Supplemental Material

Effect of the Total Isolation Reagent on the Exosome Spectrum

Preparation of exosomes is typically performed by ultracentrifugation. However, to alleviate the experimental difficulty of and length of time required for ultracentrifugation, exosome isolation kits are commercially available from several manufacturers. These kits allow exosomes to be pelleted at low centrifuge speeds, as described in the Methods section, above. In this work we use one standard kit, the Total Exosome Isolation Reagent from Life Technologies (catalog # 4478359).

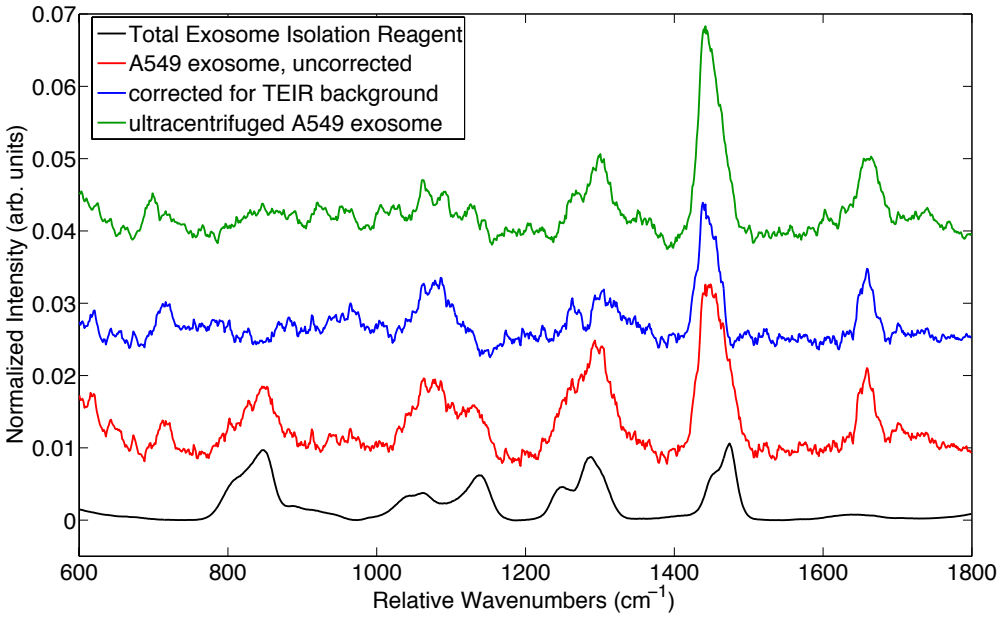

Figure S1. Raman spectra of total exosome isolation reagent (black) and trapped exosomes (colored). Numerical removal of the total isolation reagent by AsLS yields a spectrum (blue) in good agreement with exosomes prepared by ultracentrifugation (green). Spectra offset for clarity.

The spectrum of this reagent is shown as the black trace in Figure S1. Interestingly, although the exosomes are rinsed and resuspended in PBS buffer following separation and pelleting, a strong remnant of the isolation reagent's signal is clearly present in the spectra of individual trapped exosomes. The spectrum of an exosome from an a549 cell culture where the isolation reagent was not included as part of the background model described above is shown in red in Fig. S1. Including the isolation reagent's spectrum in the background model leads to the spectrum shown in blue. Finally, we isolated a549 exosomes by gold-standard ultracentrifugation, leading to the spectrum shown in green. Comparing these curves seems to suggest that the isolation reagent works by coating the exosomes, causing them to precipitate more easily during centrifugation. However, this coating remains even after washing and resuspension. A previous report of Raman measurements of exosomes in the

literature (47) did not correct for this effect, and it appears that the majority of the spectral signal they report is due not to the exosomes themselves but to the isolation reagent coating the exosomes. However, by including the isolation kit's spectrum in our background model, AsLS efficiently removes the contaminating peaks, giving rise to a corrected spectrum that agrees well with exosome spectra produced by the more laborious ultracentrifugation. The agreement between our ASLS-corrected spectra from kit-prepared exosomes and ultracentrifuged exosomes is further confirmed by the correspondence between the principal component lineshapes computed from seven cell lines shown in Fig. 1 (prepared with the TEIR reagent), and the principal component lineshapes computed from the trypsin experiment shown in Fig. 2 (computed from ultracentrifuged samples).

