Enhanced tissue compartmentalization by applying segmentation algorithms on reduced spatial omics data

Bram Heijs[1], **Ryan Marsico**[2], **Sören-Oliver Deininger**[3], **Arne Behrens**[3], **Janina Oetjen**[3], **Azad Esghi**[2], **Tobias Boskamp**[2]
[1]Bruker Nederland B.V., Leiderdorp, The Netherlands; [2]Bruker Scientific LLC, Billerica, MA, USA; [3]Bruker Daltonics GmbH & Co. KG, Bremen, Germany.

Introduction

A fundamental element of any MALDI Imaging data analysis workflow is segmentation, commonly known as spatial clustering. This process involves grouping pixels with similar biochemical properties into clusters, each assigned a numerical index for identification. By highlighting distinct morphological regions within tissues, clustering enhances our comprehension of tissue heterogeneity and the underlying mechanisms of diseases. Advancements in MALDI Imaging instrument technology have enabled the collection of data at improved spatial and chemical resolutions. The increased density of the data reveals limitations in traditional clustering algorithms. Conducting clustering on data that has undergone dimensionality reduction, such as component analysis (CA), improves computational efficiency. Since dimensionality reduction often minimizes noise, it enhances the clarity and effectiveness of clustering outcomes. This workflow illustrates how the integration of CA with clustering techniques can serve as a valuable approach for gaining deeper insights from MALDI Imaging data.

Figure 1: Schematic overview of the unsupervised multivariate statistics workflows in SCiLS Lab.

Methodology

MALDI Imaging and MALDI HiPLEX-IHC datasets were generated from rat kidney and human lung carcinoma tissues. Data was imported using a prototype version of SCiLS™ Lab 2026a. Digital scans of haematoxylin and eosin staining of the analyzed tissues were aligned with MALDI data. For MALDI HiPLEX-IHC, features were imported using a target list, while T-ReX® feature finding identified 150 features in the rat kidney dataset. The *k*-means algorithm (*k* = 10) was used as benchmark segmentation, with TIC-normalized feature intensities for MALDI Imaging data and MALDI HiPLEX-IHC data. Additional segmentation methods were evaluated using the updated multivariate statistics module in SCiLS Lab (**Figure 1**) and compared against the benchmark *k*-means results.

