Electronic Supplementary Material (ESI) for Analyst.

Colocalization of Fluorescence and Raman Microscopic Images for the Identification of Subcellular Compartments: A Validation Study

Supplementary Material

Sascha D. Krauß, Dennis Petersen, Daniel Niedieker,
Inka Fricke, Erik Freier, Samir F. El-Mashtoly,
Klaus Gerwert, Axel Mosig

1 Organelle specificity, cell line specificity, and confounders of Raman spectra

As Raman microscopy is conducted label-free, the image spectra obtained may potentially contain signals specific for different biologically relevant factors as well as signals that are irrelevant or even confounding. In fluorescence microscopy, conversely, one or few specific label(s) associated with one condition and one factor (e.g. subcellular organelle) is given in one image. It is important to notice that in a supervised classification scenario, Raman spectra may in principle be distinguishable with respect to biologically relevant factors (e.g. cell type, subcellular organelle), while at the same time being distinguishable with respect to irrelevant or confounding factors (e.g. date of experiment). Correspondingly, in order to assess the distinguishability of different factors, we trained classifiers and evaluated their accuracy in leave-one-sample-out validation in different settings:

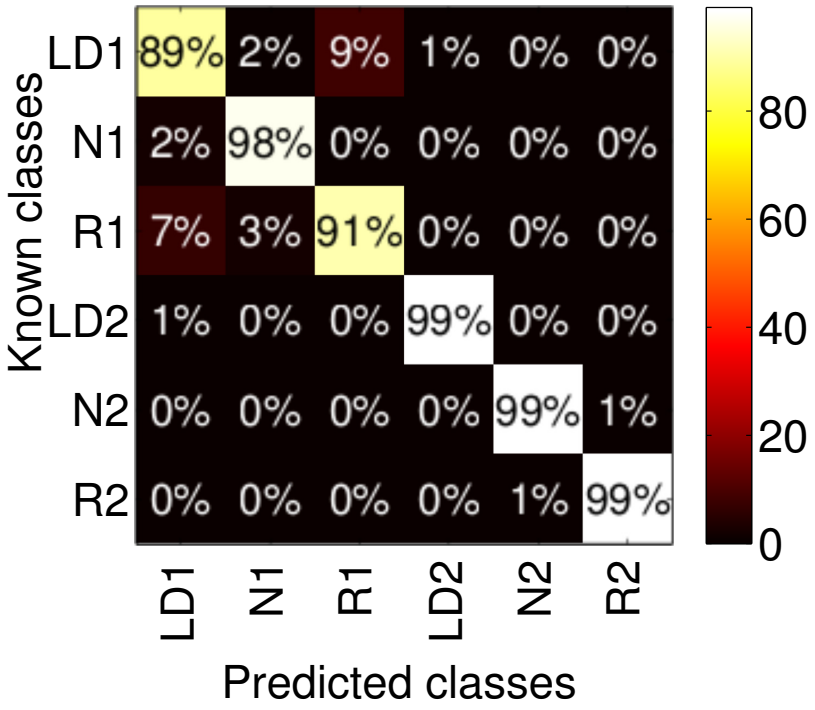
(C1) *Cell type specificity.* We trained a six-class-classifier where the classes for organelle types (nucleus, lipid droplets, rest) are further subdivided into cell-type specific classes. The resulting classifier achieves an accuracy of 98%, for a complete confusion matrix refer to Supplementary Figure 1.

(C2) *Cell type specificity of spectra from non-cellular surroundings.* Spectra from areas not covered by any cell in five Raman images of HT29 cells along with corresponding spectra in six images of MIA PaCa-2 cells were collected. A classifier to distinguish these spectra in the two types of samples achieved an accuracy of 99%.

(C3) *Confounder specificity of spectra from non-cellular surroundings.* We extracted spectra from areas not covered by any cell in Raman images of 26 MIA PaCa-2 cell samples, where 15 measurement were conducted on *Day 1*, and the reamining 11 conducted on *Day 2*. As it turns out, the spectra from the two experimental dates can be distinguished with an accuracy of 96%.

(C4) *"Null experiment"*. We trained a classifier on nucleus spectra of four HT29 cells vs. nucleus spectra from four MIA PaCa-2 cells. The labels were shuffled by flipping the labels of two

randomly selected cells of each class, resulting in an accuracy of 45%, matching the roughly 50% accuracy to be expected by chance. In particular, this supports the claim that the accuracies observed above are not due to overfitting.

