Electronic Supplementary Information

Self-supervised Clustering of Mass Spectrometry Images

Using Contrastive learning

Hang Hu¹, Jyothsna Padmakumar Bindu² and Julia Laskin^{1*}

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

2. School of Engineering Technology, Purdue University, West Lafayette, IN 47907, USA

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

Table of content

1. Methods and data analysis	
2. Additional tables	
Table S1	S6
Table S2	S6
Table S3	S6
Table S4	S6
Table S5	S7
3. Additional figures	
Figure S1	S8
Figure S2	S8
Figure S3	S9
Figure S4	S9
Figure S5	S10
Figure S6	S11 – S12
Figure S7	S13 – S14
Figure S8	S15
Figure S9	S15
Figure S10	S16 – S20
Figure S11	S21 – S26
Figure S12	S27
4. References	S27

Methods and data analysis

Mass spectrometry imaging data

Mouse uterine and brain tissue MSI datasets used as examples in this study have been previously reported.^{1,2} Uterine tissues from 2-month-old female mice were snap-frozen, and frozen sections were made with a cryostat. The mice were maintained on a C57BL6 mixed background and housed in the vivarium at the Cincinnati Children's Hospital Medical Center according to NIH and institutional guidelines for laboratory animals. This protocol was approved by the Cincinnati Children's Hospital Research Foundation Institutional Animal Care and Use Committee. Mouse uterine tissue was 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.³ Mass spectra were acquired in the *m/z* range of 133-2000 in both positive and negative ion modes with a spatial resolution of 10 μ m. Positive mode MSI data for mouse brain tissue was obtained from METASPACE,⁴ a community resource that provides open access to MSI data. The specific data were acquired using MALDI MSI in the *m/z* range of 600-1000 with a spatial resolution of 20 μ m. The dimensions of the two MSI datasets are listed in Table S1.

Data pre-processing

To generate ion images from MSI data, we used peak detection and m/z binning as described in our previous study.⁵ The signal intensity for each m/z in each pixel was extracted from the corresponding mass spectrum with a bin width of \pm 10 ppm and normalized to the total ion current. To remove visual spikes, pixels with intensities > 0.999 quantile were reassigned with the 0.999 intensity quantile value. Ion images of mouse uterine tissue were resized to 96 x 96 pixels. Meanwhile, larger-size ion images of mouse brain tissue were resized to 224 x 224 pixels. Most pre-trained CNN models and Pytorch transform functions accept RGB images. Thus, raw pixel intensities of ion images were normalized between 0 and 255 and copied to 3 channels. More specifically, we converted ion images into the PIL format before the data augmentation step. To benchmark our approach, we manually selected 367 mouse uterine ion images with distinct ion distributions and clustered them into 13 groups according to molecular colocalizations as shown in Fig. S6. For the unannotated mouse brain dataset, we detected 1101 peaks from the average spectrum and generated the corresponding ion images as shown in Fig. S8.

Architecture of the self-supervised clustering

We approached the challenge of molecular localization clustering as an image classification task. We aimed to re-train a CNN model for an individual MSI dataset to classify ion images based on the high-level spatial features without manual annotations. The model architecture is shown in Fig. 1. The pre-trained CNN (EfficientNet-B0) is re-trained by contrastive learning and self-labeling sequentially in a self-supervised manner. We use EfficientNet-B0, which has been trained on the ImageNet database. We used the EfficientNet-B0 model before the classification layer as an encoder. In our architecture, we firstly learned ion image representations through the contrastive learning. More specifically, SimCLR⁶ approach is adopted in this study. After this first phase of training, we fed ion images through the re-trained encoder to produce a set of feature vectors, which were then passed to a spectral clustering (SC) classifier to generate the initial labels for the classification task. To complete a learnable classification CNN, a linear classifier (a linear layer followed by a softmax function) was attached to the encoder and trained with the original ion images and initial labels as inputs. Finally, we utilized a self-labeling⁷ approach to fine-tune both the encoder and classifier, which allows the network to correct itself.

