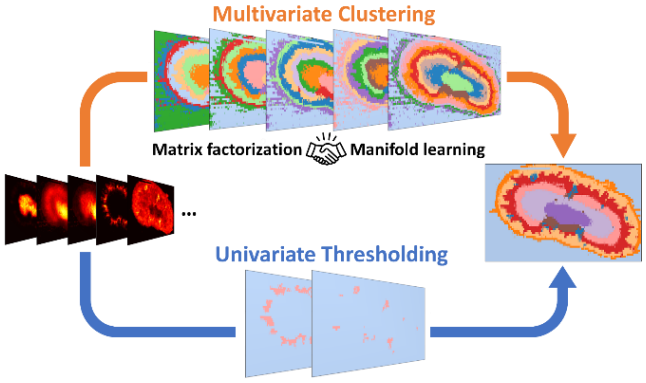
Spatial Segmentation of Mass Spectrometry Imaging Data by Combining Multivariate Clustering and Univariate Thresholding

Hang Hu, Ruichuan Yin, Hilary M. Brown, and Julia Laskin*

Department of Chemistry, Purdue University, West Lafayette, IN 47907, USA

Corresponding author: Julia Laskin, Tel: 765-494-5464, Email: jlaskin@purdue.edu

■ TOC



■ ABSTRACT

Spatial segmentation partitions mass spectrometry imaging (MSI) data into distinct regions providing a concise visualization of the vast amount of data and identifying regions of interest (ROIs) for downstream statistical analysis. Unsupervised approaches are particularly attractive as they may be used to discover the underlying subpopulations present in the high-dimensional MSI data without prior knowledge of the properties of the sample. Herein, we introduce an unsupervised spatial segmentation approach, which combines multivariate clustering and univariate thresholding to generate comprehensive spatial segmentation maps of the MSI data. This approach combines matrix factorization and manifold learning to enable high-quality image segmentation without an extensive hyperparameter search. In parallel, some ion images inadequately represented in the multivariate analysis are treated using univariate thresholding to generate complementary spatial segments. The final spatial segmentation map is assembled from segment candidates generated using both techniques. We demonstrate the performance and robustness of this approach for two MSI data sets of mouse uterine and kidney tissue sections acquired with different spatial resolutions. The resulting segmentation maps are easy to interpret and project onto the known anatomical regions of the tissue.

■ INTRODUCTION

Mass spectrometry imaging (MSI) is a powerful tool in biological and biomedical research, which enables an untargeted characterization of the spatial distribution of hundreds of molecules in tissue samples.[1–4] MSI experiments usually sample a virtual grid of pixels on a sample surface by acquiring a full mass spectrum in each spatial pixel. Typically, MSI generates hundreds of thousands of mass spectra each containing thousands of features, which are subsequently extracted and visualized as 2D ion images. Recent

developments which focused on increasing the throughput[5,6] and improving the spatial resolution[7–9] on MSI techniques have substantially increased the amount of data generated in MSI studies. Several powerful software packages have been developed to provide the MSI community with visualization and analysis tools.[10–13] In parallel, progress has been made in the development of advanced computational analysis methods, which are indispensable for the interpretation and mining of the vast MSI data.[14] In particular, concise representations of MSI data which facilitates data mining and assists human interpretation of high-dimensional data are highly desirable.[15]

Several computational methods including factorization,[16,17] co-localization,[18,19] spatial pattern clustering,[20,21] hyperspectral visualization,[22,23] and spatial segmentation[21,24–27] have been developed for the efficient visualization or fast exploration of complex MSI data. Spatial segmentation is a powerful tool that provides a concise representation of the high-dimensional data and helps identify regions of interest (ROIs) for the downstream analysis. ROIs determined using MSI have been used to understand the landscape of heterogeneous tissue samples and to link molecular signatures to biological conditions through the region-specific statistical analysis.[28–30] For example, a data-driven approach has been used to identify tumor subpopulations that are statistically linked to patient survival in gastric cancer.[31] The ROI-specific analysis has also been reported to reveal temporal lipid profile changes in a rat brain tissue following the traumatic brain injury.[32]

Unsupervised spatial segmentation of MSI data is usually conducted by clustering pixels based on their spectral similarity.[14] For this high-dimensional clustering task, a sequential combination of dimensionality reduction and clustering techniques is commonly used. In addition, a subspace clustering-based method has also been developed for clustering of MSI data.[24] Matrix factorization methods, such as principal component analysis (PCA)[33] and non-negative matrix factorization (NMF)[34] have been used to project MSI data into a lower-dimensional space in a linear manner. In addition, nonlinear manifold learning methods, such as self-organized maps (SOM),[35–37] t-distributed stochastic neighbor embedding (t-SNE),[31,38] and uniform manifold approximation and projection (UMAP)[21] have gained popularity due to their ability to preserve local structures of high-dimensional data in a low map representation.

Both dimensionality reduction techniques have downsides: matrix factorization typically requires more than three components to adequately represent the nonlinear MSI data, which limits the ability to visualize the underlying structure of high-dimensional data.[22] Despite the success of t-SNE and UMAP in the analysis of RNA sequencing[39,40] and mass spectrometry data[21,23], there is a recognition that currently emerging density-preserving versions of these techniques provide more robust visualizations of the original data architecture.[41] Finally, because of the inevitable information loss in dimensionality reduction and the tradeoff between the quality of the representation and "curse of dimensionality",[42] multivariate clustering does not necessarily describe the comprehensive patterns present in complex MSI data.

In this study, we have developed a spatial segmentation approach to address some of these limitations. The approach combines multivariate clustering and univariate thresholding to generate high-quality spatial image segmentation. In particular, the synergy of both matrix factorization (PCA) and manifold learning (UMAP) is utilized in the multivariate analysis, which generates a compressed representation of high-dimensional MSI data for Gaussian mixture model (GMM) clustering. This strategy enables good-quality clustering of pixels without extensive search of hyperparameters for the clustering algorithm. In addition, univariate multi-Otsu thresholding is applied to ion images, which are poorly represented using multivariate techniques, thereby generating complementary spatial segments. Herein, we describe the implementation of the approach using two previously reported MSI data sets of mouse uterine and kidney tissue samples, which were acquired using nanospray desorption electrospray ionization (nano-DESI)[43–45] imaging source

operated at both high (uterine tissue)[46] and moderate (kidney tissue)[47] spatial resolutions. We demonstrate the performance of the approach for different tissue types, spatial resolution, and data complexity.

■ EXPERIMENTAL SECTION

MSI Data. Two MSI data sets used as examples in this study have been previously reported. Briefly, 10 μm thick mouse uterine and kidney tissue sections were analyzed using nano-DESI MSI on a Q-Exactive HF-X Orbitrap mass spectrometer (Thermo Fisher Scientific, Waltham, MA) equipped with a custom-designed nano-DESI source.[9] Mass spectra were acquired in the range of *m/z* 133-2000. For uterine tissue sample, both positive and negative ion mode mass spectra were acquired in the same experiment, while only negative mode mass spectra were collected for kidney tissue. The spatial resolution of nano-DESI experiments was 10 μm and 50 μm for uterine and kidney samples, respectively.[46,47] The dimensions of acquired ion images are provided in the supporting information (Table S1).

