UNIVERSITY OF OKLAHOMA

GRADUATE COLLEGE

SINGLE-PROBE MASS SPECTROMETRY IMAGING: APPLICATIONS AND ADVANCED DATA ANALYSIS

A DISSERTATION

SUBMITTED TO THE GRADUATE FACULTY

in partial fulfillment of the requirements for the

Degree of

DOCTOR OF PHILOSOPHY

By

XIANG TIAN

Norman, Oklahoma

2019

SINGLE-PROBE MASS SPECTROMETRY IMAGING: APPLICATIONS AND ADVANCED DATA ANALYSIS

A DISSERTATION APPROVED FOR THE

DEPARTMENT OF CHEMISTRY AND BIOCHEMISTRY

By THE COMMITTEE CONSISTING OF

Dr. Zhibo Yang, Chair

Dr. Anthony Burgett

Dr. Si Wu

Dr. Chuanbin Mao

Dr. Yuchen Qiu

This doctoral dissertation is dedicated to my family.

Acknowledgments

About 1200 years ago, a famous Chinese philosopher and poet, Yu Han, said: “a great teacher propagates the doctrine, imparts the professional knowledge, and solves doubts/师者，传道受业解惑也”. Dr. Zhibo Yang is a great teacher, who dedicates his time and passion to teach his students and creat new ideas. I am so lucky to have him as my advisor. Without his guidance and help, none of my achievements can be done. I would like to thank him from the bottom of my heart. I cherish the time spent in this lab very much, and whenever I look back, the five and half years’ experience will continuously encourage and energize myself to confront every difficulty for the rest of my life.

I would like to thank my committee member Dr. Chuanbin Mao, Dr. Si Wu, Dr. Anthony Burgett, Dr. Yuchen Qiu and former committee member Dr. Libault Marc for attending the evaluation meeting every year and giving me lots of valuable advice.

I would like to thank all my lab members, Dr. Ning Pan, Dr. Genwei Zhang, Dr. Mei Sun, Dr. Renmeng Liu, Dr. Shawna Standke, Yanlin Zhu, Xinxiu Chen, Zhu Zou, Zongkai Peng, Yunpeng Lan, and Tra Ngyuen. I also want to thank Dr. Wei Rao who taught me a lot about mass spectrometry imaging.

Table of Contents

List of Tables

List of Figures

Abstract

Mass spectrometry imaging (MSI) is becoming a powerful tool in the bioanalytical studies owing to its unique capability to sensitively map the spatial distribution of broad ranges of molecules on biological samples. Due to the large size and complex structure of the image datasets, conventional analysis methods, such as directly mapping the selected ions, is insufficient to achieve comprehensive data analysis. To increase the data analysis efficiency and fully extract the information contained in MS image data, advanced data analysis methods are needed. This dissertation focuses on the studies using the combined Single-probe MSI method, an ambient MSI technique, with advanced data analysis, including multivariate curve resolution (MCR), machine learning (ML), and multi-modal imaging fusion. MCR is a technique to decompose the hyperdimensional dataset into major components, which possess similar spatial distributions, and extract molecules from each component. ML is becoming increasingly popular for analyzing MSI data due to its superior capability to deal with big data. Here, both supervised and unsupervised ML methods were used to segment the MSI data, providing fast and accurate approaches to image segmentation. In addition, image fusion technique was applied to enhance the MSI data analysis, in which microscope images and MS images were fused together to increase the spatial resolution and correlate the spatial information of protein biomarkers and metabolites. The integration of the Single-probe MSI experimental techniques and advanced data analysis methods can potentially benefit fundamental research and broad types of applications such as in drug discovery and studies of disease.

Chapter1. Introduction

1.1. Background

Over the past decade, mass spectrometry imaging (MSI) is an emerging technique to visualize the spatial distribution of molecules on the biological sample surface[1]. The MSI techniques can provide rich molecular information that cannot be obtained from traditional imaging methods such as MRI (Magnetic Resonance Imaging), PET (positron emission tomography), and fluorescence imaging, and can tremendously benefit biological, biochemistry and pharmaceutical studies[2, 3]. MSI has the advantage of imaging thousands of molecules including lipids, peptides, proteins, and metabolites without labeling[4]. The chemical information is usually collected by mass spectrometer via spot-to-spot analysis or rastering on the defined sample surface[5]. Then, MSI software is utilized to generate ion maps by using m/z (mass to charge ratio) values of target ions and their ion intensities, which are illustrated using color scales. To fully understand the MSI data, advanced data analysis methods were also developed to segment the image into different regions (e.g., tumor region and healthy region) basing on the differences of the mass spectrum profile[6].

1.2. Mass spectrometry imaging technique

To date, a variety of MSI methods have been developed, and they can be divided into two general categories basing on the sampling and ionization environment: ambient and non-ambient techniques.

1.2.1. Ambient ionization technique

1.2.1.1. DESI: desorption electrospray ionization

DESI was first developed by Takáts *et al*[7] in 2004, and it opens the gate of directly detecting analytes on a biological sample surface under the ambient condition. DESI uses high voltage to

generate the organic solvent spray (with the assist of nebulizer gas) on the sample surface, and the desorbed ions (e.g., lipids, peptides, proteins, and metabolites) are then transferred by the spray droplets towards the inlet of the mass spectrometer (Figure 1-1). The advantage of DESI is that it requires less sample preparation and no need for using matrix molecules. However, due to the size of spray, the spatial resolution of this technique is relatively low (about 150-250 μm)[8].

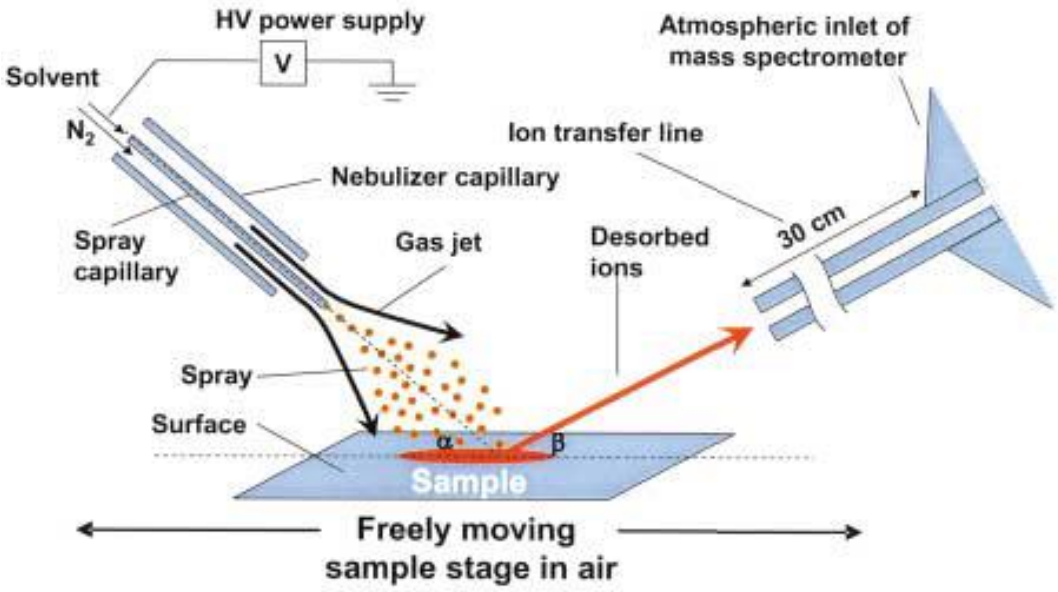


Figure 1-1. Schematic of typical DESI experiment

The sample solution was deposited from solution and dried onto a PTFE surface, and methanol-water (1:1 containing 1% acetic acid or 0.1% aqueous acetic acid solution) was sprayed at a flow rate of 3 to 15 μL/min under the influence of a high (4 kV) voltage. The nominal linear velocity of the nebulizing gas was set to 350 m/s (Takáts, Z.; Wiseman, J. M.; Gologan, B.; Cooks, R. G., *Science* **2004,** *306* (5695), 471).

1.2.1.2. Nano-DESI: Nano desorption electrospray ionization.

Julia Laskin's group developed the Nano-DESI technique,[9] which uses the liquid bridge between two capillaries to extract analytes on the substrate as shown in Figure 1-2[10]. During the sampling process, the primary capillary maintains the flow to the liquid bridge and the Nano-spray capillary transfers the liquid with analytes to the MS inlet. Instead of using gas jet in the DESI, Nano-DESI using a small liquid bridge to desorb the analyte, allowing for higher spatial resolution (up to 12 μm as reported) because of the small contact area between liquid and sample surface.

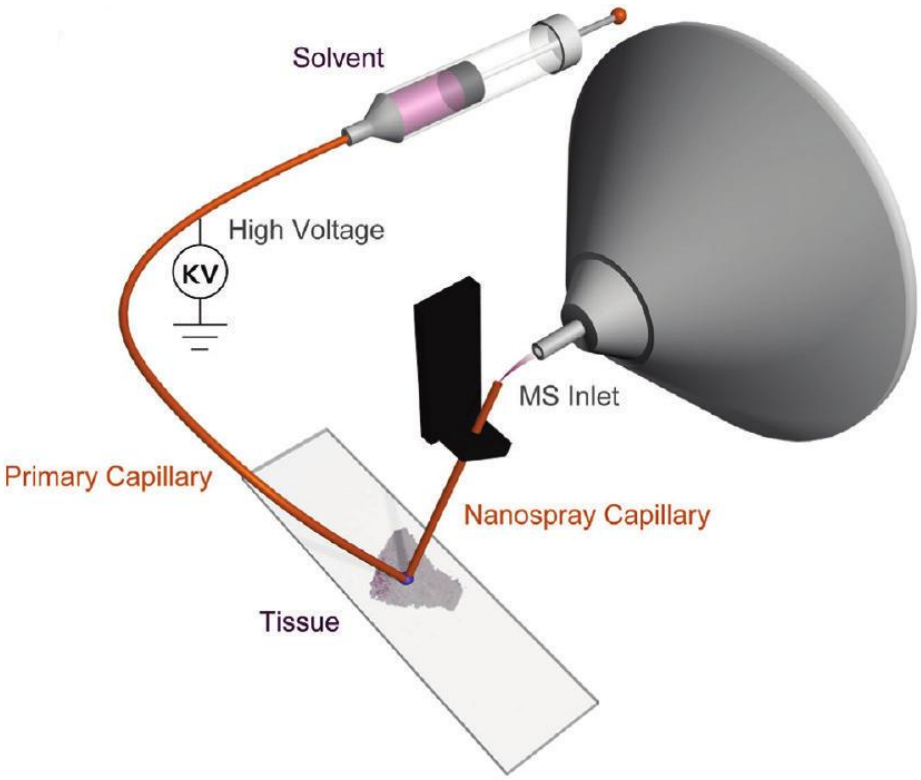


Figure 1-2. Schematic drawing of the nano-DESI ion source.

(Laskin, J.; Heath, B. S.; Roach, P. J.; Cazares, L.; Semmes, O. J., *Anal. Chem.* **2012**, 84 (1), 141-8).

1.2.1.3. LMJ-SSP: liquid micro junction surface sampling probe

LMJ-SSP was first introduced by Gary J. Van Berkel[11]. The schematic illustration of this technique was shown in Figure 1-3. LMJ-SSP uses two coaxial tubings to transfer liquid to the sampling

analyte when the probe touches the sample surface. As illustrated in Figure 1-3 (b) and (c), the channel between the inner and outer tubing is utilized to provide the sampling solvent, and the inner tubing is used to aspirate the solvent and connect to ionization interface including ESI[12] and APCI[13]. Due to the large size of the liquid junction, the resolution is generally low (around 500-1000 μm).[8]

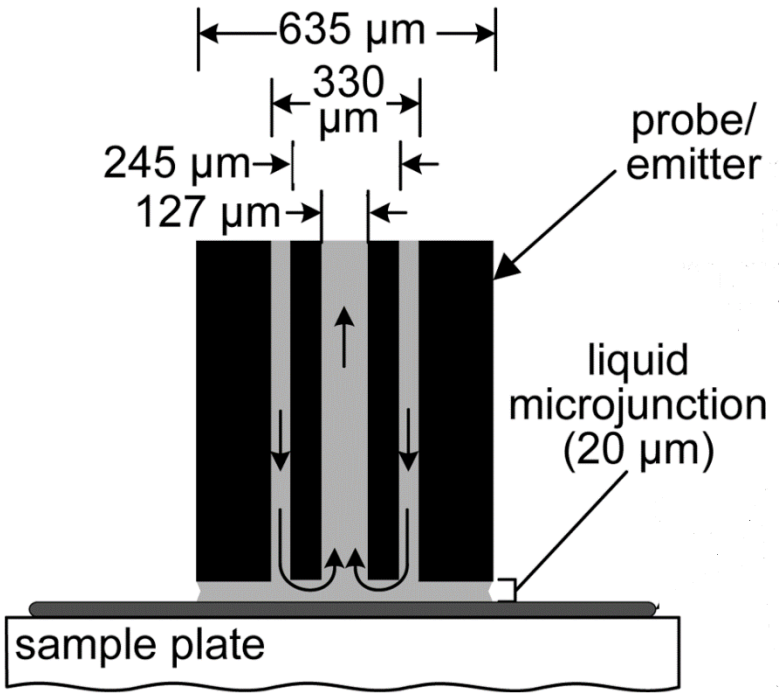


Figure 1-3. Schematic illustration of the liquid micro-junction surface.

Schematic illustration showing the close probe-to-surface spacing and narrow liquid micro-junction used for spot sampling. (Van Berkel, G. J.; Kertesz, V.; King, R. C., *Anal. Chem***. 2009**, 81 (16), 7096-7101)

1.2.1.4. LAESI: laser ablation electrospray ionization

LAESI was developed by Nemes *et al* [14] in 2007 which is an ionization technique combining the infrared laser ablation and electrospray ionization together. Similar to other ambient ionization methods, it does not need any sample preparation or pretreatment and the large mass coverage (up to 66kDa) is also one of the major advantages.

The Schematics of LAESI was shown in Figure 1-4. A mid-infrared laser source was chosen because most targets are water-rich molecules and -OH vibration has strong absorption in this wavelength region[15]. Under laser ablation, the analytes on the sample surface were converted to gas-phase neutrals and mixed with the orthogonal electrospray. Then, the charged electrospray droplet carries the analytes towards MS inlet.

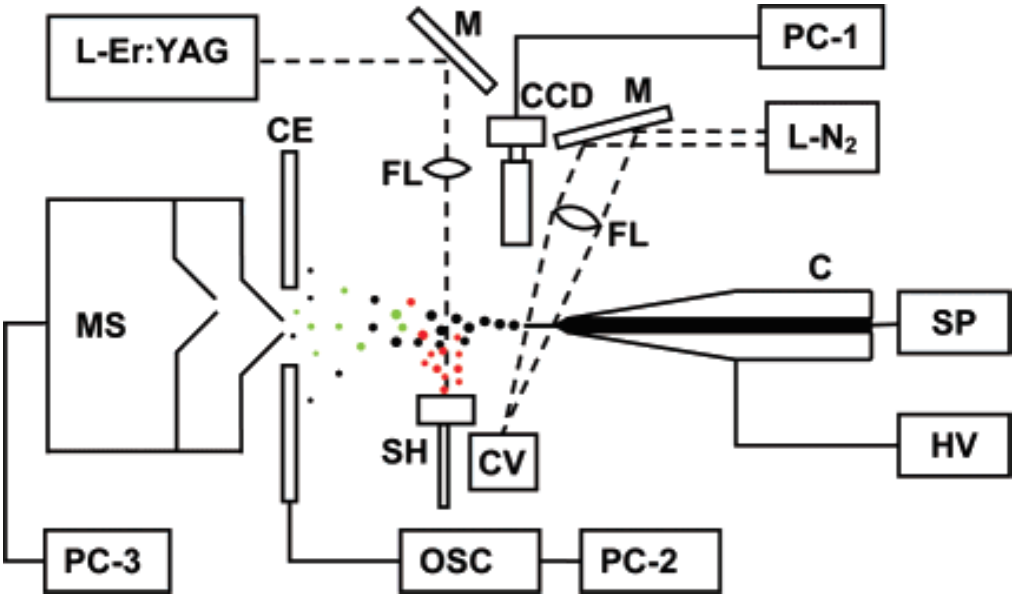


Figure 1-4. Schematics of laser ablation electrospray ionization.

(C, capillary; SP, syringe pump; HV, high-voltage power supply; L-N2, nitrogen laser; M, mirrors; FL, focusing lenses; CV, cuvette; CCD, CCD camera with short-distance microscope; CE, counter electrode; OSC, digital oscilloscope; SH, sample holder; L-Er:YAG, Er:YAG laser; MS, mass spectrometer; PC-1 to PC-3, personal computers). The cone-jet regime is maintained through monitoring the spray current on CE and adjusting the spray parameters. Black dots represent the droplets formed by the electrospray. Their interaction with the particulates and neutrals (red dots) emerging from the laser ablation produces some fused particles (green dots) that are thought to be the basis of the LAESI signal. (Nemes, P.; Vertes, A., *Anal. Chem.* **2007,** *79* (21), 8098-8106).

1.2.1.5. Single-probe imaging technique

The Single-probe was a multifunctional sampling probe developed by Zhibo Yang's group in 2014. So far, the Single-probe has been applied in many research fields including single cell analysis[16], mass spectrometry imaging,[17] and detection of the extracellular metabolites in spheroid[18].

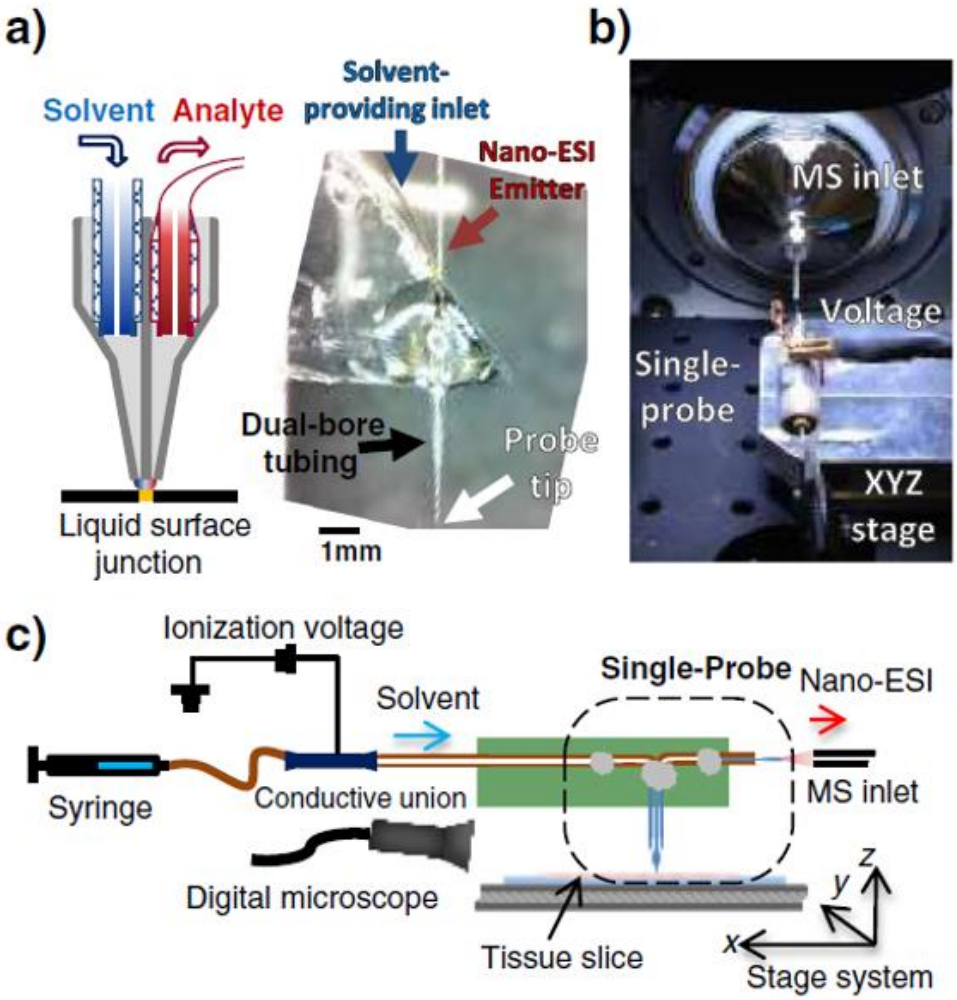


Figure 1-5. Schematic of Single-probe mass spectrometry imaging.

