A Co-registration Pipeline for Multimodal MALDI and Confocal Imaging Analysis of Stem Cell Colonies

Arina Nikitina
School of Biological Sciences, Georgia Institute of Technology, Atlanta, Georgia 30332, United States

Danning Huang, **Li Li**
School of Chemistry and Biochemistry, Georgia Institute of Technology, Atlanta, Georgia 30332, United States

Nicholas Peterman, **Sarah E. Cleavenger**
Wallace H. Coulter Department of Biomedical Engineering, Georgia Institute of Technology & Emory University, Atlanta, Georgia 30332, United States

Facundo M. Fernández,
School of Chemistry and Biochemistry and Petit Institute of Bioengineering and Biosciences, Georgia Institute of Technology, Atlanta, Georgia 30332, United States;

Melissa L. Kemp
Wallace H. Coulter Department of Biomedical Engineering, Georgia Institute of Technology & Emory University, Atlanta, Georgia 30332, United States; Petit Institute of Bioengineering and Biosciences, Georgia Institute of Technology, Atlanta, Georgia 30332, United States;

Abstract

Multimodal mass spectrometry imaging (MSI) data presents unique big data challenges in handling and analysis. Here, we present a pipeline for co-registering matrix-assisted laser desorption/ionization MSI and confocal immunofluorescence imaging data for extracting single-cell metabolite signatures. We further describe methods and introduce software for the simultaneous analysis of these concatenated data sets, which are designed to establish a connection between cell traits of interest (shape metrics, position within sample) and the cells' own metabolic signatures.

Graphical Abstract

Corresponding Authors: Facundo M. Fernańdez - Phone: 404 385 4432; facundo. Fernández@chemistry.gatech.edu; Fax: 404 385 6447, **Melissa L. Kemp** - Phone: 404 385 6341; melissa.kemp@bme.gatech.edu; Fax: 404 385 6447.
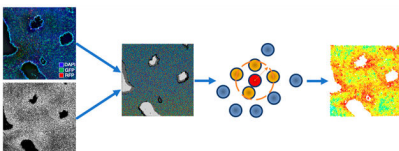
Keywords

human induced pluripotent stem cells; multimodal image analysis software; single-cell; metabolomics

INTRODUCTION

Matrix-assisted laser desorption/ionization (MALDI) mass spectrometry imaging (MSI) is a powerful tool for spatially resolved mass spectrometric investigation of metabolites and proteins on a sample surface.[1] Recent biological applications of MALDI MSI range from identification of microorganisms in a sample[2] to cancer diagnostics.[3,4] The spatial resolution afforded by routine MALDI MSI experiments, typically in the range of 10–200 μm,[5] reveals information about cell colonies rather than single cells, with the exception being higher resolution MSI approaches such as MALDI-2.[6,7] When applied to cell colonies or intact tissue samples, MALDI MSI produces no clear cell to cell separation due to spatial resolution limitations,[8] making it challenging to extract cell-specific information and to correlate such information to cell properties observed through other imaging approaches. Therefore, there is a need for analysis methods that facilitate the interpretation of MALDI when combined with other types of imaging data.

Here, we introduce software for multimodal imaging and co-registration of MALDI and confocal fluorescence data that allows linking positional basis of cellular phenotype with metabolites produced in stem cell colonies (github.com/kemplab/co-registration.git). This analytical pipeline enables insights beyond routine methods of extracting covarying relationships in mass-to-charge ratios by multivariate techniques such as principal component analysis (PCA). Prior to MALDI imaging, our experimental approach involves performing confocal live cell imaging of the sample by immuno-fluorescent pluripotency antibody labeling and Hoechst nuclei dye for consequent MALDI and confocal image alignment and segmentation. This approach allows extraction of average MALDI spectral abundances on an approximate cell-by-cell basis for each m/z value, which—in combination with cell pluripotency status and cell shape metrics information—yields a complete data set suited for multiple multivariate techniques. We introduce multivariate modeling methods that yield more informative variable dependencies and extract metabolic spatial relationships hidden in the data set.