Data Preprocessing. Detailed description of the computational approaches is provided in the Supporting Information (SI). We used different data preprocessing approaches for UMAP and PCA (SI 1.1). Specifically, peak detection and *m/z* binning were used for UMAP implementation. Meanwhile, we used a peak picking approach described in our previous study[46] to generate a list of peaks originating from the tissue for PCA analysis. The following custom-designed Python codes were developed for subsequent analysis of the MSI data. Line scan raw files were processed using pyMSfilereader, a Python binding for Thermo MSFileReader dynamic-link library, to construct a 2D MSI data array for each sample with pixel index and *m/z* bins as coordinates. The resulting data format is shown schematically in Figure S1. Signal intensity at each pixel for each *m/z* on the peak list was extracted from the corresponding mass spectrum with a bin width of ±10 ppm. These intensities were normalized to the total ion signal (TIC) of the spectrum. Because automated gain control was turned on during MSI experiments, the number of spectra varied from line to line. A linear interpolation was used to generate data with a fixed number of pixels per line (typically the average number over all the lines). Dimensionalities of the MSI data arrays for UMAP and PCA analyses are shown in Tables S1 and S3, respectively. For PCA analysis, signal intensities in each *m/z* bin were centered and scaled such that they have a mean of 0 and a standard deviation of 1, which eliminates feature's magnitude bias in the unsupervised learning algorithms described below.

Image segmentation by multivariate clustering. Given the high dimensionality of MSI data, dimensionality reduction of *m/z* bins and clustering of pixels were adopted as a general strategy to spatially partition the image. Both UMAP[48] and PCA[49] were used for dimensionality reduction. UMAP was performed to transform MSI data arrays into a low-dimensional space, in which each pixel is represented by a vector of transformed feature values (SI 1.2). Hierarchical density-based spatial clustering for application with noise (HDBSCAN)[50] was used to cluster pixels in the UMAP transformed space (SI 1.3). In parallel, PCA was utilized to reduce the dimensionality of data arrays to 2-50 dimensions. Explained variance ratio for each principal component (PC) was calculated to determine the number of PCs used for subsequent clustering. Gaussian mixture model (GMM)[51] was adopted to identify subpopulations (clusters) in the PCA-projected data; the parameters for each Gaussian mixture component were optimized using the expectation maximization (EM) algorithm. After clustering, the results were visualized by both a scatter plot of color-coded pixels in the 2D feature space and a color-coded spatial segmentation map. For UMAP algorithm, cosine distance was used as suggested in a previous study.[23] The selection of the number of mixture components in GMM was evaluated by the Akaike information criterion (AIC) and Bayesian information criterion (BIC). These two criteria are given by equations 1 and 2:

$$AIC = 2p - 2\log(L) \quad (1)$$
$$BIC = \log(N)p - 2\log(L) \quad (2)$$

where p is the number of free parameters in the model, N is the number of pixels, and L is the likelihood of the model fit. Both criteria take the negative log-likelihood of the model and a penalty term of model complexity to avoid over-fitting. In model selection, the lower value of AIC or BIC corresponds to a more favored model.[52] However, the penalty of model complexity in BIC is harsher, as it multiplies the logarithm of the number of pixels (N) by the number of free parameters (p). The implication of this difference between AIC and BIC will be discussed later in the text. In this study, UMAP was implemented using a umap-lean Python package, and other methods were implemented using Scikit-Learn, an open-source machine learning library for Python.[53]

Image segmentation by univariate thresholding. After the PCA+GMM multivariate clustering, ion distribution patterns were sorted by their PCA loadings, which helped identify outlier ion images for the univariate thresholding (SI 1.4). Multi-Otsu thresholding algorithm[54] was used to partition images into 5 classes. A subsequent despiking process was performed on the resulting segments to remove a mild noise present in the data set. In this process, a 5 x 5 pixel moving window was set to scan the full image. At each scanning step, the intensity value of the center pixel was compared with the median intensity value of pixels in the window. If the value was 3 times larger than the standard deviation, the intensity of the center pixel was replaced by the median intensity value of the surrounding 8 pixels. Multi-Otsu thresholding was implemented in Scikit-image Python package.[55]

Ensemble generation and segment assembly. Ensemble generation, which involves processing the data many times while making some perturbation to the data or hyperparameters of algorithms,[42] was applied in the image segmentation. In particular, GMM clustering was repeated with different numbers of mixture components; multi-Otsu thresholding was conducted on several ion images which provide complementary ion distributions. As a result, multiple spatial segment candidates were generated in both image segmentation modules. A majority vote method[42] was adopted to automatically assemble a segmentation map based on the co-occurrence membership agreement across the ensemble clustering results (details are provided in SI 1.5). Finally, segments generated using univariate thresholding were added to the segmentation map. A manual selection and assembly method was used to validate the final map. The source code is available at https://github.com/hanghu1024/MSI-segmentation.

■ RESULTS AND DISCUSSION

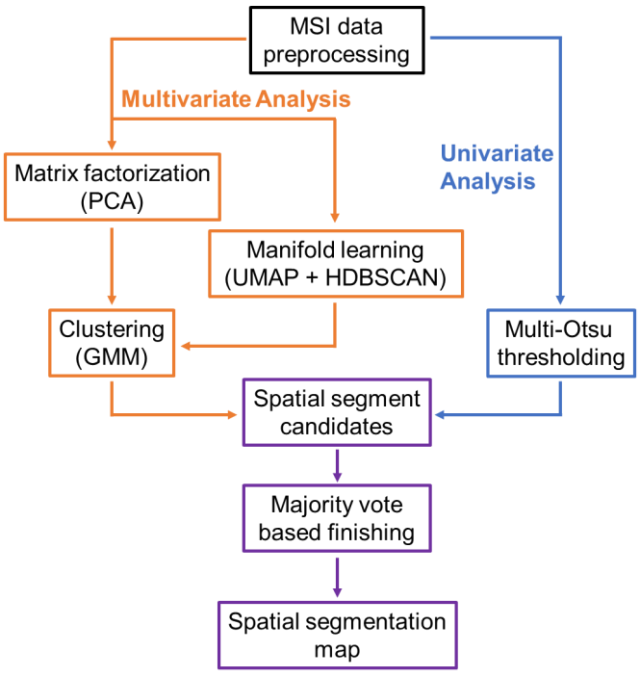


Figure 1. Overview of the image segmentation workflow. Color codes indicate independent data processing modules and arrows indicate data flow.

Spatial segmentation workflow overview. The approach developed in this study comprises four independent data processing modules, as illustrated in Figure 1. First, we perform data preprocessing to organize the MSI data for downstream data mining. Next, pixels are clustered based on their spectral similarity in multivariate analysis. Particularly, both PCA and UMAP are applied: the former generates compressed features for GMM clustering, while the latter helps estimate the number of clusters. Given these inputs, GMM is repeatedly fitted, assigned with a range of mixture component numbers around the estimated cluster number. In this study, a range of 5 is set by default. In parallel, ion images that are poorly represented in multivariate analysis are independently partitioned using multi-Otsu thresholding. As a result, ensemble generation of both multivariate and univariate analyses approximates a pool of spatial segment candidates. We adopted a co-occurrence majority vote to identify the most robust multivariate clustering result and incorporated segments generated using univariate thresholding into the final segmentation map. This ensemble generation and finishing strategy also addresses the stability issues commonly observed in multivariate clustering.[42]

