nature chemical biology

SUPPLEMENTARY INFORMATION

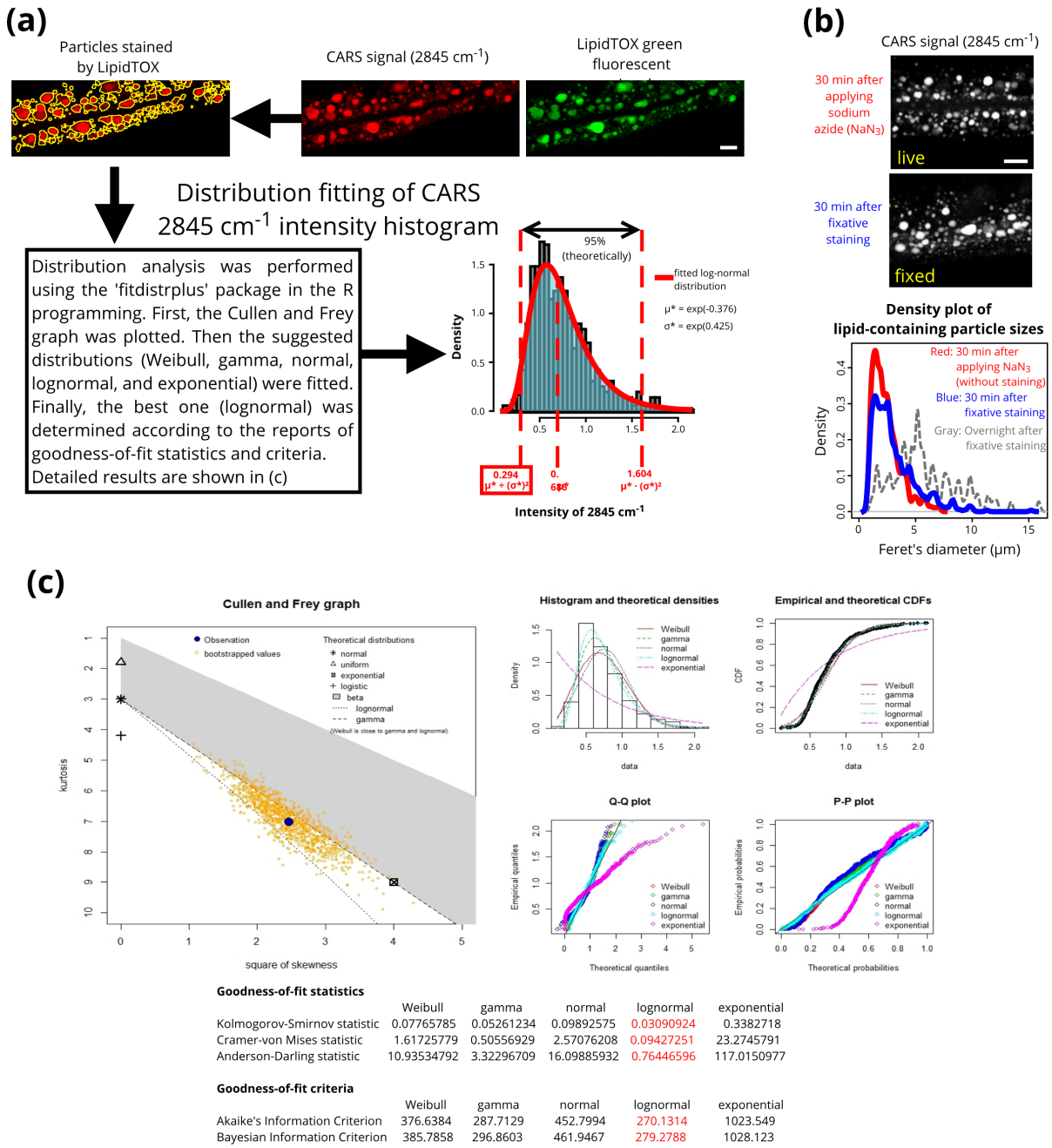
https://doi.org/10.1038/s41589-020-0565-2

In the format provided by the authors and unedited.

Spectroscopic coherent Raman imaging of *Caenorhabditis elegans* reveals lipid particle diversity

Wei-Wen Chen[1,2], George A. Lemieux[3], Charles H. Camp Jr[2], Ta-Chau Chang[4], Kaveh Ashrafi[3]✉ and Marcus T. Cicerone[1,2]✉

[1]School of Chemistry & Biochemistry, Georgia Institute of Technology, Atlanta, GA, USA. [2]Biosystems and Biomaterials Division, National Institute of Standards and Technology, Gaithersburg, MD, USA. [3]School of Medicine, University of California, San Francisco, CA, USA. [4]Institute of Atomic and Molecular Sciences, Academia Sinica, Taipei, Taiwan. ✉e-mail: Kaveh.Ashrafi@ucsf.edu; cicerone@gatech.edu



