell analysis. The microfluidic chip consists of two water-phase inlets and one oil-phase inlet. Cells and probes were encapsulated in microdroplets by regulating the flow rate and cell suspension density, and the cells within the microdroplet were allowed to incubate and lysate without the spatial crosstalk. SERS technology utilized the superposition of

various probe molecules to realize multiplexed detection. The measurement of multiplexed analytes based on single cells could be realized by combining SERS technology. As shown in Scheme 1A, a single-cell suspension was mixed with *i*SERS tags and *i*MBs in a cell lysis solution. After passing through the intersection, the mixture was divided into picoliter droplets by the oil phase. The initial state with *i*SERS tags, *i*MBs, and a single cell in one droplet is shown in Scheme 1B(a). Interestingly, we found that a spontaneous oriented movement of *i*MBs happened in each droplet without applying a magnetic bar, as displayed in Scheme 1B(b). Eventually, a great number of *i*MBs gathered on the cell membrane surface, becoming distinct dark aggregates surrounding the probed cell [Scheme 1B(c)].

Subsequently, EV-proteins were released from the lysed cell and they were captured by the *i*MBs, followed by the reporter probes to form immunosandwich structures. The incubated microdroplets were then subjected to multiplexed SERS detections (Scheme 1C). There are two main advantages to this in-drop SERS detection due to the IOA of *i*MBs. One is that a high SERS sensitivity will be achieved from the collective *i*SERS tags that supply plenty of hotspots on an interface. The other is that the *i*MB aggregates around the probed cell and give a high imaging contrast under bright-field (BF) imaging, which is convenient for laser focusing during the in-drop measurements and makes the single-cell SERS detection more robust. As shown in Scheme 1C, the tetraspanins of CD9, CD63, and FAK were selected as biomarkers for single-cell heterogeneity analysis and classification of two BC cell lines (MCF-7 and MDA-MB-231).

Characterization of *i*MBs and *i*SERS Tags. Our sensing mechanism was based on the in-drop construction of immunosandwich structures for different EV-proteins. Thus, our probes used for this study were first enriched with antibodies, followed by the unoccupied site blocking. The citrate-coated AgNPs with a size of $54.15 \pm 6.28\text{ nm}$ (Figures 1A, S6, and S7A) were chosen to prepare *i*SERS tags due to their high SERS activity. Carboxyl MBs with a size of 200–300 nm were purchased from Zhengzhou Xiyan Biotechnology Co.,

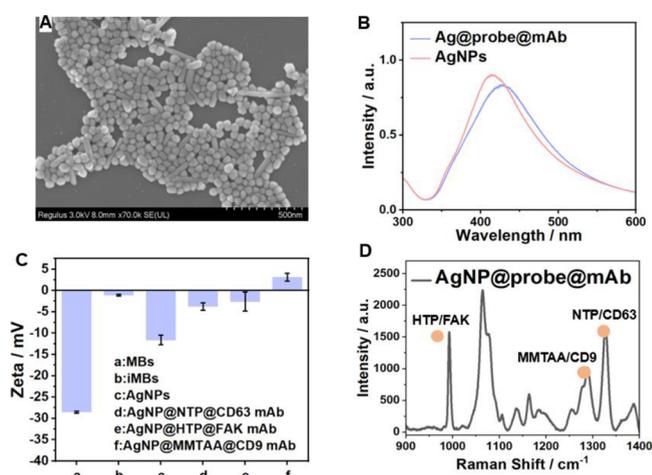


Figure 1. (A) SEM image of AgNPs. (B) UV–vis spectra of AgNPs and AgNP@probe@mAb. (C) Zeta potentials of AgNPs and MBs before and after the monoclonal antibody modification (*i*SERS tags and *i*MBs). (D) SERS spectra of a mixture of three *i*SERS tags. HTP, MMTAA, and NTP represent FAK, CD9, and CD63, respectively. The acquisition time was 2 s.

Ltd. for preparing *i*MBs (Figure S7B,C). UV–vis spectroscopy (Figure 1B) confirmed the antibody modification above the surfaces of AgNPs, owing to the Ag plasmonic band changes. The surface modifications of AgNPs and MBs with antibodies were characterized by zeta potentials. Figure 1C shows that the potential of MBs changes from -28.54 to -1.11 mV, while the potential of AgNPs varies from -11.62 to -3.77 , -2.6 , or 3.1 mV, due to the antibody modification. It should be noted that the *i*MBs display nearly neutral charges in this case. Figure 1D shows the SERS spectrum of a mixture containing three *i*SERS tags, in which HTP (993 cm^{-1} , C–C in-plane bending³¹), MMTAA (1288 cm^{-1} , CH in-plane bending³²), and NTP (1330 cm^{-1} , O–N–O stretching³³) mark FAK, CD9, and CD63 proteins, respectively, and the SERS spectra of each probe are shown in Figure S8. All three Raman reporters have large Raman cross-sections and exhibit distinguishable peaks, which makes these *i*SERS tags applicable for multiplexed detection.