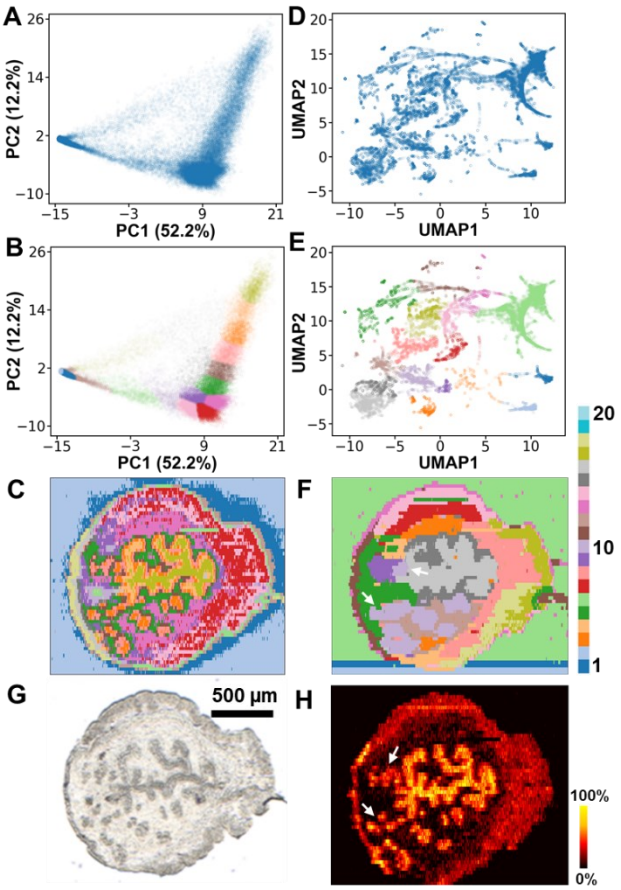


Figure 2. Multivariate clustering of MSI data of a mouse uterine tissue section. (A) 2D representation of the PCA analysis (PC1 vs. PC2). The PCA+GMM pixel clustering results color-coded by cluster assignments are visualized in a 2D feature space (B) and spatial domain (C). (D) The 2D UMAP embeddings. The UMAP+HDBSCAN pixel clustering results color-coded by cluster assignments are visualized in a 2D feature space (E) and spatial domain (F). (G) Optical image of the mouse uterine tissue section, scale bar = 500 µm. (H) A representative ion image (*m/z* 746.5106), with intensity scale changes from black (low) to yellow (high). The color bar on the right illustrates integer clustering labels (1-20) for color codes. Integer labels were randomly assigned to clusters, and clustering labels in (B, C) and (E, F) are different. Dots in the scatter plots are set to be 90% transparent. White arrows in panel H indicate glandular epithelium region.

Dimensionality reduction for multivariate clustering. Extensive search of the parameter space of both UMAP and PCA implementation indicated that best UMAP result is obtained when all the peaks including tissue-related and solvent signals are included. We found that peak detection and *m/z* binning provides good-quality UMAP representations with a reasonable runtime on a desktop computer (Figure S4). Meanwhile, best PCA representation was obtained using a filtered list of predominately tissue-related peaks (Figure S5).

Our approach, which synergistically combines UMAP and PCA methods, is illustrated in Figure 2 for mouse uterine MSI data. Figures 2A and 2D depict the projection of all 52954 pixels in the mouse uterine tissue data onto a 2D feature space generated using PCA and UMAP, respectively. PCA preserves the relative distances between objects by linear projection. Therefore both the separation and compactness of

the projected patterns describe their association in the original space.[56] However, a substantial information loss occurs when only 2 PCs are employed. As shown in Figure 2A, there is insufficient separation of the clusters in the PCA projection. In contrast, UMAP creates a nearest-neighbors graph in the original space and arranges a low-dimensional embedding according to the distances between the neighboring points in the graph, which thereby preserves the local structure of the high-dimensional data.[40,48] Therefore, UMAP efficiently separates MSI data into a number of distinct clusters (Figure 2D), which have better compactness and separation in comparison with PCA.

In the next step, we use HDBSCAN and GMM to cluster the UMAP and PCA 2D projections, respectively. The HDBSCAN views clusters as areas of high density separated by areas of low density[50] and is therefore well-suited to cluster the UMAP embedding. In contrast, the GMM models PCA-projected data using a number of Gaussian mixture components with variable parameters. In this HDBSCAN implementation, we used a soft clustering mode (min_cluster_size = 300, min_samples = 30), which classified pixels into 18 clusters (details are provided in SI 1.3). For direct comparison, we use the same number of mixture components (18 components) in GMM.

The clustering results are visualized in both 2D feature space (Figures 2B, E) and spatial domain (Figures 2C, F). Individual pixels are color-coded based on their cluster assignments indicated by the color bar on the right of Figure 2. We note that independent coloring is used to visualize the PCA and UMAP results. The clustering analysis reveals the spectral similarity of pixels in subregions of the uterine tissue section. As seen in the PCA+GMM plots (Figure 2B, C), pixels on the glass slide (light blue: label 2) generate a more compact cluster than pixels in tissue subregions. There is also a large distance between these two kinds of pixels in Figure 2B. Meanwhile, the clusters observed on the tissue edge (light brown: label 12, light green: label 6, light yellow: label 18) are lined up in the middle of the PCA plane. Clustering analysis in the PCA feature space indicates that spectra from adjacent subregions are more similar than those from distant subregions, which is consistent with the expected chemical gradients presenting in biological tissue samples. The combination of PCA and GMM provides a reasonable image segmentation, in which segments reproduce biologically interesting patterns observed in the optical and ion images (Figure 2G, H and Figure S5). Similar cluster architectures are also observed in UMAP+HDBSCAN plots (Figure 2E and 2F). The UMAP spatial segmentation map captures major features but does not capture some fine patterns observed in the ion images. For example, segment 10 highlighted with arrows in Figure 2F does not include some of the glandular epithelium pixels but includes some stroma pixels, which affects quantification results discussed later.

We used the ground truth color-coding approach reported in the literature[41] to examine the quality of multivariate representations (Figure S9). Specifically, we color-coded the final segmentation labels, obtained at the end of the workflow, to PCA and UMAP 2D representations. Pixels from the same tissue region uniformly agglomerate together in PCA. Meanwhile, some of the groups are split and outlier pixels are observed outside the groups generated by UMAP. Additional segmentation results obtained using UMAP embeddings with more dimensions are shown in Figure S10. We found that the higher-dimensional analysis does not improve the segmentation results. Based on the MSI data obtained from the highly heterogeneous mouse uterine tissue, the comparison of the performance of UMAP and PCA reveals that PCA is well-suited for image segmentation. Meanwhile, UMAP is a powerful technique for the visualization of the embedded local clusters that present in the high-dimensional MSI data.

Hyperparameter search for PCA+GMM clustering. Next, we examine the effect of hyperparameters on image segmentation using PCA+GMM. In particular, we use both mouse uterine and kidney tissue MSI data sets to examine the effect of the: (1) number of PCs in PCA, and (2) number of mixture components in GMM on the performance of multivariate clustering.

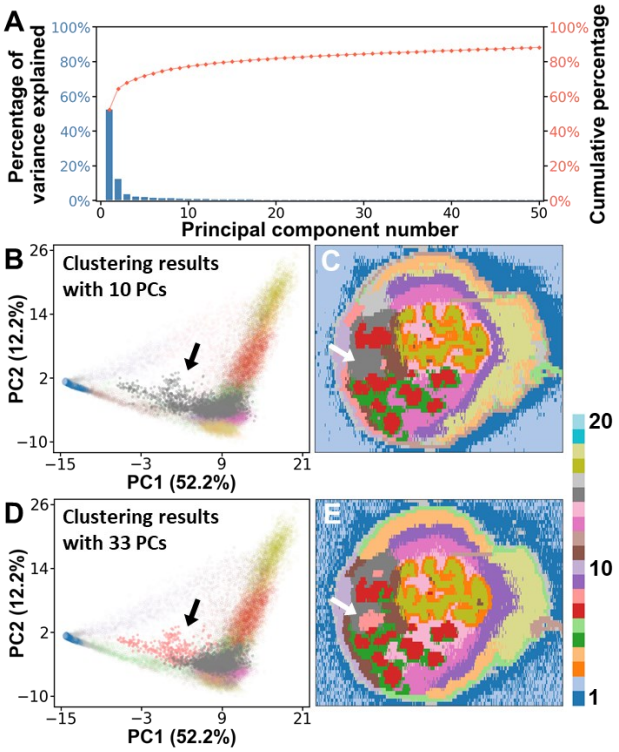


Figure 3. Comparison of the GMM pixel clustering results for different number of PCs. (A) Scree plot on the percentage of variance explained by PCs. GMM with 10 PCs clustering result visualized in the 2D feature space (B) and spatial domain (C). GMM with 33 PCs clustering result visualized in the 2D feature space (D) and spatial domain (E). Most dots in scatter plots are set to be 90% transparent, while gray (label 15) coded pixels in (B), gray and salmon (label 8) coded pixels in (D) are set to be opaque for comparison. White arrows in panels C and E indicate pixels of a fine pattern in spatial domain, while black arrows in panels B and D indicate them in the 2D feature space.