EXPERIMENTAL SECTION

All used materials and cell culture protocols are described in the Supplemental Sections S1 and S2. Human-induced pluripotent stem cell (hiPSC) colonies seeded on indiumtin oxide coated slides were subject to a spontaneous differentiation protocol (Supplemental Section

S3). Measurements were taken on days 4, 5, and 6 of the protocol. For each day analyzed, live hiPSCs were stained with Hoechst (Figure 1a), NL493-conjugated Mouse Anti-Human SSEA4, and NL557-conjugated Mouse Anti-Human TRA-1–60(R), and confocal images were acquired. 1,5-Diaminonaphthalene was used as MALDI matrix, deposited via sublimation. Samples were analyzed in reflectron mode using a RapifleX Tissuetyper time-of-flight mass spectrometer with a laser raster size of 10 μm. Data was collected in negative-ion mode in the m/z 200–1600 range (Supplemental Section S4). Automated peak picking was first performed using SCiLS Lab, followed with manual peak curation to select the m/z features that were best associated with the cell colony regions. A mutual information metric was used as an alignment scoring function. MALDI and confocal images were passed on to a dual annealing global minimum approximation function[9] which rotated, scaled, and translated the average MALDI ion image until the best alignment with confocal colony image was found. Next, MALDI and confocal images were cropped by the aligned matching region (Figure 1c). Finally, we identified nuclei boundaries using confocal image stained with Hoechst nuclei live dye. A custom python script that used the scikit-image library[10] was applied to identify and tag the location and shape of each cell in the aggregate (Figure 1d). Next, these location and shape data were overlaid on each of the aligned m/z images (Figure 1e) allowing for the extraction of the intensity data at each m/z value on a cell by cell basis (Figure 1f) along with the corresponding fluorescence intensities from the confocal image.

RESULTS AND DISCUSSION

In multicellular differentiation, an individual cell state depends heavily on its surrounding environment. Thus, each co-registered image contains much more hidden information than just signals from individual cells; each cell has a unique set of neighboring cells that may influence cell fate decisions.[11] As such, it is useful to define a pair of cells as "neighbors" if they are closer to each other than a connection distance parameter (Figure S2a). To extract spatial information, additional neighborhood metrics designed to describe individual cells were introduced.

Number of Neighbors, Average Connection Length, and Local Density.

For every cell, the number of neighbors (N), the average distance between a cell and its neighbors (d), and the local density which equals the number of cells per area unit $\left(\frac{N}{d^2}\right)$ were calculated (Figure 2f).

Edge, Edge Distance, And Center Distance.

To determine if a cell was located on the edge of the colony, it was first checked if a line could be drawn from the current cell to one of its neighbors that would contain all of the neighbors of the initial cell on one side of the line (Figure S2b). Next, an *edge distance* metric was calculated as a distance to the closest edge cell. The distance from the center of the colony was approximated by the formula $\frac{1}{\text{edge\_distance} + 1}$.

Average Neighbor Value.

For each initial variable of every cell, the values among neighboring cells were averaged as a metric (Figure 2b).

Standard Deviation of Neighboring Values.

For each cell, a standard deviation of neighboring values was calculated (Figure 2e).

Self-Value over Average Neighbor Value.

For each initial variable of every cell, the ratio of the cell's signal value and average neighbor signal value was calculated.

Downsampling.

An alternative approach to data handling in case perfect alignment is not achievable due to imaging artifacts is to divide images into blocks by applying a rectangular grid and averaging the intensity data within the blocks to smooth out potential misalignment (Figure S2).

Randomness Metric.

For certain m/z values, MALDI MSI produces a signal distribution that resembles background noise (Figure 3a), whereas for other m/z values there are visibly organized clusters (Figure 3d). To quantify the level of such randomness, a spatial clustering algorithm that looks for continuous clusters of cells with similar levels of signal and provides the number of such clusters normalized by the number of cells was developed. This number ranges from a minimum value of $\frac{1}{n}$ (where n is the number of cells in the colony) when the distribution of signal is uniform and every cell has the same value, to a maximum of 1 when the signal is completely random with every cell being its own cluster. First, an average neighbor smoothing is applied to even out the high-frequency noise. Second, k-means clustering is used to divide the colony into k layers based on the signal intensity (e.g., $k = 2$ on Figure 3b,e). Third, each layer is treated as a graph with cells representing the nodes and with edges connecting all the neighboring cells. Next, every graph is passed on to the DBSCAN[12] clustering algorithm that outputs connectivity components (Figure 3c,f). There are two crucial parameters required—the length of connection e and the minimum cluster size. The latter allows to focus on bigger clusters if necessary, leaving multiple cells with no cluster assignment—in this case adjustments should be made to the randomness metric calculation to account for these cells (e.g., penalty for every unlabeled cell). Apart from the randomness metric, this algorithm yields other useful metrics, including the size of a component a cell belongs to, the average variable value in that component (Figure 2c), and the signal value of the cell over the average signal value of the component (Figure 2d).