(a) Diagram and photograph of the Single-probe showing the various features of the device. (b) Photography of the general setup of the Single-probe during MSI measurement. (c) Schematic of the Single-probe MSI system. (Rao, W.; Pan, N.; Yang, Z. B., *J. Am. Soc. Mass Spectrom.* **2015,** *26* (6), 986-993.)

The working principle of the Single-probe MSI is illustrated in Figure 1-5. In general, during the sampling, the sampling solvent is continuously delivered from one channel to form a liquid surface junction between the probe tip and sample surface, where the liquid junction extracts the analytes. The extracted molecules are delivered to the nano-ESI emitter through the other channel for MS analysis. The highest spatial resolution can be obtained this technique is 8.5 μm[17]. It worths to note that the tip size is adjustable according to different experiments. The tip size is usually smaller than 10 μm in the single cell analysis to allow the tip insert into an individual cell; in the imaging experiment, the tip size is about 50 μm to eliminate the clogging issue.

1.2.2. Non-ambient ionization technique

1.2.2.1. MALDI: matrix-assisted laser desorption ionization

MALDI, one of the most popular non-ambient ionization techniques due to the ability to measure intact masses of a wide molecular mass range of biological molecules[19], was first named by Franz Hillenkamp

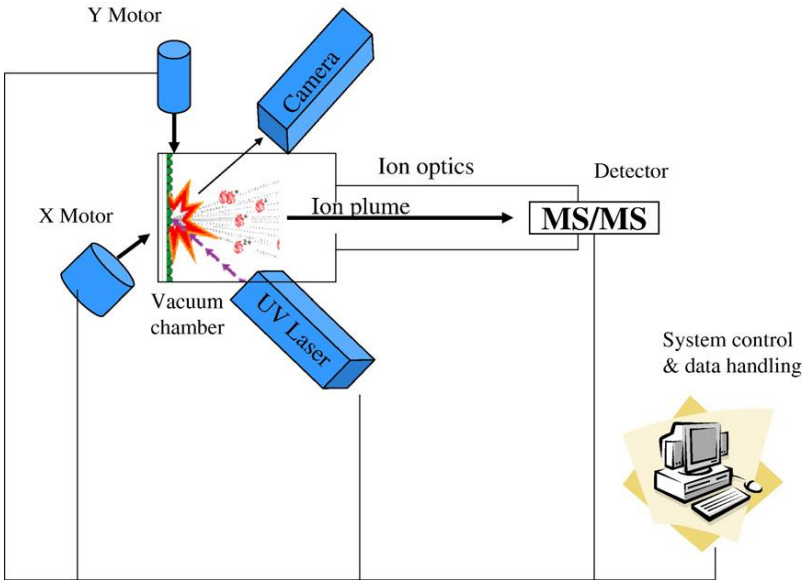


Figure 1-6. Diagram for an automated MALDI-IMS system used for direct analysis of pharmaceuticals in tissues

(Hsieh, Y.; Chen, J.; Korfmacher, W. A., J. Pharmacol. Toxicol. Methods 2007, 55 (2), 193-200).

and Michael Karas [20] in 1985. The first MALDI mass spectrometry imaging was conducted by Richard M. Caprioli to locate the peptide and protein on the biological sample surface[21].

The schematic of the MALDI is shown in Figure 1-6. The matrix crystal absorbs the laser beam energy, resulting in the desorption and ionization of the analytes embedded into the crystals [21, 22].

1.2.2.2. SIMS: secondary ion mass spectrometry

SIMS is another well-known non-ambient ionization technique, and the first development of this technique is mainly contributed by Benninghoven in 1970s[23]. As shown in Figure 1-7, SIMS uses the primary ion beam (e.g. He^+, Ne^+, and C_{60}^+) to bombard the sample surface, and the secondary ions (fragments) are generated if the kinetic energy is high enough (e.g., higher than the bond energies of analytes) [24]. Due to the extremely small size of the ion beam, SIMS has the highest spatial resolution (nm level) as reported[25]

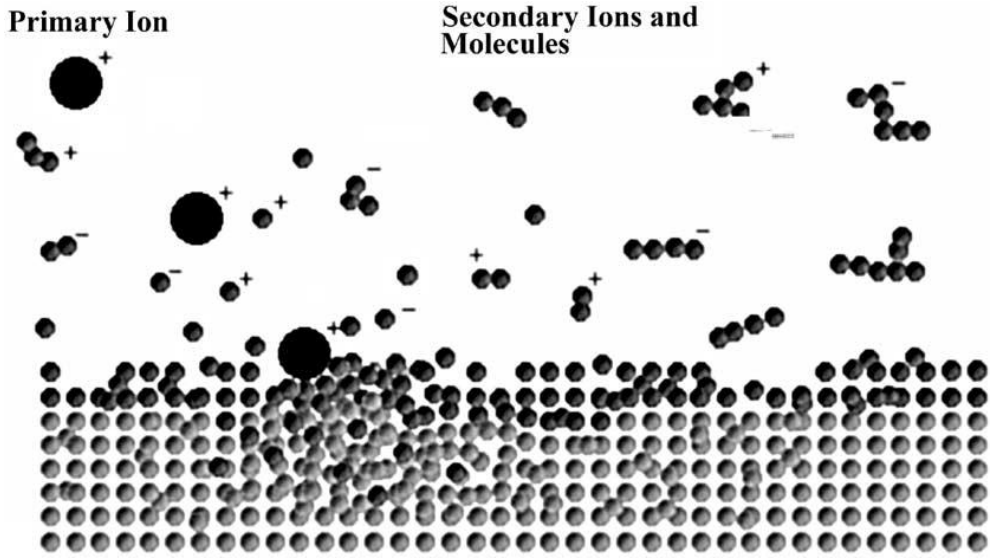


Figure 1-7.A schematic drawing of the secondary ion emission process initiated by the impact of a primary ion.

Extensive fragmentation occurs near the collision site producing mainly atomic particles. Away from the point of impact collisions become less energetic resulting in the emission of larger molecular fragments. (Belu, A. M.; Graham, D. J.; Castner, D. G., *Biomaterials* **2003,** *24* (21), 3635-3653).

1.3. MSI Application

Due to the increasing amounts of demands in biological research, a number of MSI methods have been developed for broad ranges of applications. So far, MSI techniques play important roles in many research fields (e.g. metabolomics[26, 27], proteomics[28], natural product[29], and drug discovery[30]) mainly due to its unique capability of localizing molecules on the biological sample.

1.3.1. Metabolomics

Metabolomics is the systematic study focusing on small molecules (< 1500 Da) in biological samples. As the endpoint of "omics cascade", metabolites can directly reflect the status of a cell, tissue, organ, and whole-body under the different environments[31], and they become the interest of biology and biochemistry studies. Although, to date, the mainstream of metabolomics methods are still dominated by MS[31] (Mass Spectrometry), especially LC-MS (Liquid Chromatography Mass Spectrometry) and GC-MS (Gas Chromatography Mass Spectrometry), imaging techniques (e.g. NMR (Nuclear Magnetic Resonance), PET (Positron emission tomography), and fluorescence microscopy) also serve as indispensable tools in the metabolomics study, mainly because that the spatial distribution of molecules within the organism is essential to decode the complex biochemistry process[26]. However, these above imaging techniques have low coverage of molecule even though with high sensitivity of mapping molecules location[32]. MSI, with high detection sensitivity and broad molecular coverage, becomes a powerful platform to study metabolomics. For example, DESI was implemented to determine the spatial distribution of fatty acids and phospholipid in mouse brain[33]. Although MALDI is rarely used for low-molecule-weight (<500) metabolites because of the matrix effect, it was still utilized to map some lipids molecules[34].

1.3.2. Proteomics

Proteomics focuses on the large-scale study of proteins or peptides in biological samples[35], and it is a very important platform for biological and biochemistry studies[36]. Like metabolomics, the conventional methods almost based on mass spectrometry coupled with some prior separation techniques, such as GC-MS, LC-MS, and CE (capillary electrophoresis)-MS[37]. However, due to the homogenized sample preparation process, these methods lack the protein spatial information, which is also very significant in the proteomics study[38]. MSI, as the MS-based method, has the intrinsic advantages over the traditional protein imaging methods on chemical specificity and coverage. For example, immunohistochemistry (IHC) and fluorescence microscope image need target proteins prior to experiments[28].

So far, there are many proteomics studies using MSI technique especially MALDI MSI due to its ability of detecting peptides and proteins. For example, Groseclose *et.al*[28] used MALDI MSI to determine proteins' expression and relative quantification of multiple lung cancer patient tissues. The comparison of protein profiles may provide important information of different disease status and evaluation of clinical treatment effectiveness.

1.3.3. Natural product and drug discovery

Plants, microbes, and animals are always the main sources of natural products in drug discovery. Since 1994, more than half of the discovered drugs belong to natural products or their derivatives [39, 40]. Traditionally, drug discovery from natural product requires complex sample preparation, such as multiple times extraction from bulk sample, concentration, and separation basing on molecules' different properties (e.g., affinity, polarity), before determining the structure of compounds.[29] Recently, MSI becomes increasingly more important in drug screening and detection of drug metabolites in tissue. Compared with conventional methods, MSI has unique

advantages such as less sample preparation process[26] and determining the spatial distribution of natural products or drugs on tissue[41] as mentioned in section 1.2. To date, a number of studies have been done using MSI to achieve drug distribution in animal samples. For example, MALDI MSI was implemented to determine the spatial distribution of Raclopride, a dopamine D2 receptor-selective antagonist[42], in multiple organs of dosed mice[43]. The results of MALDI MS images provide detailed information on Raclopride accumulated differently at different organs, which can potentially benefit early drug development. Another example is using DESI imaging to identify the distribution of Clozapine and its metabolites in mice after oral dose (50mg/kg)[44]. From these studies, Clozapine was found exist in whole body, however, the N-dimethyl metabolites of Clozapine were only observed in the lung sections, indicating the unique ability of MSI in the pharmaceutical research.

1.4. Data analysis

MSI data generally contain large data size and high dimensional structure due to their inherent complex chemical and special information[5]. To comprehensively analyze the MSI data and extract the hidden information, many methods and software packages are developed.

1.4.1. Methods

MSI data contain rich chemical information (e.g., m/z values and intensities of large numbers of ions) and spatial information (e.g., 2D or 3D coordinates). The most common method to construct an ion image is to select a certain m/z value and generate its spatial distribution along with its intensity[7]. However, this method needs prior knowledge to select the target ions. In addition, is relatively time-consuming to generate every ion's image in order to fully achieve the analysis of MSI data[45]. To overcome this drawback, multivariate analysis has been utilized to mine MSI data. Among them, the most commonly used is the Principle Component Analysis (PCA)[46]. In PCA

results, most information of the image can be represented by the score images of first a few components. For example, Allison L. Dill *et al*[47] used PCA to successfully differentiate the bladder cancer tissue from normal tissue in the DESI imaging data. However, the intrinsic disadvantage of PCA is that the negative value in the score plot does not have any physical meaning[6]. Probabilistic latent semantic analysis (pLSA) was utilized to solve this problem; however, it needs define the numbers of the component before analysis[48]. In addition to PCA and pLSA, other clustering methods (Hierarchical Clustering[49], *k*-means[50], and fuzzy c-means[51]) are used to segment the tissue according to the mass spectra profiles, which are very important for MSI data analysis. Recently, Machine learning methods, including both supervised and unsupervised algorithms, have also been implemented to MSI data analysis due to their advantage of dealing big data. Unsupervised machine learning is used to cluster imaging pixels into different groups according to their similarity and does not require any prior knowledge[52]. For supervised machine learning, training data are needed to train and optimize the machine learning model, prior to predicting large unknown datasets. For example, Zhou *et al*[53] using supervised machine learning to differentiate the fingerprints of Chinese male and India female from 194 samples with about 90% accuracy.

1.4.2. Commercial and in-house developed software packages

So far, a number of software packages focusing on MS imaging data analysis have been developed and they can be generally divided into commercialized software packages or in-house developed data analysis methods. Commercialized software is usually developed by companies. For example, BioMaps, a free commercial imaging data analysis software (https://ms-imaging.org/wp/biomap/), was first developed by Novartis to analyze MRI data since 1996[54]. After decades of modification, it was also utilized to process different types of imaging data such as optical, PET, CT (Computed tomography), NIRF (Near-infrared Fluorescence Imaging) and MSI. Besides generating the target

ion map, BioMaps has other important functions, including Coregistration, analysis spectra difference among different ROI (region of interest), and database searching. FlexImaging is an instrument software developed by Bruker to acquire and process MALDI imaging data such as data classification in proteomics researches.[28] Other instrument manufacturers also developed software packages compatible with their own instruments. For example, Shimadzu and Thermo Fisher developed Axima 2 Analyze and ImageQuest, respectively[55]. Generally, all these software packages have ion image reconstruction functions and some common data analysis methods.

In-house developed tools are also commonly used for MSI data analysis. A number of research labs utilize their own in-house developed script, which is written in MATLAB, R and Python languages, to process the MSI data. For example, MCR-ALS (multivariate curve resolution alternating least squares), a multivariate data analysis in Bioinformatics ToolBox of MATLAB, is developed by Tauler *et al*[56] in 2005 and it has been applied to analysis MSI data[57]. The merits of this method are that it not only extracts the major component of the image but also group the ions in each component which has a similar distribution. In addition, some data segmentation methods, which are also known as unsupervised machine learning (e.g., k-means[50] and Hie rarchical Clustering[49]), were coded in R or MATLAB to process the MSI data.

References

(1) Stoeckli, M.; Chaurand, P.; Hallahan, D. E.; Caprioli, R. M., *Nat. Med.* **2001,** *7* (4), 493-496.

(2) Franck, J.; Arafah, K.; Elayed, M.; Bonnel, D.; Vergara, D.; Jacquet, A.; Vinatier, D.; Wisztorski, M.; Day, R.; Fournier, I.; Salzet, M., *Molecular & Cellular Proteomics* **2009,** *8* (9), 2023.

(3) Beger, R. D., *Metabolites* **2013,** *3* (3), 552-74.

(4) Buchberger, A. R.; DeLaney, K.; Johnson, J.; Li, L., *Anal. Chem.* **2018,** *90* (1), 240-265.

(5) Amstalden van Hove, E. R.; Smith, D. F.; Heeren, R. M., *J. Chromatogr. A* **2010,** *1217* (25), 3946-54.

(6) Alexandrov, T., *BMC Bioinformatics* **2012,** *13 Suppl 16*, S11.

(7) Takáts, Z.; Wiseman, J. M.; Gologan, B.; Cooks, R. G., *Science* **2004,** *306* (5695), 471.

(8) Laskin, J.; Lanekoff, I., *Anal. Chem.* **2016,** *88* (1), 52-73.

(9) Roach, P. J.; Laskin, J.; Laskin, A., *Analyst* **2010,** *135* (9), 2233-6.

(10) Laskin, J.; Heath, B. S.; Roach, P. J.; Cazares, L.; Semmes, O. J., *Anal. Chem.* **2012,** *84* (1), 141-8.

(11) Van Berkel, G. J.; Kertesz, V.; King, R. C., *Anal. Chem.* **2009,** *81* (16), 7096-7101.

(12) Van Berkel, G. J.; Sanchez, A. D.; Quirke, J. M. E., *Anal. Chem.* **2002,** *74* (24), 6216-6223.

(13) Asano, K. G.; Ford, M. J.; Tomkins, B. A.; Van Berkel, G. J., *Rapid Commun. Mass Spectrom.* **2005,** *19* (16), 2305-2312.

(14) Nemes, P.; Vertes, A., *Anal. Chem.* **2007,** *79* (21), 8098-8106.

(15) Vogel, A.; Venugopalan, V., *Chem. Rev.* **2003,** *103* (5), 2079-2079.

(16) Pan, N.; Rao, W.; Kothapalli, N. R.; Liu, R.; Burgett, A. W.; Yang, Z., *Anal. Chem.* **2014,** *86* (19), 9376-80.

(17) Rao, W.; Pan, N.; Yang, Z. B., *J. Am. Soc. Mass Spectrom.* **2015,** *26* (6), 986-993.

(18) Sun, M.; Tian, X.; Yang, Z., *Anal. Chem.* **2017,** *89* (17), 9069-9076.

(19) Angel, P. M.; Caprioli, R. M., *Biochemistry* **2013,** *52* (22), 3818-28.

(20) Karas, M.; Bachmann, D.; Hillenkamp, F., *Anal. Chem.* **1985,** *57* (14), 2935-2939.

(21) Caprioli, R. M.; Farmer, T. B.; Gile, J., *Anal. Chem.* **1997,** *69* (23), 4751-4760.

(22) Hsieh, Y.; Chen, J.; Korfmacher, W. A., *J. Pharmacol. Toxicol. Methods* **2007,** *55* (2), 193-200.

(23) Benninghoven, A., *Surf. Sci.* **1973,** *35*, 427-457.

(24) Belu, A. M.; Graham, D. J.; Castner, D. G., *Biomaterials* **2003,** *24* (21), 3635-3653.

(25) Levi-Setti, R.; Hallégot, P.; Girod, C.; Chabala, J. M.; Li, J.; Sodonis, A.; Wolbach, W., *Surf. Sci.* **1991,** *246* (1), 94-106.

(26) Miura, D.; Fujimura, Y.; Wariishi, H., *J. Proteomics* **2012,** *75* (16), 5052-60.

(27) Tian, X.; Xie, B.; Zou, Z.; Jiao, Y.; Lin, L. E.; Chen, C. L.; Hsu, C. C.; Peng, J.; Yang, Z., *Anal. Chem.* **2019,** *91* (20), 12882-12889.