In general, the smallest number of PCs resulting in sufficient explained variance should be used to reduce data dimensionality with a minimal loss of information. Empirically, we set an 85% threshold for cumulative percentage of variance explained (CVE) and a maximum PC number of 50. For the mouse uterine tissue data, a scree plot is shown in Figure 3A. According to our criterion, 33 PCs describing 84.9% CVE are selected and modeled by GMM with 18 mixture components. The results are compared with the GMM clustering performed using 10 PCs describing 75% CVE and the same number of GMM mixture components. Clustering results are visualized in both the 2D feature space and spatial domain with color-coded pixels in Figures 3B-E. In comparison to clustering results obtained using two PCs shown in Figure 2B, C, cleaner spatial segments are obtained using a larger number of PCs. This is attributed to a better separation of pixels residing at the classification boundaries with increase in the number of PCs. We obtain similar results using 10 and 33 PCs. However, some of the finer patterns discovered using 33 PCs are not observed in the segmentation obtained using 10 PCs. For example, a cluster shown in salmon (label 8) and highlighted by the black and white arrows in Figures 3D and 3E, respectively, is not captured using 10 PCs. This subregion highlighted in Figure S23 has a lower level of alkali ion concentrations compared to the surrounding stroma cells and is only observed in ion images normalized to TIC.[46]

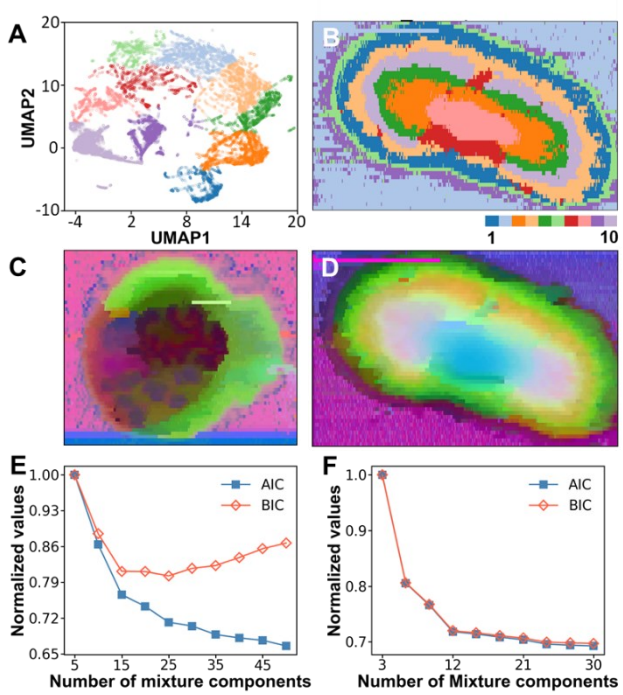


Figure 4. Estimation of the optimal number of mixture components in GMM. (A) A 2D UMAP embedding of the mouse kidney MSI data. (B) A PCA+GMM pixel clustering result visualized in the spatial domain using 5 PCs and 10 mixture components. RGB hyperspectral visualizations of (C) mouse uterine and (D) kidney tissue datasets. Normalized AIC and BIC values as a function of the number of mixture components for (E) mouse uterine and (F) kidney tissue MSI data.

One of the major challenges in unsupervised clustering is to determine the number of clusters.[57] In MSI image segmentation, an extensive search is usually adopted to estimate the optimal number of clusters.[20,24,33,58] This approach increases the complexity and time of the analysis. In this study, we use UMAP analysis to guide the selection of the number of mixture components for GMM clustering. Since UMAP effectively separates complex MSI data in 2D/3D feature space, the optimal number of clusters may be estimated using UMAP visualizations. Furthermore, we use the information criteria described in the experimental section to validate this parameter search strategy. Since GMM is a probabilistically grounded method, the probability of the clustering assignment for each pixel is traceable, which enables the calculation of the likelihood and AIC/BIC values in the fitted model.

We demonstrate the performance of this approach for both mouse uterine and kidney MSI data, in which 33 and 5 PCs (85% CVE from Figure S24) are used in the GMM clustering, respectively. First, an intuitive estimation of the number of clusters could be made based on the 2D UMAP plots (Figure S11). A HDBSCAN clustering analysis with similar cluster identification mechanism can be used to assist this decision making as shown in Figures 2E, 4A and S13. Furthermore, UMAP-based RGB hyperspectral visualizations[23] shown in Figures 4C and 4D help visualize major features of the MSI data. Based on the 2D UMAP plot, we estimate the presence of 18 and 10 clusters in the uterine and kidney tissue data, respectively. The corresponding GMM clustering results are shown in Figure 3E and Figure 4B, which reasonably represent ion distribution patterns in Figure S5. To statistically evaluate it, normalized AIC/BIC values, calculated from independently fitted GMMs, are plotted against the number of mixture component for uterine (Figure 4E) and kidney (Figure 4F) MSI data. For mouse uterine tissue, the AIC value reaches an inflection point at 15 mixture components. In contrast, a shallow minimum is observed in the BIC plot

spanning over 15-25 mixture components. This distinct difference in the AIC and BIC trends may be attributed to the harsher penalty over model's complexity in BIC calculation. Interestingly, both BIC and AIC show a similar gradual decrease with the number of mixture components in the analysis of the mouse kidney tissue with an inflection point at 12 mixture components shown in Figure 4D. The steady downward trend of both AIC and BIC may be attributed to the presence of non-Gaussian components in the data, which cannot be described using a single Gaussian distribution.[59] We propose that the inflection point is a reasonable choice for selecting the optimal number of mixture components in GMM. For both mouse uterine and kidney data, the number of mixture component determined by analyzing AIC/BIC values is consistent with the number of clusters estimated by UMAP visualizations. Our results indicate that regardless of the complexity and spatial resolution of the MSI data, UMAP can be used to determine the number of clusters for PCA+GMM clustering.