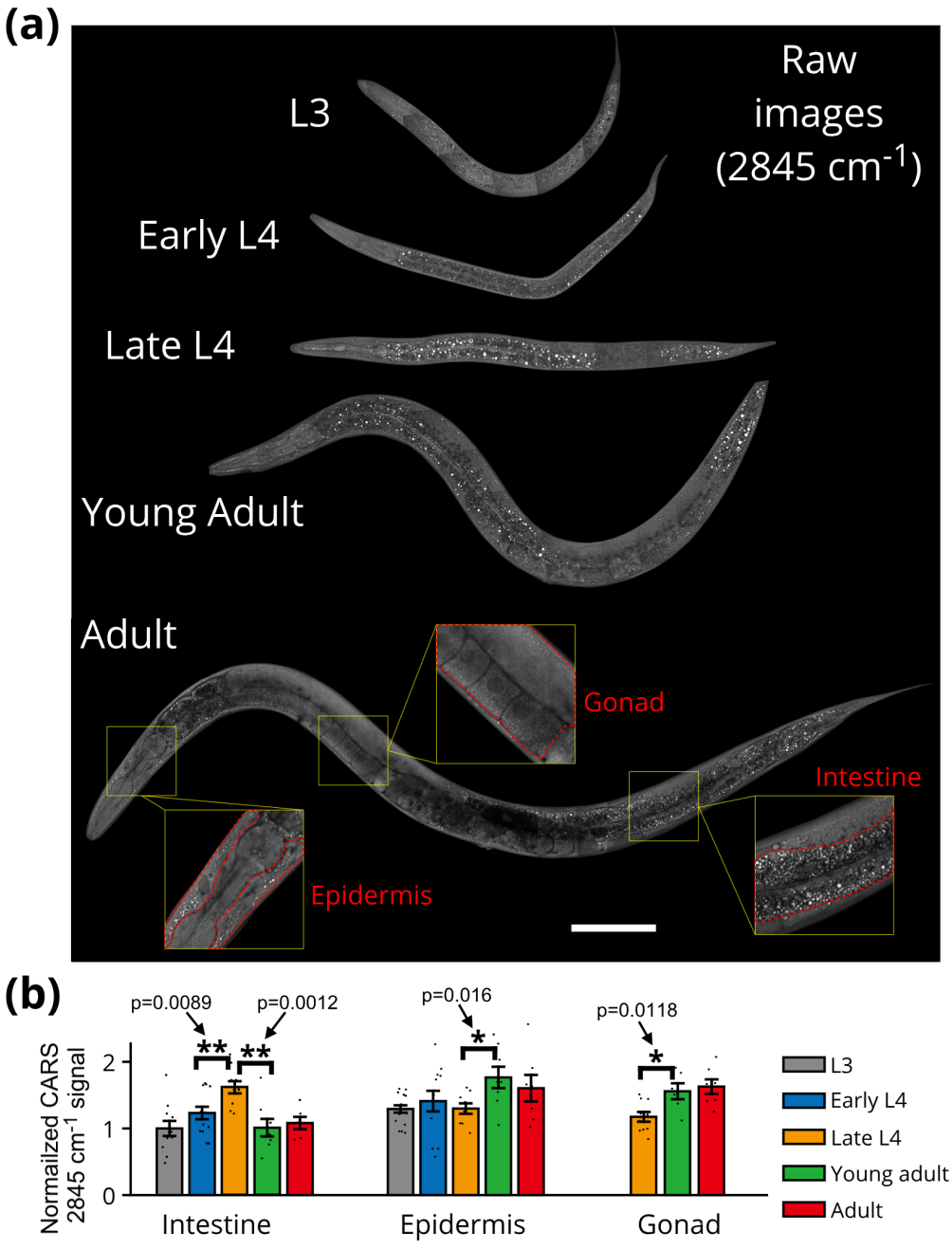
Goodness-of-fit statistics

	Weibull	gamma	normal	lognormal	exponential
Kolmogorov-Smirnov statistic	0.07765785	0.05261234	0.09892575	0.03090924	0.3382718
Cramer-von Mises statistic	1.61725779	0.50556929	2.57076208	0.09427251	23.2745791
Anderson-Darling statistic	10.93534792	3.32296709	16.09885932	0.76446596	117.0150977