Hierarchical Clustering of Cell Line Spectra

Figure S2. The first ten principal component loading vectors, calculated from the full exosome spectral dataset from all seven cell lines.

Following principal components decomposition, we plot the first 10 principal component loadings, shown in Figure S2 (first three curves identical to Fig. 1 (D)). Projecting the Raman spectra onto these 10 principal component axes yields 10 scores. A hierarchical clustering analysis, performed as described in the Methods section, above, using these scores, yields a dendrogram as shown in Figure S3. The dendrogram describes spectral distance, where branches of the dendrogram tree are connected based on the similarity of one spectrum to another. This similarity is assessed by computing pairwise distances between each spectrum in the 10-space described by first 10 principal components. Spectra with small pairwise distances are more similar to each other (have more similar PC scores). The dendrogram tree

has been extended to the level of individual exosomes, with each terminus label color-coded as in Figure 2. The dendrogram reveals four major clusters, with clusters 2 and 3 having a heavy representation from exosomes of non-tumor cell lines. Spectra from each cluster can be averaged to yield cluster-averaged spectra as shown in Figure 2 (E).

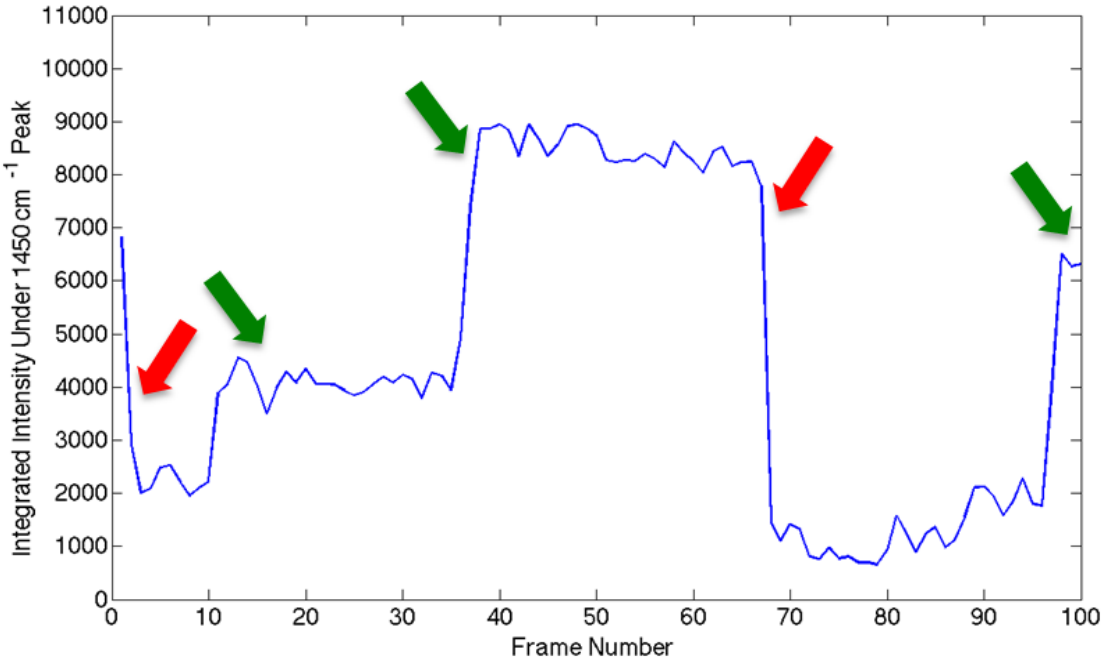

Figure S3. Dendrogram showing the clustering result using the first 10 principal components. Labels are color coded as in Figure 2.