(28) Groseclose, M. R.; Massion, P. P.; Chaurand, P.; Caprioli, R. M., *Proteomics* **2008,** *8* (18), 3715-24.

(29) Esquenazi, E.; Yang, Y. L.; Watrous, J.; Gerwick, W. H.; Dorrestein, P. C., *Nat. Prod. Rep.* **2009,** *26* (12), 1521-34.

(30) Prideaux, B.; Stoeckli, M., *J. Proteomics* **2012,** *75* (16), 4999-5013.

(31) Dettmer, K.; Aronov, P. A.; Hammock, B. D., *Mass Spectrom. Rev.* **2007,** *26* (1), 51-78.

(32) Judenhofer, M. S.; Wehrl, H. F.; Newport, D. F.; Catana, C.; Siegel, S. B.; Becker, M.; Thielscher, A.; Kneilling, M.; Lichy, M. P.; Eichner, M.; Klingel, K.; Reischl, G.; Widmaier, S.; Rocken, M.; Nutt, R. E.; Machulla, H. J.; Uludag, K.; Cherry, S. R.; Claussen, C. D.; Pichler, B. J., *Nat. Med.* **2008,** *14* (4), 459-65.

(33) Wiseman, J. M.; Ifa, D. R.; Song, Q.; Cooks, R. G., *Angew. Chem. Int. Ed. Engl.* **2006,** *45* (43), 7188-92.

(34) Jackson, S. N.; Ugarov, M.; Egan, T.; Post, J. D.; Langlais, D.; Albert Schultz, J.; Woods, A. S., *J. Mass Spectrom.* **2007,** *42* (8), 1093-8.

(35) Ong, S. E.; Mann, M., *Nat. Chem. Biol.* **2005,** *1* (5), 252-62.

(36) Bantscheff, M.; Schirle, M.; Sweetman, G.; Rick, J.; Kuster, B., *Anal. Bioanal. Chem.* **2007,** *389* (4), 1017-31.
(37) Aebersold, R.; Mann, M., *Nature* **2003,** *422* (6928), 198-207.
(38) Meding, S.; Nitsche, U.; Balluff, B.; Elsner, M.; Rauser, S.; Schone, C.; Nipp, M.; Maak, M.; Feith, M.; Ebert, M. P.; Friess, H.; Langer, R.; Hofler, H.; Zitzelsberger, H.; Rosenberg, R.; Walch, A., *J. Proteome Res.* **2012,** *11* (3), 1996-2003.
(39) Harvey, A. L., *Drug Discov. Today* **2008,** *13* (19-20), 894-901.
(40) Newman, D. J.; Cragg, G. M., *J. Nat. Prod.* **2007,** *70* (3), 461-477.
(41) Rubakhin, S. S.; Jurchen, J. C.; Monroe, E. B.; Sweedler, J. V., *Drug Discov. Today* **2005,** *10* (12), 823-837.
(42) Barth, V. N.; Chernet, E.; Martin, L. J.; Need, A. B.; Rash, K. S.; Morin, M.; Phebus, L. A., *Life Sci.* **2006,** *78* (26), 3007-12.
(43) Takai, N.; Tanaka, Y.; Inazawa, K.; Saji, H., *Rapid Commun. Mass Spectrom.* **2012,** *26* (13), 1549-56.
(44) Wiseman, J. M.; Ifa, D. R.; Zhu, Y.; Kissinger, C. B.; Manicke, N. E.; Kissinger, P. T.; Cooks, R. G., *Proceedings of the National Academy of Sciences* **2008,** *105* (47), 18120-18125.
(45) Alexandrov, T.; Becker, M.; Deininger, S.-O.; Ernst, G.; Wehder, L.; Grasmair, M.; von Eggeling, F.; Thiele, H.; Maass, P., *J. Proteome Res.* **2010,** *9* (12), 6535-6546.
(46) Klerk, L. A.; Broersen, A.; Fletcher, I. W.; van Liere, R.; Heeren, R. M. A., *Int. J. Mass Spectrom.* **2007,** *260* (2-3), 222-236.
(47) Dill, A. L.; Eberlin, L. S.; Costa, A. B.; Zheng, C.; Ifa, D. R.; Cheng, L. A.; Masterson, T. A.; Koch, M. O.; Vitek, O.; Cooks, R. G., *Chem. Eur. J.* **2011,** *17* (10), 2897-2902.
(48) Hanselmann, M.; Kirchner, M.; Renard, B. Y.; Amstalden, E. R.; Glunde, K.; Heeren, R. M. A.; Hamprecht, F. A., *Anal. Chem.* **2008,** *80* (24), 9649-9658.
(49) Deininger, S.-O.; Ebert, M. P.; Fütterer, A.; Gerhard, M.; Röcken, C., *J. Proteome Res.* **2008,** *7* (12), 5230-5236.
(50) McCombie, G.; Staab, D.; Stoeckli, M.; Knochenmuss, R., *Anal. Chem.* **2005,** *77* (19), 6118-24.
(51) Mansfield, J. R.; Sowa, M. G.; Scarth, G. B.; Somorjai, R. L.; Mantsch, H. H., *Anal. Chem.* **1997,** *69* (16), 3370-3374.
(52) Alexandrov, T.; Kobarg, J. H., *Bioinformatics* **2011,** *27* (13), i230-i238.
(53) Zhou, Z. P.; Zare, R. N., *Anal. Chem.* **2017,** *89* (2), 1369-1372.
(54) Trim, P. J.; Djidja, M. C.; Muharib, T.; Cole, L. M.; Flinders, B.; Carolan, V. A.; Francese, S.; Clench, M. R., *J. Proteomics* **2012,** *75* (16), 4931-4940.
(55) Jardin-Mathe, O.; Bonnel, D.; Franck, J.; Wisztorski, M.; Macagno, E.; Fournier, I.; Salzet, M., *J. Proteomics* **2008,** *71* (3), 332-45.
(56) Jaumot, J.; Gargallo, R.; de Juan, A.; Tauler, R., *Chemom. Intell. Lab. Syst.* **2005,** *76* (1), 101-110.
(57) Jaumot, J.; Tauler, R., *Analyst* **2015,** *140* (3), 837-846.

Chapter2. Towards Enhanced Metabolomic Data Analysis of Mass Spectrometry Image: Multivariate Curve Resolution and Machine Learning

This project was a collaborative work that consists of the following authors: Tian, Xiang; Zhang, Genwei; Shao, Yihan; Yang, Zhibo

Genwei Zhang conducted the machine learning analysis in Figure 2-4 and Figure 2-5

2.1. Abstract

Large amounts of data are generally produced from mass spectrometry imaging (MSI) experiments obtaining the molecular and spatial information of biological samples. Traditionally, MS images are constructed using manually selected ions, and it is very challenging to comprehensively analyze MSI results due to their large data sizes and highly complex data structures. To overcome these barriers, it is obligatory to develop advanced data analysis approaches to handle the increasingly large MSI data. In the current study, we focused on the method development of using Multivariate Curve Resolution (MCR) and Machine Learning (ML) approaches. We aimed to effectively extract the essential information present in the large and complex MSI data and enhance the metabolomic data analysis of biological tissues. Multivariate Curve Resolution-Alternating Least Squares (MCR-ALS) algorithm was used to obtain major patterns of spatial distribution and grouped metabolites with the same spatial distribution patterns. In addition, both supervised and unsupervised ML methods were established to analyze the MSI data. In the supervised ML approach, the Random Forest method was selected, and the model was trained using the selected datasets based on the distribution pattern obtained from MCR-ALS analyses. In the unsupervised

ML approach, both DBSCAN (Density-based Spatial Clustering of Applications with Noise) and CLARA (Clustering Large Applications) were applied to cluster the MSI datasets. It is worth noting that similar patterns of spatial distribution were discovered through MSI data analysis using MCR-ALS, supervised ML, and unsupervised ML. Our protocols of data analysis can be applied to process the data acquired using many other types of MSI techniques, and to extract the overall features present in MSI results that are intractable using traditional data analysis approaches.

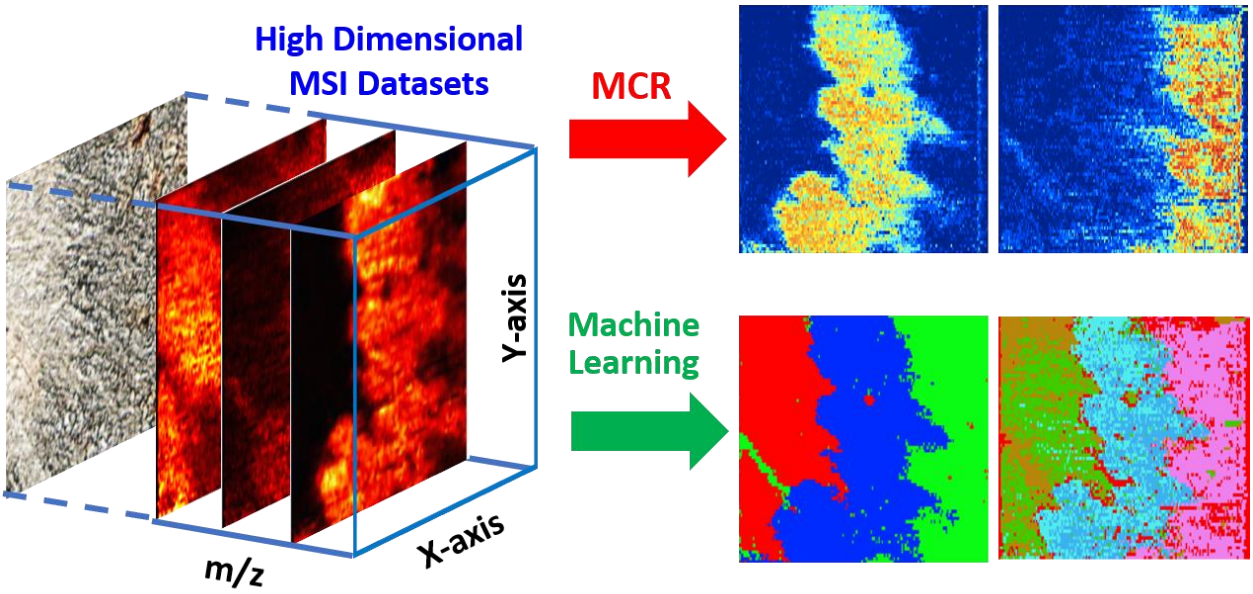


Figure 2-1. Graphic abstract

2.2. Introduction

MS imaging (MSI) is a powerful tool to construct the spatial distribution of wide ranges of molecules in biological tissue samples. Among all MSI methods that have been developed, matrix-assisted laser desorption ionization (MALDI) and secondary ion MS (SIMS) are the most widely used sampling and ionization approaches in non-ambient methods [1]. MALDI has broad detection ranges of molecular weight (>500,000 Da)[2], whereas SIMS provides the highest spatial resolution (<700 nm) [3]. Ambient MSI methods, such as DESI (desorption electrospray ionization) and LAESI

(laser ablation electrospray ionization), require minimum or no sample-preparation, and they can be conveniently used in atmospheric pressure conditions [4]. MSI experiments provide rich chemical information of the biological sample surface, and the experimental datasets generally have very large sizes and complex high-dimensional data structures [5, 6]. For example, depending on the spatial resolution (100−5 μm), a 1 mm^2 MADLI MS image is composed of pixels ranging from 100 to $4x10^4$ [7]. Particularly, each pixel represents a complex chemical profile with the specific spatial information, and the size of a MS image can range from hundreds of megabytes to several gigabytes[8]. Conventional data analyses are typically carried out by focusing on the selected ions among all species detected from samples. However, it is extremely difficult to manually conduct a comprehensive analysis to extract overall features from MSI data through traditional approaches.

To effectively extract essential chemical and spatial information from large and complex MSI data, a number of statistical analysis methods, such as Principal Component Analysis (PCA)[9, 10], clustering[8], and other multivariate analysis[11, 12], have been utilized in previous studies. PCA can be used to determine the major molecular components contributing to the differences of spatial distributions. However, the applications of PCA to MSI data analysis are limited due to same drawbacks: the negative values of PCA score plots have no physical meaning (i.e., the mass intensity cannot be negative), incapability to define the region of interest, and inconsistence with the corresponding loading plot [13, 14]. To overcome these drawbacks, other methods such as probabilistic latent semantic analysis (pLSA) and non-negative parallel factors analysis (NN-PARAFAC), can be applied to achieve non-negative components decomposition; however, the number of components needs to be specified prior to analysis[1, 15]. Unsupervised clustering methods, such as Hierarchical Clustering [16], *k*-means [17], fuzzy c-means [18], and Iterative Self-Organizing Data Analysis Technique (ISODATA) [19], share a common drawback: they cannot yield the spectra

(i.e., molecular information) in each cluster[8]. In contrast, Multivariate Curve Resolution (MCR) algorithms, a family of methods for analyzing mixtures, have been proven as effective approaches to overcome this limitation. MCR techniques resolve mixed datasets by obtaining the number of components, and the signal profile and abundance of each component [11]. The first application of MCR was to investigate two-compound mixtures in UV spectroscopy [20]. Then, a number of MCR algorithms have been developed to analyze complex data. For example, Multivariate Curve Resolution-Alternating Least Squares (MCR-ALS), one of the most popular MCR methods utilizing ALS algorithm, has been used in many research fields [21] such as analyzing data from LC-MS [22], fluorescence [23], and MSI experiments [12]. Among all these applications, MCR-ALS has been successfully used to group major chemical components possessing similar spatial distributions from MSI studies [11]. This method can significantly improve the data analyzing efficiency, and potentially provide more information for a better understanding of biologically relevant metabolites in terms of their spatial distributions in tissues [24, 25]. However, due to relatively intensive computation needed, MCR-ALS is unlikely to be the most efficient approach for the analyses of large sizes of MSI results [11]. For example, MCR needs to be separately performed to analyze individual sets of data obtained from multiples slices of the same tissue (i.e., 3D MSI), which share a large amount of similarities.

Due to the rapid growth of applications of AI (Artificial Intelligence) in a broad range of areas, using ML approaches to analyze MSI data has become an emerging trend [26-29]. ML uses computational approaches to learn from complex instance data, identify patterns and relationships present within the instance data, and achieve predictive data mining [30]. ML has been applied in many branches of biomedical research, including cancer prediction and prognosis [31], genetics and genomics [32], proteomics [33], and medical imaging [34]. However, due to the large size and high

complexity of MSI data as well as the infancy of ML applications, this promising data analysis method has only been utilized in a very few MSI studies [26-29].

In general, ML methods can be divided into two categories: supervised ML and unsupervised ML. Supervised ML predicts features of a dataset after the model is trained using the labeled training data (i.e., datasets of ions with assigned tissue types or regions). Data analysis using supervised ML usually involves three steps: training data selection, model optimization/validation, and prediction of new dataset. Supervised ML has been used in MSI data analysis such as in DESI MSI studies to define the histological subtypes and estimate tumor cell abundances in a gliomas tissue [26]. Most importantly, a trained ML model can be employed to rapidly process 3D MSI data collected from a series of sections of the same tissue, eliminating time-consuming procedures needed in the multivariate analysis of individual MS images. Multiple supervised ML methods, such as support vector machines (SVM)[35] and Random Forest[36], have been developed for data analysis. In our study, the Random Forest algorithm was selected as the supervised ML tool because of its robustness of classification without over-fitting issues[37], high prediction accuracy, and capability of handling a large number of input variables[38]. Compared with supervised ML, unsupervised ML (i.e., unsupervised clustering) methods can be used to directly cluster data into major components without known sample labels in advance, suggesting they are appropriate approaches to analyzing MSI experimental results without labels. Common unsupervised ML methods include *k*-means, PAM (Partitioning Around Medoids), CLARA (Clustering Large Applications), and DBSCAN (Density-based Spatial Clustering of Applications with Noise). k-means is one of the most widely used methods, but it is sensitive to anomalous data points and outliers (i.e., data points that are distant from the mass of data)[39]. Compared with k-means, PAM is a more robust method to process data containing outliers. As an extension of PAM, CLARA has

been developed to effectively analyze big data (e.g., more than several thousands of observations), and it is superior to other methods for its capability of self-optimizing the number of major components [40]. With its superior capability of picking up system outliers, DBSCAN has also become one of the most cited data clustering algorithms [41].

Generally, unsupervised ML approaches are more likely to be suitable to analyze large sizes of MSI data. However, due to the intrinsic high-dimensionality of MSI datasets, which contain both spatial (i.e., x, y-coordinates) and molecular (i.e., m/z values with their intensities) information of each pixel, these methods cannot be directly applied to process the raw MSI data. Instead, the dimensionality reduction of MSI data is needed before unsupervised ML can be conducted. The common dimensionality reduction techniques include the aforementioned PCA, SOM (Self-Organized Maps), and *t*-SNE (*t*-distributed Stochastic Neighbor Embedding). Among these techniques, PCA, as a common tool for linear dimensionality reduction, is widely used to extract key information from complex data, and to visualize them in low-dimensional score plots of top major components. However, the drawbacks of the PCA score plot as mentioned above indicate that it is an inadequate tool for the dimensionality reduction of MSI data. SOM reduces the dimensionality of imaging data by projecting each intensity imaging to a fixed grid (hexagonal or rectangular topology), resulting in a loss of the variants relationship (i.e., spatial information of clusters) [28, 42]. As a nonlinear dimensionality reduction technique, *t*-SNE is particularly suitable for clustering and visualizing highly complex multi-dimensional data [43]. With crucial information in the high-dimensional datasets retained in the low-dimensional datasets upon dimensionality reduction [28, 44], *t*-SNE shows a more desirable dimension-reducing ability.

The potential applications of MCR and ML methods to a variety of areas, such as pharmacometabolomics, biomarker discovery, disease diagnosis, and clinic efficacy monitoring,

have been reported [45-49]. Particularly, MCR-ALS [11] and ML [27, 29] techniques have been recently applied to MSI data analysis. In the current study, we used MCR-ALS and ML (both supervised and unsupervised ML) approaches to carry out enhanced metabolomics analysis of high-spatial resolution (8 ± 2 μm) MSI data of mouse kidney sample. The datasets analyzed in this study were adopted from our previously published work using the Single-probe MSI method [50], which is a versatile technique that has been used in a variety of studies such as MS imaging [50, 51], live single cell analysis [52, 53], and extracellular metabolites measurement in live multicellular spheroids[54]. In the current study, MCR-ALS was first utilized to group the molecular species possessing similar spatial patterns within MS images. Then, to establish an efficient platform for more comprehensive analysis, we performed both supervised and unsupervised ML studies. The Random Forest method was used in the supervised ML approach after MCR-ALS analysis, whereas DBSCAN and CLARA were utilized to cluster datasets upon dimensionality reduction using *t*-SNE. Our application of machine learning methods can be greatly helpful to reveal the hidden information present in large amounts of MSI data, which can potentially benefit biochemical and medical studies.