SERS Amplification due to the IOA Effect. In many in-drop immunoassays, additional carriers are needed, such as polymer beads and MBs, which are enveloped with a single cell in one drop to collect proteins and reporters.²⁸ In this SERS sensing system, we adopted MBs due to their controllable collection, which can provide additional SERS amplification from the formation of multiscale SERS hotspots.²³ Interestingly, we found that without a magnetic bar, they can start an IOA behavior, which allows most *i*MBs to gather around the target cell (Figures 2 and S9B), providing identifiable dark aggregates around the probed cell.

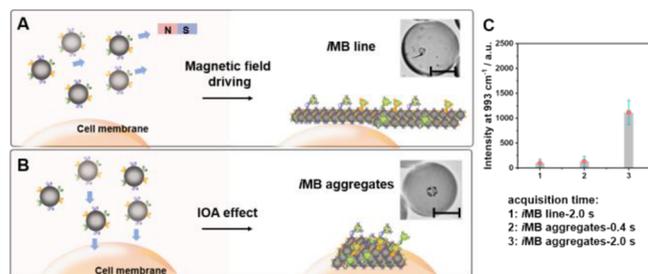


Figure 2. (A, B) Schematic diagram of the magnetic field driving and the IOA effect, and the BF images of the *i*MB line or *i*MB aggregates on the cytomembrane in a single microdroplet. The scale bars are $20\ \mu\text{m}$. (C) Comparison of SERS signals of the immunosandwich structures in droplets, enhanced by a magnetic action-driven *i*MB line (1) or *i*MB aggregates formed under two different incubation periods, 20 spectra, respectively (1 and 2).

We conjecture that the electrostatic attraction between the neutrally charged *i*MBs and the negatively charged cytomembrane dominates this self-driven phenomenon.^{34,35} The aggregates of *i*MBs at the biological interface are caused by electrostatic attraction, and the binding dynamics of this process can be explained by the Langmuir adsorption equation.^{36,37} The negatively charged part of the cell surface can absorb electrically neutral *i*MBs through electrostatic interactions.

In the IOA effect, the neutral magnetic nanoparticles are driven toward the negatively charged cell membrane by electrostatic attraction and eventually accumulate on the surface of the cell membrane, while this effect will fail under the disturbance of external forces. When a magnet was applied

around the microdroplet, a noticeable MB line was formed inside the microdroplet (Figures 2A and S9A).

During the sensing step, the *i*MBs will capture the EV-proteins, followed by the recognition of *i*SERS tags, the immunosandwich structures. When focusing a laser on the *i*MB aggregates, the SERS signals from the *i*SERS tags will be obtained. A comparison of SERS recorded above the *i*MB line and IOA aggregates was carried out, and the results are displayed in Figure 2C. It can be found that the IOA aggregates exhibit higher SERS intensity than the *i*MB lines. Such strong SERS enhancement can be attributed to the more concentrated, multiscale hotspots formed in IOA aggregates, involving nanogaps between the *i*SERS tags and the *i*MBs, or among *i*SERS tags. For the *i*MB lines, nanogaps are dispersed along the line, which leads to a relatively mild enhancement activity for SERS.

During the SERS measurement, we can easily focus laser spots on the collective *i*MBs in each microdroplet due to their high imaging contrast. Differently, we found that the *i*MB line was quickly destroyed by laser irradiation when recording SERS spectra, but the IOA aggregate exhibits relatively higher stability. Therefore, the subsequent experiments were carried out in the IOA system, and the integration time was set as 0.4 s to avoid photodamage.