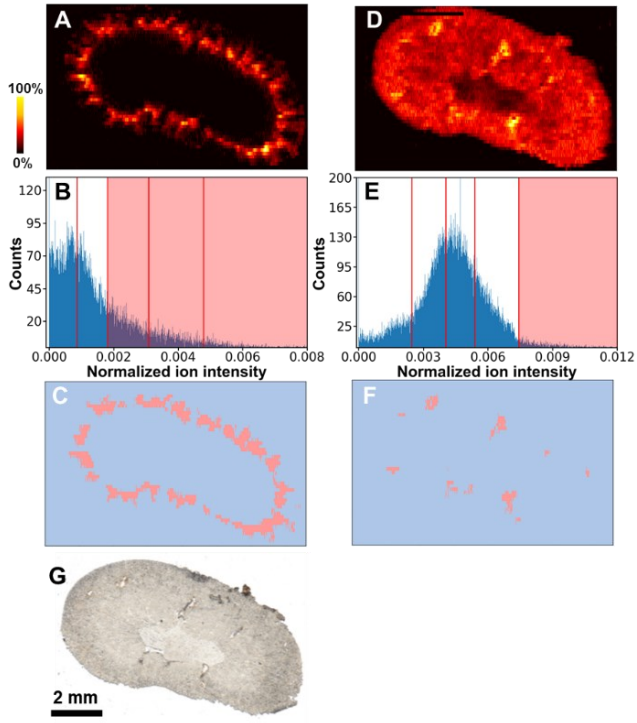


Figure 5. Single ion image thresholding generates complementary spatial segments for mouse kidney tissue MSI data. (A) Ion image of *m/z* 439.2441. (B) Histogram of (A) with thresholds illustrated by red vertical lines. (C) Spatial segment (salmon) selected from 3 groups of pixels, as illustrated by salmon coded blocks in (B). (D) Ion image of *m/z* 236.0411. (E) Histogram of (D) with thresholds. (F) Spatial segment (salmon) selected from the last group of pixels, as illustrated by salmon coded block in (E). (G) Optical image of the kidney tissue section.

Univariate thresholding of outlier ion images for complementary image segmentation. Although a good-quality spatial segmentation may be achieved by using multivariate clustering, some ion distribution patterns may not be captured by the clustering models. In mouse kidney tissue MSI data, two ion images of *m/z* 439.2441 and 236.0411, shown in Figures 5, represent two outliers not identified by the clustering analysis. They were detected and selected through a Ward's hierarchical clustering-based outlier detection process described in SI 1.4. The ion at *m/z* 439.2441 (Figure 5A) is observed in the tubules of the inner cortex of the kidney. This region is located in close proximity to the gold segment (label 4) in Figure 4B, which was generated by PCA+GMM, but its outline poorly matches the ion distribution of *m/z* 439.2441.

Another unusual pattern is observed for *m/z* 236.0411 (Figure 5D), which is substantially enhanced in lymphatic vessels clearly seen in the optical image (Figure 5G). The presence of such outliers could be attributed to the insufficient representation of minor spatial patterns among selected PCs in multivariate analysis module. In PCA analysis, the loading is the correlation between the variable and component, which estimates the information they share.[49] Accordingly, the sum of squares of loadings (SSL) represents the proportion of the variance of the original ion distribution explained by the selected PCs. For *m/z* 439.2441 and 236.0411 in mouse kidney data the values of SSL are 4.7% and 1.7%, respectively, with the total SSL of the corresponding classes of spatial patterns of 10.1% and 17.3% (Figure S16). These two classes are ranked as 21th and 14th among all ion distribution patterns (23 classes in total). Due to the low SSL values, these spatial patterns are not incorporated into the spatial segments generated by multivariate clustering, which are dominated by higher-ranked features.

In this study, we used the multi-Otsu thresholding algorithm to obtain spatial segments based on single ion images shown in Figures 5A and 5D. The multi-Otsu thresholding method partitions images by separating pixels into several classes while maximizing the between-class variance of pixel intensities.[54] As a preliminary step to thresholding analysis, the histograms (Figure 5B, E) of two ion images are examined. The histogram provides the statistics of pixels in an image, which helps to discover ion distribution properties. For example, there are three noticeable pixel distributions in Figure 5E: (1) pixels observed below the intensity of 0.0025 correspond to pixels on the glass slide with close to zero intensity and a small number of low intensity pixels in the kidney medulla region; (2) a Gaussian-like distribution observed between 0.0025 and 0.0075 corresponds to pixels localized on the tissue; (3) pixels observed above the intensity of 0.0075 are localized at lymphatic vessels. To implement multi-Otsu thresholding for ion images, we predefine the number of pixel classes as 5, in order to classify all possible features. As illustrated by red vertical lines in Figure 5E, thresholds generated by this method successfully differentiate components in the ion image of *m/z* 236.0411. By selecting and merging the classified pixels, we obtained two segments shown in Figure 5C and 5F, which are complementary to the multivariate clustering results. Other state-of-the-art univariate segmentation methods such as spatially aware Dirichlet Gaussian mixture models may be utilized in this step.[27]

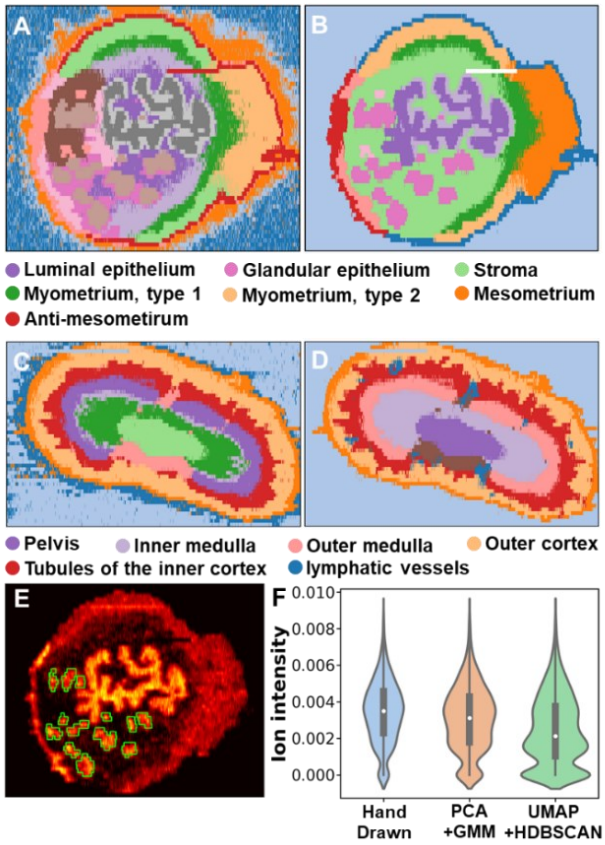


Figure 6. Assembly of the final segmentation map and subsequent region-specific quantitative analysis. Automatically assembled segmentation maps of mouse uterine (A) and kidney (C) tissues. Segments are identified using the majority vote method from ensemble clustering results. Simplified maps showing key anatomical features in mouse uterine (B) and kidney (D) tissues. Anatomical annotations are only shown for simplified segmentation maps. (E) Hand-drawn ROI for the glandular epithelium using an ion image of *m/z* 746.5106. (F) Violin plots of ion intensities in ROIs generated using three different methods.