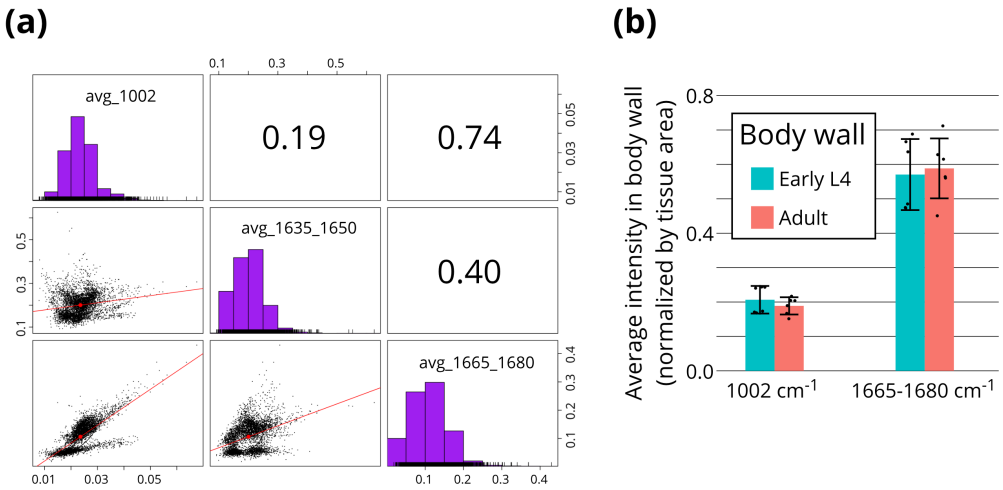
Goodness-of-fit criteria

	Weibull	gamma	normal	lognormal	exponential
Akaike's Information Criterion	376.6384	287.7129	452.7994	270.1314	1023.549
Bayesian Information Criterion	385.7858	296.8603	461.9467	279.2788	1028.123

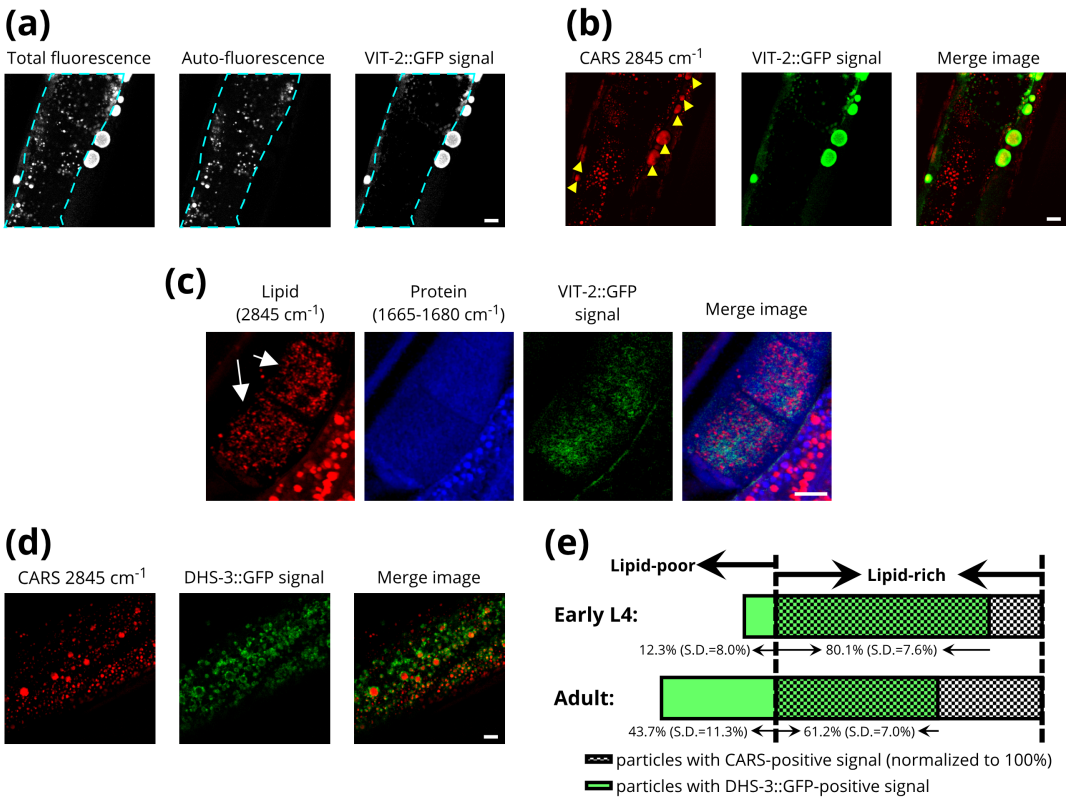
Supplementary Figure 1. Selection of lipid-rich particles based on the results of fixative lipid dye staining. (a) The CARS and TPEF images of a wild-type worm with lipidTOX green fixative staining. The stained lipid particles in the LipidTOX green fluorescence image were first converted into a binary mask. Then, the mask of selected particles was applied to CARS 2845 cm^{-1} images and the histogram of CARS 2845 cm^{-1} intensity was plotted. After distribution analysis and curve fitting, a 2845 cm^{-1} intensity threshold $(\mu^* \div (\sigma^*)^2)$ was determined. This threshold excluded the lipid-poor region in the CARS images. The particles with the CARS 2845 cm^{-1} intensity above this threshold are referred as the lipid-rich particles. (b) The comparison of the size of lipid-rich particles in live and fixative-stained worms. The CARS 2845 cm^{-1} images and the density plot of particle sizes show that larger lipid-rich particles can be observed 30 minutes after fixative staining. Feret's diameter is a measure of an object size along its longest axis direction (or the maximum diameter). Scale bar = 10 μm. The experiments were repeated three times independently with similar results. (c) The detailed reports of distribution analysis (fitdistrplus package version 1.0-14 , R programming). Total 716 lipid-rich particles obtained from n=3 biologically independent animals were plotted in the Cullen and Frey graph and analyzed.



