Supplementary Material to:

Combining Tine-of-Flight Secondary Ion Mass Spectrometry imaging mass spectrometry and CARS microspectroscopy reveals lipid patterns reminiscent of gene expression patterns in the wing imaginal disc of *Drosophila melanogaster*

Authors: Florian Marty[1]†, Gianluca Rago[2,3], Donald F. Smith[2], Xiaoli Gao[4], Gert B. Eijkel[2,5], Luke MacAleese[2]††, Mischa Bonn[2,3], Erich Brunner[1], Konrad Basler[1]*, Ron M.A. Heeren[2,5]*

Affiliations:

[1]Institute of Molecular Life Sciences, University of Zürich, Winterthurerstrasse 190, CH-8057 Zürich, Switzerland

[2]FOM-Institute AMOLF, Science Park 104, 1098 XG Amsterdam, The Netherlands

[3]Max Planck Institute for Polymer Research, Ackermannweg 10, 55128 Mainz, Germany

[4]Institutional Mass Spectrometry Laboratory, The University of Texas Health Science Center at San Antonio, 8403 Floyd Curl Drive, MC-7760 San Antonio, TX, USA

[5]The Maastricht Multimodal Molecular Imaging Institute, Maastricht University, Universiteitssingel 50, 6229 ER Maastricht, The Netherlands

*Correspondence to: r.heeren@maastrichtuniveristy.nl, konrad.basler@imls.uzh.ch

Current address:

† Biognosys AG, Wagistrasse 21, 8952 Schlieren

†† Groupe de Spectrométries des Biomolécules et Agrégats
Institut Lumière Matière
Cité Lyonnaise de l'Environnement et de l'Analyse
5 rue de la Doua, 69100 Villeurbanne, France

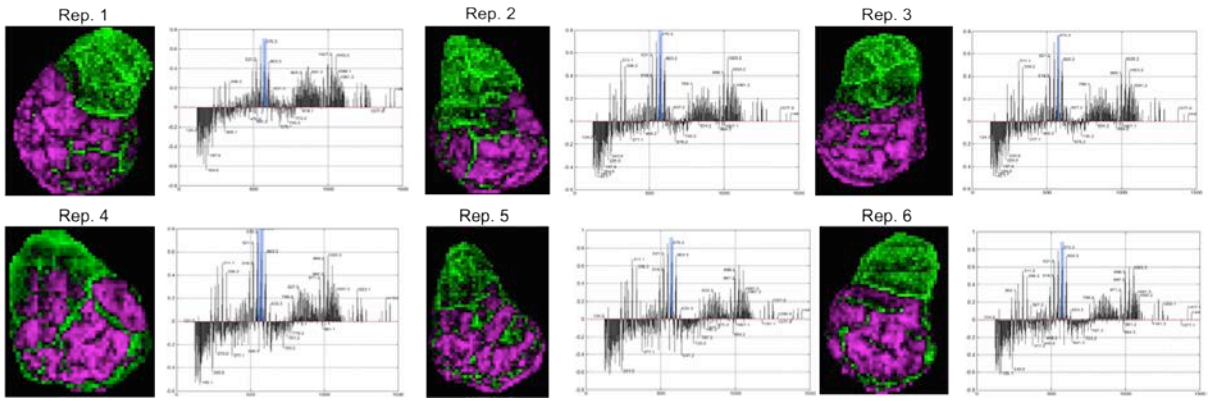

Figure S1: Repetitive measurement of individual wing imaginal discs to confirm the B/W pattern in positive ion mode. Green = body wall, purple = wing blade

		CARS	SIMS
CV	Correlation	Variance described	
1	0.394264	4.984	0.69102
2	0.288557	3.600	0.54064
3	0.057636	3.209	0.50543
4	-0.048119	9.538	0.97901
5	-0.020349	7.359	0.34812
6	-0.021041	1.867	0.72842
7	-0.005737	2.181	0.39812

Supplementary Table 1. Canonical correlation analysis (CCA) scores for patterns observed by CARS and ToF-SIMS. Correlation was examined with two independent methods, both indicating that the first two CV significantly correlate.