The IOA of *i*MBs can be observed under the BF imaging. As shown in Figure 3A, after a single cell mixed with *i*MBs, *i*SERS

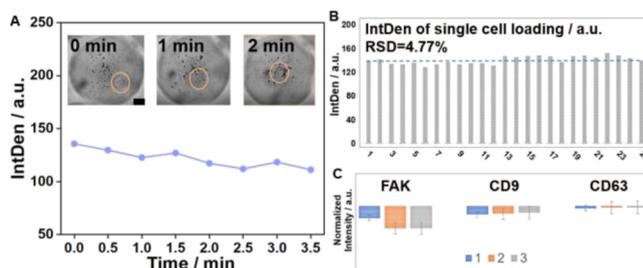


Figure 3. (A) Loading time of the *i*MBs self-driving to the cell membrane. The scale bar is $20\ \mu\text{m}$. (B) IntDen of single-cell loading. (C) Reproducibility of three experiments.

tags, and lysate, the *i*MBs would be loaded onto the cell surface within 5 min. At the same time, the lysate-stimulated cell starts to release EV-proteins. Within the following 1 h, the *i*MBs captured these EV-protein, followed by the *i*SERS tag fixing, to form immunosandwich structures.

The reproducibility of the *i*MB loading above the cell surface was verified. First, we counted the number of *i*MBs loaded on the surface of 25 single cells by converting the BF images by ImageJ software, since it determined the subsequent capture of exosome proteins and SERS labels. Figure 3B shows that the number of *i*MBs loaded on a single cell fluctuated around 140, with a relative standard deviation of 4.77%. Subsequently, three repeated experiments were conducted on the level test of three proteins in a single cell, as shown in Figure 3C, and the results showed good repeatability.

Single-Cell Heterogeneity Analysis by SERS. By using the microdroplet SERS platform, we started the cell heterogeneity analysis of two kinds of BC cells according to their EV-proteins. Before in-drop SERS recording of immunoassays, DMEM and mineral oil used for preparing microdroplets were measured. Figure S10 shows the SERS spectra of *i*SERS tags in DMEM, demonstrating that DMEM brings no

additional peaks that overlap with the Raman bands of reporters. Also, mineral oil indicates no interference with the spectral profiles of Raman reporters.

The single-cell heterogeneity analysis started with the in-drop SERS measurements of *i*SERS tags collected by *i*MBs above the cell surface. The cell was cocultured with lysate to assist the EV-protein release. It should be noted that the cell lysis buffer dose was set as 10 μ L to keep the cell morphology integrity. The effect of lysate for the in-drop single-cell assay was also assessed. Figure S11 shows that the mean SERS spectra from single-cell droplets with lysis and without lysis display different natures, which indicates that the lysate is conducive to the release of EV-proteins that can be bridges for *i*SERS tags.

After this, spectral pretreatments were performed to obtain the single-cell data sets, followed by machine learning algorithmic tools. First, we performed the violin plots to analyze the mean levels of three EV-proteins based on two types of cells. Figure 4A shows the violin plots of the three EV-

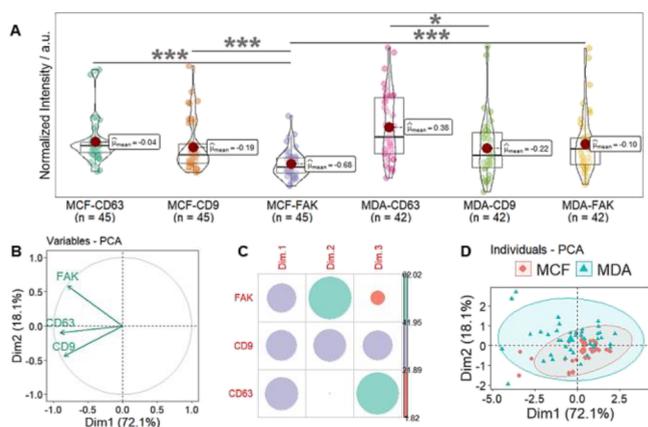


Figure 4. (A) Violin plots of the expression levels of three EV-proteins from MCF-7 and MDA-MB-231 cells. The vertical coordinate was a normalized value. (B) Variable correlation plot. (C) Heat map of the contribution of variables to the principal component. (D) PCA plot, the pink points represent the MCF-7 cell line and the blue points represent the MDA-MB-231 cell line.

protein levels of MCF-7 and MDA-MB-231 cell lines with the normalized levels on the vertical axis. The median value in the box (bright red dot) represents the overall situation (mean) of the single-cell data, the scatter reflects the protein distribution of individual cells, the violin waveform represents the degree of concentration, and the box length of the violin shows the degree of dispersion of the single-cell data. The *p* value analysis shows a significant difference in the secretion of CD63/CD9 and FAK proteins in MCF-7 cells (*p* value < 0.05), and FAK levels show significant differences between MDA-MB-231 and MCF-7 cells. No significant differences are observed between other proteins. Moreover, CD63 and CD9 show heterogeneous distributions among single cells, and the presence of multiple peaks implies the existence of different subpopulations, and these conditions may be related to the heterogeneous distribution of tetraspanins in the exosomes of individual cells.³⁸