Ensemble generation and assembly of spatial segmentation map. After the validation of approaches adopted in multivariate and univariate analyses, the last step is to assemble a spatial segmentation map from segment candidates. To construct the candidate pool, ensemble generation is applied for both image segmentation modules. It is especially useful in multivariate clustering, since neither UMAP visualizations nor the analysis of AIC/BIC values provides a unique value for the number of mixture components. In order to evaluate different clustering scenarios, we independently performed GMM clustering over a range of numbers of mixture components. Using the procedure described earlier, GMM was repeatedly fitted with 16-20 and 8-12 mixture components for uterine and kidney data, respectively. Multi-Otsu thresholding was applied to two ion images for kidney data but not for uterine data, which are adequately described using GMM ensemble clustering (Figure S5 and S18). Ensemble generation results for both data sets are shown in Figures S26 and S27. All clustering generated segmentation maps reveal reasonable spatial segments on tissues, and thresholding generated segments are unique and complementary. With the increase of the number of mixture component in GMM, finer patterns emerge on segmentation maps, while some patterns remain invariant throughout all models, which indicate robust clusters. These individual segments were identified using a co-occurrence majority vote-based approach (Figure S20 - S22). The resulting segmentation maps shown in Figures 6A and 6C closely resemble spatial ion distributions in the original

data (Figure S5). Furthermore, the identified segments are closely related to key anatomical features of mouse uterine and kidney sections highlighted in the simplified maps in Figures 6B and 6D, respectively. Automatically generated results are in good agreement with a manual segment assembly shown in Figures S26 and S27. This process also addresses the clustering stability issue. With random initialization, the EM algorithm may converge to a local rather than global minimum.[52] The optimization process is also affected by the number of mixture components assigned. Ensemble generation and finishing help identify robust clustering results and avoid errors originating from algorithm perturbations.[42] Collectively, this approach provides comprehensive high-quality ROIs without extensive search of hyperparameters, which benefits the downstream quantitative analysis. One example for ion at *m/z* 746.5106 is shown is Figure 6F. The statistics of TIC-normalized ion signals in the hand-drawn ROI of the glandular epithelium along with ROIs obtained using PCA+GMM (Figure 6B) and UMAP+HBDSCAN (Figure 2F) are visualized using violin plots. Both the average and the distribution of ion signals obtained using PCA+GMM ROI closely resemble the results obtained using a hand-drawn ROI. Meanwhile UMAP+HDBSCAN provides a lower average value and a substantially distorted distribution of ion signals due to the large number of the stroma signals included in this ROI.

■ CONCLUSION

We have developed and validated a robust approach for image segmentation of high-dimensional MSI data by combining multivariate clustering and univariate thresholding. We discuss the trade-off between matrix factorization (PCA) and manifold learning methods (UMAP) for the high-dimensional clustering of MSI data. Specifically, PCA reduces the dimensionality and preserves the relative distances of the high-dimensional data. As a result, a PCA plot provides a readily interpretable map for spectral similarity of pixels in the MSI data. In contrast, UMAP preserves the local structure of the data and provides a better separation of the groups of pixels making it possible to estimate the number of segments in the complex data. PCA analysis also generates a compressed representation of the high-dimensional data for GMM clustering. It provides an optimized probability distribution for each segment, which is further validated using the AIC and BIC analysis. A combination of these methods enables high-quality image segmentation without extensive hyperparameter search, which captures a majority of spatial segments in MSI data.

Univariate thresholding is adopted to partition outlier ion distribution patterns, which are missed by the multivariate clustering due to the relatively low explained variance in the transformed representations. The identification of such patterns may be important for understanding biological processes based on MSI data. Furthermore, quantitative analysis of MSI data benefits from an accurate representation of both major and minor spatial patterns. The integrated strategy developed in this study, assisted by ensemble generation and finishing, has been used to construct comprehensive spatial segmentation maps of distinct sets of MSI data of varying complexity acquired with different spatial resolutions for different tissue types.

Our approach can be readily expanded to incorporate other image processing techniques. For example, denoising of ion images may be used to improve the performance of the multivariate/univariate analysis. Furthermore, other imaging modalities such as optical, fluorescence, or confocal Raman imaging, could be incorporated into the workflow with appropriate spatial registration to improve the information content of different imaging techniques.

■ ACKNOWLEDGMENTS

The authors acknowledge support from the National Science Foundation (NSF-1808136), the National Institutes of Health (NIH) Common Fund, through the Office of Strategic Coordination/Office of the NIH

Director under award UG3HL145593 and UH3CA255132 (HuBMAP Program), and Merck & Co. (Grant 40002399).

■ ASSOCIATED CONTENT

Supporting Information

The Supporting Information is available free of charge on the ACS Publications website.

Contents in the Supporting Information:

Supporting data analysis, results and discussions.

Supporting figures and references.

■ AUTHOR INFORMATION

Corresponding Author

Address: 560 Oval Drive, West Lafayette, IN 47907-2084. E-mail: jlaskin@purdue.edu

Notes

The authors declare no competing financial interest.

■ REFERENCES

(1) Norris, J. L.; Caprioli, R. M. Analysis of Tissue Specimens by Matrix-Assisted Laser Desorption/Ionization Imaging Mass Spectrometry in Biological and Clinical Research. *Chem. Rev.* **2013**, *113* (4), 2309–2342.

(2) Vaysse, P. M.; Heeren, R. M. A.; Porta, T.; Balluff, B. Mass Spectrometry Imaging for Clinical Research-Latest Developments, Applications, and Current Limitations. *Analyst* **2017**, *142* (15), 2690–2712.

(3) Wu, C.; Dill, A.; Eberlin, L.; Cooks, G.; Ifa, D. Mass Spectrometry Imaging Under Ambient Conditions. *Mass Spectrom. Rev.* **2013**, *32*, 218–243.

(4) Laskin, J.; Lanekoff, I. Ambient Mass Spectrometry Imaging Using Direct Liquid Extraction Techniques. *Anal. Chem.* **2016**, *88* (1), 52–73.

(5) Prentice, B. M.; Chumbley, C. W.; Caprioli, R. M. High-Speed MALDI MS/MS Imaging Mass Spectrometry Using Continuous Raster Sampling. *J. Mass Spectrom.* **2015**, *50* (4), 703–710.

(6) Basu, S. S.; Regan, M. S.; Randall, E. C.; Abdelmoula, W. M.; Clark, A. R.; Gimenez-Cassina Lopez, B.; Cornett, D. S.; Haase, A.; Santagata, S.; Agar, N. Y. R. Rapid MALDI Mass Spectrometry Imaging for Surgical Pathology. *npj Precis. Oncol.* **2019**, *3* (1), 1-5.

(7) Zavalin, A.; Yang, J.; Hayden, K.; Vestal, M.; Caprioli, R. M. Tissue Protein Imaging at 1 Mm Laser Spot Diameter for High Spatial Resolution and High Imaging Speed Using Transmission Geometry MALDI TOF MS. *Anal. Bioanal. Chem.* **2017**, *407* (8), 2337–2342.

(8) Kompauer, M.; Heiles, S.; Spengler, B. Atmospheric Pressure MALDI Mass Spectrometry Imaging of Tissues and Cells at 1.4-Mm Lateral Resolution. *Nat. Methods* **2016**, *14* (1), 90–96.

(9) Yin, R.; Kyle, J.; Burnum-Johnson, K.; Bloodsworth, K. J.; Sussel, L.; Ansong, C.; Laskin, J. High

Spatial Resolution Imaging of Mouse Pancreatic Islets Using Nanospray Desorption Electrospray Ionization Mass Spectrometry. *Anal. Chem.* **2018**, *90* (11), 6548–6555.

(10) Bemis, K. D.; Harry, A.; Eberlin, L. S.; Ferreira, C.; Van De Ven, S. M.; Mallick, P.; Stolowitz, M.; Vitek, O. Cardinal: An R Package for Statistical Analysis of Mass Spectrometry-Based Imaging Experiments. *Bioinformatics* **2015**, *31* (14), 2418–2420.

(11) Källback, P.; Nilsson, A.; Shariatgorji, M.; Andrén, P. E. MsIQuant - Quantitation Software for Mass Spectrometry Imaging Enabling Fast Access, Visualization, and Analysis of Large Data Sets. *Anal. Chem.* **2016**, *88* (8), 4346–4353.

(12) Bokhart, M. T.; Nazari, M.; Garrard, K. P.; Muddiman, D. C. MSiReader v1.0: Evolving Open-Source Mass Spectrometry Imaging Software for Targeted and Untargeted Analyses. *J. Am. Soc. Mass Spectrom.* **2018**, *29* (1), 8–16.