Clustering Algorithms.

Raw MALDI MSI data stores a whole mass spectrum for each pixel. However, only select peaks are typically considered in data analysis pipelines. One reason is that not all of the peaks can be assigned to specific chemical species; another reason is an increase in

computational cost. One way to leverage the whole spectrum information for each pixel is through Pearson's correlation coefficient-based clustering.[13] First, the on-colony region of the image is segmented via histogram-based thresholding. Next, Pearson's correlation coefficient is calculated for each pixel in the image with an on-colony reference pixel chosen randomly. The Pearson's correlation coefficient spatial distribution (Figure 4b) clearly separates on-colony and off-colony regions as well as clusters of pixels with similar correlation coefficient value. To leverage this cluster information, k-means clustering with Pearson's correlation coefficient used instead of Euclidean distance is applied to MALDI MSI spectra (Figure 4c), demonstrating that distinct clusters with $k = 0, 1$ correspond to the off-colony pixels as compared to the confocal image. This demonstrates the potential of these methods for identifying occurrence of biologically relevant processes, such as loss of pluripotency, which can be verified through the aligned confocal image.

CONCLUSIONS

The described methods provide rich and robust data sets with an X-block comprised of m/z values and derived metrics, as well as cell shape and density characteristics, and a Y-block composed of fluorescence levels as a continuous label and days of the differentiation as a categorical label. Importantly, spatial features associated with MSI are not lost in subsequent analysis of m/z relations. Running regression or classification on the multimodal data set as a whole, or making use of the clustering technique to focus on a specific group of cells may reveal hidden biological interconnections between cell pluripotency and metabolic signatures.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

ACKNOWLEDGMENTS

REFERENCES

(1). Schwamborn K; Kriegsmann M; Weichert W MALDI imaging mass spectrometry—from bench to bedside. Biochim. Biophys. Acta, Proteins Proteomics 2017, 1865 (7), 776–783. [PubMed: 27810414]

(2). Wieser A; Schneider L; Jung J; Schubert S MALDI-TOF MS in microbiological diagnostics—identification of microorganisms and beyond. Appl. Microbiol. Biotechnol 2012, 93, 965. [PubMed: 22198716]

(3). Kriegsmann J; Kriegsmann M; Casadonte R MALDI TOF imaging mass spectrometry in clinical pathology: a valuable tool for cancer diagnostics. Int. J. Oncol 2015, 46 (3), 893–906. [PubMed: 25482502]

(4). Liu X; Flinders C; Mumenthaler SM; Hummon AB MALDI Mass Spectrometry Imaging for Evaluation of Therapeutics in Colorectal Tumor Organoids. J. Am. Soc. Mass Spectrom 2018, 29 (3), 516–526. [PubMed: 29209911]

(5). Gessel MM; Norris JL; Caprioli RM MALDI Imaging Mass Spectrometry: Spatial Molecular Analysis to Enable a New Age of Discovery. J. Proteomics 2014, 107, 71–82. [PubMed: 24686089]

(6). Soltwisch J; Kettling H; Vens-Cappell S; Wiegelmann M; Müthing J; Dreisewerd K Mass spectrometry imaging with laser-induced postionization. Science 2015, 348 (6231), 211–5. [PubMed: 25745064]

(7). Niehaus M; Soltwisch J; Belov ME; Dreisewerd K Transmission-mode MALDI-2 mass spectrometry imaging of cells and tissues at subcellular resolution. Nat. Methods 2019, 16 (9), 925–931. [PubMed: 31451764]

(8). Shimizu Y; Satou M; Hayashi K; Nakamura Y; Fujimaki M; Horibata Y; Ando H; Watanabe T; Shiobara T; Chibana K; Takemasa A Matrix-assisted laser desorption/ionization imaging mass spectrometry reveals changes of phospholipid distribution in induced pluripotent stem cell colony differentiation. Anal. Bioanal. Chem 2017, 409 (4), 1007–1016. [PubMed: 27815610]

(9). Xiang Y; Sun DY; Fan W; Gong XG Generalized Simulated Annealing Algorithm and Its Application to the Thomson Model. Phys. Lett. A 1997, 233, 216–220.