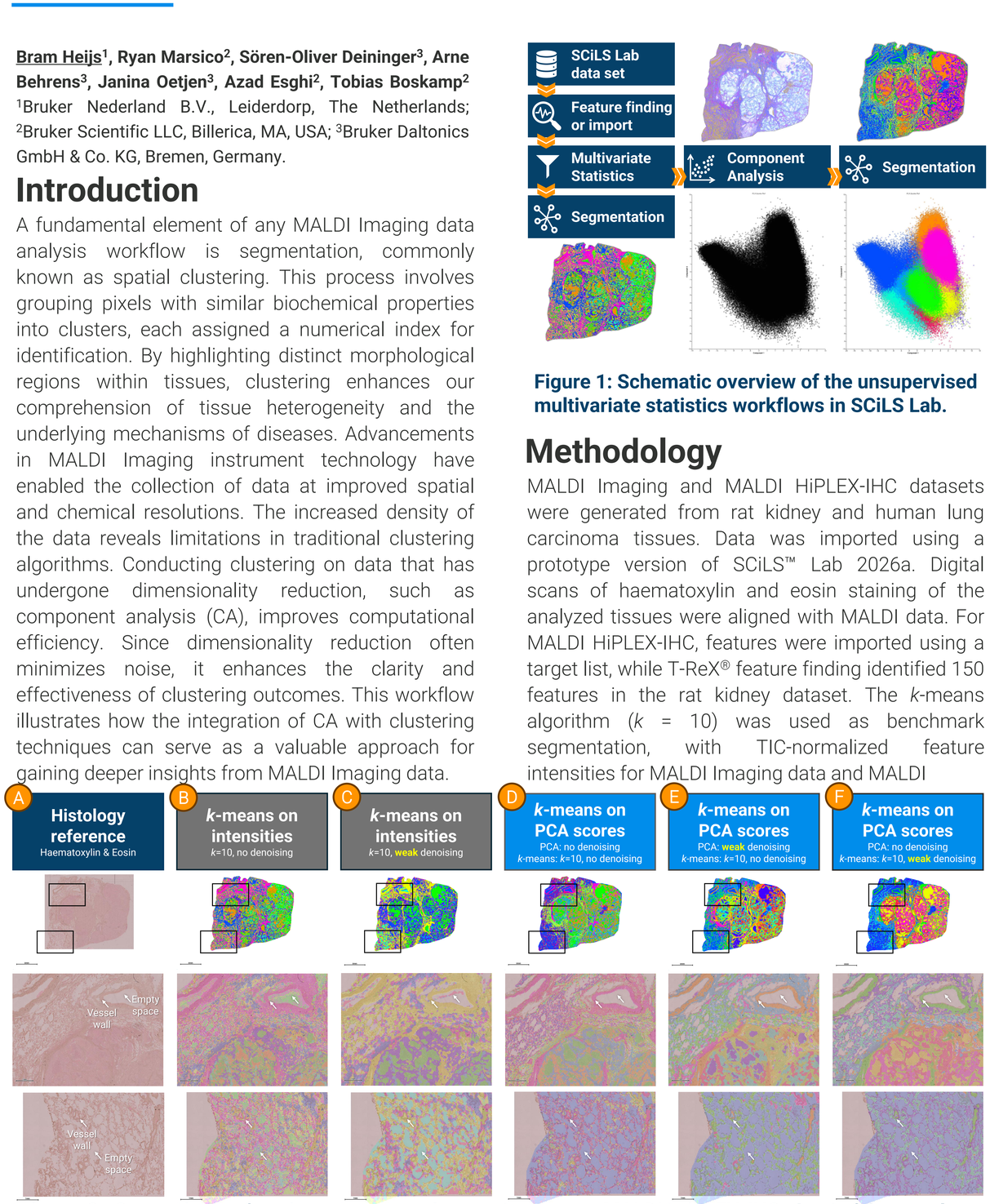

Figure 2: Comparison of various segmentation strategies on MALDI HiPLEX-IHC data of human lung carcinoma. (A) H&E reference image, *k*-means clustering (k=10) on TIC normalized feature intensities **(B)** without spatial denoising and **(C)** with weak spatial denoising. *k*-Means clustering (k=10) on PCA scores (30 components, unit variance scaling) **(D)** without spatial denoising and **(E)** with spatial denoising applied to the PCA and **(F)** segmentations, respectively. The cluster containing predominantly "empty" pixels has been hidden in the middle row. Scale bar overview images 6mm. Scale bars zoomed in images: 1mm. Pixel size 30 × 30µm^2.

Results

When performing segmentation on feature intensities, one of the main problems is caused by noise. Sparse spatial distributions, predominantly caused by low-intensity features, result in an equally sparse clustering (**Figure 2B**). One strategy to avoid this is to perform spatial denoising to the feature ion images prior to clustering (**Figure 2C**),

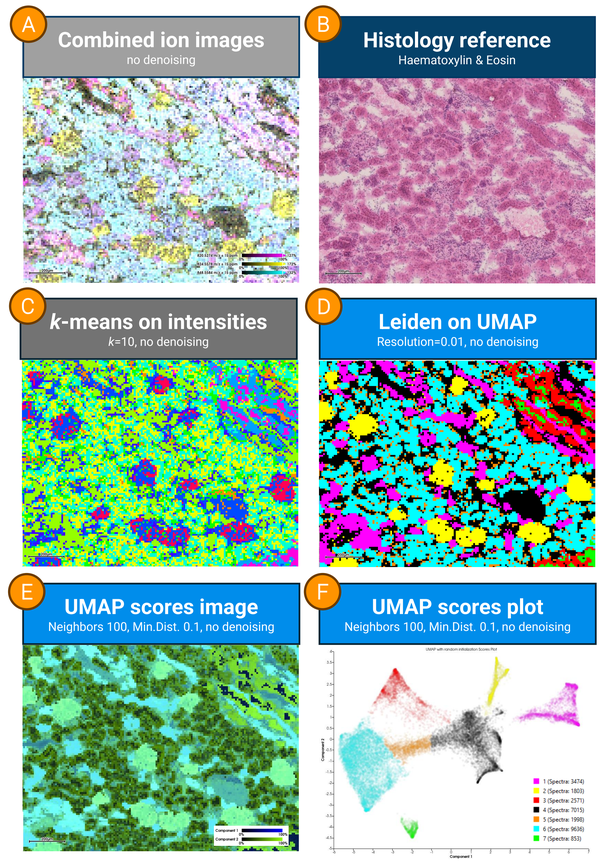

Figure 3: Combined UMAP and Leiden clustering on lipid MALDI Imaging data in rat kidney. (A) Three combined lipid ion images, **(B)** H&E reference image, **(C)** *k*-means clustering (*k*=10) on TIC normalized feature intensities, **(D)** Leiden clustering on UMAP scores, **(E)** UMAP component score images, and **(F)** UMAP scores plot with datapoints colored by the Leiden clustering result. Scale bar = 200 µm. Pixel size 10 × 10µm^2.

although it can come at the cost of losing spatial detail in the cluster image. CAs are designed to identify and extract the most important features in a dataset, and by focussing on these, they efficiently filter noise and irrelevant information from the data.

Figure 2D shows that presenting the same segmentation algorithm with component scores obtained from a CA on the feature intensities results in a superior clustering with better preserved biological structures. Spatial smoothing to either feature intensities prior to CA or on the component scores prior to segmentation can help to generate smoother clustering results (**Figure 2EF**).

While classical algorithms for both CA (e.g. principal component analysis, PCA) and segmentation (e.g. *k*-means) produce relevant and high-quality clustering results (**Figure 2**), modern CA and segmentation algorithms, like uniform manifold approximation and projection (UMAP) and Leiden clustering respectively, have been shown to produce superior clustering results as these tools are more capable of capturing intricate (and non-linear) relationships between data points, commonly observed in biological data (**Figure 3**).

Conclusions

- **Segmentation on component analysis scores provides superior clustering results without losing spatial detail.**
- **Combining modern-day component analysis and segmentation methods better capture intricate relationships observed in MALDI Imaging data of biological systems.**
- **These workflows will be available in the upcoming SCiLS Lab 2026a version.**

Imaging MS: Computational Methods, Software & Analysis

*Miralys™ photocleavable mass tag imaging probes plus their high-plex & multimodal oriented workflows are patented by:

AmberGen

Conflict of interest statement

All coauthors are employees of Bruker, a vendor of analytical equipment, including mass spectrometers, and associated software.

© 2025 Bruker