

Supporting Information

Combined morpho-chemical profiling of individual extracellular vesicles and functional nanoparticles without labels

Yichuan Dai^{1,#}, *Suwen Bai*^{2,#}, *Chuanzhen Hu*¹, *Kaiqin Chu*^{1,3}, *Bing Shen*^{2*} and *Zachary J. Smith*^{1*}

1. Key Laboratory of Precision Scientific Instrumentation of Anhui Higher Education Institutes, Dept. of Precision Machinery and Precision Instrumentation, University of Science and Technology of China
2. Dept. of Basic Medicine Anhui Medical University
3. Hefei National Laboratory for Physical Sciences at the Microscale, University of Science and Technology of China, Hefei, Anhui, China.

* Correspondence to author: Bing Shen, E-mail: shenbing@ahmu.edu.cn.
Zachary J. Smith, E-mail: zsmith@ustc.edu.cn.

Table of content

1. Construction of R-NTA system.....	- 2 -
2. Spectral data processing and analysis.....	- 2 -
3. R-NTA quantitative chemical measurement.....	- 2 -
4. R-NTA sizing theory and method.....	- 3 -
5. Optimizing Size estimation	- 4 -
6. Mie theory.....	- 6 -
7. Precision of RI measurement.....	- 7 -
8. Precision of quantitative Raman measurement	- 8 -
10. R-NTA analysis of liposomes with additional extrusion.....	- 9 -
11. EVs sample preparation and purification.....	- 10 -
12. Western Blotting of cells and SEC-isolated EV samples.....	- 10 -
13. Stability of Raman measurements of single EVs over measurement period	- 11 -
14. Pure component spectra	- 11 -
15. Principal components spectra and fittings with pure component spectra.....	- 12 -
16. Hierarchical clustering analysis of EVs	- 13 -
17. EVs clustering by Raman information only	- 14 -
18. Biochemical concentration analysis of EV spectra	- 14 -
Table S1 EVs Spectral Interpretation	- 15 -
Table S2 Fitting coefficients of each cluster	- 15 -
References	- 16 -

1. Construction of R-NTA system

A 785 nm single-mode continuous wave laser with controllable output power (Sacher Lasertechnik Serval Plus, Marburg, Germany) is used to create a tightly confined laser trap via a high NA water immersion objective (60x / NA = 1.27, Nikon Plan Apo SR IR). A bandpass filter (LL01-785-12.5, Semrock, Rochester, USA) is used to ensure monochromatic emission of the laser source. Backward scattered light from trapped particles is separated into two channels: elastic and Raman scattering. Elastically scattered light is separated by means of a 5% laser pickoff beamsplitter (BSF10-B, Thorlabs, New Jersey, USA), and directed towards a high-speed sCMOS detector (zyla 4.2, Andor, Belfast, UK) placed in a conjugate plane to the sample plane. The images recorded by the detector are used to monitor the motion of trapped particles. An $f = 500$ mm lens (AC508-500-B-ML, Thorlabs, New Jersey, USA) is coupled with the camera and the pixel to micron conversion factor of the s-CMOS image is 0.0443 micron/pixel. Inelastic Raman scattering is separated from the elastic scattering through the use of a dichroic beamsplitter (LPD02-785RU-25x, Semrock, Rochester, USA) and a long-pass filter (LP02-785RU-25, Semrock, Rochester, USA). The Raman scattered light is focused by a $f = 75$ mm lens (AC254-75-B-ML, Thorlabs, New Jersey, USA) and collected via a 105 μm -core multimode fiber acting as a confocal pinhole. The fiber is attached to a 150 lp/mm grating spectrometer (Shamrock 500i, Andor, UK) and spectra were acquired using a thermoelectrically cooled back-illuminated CCD camera (iDus DU401-DD, Andor, Belfast, UK) with a spectral resolution of 14.1 cm^{-1} and spectral range from 330 to 3950 cm^{-1} . Bright field image of R-NTA was captured via an CCD camera (DCU224M, Thorlabs, New Jersey, USA).

2. Spectral data processing and analysis

All Raman spectral data processing was performed in Matlab (R2018a). 5 frames of 30s data acquisition were acquired for each nanoparticle. The 5 frames from each nanoparticle were averaged for an equivalent integration time of 2.5 min per nanoparticle. All raw spectra data were corrected for cosmic rays using a median threshold filter. The spectra were smoothed using the Whittaker smoother with a Lagrange parameter of 2^1 . Then the spectra were background corrected to remove contributions from quartz, PBS buffer/DI water and a polynomial term to account for auto-fluorescence, using asymmetric least squares fitting (AsLS) using an asymmetry parameter of $p=0.01$. Coefficients for each component in NIST standard particles and liposomes were determined using general least squares fitting, and for that in EVs were determined by using the AsLS algorithm, with $p=0.01$. For comparison of size-normalized and vector-normalized Raman spectra shown in the Supplemental Material, the wavenumber range for vector normalization was $600\text{-}1800 \text{ cm}^{-1}$ and $2600\text{-}3100 \text{ cm}^{-1}$. Principal components analysis was performed using the built-in MATLAB function `pca`. Hierarchical clustering analysis is used for creating linkages based on Euclidean pairwise distances².

3. R-NTA quantitative chemical measurement

In R-NTA, a confocal detection setup collects Raman scattering only from a precise focal

volume, allowing the study of individual nanoparticle composition. Once Raman spectrum of a nanoparticle is acquired, the number of molecules represented by the Raman spectra could be quantitatively extracted by the ASLS fit coefficients and the number of molecules represented in the pure spectra used in the ASLS fitting process. Given this, the amount of chemical being represented by pure spectra could be quantified by:

$$N = \frac{V_{fv}(ml) \times \rho(g/ml)}{M(g/mol)} \quad \text{E. S1}$$

Where ρ and M represents the density and molecular weight of pure components respectively. V_{fv} represents the unchanged focal volume in LTRS. Since the focal volume in LTRS is unchanged during measurements, an accurate measure of the focal volume V_{fv} could be determined by comparing Raman intensity of small nanoparticles and relatively large particles (occupying the full focal volume) made of the same material, i.e. 100 nm and 5 micron polystyrene beads. Thus, the focal volume could be written as $V_{fv} = I_{ratio} \times V_{PS100}$, where V_{PS100} represents the volume of 100 nm sphere, and I_{ratio} is the intensity ratio between two Raman spectra. The absolute concentration within the nanoparticle, then, can be simply calculated by taking the number of molecules of the analyte N and dividing by the size of the nanoparticle, as determined below.

4. R-NTA sizing theory and method

In optical tweezers, the gradient force acts as a harmonic optical restoring force that constrains the particle motion at long time scales. Thus, the traditional NTA sizing technique, which depends on the Stokes-Einstein equations governing Brownian motion operating at relatively long time-scales, fails in this constrained motion model. In fact, the probability density function of the constrained nanoparticle position over time could be described by the Smoluchowski equation³.

$$\frac{\partial}{\partial t} P(x,t|r_0,t_0) = \nabla \cdot D[\nabla - \beta F(r)]P(x,t|r_0,t_0) \quad \text{E. S2}$$

Where D is the diffusion coefficient, $\beta = 1/k_b T$, and $F(r)$ is the force applied on the particle.