(10). van der Walt S; Schonberger JL; Nunez-Iglesias J; Boulogne F; Warner JD; Yager N; Gouillart E; Yu T scikit-image: Image processing in Python. PeerJ 2014, 2, No. e453. [PubMed: 25024921]

(11). Glen CM; McDevitt TC; Kemp ML Dynamic intercellular transport modulates the spatial patterning of differentiation during early neural commitment. Nat. Commun 2018, 9 (1), 4111. [PubMed: 30291250]

(12). Ester M; Kriegel HP; Sander J; Xu X A Density-Based Algorithm for Discovering Clusters in Large Spatial Databases with Noise In Proceedings of the 2nd International Conference on Knowledge Discovery and Data Mining; AAAI Press: Portland, OR, 1996; pp 226–231.

(13). Zhvansky E; Sorokin A; Ivanov D; Eliferov V; Bugrova A; Pekov S; Popov I; Nikolaev E Dissimilarity metrics mapping algorithm for assist region detection in mass spectrometry imaging Proceedings of the 67th ASMS Conference on Mass Spectrometry and Allied Topics; ASMS, 2019; 297092.
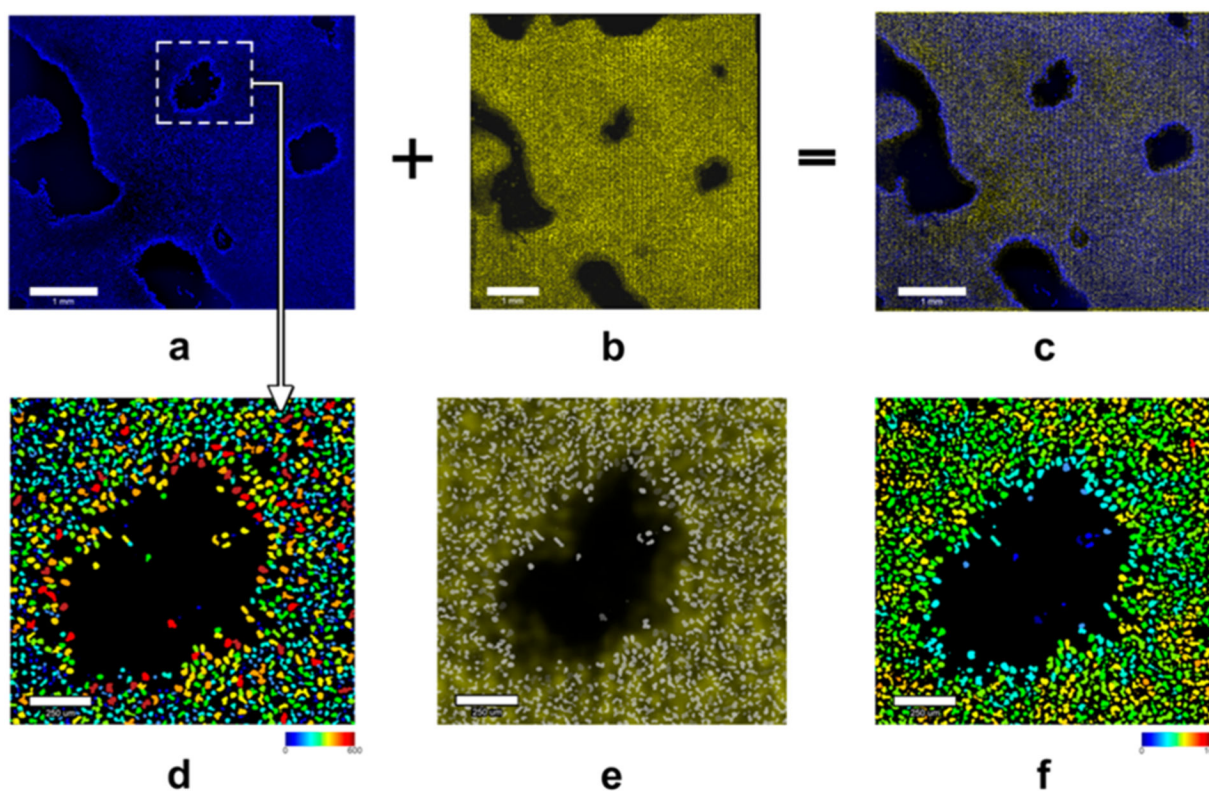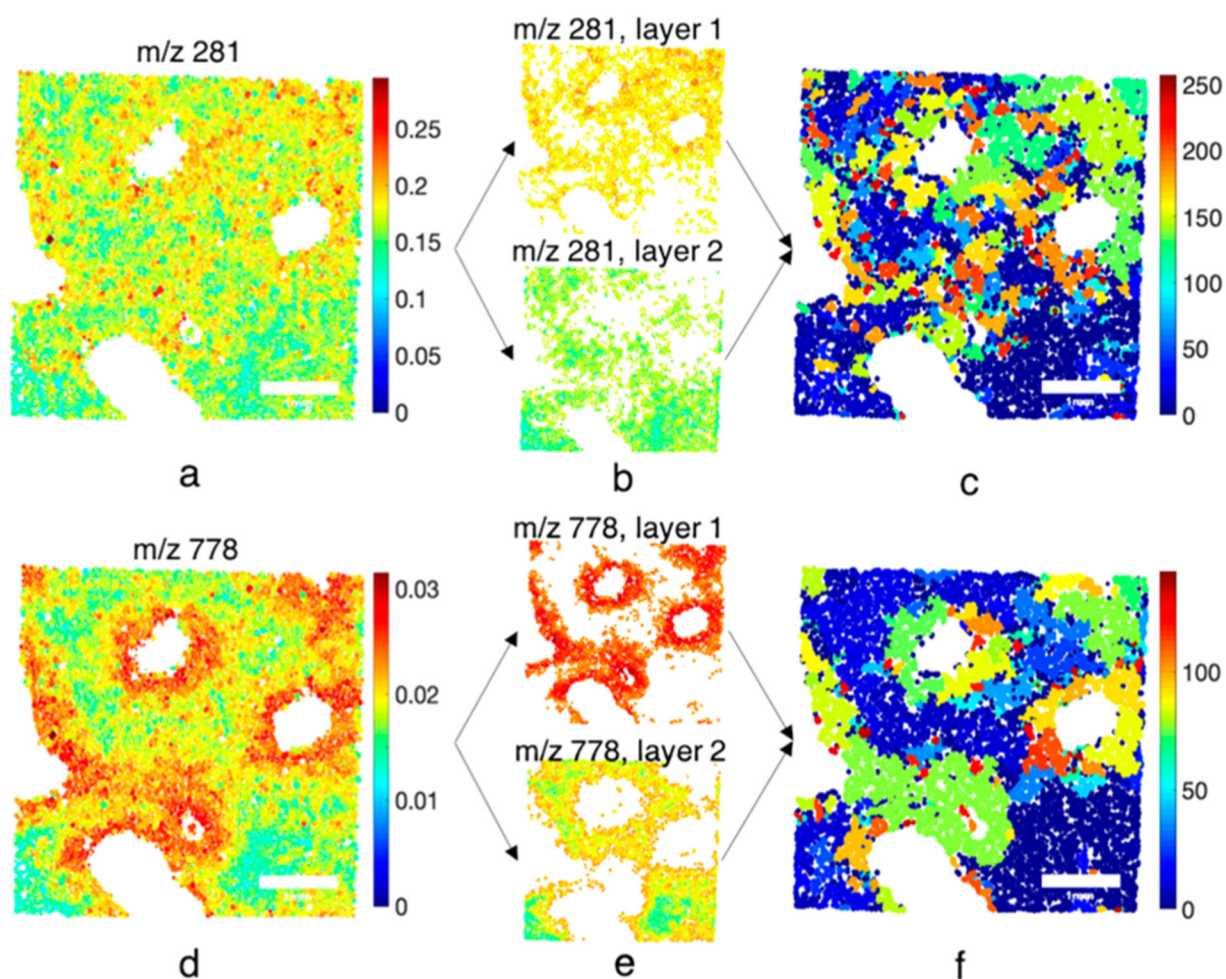
Figure 1.