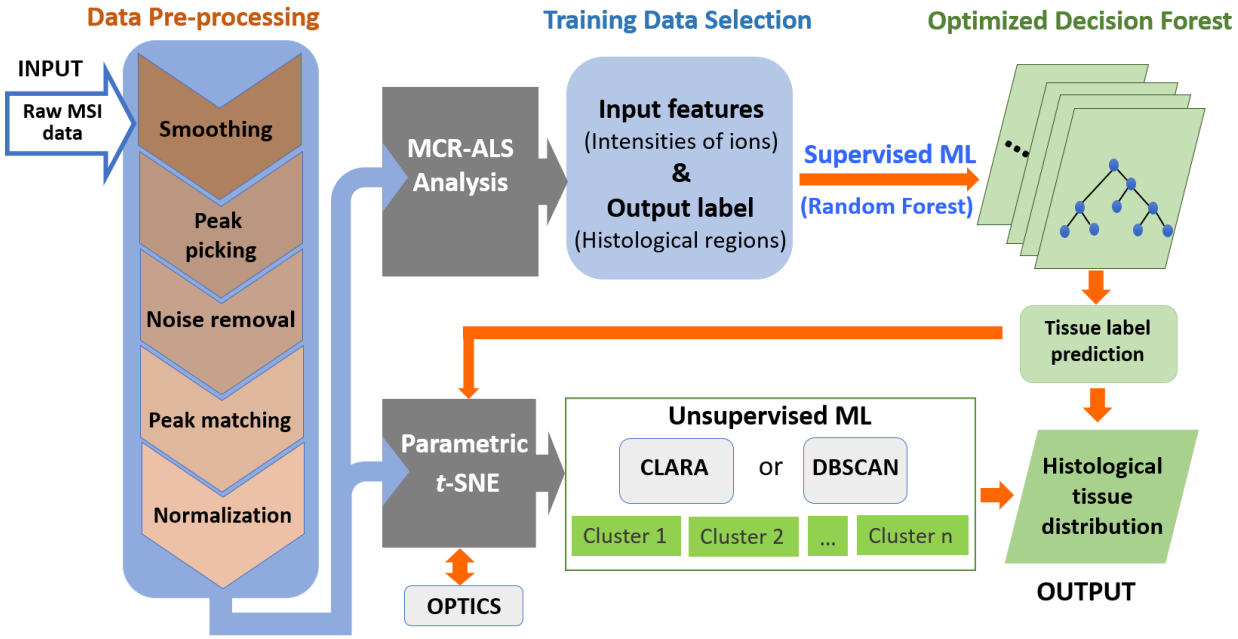
2.3. Data Pre-Processing and Analysis

The experimental data used in the current study were obtained from our previous work[50], which describes detailed experimental protocols of the Single-probe fabrication, tissue sample preparation, and MSI data collection. A brief description of experimental protocols is provided in the Supporting Information.

2.3.1. Flow of MSI Data Analysis

As summarized in Figure 2-2, the complete procedures of our MSI data analysis include four major steps: data pre-processing, MCR-ALS analysis, supervised ML, and unsupervised ML. 1) Data

Pre-processing. We carried out the pre-processing of MSI raw data to obtain datasets that are suitable for the subsequent analyses. 2) MCR-ALS analysis. MCR-ALS algorithm was used to analyze datasets upon accomplishing the pre-processing step. 3) Supervised ML. We labeled partial MCR-ALS results, and used them as the training data (Training Data Selection) for the optimization of the supervised ML model (Random Forest). The trained model was then used to classify the entire histological regions of MSI data (Tissue label prediction). The spatial distribution of clusters was constructed using the classified datasets according to their tissue labels. Supervised ML results were compared with manually assigned tissue labels under the assistance of parametric *t*-SNE (henceforth referred visualization). 4) Unsupervised ML. We used *t*-SNE algorithm to reduce the dimensionality of the high-dimensional datasets obtained from the pre-processing step, and then used unsupervised ML (CLARA and DBSCAN clustering) to analyze



the lower-dimensional datasets. Similarly, the spatial distribution of clustered datasets was constructed according to their tissue labels.

Figure 2-2. Flow chart of MSI data analysis

2.3.1.1. MSI Data pre-processing

In the data pre-processing step, the original MSI data were converted into a format that is suitable for subsequent processing. Technical details of the data pre-processing were provided in the Supporting Information. To import the MSI data into the pre-processing platform in MATLAB, the original MSI data (.raw) was converted to the imzML format using imzMLconverter software [55]. Data pre-processing (Figure. 2-2), including smoothing, peak identification, noise removal, peak alignment, and normalization, was carried out using the built-in functions of MATLAB Bioinformatics Toolbox by following the general procedures described by Robert C. Glen *et al*. [28]. Specifically, Savitzky-Golay filtering was used to perform the smoothing of spectral profiles. Peak picking was conducted using the 'mspeaks' function, and the peak assignment was determined by the sign changes of the first derivatives of the spectral profiles. Noise signal was labeled through median absolute deviation (MAD) estimation and removed by setting the threshold of noise as the signal-to-noise ratio (S/N) less than five (S/N<5). For peak alignment, the 'mspalign' command was applied to the spectra obtained from the previous steps. The default estimation method, histogram (kernel density function), was used to determine the peak locations (i.e., *m/z* values). For the comparison of relative ion intensities among different MS scans, peak intensities were normalized to the total ion current (TIC). Upon accomplishing the data pre-processing, an aligned data matrix (14,100 × 182) was obtained for subsequent MCR-ALS analysis: one dimension of the matrix is the number of aligned *m/z* values (182), and the other is the number of pixels (188 pixel x 75 line =14100).

2.3.1.2. MCR-ALS analysis

To conduct MCR-ALS analysis, Multivariate Curve Resolution Toolbox, developed by Tauler *et al*.[56], running under MATLAB environment was used to analyze the data matrix obtained from the

pre-processing step. To better differentiate low intensity ions in the MCR plots, a logarithmic transformation ($log_2(x)$) of ion intensities in the data matrix (182 x 14100) was conducted prior to the MCR-ALS analysis. The Singular Value Decomposition (SVD) approach was used to estimate the appropriate number of eigenvalues (i.e., components), and each component represents a unique pattern of the spatial distribution of a group of ions. Ions with similar spatial distribution were assigned to the same group. Five major components (explained variance of 97.7%) were found to be an optimal number of coexisted patterns. The spatial distribution of each component was exported from the MCR-ALS through a user-friendly interface, and the corresponding ions were shown as mass spectra using R language (details are provided in the SI).

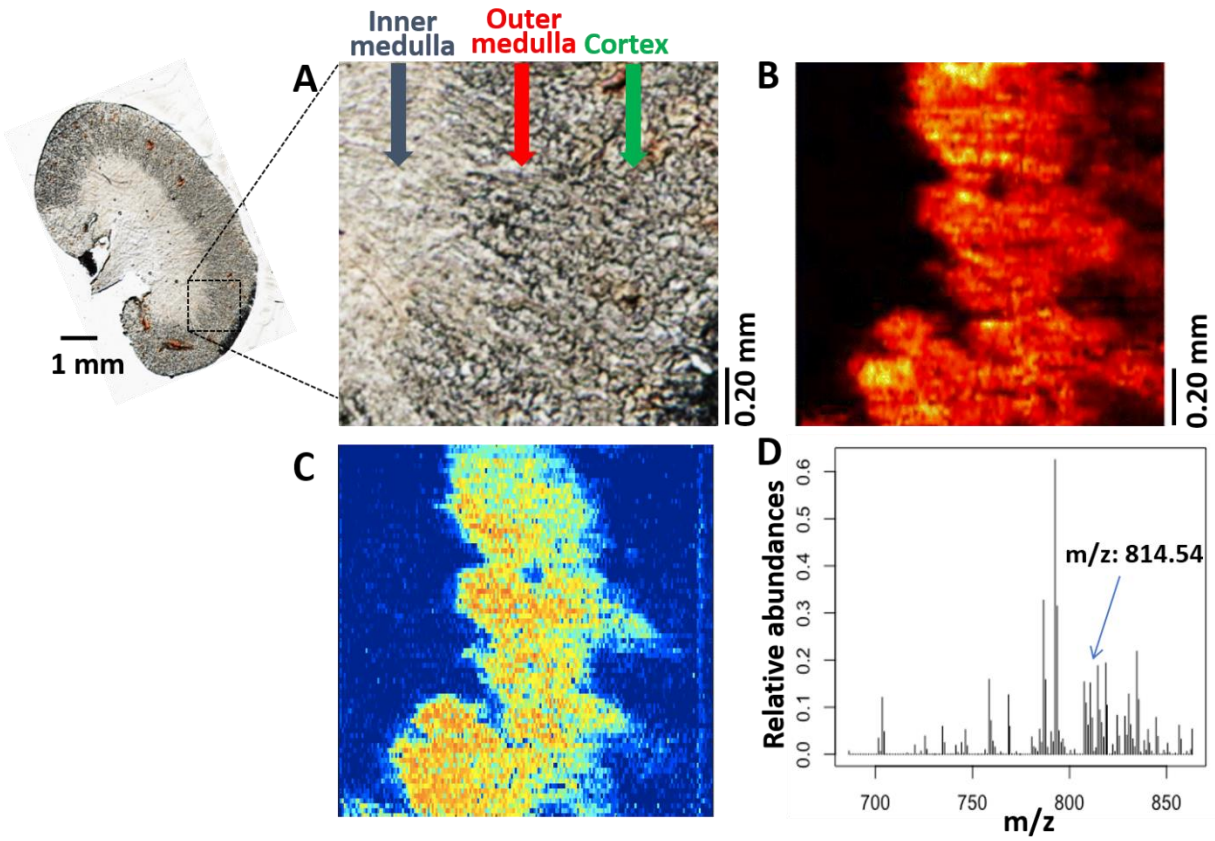


Figure 2-3. MSI data analysis using selected ion or MCR methods

(A) Optical image of a mouse kidney illustrating three histological regions: inner medulla, outer medulla, and cortex. (B) An example of MS image constructed using the selected ion ([PC (38:5) + Na]$^+$; *m/z* 814.5726). (C) An example of spatial distribution pattern obtained from MCR-ALS analysis. (D) The mass spectrum of the grouped molecules possessing the same pattern of spatial

distribution shown in C. Note: Figs. A and B are adapted from "High Resolution Tissue Imaging Using the Single-probe Mass Spectrometry under Ambient Conditions," by Wei Rao, Ning Pan, and Zhibo Yang, 2015, *Journal of the American Society of Mass Spectrometry*, 26, 986-993 [50]. Copyright 2015 by *The American Society of Mass Spectrometry*.

2.3.1.3. Supervised Machine Learning (ML): Random Forest

Training data and testing data (Training Data Selection; Table S2-1) were manually selected from three classified histological regions of mouse kidney (Figure 2-4A) determined by MCR-ALS analysis to provide input features (intensities of ions) and output labels (inner medulla, outer medulla, or cortex). The Random Forest approach provided in R language was trained using the selected training data. To validate this supervised ML model, testing data were used to test the accuracy of prediction. Three trails have been executed, and the average prediction accuracy is > 99% (Table S2-2). Then, the optimized model (Optimized Decision Forest) was used to process the rest of the datasets for tissue label prediction, i.e., to classify ions in the MSI data into each of three regions. To provide clear physical meanings of the overall results obtained from the supervised ML approach, the predicted tissue labels (Figure 2-4B) were transformed into histological tissue distribution using the R language.

The supervised ML results were verified through *t*-SNE visualization. *t*-SNE has become a popular method to transform the high-dimensional structures of data into two- or three-dimensional formats that can be conveniently visualized. In our studies, we used *t*-SNE, which is provided as a built-in function in Statistics and Machine Learning Toolbox of MATLAB, to visualize the results from both manually assigned tissues labels (Figure 2-4A) and ML-generated (Figure 2-4B) upon finishing the dimensionality reduction (Figure 2-4C and 2-4D). As a validation of *t*-SNE results, the OPTICS (Ordering Points To Identify the Clustering Structure) approach was also used to

index the data points in order to identify the clustering structure (Figure S2-2B). The optimized parameters of *t*-SNE and OPTICS were detailed in the Supporting Information.

2.3.1.4 Unsupervised machine learning (ML): CLARA and DBSCAN

We performed CLARA analysis using R, provided within package 'cluster', with six optimal clusters and the sample size equals to 50. CLARA results were visualized using the R language (Figure 2-5A). DBSCAN, another popular unsupervised ML method with an excellent capability of picking up system outliers, was used to perform a comparison with CLARA. Similarly, DBSCAN results were visualized using the R. Finally, to interpret the physical meanings of all the clustered data from CLARA and DBSCAN analyses, the classification maps of mouse kidney tissue were constructed using the R (all packages can be found at https://cran.r-project.org/web/packages). As a comparison of *t*-SNE results, PCA was also used to reduce the dimensionality of the MSI data; however, the boundaries among clusters (Figure S2-7) are not as clear as those from *t*-SNE results (Figure 2-4C), indicating that PCA is inadequate to visualize high dimensional MSI dataset.

2.4. Results and Discussion

2.4.1. MCR-ALS approach grouped molecules with similar spatial distribution pattern from MS images.

Figure 2-3A and 2-3B, which are adopted from our previous publication [50], illustrate the optical image of kidney slice (with the analyzed area in MSI experiments) and the MS image of a selected ion ($[PC(38:5) + Na]^+$, m/z 814.5726), respectively. Three different histological regions (i.e., inner medulla, outer medulla, and cortex) can be generally observed from the optical image. Using the high-spatial resolution (8.5 μm) of our Single-probe MSI techniques, 14100 pixels (equivalent to

14,100 mass spectra) were obtained from the MS image with an area of ~1.0 x 1.0 mm^2 (Figure 2-3A). Data analysis was carried out using a Dell Precision T5500 work station (processor: dual Intel (R) Xeon(R) CPU X5650 2.66 GHz; memory (RAM): 72.0 GB). 182 common ions were generated from pre-treatment for the subsequent analysis as outlined in Method 2.1.2. A complete MCR analysis of the pretreated MSI data took about 20 minutes to accomplish. MCR-ALS approach was used to determine the number of major components present in the data matrix. Using the Singular Value Decomposition (SVD) method (Figure S2-5), we concluded that five eigenvalues (i.e., components) were sufficient to represent the majority information (explained variance of 97.7%, Figure S2-6 and Table S2-7) of the entire MSI dataset. Species in each component (i.e., m/z values with their corresponding relative intensities) were extracted from the data matrix. For example, a component obtained from MCR-ALS analysis (Figure 2-3C) exhibits very similar spatial features as a MS image constructed using a selected ion (m/z 814.5726) within outer medulla (Figure 2-3B) [50]. In fact, this selected ion is among many others with very similar spatial features as summarized in Figure 2-3D. It is very likely that compounds present similar spatial distribution have correlated metabolomic functions. For example, previous studies indicate that PC (40:6) and PC (36:4) have high abundances in the kidney outer medulla, and they both are upregulated in response to cisplatin treatment by shaping membrane-protein function[57, 58]. The MCR-ALS classification can potentially benefit the future discovery of the metabolomic pathways among species possessing similar patterns of spatial distribution. We summarized the top 15 most abundant ions grouped in the component 1 in Table 2-1, and included *m/z* values from MCR analysis, experimental *m/z* values, and tentatively assigned metabolites acquired from METLIN (https://metlin.scripps.edu/landing_page.php?pgcontent=mainPage).[59] In addition, the loading scores were provided to illustrate the relative contribution of each metabolite to component 1.

Using similar approaches, the other four major components representing the spatial and molecular features of the MS images for the inner medulla and cortex regions were provided in the Supporting Information (Figure S2-1 and Table S3-S6).

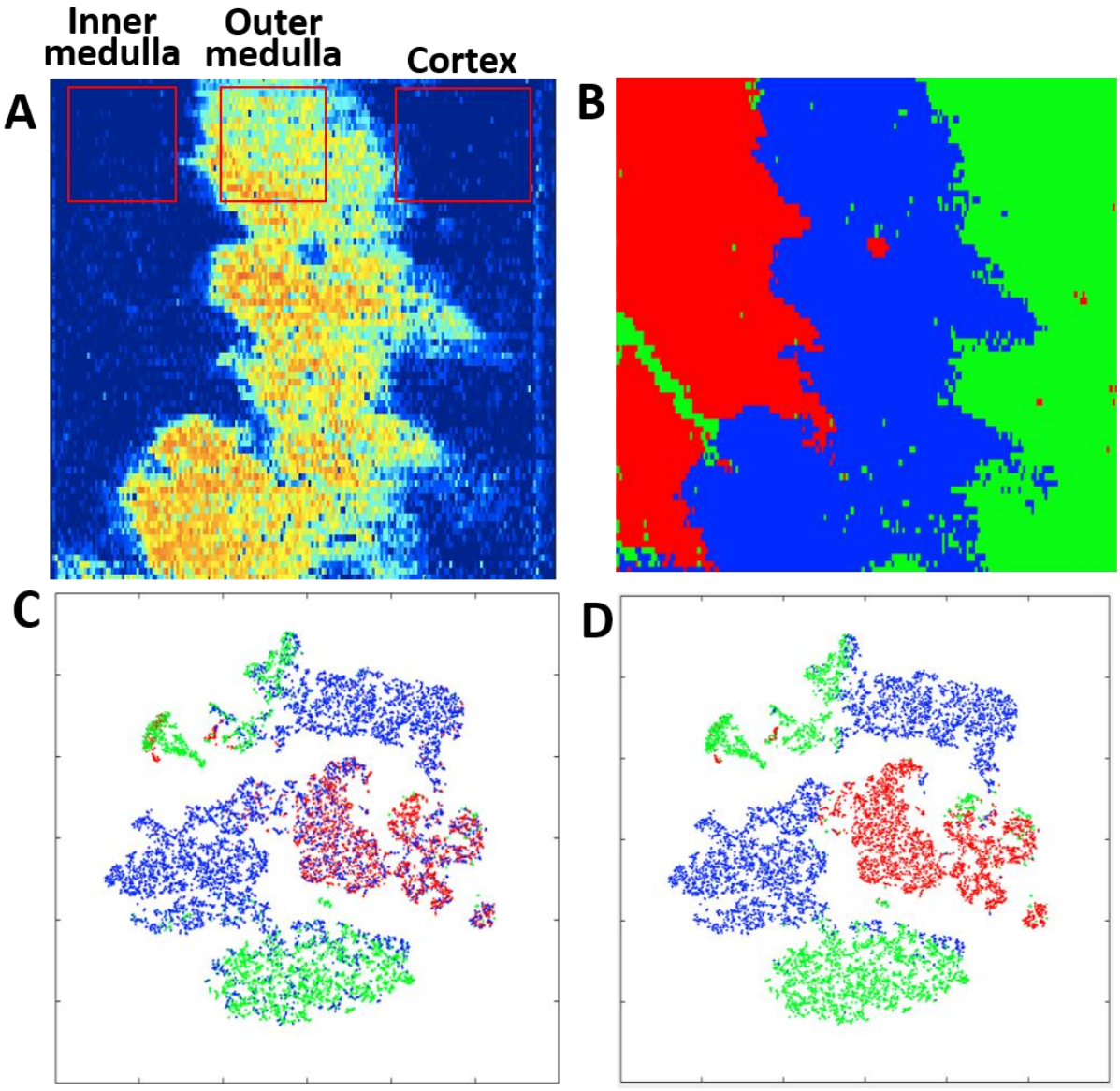


Figure 2-4. MSI data analysis using supervised ML method.