The analytical solution to this partial differential equation is a Gaussian function with center and variance depending on trapping stiffness k_{trap} , diffusive coefficient D and time lag Δt , the former two parameters both depend on the particle size. The variance of the probability distribution versus time-lag is given by Equation S3⁴:

$$\sigma^2(\Delta t) = \frac{k_b T}{k_{trap}} \left[1 - \exp\left(-\frac{2(k_{trap} D)\Delta t}{k_b T}\right) \right] \quad \text{E. S3}$$

The dynamic of constrained nanoparticles could be divided into three different domains (trap stiffness dominated, intermediate, and diffusion regimes) based on the relation between the time lag and relaxation time ($\tau = 3\pi\eta d / k_{trap}$)^{3,4}. When the time lag is very short and optical stiffness k_{trap} , which is inversely proportional to relaxation time, is weak, the influence caused by the optical force is negligible, and the diffusion dominates the nanoparticle's motion. Following O'Dell et al. for diffusion-dominated Brownian motion⁵, the variance of displacement in these short time lags is equal to:

$$\sigma^2(\Delta t) \approx 2D\Delta t \quad \text{E. S4}$$

by taking the first-order Taylor approximation of E.S(4), as expected for normal diffusion. Thus, unlike traditional NTA, where particles are tracked using relatively slow video-rate acquisitions, due to the influence of the optical trap, the R-NTA system must use extremely fast acquisition times (thousands of frames per second) in order to record motion that fits within the diffusion regime. With very fast frame rates, therefore, an effective diffusion coefficient could be determined by fitting a line through the first three frame lags and finding the slope⁶ according to the Eq(2) in the main text. In practice, the particle's position over time could be extracted from the two seconds video sequence. The particle tracking analysis was achieved using the open source Image-J plugin "MOSAIC"⁷. Its basic function is to determine the sub-pixel level centroid of each particle in each frame and then re-link the individual particles between frames. Despite the particle motion being below the imaging system diffraction limit, the centroid of the particle in each frame can still be determined with very high accuracy⁸. The output provides the two dimensional coordinate of centroid in each frame, referring to Figure S1(a). Since the temperature increase at the focus caused by laser heating is relatively small due to our choice of near-infrared wavelength⁹, we regard T in the Eq(1) to be the controlled room temperature of the experimental laboratory (295 K).

However, the threshold frame rate between the diffusion and intermediate regimes depends on the trap stiffness, which is a function of optical trapping power and particle size. In the experiments we perform in the Results section, the threshold frame rate at high laser powers is beyond the maximum frame rate of the sCMOS camera. As the power of the laser is reduced, the size estimation changes because the motion captured by the camera moves from the intermediate to diffusion regimes. As the optical stiffness is proportional to the third power of particle diameter, size heterogeneities of nanoparticles make it difficult to determine an optical trapping power for all particles *a priori*. Hence, while measuring nanoparticles with unknown size, the measurement proceeds as follows: A Raman spectrum is acquired using relatively high trapping power. The trapping power is then successively reduced, with a short video (2s) of the particle motion recorded at each power, until the trapping strength is too low to stably trap the nanoparticle. We then extract the size to which the low-power trapping measurements converge.

5. Optimizing Size estimation

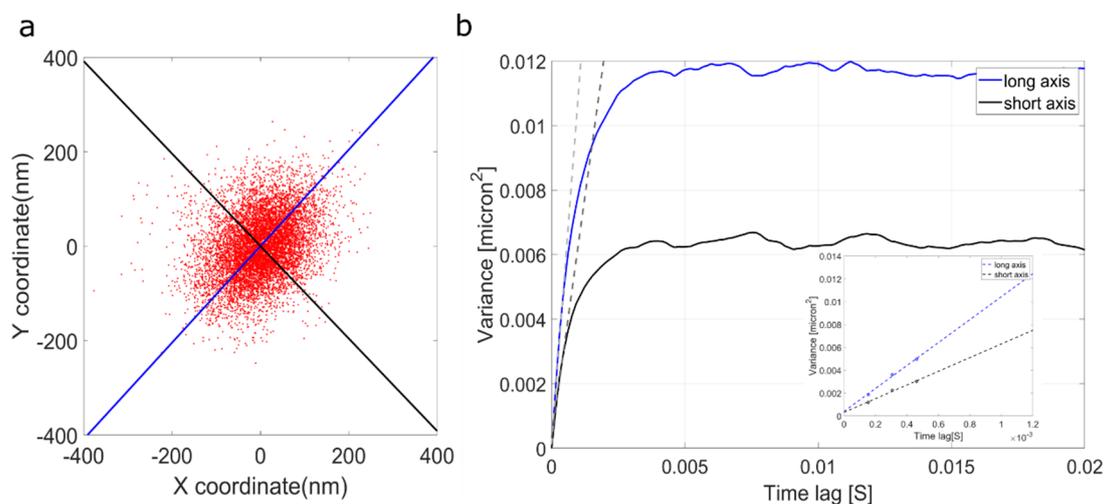


Figure S1. Size estimation versus calculation direction. **a** shows the position distribution of PS 100nm nanoparticle in 2D and **b** its variance in two perpendicular directions. Inset shows the linear fit through the first three points.

Ideally, the position of the particle should have an isotropic probability distribution in x and y . However, the non-circular symmetric polarization of the laser makes the optical stiffness higher along the axis of stronger polarization. As a result, we can clearly see that the contour shape shown in Figure S1(a) is not circular but elliptical. Considering that our constrained nanoparticle motion is close to the intermediate regime, we calculate the size along the direction in which the trapping stiffness is the lowest. In this direction, the nanoparticle motion most rigorously satisfies the diffusion approximation, so that the size estimation would be the most accurate. Hence, the long axis of the contour ellipse (blue line) would be the optimal direction choice.