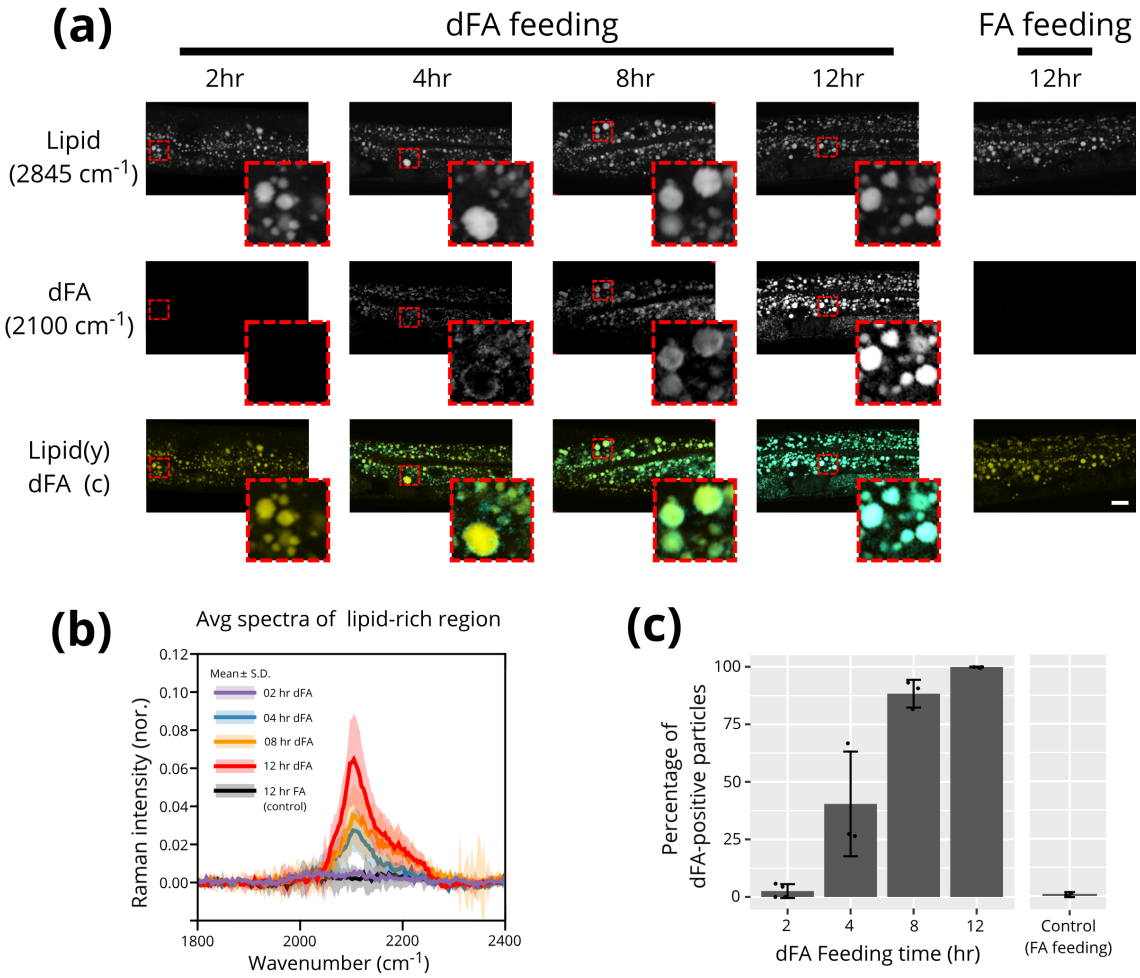
Supplementary Figure 2. Lipid distribution at different developmental stages. (a) The CARS raw images (at 2845 cm^{-1}) of worms at different developmental stages. The non-resonant background (NRB) that accompanies with CARS signal can provide a bright-field-like scattering image that helps morphological recognition in the raw CARS images. Three specific tissue including skin-like epidermis near pharynx, gonad, and intestine were examined. Scale bar = 100 μm. (b) The normalized 2845 cm^{-1} signal of tissues (normalized by the area of tissue) at different developmental stages. Each measurement is presented as a dot. The error bars represent the standard error of the mean. P-value: *p < 0.05, **p < 0.01, two-tailed Student's t-test was used. From developmental time L3 to adult, n=11, 13, 10, 8, 6 for intestine and n=17, 12, 9, 8, 7 biologically independent animals, respectively. For gonad, n=10, 6, 7 biologically independent animals from late L4 to adult, respectively.)