(A) Training data for supervised ML were manually selected from three regions (inner medulla, outer medulla, and cortex) based on the MCR-ALS results. (B) Histological tissue distribution constructed using supervised ML (Random Forest) results. (C) The clustering capability of MCR-

ALS method was evaluated using *t*-SNE. (D) The clustering capability of the supervised ML method was evaluated using *t*-SNE.

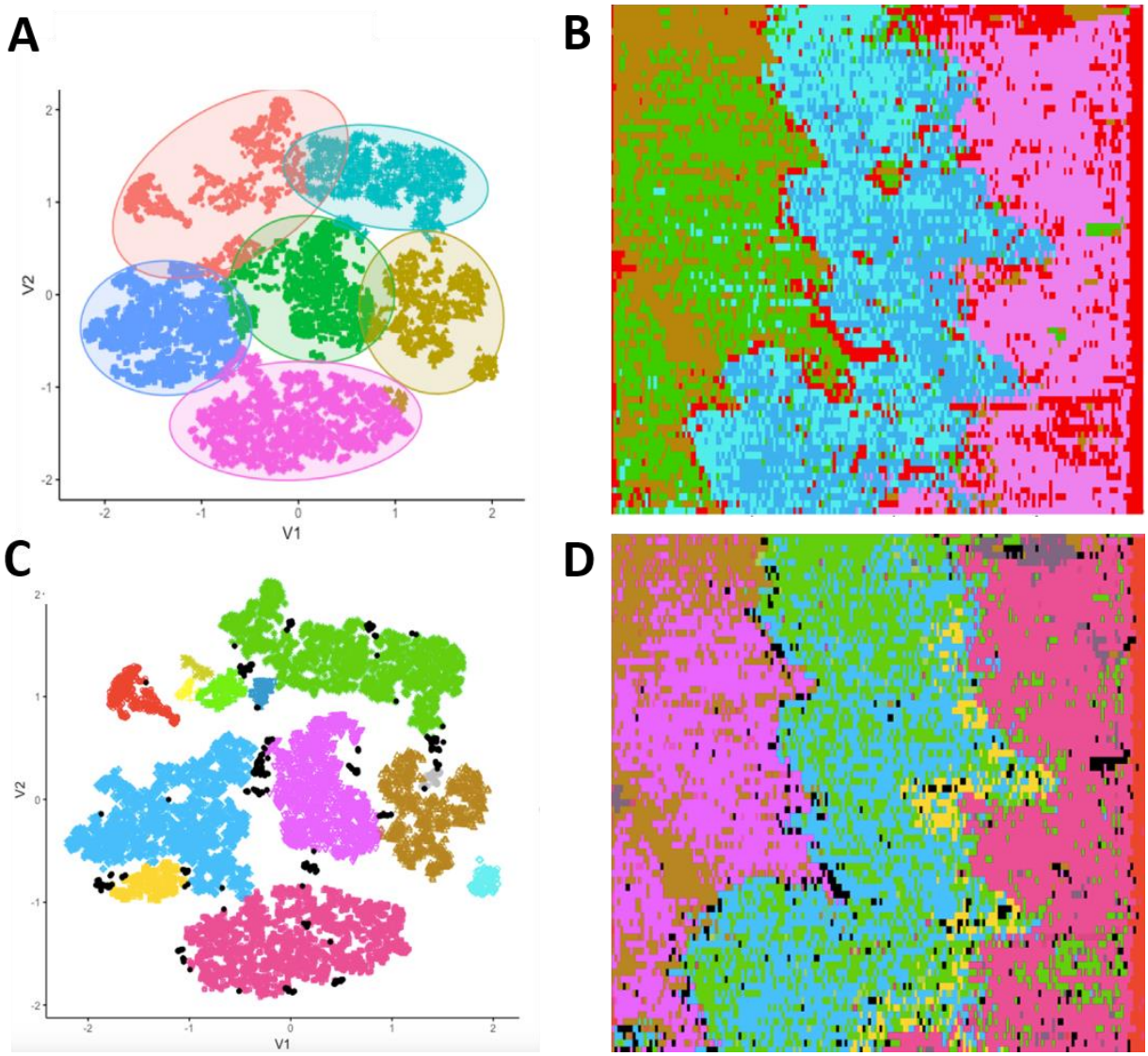


Figure 2-5. MSI data analysis using unsupervised ML methods.

(A) CLARA was used to cluster the MSI data into six major groups upon optimization. (B) Histological tissue distribution constructed using CLARA results. (C) DBSCAN was applied to cluster the MSI data into seven major groups. (D) Histological tissue distribution constructed using DBSCAN results.

2.4.2. Supervised ML utilized MCR-ALS results and improved the discovery of subtle features from MSI data.

Since the quality of training data can significantly affect the ML prediction capability, it is crucial to carefully select the appropriate training datasets representing characteristics of the overall data. We selected the datasets according to results obtained from the MCR-ALS analysis, and divided them into training and testing datasets (Table S2-1), which are grouped *m/z* values of species present on each of those three regions on the tissue slice (i.e., inner medulla, outer medulla, and cortex) (Figure 2-4A). A Random Forest classification model was trained using the training datasets. We evaluated the accuracy of this trained model using the testing datasets and achieved a high predictive accuracy (> 99%; details are provided in Table S2-2). This trained Random Forest model was then applied to the rest of the MSI datasets to predict their tissue labels, i.e., to classify the detected molecules to each of those three physiological regions on tissue slice. To interpret the physical meanings of the supervised ML results, classified data were used to generate the tissue classification map, in which three different colors represent three physiological regions on tissue (Figure 2-4B). Because the training datasets were selected based on MCR-ALS results, the supervised ML generally reproduced the overall features of three histological regions. Particularly, the spatial distribution of molecules in the inner medulla exhibited similar features that can be observed in a spatial pattern obtained from MCR-ALS analysis (Figure 2-4A). Moreover, our supervised ML analysis enhanced some subtle features in MS images, and generated clearer boundaries between different regions.

To verify that the supervised ML method provides more capability of classification than MCR-ALS approach, *t*-SNE has been employed to process the MCR-ALS results. Since *t*-SNE requires tissue labels to be known prior to plotting the *t*-SNE clustering results, three types of tissue labels

(i.e., inner medulla, outer medulla, and cortex) were assigned to the 14,100 scans based on the MCR-ALS results (Figure 2-4A). Using the *t*-SNE to cluster these labels, overlapped color dots were observed between the outer medulla and inner medulla or cortex (Figure 2-4C), indicating unclear assignments of boundary labels can arise from manually assigned tissue labels solely based on the MCR-ALS results. In contrast, for tissues labels predicted from the supervised ML analysis, an improved *t*-SNE cluster map was generated (Figure 2-4D), implying less biased determination of boundary pixels was achieved using supervised ML rather than human intuition. More importantly, the established ML models can be directly used for efficient analyses of many other MS images obtained from the same tissues such as slices used in 3D MSI studies of a given tissue.

Table 2-1. Top 15 most abundant ions grouped in component 1 (Fig. 2-3D)

*PC: Phosphatidylcholines, PE: Phosphatidylethanolamine, SM: Sphingomyelin, PG: Phosphatidylglycerol

MCR (m/z)	Experiment(m/z)	Tentative assignment*	Exact (m/z)	ppm	Loading Score
792.5428	792.5911	PC(18:1/20:4)	792.5902	1	0.6251
786.6442	786.6018	PC(18:1/18:1)	786.6007	1	0.3268
793.6425	793.5951	PG(22:0/15:0)	793.5953	0	0.3146
834.6404	834.6015	PC(20:4/20:2)	834.6007	0	0.2185
818.6404	818.6063	PC(18:1/22:5)	818.6058	0	0.1934
814.5414	814.5729	PC(18:1/20:4)	814.5721	0	0.1878
758.5451	758.5705	PE(22:2/15:0)	758.5694	1	0.1588
810.6423	810.6019	PC(20:3/18:1)	810.6007	1	0.1511
830.5394	830.5464	PC(22:5/16:0)	830.5460	0	0.1276
768.5447	768.5550	PE(22:4/16:0)	798.5538	1	0.1262
703.5506	703.5757	SM(18:1/16:0)	703.5749	1	0.1206
808.5428	808.5864	PC(20:4/18:1)	808.5851	1	0.1086
819.3422	819.6100	PG(21:0/18:1)	819.6110	1	0.1037
815.5661	815.5763	PG(18:0/19:0)	815.5773	1	0.0939
824.5409	824.5578	PC(18:2/18:0)	824.5566	1	0.0824

2.4.3. Unsupervised ML extracted more molecular and spatial information from MSI data.

Supervised ML requires training data to optimize the model prior to any applications. However, additional experiments, such as MCR-ALS or H&E staining, are needed to provide histological

information for selecting the training data. In contrast, such training data selection is not required for unsupervised ML methods. To generate datasets with suitable size and structure, *t*-SNE algorithm was utilized to reduce the high-dimensional MSI data. Notably, directly applying unsupervised ML (without *t*-SNE processing) to analyze our original MSI dataset cannot produce any optimal clusters through parameter-changing attempts (Figure S2-4), indicating that using *t*-SNE for dimensionality reduction is a key step to obtain the classification of high-dimensional MSI dataset. To support the effectiveness of *t*-SNE for dimensionality reduction, OPTICS algorithms were used to generate the point reachability plot of the lower-dimensional datasets (Figure S2-3B), in which the reachability indicates the extent of separation in the *t*-SNE plot (Figure 2-4) (details of OPTICS are in the Figure S2-2B)

Two unsupervised ML approaches, CLARA and DBSCAN, were chosen to perform the classification. CLARA is suitable to deal with big dataset and capable of optimizing the number of clusters, whereas DBSCAN has the advantage of picking up system outliers. The results of CLARA and DBSCAN were shown in Figure 2-5A and 2-5C, respectively. The physical meanings of the unsupervised ML-generated data clusters were further investigated through reconstructing their spatial distributions (Figure 2-5B and 2-5D). Interestingly, without given information of tissue distribution patterns, the CLARA approach generated an optimal number (six) of clusters possessing spatial distributions that are relevant to the histological regions shown on the optical image (Figure 2-3A). These six optimal clusters coexist in our dataset, and two sub-regions were discovered in both the inner and outer medulla regions (Figure 2-5A and 2-5B). DBSCAN analysis also discovered similar cluster patterns as well as several minor sub-regions that are comparable to CLARA results. In addition, another advantage of using DBSCAN was to identify the system outliers, labeled as the black dots in Figure 2-5C and 2-5D, that were absent from the CLARA

analysis. Noteworthy, clusters assigned with red color from both CLARA and DBSCAN results are basically classified as margins of the tissue slice (Figure 2-5).

The reconstructed histological tissue distributions were obtained from both supervised (Figure 3B) and unsupervised (Figure 2-5B and 2-5D) ML approaches, and similar features of spatial distributions were obtained. In supervised ML method, a portion of MCR-ALS results was used as the training data to optimize the model, which was further used to reproduce all three clusters. In unsupervised ML methods, including CLARA and DBSCAN, the *t*-SNE results were used to perform the tissue classification (i.e., to allocate the spatial locations of the detected molecule). Encouragingly, two different approaches led to similar results, indicating that the clusters we identified and the analysis protocols in use were reliable.

Although both supervised and unsupervised ML methods have been successfully applied to our MSI data analyses, they have their own inherent advantages and disadvantages. Previous studies indicate that supervised ML is a suitable tool to identify different tissue features or potentially distinguish pathological cells (e.g., cancer cells) from normal cells in MSI data analysis [60]. Although supervised ML algorithm shows adaptability for rapid prediction of unknown MS images, careful selection of training data, which requires additional information, is critical for the optimization of models. This type of training process is obligatory before supervised ML can be applied for any MSI data analysis. In contrast, unsupervised ML can be conveniently utilized to analyze MSI data without model training. However, the results need to be validated by comparing with those obtained from other labeling studies or techniques such as tissue staining [28, 61]. Nevertheless, it is still possible that some subtle physical or chemical features can be overlooked, or experiments cannot be performed due to limitations of available techniques or samples. Therefore, to significantly increase the data-analyzing efficiency while minimizing uncertainties

during the MSI data analyses, we suggest using a combined method including both supervised and unsupervised ML approaches.

2.5. Conclusion

As an emerging molecular imaging technique used for biological tissue analysis in fundamental research and biomedical applications, MSI experiments usually generate huge amounts of data. Traditionally, MSI data analysis is generally carried out for manually selected ions. Due to the large size and high-dimensionality of MSI datasets, it has been very challenging to conduct comprehensively data analysis to extract overall features representing essential molecular and spatial information present in biological tissues[14]. Therefore, advanced data analysis methods are needed to perform a more efficient analysis of large sizes of MSI data [62].

MCR-ALS is a multivariate analysis method that can decompose the complex MSI dataset into major components with spatial distribution patterns and grouped ions [11, 12]. The application of MCR-ALS can enhance MSI data analysis without tissue-histological knowledge [63]. To further increase the efficiency of data mining from larger sizes of MSI results, machine learning (ML) methods are likely to be more efficient approaches. In the current study, both supervised and unsupervised ML methods have been utilized to distinguish histological regions from the high-spatial resolution MS images of mouse kidney slice. Two programming languages, R and MATLAB, were cooperatively used to implement the methods for MSI data processing. Regardless of the type of ML approach used in the analysis of MSI data, reducing the dimensionality of target datasets was an obligatory step. As a prevailing dimensionality reduction tool, *t*-SNE was employed in both supervised and unsupervised ML studies. In supervised ML (i.e., Random Forest) studies, the histological regions determined by MCR-ALS analysis were used as a guide to select the defined training datasets, and the *t*-SNE algorithm was utilized to reduce the

high-dimensional datasets allowing us to visualize the results obtained from Random Forest analyses. In unsupervised ML studies (CLARA and DBSCAN), *t*-SNE has been proven as an effective approach to process the pre-treated MSI datasets and ensure them to be suitable for subsequent unsupervised ML processes. Both supervised and unsupervised ML approaches are effective for MSI data analysis. However, combined supervised and unsupervised ML studies are likely to be more effective to extract the overall chemical and spatial features from complex MSI data with minimum overlooked information. Our studies indicate that advanced data analysis methods, including MCR-ALS and ML approaches, are efficient tools for comprehensive analysis of large amounts of high-spatial resolution MS images obtained using our Single-probe MSI techniques. These emerging methods can be broadly utilized for many other MSI studies conducted using different techniques, and to promote the growth of data-analyzing tools needed for big data science.

The material in chapter 2 is adapted from Tian, X.; Zhang, G.; Shao, Y.; Yang, Z., *Anal. Chim. Acta* **2018,** *1037*, 211-219. Copyright permission is obtained from Elsevier.

References

(1) Amstalden van Hove, E. R.; Smith, D. F.; Heeren, R. M., *J. Chromatogr. A* **2010,** *1217* (25), 3946-54.

(2) Cohen, A., *Appl. Spectrosc. Rev.* **2009,** *44* (4), 362-362.

(3) Brison, J.; Robinson, M. A.; Benoit, D. S. W.; Muramoto, S.; Stayton, P. S.; Castner, D. G., *Anal. Chem.* **2013,** *85* (22), 10869-10877.

(4) Ifa, D. R.; Wiseman, J. M.; Song, Q. Y.; Cooks, R. G., *Int. J. Mass spectrom.* **2007,** *259* (1-3), 8-15.

(5) McDonnell, L. A.; Heeren, R. M., *Mass Spectrom. Rev.* **2007,** *26* (4), 606-43.

(6) Wu, C.; Dill, A. L.; Eberlin, L. S.; Cooks, R. G.; Ifa, D. R., *Mass Spectrom. Rev.* **2013,** *32* (3), 218-43.

(7) Spraggins, J. M.; Caprioli, R., *J. Am. Soc. Mass Spectrom.* **2011,** *22* (6), 1022-1031.

(8) McCombie, G.; Staab, D.; Stoeckli, M.; Knochenmuss, R., *Anal. Chem.* **2005,** *77* (19), 6118-24.

(9) Yao, I.; Sugiura, Y.; Matsumoto, M.; Setou, M., *Proteomics* **2008,** *8* (18), 3692-701.

(10) Dill, A. L.; Eberlin, L. S.; Costa, A. B.; Zheng, C.; Ifa, D. R.; Cheng, L. A.; Masterson, T. A.; Koch, M. O.; Vitek, O.; Cooks, R. G., *Chem. Eur. J.* **2011,** *17* (10), 2897-2902.

(11) Jaumot, J.; Tauler, R., *Analyst* **2015,** *140* (3), 837-846.

(12) Rao, W.; Scurr, D. J.; Burston, J.; Alexander, M. R.; Barrett, D. A., *Analyst* **2012,** *137* (17), 3946-3953.

(13) Alexandrov, T.; Becker, M.; Deininger, S. O.; Ernst, G.; Wehder, L.; Grasmair, M.; von Eggeling, F.; Thiele, H.; Maass, P., *J. Proteome Res.* **2010,** *9* (12), 6535-6546.

(14) Alexandrov, T., *BMC Bioinformatics* **2012,** *13 Suppl 16*, S11.

(15) Hanselmann, M.; Kirchner, M.; Renard, B. Y.; Amstalden, E. R.; Glunde, K.; Heeren, R. M.; Hamprecht, F. A., *Anal. Chem.* **2008,** *80* (24), 9649-58.

(16) Deininger, S. O.; Ebert, M. P.; Futterer, A.; Gerhard, M.; Rocken, C., *J. Proteome Res.* **2008,** *7* (12), 5230-6.

(17) Trede, D.; Schiffler, S.; Becker, M.; Wirtz, S.; Steinhorst, K.; Strehlow, J.; Aichler, M.; Kobarg, J. H.; Oetjen, J.; Dyatloy, A.; Heldmann, S.; Walch, A.; Thiele, H.; Maass, P.; Alexandrov, T., *Anal. Chem.* **2012,** *84* (14), 6079-6087.

(18) Chuang, K. S.; Tzeng, H. L.; Chen, S.; Wu, J.; Chen, T. J., *Comput. Med. Imaging Graph.* **2006,** *30* (1), 9-15.

(19) Klinkert, I.; Chughtai, K.; Ellis, S. R.; Heeren, R. M. A., *Int. J. Mass spectrom.* **2014,** *362*, 40-47.

(20) Ruckebusch, C.; Blanchet, L., *Anal. Chim. Acta* **2013,** *765*, 28-36.

(21) Jaumot, J.; de Juan, A.; Tauler, R., *Chemom. Intell. Lab. Syst.* **2015,** *140*, 1-12.

(22) Perez, I. S.; Culzoni, M. J.; Siano, G. G.; Garcia, M. D. G.; Goicoechea, H. C.; Galera, M. M., *Anal. Chem.* **2009,** *81* (20), 8335-8346.

(23) Neves, A. C.; Tauler, R.; de Lima, K. M., *Anal. Chim. Acta* **2016,** *937*, 21-8.

(24) Bedia, C.; Tauler, R.; Jaumot, J., *Talanta* **2017,** *175*, 557-565.

(25) Bean, H. D.; Zhu, J. J.; Hill, J. E., *Jove-Journal of Visualized Experiments* **2011,** (52).