Fitting Principal Component Axes from Seven Cell Lines to Pure Components and Principal Components of Control Experiments

Pure component spectra of membrane lipid components were fit to the shape of the 3rd principal component axis from the seven cell lines experiments using standard least-squares fitting, with the result shown in Figure S4. Note that the pure spectra in Figure S4 are scaled by their fit coefficient, so cholesterol, for example, has a negative fit coefficient and therefore is the negative of the lineshape shown in Fig. 2, above. A 5th order polynomial was also included in the model to account for baseline fluctuations between the two experiments. The fit has a high correlation (r=0.82) with the PC_{7c} 3. The fitting coefficients can be used to quantitatively determine the molecular relationship between cholesterol and phospholipid represented by the principal component. The pure components were measured from solid forms of the chemicals that fully filled the confocal measurement volume of the LTRS system. Thus, the focal volume was filled with a certain number of molecules related to the density and molecular weight of the chemical being measured. Given this, the total number of molecules measured per focal volume can be calculated following Equation S1, below, where N_A is Avogadro's number:

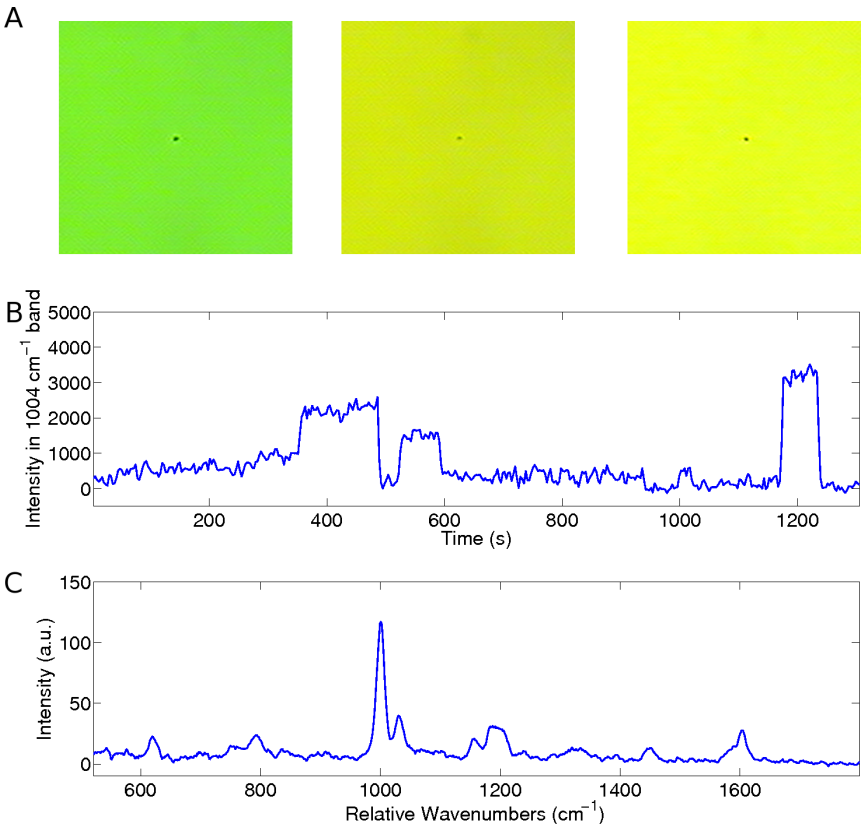

Figure S4. Least squares fit of the 3rd principal component from Figure 1 (D) to the pure spectra of cholesterol, phosphatidylcholine, and phosphatidylethanolamine shown in Figure 2. Spectra offset for clarity.

$$\frac{molecules}{Vmeas} = density(\frac{g}{mL}) \times \frac{1}{Mol.Wt.}(\frac{mol}{g}) \times N_A(\frac{molecules}{mol}) \qquad \text{(Equation S1)}$$

Cholesterol, phosphatidylcholine, and phosphatidylethanolamine have similar densities (1.067, 1.019, and 1.0, respectively). Thus, the primary determinant in the number of molecules per measurement volume is the molecular weight of each component (386.95 for cholesterol, 768 for phosphatidylcholine, and 744 for phosphatidylethanolamine). Thus, each pure component spectrum represents the signal from a certain number of molecules of that pure chemical, which could be determined by multiplying Eq. S1 by a highly accurate measure of the measurement volume V_{meas}. However, even in the absence of that information, because V_{meas} is constant between experiments, ratios of fit coefficients can be used to quantitatively determine the ratios of different molecules represented by PC_{7c} 3. The coefficients of the fit are:

$$-0.42 \times chol. + 0.28 \times phosphatidylcholine + 0.6 \times phosphatidylethanolamine$$

where "chol," "phosphatidylcholine" and "phosphatidylethanolamine" represent one unit spectrum of pure crystals of these compounds. Taking the ratio of the fit values for cholesterol and phospholipids, where each fit value is scaled by the number of molecules represented by each unit spectrum using Eq. S1, we can determine the

ratio of cholesterol to phospholipid represented by PC_{7c} 3, as shown in Eq. S2, below:

$$\frac{-0.42}{386.95} \div \left(\frac{.28}{768} + \frac{.6}{744}\right) \approx 1 \qquad \text{(Equation S2)}$$

Therefore, PC_{Tr} 3 represents an exchange of cholesterol for phospholipids, with an exchange ratio of 1 cholesterol molecule for 1 phospholipid.