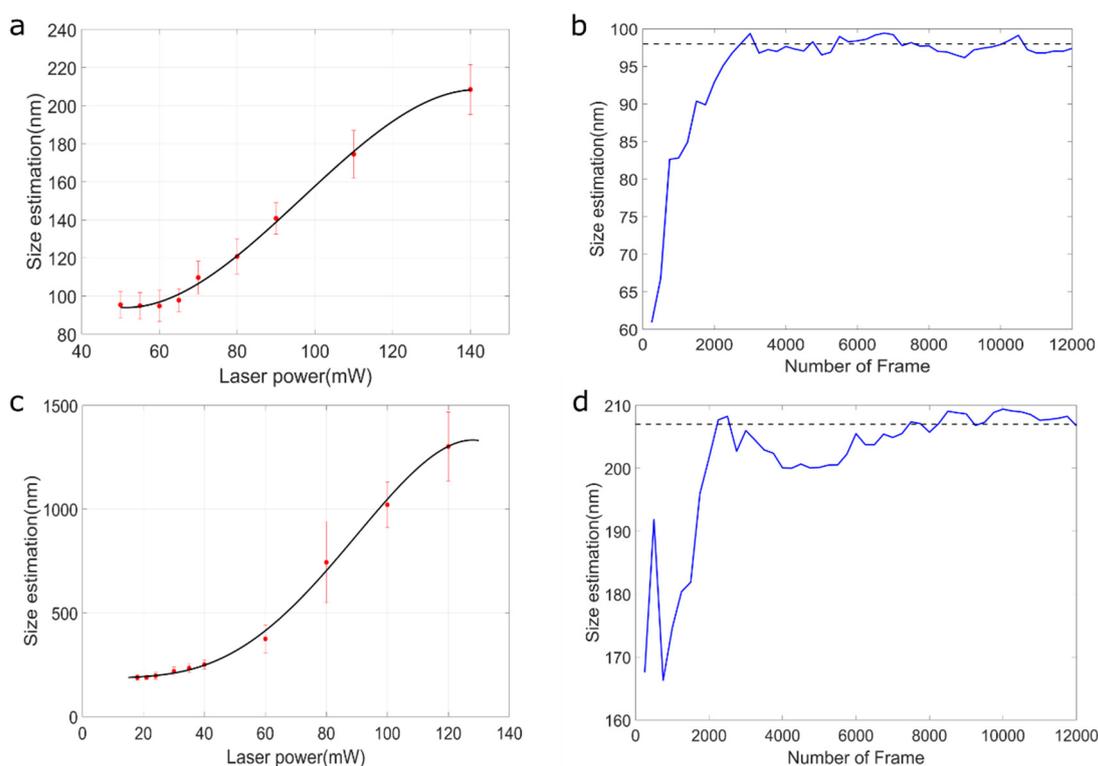


Figure S2 Size estimation vs. laser power and video length. **a c** image shows the size estimation of PS 100nm and PS 200nm, respectively, versus laser power. **b d** image shows the relationship between estimated diameter and number of frames used for calculation. Dotted line represents the final size value.

Further, at high laser powers even at the maximum frame rate of the camera, large particles or those with high refractive index contrast may not be in the diffusion regime. As shown in Supplemental Figure S2, while laser power is successively decreased (reducing the trapping strength and allowing the particle to more rigorously satisfy the diffusion approximation), the estimated size converges to a stable value. At this point a residual 3-5% fluctuation in sizing remains, indicating the precision limit of the sizing. Besides, size estimation increased to a convergence as laser power continuously increase, which means that the high laser power trapping leads to an overestimation of size. Theoretical explanation of this result is given as follow.

Gaussian width of particle position could be given by Smoluchowski equation.

$$\sigma^2 = \frac{k_b T \{1 - \exp[-2KD\Delta t / k_b T]\}}{K} \quad \text{E. S5}$$

Where K and D represent optical stiffness and diffusion coefficient, respectively, and Δt represents the time interval between measurements. From this equation we could see that as stiffness increases, variance would decrease and finally approach zero, which means that particles are trapped so tightly that they hardly move in the trap. Also we have $\sigma^2 = 2D\Delta t$ and $D = \frac{K_b T}{3\pi\eta R}$ in diffusive regime, where η is the viscosity of the solvent and R is diameter of nanoparticles. Therefore, it is easy to draw the conclusion that the estimated R increases as laser power (and consequently trap stiffness) increases.

Furthermore, as shown in Figure S2(a) and (c), the size estimation of larger particles is more sensitive to the laser power. Conversely, and fortuitously, it is easier to achieve a precise size measurement for small nanoparticles with smaller refractive index contrast, such as EVs and liposomes, for which the trapping forces are lower. Meanwhile, the effect of the number of frames on the sizing is shown in Figure S2(b) and (d), showing that once the video reaches about 1 second in length (~ 6000 frames) the size converges to a stable value.

6. Mie theory

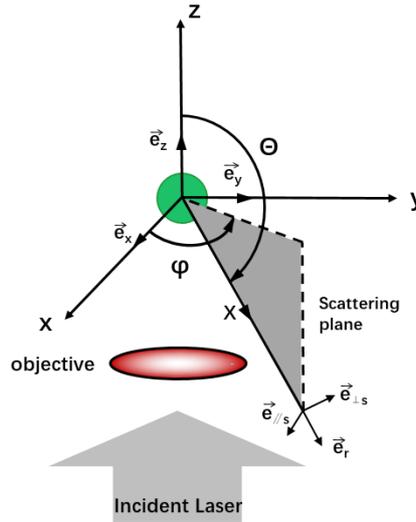


Figure S3. Mie scattering geometry. A spherical particle with diameter d and refractive index n_p that is illuminated by incident laser (approximated here as a plane wave) propagating along the z direction.

Since the size of nanoparticles is much smaller than laser beam waist, at the laser focus we could simply regard the incident laser as a plane wave instead of spherical wave. Full consideration of the Gaussian nature of the beam profile could be obtained using Generalized Lorenz-Mie Theory (GLMT), pioneered by Grehan and Gouesbet¹⁰, but the difference in scattering intensities for the particle size and focal volumes considered here are negligible. Using standard Mie theory, then, the vector scattering amplitude X could be written as

$$X = (S_2 \cos \phi) \vec{e}_{\parallel s} + (S_1 \sin \phi) \vec{e}_{\perp s} \quad \text{E. S6}$$

Where $\hat{e}_{\parallel s}$ and $\hat{e}_{\perp s}$ are parallel and perpendicular unit vectors to the scattering plane. The parameters S_1 and S_2 are the scattering matrix elements and calculated using the Matlab routines of Mätzler¹¹ that are based on the Mie theory formalism as defined in Bohren and

Huffman's classic text¹². In case of back-scattering detection, the z-axis is an axis of symmetry and the backward scattered intensity is given by¹³:

$$I_{BS} = \frac{\pi\alpha_{BS}}{k^2} \int_{\theta_{min}}^{\theta_{max}} \frac{1}{2} (|S_1|^2 + |S_2|^2) \sin\theta d\theta \quad E. S7$$

Where k and θ represent wave number and polar angle, respectively. In our R-NTA system, θ is limited by numerical aperture (NA) of the objective. NA characterizes the range of angles over which the objective collects light and is defined as $NA = n_m \sin\alpha$, where n_m is the refractive index of the medium and α is the maximum propagation angle. Hence, θ is integrated from $\theta_{min} = \pi - \sin^{-1}\alpha$ to $\theta_{max} = \pi + \sin^{-1}\alpha$. Since θ is symmetrical along the z-axis, we have

$$I_{BS} = \frac{2\pi\alpha_{BS}}{k^2} \int_{\theta_{min}}^{\pi} \frac{1}{2} (|S_1|^2 + |S_2|^2) \sin\theta d\theta \quad E. S8$$

And the scalar α_{BS} is introduced in order to take the power of the laser beam and the camera setting into account in the numerical computations. Practically, it represents the linear calibration factor between theory and experimental data. The calibration of this factor is done by matching theoretical Mie scattering intensities with experimental measurements on NIST standard nanoparticles. Following calibration consistent settings for camera and laser power are required for experimental measurements. Backward scattering intensity of trapped nanoparticles could be quantified by using the 16-bit images after removing the background caused by the reflection of coverslip, and the scattering intensity is the sum of all pixel intensities within the circular window with its centroid determined by "MOSAIC" plugin. The key parameters setting used in "MOSAIC" is as follows: particle radius: 8 pixels; the cutoff threshold for particle detection: 0.01; the dynamic particle motion model: Brownian motion; particle re-linking range: 10~20 frames; maximum jump distance: 20 pixels. Because the measured scattering intensity depends on axial position of the nanoparticle, which fluctuated due to constrained Brownian motion, we regard that nanoparticle is at focus when the maximum scattering intensity is detected.