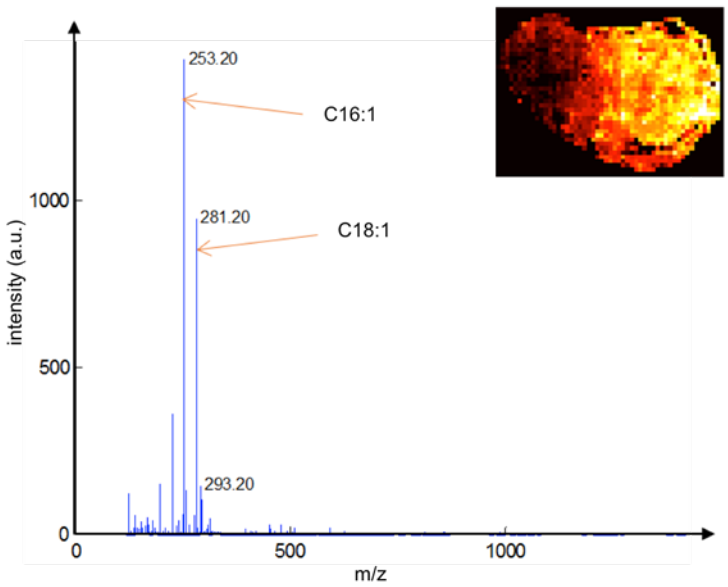

Figure S2: The wing blade part of the loadings spectra of the negative ion mode ToF-SIMS PCA is dominated by the peaks of the two fatty acids C16:1 and C18:1 further confirming the identity of the DAG(34:2) identified in positive ion mode and orthogonally validated with the LC-MS/MS measurements.

Data analysis

ToF-SIMS data: The TRIFT-associated software WinCadence (v4.4 Physical Electronics, Chanhassan, USA) was used to pick peaks on the total average mass spectrum and produce manually mass-selected ion images of the wing discs. RAW images were converted to datacubes with the MSTools software suite developed at AMOLF (http://www.maldi-msi.org, in imzML / Software Tools). The datacube format enables direct observation of mass-selected images while browsing through *m/z* values, and permits fast region-of-interest (ROI) selection accompanied by comparison between ROI average mass spectra. For principal component analysis (PCA), the MATLAB-based ChemomeTricks[1] software package was used, as previously described. Briefly, RAW files were first converted to 1 × 1 pixel into .mat file formats to generate an average spectrum with a mass resolution of 0.1 Da for the entire wing disc. The average mass spectrum was peak-picked and approximately 1000 peaks were stored for every dataset. The peak list together with the initial RAW file was reconverted with 64 × 64 pixel spatial resolution for PCA. Output of the TRIFTtricks software was manually investigated for patterns reminiscent of underlying known patterns. PCA patterns reminiscent of biologically known patterns were displayed in false color images.

CARS data: The quantitative information contained in the vibrational resonant term of the CARS spectra was hidden by the convolution with a non-resonant term arising from the electronic response of the material. The imaginary part of the third-order nonlinear susceptibility, Im[$\lambda(3)$], was retrieved from the raw CARS spectra by the maximum entropy method, as described elsewhere [2]. Im[$\lambda(3)$] spectra, referred to as retrieved CARS spectra, are directly comparable to spontaneous Raman spectra. Thus, the peak amplitude or integrated intensities are proportional to the number of scatterers in the focal volume.

Data depth alignment

Static SIMS measurements image the topmost layer of the sample, independent of slight differences in height at the sample surface. Confocal CARS measurements, however, image a focal plane independent of differences in surface height. In order to create datasets that reflect information from similar height regions as precise as possible, five CARS depth slices were used. Of each subsequent CARS layer starting at 0 μm with a 2-μm step size, spectra with a summed intensity significantly higher than the background signal were stored. Combination

of these data from the lowest level up to the highest level produced the CARS dataset that was used for PCA and Canonical Cross Correlation Analysis (CCA).