(26) Eberlin, L. S.; Norton, I.; Dill, A. L.; Golby, A. J.; Ligon, K. L.; Santagata, S.; Cooks, R. G.; Agar, N. Y., *Cancer Res.* **2012,** *72* (3), 645-54.

(27) Galli, M.; Zoppis, I.; Smith, A.; Magni, F.; Mauri, G., *Expert Rev Proteomic* **2016,** *13* (7), 685-696.

(28) Inglese, P.; McKenzie, J. S.; Mroz, A.; Kinross, J.; Veselkov, K.; Holmes, E.; Takats, Z.; Nicholson, J. K.; Glen, R. C., *Chem. Sci.* **2017,** *8* (5), 3500-3511.

(29) Zhou, Z. P.; Zare, R. N., *Anal. Chem.* **2017,** *89* (2), 1369-1372.

(30) Kotsiantis, S. B.; Zaharakis, I. D.; Pintelas, P. E., *Artif Intell Rev* **2006,** *26* (3), 159-190.

(31) Kourou, K.; Exarchos, T. P.; Exarchos, K. P.; Karamouzis, M. V.; Fotiadis, D. I., *Comput. Struct. Biotechnol. J.* **2015,** *13*, 8-17.

(32) Libbrecht, M. W.; Noble, W. S., *Nat. Rev. Genet.* **2015,** *16* (6), 321-332.

(33) Swan, A. L.; Mobasheri, A.; Allaway, D.; Liddell, S.; Bacardit, J., *OMICS* **2013,** *17* (12), 595-610.
(34) Wernick, M. N.; Yang, Y. Y.; Brankov, J. G.; Yourganov, G.; Strother, S. C., *IEEE Signal Process Mag* **2010,** *27* (4), 25-38.
(35) Schölkopf, B.; Smola, A. J., *Learning with kernels : support vector machines, regularization, optimization, and beyond*. MIT Press: Cambridge, Mass., 2002; p xviii, 626 p.
(36) Breiman, L., *Machine Learning* **2001,** *45* (1), 5-32.
(37) Svetnik, V.; Liaw, A.; Tong, C.; Culberson, J. C.; Sheridan, R. P.; Feuston, B. P., *J. Chem. Inf. Comput. Sci.* **2003,** *43* (6), 1947-58.
(38) Hanselmann, M.; Kothe, U.; Kirchner, M.; Renard, B. Y.; Amstalden, E. R.; Glunde, K.; Heeren, R. M. A.; Hamprecht, F. A., *J. Proteome Res.* **2009,** *8* (7), 3558-3567.
(39) Ertoz, L.; Steinbach, M.; Kumar, V., *Proceedings of the Third Siam International Conference on Data Mining* **2003**, 47-58.
(40) Nagpal, A.; Jatain, A.; Gaur, D., *2013 Ieee Conference on Information and Communication Technologies (Ict 2013)* **2013**, 298-303.
(41) Ester, M.; Kriegel, H.-P.; Sander, J.; Xu, X. In *A density-based algorithm for discovering clusters in large spatial databases with noise*, Kdd, 1996; pp 226-231.
(42) Franceschi, P.; Wehrens, R., *Proteomics* **2014,** *14* (7-8), 853-61.
(43) van der Maaten, L.; Hinton, G., *Journal of Machine Learning Research* **2008,** *9*, 2579-2605.
(44) Abdelmoula, W. M.; Balluff, B.; Englert, S.; Dijkstra, J.; Reinders, M. J. T.; Walch, A.; McDonnell, L. A.; Lelieveldt, B. P. F., *Proc. Natl. Acad. Sci. U. S. A.* **2016,** *113* (43), 12244-12249.
(45) Connor, J. P.; Symons, M.; Feeney, G. F.; Young, R. M.; Wiles, J., *Subst. Use Misuse* **2007,** *42* (14), 2193-206.
(46) Kim, Y. K.; Na, K. S., *Prog. Neuropsychopharmacol. Biol. Psychiatry* **2017**.
(47) Thottakkara, P.; Ozrazgat-Baslanti, T.; Hupf, B. B.; Rashidi, P.; Pardalos, P.; Momcilovic, P.; Bihorac, A., *PloS one* **2016,** *11* (5), e0155705.
(48) Zhang, Z., *Ann Transl Med* **2016,** *4* (6), 125.
(49) Lin, G.; Chung, Y. L., *Biomed Res Int* **2014,** *2014*, 625095.
(50) Rao, W.; Pan, N.; Yang, Z. B., *J. Am. Soc. Mass Spectrom.* **2015,** *26* (6), 986-993.
(51) Rao, W.; Pan, N.; Tian, X.; Yang, Z., *J. Am. Soc. Mass Spectrom.* **2016,** *27* (1), 124-34.
(52) Pan, N.; Rao, W.; Kothapalli, N. R.; Liu, R.; Burgett, A. W.; Yang, Z., *Anal. Chem.* **2014,** *86* (19), 9376-80.
(53) Pan, N.; Rao, W.; Standke, S. J.; Yang, Z., *Anal. Chem.* **2016,** *88* (13), 6812-9.
(54) Sun, M.; Tian, X.; Yang, Z., *Anal. Chem.* **2017,** *89* (17), 9069-9076.
(55) Race, A. M.; Styles, I. B.; Bunch, J., *J. Proteomics* **2012,** *75* (16), 5111-5112.
(56) Jaumot, J.; Gargallo, R.; de Juan, A.; Tauler, R., *Chemom. Intell. Lab. Syst.* **2005,** *76* (1), 101-110.
(57) Moreno-Gordaliza, E.; Esteban-Fernandez, D.; Lazaro, A.; Humanes, B.; Aboulmagd, S.; Tejedor, A.; Linscheid, M. W.; Gomez-Gomez, M. M., *Talanta* **2017,** *164*, 16-26.
(58) Phillips, R.; Ursell, T.; Wiggins, P.; Sens, P., *Nature* **2009,** *459* (7245), 379-385.
(59) Smith, C. A.; O'Maille, G.; Want, E. J.; Qin, C.; Trauger, S. A.; Brandon, T. R.; Custodio, D. E.; Abagyan, R.; Siuzdak, G., *Ther. Drug Monit.* **2005,** *27* (6), 747-751.
(60) Heylman, C.; Datta, R.; Sobrino, A.; George, S.; Gratton, E., *PloS one* **2015,** *10* (12), e0144572.
(61) Fuchs, T. J.; Buhmann, J. M., *Comput. Med. Imaging Graph.* **2011,** *35* (7-8), 515-530.
(62) Jones, E. A.; Deininger, S. O.; Hogendoorn, P. C.; Deelder, A. M.; McDonnell, L. A., *J. Proteomics* **2012,** *75* (16), 4962-89.
(63) Duponchel, L.; Elmi-Rayaleh, W.; Ruckebusch, C.; Huvenne, J. P., *J. Chem. Inf. Comput. Sci.* **2003,** *43* (6), 2057-2067.

Chapter3. Anticancer Drug Affects Metabolomic Profiles in Multicellular Spheroids: Studies Using Mass Spectrometry Imaging Combined with Machine Learning

This project was a collaborative work that consists of the following authors: Tian, Xiang; Zhang, Genwei; Zou, Zhu and Yang, Zhibo.

Genwei Zhang conducted some data analysis in Figure 3-6

Zhu Zou conducted the contour map in Figure S3-2 and determined the morphology change in spheroids in Figure S3-3

3.1. Abstract

Multicellular spheroids (hereinafter referred to as spheroids) are 3D biological models. The metabolomic profiles inside spheroids provide crucial information reflecting the molecular phenotypes and microenvironment of cells. To study the influence of anticancer drugs on the spatially resolved metabolites, spheroids were cultured using HCT-116 colorectal cancer cells, treated with anticancer drug Irinotecan under a series of time- and concentration-dependent conditions. The Single-probe mass spectrometry imaging (MSI) technique was utilized to conduct the experiments. The MSI data were analyzed using advanced data analysis methods to efficiently extract metabolomic information. Multivariate Curve Resolution Alternating Least Square (MCR-ALS) was used to decompose each MS image into different components with grouped species. To improve the efficiency of data analysis, both supervised (Random Forest) and unsupervised (Cluster Large Application (CLARA)) machine learning (ML) methods were employed to cluster

MS images according to their metabolomic features. Our results indicate that the anticancer drug significantly affected the abundances of a variety of metabolites in different regions of spheroids. This integrated experiment and data analysis approach can facilitate the studies of metabolites in different types of 3D tumor models and tissues, and potentially benefit the drug discovery, therapeutic resistance, and other biological research fields.

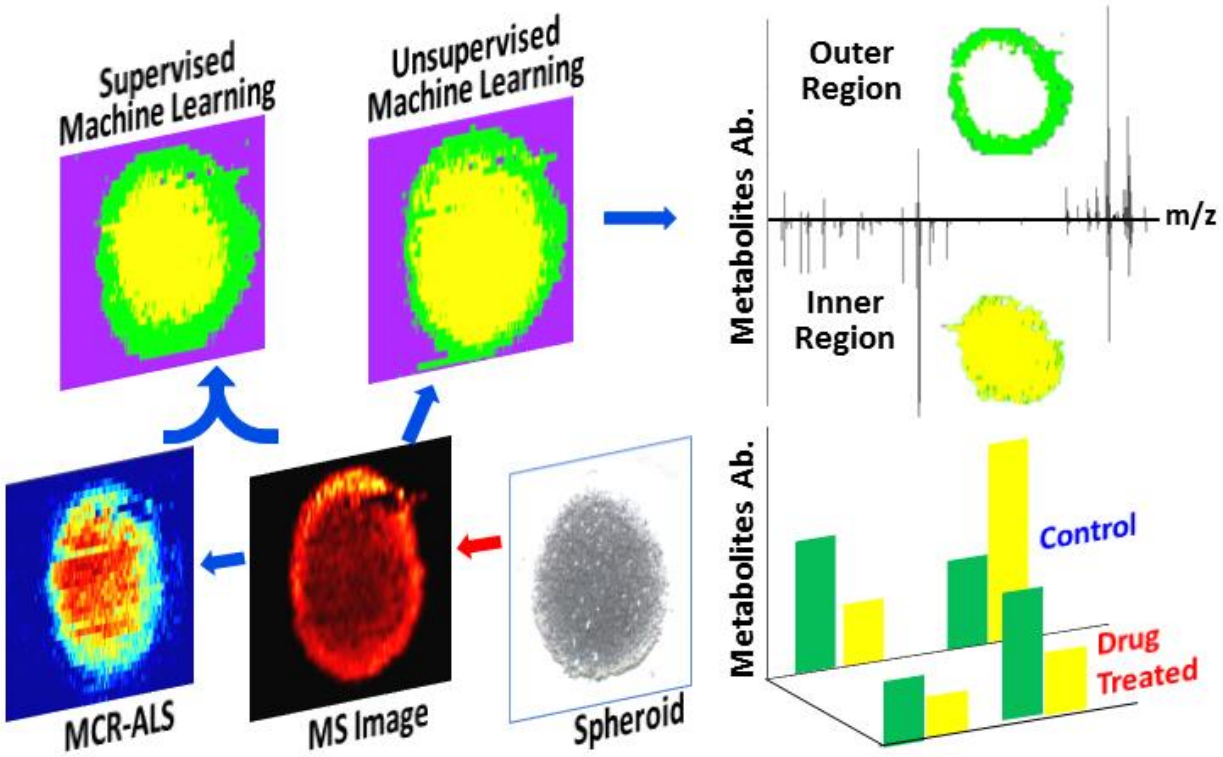


Figure 3-1. Graphic abstract

3.2. Introduction

Spheroids, the spherical aggregates of tumor cells, fill the gap between the simplified 2D cell culture models and very complex real tissues.[1] Compared with common 2D-cultured cells,

spheroids provide more vivid and cost-efficient models with a higher degree of relevance to clinical and biological applications.[2] Particularly, the 3D-structured spheroids can mimic the microenvironment of cells with higher fidelity.[3, 4] For example, the gradients of nutrients, oxygen, and pH value result in different proliferation status of cells from the inside to the outside regions of spheroids.[5] Spheroid has become an important platform for broader ranges of *in-vitro* studies such as proteomics,[6] drug screening[7], and metabolomics.[8, 9] Among these applications, metabolomics focuses on small molecules (e.g., M.W. <1500 Da),[10] with both endogenous and exogenous origins, in the biological samples such as cells, tissues, and biofluids.[11] The changes of metabolites can rapidly and directly reflect the state of biological systems affected by a variety of factors, including microenvironment perturbation, genetic mutation, kinetic activity of enzymes, and changes in metabolic reactions.[12-14] Metabolomics studies utilizing spheroids have a broad influence on drug discover, toxicology, and disease diagnosis.[7, 15]

Current metabolomics studies of biological tissues are primarily carried out using lysates prepared from samples, and measurements are conducted using mass spectrometry (MS), which is usually coupled to liquid chromatography (LC) or gas chromatography (GC) separation techniques,[16] or nuclear magnetic resonance (NMR), typically ^{1}H NMR.[17] However, because lysates need to be prepared from homogenized samples,[18] the spatial distribution of metabolites, which is critical to understand the complex biological process and the pathophysiology, is inevitably lost.[19, 20] To obtain the spatially resolved metabolites, molecular imaging techniques, such as positron emission tomography (PET), magnetic resonance imaging (MRI), and MS imaging (MSI), have been developed. PET can locate tumor areas using certain target molecules (e.g., radiolabeled glucose (2-[^{18}F]fluoro-2-deoxy-D-glucose (FDG)) owing to their accumulations in tumors.[21] MRI can diagnose many types of cancers by visualizing certain metabolomic biomarkers.[22, 23] However, the

broader applications of PET and MRI are largely limited by their relatively low coverage of molecular types[24]. MS imaging (MSI), with high sensitivity and wide ranges of molecular coverage, is a powerful technique to visualize the distribution of metabolites on tissue slices.[20] MSI has been applied to numerous metabolomics studies of plants,[25] drugs,[26] and diseases such as cancers.[27, 28] Among all developed MSI techniques, vacuum based sampling and ionization techniques, such as matrix assist laser desorption ionization MS (MALDI-MS) and secondary ion MS (SIMS), provide superior sensitivity and excellent spatial resolution,[29] whereas ambient MSI techniques, such as desorption electrospray ionization (DESI)[30] and laser ablation electrospray ionization (LAESI)[31], require minimum sample preparation and allow for experiments to be conducted under convenient conditions.[32] Particularly, the absence of matrix molecules in sample preparation enables ambient MSI techniques to effectively detect small molecules such as metabolites.[33]

In addition to the rapid development of MSI experimental techniques, advanced data analysis methods become increasingly important to effectively extract chemical information from a large amount of MSI data (e.g., several to hundred GB can be typically generated from MALDI MSI experiments[34]). Although the traditional methods, such as Principle Component Analysis (PCA)[35] and Partial Least Squares Discriminant Analysis (PLS-DA),[36] have been applied to the analysis of MSI data, they have some intrinsic limitations[4]. For example, the negative values in PCA and PLS-DA score plot have no physical meaning (i.e., ion intensities in mass spectra cannot be negative).[37, 38] In contrast, Multivariate Curve Resolution (MCR-ALS), a multivariate data analysis method, has the advantages of not only decomposing the image data into major components but also extracting the ions contributing to each component.[39, 40] However, because of intensive computing is needed during the MCR-ALS analysis, this method is less than ideal when analyzing large

amounts of MSI data, particularly for samples possessing similar features such as slices obtained from the same tissue.

Machine learning (ML) methods, including unsupervised and supervised ML, are powerful tools for efficient MSI data processing.[41, 42] Unsupervised ML is utilized to cluster MS image data into different groups of molecules according to their similarities of MS profiles without any prior knowledge.[38] Common unsupervised ML algorithms include hierarchical clustering, k-means, and Cluster Large Application (CLARA).[34] These common ML methods have been employed for MSI data analysis. Hierarchical clustering is an algorithm to build hierarchical cluster tree, and each branch is split and followed down to individual groups of data such as mass spectra in MS imaging datasets.[43] K-means is a widely used method; however, it is sensitive to outliers, thus small numbers of pixels (i.e., potential outliers) in MS image data might be unnecessarily grouped into clusters[34]. CLARA is superior to other unsupervised ML methods for MSI data analysis due to its capability of optimizing the cluster numbers of MS image data.[40, 44] On the other hand, supervised ML methods, including Supporting Vector Machine (SVM) and Random Forest, can be employed to classify MS image data into several predefined groups after model training.[45] SVM is used to find a hyperplane that can separates one or more classes based on their mass spectrum.[46] Random Forest has less over-fitting issue, and it is efficient in analyzing large dataset.[47, 48] Both unsupervised and supervised ML methods can facilitate the analysis of metabolites in MS imaging data.

Here we used the Single-probe MSI technique to measure the spatial distributions of metabolites in spheroids, which were cultured using HCT-116 colorectal cancer cells and treated with anticancer drug Irinotecan. The Single-probe is a multifunctional sampling and ionization device that has been applied into many research fields, including high resolution MSI,[49, 50] live single cell

analysis,[51-56] and the measurement of extracellular metabolites in live spheroids.[9] MSI data were then analyzed using MCR-ALS, unsupervised ML (CLARA), and supervised ML (Random Forest) methods. MCR-ALS was used to decompose the MS images into major components with grouped molecules, whereas Random Forest and CLARA were utilized to efficiently classify the MS images into different regions in spheroids. In addition, we investigated the metabolites altered by anticancer drug treatment in different regions of spheroids. Our studies provided a combined MSI experiment and advanced data analysis to effectively analyze metabolomics and investigate the influence of microenvironment on metabolite change in tissues. This comprehensive method can potentially benefit the biomarker discovery and metabolomics studies.

3.3. Experimental section

3.3.1. Chemicals and materials

Chemicals used in the experiments include reagents, such as methanol, water, agarose (Sigma-Aldrich, St. Louis, MO, USA), and anticancer drug Irinotecan (Thermo Scientific, Ward Hill, MA, USA). Materials needed to fabricate the Single-probe include fused silica capillary (O.D. 105 µm; I.D. 40 µm, Polymicro Technologies, Phoenix, AZ, USA) and dual-bore quartz tubing (O.D. 500 µm; I.D. 127 µm, Friedrich & Dimmock, Inc., Millville, NJ, USA). Reagents used to culture HCT-11 cells and spheroids (ATCC, Manassas, VA, USA) include McCoy's 5A cell culture media, FBS (fetal bovine serum), and Pen Strep (Life Technologies, Grand Island, NY, USA).