7. Precision of RI measurement

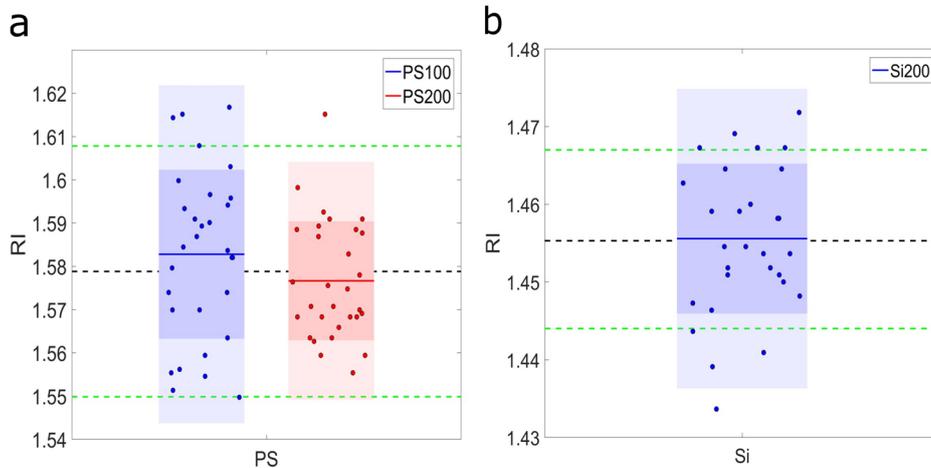


Figure S4. Precision of RI measurements using NIST standard polystyrene (a) and silica (b) particles.

Green dashed lines indicate the 95% confidence interval due to R-NTA sizing error and laser power fluctuations. Dark and light color bars represent one and two standard deviations of the RI determinations, respectively.

The limitation of the RI determination should depend on two main factors: (1) as the RI is determined using the size calculated from the Brownian motion, sizing accuracy will affect the RI accuracy; (2) As the RI is directly determined from the scattering power, laser power fluctuation is another non-negligible factor. Laser power fluctuation was measured to be approximately 3%, leading to a 0.4% error in RI, while a theoretical analysis of silica and polystyrene spheres shows a 3–5% sizing error yields an approximately 1.5% (1.5788 ± 0.02) and 0.8% (1.4553 ± 0.01) fluctuation in the extracted RI for polystyrene and silica, respectively. Thus, the theoretical limit should be 1–2% error assuming the size and laser power errors are uncorrelated and add in quadrature.

8. Precision of quantitative Raman measurement

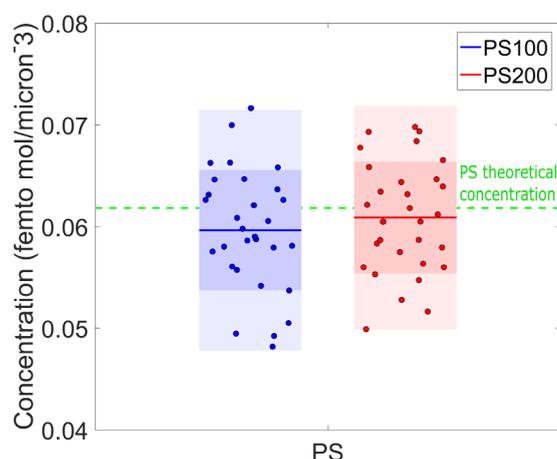


Figure S5. Precision analysis for absolute Raman concentration measurements. Concentration of polystyrene in PS 100nm and PS 200nm samples measured by R-NTA match well with the theoretical value $0.0618 \text{ fmol}/\mu\text{m}^3$.

The slight deviation between measured and theoretical concentrations are likely caused by three factors: (1) spectral noise leading to uncertainty in the least squares fitting, (2) sizing errors as described above, and (3) variations in the production of the PS nanoparticles leading to density variations due to variations in crosslinking or other production factors like temperature¹⁴.

9. Preparation process of liposomes

Liposomes were prepared according to the standard procedure. 4 mg mL^{-1} of DPPC, 6 mg mL^{-1} egg PC, and 3 mg mL^{-1} cholesterol (from Aladdin, China, SCR HuShi, China and Sigma-Aldrich, USA, respectively) were dissolved in chloroform. Additionally, 1 mg mL^{-1} curcumin was dissolved by chloroform. Subsequently, the membrane was formed in a 15 ml round bottom flask with the molecular ratio of 1.45:1.45:1:0.43 corresponding to cholesterol: PC: DPPC: curcumin. The chloroform was evaporated under air flow to form a thin lipid film. The films were hydrated with 1.5 ml DI water and sonicated for 10 min under 65°C . The solutions were then extruded through a column filled with glucan gel with 2 min of 2000 rpm/min centrifugation to isolate the undissolved curcumin. Liposomes were extruded using

a membrane with 200 nm pore size and we repeated the extrusion process 3 times.