(13) Tortorella, S.; Tiberi, P.; Bowman, A. P.; Claes, B. S. R.; Ščupáková, K.; Heeren, R. M. A.; Ellis, S. R.; Cruciani, G. LipostarMSI: Comprehensive, Vendor-Neutral Software for Visualization, Data Analysis, and Automated Molecular Identification in Mass Spectrometry Imaging. *J. Am. Soc. Mass Spectrom.* **2020**, *31* (1), 155–163.

(14) Verbeeck, N.; Caprioli, R. M.; Van de Plas, R. Unsupervised Machine Learning for Exploratory Data Analysis in Imaging Mass Spectrometry. *Mass Spectrom. Rev.* **2019**, 1–47.

(15) Alexandrov, T. Spatial Metabolomics and Imaging Mass Spectrometry in the Age of Artificial Intelligence. *Annu. Rev. Biomed. Data Sci.* **2020**, *3* (1), 61–87.

(16) Verbeeck, N.; Spraggins, J. M.; Murphy, M. J. M.; Wang, H. dong; Deutch, A. Y.; Caprioli, R. M.; Van de Plas, R. Connecting Imaging Mass Spectrometry and Magnetic Resonance Imaging-Based Anatomical Atlases for Automated Anatomical Interpretation and Differential Analysis. *Biochim. Biophys. Acta - Proteins Proteomics* **2017**, *1865* (7), 967–977.

(17) Van De Plas, R.; Ojeda, F.; Dewil, M.; Van Den Bosch, L.; De Moor, B.; Waelkens, E. Prospective Exploration of Biochemical Tissue Composition via Imaging Mass Spectrometry Guided by Principal Component Analysis. *Pacific Symp. Biocomput. 2007, PSB 2007* **2007**, *469*, 458–469.

(18) McDonnell, L. A.; Van Remoortere, A.; Van Zeijl, R. J. M.; Deelder, A. M. Mass Spectrometry Image Correlation: Quantifying Colocalization. *J. Proteome Res.* **2008**, *7* (8), 3619–3627.

(19) Ovchinnikova, K.; Stuart, L.; Rakhlin, A.; Nikolenko, S.; Alexandrov, T. ColocML: Machine Learning Quantifies Co-Localization between Mass Spectrometry Images. *Bioinformatics* **2020**, *36* (10), 3215–3224.

(20) Alexandrov, T.; Chernyavsky, I.; Becker, M.; Von Eggeling, F.; Nikolenko, S. Analysis and Interpretation of Imaging Mass Spectrometry Data by Clustering Mass-to-Charge Images According to Their Spatial Similarity. *Anal. Chem.* **2013**, *85* (23), 11189–11195.

(21) Smets, T.; Waelkens, E.; De Moor, B. Prioritization of m / z -Values in Mass Spectrometry Imaging Profiles Obtained Using Uniform Manifold Approximation and Projection for Dimensionality Reduction . *Anal. Chem.* **2020**, *92* (7), 5240–5248.

(22) Fonville, J. M.; Carter, C. L.; Pizarro, L.; Steven, R. T.; Palmer, A. D.; Griffiths, R. L.; Lalor, P. F.; Lindon, J. C.; Nicholson, J. K.; Holmes, E.; Bunch, J. Hyperspectral Visualization of Mass Spectrometry Imaging Data. *Anal. Chem.* **2013**, *85* (3), 1415–1423.

(23) Smets, T.; Verbeeck, N.; Claesen, M.; Asperger, A.; Griffioen, G.; Tousseyn, T.; Waelput, W.;

Waelkens, E.; De Moor, B. Evaluation of Distance Metrics and Spatial Autocorrelation in Uniform Manifold Approximation and Projection Applied to Mass Spectrometry Imaging Data. *Anal. Chem.* **2019**, *91* (9), 5706–5714.

(24) Alexandrov, T.; Becker, M.; Deininger, S. O.; Ernst, G.; Wehder, L.; Grasmair, M.; Von Eggeling, F.; Thiele, H.; Maass, P. Spatial Segmentation of Imaging Mass Spectrometry Data with Edge-Preserving Image Denoising and Clustering. *J. Proteome Res.* **2010**, *9* (12), 6535–6546.

(25) Alexandrov, T.; Kobarg, J. H. Efficient Spatial Segmentation of Large Imaging Mass Spectrometry Datasets with Spatially Aware Clustering. *Bioinformatics* **2011**, *27* (13), 230–238.

(26) Bemis, K. D.; Harry, A.; Eberlin, L. S.; Ferreira, C. R.; Van de Ven, S. M.; Mallick, P.; Stolowitz, M.; Vitek, O. Probabilistic Segmentation of Mass Spectrometry (MS) Images Helps Select Important Ions and Characterize Confidence in the Resulting Segments. *Mol. Cell. Proteomics* **2016**, *15* (5), 1761–1772.

(27) Guo, D.; Bemis, K.; Rawlins, C.; Agar, J.; Vitek, O. Unsupervised Segmentation of Mass Spectrometric Ion Images Characterizes Morphology of Tissues. *Bioinformatics* **2019**, *35* (14), i208–i217.

(28) Bergman, H. M.; Lundin, E.; Andersson, M.; Lanekoff, I. Quantitative Mass Spectrometry Imaging of Small-Molecule Neurotransmitters in Rat Brain Tissue Sections Using Nanospray Desorption Electrospray Ionization. *Analyst* **2016**, *141* (12), 3686–3695.

(29) Yajima, Y.; Hiratsuka, T.; Kakimoto, Y.; Ogawa, S.; Shima, K.; Yamazaki, Y.; Yoshikawa, K.; Tamaki, K.; Tsuruyama, T. Region of Interest Analysis Using Mass Spectrometry Imaging of Mitochondrial and Sarcomeric Proteins in Acute Cardiac Infarction Tissue. *Sci. Rep.* **2018**, *8* (1), 1–10.

(30) Andersen, M. K.; Krossa, S.; Høiem, T. S.; Buchholz, R.; Claes, B. S. R.; Balluff, B.; Ellis, S. R.; Richardsen, E.; Bertilsson, H.; Heeren, R. M. A.; Bathen, T. F.; Karst, U.; Giskeødegård, G. F.; Tessem, M. B. Simultaneous Detection of Zinc and Its Pathway Metabolites Using MALDI MS Imaging of Prostate Tissue. *Anal. Chem.* **2020**, *92* (4), 3171–3179.

(31) Abdelmoula, W. M.; Balluff, B.; Englert, S.; Dijkstra, J.; Reinders, M. J. T.; Walch, A.; McDonnell, L. A.; Lelieveldt, B. P. F. Data-Driven Identification of Prognostic Tumor Subpopulations Using Spatially Mapped t-SNE of Mass Spectrometry Imaging Data. *Proc. Natl. Acad. Sci. U. S. A.* **2016**, *113* (43), 12244–12249.

(32) Roux, A.; Muller, L.; Jackson, S. N.; Post, J.; Baldwin, K.; Hoffer, B.; Balaban, C. D.; Barbacci, D.; Schultz, J. A.; Gouty, S.; Cox, B. M.; Woods, A. S. Mass Spectrometry Imaging of Rat Brain Lipid Profile Changes over Time Following Traumatic Brain Injury. *J. Neurosci. Methods* **2016**, *272*, 19–32.

(33) Sarkari, S.; Kaddi, C. D.; Bennett, R. V.; Fernandez, F. M.; Wang, M. D. Comparison of Clustering Pipelines for the Analysis of Mass Spectrometry Imaging Data. *2014 36th Annu. Int. Conf. IEEE Eng. Med. Biol. Soc. EMBC 2014* **2014**, 4771–4774.