Contrastive learning

Details of SimCLR implementation are shown in Fig. S1. The representation of an ion image is the output before the classification layer of EfficientNet-B0. A small multilayer perceptron with one hidden layer maps the representations to the projection space (Z) where the contrastive loss is applied. In the training step each ion image is used to generate a pair of augmentations. We treat two augmentations of one ion image as a positive pair (i, j). Then the loss function is defined as

$$\ell_{i,j} = -\log \frac{\exp\left(\sin\left(\mathbf{z}_{i}, \mathbf{z}_{j}\right) / \tau\right)}{\sum_{k=1}^{2N} \mathbf{1}_{[k \neq i]} \exp\left(\sin\left(\mathbf{z}_{i}, \mathbf{z}_{k}\right) / \tau\right)}$$

where 2N is the number of augmented images, $\mathbf{z}_i, \mathbf{z}_j \in \mathbb{Z}$, sim denotes the cosine similarity, $\mathbf{1}_{[k\neq i]}$ denotes an indicator function evaluating to 1 iff $k \neq i$ and τ is a temperature parameter with a default value of 0.5. The final loss is computed across all positive pairs in the minibatch. To evaluate the quality of learned representations for mouse uterine benchmark, we used a linear evaluation and the nearest neighbor mining. In the linear evaluation, one average ion image was generated from each manually classified group of images as the centroid of the cluster. A linear classifier was subsequently trained on top of the frozen retrained encoder, with 13 average ion images and their corresponding annotated labels. Next, all original ion images were classified by the encoder and the updated linear classifier. The resulting classification accuracy is used as a proxy for the quality of image representation.⁶ In the nearest neighbor mining protocol, for each ion image, we searched its K nearest neighbors (K $\in [1, 30]$) based on cosine distance in the representation space. We quantified the purity of the neighborhood by counting the annotation-matching pairs for each image and their nearest neighbors. We also visualized the learned representations using t-SNE with scikitlearn default settings.

Image clustering

It has been demonstrated in a previous report⁷ and this study (Fig. 2e) that images with similar high-level spatial features are mapped together in the representation space using contrastive learning. To leverage the meaningful local neighborhoods, SC was adopted to cluster ion images as the classification pretext task. Based on the cosine distance, 10 nearest neighbors of each ion image were identified to construct a graph. Next, we used the discretization approach to cluster nodes in the graph after the Laplacian embeddings.⁸ After the SC, we used the resulting classification labels to initialize a learnable linear classifier on top of the CNN encoder, which enables the following self-labeling process to fine-tune the model (step 2 in Fig. 1). More specifically, the linear classifier is composed of a linear layer and a softmax function. We set the encoder in frozen mode and trained the classifier with original ion images and initial labels obtained from SC.

Self-labeling

As reported in a previous study, self-labeling improves the CNN model using a plain criterion: two independently augmented images from one ion should be classified into the same cluster. With this principle, the self-supervised training is able to enhance the generalization power of the model. In addition, only the confidently-classified ion images are included in this training.⁷ As shown in Fig. 2f and Fig. S5, we observed that the classification accuracy obtained for selected mouse uterine ion images increases with an increase in the softmax probability threshold. This indicates that the softmax probability threshold may be used to exclude falsely classified ion images from the training, which enhances the accuracy of the CNN

model in the self-labeling step. In the implementation, we empirically selected the probability threshold with respect to the sample population. More specifically, after the initialization of the linear classifier, we selected the 40% quantile of softmax probabilities as the threshold. The selected training data containing 60% of the original ion images has a larger fraction of correctly classified images than the original data. Training samples were updated with the same probability threshold at every epoch, thus we gradually included more samples into the training (Fig. 2g). For each selected ion image, we applied a weak and a strong augmentations, respectively. Two pseudo labels were obtained after the encoder and classifier. A weighted cross-entropy loss was then applied to the minibatch of weakly augmented ion images to update the parameters of the CNN model.