We performed the PCA method to evaluate the contribution of the three variables to the overall difference of the single-cell data set. The PCA results are shown in Figure 4B–D. The interpretation degree of PC 1 is 72.1%, and the interpretation

degree of PC 2 is 18.1%. The variable correlation plots reveal the correlation between the three exosomal proteins (Figure 4B), which helps find subpopulations that exhibit functional similarity. The weakest correlation lies between CD9 and FAK, indicating a lack of interaction between two proteins.³⁹ There was some positive correlation between CD9 and CD63, indicating that their expressions are interrelated.⁴⁰ These facts emphasize the presence of a ^{CD9+CD63+}EV subpopulation.⁴¹ In addition, we noticed that the difference between CD9 and CD63 may be attributed to different intracellular origins, that is, CD9 originates mainly from the plasma membrane, and a small number of EVs carry CD9 in the endosomal machinery, whereas CD63 is abundantly enriched in late endosomes.⁴² The heat map of the contribution of variables to the principal component measured the importance of the original variables to the PC (Figure 4C). FAK and CD9 have a strong correlation for both PC1 (Dim.1) and PC2 (Dim.2) and need to be prioritized in explaining the differences between these two cell lines. In particular, results highlight the involvement of FAK and the signaling events of FAK leading to the invasion and metastasis of BC.⁴³ This suggests that FAK proteins may be a potential marker in distinguishing between different subtypes of BC (metastatic vs nonmetastatic). The scatter plot based on PCA analysis (Figure 4D) divides two cell lines according to the concentrated ellipse, which provides poor dimensionality reduction due to the weak correlation between the original variables and the linear dimensionality of PCA.

To further assess the probability of heterogeneity in the single-cell data set, we performed clustered heat maps, *t*-SNE, and PhenoGraph clustering to find potential subgroups. The subgroup clustering heat map reflects the differences and similarities between samples through color changes and dendrogram classification. Differences in single-cell protein levels are noticed by color depths, with green and purple representing high and low abundance. The dendrogram was obtained by clustering based on the similarity of abundance. Figures S12 and 5A show that the single-cell clustering heat map and the population cell clustering heat map cluster the data set into two subpopulations. The statistical results at the single-cell level and the average level of the population cells show that the levels of three EV-proteins are significantly different between MCF-7 and MDA-MB-231 cells, where, regarding the levels of tetraspanins, the CD9 level of MCF-7 is higher than that of MDA-MB-231. The CD63 level of MDA-MB-231 is also higher than that of CD63 levels of MCF-7. There was no significant difference between CD63 and CD9 in MCF-7 cells. The CD63 levels of MDA-MB-231 are much higher than the CD9 levels. The results are consistent with other EV-protein studies.⁴⁴ Overall, CD63 levels are highly expressed in both malignant BC subtypes, which is consistent with previous reports.⁴⁵ FAK is more frequently found in metastatic BC subtypes, and the overexpression of FAK highlights their role in BC invasion and metastasis.⁴³

To further validate the heterogeneous distribution characteristics of the single-cell data set, the data set was dimensionally reduced. The *t*-SNE method takes the data on the relative secretion levels of the three EV-proteins in each cell line group as variables and projects them onto a two-dimensional plane, where individuals with high similarity start clustering. The cell colors encode the cell line types (Figure 5C) and their EV-protein levels (Figure 5D–F). As seen in the *t*-SNE plots, unlike the PCA results, two cell types can be well distinguished. The distribution characteristics of MCF-7 cells

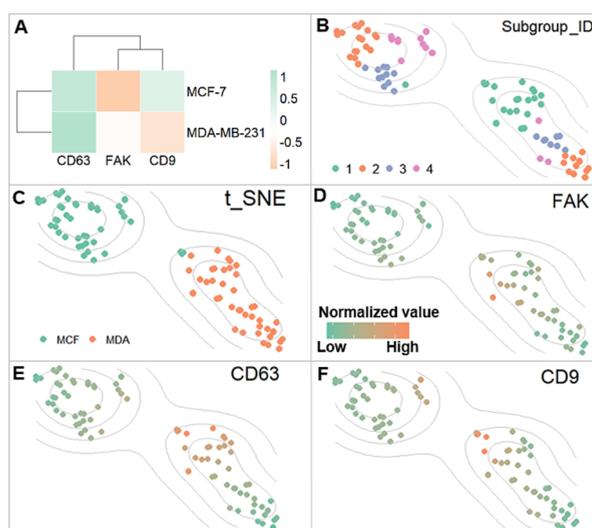


Figure 5. (A) Group clustering heat map of population cell data, clustering by rows. (B) PhenoGraph clustering resulted from single-cell data sets in the coordinate of *t*-SNE, colored by four subgroups. (C) *t*-SNE plot, each point represents a single cell, turquoise points are MCF-7 cells and pink points are MDA-MB-231 cells. (D, E, and F) EV-protein of FAK, CD63, and CD9 levels in the two cell lines in the coordinate of *t*-SNE.

are different from those of MDA-MB-231 cells, which are more dispersed overall. Figure 5D–F directly shows the horizontal distributions of three proteins, and it can be seen that some subpopulations are similar in the expression levels of the three proteins.