Data spatial alignment

To be able to perform CCA, these data must be spatially aligned as optimally as possible. Therefore, the SIMS data were translated, rotated and scaled using MATLAB tools to obtain maximum overlap with the CARS dataset. The final data-preprocessing step involved the selection of the on-tissue pixel based on the total intensity obtained by the CARS measurement. All signals originating outside the tissue were removed from the datasets.

Principal component analysis (PCA)

PCA is a mathematical technique that describes the maximum amount of variance in a minimal amount of functions, the PCs. These PCs are linear combinations of the original variables (in this case, *m/z* channels for ToF-SIMS and wave numbers for CARS). The correlation between each of the original variables and the calculated PCs is given by the PC loading. A plotted version of all loadings for a given PC is often used as graphical representation of that PC, the loading spectrum. By calculating the projection of each of the original spectra on the PC, we obtain PC scores. The scores are used for creating a PC score image. PCA typically reduces the number of variables in a dataset to a limited number of PCs that contain the major part of the variance in the data. The calculated PCs are ranked based on their total length, their Eigenvalue (EV). The number of relevant PCs is determined by using methods as the removal of PCs with EVs less than 1 [3], using either a Scree plot [4] or choosing the *n* highest-ranked PCs that account for a certain percentage of the total variance. PCA is a very efficient data-compression technique as well as a noise-reduction technique since noise in the data is uncorrelated. Uncorrelated signals contain very little variance and thus will be described in the lower-ranked PCs. Removal of these lower-ranked PCs using one of the above-mentioned methods results in a smaller reconstructed dataset containing fewer variables and less noise.

Canonical correlation analysis (CCA)

CCA is a mathematical technique that determines the relationship between two separate sets of variables. The PC score matrices from ToF-SIMS and CARS datasets were used to calculate their cross-covariance matrix. PCA was performed on this cross-covariance matrix to obtain a new set of independent linear combinations of the original variables (in this case, PC scores), the Canonical Variates. The Canonical Variates represent the common data space

for both datasets. Thus, the scores for both datasets can be calculated in the same data space and directly compared by calculating the correlation. Spatial binning of the data was necessary to increase the signal-to-noise ratio and to reduce computer memory requirements. For ToF-SIMS, the PHI RAW data format was converted to MATLAB/ChemomeTricks format using the in-house–built sims2tricks software. Peak picking was performed on the data using the PEAPI algorithm[1].

For CARS, data sets were exported from the analysis software Igor Pro (Wavemetrics, USA) as text data. The text data was converted to MATLAB/ChemomeTricks data format using the in-house–developed CARS2Tricks software. The separate CARS measurement tiles were stitched using the in-house–developed StitchCARS software.

Lipid analysis by HPLC-ESI-MS/MS

Progenesis CoMet (Nonlinear Dynamics Limited, Newcastle, UK) was used to process the raw data files. Peak alignment and integration were performed, and the relative abundance was generated. Lipid species were identified by searching the following databases: METLIN (http://metlin.scripps.edu/index.php); lipid maps (http://www.lipidmaps.org/data/structure/); Human Metabolome Database (http://www.hmdb.ca/) using a 5-ppm mass tolerance. The putative lipid identifications were manually verified through examination of the neutral losses and acyl chain diagnostic ions in the MS/MS spectra and by comparison with the retention times from commercially available standards (Avanti Polar Lipids, Alabaster, AL).

Supplemental references

1. Eijkel, G.B., Kaletas, B.K., van der Wiel, I.M., Kros, J.M., Luider, T.M., Heeren, R.M.A.: Correlating MALDI and SIMS imaging mass spectrometric datasets of biological tissue surfaces. Surf Interface Anal. **41**, 675-685 (2009)
2. Vartiainen, E.M., Rinia, H.A., Muller, M., Bonn, M.: Direct extraction of Raman line-shapes from congested CARS spectra. Optics Express. **14**, 3622-3630 (2006)
3. Kaiser, H.: The application of electronic computers to factor analysis. Educational and psychological Measurement. **20**, 141-151 (1960)
4. Cattell, R.: The Scree test for the number of factors. Multivariate Rehavioral Research. **1**, 245-276 (1966)