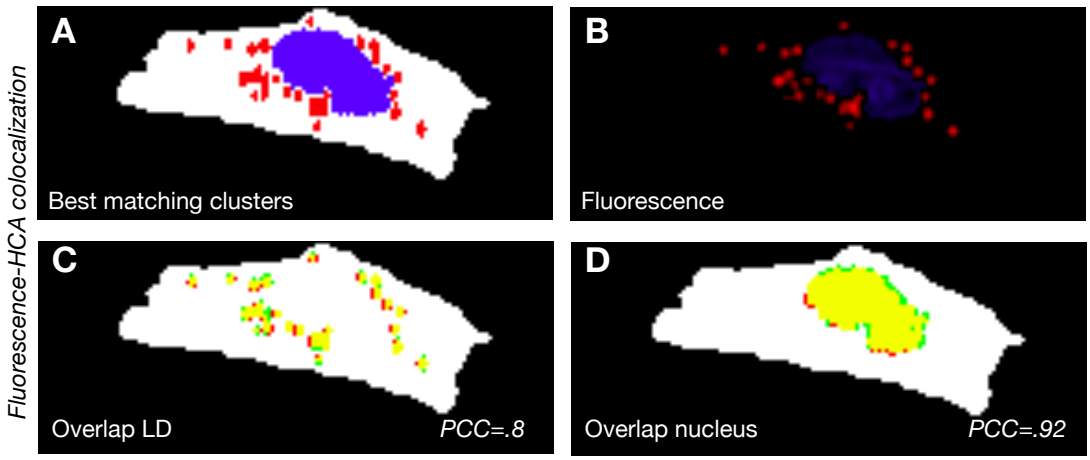


Supplementary Figure 1: *Row-by-row relative confusion matrix of a random forest trained on two cell lines with three organelles each.* The classes in this random forest are *N*ucleus, *L*ipid *D*roplets and *R*est, each class in group 1 from MIA PaCa-2 and group 2 from HT29 cells. The values were calculated according to the concept of leave-one-sample-out cross-validation, the sensitivities can be seen in the diagonal.

2 Effect of background correction on segmentation

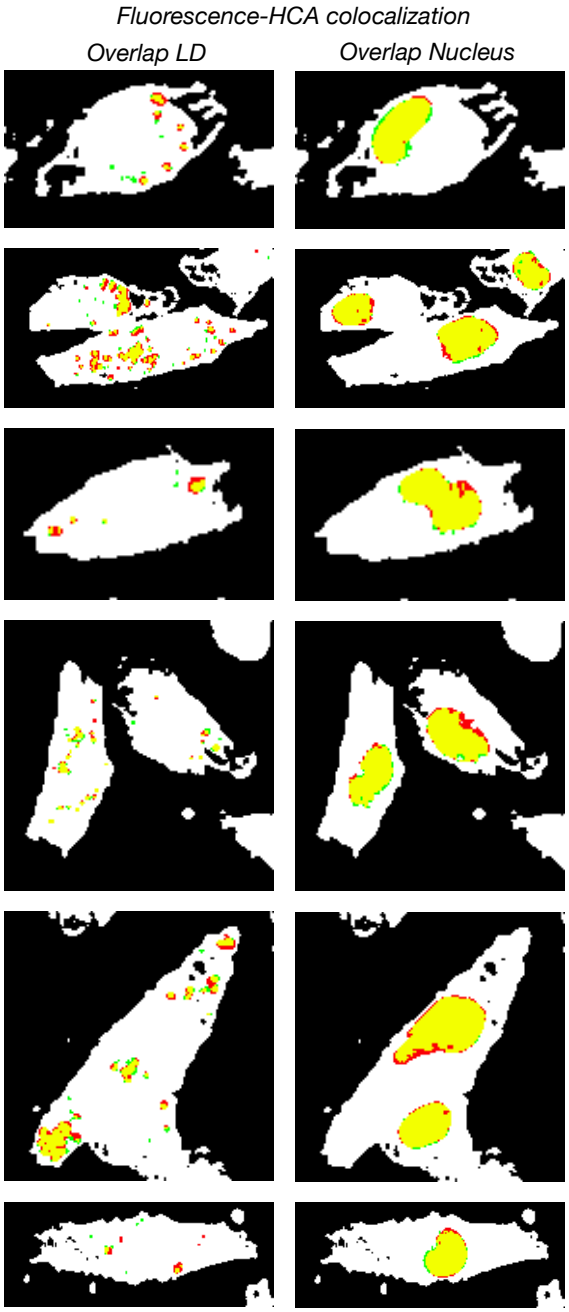
Several groups have shown that there is no or a weak fluorescence background in Raman measurements of cells [3, 1] and that cluster analysis on normalized data allows a good distinction for a certain number of cell compartments [2]. Nevertheless, we compared cluster analysis on baseline corrected and uncorrected data and concluded that there are minor differences for our setup on cell measurements (Supplementary Figure 2, Figure 3 in the main paper). No peaks are present in the region selected for normalization and data analysis of the substrate.