Confocal and MALDI image co-registration pipeline. (a) Original confocal image stained with Hoechst nuclei live dye acquired on day 4. (b) Spatial distribution of m/z 887.553 (PI(38:3), [M - H]$^-$/PG(45:8), [M - H]$^-$) acquired via MALDI MSI on day 4. (c) MALDI and confocal images aligned. (d) Confocal image is processed with a custom Python script utilizing the *scikit-image* library to identify nuclei boundaries. Cell objects are colored according to their area in μm^2. (e) Cell objects are mapped onto the aligned MALDI MS image. Any observed mismatch may be explained by partial cell loss and leak of metabolites during the matrix deposition. (f) MALDI MSI data converted to single cell resolution data for extraction of network metrics. Cell objects are colored according to the intensity of m/z 887.553 peak.
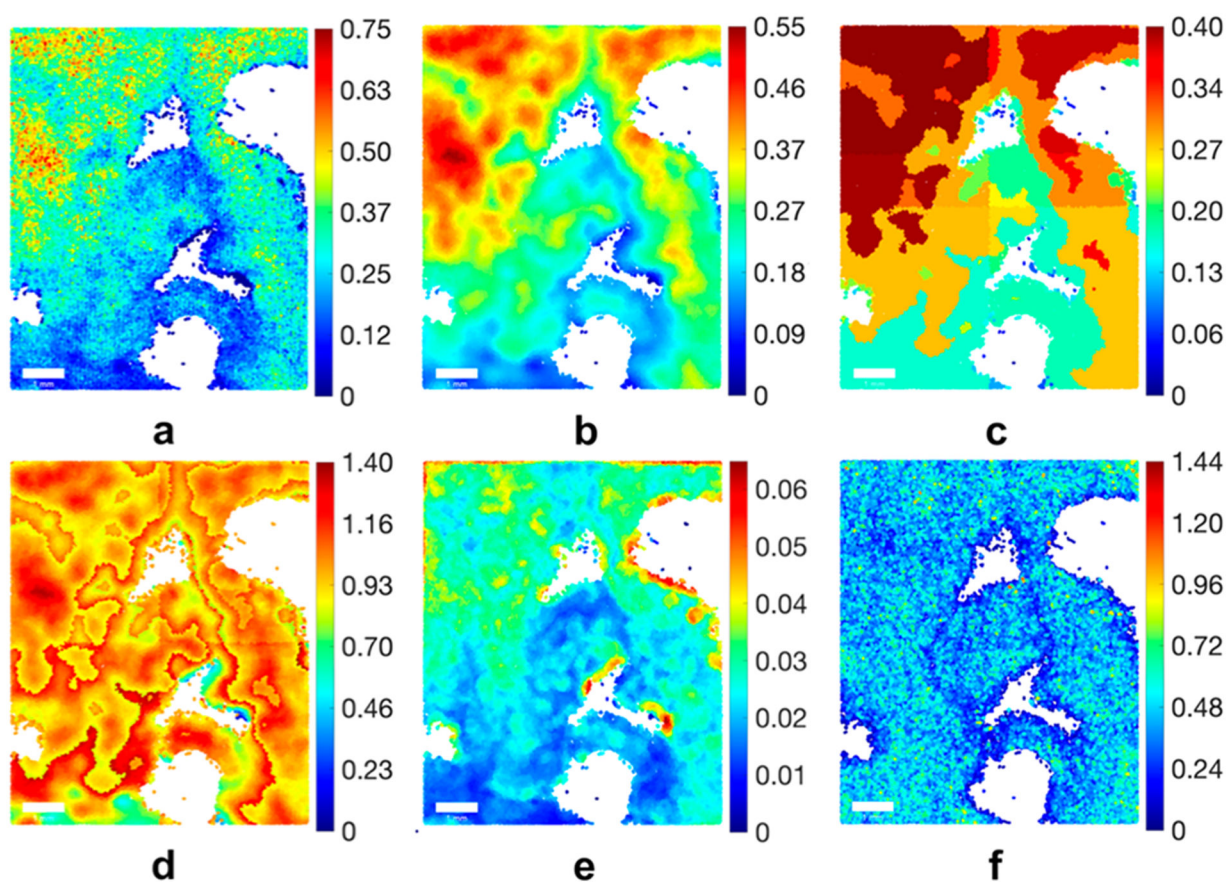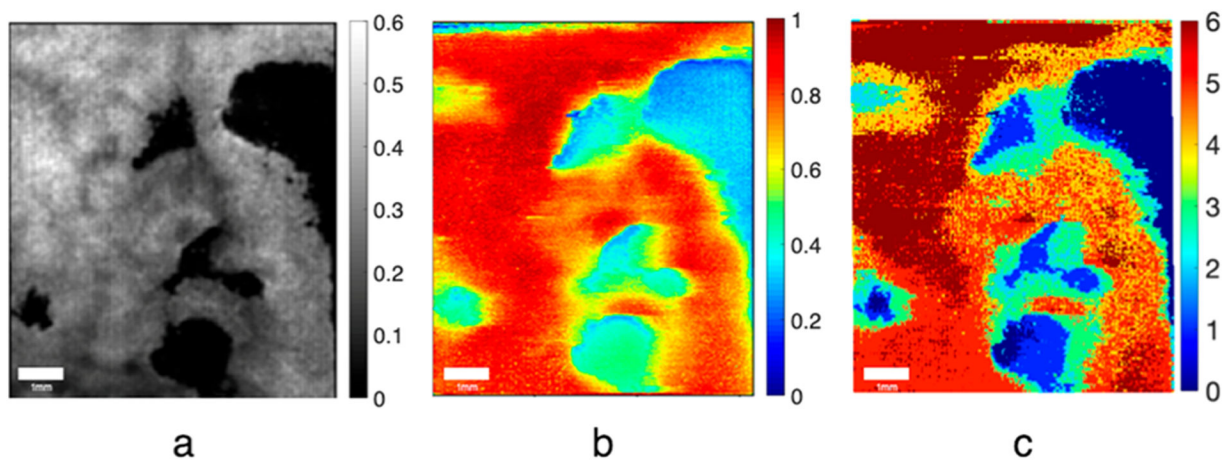
Figure 2.