10. R-NTA analysis of liposomes with additional extrusion

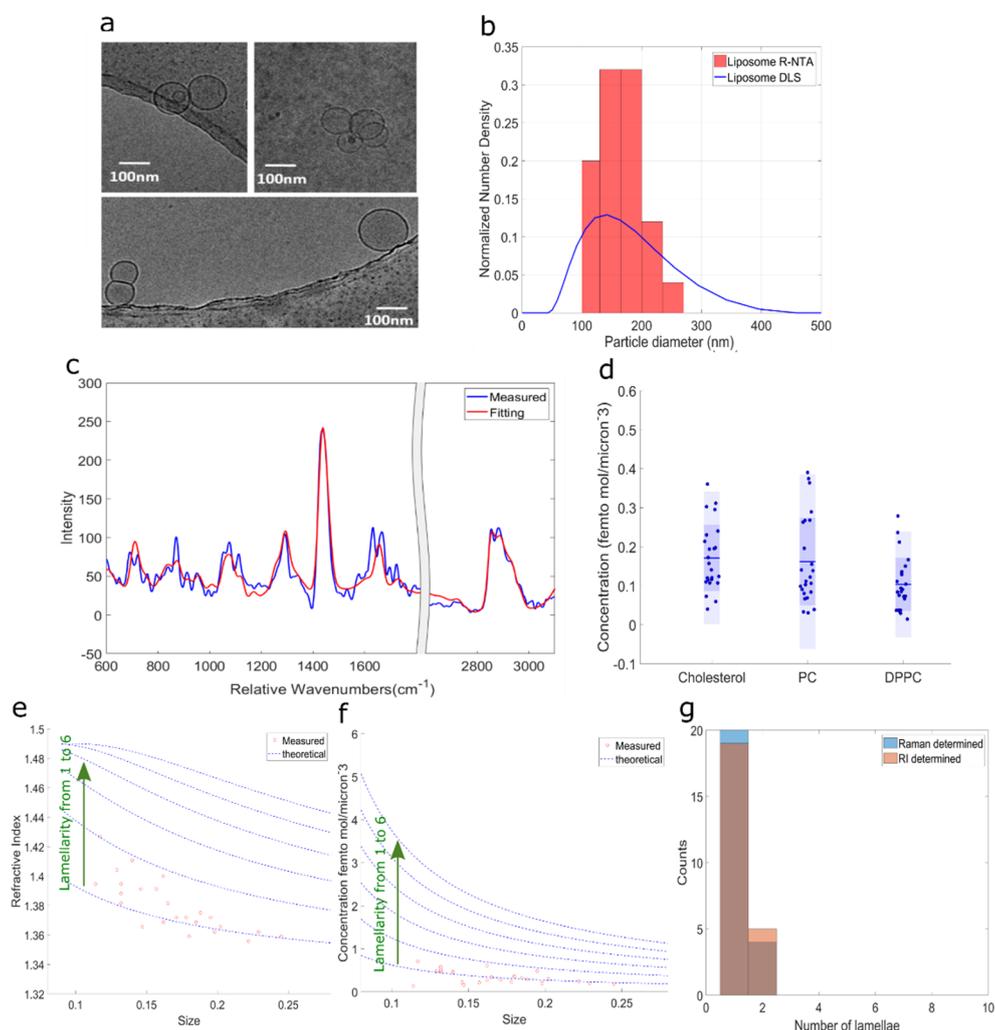


Figure S6. Size, chemical concentration and RI analysis of extruded free drug-loading liposomes (N = 24) based on R-NTA. **a** Cryo-TEM images of extruded liposomes without loaded curcumin. Sample solution has been diluted before extrusion. **b** Histogram of particle sizes measured by R-NTA (blue) and DLS (red). **c** Mean Raman spectra of liposomes (blue) and the corresponding fit by pure components (red) using least-squares fitting. **d** Fitting concentration of each basic component on a per-particle basis. Light and dark shaded areas show ± 1 and 2 SD, respectively. **e,f** Lamellarity fitting of liposomes to theoretical models (blue dashed line) based on Raman-extracted total lipid concentration and scattering-extracted RI, respectively.

As demonstrated in Figure S6(b), the size analyzed by R-NTA also shows good agreement with DLS results for extruded liposomes. Quantitative chemical components determination of extruded liposomes is shown in Figure S6(c) and (d), verifying the chemical profiling capability of R-NTA again. As shown in Figure S6(e) and (f), it is clearly seen from both the Raman and RI analysis results that after the extrusion, most liposomes were unilamellar, which was confirmed via cryo-TEM, shown in Figure S6(a). However, it is indicated in Figure S6(e) and (f) that some liposomes don't fit perfectly with the theoretical curves, which are located between unilamellar and bilamellar curves. This is speculated to be related to

heterogeneous concentration of cholesterol within the lipid membrane, as cholesterol leads to a “condensing effect”¹⁵, where increasing cholesterol increases the density of the membrane, especially for bilayers¹⁶ containing unsaturated lipid like PC. It could also be due to bi-lamellar liposomes where the “inner bubble” is small, as our model assumes that the spacing between the different lipid bilayers is small.

11. EVs sample preparation and purification

HN4 cells were cultured in Dulbecco’s modified Eagle medium (DMEM, 4.5 g/L glucose) (Biological Industries, State of Israel) supplemented with 10 % fetal bovine serum and antibiotics (100 KU/L penicillin and 100 mg/L streptomycin) in an incubator at 37 °C with 5 % CO₂. HN4 cells were transiently transfected with TRPP2 shRNA using Lipofectamine 3000 following the manufacturer’s instruction and with or without 150 µg/ml G418 for 48 h. NP69 purchased from Fenghui Bio (Hunan, China) were cultured in Roswell Park Memorial Institute 1640 (RPMI-1640, Biological Industries, Israel) supplemented with 10 % fetal bovine serum and antibiotics (100 KU/L penicillin and 100 mg/L streptomycin) in an incubator at 37 °C with 5 % CO₂. When the number of cells reached about 80%, the FBS-containing medium was removed and substituted with FBS-free medium and cultured for 24h. For the size-exclusion chromatography (SEC) process the isolation column was filled with Sepharose CL-2B (Bomeibio, China). Cell-free supernatant was loaded on the column, followed by elution with PBS. The eluate was collected in 18 sequential fractions of 1 mL. Fractions 6-11, which has the highest particle concentrations, were utilized for R-NTA characterization.

12. Western Blotting of cells and SEC-isolated EV samples

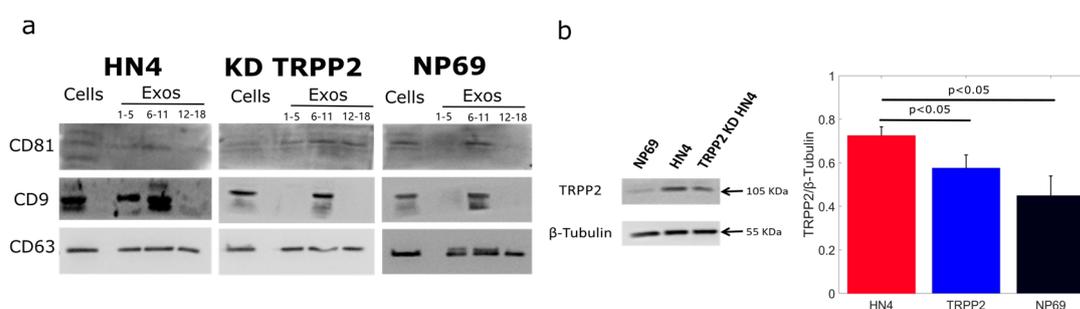


Figure. S7 Western blotting (WB) of extracellular vesicles and cells. a WB analysis of CD81, CD9 and CD63 in HN4, TRPP2 knock down and NP69 cell lines and representative EVs. **b** Validation of TRPP2 knockdown based on WB.

Procedure of western blotting is as follow. Samples were mixed with Laemmli Sample Buffer (0.125 M TrisHCl (pH 6.8), 10% glycerol, 2.3% sodium dodecyl sulphate (SDS)), and then boiled for 10 min at 95 °C firstly. Then proteins were separated by 12% SDS-PAGE (SDS-polyacrylamide gel electrophoresis) and blotted onto polyvinylidene fluoride (PVDF) membranes (Catalog No: ISEQ00010, Millipore, USA). After blotting, membranes were blocked in 5% non-fat milk for 2 hours in room temperature. Membranes were then incubated with a primary antibody overnight at 4 °C. After six five-minute washes in 1xTBST 0.1% Tween-20, membranes were incubated with secondary antibody for 1 h at room temperature. Finally, membranes were imaged by an ECL Imaging System (Shanghai Peiqing, China) and

analyzed using Quantity One software.