Clustering accuracy evaluation

In this study, we used over-clustering to effectively capture the intra-class variance. The clustering accuracy was calculated using the following equation:

$$Accuracy(\mathbb{C},\mathbb{T}) = \frac{1}{N} \sum_{i} \max_{j} |c_{i} \cap t_{j}|$$

Where $\mathbb{C} = c_1, c_2, ..., c_I$ is the set of predicted clusters, $\mathbb{T} = t_1, t_2, ..., t_J$ is the set of ground truth classes, N is the total number of ion images. In each cluster, the most frequent ground truth class was identified and assigned as a predicted class for the whole cluster. The accuracy was calculated by counting the correctly predicted ion images and dividing by N.

Comparison of the self-supervised clustering with vector-based methods

First, we evaluated the pairwise similarity measurements with different input data. In the current method, we used CNN feature vectors for the downstream classification task. For a comparative study, image vectors were generated by flattening the original images after 0.999 quantile hot spot removal as a pre-processing step. Three metrics, including Euclidean distance, cosine similarity, and Pearson correlation were calculated for comparison.

Next, we evaluated the clustering accuracy of two established machine learning methods that rely on image vectors: a density based clustering analysis with nonlinear dimensionality reduction⁹ and Ward hierarchical clustering⁵. The Uniform Manifold Approximation and Projection (UMAP) was performed using the umaplearn package with the following parameters: n_components = 3, n_neighbors=5, min_dist=0.5, metric = 'cosine'. We note that these parameters were optimal for visualizing and grouping of image vectors, but were not best suited for the CNN feature vectors (Fig. 2d as reference). Nevertheless, we used the same UMAP parameters in Fig. 4c and 4d for the visualization of both results for an objective comparison between them. Hierarchical Density-based Spatial Clustering for Applications with Noise (HDBSCAN) was performed to cluster UMAP representations using default parameters except for "min_cluster_size". We tuned "min_cluster_size" in order to obtain either 13 or 20 clusters. In all the HDBSCAN clustering results, the number of noise data points was less than ten. In the Ward hierarchical clustering, each vector was normalized to the range of [0, 1]. The method was implemented using scipy package (scipy.cluster.hierarchy.linkage) with the following parameters: method='ward', metric='euclidean'."

Isotopic recall evaluation

Isotopic recall is a metric reported in a previous study¹⁰ to evaluate the performance of ion image clustering. Isotopic ion images were identified based on the m/z shift and Pearson correlation of the ion image with that of the candidate monoisotopic peak. Specifically, mass spectral features separated from each other by

m/z 1.003 with a tolerance of m/z 0.01 were assigned as candidate isotopic peaks. Second, ion images of isotopic peaks should have a Pearson correlation greater than 0.5. For the self-supervised clustering results, we counted the number of isotopic images that were correctly clustered together and divided it by the total number of identified isotopes. This fraction gives the isotopic recall, which is in the range of 0 to 1.

Data augmentation

Data augmentation is crucial for contrastive learning and self-labeling. We systematically studied the impact of data augmentation operators as listed in Table S3. For each single augmentation operation, we randomly sampled a parameter interval for an image transformation function. Moreover, we classified the compositions of data augmentation operators into three classes. Medium augmentation was used in SimCLR, while weak and strong augmentations were used in self-labeling.

Model training protocal

For SimCLR training step, we used Adam optimizer with the initial learning rate of 0.001. A cosine annealing with a period of training epochs was used to decay the learning rate. According to the previous report, SimCLR benefits from larger batch sizes.⁶ In our implementation, we used the largest batch size allowed by the GPU memory and trained for 100 iterations. For the self-labeling step, an Adam optimizer with the initial learning rate of 0.0001 and the same cosine annealing schedular were used. We trained the CNN for 300 iterations. We implemented the model training on Google Colaboratory, a cloud computing platform. Using the NVIDIA Tesla P100-16GB GPU, the total training time for mouse uterine MSI benchmark and mouse brain MSI dataset was about 15 and 45 minutes, respectively.