3.3.2. The Single-probe fabrication and experimental setup

The detailed fabrication protocols are described in our previous work,[50] and only the outlined procedures are provided here. The Single-probe (Figure S3-1) has three components: a laser-pulled dual-bore quartz needle, a fused silica capillary (solvent-providing capillary), and a nano-electrospray (nano-ESI) emitter. A Sutter P-2000 laser micropipette puller (Sutter Instrument,

Novato, CA, USA) is used to prepare the dual-bore needle and nano-ESI emitter. A Single-probe is fabricated by embedding within a dual-bore quartz needle with one fused silica capillary and one nano-ESI emitter. The experiment setup is largely adopted from our previous MS single cell and MSI studies[49, 50] (Figure 3-2). To precisely control the movement of tissue slice, the sample was attached to a XYZ-translational stage system (CONEX-MFACC, Newport Co., Irvine, CA, USA) controlled using a LabView software package.[57] A digital microscope was placed next to the Single-probe to adjust the distance between the Single-probe tip and tissue slice surface, and to monitor the sampling process. MS spectra were collected using a Thermo LTQ XL mass spectrometer (Thermo Scientific, Waltham, MA, USA) with the following parameters: mass resolution 60,000 (m/Δm), 4.5 kV ionization voltage (positive ion mode), 1 microscan, 100 ms max injection time, and AGC on. The sampling solvent (i.e., 85% methanol/15%water (v/v)) was continuously delivered (flowrate 200 nL/min) by a syringe pump (PHD ULTRA, Harvard Apparatus, Holliston, MA, USA). The MS images were generated using MSI QuickView software.[58]

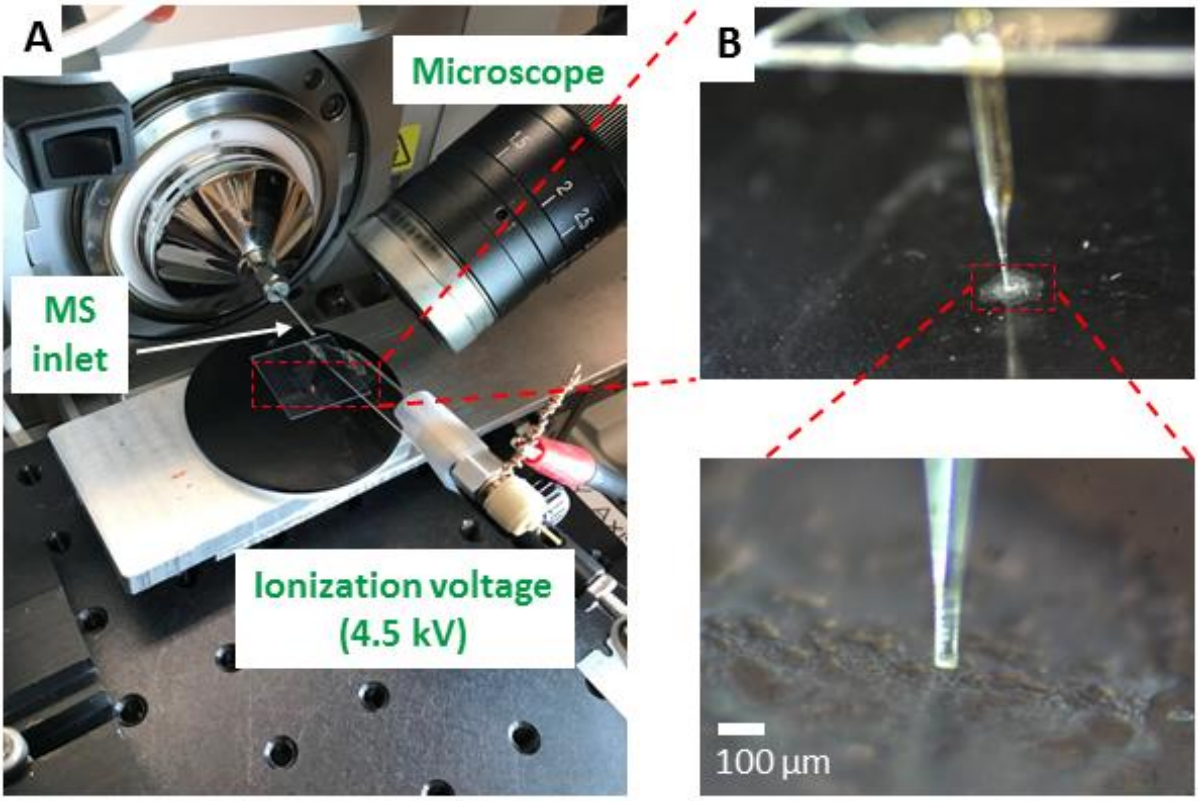


Figure 3-2. The Single-probe MSI setup.

3.3.3. Spheroids culture and sectioning

The colon carcinoma cell line HCT-116 was used to culture spheroids. Cells were maintained in McCoy's 5A cell culture media containing 10% FBS and 1% Pen Strep. Cells were grown in an incubator (HERA cell, Heraeus, USA) under well controlled conditions (5% CO_2 at 37 °C), and cell passage was performed every two days. Spheroids were cultured using the modified protocols based on previous publications[59, 60]. Briefly, 60 µL agarose gel (1.7% agarose in plain McCoy's culture media) was used to coat wells in a U-bottom 96-well plate (VWR, Radnor, PA, USA), and about 10,000 HCT-116 cells were seeded into each agarose coated well. To improve the success rate of spheroid culture, we used the U-bottom 96-well plates to promote the accumulation of cells

during spheroid growth. Cells were incubated for 2-3 days to allow them aggregating into solid spheroids. Culture medium was changed every two days, and spheroids were harvested after being cultured for 10 days. Spheroids were treated using Irinotecan under a series of conditions (i.e., 5 μM for 1 h, 10 μM for 1 h, 20.6 μM for 1 h, 20.6 μM for 10 h, and 20.6 μM for 24 h), and rinsed twice using PBS (phosphate-buffered saline) to remove the drug residue on spheroid surface. Spheroids were then embedded in 10% HMPC ((Hydroxypropyl)methyl cellulose, Sigma-Aldrich St. Louis, MO)) and frozen on dry ice. Depending on their sizes, spheroids were sectioned into 20 to 30 slices (~15 μm thickness for each slice) using a cryotome (American Optical 845 Cryo-cut Mictorome, Southbridge, MA, USA) at –15°C. To better represent the symmetric histological structures of spheroids (e.g., inner region and outer region), slices in the middle part of well-aligned spheroids (e.g., slice number ~10 to ~15) were selected and inspected using microscope. The selected slices were attached onto plastic microscope slides (VWR, Radnor, PA, USA), dried in air, and stored at –80°C before usage. The optical images of slides were taken using a PathScan Enabler IV histology slide scanner (Meyer Instruments, Houston, TX, USA).

3.3.4. Data Analysis

3.3.4.1. Data pre-processing.

Before conducting the MCR-ALS analysis, MSI data (.raw) need to be converted into an appropriate format that can be input into a home-built MATLAB processing platform. First, MSConvert (a tool in ProteoWizard) was used to convert the data format from .raw to .mzML, which was further converted into .imzML format using imzML Converter[61]. Second, data pre-processing, including smoothing, noise removal, peak alignment, peak picking, and insensitivity normalization, was executed using Bioinformatics Toolbox, a built-in function of MATLAB. Third, each set of MS imaging data was exported as a data matrix with high dimensionalities (Table

S3-1). For example, the data matrix of the MS image of a control sample slice is composed of 7840 X 307, i.e., 7840 pixels (196 (scans/line) X 40 lines) with 307 aligned common ions, and each aligned peak stands for a dimension of the dataset. The number of pixels (ranging from 6,000 to 10,000) of each MS image depend on the size and the spatial resolution of the MS image. A log-transformation (log 2) was applied to the data matrix produced from the pre-processing step to better represent features of low-intensity ions. The details of data pre-processing are provided in the Supporting Information.

3.3.4.2. MCR-ALS analysis

Multivariate Curve Resolution Toolbox, developed by Tauler et al.[62], was utilized to group the ions with similar spatial distribution patterns in MS images. Singular Value Decomposition (SVD) was used to determine the number of major components, which usually represent the majority of variations (> 80%) in the data matrix. The spatial distribution of each component was carried out using a user-friendly interface, and the mass spectrum of each component was generated using R (Supporting Information).

3.3.4.3. Unsupervised ML (CLARA).

CLARA was used as an unsupervised algorithm to analyze the MSI datasets. To generate datasets that are suitable for CLARA model, *t*-SNE (t-Distributed Stochastic Neighbor Embedding) was used to reduce the high dimensionality of the data matrix generated from the pre-processing into 2D space (parameters are listed in the Supporting Information). CLARA (provided within 'cluster' package[63]) was then used to analyze the datasets with 2D space, and the optimal numbers of clusters were determined using the average silhouette width. To illustrate the CLARA results according to the spatial distribution of ions in each group, their histological distributions were reconstructed using a homebuilt R program.

3.3.4.4. Supervised ML (Random Forest).

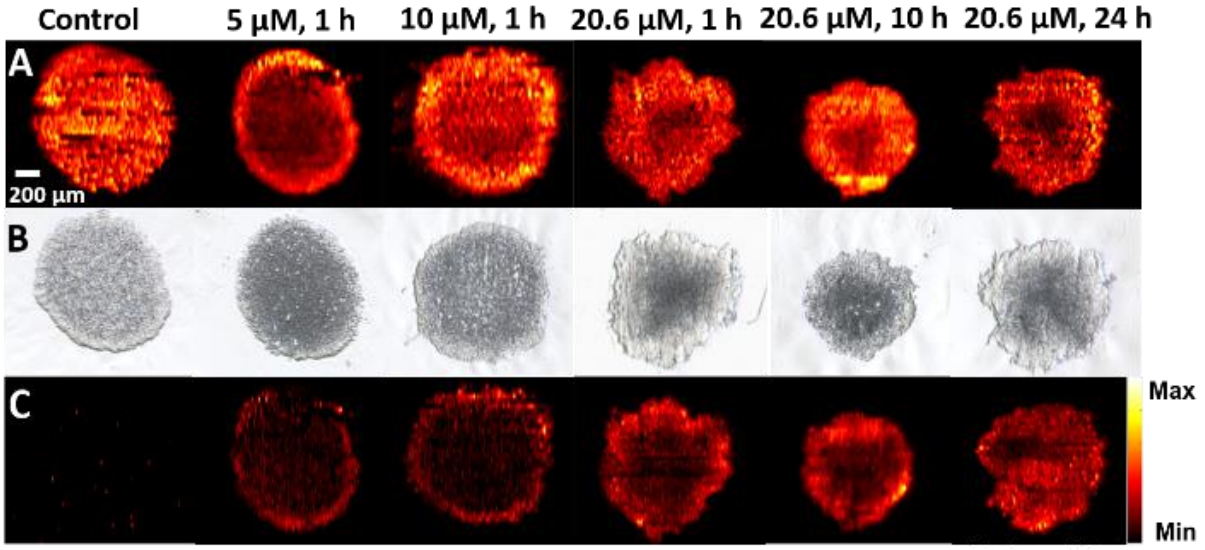
Random forest provided in R language (the package 'randomForest'[64]) was used a supervised ML algorithm to classify the MSI datasets. Based on MCR-ALS results, data were selected from each region (i.e., inner or outer) of spheroid to serve as the training data and testing data. The training data were used to train the Random Forest model (details are provided in the Supporting Information), whereas the testing data were utilized to evaluate prediction accuracy (Table S3-2). Once the prediction accuracy was satisfactory (> 90%), the optimized model was employed to predict the rest of the dataset. To visualize overall results obtained from the supervised ML approach, the predicted spatial distribution features of the metabolites on spheroid slices were constructed using R language. (All packages used in ML models can be found at https://cran.r-project.org).

3.4. Results and discussion

3.4.1. Mass spectrometry images of spheroids

Irinotecan is a common anticancer drug that has been widely used in clinic treatment of broad ranges of cancers (e.g., colorectal, pancreatic, and lung cancers) and fundamental research. This drug compound was selected to treat spheroids cultured using a colorectal cancer cell line (HCT-116). To investigate the influence of treatment time and concentration on the change of metabolites, experiments were conducted using spheroids in three different groups: control (no drug treatment), concentration-dependent treatment, and time-dependent treatment. In concentration-dependent experiments, the treatment time was fixed at 1 hour, whereas three different concentrations (5.0, 10.0, and 20.6 μM) were selected. In the time-dependent experiment, the concentration was 20.6 μM, which is the IC_{50} of Irinotecan for HCT-116 spheroids[8], whereas the treatment time was varied (1, 10, and 24 h). MSI experiments were carried out using the Single-probe MSI technique, and

MS images of selected ions were constructed (Figure 3-3). Because phosphatidylcholine (34:1) (PC(34:1)) is a very common and abundant lipid (with the ion intensity ~10^7), its MS image was



used to represent the shape of the spheroid slice. The MS images of this representative phospholipid ($[PC(34:1) + H]^+$, Figure 3-3A) and the drug compound ($[Irinotecan + H]^+$, Figure 3-3C) were shown along with the corresponding optical images of the spheroid slices (Figure 3-3B), indicating a generally good match of spatial features obtained from both techniques. It is worth noting that all spheroids were cultured under the same conditions and their slices were prepared using the same protocols; however, the shapes of slices can be slightly different, likely due to the variance in multiple factors such as the number of cells seeded for spheroid culture, orientation of spheroids in bedding material, and sectioning positions in spheroids. Interestingly, the morphology of spheroids was not obviously changed within our drug treatment time (24 h) as shown in Figure S3-2.

Figure 3-3. MS and optical images of spheroid slices.

(A) A common ion $[PC(34:1) + H]^+$ (*m/z* 760.5964) illustrating shapes of MS images of spheroid slices in the control and drug treatment groups. (B) Optical images of the corresponding spheroid slices. (C) MS images of $[Irinotecan + H]^+$ (*m/z* 587.2923) in drug-treated spheroids.

As expected, Irinotecan was detected in spheroids subjected to the treatment using this drug compound, but not observed in the control sample. In the concentration-dependent experiments (Figure 3-3C), Irinotecan was only detected in a thin layer of the spheroid for 1 h treatment; the drug penetration depth was slightly increased as the concentration was increased from 5.0 to 20.6 μM. Alternatively, a contour plot (Figure S3-3) provides a more quantitative description of Irinotecan distribution of the same sample.

In the time-dependent experiments, this drug compound was observed in much deeper regions as the treatment time was increased. For example, Irinotecan mainly distributed in the outer region of the spheroid for 10 h treatment, whereas it was detected in the inner region after 24 h treatment. The distribution patterns of Irinotecan can likely be attributed to the mechanisms of molecular diffusion in tissues. Due to the lack of developed blood vascular systems inside tumors, molecules (e.g., nutrients and drugs) penetrate tumors mainly through molecular diffusion mechanism, which is primarily determined by drug concentration[65] and other factors such as drug molecular weight, cell density in tumors, and cells' microenvironment.[66] Similar to tumors, the 3D structure of spheroids possesses histological heterogeneity due to heterogeneous distribution of nutrients, oxygen, carbon dioxide, wastes, etc. For example, nutrients and oxygen can be absorbed more efficiently by the cells in the outer region than those in the inner region, whereas wastes and carbon dioxide are more abundant in the inner region than the outer region.[2, 67]

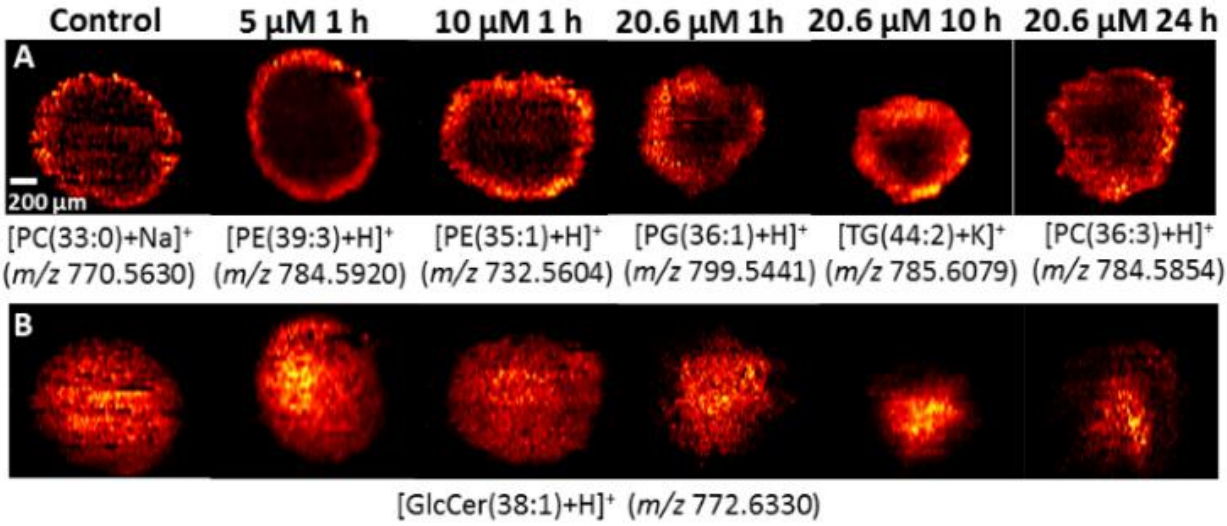


Figure 3-4. MS images of spheroid slices from control and drug-treated samples.

(A)MS images of representative metabolites primarily present in the outer region. (B) The influence of anticancer drug treatment on the spatial distribution of a selected metabolite GlcCer(38:1).

The heterogeneous cell microenvironment inside spheroids can affect the cell metabolic pathways and result in metabolites changes in different regions. To illustrate the heterogeneity of spatial distribution of molecules in spheroids, we constructed the MS images of a number of representative lipids (e.g., PC, phosphatidylethanolamine (PE), and phosphatidylglycerol (PG)) that are more abundant in the outer regions (Figure 3-4A). In addition, we investigated the influence of anticancer drug treatment on the spatial distribution of metabolites such as glucosylceramides $[GlcCer(38:1) + H)]^+$ (Figure 3-4B). Previous studies indicate that the upregulation of GlcCer in cancer cells is attributed to their hypoxia environment, in which the glucosylceramide synthase (GCS) possesses higher activities transferring ceramides to glucosylceramides.[68,69] Our results (Figure 3-4B) indicate that GlcCer(38:1) has relatively higher abundances in the inner regions of spheroids, in which cells are in hypoxia environment. Compared

with cells in the outer regions, Irinotecan treatment has less influence on the metabolism and microenvironment of cells in the inner regions.

3.4.2. Multivariate Curve Resolution analysis

It is expected that there are large numbers of grouped metabolites possessing the similar spatial distributions, whereas manual selection approach is unlikely to be effective to acquire all metabolites in each group. For more comprehensive and efficient data analysis, MCR-ALS was employed to classify ions into different groups according to their spatial distribution features. The data generated from the pre-processing step (see **Data Pre-processing**) was directly analyzed using MCR-ALS algorithm. The MSI data of drug-treated samples (5 μM Irinotecan, 1 h) were decomposed into three major components, which represent the majority molecular information present in the MSI data, i.e., >80% of variance was explained (Figure S3-4). These three components (i.e., grouped ions) represent species in the outer region, the inner region, and the background, respectively (Figure 3-5). The alternative contour plots of Figure 3-5 were constructed based on the MCR analysis results (Figure S3-5). The top-ten abundant ions in each component were summarized in Table S3-3 and Table S3-4. Similarly, the MSI data obtained from

spheroids under other treatment conditions were also decomposed into these three groups (Figure S3-6).