13. Stability of Raman measurements of single EVs over measurement period

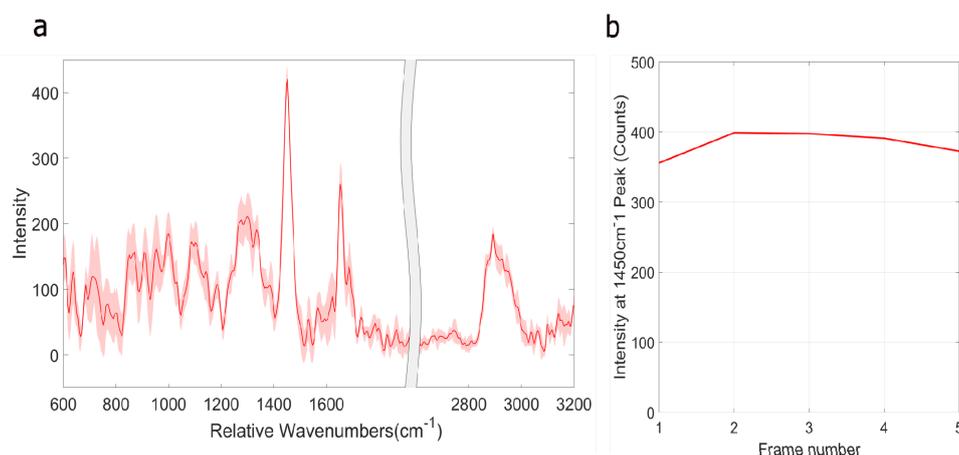


Figure. S8 Time-sequence analysis verifies that no additional vesicles enter the trap during a Raman measurement. **a** Averaged Raman spectra (solid lines) and ± 1 standard deviation (shaded areas) of a single (142nm) optically-trapped EV over 5 frames. **b** Intensity at the 1450 cm^{-1} peak for the trapped EV, showing around 5% variation, consistent with laser fluctuation as well as the particle exploring different z-locations within the trap.

Raman signal fluctuation over 5 frames was checked after spectral processing, as shown in Figure S8(a). Spectra intensity at the 1450 cm^{-1} (CH_2 vibrations) shows around 5% fluctuation in the signal over these frames, as demonstrated in Figure S8(b). The main factor that causes this fluctuation is laser power instability, Gaussian readout noise, and the particle exploring different z-positions within the trap.

14. Pure component spectra

To analyze the chemical composition in samples quantitatively, spectra of several pure components were measured by the R-NTA system. Here, cholesterol, phosphatidylethanolamine, phosphatidylcholine and sphingomyelin were measured in the liquid state, DNA, albumin and glycogen were measured as dissolved in DI water. DPPC and Curcumin were measured in liquid form by dissolving in Chloroform. In the case of dissolved components, chloroform and DI water spectra were subtracted to obtain the pure component spectrum.

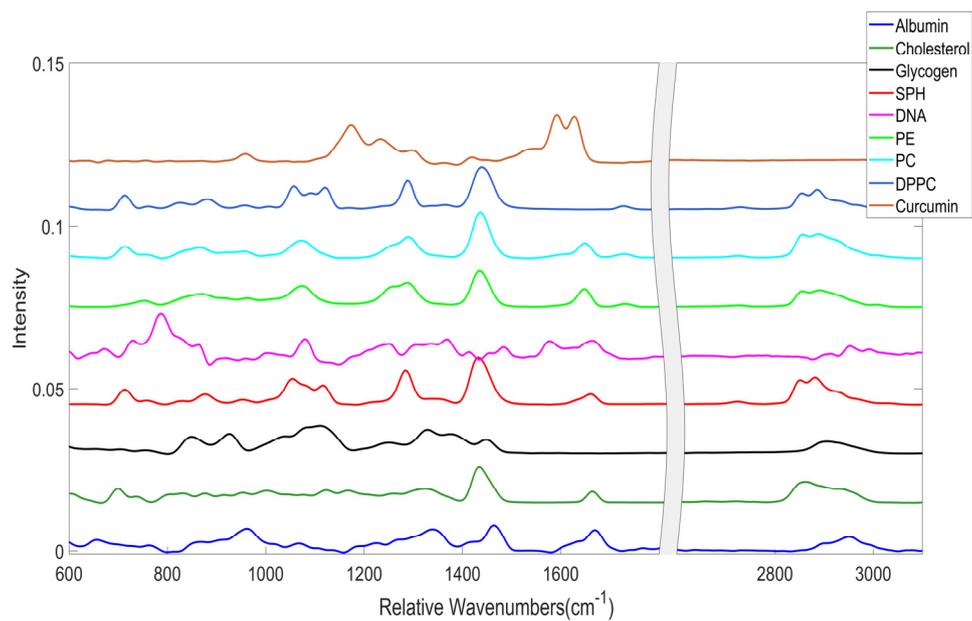


Figure S9. Normalized pure components spectra measured by R-NTA system.