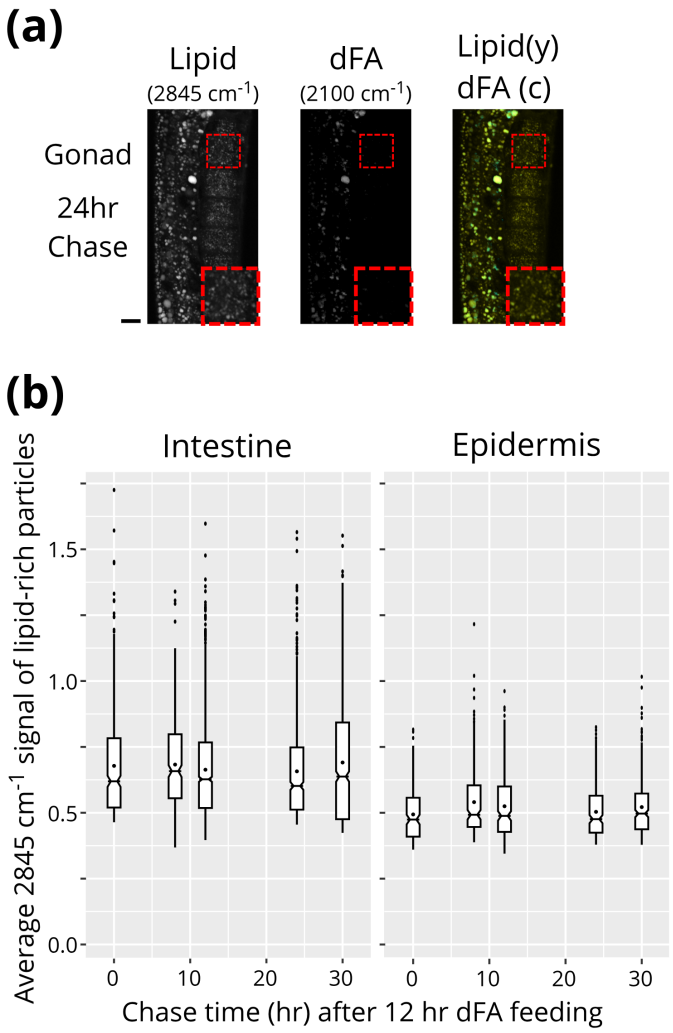
Supplementary Figure 3. The control measurements for protein-related signals. (a) The correlation between the average intensity per particle for phenylalanine at 1002 cm^{-1} (avg_1002), C=C at 1635-1650 cm^{-1} (avg_1635_1650), and amide I at 1665-1680 cm^{-1} (avg_1665_1680). Total 4700 lipid-rich particles obtained from n=10 biologically independent animals (n=5 for early L4 and n=5 for adult worms) were plotted. The three plots on the lower-left side are the scatter plots between them, where the red lines represent linear regression. The numbers shown on the upper-right side represent Pearson correlation coefficients (r), where r=1 for a perfectly positive linear relationship; r=0 for no relationship; r=-1 for a perfectly negative linear relationship. (b) The average intensities of protein-related signals (1002 cm^{-1} and 1665-1680 cm^{-1}) in the body wall of both early L4 and adult worms (n=6 biologically independent animals for both). Both protein-related signals were at similar level between the two developmental stages examined. The intensity was normalized by the tissue area. Each measurement is presented as a dot. The error bars represent the standard deviation.