Furthermore, cluster analysis was performed using the single-cell data set. Unlike the clustering results of the single-cell heat map, as shown in Figure 5B, the single-cell data sets of two cell lines are divided into four subgroups by PhenoGraph clustering, in which, in subgroup 1, the levels of all three proteins are at high levels in single cells (Figure 5F–D) and most of them are derived from MDA-MB-231 cells. The other subgroups (2, 3, and 4) distribute in the coordinates of both cell lines, suggesting that the exosome protein level may be close in two cell lines. Among them, CD9, CD63, and FAK in subgroup 3 show moderate levels, while they present low levels in subgroup 2. In group 4, CD63 and CD9 are at moderate levels, and FAK is at low levels. The heterogeneity of EV-protein explained these differences in protein levels between subgroups.⁴⁶

Classification of Two BC Cell Lines. We constructed classifiers for distinguishing different subtypes of BC cell lines. As shown in Figure 6A, 500 decision trees are generated by randomly selecting 70 observations by the random forest algorithm and 1 variable from the classification nodes. The error rate of the classifier is 18.57%, in which 88.89% of MCF-7 cells and 73.53% of MDA-MB-231 cells are correctly classified. The constructed classifier was tested on the training and test set samples, as shown in Figure 6B,C. The random forest classifier could classify the training set samples accurately, both 100%, and the classification accuracy of the test set samples are 100% (MCF-7 cells) and 87.5% (MDA-MB-231 cells), respectively. In Figure 6D, the ROC curve was used for evaluating the accuracy of the classifier with an AUC (area under the curve) value of 0.938, which is close to 1, indicating the good performance of this classifier. To assess the importance of each protein in the classification model, the

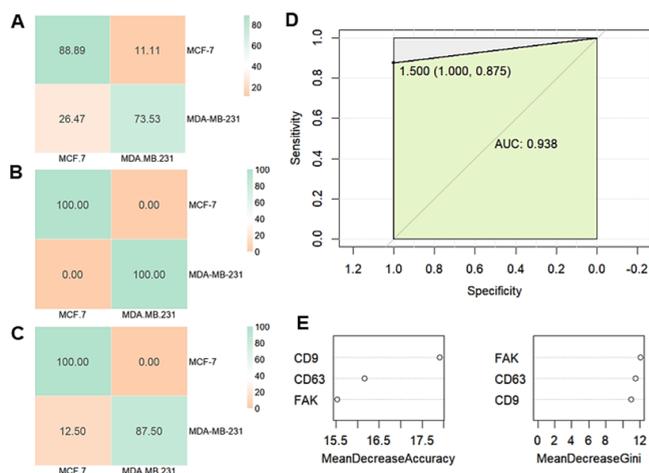


Figure 6. (A) Random forest error rate matrix. (B) Classifier tested with the training set. (C) Classifier tested with the test set, the units are in %. The training set and a test set are randomly divided by the “createDataPartition” function in R language, setting 80% as the training set and 20% as the test set. (D) Dichotomous receiver operator characteristic (ROC) curve of a classifier. (E) Importance rankings of three variables using the MDA and MDG index.

variables were characterized using MDA (mean decrease accuracy) and MDG (mean decrease Gini), and CD9 proteins show critical importance with FAK proteins, as shown in Figure 6E, which are consistent with the PCA results. In addition, we can observe that CD63 is not distinguishable between the two cell lines, which may attribute to the fact that CD63 expression is comparable in all malignant cells.⁴² In conclusion, with the random forest model, key EV-proteins could be screened as biomarkers to distinguish between these two BC cell lines.