Our baseline correction of Raman spectra from cell measurements in media was executed in two steps. At first the water band at 1645 cm^{-1} and 3100-3600 cm^{-1} was subtracted and in the second step the residual baseline was removed by a polynomial baseline correction. Afterwards the spectra are normalized in a region between 700 cm^{-1} and 3100 cm^{-1}. For the subtraction of the water band endmembers of a Vertex Component Analysis (VCA) were calculated from a set of pure buffer spectra. In a least square solution these endmembers were fitted to the single spectra and subtracted from them. In the second step a sweep algorithm was applied on wavelet-denoised spectra (Daubechies wavelet D4) to gain supporting points for the polynomial baseline correction. A fifth order polynomial was fitted to these supporting points and substracted from the spectra.



Supplementary Figure 2: *Effect of background correction on segmentation.* Panels *A* and *B* display best matching clusters and fluorescence, respectively. Panels *C* and *D* show overlays between the best matching cluster (green) and the corresponding thresholded fluorescence image (red) along with the PCC between the two. The PCC values of .8 and .92 compare to PCCs of .79 and .93, respectively, obtained for uncorrected spectra (see Figure 3 in the main paper).

3 Dataset for nucleus and lipid droplet identification



Supplementary Figure 3: Complete dataset used for the identification of nucleus and lipid droplets. The left column displays overlays between the best matching cluster (green) and the corresponding thresholded fluorescence (red) for lipid droplets, the right column shows the overlay for nucleus.

4 Parameter-free extraction of training spectra

In this supplementary section, we provide details on the extraction of spectra for training supervised classifiers. In our approach, the only input utilized for extracting training spectra are

- The dendrogram D and
- the fluorescence image I.

The output of the function consists of a set of training spectra. Extracting training spectra involves no further parameters, in particular no fluorescence threshold or number of clusters.

When extracting training spectra from a fluorescence image I and spectral image J, we perform hierarchical clustering using Ward's method on the pixel spectra in J to obtain a *dendrogram* D. If N denotes the number of pixels in J, dendrogram D is a binary tree with N leaf vertices and $N-1$ internal vertices, i.e., $2N-1$ vertices in total. As a notational convention, we let V denote the set of all $2N-1$ vertices. Each vertex $v \in V$ is associated with a binary image $G(v)$, which exhibits intensity 1 (white) at the coordinates of all pixel spectra at the leaf vertices below v, and intensity 0 (black) at all other coordinates. The fluorescence image J can be turned into a binary image by setting a threshold t, so that the thresholded image $R_I(t)$ has intensity 1 at all positions where the fluorescence intensity in I exceeds t, and 0 at all other positions.

4.1 Determining a best matching cluster.

Our approach starts with determining a *best matching cluster* in D. To this end, all cluster images $G(v)$ for *all possible* $v \in V$ constitute potential best matching clusters, as well as the thresholded images $R_I(t)$ for *all possible* thresholds t constitute potential thresholds, where both v and t are to be determined automatically without utilizing any further parameters. Note that for any given pair of a vertex v and threshold t, we can determine the Pearson correlation $PCC(v,t)$ between $G(v)$ and $R_I(t)$ using Equation (1) from the main paper. For pairs (v,t) where the images $G(v)$ and $R_I(t)$ display highly similar areas, $PCC(v,t)$ will yield a correlation close to 1. If the cluster image and the thresholded image disagree, the correlation will be close to 0 or even negative. To identify the highest possible degree of colocalization between I and J, an optimal pair of vertex v_{opt} and threshold t_{opt} are obtained through

$$(v_{\text{opt}}, t_{\text{opt}}) = \arg\max_{v \in V, t} \; PCC_{v,t}. \tag{1}$$