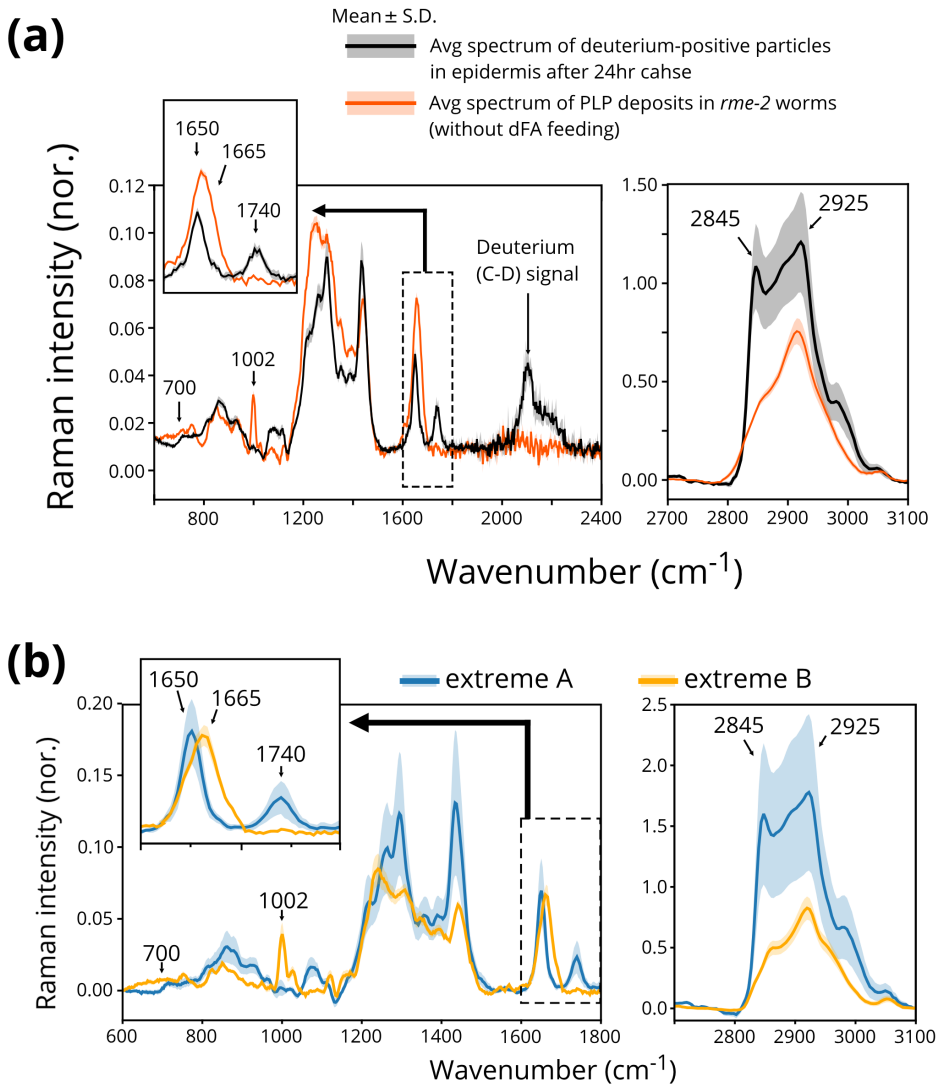
Supplementary Figure 4. Two-photon excitation fluorescence (TPEF) and CARS (at 2845 cm^{-1}) imaging of wild-type worms expressing VIT-2::GFP and DHS-3::GFP. (a) The TPEF images of a wild-type gravid worm. The auto-fluorescence signal can be removed either by using the fluorescent life-time gated imaging because the lifetime of autofluorescence (typically < 1 ns) is much shorter than that of GFP (∼3 ns), or by the subtraction of the region with auto-fluorescence signal appeared at 560-600 nm (for this case). (b) The CARS lipid signal overlapped well with VIT-2::GFP signal in the PLP region (indicated by yellow arrows). However, after excluding the interference of auto-fluorescence, the VIT-2::GFP signal was weak in the intestine. Moderate VIT-2::GFP signal was found being restricted to the baso-lateral surface. (c) Most of the VIT-2::GFP signal was not overlapping with CARS lipid signal in the oocytes. About 1200 (for adult worms) and 760 (for early L4) intestinal particles were collected (from three different experiments) and analyzed. For (a) to (c), the experiments were repeated three times independently with similar results. (d) The overlaid images of 2845 cm^{-1} signal (red) and GFP signal (green) of a wild-type worm expressing DHS-3::GFP, a GFP mark for lipid droplet associated protein DHS-3. (e) Normalized average percentage of lipid-rich particles (CARS-positive) and DHS-3::GFP-positive pariticles (green). About 770 and 1200 particles were collected from early L4 (n=4) and adult (n=3) worms, respectively, and analyzed. Scale bar = 10 μm for (a) to (d).



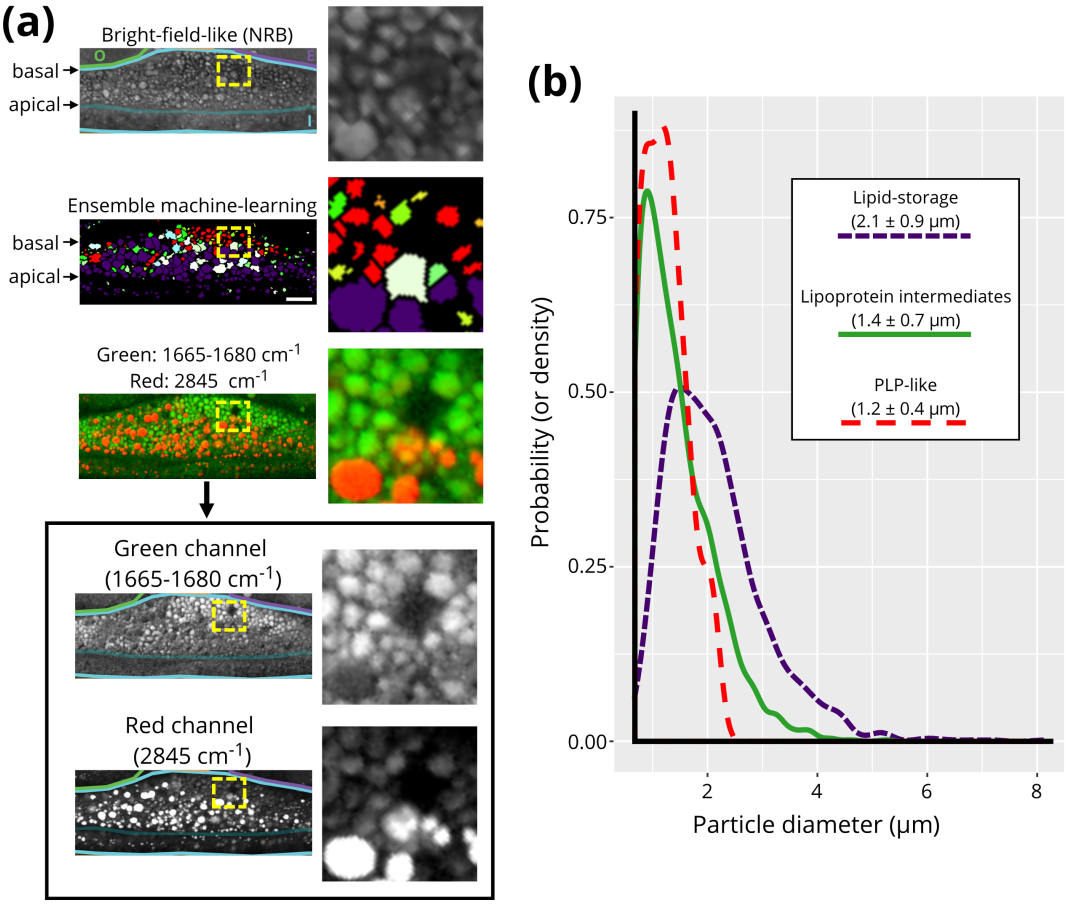
Supplementary Figure 5. The deuterium labeling of digested lipids. (a) The CARS images of total lipids (2845 cm^{-1}) and dueterium-labeled lipids (dFA, 2100 cm^{-1}). The fatty acid without deuterium (FA) was used as a control. The third row images are the overlaid images of 2845 cm^{-1} (yellow) and 2100 cm^{-1} (cyan) images. Scale bar = 10 μm. The experiments were repeated three times independently with similar results. (b) The retrieved Raman spectra of deuterium (C-D) signals and (c) the average percentage of deuterium-labeling particles. At each time point about 600-1500 particles collected from n= 4, 3, 3, 3, 2 biologically independent animals for 2, 4, 8, 12 hr dFA feeding and 12hr FA feeding, respectively. Each measurement is presented as a dot. The error bars represent the standard deviation.