CONCLUSIONS

In summary, we established a microfluidic droplet-based multiplexed SERS method to analyze EV-proteins at the single-cell level. The key to this method is the IOA effect, which can collect the target protein as well as the *i*SERS tags by converging *i*MBs on the probed cell membrane surface, enabling the enrichment of dispersed proteins to the detection spots, while allowing signal amplification by forming hotspots to improve the in-drop detection sensitivity. The IOA process was easy to prepare samples and fast to detect. Using this technology, we studied three EV-proteins from two BC cell lines and the results evidenced that this IOA-based platform can trace the EV-proteins of an individual cell. We also introduced machine learning algorithms to discover the presence of different subpopulations in the single-cell data set and to analyze the association between different proteins. Based on the single-cell data set, different performance characteristics of different subpopulations were found, confirming the presence of heterogeneity in single cells. Furthermore, we validated FAK and CD9 to be key biomarkers in distinguishing two BC cell lines. In addition to the detection of different EV-proteins presented in this protocol, the use of this platform also allows metabolic analysis at the single-cell level as well as drug analysis through probe molecule–analyte interactions.

■ ASSOCIATED CONTENT

SI Supporting Information

The Supporting Information is available free of charge at <https://pubs.acs.org/doi/10.1021/acs.analchem.3c01273>.

Additional experimental details including reagents, cell culture, and preparation of cell suspension, characterizations, and processing images by ImageJ; the schematic diagram of the preparation process of labeling probes and capturing probes, the microfluidic chip fabrication process, the droplet formation chip pattern, single-cell data set preprocessing procedures; experimental data involve the statistics of the probability of droplet encapsulation, the size distribution of AgNPs counted from the SEM image, the size distribution of AgNPs and MBs by the nanoparticle and zeta potential analyzer, the SERS spectra of each probe, the microscopic images of the IOA aggregates on the cytomembrane in the microdroplet and the iMB lines in the microdroplets; SERS spectra of AgNP@probe@mAb in DMEM and mineral oil in the microdroplets, mean SERS spectra of the single-cell droplet with and without lysis, and group clustering heat map of single-cell data sets (PDF)

■ AUTHOR INFORMATION

Corresponding Author

Shuping Xu – State Key Laboratory of Supramolecular Structure and Materials, College of Chemistry, Center for Supramolecular Chemical Biology, College of Chemistry, and Institute of Theoretical Chemistry, Jilin University, Changchun 130012, P. R. China; State Key Laboratory of Applied Optics, Changchun Institute of Optics, Fine Mechanics and Physics, Chinese Academy of Sciences, Changchun 130033, P. R. China; orcid.org/0000-0002-6216-6175; Email: xusp@jlu.edu.cn

Authors

Jiaqi Wang – State Key Laboratory of Supramolecular Structure and Materials, College of Chemistry, Jilin University, Changchun 130012, P. R. China

Lili Cong – State Key Laboratory of Supramolecular Structure and Materials, College of Chemistry, Jilin University, Changchun 130012, P. R. China

Wei Shi – Key Lab for Molecular Enzymology & Engineering of Ministry of Education, Jilin University, Changchun 130012, P. R. China

Weiqing Xu – State Key Laboratory of Supramolecular Structure and Materials, College of Chemistry and Institute of Theoretical Chemistry, Jilin University, Changchun 130012, P. R. China; orcid.org/0000-0002-1947-317X

Complete contact information is available at:

<https://pubs.acs.org/doi/10.1021/acs.analchem.3c01273>

Notes

The authors declare no competing financial interest.

■ ACKNOWLEDGMENTS

This work was supported by the National Natural Science Foundation of China NSFC (Nos. 21873039, 21827805, and 22173035), Science and Technology Development Program Projects of Jilin Province (20220101046JC), Opening Project of the State Key Laboratory of Applied Optics

(SKLAO2021001A14), and Interdisciplinary Integration Innovation Project of Jilin University (JLUXKJC2020106).