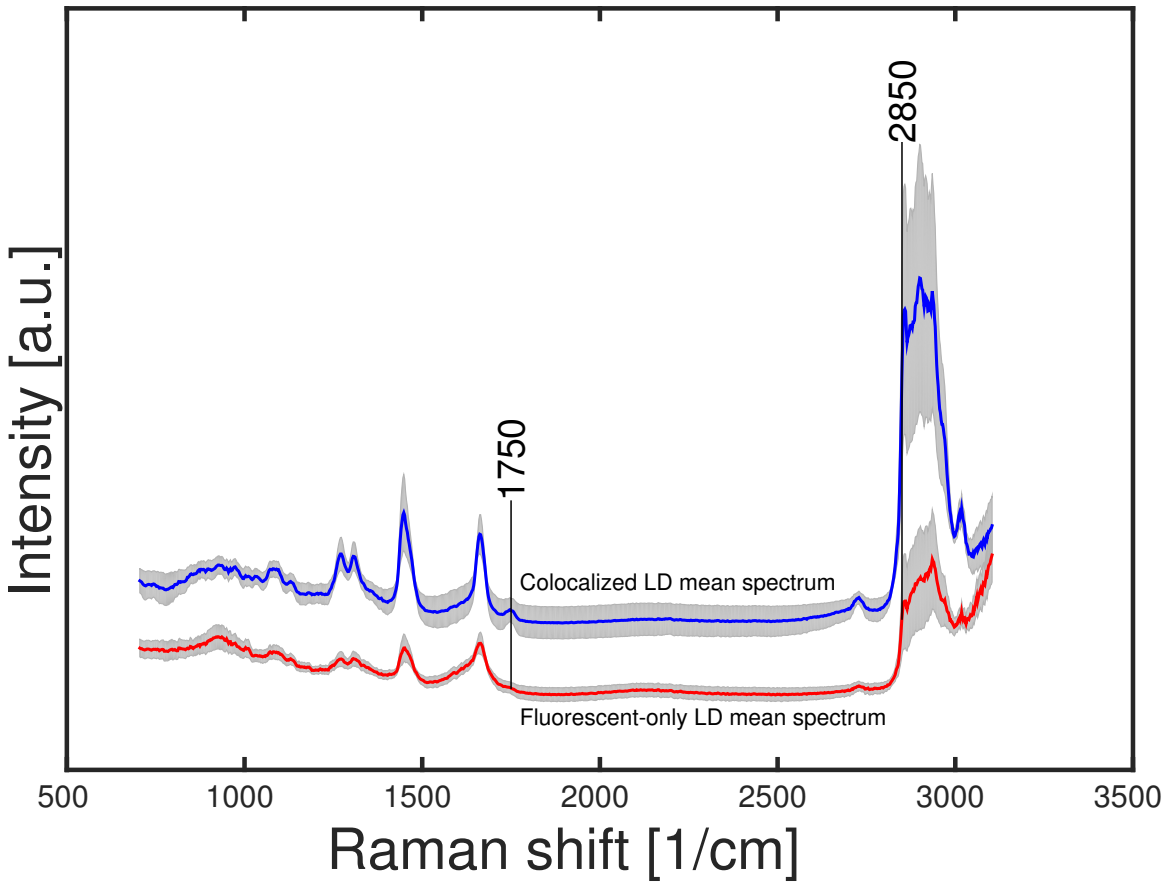
The vertex v_{opt} obtained using Eqn. (1) is the formal definition of what is referred to as the *best matching cluster* in the main paper. Due to the statistical origin of Pearson correlation – and as well-established in the corresponding literature on colocalization between fluorescence images (References 1,26,27 in the main paper) – a value of $PCC(v,t)$ close to 1 allows the interpretation that there is a statistically significant high degree of agreement between I and J.

It is important to notice the following:

- No user-adjustable parameters are involved in determining the best matching cluster.
- Identification of the best matching cluster is performed separately for each fluorescence image (i.e., fluorescently labelled organelle) for each sample.
- Once a vertex v has been identified with an organelle, v as well as all descendant vertices of v in the dendrogram are removed and thus cannot be identified with other organelles.

4.2 Obtaining training data from best matching clusters

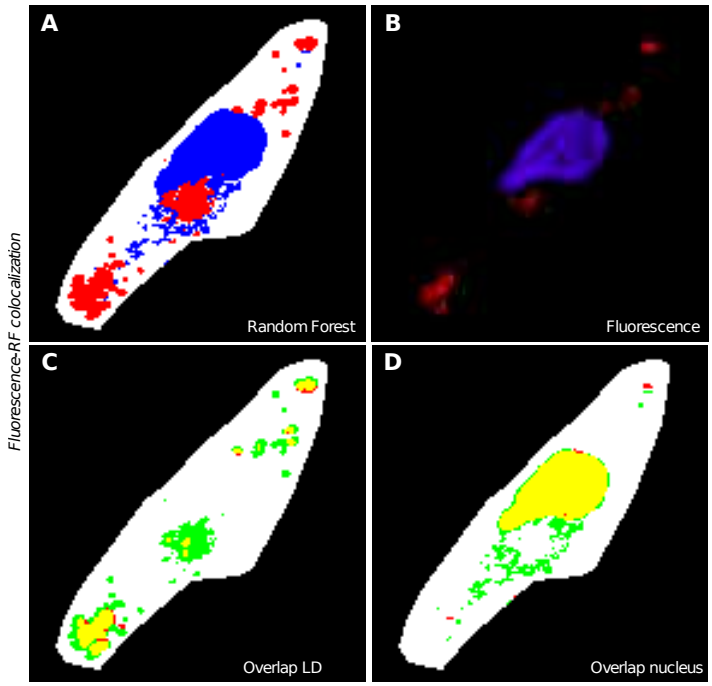
In our approach, training data for a given subcellular organelle are obtained on the grounds of the best matching cluster. More precisely, we extract training spectra from all positions where both the thresholded fluorescence image $G_I(t_{\text{opt}}) = 1$ and the cluster image $R(v_{\text{opt}}) = 1$, corresponding to the *yellow* pixels in panels **C** and **D** of Fig. 3 in the main paper.