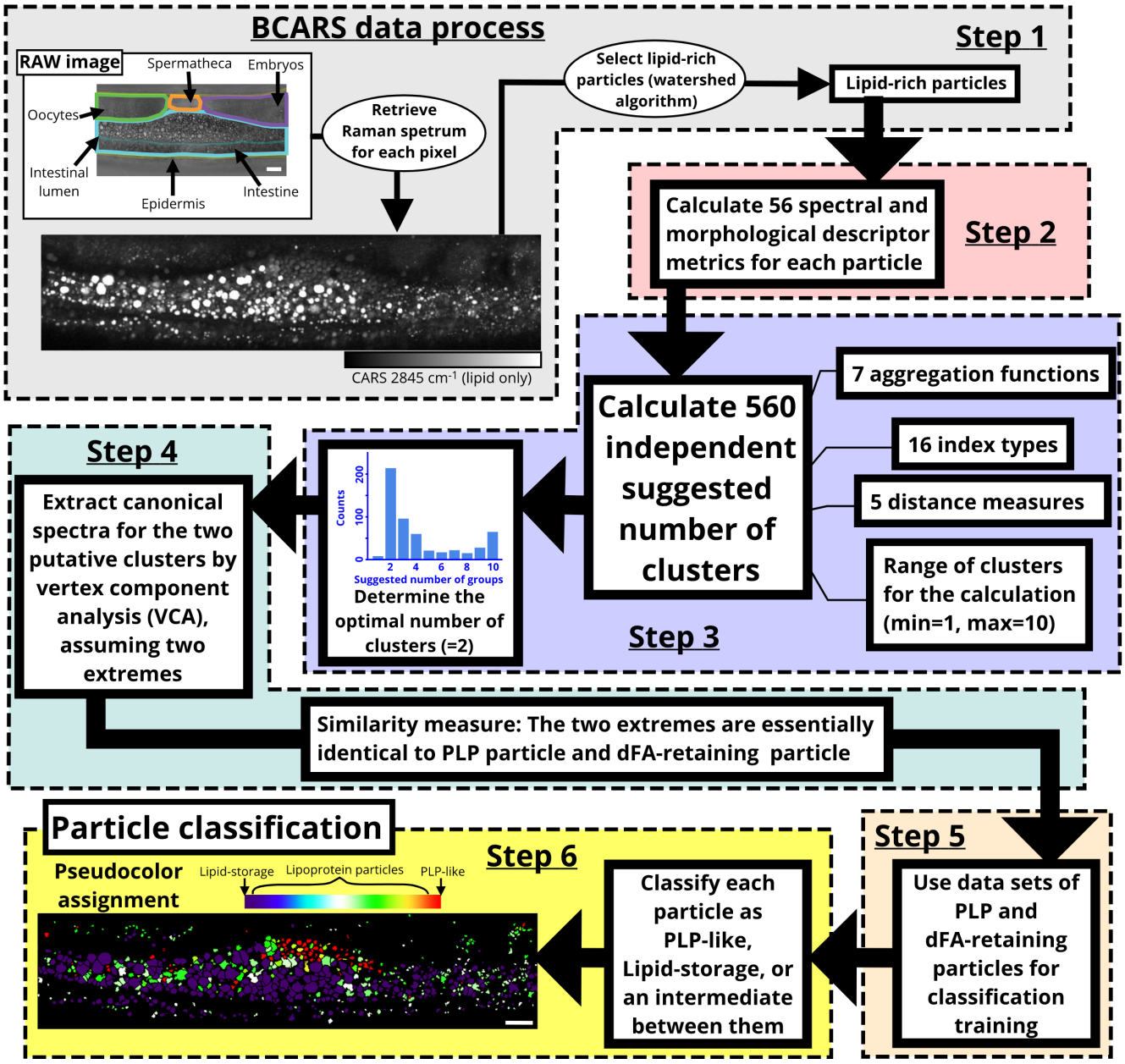
Supplementary Figure 6. The lipid and dFA signals at chase time points. (a) Lipid and dFA images of gonad at 24 hour chase time. CARS lipids (2845 cm^{-1}), dFA (2100 cm^{-1}), and overlaid images of gonad at 24 hour chase time point. Scale bar = 10 μm. The experiments were repeated three times independently with similar results. (b) Average 2845 cm^{-1} signal of lipid-rich particles in intestine and skin-like epidermis at different chase time points. At each time points, at least 1000 particles were collected from n=3 biologically independent animals, and were analyzed. The boxes denote 25th to 75th percentiles; The whiskers indicate the maximal data points present within 1.5 $\times$ the interquartile range (IQR) from the top and bottom of the boxes; The notch displays the 95% confidence interval around the median (the line across the notch in the box) and the dot in the box represents the mean value.