(34) Reindl, W.; Bowen, B. P.; Balamotis, M. A.; Green, J. E.; Northen, T. R. Multivariate Analysis of a 3D Mass Spectral Image for Examining Tissue Heterogeneity. *Integr. Biol.* **2011**, *3* (4), 460–467.

(35) Franceschi, P.; Wehrens, R. Self-Organizing Maps: A Versatile Tool for the Automatic Analysis of Untargeted Imaging Datasets. *Proteomics* **2014**, *14* (7–8), 853–861.

(36) Gardner, W.; Cutts, S. M.; Muir, B. W.; Jones, R. T.; Pigram, P. J. Visualizing ToF-SIMS

Hyperspectral Imaging Data Using Color-Tagged Toroidal Self-Organizing Maps. *Anal. Chem.* **2019**, *91* (21), 13855–13865.

(37) Gardner, W.; Maliki, R.; Cutts, S. M.; Muir, B. W.; Ballabio, D.; Winkler, D. A.; Pigram, P. J. Self-Organizing Map and Relational Perspective Mapping for the Accurate Visualization of High-Dimensional Hyperspectral Data. *Anal. Chem.* **2020**, *92* (15), 10450–10459.

(38) Abdelmoula, W. M.; Pezzotti, N.; Hölt, T.; Dijkstra, J.; Vilanova, A.; McDonnell, L. A.; Lelieveldt, B. P. F. Interactive Visual Exploration of 3D Mass Spectrometry Imaging Data Using Hierarchical Stochastic Neighbor Embedding Reveals Spatiomolecular Structures at Full Data Resolution. *J. Proteome Res.* **2018**, *17* (3), 1054–1064.

(39) Becht, E.; McInnes, L.; Healy, J.; Dutertre, C. A.; Kwok, I. W. H.; Ng, L. G.; Ginhoux, F.; Newell, E. W. Dimensionality Reduction for Visualizing Single-Cell Data Using UMAP. *Nat. Biotechnol.* **2019**, *37* (1), 38–47.

(40) Dorrity, M. W.; Saunders, L. M.; Queitsch, C.; Fields, S.; Trapnell, C. Dimensionality Reduction by UMAP to Visualize Physical and Genetic Interactions. *Nat. Commun.* **2020**, *11* (1), 1–6.

(41) Narayan, A.; Berger, B.; Cho, H. Assessing Single-Cell Transcriptomic Variability through Density-Preserving Data Visualization. *Nat. Biotechnol.* **2021**, 1-10.

(42) Ronan, T.; Qi, Z.; Naegle, K. M. Avoiding Common Pitfalls When Clustering Biological Data. *Sci. Signal.* **2016**, *9* (432), 1–13.

(43) Roach, P. J.; Laskin, J.; Laskin, A. Nanospray Desorption Electrospray Ionization: An Ambient Method for Liquid-Extraction Surface Sampling in Mass Spectrometry. *Analyst* **2010**, *135* (9), 2233–2236.

(44) Laskin, J.; Heath, B. S.; Roach, P. J.; Cazares, L.; Semmes, O. J. Tissue Imaging Using Nanospray Desorption Electrospray Ionization Mass Spectrometry. *Anal. Chem.* **2012**, *84* (1), 141–148.

(45) Li, X.; Yin, R.; Hu, H.; Li, Y.; Sun, X.; Dey, S. K.; Laskin, J. An Integrated Microfluidic Probe for Mass Spectrometry Imaging of Biological Samples. *Angew. Chem. Int. Ed.* **2020**, 59, 22388–22391.

(46) Yin, R.; Burnum-johnson, K. E.; Laskin, J. High Spatial Resolution Imaging of Biological Tissues Using Nanospray Desorption Electrospray Ionization Mass Spectrometry. *Nat. Protoc.* **2019**, *14*, 3445–3470.

(47) Brown, H. M.; Sanchez, D. M.; Yin, R.; Chen, B.; Vavrek, M.; Cancilla, M. T.; Zhong, W.; Shyong, B.; Zhang, R.; Li, F. Mass Spectrometry Imaging of Diclofenac and Its Metabolites in Tissues Using Nanospray Desorption Electrospray Ionization . *ChemRxiv* **2020**, 13194422.v1.

(48) McInnes, L.; Healy, J.; Melville, J. UMAP: Uniform Manifold Approximation and Projection for Dimension Reduction. *arXiv* **2018,** 1802.03426.

(49) Abdi, H.; Williams, L. J. Principal Component Analysis. *Wiley Interdiscip. Rev. Comput. Stat.* **2010**, *2* (4), 433–459.

(50) McInnes, L.; Healy, J. Accelerated Hierarchical Density Based Clustering. *IEEE Int. Conf. Data Min. Work. ICDMW* **2017**, *2017-Novem*, 33–42.

(51) Chellappa, R.; Veeraraghavan, A.; Ramanathan, N.; Yam, C.-Y.; Nixon, M. S.; Elgammal, A.; Boyd, J. E.; Little, J. J.; Lynnerup, N.; Larsen, P. K.; Reynolds, D. Gaussian Mixture Models. *Encycl. Biometrics* **2009**, No. 2, 659–663.

(52) Li, R.; Perneczky, R.; Yakushev, I.; Förster, S.; Kurz, A.; Drzezga, A.; Kramer, S. Gaussian Mixture Models and Model Selection for [18F] Fluorodeoxyglucose Positron Emission Tomography Classification in Alzheimer's Disease. *PLoS One* **2015**, *10* (4), 1–22.

(53) Pedregosa, F.; Varoquaux, G.; Gramfort, A.; Michel, V.; Thirion, B.; Grisel, O.; Blondel, M.; Prettenhofer, P.; Weiss, R.; Dubourg, V.; Vanderplas, J.; Passos, A.; Cournapeau, D.; Brucher, M.; Perrot, M.; Duchesnay, E. Scikit-Learn: Machine Learning in Python. *J. Mach. Learn. Res.* **2011**, *12*, 2825–2830.

(54) Liao, P. S.; Chen, T. S.; Chung, P. C. A Fast Algorithm for Multilevel Thresholding. *J. Inf. Sci. Eng.* **2001**, *17* (5), 713–727.

(55) Van Der Walt, S.; Schönberger, J. L.; Nunez-Iglesias, J.; Boulogne, F.; Warner, J. D.; Yager, N.; Gouillart, E.; Yu, T. Scikit-Image: Image Processing in Python. *PeerJ* **2014**, *2014* (1), 1–18.

(56) Domeniconi, C.; Papadopoulos, D.; Gunopulos, D.; Ma, S. Subspace Clustering of High Dimensional Data. *SIAM Proc. Ser.* **2004**, *6* (1), 517–521.

(57) Tibshirani, R.; Walther, G.; Hastie, T. Estimating the Number of Data Clusters via the Gap Statistic. *Journal of the Royal Statistical Society: Series B*. 2001, pp 411–423.

(58) Widlak, P.; Mrukwa, G.; Kalinowska, M.; Pietrowska, M.; Chekan, M.; Wierzgon, J.; Gawin, M.; Drazek, G.; Polanska, J. Detection of Molecular Signatures of Oral Squamous Cell Carcinoma and Normal Epithelium – Application of a Novel Methodology for Unsupervised Segmentation of Imaging Mass Spectrometry Data. *Proteomics* **2016**, *16* (11–12), 1613–1621.

(59) Steele, R. J.; Raftery, A. E. Performance of Bayesian Model Selection Criteria for Gaussian Mixture Models. *Front. Stat. Decis. Mak. bayesian Anal.* **2010**, 113–130.