Supplementary Figure 4: *Blue spectrum:* Mean spectrum of training spectra obtained from *colocalization* of fluorescence and best-matching cluster. The spectra exhibit strong lipid bands at 1750 and 2850 cm^{-1}. *Red spectrum:* Corresponding mean spectrum from fluorescence positions not colocalized with the best-matching cluster. Obviously, these spectra do not exhibit clear signals of lipid droplets, as in particular the peak at the lipid band 1750 cm^{-1} is not present, and the peak at 2850 cm^{-1} only very weak. The shading of the spectra shows the standard deviation.

Note that *red* pixels in these overlap images correspond to pixels where fluorescence is above threshold, i.e., $R(v_{\text{opt}}) = 1$, but which are not covered by the best matching cluster. We do not include such spectra in the training data set. The motivation to do so is that overlap between fluorescence and Raman image will never be perfect due to at least three factors, namely (i) differences in confocal volume between Raman and fluorescence microscopy, (ii) slight deviations in the z-layer, and (iii) small distortions of the sample due to fluorescence staining. Consequently, some positions in border regions of the organelles that exhibit fluorescence cover positions in the spectral image where the organelle is actually *not* present. This will obviously lead to false training spectra, and reduce the

accuracy of the resulting classifier. Overlap with the best matching cluster yields, conversely, training spectra that exhibit a high degree of spectral consistency, so that these positions of misaligned overlap can be expected not to be part of the best matching cluster.



Supplementary Figure 5: Classification result of classifier trained on fluorescence above-threshold positions directly in comparison with fluorescence labeling. The cell shown is the same cell as Fig. 3 in the main paper. *(A):* Classification result with nucleus positions indicated in blue and lipid droplets in red. The lipid droplets close to the nucleus are obviously not well resolved, as well as larger regions that are obviously not part of the nucleus are recognized as nucleus. This leads to a drop in the PCC between the predicted lipid droplet regions and the respective fluorescence image (panel *(B)*) from .79 to .5 (see panel *(C)* for an overlay). The PCC between predicted nucleus and unthresholded nucleus fluorescence drops from 0.96 to 0.79 (overlay shown in panel *(D)*).

Beyond the observed drop in accuracy, we assessed the influence of choosing overlap ("yellow") positions for training data compared to choosing all fluorescence foreground ("yellow and red") positions as training data in further detail. The spectra shown in Figure 4 indicate that spectra in the fluorescent non-overlap positions show no clear signals of lipid droplets. For either supervised classifiers or regression based approaches, falsely assigned spectra in the training data will obviously lead to weaker classifiers with reduced accuracy during cross-validation.

It is worthwhile to notice that high accuracies of classification are obtained only if the agreement between the best matching cluster and the fluorescence image is high across all data sets in the training data. As can be seen from the cross validation for perturbed data Figure 5 of the main paper, a low agreement between fluorescence and HCA will lead to low cross-validation accuracies.

References

[1] Christoph Krafft, Benjamin Dietzek, and Jürgen Popp. Raman and CARS microspectroscopy of cells and tissues. *Analyst*, 134(6):1046–1057, 2009.

[2] Miloš Miljković, Tatyana Chernenko, Melissa J Romeo, Benjamin Bird, Christian Matthäus, and Max Diem. Label-free imaging of human cells: algorithms for image reconstruction of Raman hyperspectral datasets. *Analyst*, 135(8):2002–2013, 2010.

[3] Henk-Jan van Manen, Yvonne M Kraan, Dirk Roos, and Cees Otto. Single-cell Raman and fluorescence microscopy reveal the association of lipid bodies with phagosomes in leukocytes. *Proceedings of the National Academy of Sciences of the United States of America*, 102(29):10159–10164, 2005.