15. Principal components spectra and their fittings with pure component spectra

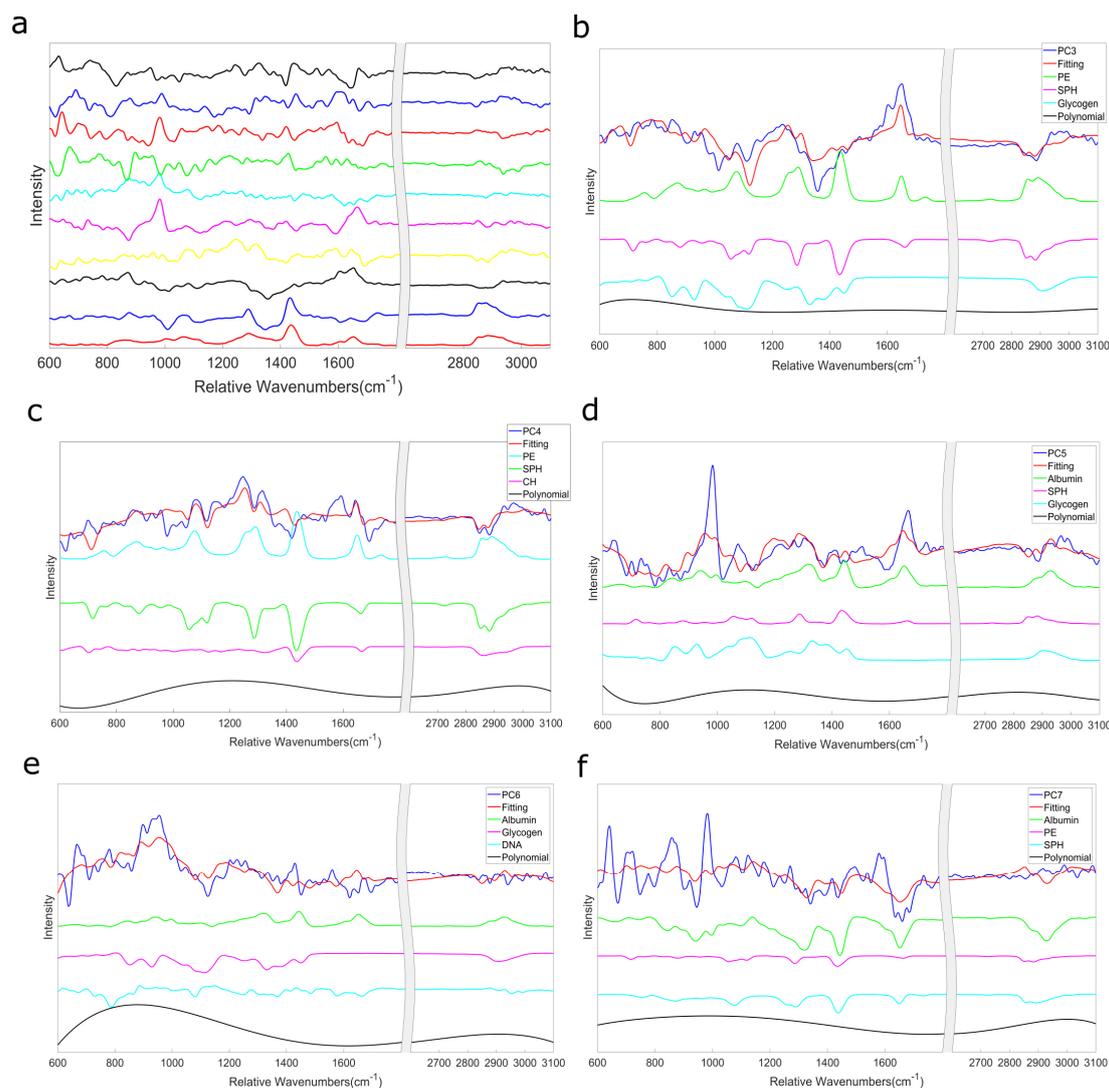


Figure S10. PCs and spectral fitting to PCs for EV dataset. **a** first 10 principal components of EVs Raman spectra. Spectra are offset for clarity, PC 1 at the bottom and PC 10 at the top. **b-f** Least squares fit of principal component 3~7 fitted with pure components spectra of sphingomyelin (SPH), phosphatidylethanolamine (PE), cholesterol (CH), glycogen. Pearson correlation coefficients are 0.77, 0.79, 0.83, 0.84 and 0.52, respectively. If ignoring the contribution from polynomial terms, Pearson correlation coefficients of PC 3-7 are 0.53, 0.53, 0.64, 0.01 and 0.30.

Visual comparison of pure chemical fits to PC1 and PC2 are shown in Figure S10(b) and Figure S10(c), respectively. Through least-squares fitting, the correlation coefficients between the principal components and the spectral model was determined to be 0.98 and 0.90, which means that the spectral model faithfully reproduces both loadings' major peaks. The fit coefficients can be used to determine quantitative relationships between chemicals represented by this principal component. The increasingly noisy character and relatively low correlation coefficient (high correlation coefficient of PC6 is mainly contributed by the polynomial terms) of the loadings PC6 and higher indicate that these PCs primarily represent both the noise during Raman measurement as well as lower abundance chemical compounds not accounted for in our spectral library (for example, specific surface proteins, hormones, and other small molecules).

16. Hierarchical clustering analysis of EVs

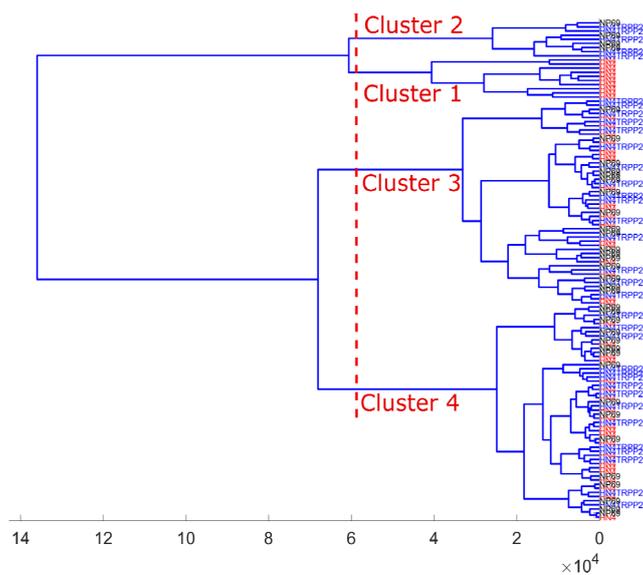


Figure S11. Hierarchical Cluster Analysis of EV dataset. The dendrogram shows the clustering result using the first 4 principal components and size information as inputs. Labels are color coded as in Figure 5(c).

17. EVs clustering by Raman information only

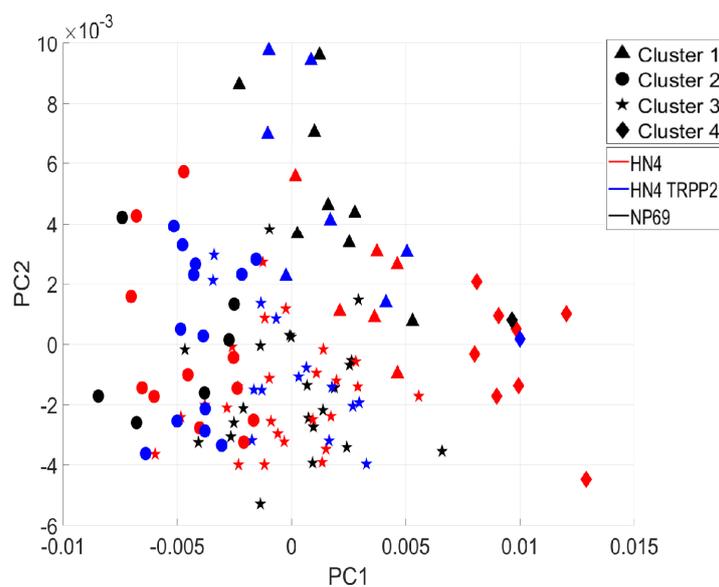


Figure S12. First and second principal component scores for each cell line based on vector normalized Raman spectra (without size information), showing substantially reduced separation between clusters.

18. Biochemical concentration analysis of EV spectra

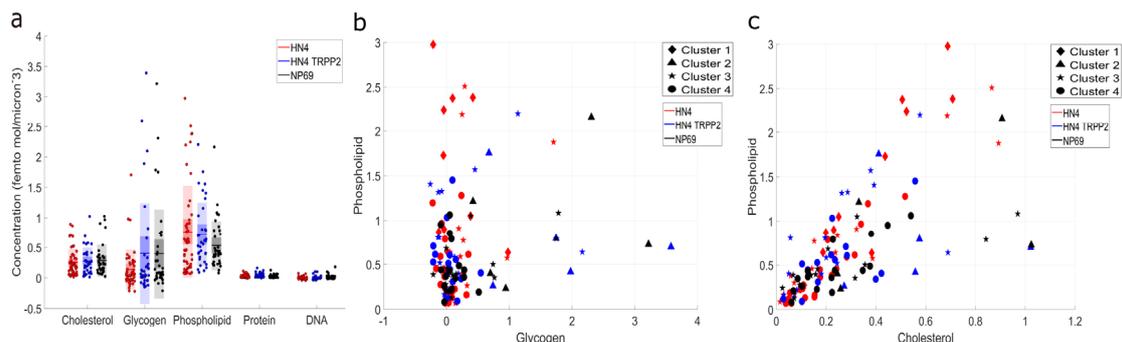


Figure S13. **a** Biochemical modeling of 3 groups of EVs via known pure chemical components. Difference between cell lines are dominated by cholesterol, glycogen and lipid content of individual outliers. **b-c** Phospholipid versus glycogen concentration and phospholipid versus cholesterol concentration, respectively, for EVs of each cell line. Colors and shapes are as in (a). Outlier vesicles (Clusters 1 and 2) show diverse chemical contents, with HN4 outliers expressing high phospholipid but average glycogen content, while NP69 and TRPP2-knockdown outliers show high glycogen and cholesterol, yet average-to-low phospholipid content.