■ REFERENCES

- (1) Tan, Y.; Luo, X.; Lv, W.; Hu, W.; Zhao, C.; Xiong, M.; Yi, Y.; Wang, D.; Wang, Y.; Wang, H.; Wu, Y.; Zhang, Q. *Cell Death Dis.* **2021**, *12*, 547.
- (2) Liang, Y.; Zhang, H.; Song, X.; Yang, Q. *Semin. Cancer Biol.* **2020**, *60*, 14–27.
- (3) Dong, M.; Liu, Q.; Xu, Y.; Zhang, Q. *Front. Cell Dev. Biol.* **2022**, *10*, No. 842898.
- (4) Galindo-Hernandez, O.; Villegas-Comonfort, S.; Candanedo, F.; González-Vázquez, M. C.; Chavez-Ocaña, S.; Jimenez-Villanueva, X.; Sierra-Martinez, M.; Salazar, E. P. *Arch. Med. Res.* **2013**, *44*, 208–214.
- (5) Meng, L.; Song, K.; Li, S.; Kang, Y. *Membranes* **2022**, *12*, 775.
- (6) 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. *Cell* **2019**, *177*, 428.e18–445.e18.
- (7) Willms, E.; Cabañas, C.; Mäger, I.; Wood, M.; Vader, P. *Front. Immunol.* **2018**, *30*, 738.
- (8) Andreu, Z.; Yáñez-Mó, M. *Front. Immunol.* **2014**, *5*, 442.
- (9) Kummer, D.; Steinbacher, T.; Schwietzer, M. F.; Thölmann, S.; Ebneth, K. *Med. Microbiol. Immunol.* **2020**, *209*, 397–405.
- (10) Termini, C. M.; Gillette, J. M. *Front. Cell Dev. Biol.* **2017**, *5*, 34.
- (11) Zhou, J.; Yi, Q.; Tang, L. J. *Exp. Clin. Cancer Res.* **2019**, *38*, 250.
- (12) Luo, M.; Guan, J. L. *Cancer Lett.* **2010**, *289*, 127–139.
- (13) Vinik, Y.; Ortega, F. G.; Mills, G. B.; Lu, Y.; Jurkiewicz, M.; Halperin, S.; Aharoni, M.; Gutman, M.; Lev, S. *Sci. Adv.* **2020**, *6*, No. eaba5714.
- (14) Tian, F.; Zhang, S.; Liu, C.; Han, Z.; Liu, Y.; Deng, J.; Li, Y.; Wu, X.; Cai, L.; Qin, L.; Chen, Q.; Yuan, Y.; Liu, Y.; Cong, Y.; Ding, B.; Jiang, Z.; Sun, J. *Nat. Commun.* **2021**, *12*, 2536.
- (15) Wang, Wuethrich, A.; Sina, A. A.; Lane, R. E.; Lin, L. L.; Wang, Y.; Cebon, J.; Behren, A.; Trau, M. *Sci. Adv.* **2020**, *6*, No. eaax3223.
- (16) Chin, L. K.; Son, T.; Hong, J. S.; Liu, A. Q.; Skog, J.; Castro, C. M.; Weissleder, R.; Lee, H.; Im, H. *ACS Nano* **2020**, *14*, 14528–14548.
- (17) Moura, S. L.; Martín, C. G.; Martí, M.; Pividori, M. I. *Biosens. Bioelectron.* **2020**, *150*, No. 111882.
- (18) Yue, S.; Fang, J.; Xu, Z. *Biosens. Bioelectron.* **2022**, *198*, No. 113822.
- (19) Keyes, T. J.; Domizi, P.; Lo, Y. C.; Nolan, G. P.; Davis, K. L. *Cytometry, Part A* **2020**, *97*, 782–799.
- (20) Kourou, K.; Exarchos, T. P.; Exarchos, K. P.; Karamouzis, M. V.; Fotiadis, D. I. *Comput. Struct. Biotechnol. J.* **2014**, *13*, 8–17.
- (21) Weber, L. M.; Robinson, M. D. *Cytometry, Part A* **2016**, *89*, 1084–1096.
- (22) Shen, Y.; Yue, J.; Xu, W.; Xu, S. *Theranostics* **2021**, *11*, 4872–4893.
- (23) Sun, D.; Cao, F.; Xu, W.; Chen, Q.; Shi, W.; Xu, S. *Anal. Chem.* **2019**, *91*, 2551–2558.
- (24) Zong, C.; Xu, M.; Xu, L. J.; Wei, T.; Ma, X.; Zheng, X. S.; Hu, R.; Ren, B. *Chem. Rev.* **2018**, *118*, 4946–4980.
- (25) Cong, L.; Wang, J.; Li, X.; Tian, Y.; Xu, S.; Liang, C.; Xu, W.; Wang, W.; Xu, S. *Anal. Chem.* **2022**, *94*, 10375–10383.
- (26) Lee, P. C.; Meisel, D. *J. Phys. Chem.* **1982**, *86*, 3391–3395.
- (27) Shakeri, A.; Khan, S.; Didar, T. F. *Lab Chip* **2021**, *21*, 3053–3075.
- (28) Eyer, K.; Doineau, R. C. L.; Castrillon, C. E.; Briseño-Roa, L.; Menrath, V.; Mottet, G.; England, P.; Godina, A.; Briant-Litzler, E.; Nizak, C.; Jensen, A.; Griffiths, A. D.; Bibette, J.; Bruhns, P.; Baudry, J. *Nat. Biotechnol.* **2017**, *35*, 977–982.
- (29) Cheng, H.; Li, Z.; Guo, Z.; Shao, S.; Mo, L.; Wei, W.; Xue, M. *Biosens. Bioelectron.* **2021**, *190*, No. 113368.
- (30) Levine, J. H.; Simonds, E. F.; Bendall, S. C.; Davis, K. L.; Amir, E.; Tadmor, M. D.; Litvin, O.; Fienberg, H. G.; Jager, A.; Zunder, E. R.; Finck, R.; Gedman, A. L.; Radtke, I.; Downing, J. R.; Pe'er, D.; Nolan, G. P. *Cell* **2015**, *162*, 184–197.