A similar fitting can be carried out for the 2nd principal component from the 7 cell lines experiment (PC_{7c} 2), which is hypothesized to report on membrane protein. In this case, we have *a priori* knowledge that PC_{Tr} 1 and PC_{Tr} 2 from the trypsin experiment are related to membrane protein (as they are the dominant sources of spectral difference between native and trypsinized EVs). Therefore, we can attempt to fit PC_{7c} 2 with PC_{Tr} 1 and PC_{Tr} 2 of the trypsin experiment. The results of this fit are shown below, in Figure S5 (fit coefficients = -0.22*PC_{Tr} 1 + 0.43*PC_{Tr} 2). Similar to above, note that PC_{Tr} 1 in this case has a negative fit coefficient and is therefore the inverse lineshape compared to its presentation in Fig. 3(B). The fit has a high correlation with PC_{7c} 2 (r = 0.8), and provides strong support for the hypothesis that PC_{7c} 2 and PC_{Tr} 1 and 2 report on the same quantity: surface protein expression.

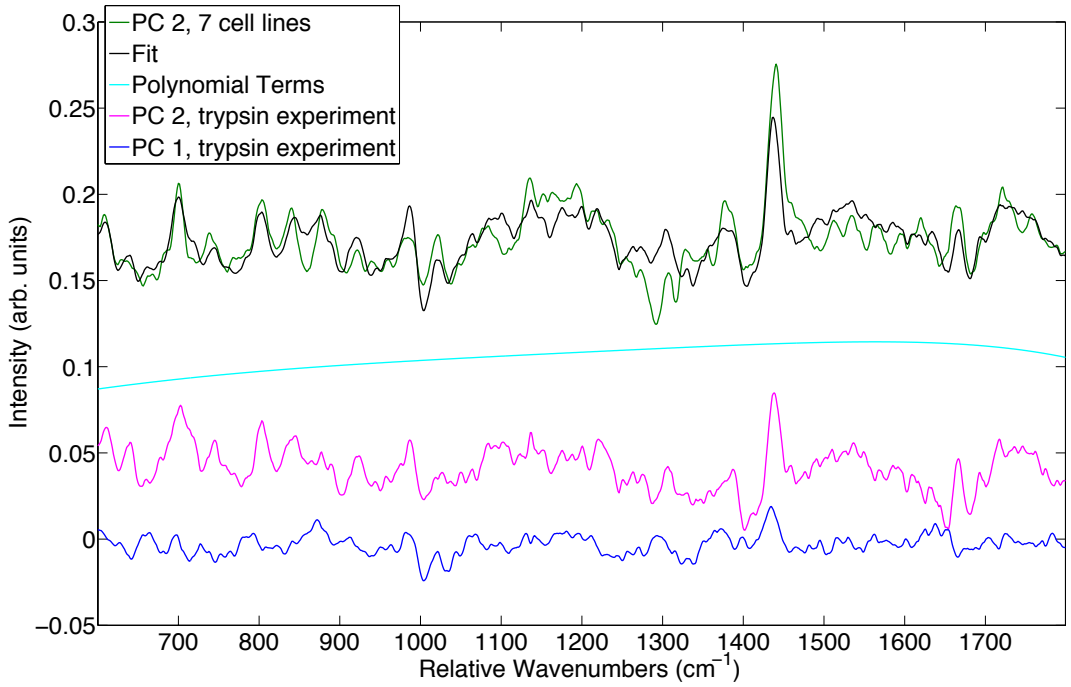

Figure S5. Least squares fit of the 2nd principal component from Figure 1 (D) by the 1st and 2nd principal components from the trypsin control experiment shown in Figure 3 (B). Spectra offset for clarity.