Parameters	Benchmark dataset	Unannotated dataset
Tissue type	Mouse uterine	Mouse brain
Ionization methods	nano-DESI	MALDI
Spatial resolution (µm)	10	20
Raw pixel number	58×920	240×220
Resized pixel number	96×96	224×224
Polarity mode	Positive and negative	Positive
Mass range (m/z)	135-1000	600-1000
Image number	367	1101

Table S1. Information of benchmark mouse uterine dataset and unannotated mouse brain dataset.

Table S2. Accuracy of spectral clustering results with 20 clusters at conditions of varying neighborhood sizes. Ion image representations were generated by encoder shown in Fig. 2d.

Number of neighbors	3	5	8	10	15	20	30
Accuracy (%)	67.8 ± 2.4	92.3 ± 1.0	92.8 ± 0.1	92.2 ± 0.3	90.8 ± 0.7	88.8 ± 2.1	85.8 ± 0.8

Table S3. A summary of image augmentation operator parameters. The kernel size in Gaussian blur is set to be 10% of the image height/width. Off-tissue mask was not applied to unannotated dataset.

Augmentation class	Transformation	Parameter	Interval	
Weak	Gaussian blur	sigma	[0.001, 0.4]	
	Gaussian noise	sigma	[0.001, 0.1]	
	Color Jitter	brightness, contrast, saturation, hue	0.25, 0.25, 0.25, 0.2	
	Off-tissue mask	N/A	N/A	
Medium	Gaussian blur	sigma	[0.01, 0.75]	
	Gaussian noise	sigma	[0.001, 0.2]	
	Color Jitter	brightness, contrast, saturation, hue	0.5, 0.5, 0.5, 0.2	
Strong	Gaussian blur	sigma	[0.1, 2]	
	Gaussian noise	sigma	[0.001, 0.4]	
	Color Jitter	brightness, contrast, saturation, hue	0.5, 0.5, 0.5, 0.2	
	Off-tissue mask	N/A	N/A	

Table S4. Comparison of the classification accuracy of different image vector-based clustering methods on benchmark data.

Clustering method	Number of clusters	Accuracy (%)
Ward Hierarchical Clustering	13	70.2
UMAP + HDBSCAN	13	58.3 ± 2.1
Self-Supervised Clustering	13	84.0 ± 3.1
Ward Hierarchical Clustering	20	80.9
UMAP + HDBSCAN	20	68.4 ± 2.0
Self-Supervised Clustering	20	92.7 ± 2.1

Group	Reference ion	Other ion	Image vector			CNN feature vector		
			Euclidean distance	Cosine similarity	Pearson similarity	Euclidean distance	Cosine similarity	Pearson similarity
1 pos, 87	nos 975 5700	pos, 739.4681	35.5	0.720	0.609	23.2	0.420	0.406
	pos, 875.5700	pos, 868.5243	38.7	0.685	0.564	18.6	0.695	0.687
2 neg, 471.261	nog 171 2616	neg, 684.6076	23.5	0.745	0.724	25.6	0.626	0.560
	neg, 471.2010	pos, 793.5930	35.7	0.595	0.553	18.4	0.785	0.759
3 pos, 820.61	nos 920 6191	pos, 594.3172	32.8	0.596	0.552	22.0	0.530	0.496
	pos, 820.0181	pos, 707.5007	33.7	0.563	0.521	10.3	0.873	0.865
4	neg, 293.5339	neg, 309.5065	44.3	0.573	0.466	32.8	0.232	0.186
		pos, 838.5600	47.6	0.513	0.398	27.5	0.420	0.399

Table S5. A variety of pairwise similarity measurements for image vector and CNN feature vectors.