Using each PC loading's chemical interpretation, multiplied with the size-normalized score of each EV on that axis, quantitative biochemical concentrations for different chemical components were determined on a per-vesicle basis, as shown in Figure S13(a). In order to understand the chemical contents of vesicles in different clusters, we plot the individual vesicles in a quantitative biochemical space formed by the fit coefficients for cholesterol, phospholipid, and glycogen, as shown in Figure S13(b) and (c). Here we see groupings of the vesicles that largely recapitulate the clustering results described in the main text, but with enhanced biochemical significance. It could be clearly seen that outlier vesicles are not simply "dense" vesicles (where all chemicals are enriched) but have diverse chemical content. For instance, most of the vesicles clustered in Cluster 1 have abundant phospholipid but average glycogen expression level, while those vesicles clustered in Cluster 2 predominantly feature heightened glycogen content but with average phospholipid concentration.

Table S1 EVs Spectral Interpretation

Position(cm^{-1})	Peak Assignment	Reference
700	Cholesterol ester	(17)
852	Glycogen	(18)
935	C-C stretching of protein backbone in glycogen	(18)
1002	Phenylalanine in proteins	(19)
1066	Chain C-C stretching in lipids	(17)
1296	CH ₂ deformation in lipids	(18)
1320	CH ₂ -CH ₃ stretching vibration in nucleic acid	(20)
1651	C=C stretching in lipids	(17)
1668	Amide I vibrations in proteins Cholesterol ester	(17)
2900	CH stretch in lipids and proteins	(19)
2928	CH ₂ antisymmetric stretch in lipids	(19)
2970	CH ₃ antisymmetric stretch in cholesterol ester	(20)

Table S2 Fitting coefficients of each cluster

	Cholesterol	Glycogen	Phospholipid	Protein	DNA
Cluster1	0.29	0.07	1.22	0.03	0.01
Cluster2	0.42	1.24	0.61	0.05	0.04
Cluster3	0.24	0.07	0.38	0.01	0.01
Cluster4	0.10	0.04	0.34	0.01	0.01

Unit fmol/micron^{^3}

References

- Eilers, P.H.C. A Perfect Smoother. *Anal. Chem.* **2003**, 75, 3631-3636.
- Jiang, S.; Zhang, X.; Zhang, Y.; Hu, C.; Zhang, R.; Zhang, Y.; Liao, Y.; Smith, Z.J.; Dong Z.; Hou J. Subnanometer-resolved chemical imaging via multivariate analysis of tip-enhanced Raman maps. *Light-Sci. Appl.* **2017**, 16, e17098.
- Lindner, M.; Nir, G.; Vivante, A.; Young, I.T.; Garini, Y. Dynamic analysis of a diffusing particle in a trapping potential. *Phys. Rev. E.* **2013**, 87, 022716.
- Lukic, B.; Jeney, S.; Tischer, C.; Kulik, A.J.; Forro, L.; Florin, E.-L. Direct observation of nondiffusive motion of a Brownian particle. *Phys. Rev. Lett.* **2005**, 95, 160601.
- O'Dell, D.; Schein, P.; Erickson, D. Simultaneous Characterization of Nanoparticle Size and Particle-Surface; Interactions with Three-Dimensional Nanophotonic Force Microscopy. *Phys. Rev. Appl.* **2016**, 6, 034010.
- Ernst, D.; Kohler, J. How the number of fitting points for the slope of the mean-square displacement influences the experimentally determined particle size distribution from single-particle tracking. *Phys. Chem. Chem. Phys.* **2013**, 15, 3429-3432.
- Sbalzarini, I. F.; Koumoutsakos, P. Feature Point Tracking and Trajectory Analysis for Video Imaging in Cell Biology. *J. Struct. Biol.* **2005**, 151, 182-195.
- Thompson, R.E.; Larson, D.R.; Webb, W.W. Precise nanometer localization analysis for individual fluorescent probes. *Biophys. J.* **2002**, 82, 2775-2783.
- Peterman, E.J.G.; Gittes, F.; Schmidt, C.F. Laser-Induced Heating in Optical Traps. *Biophys. J.* **2003**, 84, 1308-1316.
- Hodges, J. T.; Grehan, G.; Gouesbet, G.; Presser, C. Forward scattering of a Gaussian beam by a nonabsorbing sphere. *Appl. Optics.* **1995**, 34, 2120-2132.
- Marzler, C. MATLAB Functions for Mie Scattering and Absorption. *Institut fur Angewandte Physik.* 02, 2002.
- Bohren, C.F.; Huffman, D.R. Absorption and scattering of light by small particles. *Wiley: New York*, 1983.
- F, Jacques.; Baudry, J.; Emerard, J.-E.; Bertrand, E.; Goubault, C.; Henryband, N.; Bibette, J. Size and fluorescence measurements of individual droplets by flow cytometry. *Soft Matter.* **2009**, 5, 2232.
- Pugh, T. L.; Heller, W. Density of polystyrene and polyvinyltoluene latex particles. *J. Colloid Sci.* **1957**, 12, 173-180.
- Emilova, I.; Lyubartsev, A.P. Cholesterol in phospholipid bilayers: positions and orientations inside membranes with different unsaturation degrees. *Soft Matter.* **2019**, 15, 78-93.
- Hung, W.-C.; LeeM.-T.; Chen, F.-Y.; Huang, H.W. The Condensing Effect of Cholesterol in Lipid Bilayers. *Biophys. J.* **2007**, 92, 3960-3967.

- 17 Edwards, H.G.W.; W. Farwell, D.W.; Williams, A.C.; W. Barry, B.W.; Rull, F. Novel spectroscopic deconvolution procedure for complex biological-systems-vibrational components in the Ft-Raman spectra of ice-man and contemporary skin. *J. Chem. Soc Faraday. Trans.* **1995**, 91, 3883-3887.
- 18 Movasaghi, Z.; S Rehman,S.; Rehman, I.U. Raman Spectroscopy of Biological Tissues. *Appl. Spectrosc. Rev.* **2007**, 42, 493-541.
- 19 Chen,Y.; Dai, J.; Zhou, X.; Liu, Y.; Zhang, W.; Peng, G. Raman Spectroscopy Analysis of the Biochemical Characteristics of Molecules Associated with the Malignant Transformation of Gastric Mucosa. *Plos One.* **2014**, 9, e93906.
- 20 Notingher I.; Verrier, S.; Haque, S.; Polak, J.M.; Hench, L.L. Spectroscopic study of human lung epithelial cells (A549) in culture: living cells versus dead cells. *Biopolymers.* **2003**, 72, 230-240.