Examples of spatial metrics distributions for a colony undergoing day 5 of a spontaneous differentiation protocol. (a) Cell-by-cell intensities extracted from m/z 887.553 (PI(38:3), [M - H]$^-$/PG(45:8), [M - H]$^-$) ion image co-registered with the corresponding confocal image. (b) Average neighbor metric applied to (a). (c) Average cluster value metric applied to (b). (d) Dividing initial values from (a) by average cluster values from (c) shows how cells' m/z signal relates to average m/z signal in their respective clusters. (e) Standard deviation in the neighborhoods of cells from (a) highlights regions with local variability in levels of the m/z signal. (f) Local density distribution calculated as the number of cells per area unit $\left(\frac{N}{d^2}\right)$ where N is the number of cell neighbors and d is the average distance between a cell and its neighbors. As expected, density is lower along the edges.

Figure 3.
Two examples (a–c and d–f) of spatial clustering of m/z signal distributions yielding different randomness metrics. Spatial distribution of m/z 281.330 (FA(18:1), [M - H]⁻) and m/z 778.606 (PE(39:5), [M - H]⁻) following co-registration as described in Figure 1. (b, e) Colony divided into $k = 2$ intensity layers via k-means clustering. (c, f) Each layer undergoes separation into connectivity components via DBSCAN clustering algorithm-different colors represent distinct connectivity components. Clusters of connected cells with similar values are highlighted by joining k layers together. The image in Figure 2c yields many small clusters that results in a randomness value of 0.45, twice as random as Figure 2f with bigger clusters and a randomness value of 0.23. Scalebars are 1 mm.

Figure 4.
Clustering by Pearson's correlation coefficient with a reference pixel. (a) Spatial distribution of m/z 887.553 (PI(38:3), $[M - H]^-$/PG(45:8), $[M - H]^-$) acquired via MALDI MSI on day 5. (b) Pearson's correlation coefficient distribution with an on-colony reference pixel selection. Off-colony/background pixels have clearly distinct lack of correlation compared to on-colony regions. (c) K-means clustering ($k = 7$) of pixels based on their Pearson's correlation coefficient.