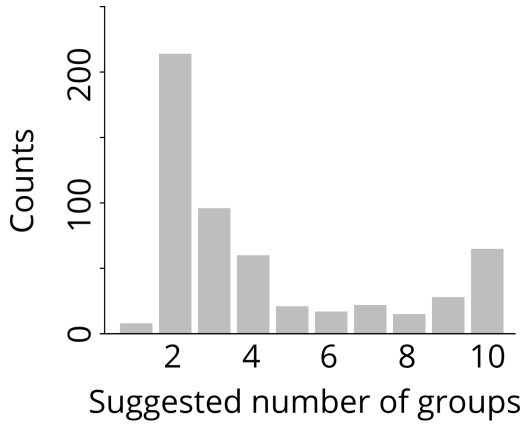
Supplementary Figure 7. The average Raman spectra of deuterium-positive particles, PLP deposits, and two canonical spectra of the two extremes extracted by performing vertex component analysis (VCA). (a) The retrieved average Raman spectra of deuterium-positive particles in epidermis after 24-hour chase (black line, total 217 particles obtained from n=4 biologically independent animals) and the PLP accumulation in *rme-2* mutants (orange line, total 54 particles obtained from n=3 biologically independent animals). (b) The spectra of two extremes extracted by VCA. Each spectrum represents the average Raman spectrum (mean ± 1 s.d.) of the single lipid-rich particle (extreme A or extreme B).



Supplementary Figure 8. The images and particle sizes after applying ensemble machine-learning analysis. (a) The non-resonant background (NRB), protein (1665-1680 cm^{-1}, and lipid (2845 cm^{-1} images, as well as the image of pseudo-color assignment after applied ensemble machine-learning analysis. In the bright-field-like (NRB) image, the intestine, embryo, and oocyte are denoted by "I", "E", "O", respectively. The raw images containing NRB that can be used to identify the different tissues of *C. elegans*. Since the NRB signal provides no chemical contrast, it can be used as a bright-field-like signal for morphological recognition of tissues. Some particles shown in the NRB image can be only found in either lipid (2845 cm^{-1}) or protein (1665-1680 cm^{-1} image, suggesting chemically distinct compositions in these particles. The experiments were repeated five times independently with similar results. (b) The size distribution of classified particles in the intestine of adult worms. About 2300 particles collected from n=3 biologically independent animals were classified and compared. The numbers shown in the legend represent the average diameter (mean ± s.d. μm).



Supplementary Figure 9. The workflow of ensemble machine-learning analysis. The 56 spectral and morphological descriptor metrics for each particle included the amplitude of assigned Raman bands at different wavenumbers (1002 cm^{-1}, 1635-1650 cm^{-1}, 1665-1680 cm^{-1}, 1740 cm^{-1}, 2845 cm^{-1}, and 2925 cm^{-1}) and the peak ratios (1002cm^{-1} / 2845cm^{-1}, 1665-1680cm^{-1} / 2845cm^{-1}, and 2925cm^{-1} / 2845cm^{-1}) were first retrieved. Then, the skew and kurtosis of intensity histogram, mean intensity, median intensity, standard deviation of intensity, and integrated intensity per particle at different wavenumbers and ratios, plus the particle size and the Feret's diameter, were measured. For each particle in the data set, total 56 spectral and morphological descriptors were calculated. These descriptors not only reflect the chemical composition and concentration of the particle (from various Raman peaks) but also contain the information of physical properties including distribution of different species (the skew and kurtosis of intensity histogram) and particle size (area and Feret's diameter). See Method for the details for each step. Scale bar =10 μm.



Supplementary Figure 10. The histogram of suggested number of groups aggregated from 560 independent calculations by performing NbClust (v 3.0, R programming). These analysis indicated that the optimal number of groups in the pooled data of wild-type worms is 2. About 10^4 particles were calculated.