- (31) Gu, X.; Wang, H.; Schultz, Z. D.; Camden, J. P. *Anal. Chem.* **2016**, *88*, 7191–7197.
- (32) Li, J.; Wuethrich, A.; Sina, A.; Cheng, H. H.; Wang, Y.; Behren, A.; Mainwaring, P. N.; Trau, M. *Nat. Commun.* **2021**, *12*, 1087.
- (33) van Schrojenstein Lantman, E. M.; Gijzeman, O. L.; Mank, A. J.; Weckhuysen, B. M. *ChemCatChem* **2014**, *6*, 3342–3346.
- (34) Chen, B.; Le, W.; Wang, Y.; Li, Z.; Wang, D.; Ren, L.; Lin, L.; Cui, S.; Hu, J. J.; Hu, Y.; Yang, P.; Ewing, R. C.; Shi, D.; Cui, Z. *Theranostics* **2016**, *6*, 1887–1898.
- (35) Alves, A. C.; Ribeiro, D.; Nunes, C.; Reis, S. *Biochim. Biophys. Acta, Biomembr.* **2016**, *1858*, 2231–2244.
- (36) Lesniak, A.; Salvati, A.; Santos-Martinez, M. J.; Radomski, M. W.; Dawson, K. A.; Åberg, C. *J. Am. Chem. Soc.* **2013**, *135*, 1438–1444.
- (37) Wilhelm, C.; Gazeau, F.; Roger, J.; Pons, J. N.; Bacri, J.-C. *Langmuir* **2002**, *18*, 8148–8155.
- (38) Mizenko, R. R.; Brostoff, T.; Rojalin, T.; Koster, H. J.; Swindell, H. S.; Leiserowitz, G. S.; Wang, A.; Carney, R. P. *J. Nanobiotechnol.* **2021**, *19*, 250.
- (39) Steelant, W. F.; Kawakami, Y.; Ito, A.; Handa, K.; Bruyneel, E. A.; Mareel, M.; Hakomori, S. *FEBS Lett.* **2002**, *531*, 93–98.
- (40) Cavallaro, S.; Pevere, F.; Stridfeldt, F.; Görgens, A.; Paba, C.; Sahu, S. S.; Mamand, D. R.; Gupta, D.; El Andaloussi, S.; Linnros, J.; Dev, A. *Small* **2021**, *17*, No. e2008155.
- (41) Ji, Y.; Qi, D.; Li, L.; Su, H.; Li, X.; Luo, Y.; Sun, B.; Zhang, F.; Lin, B.; Liu, T.; Lu, Y. *Proc. Natl. Acad. Sci. U. S. A.* **2019**, *116*, 5979–5984.
- (42) Kowal, J.; Arras, G.; Colombo, M.; Jouve, M.; Morath, J. P.; Primdal-Bengtson, B.; Dingli, F.; Loew, D.; Tkach, M.; Théry, C. *Proc. Natl. Acad. Sci. U. S. A.* **2016**, *113*, E968–E977.
- (43) Shtam, T.; Naryzhny, S.; Samsonov, R.; Karasik, D.; Mizgirev, I.; Kopylov, A.; Petrenko, E.; Zbrodskaya, Y.; Kamyshinsky, R.; Nikitin, D.; Sorokin, M.; Buzdin, A.; Gil-Henn, H.; Malek, A. *Breast Cancer Res. Treat.* **2019**, *174*, 129–141.
- (44) Kugeratski, F. G.; Hodge, K.; Lilla, S.; McAndrews, K. M.; Zhou, X.; Hwang, R. F.; Zanivan, S.; Kalluri, R. *Nat. Cell Biol.* **2021**, *23*, 631–641.
- (45) Yoshioka, Y.; Konishi, Y.; Kosaka, N.; Katsuda, T.; Kato, T.; Ochiya, T. *J. Extracell. Vesicles* **2013**, *2*, 20424.
- (46) Morales, R. T.; Ko, J. *ACS Nano* **2022**, *16*, 11619–11645.