Figure S1. The framework of SimCLR.



Figure S2. Contrastive learning with strong geometry-changing augmentations. (a) Illustration of strong geometry-changing augmentations used in current study. (b) Linear evaluation of re-trained CNN encoder with individual augmentation shown in panel a. Results of transfer learning and SimCLR without data augmentation are also listed for comparison.



Figure S3. Generalization power provided by composite image augmentation strategy.



Figure S4. The impact of training time on learned image representations during contrastive learning. (a) t-SNE visualizations of learned representations with different training epochs. (b) Contrastive losses during trainings with different epoch settings. (c) Linear evaluation of re-trained CNN encoder with different epoch settings.



Figure S5. Relationships between classification accuracy and confidently classified image fraction obtained from (a) CNN model which was trained by initial labels and classified ion images into 20 groups. (b) CNN model which was trained by self-labeling and classified ion images into 13 groups and (c) CNN model which was trained by self-labeling and classified ion images into 20 groups. Each data point was calculated from an ion image subset which was selected by a softmax probability threshold.



43¥)



Figure S6. Manual classification of 367 ion images in the mouse uterine benchmark.





Figure S7. Self-supervised clustering results of 367 ion images in the mouse uterine benchmark.



Figure S8. Four groups of ion images for pairwise similarity measurement discussion.



Figure S9. Dendrogram and sorted image vectors in Ward hierarchical clustering results. A zoom-in image shows that three ion images (shown in Fig. 4a) are clustered together. Although Ward hierarchical clustering is able to capture major patterns in input data, it cannot differentiate between the ion image of m/z 739.4681 and two other ion images used in this example.











Figure S10. 1101 ion images in the MALDI mouse brain dataset.













Figure S11. Self-supervised clustering results of 1101 ion images in the MALDI mouse brain dataset.



Figure S12. t-SNE visualizations of ion image representations obtained from (a) EfficientNet-B0 and (b) re-trained CNN encoder.

References

- Yin, R., Burnum-johnson, K. E. & Laskin, J. High spatial resolution imaging of biological tissues using nanospray desorption electrospray ionization mass spectrometry. *Nat. Protoc.* 14, 3445–3470. (2019).
- 2. Kompauer, M., Heiles, S. & Spengler, B. Atmospheric pressure MALDI mass spectrometry imaging of tissues and cells at 1.4-µm lateral resolution. *Nat. Methods* 14, 90–96 (2016).
- 3. Yin, R. *et al.* High Spatial Resolution Imaging of Mouse Pancreatic Islets Using Nanospray Desorption Electrospray Ionization Mass Spectrometry. *Anal. Chem.* 90, 6548–6555 (2018).
- 4. Palmer, A. *et al.* FDR-controlled metabolite annotation for high-resolution imaging mass spectrometry. *Nat. Methods* 14, 57–60 (2016).
- Hu, H., Yin, R., Brown, H. M. & Laskin, J. Spatial Segmentation of Mass Spectrometry Imaging Data by Combining Multivariate Clustering and Univariate Thresholding. *Anal. Chem.* 93, 3477– 3485 (2021).
- 6. Chen, T., Kornblith, S., Norouzi, M. & Hinton, G. A simple framework for contrastive learning of visual representations. In *Int. Con. on Mach. Learn.* 1575–1585 (ICML 2020).
- 7. Van Gansbeke, W., Vandenhende, S., Georgoulis, S., Proesmans, M. & Van Gool, L. SCAN: Learning to Classify Images Without Labels. In *Eur. Conf. on Comput. Vis.* 268–285 (ECCV 2020).
- 8. Von Luxburg, U. A tutorial on spectral clustering. *Stat. Comput.* 17, 395–416 (2007).
- 9. 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.
- 10. Zhang, W. *et al.* Spatially aware clustering of ion images in mass spectrometry imaging data using deep learning. *Anal. Bioanal. Chem.* 413, 2803–2819 (2021).