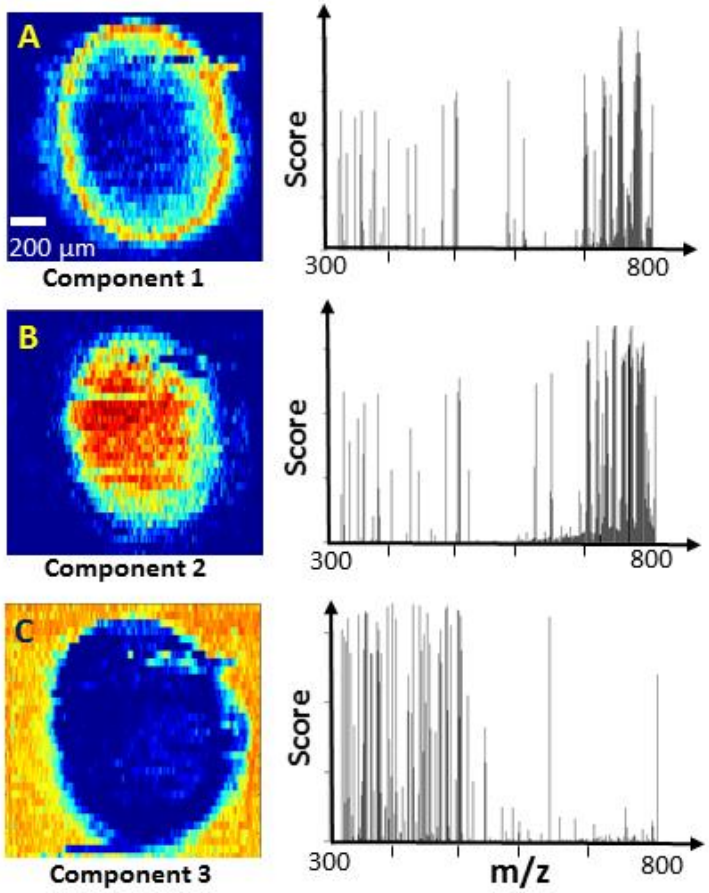


Figure 3-5. Results of MCR-ALS analysis of MSI data obtained from Irinotecan treated spheroid (5 µM, 1 h).

(A) The spatial distribution patterns and the corresponding ions of (A) component 1 (outer region of spheroid), (B) component 2 (inner region of spheroid), and (C) component 3 (background).

The spatial distribution patterns of metabolites are likely related to the inherent heterogeneity of spheroids. Previous studies indicate that spheroids can be divided into two major regions: non-viable and viable regions.[70] In the non-viable region, nutrients, and oxygen are insufficient to sustain cells' viability. Because cells in this region undergo necrosis and apoptosis due to the hypoxia microenvironment, the non-viable region can be further divided into quiescent layer and necrotic core.[66] In contrast, cells in the viable region can acquire adequate nutrients and oxygen

for proliferation. Due to the diffusion-limited transport of oxygen,[2] the thickness of viable region generally ranges from 150 to 200 μm, which is similar to the thickness of the outer region measured from MS images of spheroids in the present study (Figure 3-5A). Therefore, those two regions in spheroids obtained from MCR-ALS analysis (Figures 3-5A and 3-5B) are very likely to be the viable (outer) and non-viable (inner) regions, respectively. However, we were unable to differentiate the quiescent layer and necrotic core in the non-viable region in the current MSI studies. Similar results were obtained from spheroids in the control and other drug-treated groups using higher drug concentrations and longer treatment times (Figure S3-6).

3.4.3. ML: CLARA and Random Forest

For more efficient data analysis, the unsupervised ML was applied to classify the MSI data. First, *t*-SNE was used to reduce the high dimensionality of the pre-processed data matrix into a 2D space (Figure S3-7B) along with the optimal number (i.e., three) of clusters (Figure S3-7A); the same number of grouped ions obtained from MCR-ALS analysis. In fact, the 2D *t*-SNE shows a highly stable performance for dimensionality reduction, indicated by a narrow range (0.85 - 0.91) of the calculated Adjusted Rand Index (Figure S3-8). We also tested the 3D *t*-SNE and obtained similar result as the 2D *t*-SNE (Figure S3-9). Second, we utilized CLARA, a common classification method primarily used for grouping large datasets with optimal numbers of clusters[44], to analyze the MSI data. Third, to clearly illustrate the spatial locations of clustered ions on spheroid slice, we assigned a color to each cluster (i.e., color coding), and reconstructed MS images (Figure 3-6A) using these clustered species (Figure 3-6B), in which green, yellow, and purple represent the outer region, the inner region, and the background, respectively. As a popular method, PCA has been utilized for the analysis of MS images.[4, 35] We tested PCA for dimensionality reduction of our MSI datasets, and then performed CLARA. The results are not satisfactory: only two major

components (i.e., the entire spheroid and the background) can be obtained, whereas the inner and outer regions of the spheroid cannot be distinguished (Figure S3-10).

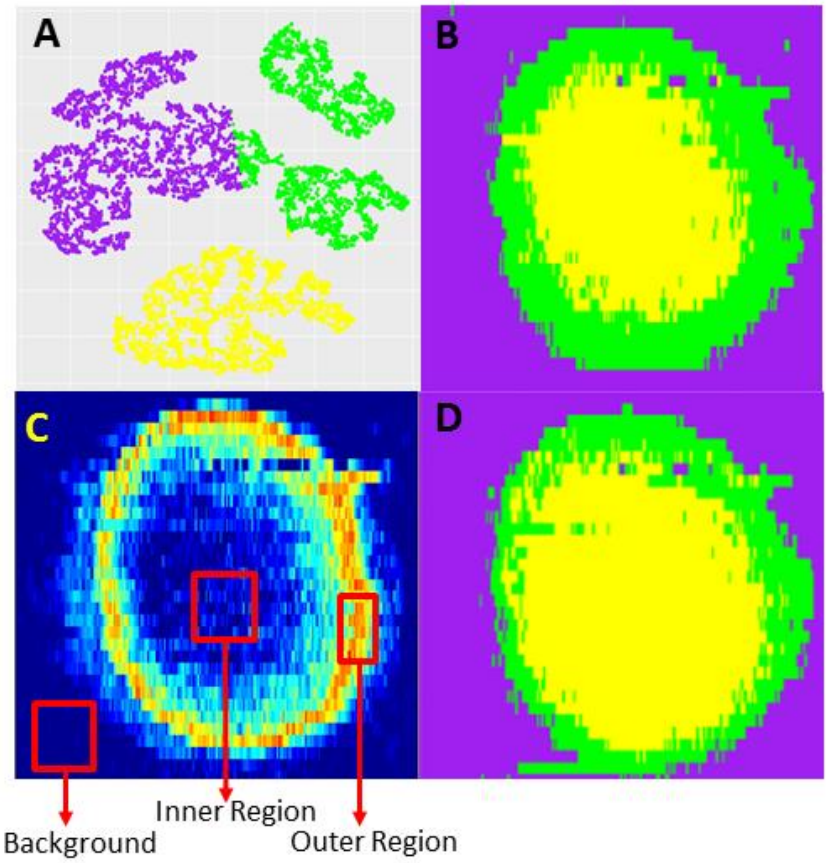


Figure 3-6. MSI data analysis using ML methods for an Irinotecan treated spheroid (5 µM, 1 h).

(A) Classification of MSI data obtained from unsupervised ML (CLARA). (B) Three clustered MSI data from CLARA analysis were color-coded to reconstruct their spatial distributions representing the inner region (yellow), the outer region (green), and the background (purple), respectively. (C) The selection of the training data for supervised ML (Random Forest) was based on MCR-ALS results. (D) Color-coded spatial distributions obtained from Random Forest are similar to CLARA results shown in 5B.

Supervised ML (Random Forest) was also used to classify the MS images of spheroid slices. The selection of training data was based on the MCR-ALS results (Figure 3-6C, details are provided in Table S3-5), and the color-coded spatial distributions (Figure 3-6D) were reconstructed using

the same approach as mentioned above. Similar to CLARA results, we were able to predict all MSI data and classify them into those three regions using Random Forest (Figure S3-11B). Interestingly, as shown in all six color-coded classification images generated using unsupervised (Figure S3-11A) and supervised (Figure S3-11B) ML methods, the thickness of outer region of spheroids are not significantly affected by the drug treatment. This observation is similar to the results obtained from MCR-ALS analysis.

Because training data selection may affect the prediction accuracy of supervised ML, we compared the performance of MCR-ALS with the other two common methods: k-means and hierarchical clustering. Our results indicate the training data selection based on three different methods (i.e., MCR-ALS, k-means, and hierarchical clustering) resulted in very similar prediction of MSI data (Figure S3-12).

3.4.4. Changes of metabolites inside spheroids induced by Irinotecan treatment

Anticancer drug treatment can change the microenvironment of cells, affect metabolomic pathways, and result in changes of metabolites. To obtain metabolites altered by Irinotecan inside spheroids, we generated two averaged mass spectra representing metabolites in the inner and outer regions of spheroids in both control and treatment groups. Specifically, according to results obtained from unsupervised and supervised ML models, mass spectra representing inner and outer regions, respectively, were averaged to generate these two mass spectra, and the intensity of each ion was then normalized to TIC (total ion current). By conducting *t*-test of species in the averaged spectra, metabolites with significantly changed abundances can be determined (details are provided in the Supporting Information).

For the time-dependent experimental results, supervised ML analyses provided a number of metabolites with relatively higher abundances (i.e., upregulation) in the drug-treated spheroids

than those in the control sample. For example, the relative intensities of $[PC(32:1) + H]^+$ (m/z 732.5483) and $[PE(O\text{-}35:0) + H]^+$ (m/z 720.5908) in the inner and outer regions of the drug-treated spheroids are about 10 and 3─4 times higher, respectively, than those in the same regions in the control (Figures 3-7A and 3-7B). Similarly, in the concentration-dependent experiments, a group of metabolites possess significantly different abundances in the outer and inner regions between the drug-treated and control samples. For example, the relative ion intensity of $[PC(34:2) + H]^+$ (m/z 758.5685) is about five times higher in the inner region of the control than that in the treated spheroid (20.6 μM Irinotecan). Similar trends were obtained from unsupervised ML analysis (Figure 3-7G) with most metabolites were upregulated due to the drug treatment. Because the comparison of ion intensities (Figure 3-7) was based on the grouped data (i.e., pixels in inner and outer regions) from the supervised and unsupervised ML methods (e.g., Figures 3-6B and 3-6D), minor changes of data grouping using different approaches may lead to variances in relative ion intensities. For example, the supervised ML results (Figure 3-7C) indicate that the relative intensity of $[PC(34:2) + H]^+$ in the outer region of the spheroid increased as the drug concentration was increased (from 5 μM to 10 μM), whereas this order was reversed in the unsupervised ML results (Figure 3-7G).

All ions significantly altered by the drug treatment in the outer or inner region are summarized in Table S3-6 and Table S3-7. The identifications of all ions were carried out using tandem MS (MS/MS) measurements (Figure S3-13), and these spectra were compared with online database Metlin[71]. To further confirm the metabolites' structures, MS^n analysis can be carried out using standard compounds for comparison. For example, the detection of $[PC (34:1) + H]^+$ from spheroid slice was confirmed by comparing its MS^3 spectra with those obtained from the standard compound (Figure S3-14).

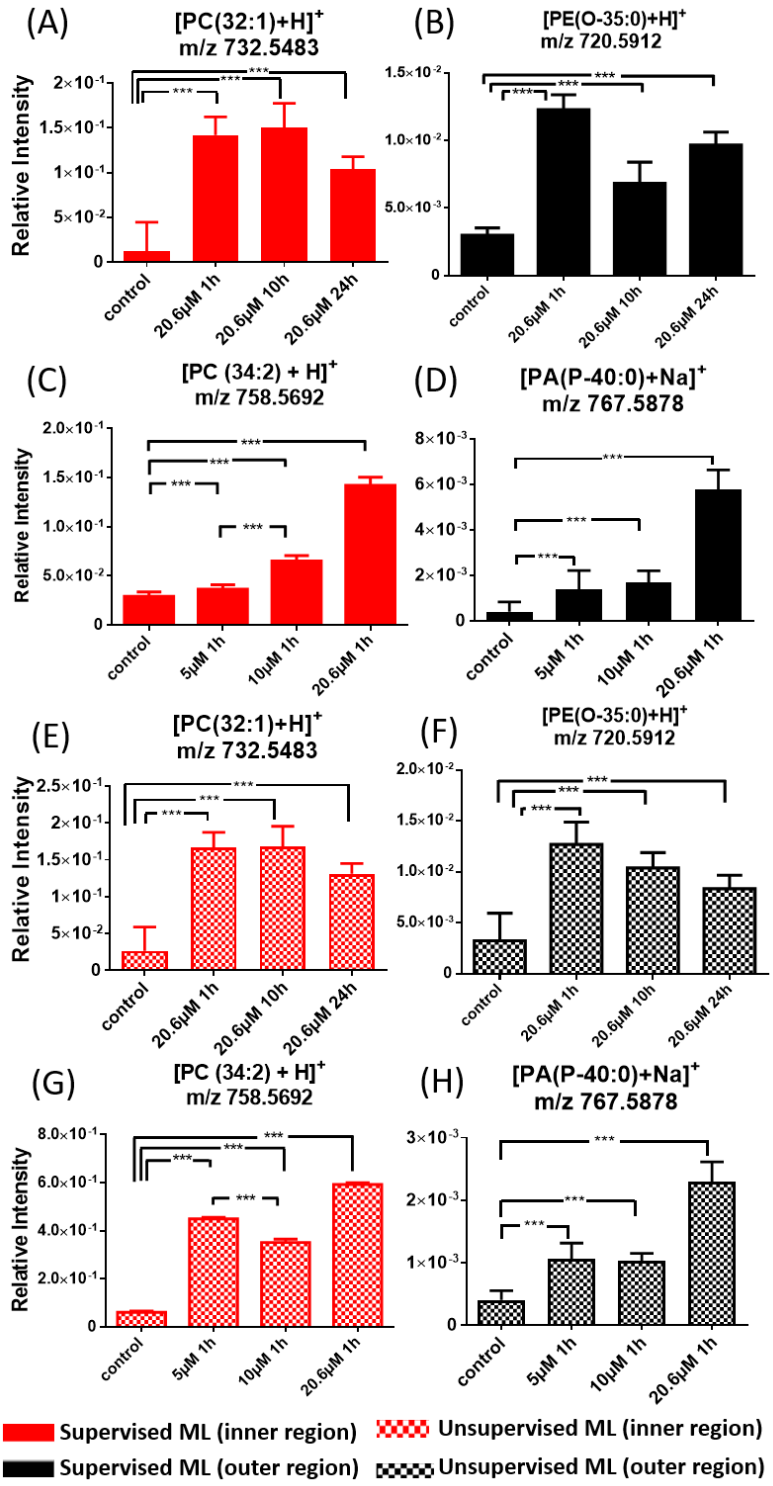


Figure 3-7. Representative common metabolites in the inner and outer regions upregulated by Irinotecan treatment.

Results were obtained from (A)-(D) supervised ML (Random Forest) and (E)-(H) unsupervised ML (CLARA) analyses. (*** $p < 0.001$).

Large numbers of phospholipids were observed in our experiments. Phospholipids are structural building blocks of cell membranes. In addition, they have important biological functions (e.g., signaling, energy storage, and disease biomarker[72]), and play key roles in the cell motility, invasion, and metastasis.[73, 74] The composition of phospholipids can be affected by the extracellular stimuli such as growth factor, oncogene, and hypoxia conditions.[75] In our study, the relative abundances of a series of phospholipids, including $[PC(32:1) + H]^+$ (m/z 732.5483) and $[PC(34:2) + H]^+$ (m/z 758.5692), were significantly higher in the Irinotecan treated groups compared with the control. Previous studies indicate that tumor cells can alter their metabolisms to reinforce drug resistance.[76] For example, nuclear receptor PXR (Pregnant X receptor) in cancer cells is activated upon their exposure to Irinotecan, resulting in an overexpression of CYP3A4, which is an oxidizing enzyme for xenobiotics such as drug molecules. Overexpressed CYP3A4 further reduces the abundance of Irinotecan inside cells and lead to strengthened drug resistance.[77] Meanwhile, the *de novo* lipogenesis is enhanced the by the activated PXR to increase the level of phospholipids,[78] which are also believed to promote the cell survival.[79, 80] Similarly, our previous studies of the extracellular species in spheroids indicate that a large number of phospholipids (e.g., PC, PE, and PA) are increased by Irinotecan treatment.[9] Therefore, the increased levels of phospholipids are very likely attributed the metabolomic response of cancer cells to the anticancer drug in the microenvironment inside spheroids.

3.5. Conclusion

In this study, we cultured HCT-116 spheroids as tumor models and treated them using anticancer drug Irinotecan. Experiments were carried out using the Single-probe MSI technique. Advanced data analysis methods, including MCR-ALS and ML, were employed to analyze spatially resolved

metabolites on spheroid slices, particularly for those significantly regulated by the drug compound. MCR-ALS algorithm was utilized to decompose each MS image data into three major components: the inner region, the outer region, and the background. Both unsupervised (Random Forest) and supervised (CLARA) ML algorithms were employed to classify each MS image, and similar results were obtained. We further compared the metabolites between the inner and outer regions for each set of MS image data obtained from spheroids in the control and drug treatment groups. We acquired the grouped species in both inner and outer regions that were significantly regulated by the anticancer drug. Compared with conventional methods based on bulk analysis, our method can potentially extract crucial metabolomic information from MSI data, and provide spatially resolved metabolites and biomarkers reflecting the influence of drug treatment. The fully established methods may benefit the drug screening, therapeutic resistance, and biomarker discovery.

The material in chapter 3 is adapted from Tian, X.; Zhang, G.; Zou, Z.; Yang, Z., *Anal. Chem.* **2019,** *91* (9), 5802-5809. The copyright permission is obtained from ACS.

References

(1) Fennema, E.; Rivron, N.; Rouwkema, J.; van Blitterswijk, C.; de Boer, J., *Trends Biotechnol.* **2013,** *31* (2), 108-15.

(2) Lin, R. Z.; Chang, H. Y., *Biotechnol. J.* **2008,** *3* (9-10), 1172-84.

(3) Hirschhaeuser, F.; Menne, H.; Dittfeld, C.; West, J.; Mueller-Klieser, W.; Kunz-Schughart, L. A., *J. Biotechnol.* **2010,** *148* (1), 3-15.

(4) Weaver, E. M.; Hummon, A. B.; Keithley, R. B., *Anal. Methods* **2015,** *7* (17), 7208-7219.

(5) Minchinton, A. I.; Tannock, I. F., *Nat. Rev. Cancer* **2006,** *6* (8), 583-92.

(6) Gaedtke, L.; Thoenes, L.; Culmsee, C.; Mayer, B.; Wagner, E., *J. Proteome Res.* **2007,** *6* (11), 4111-4118.

(7) Friedrich, J.; Seidel, C.; Ebner, R.; Kunz-Schughart, L. A., *Nat. Protoc.* **2009,** *4* (3), 309-24.

(8) Liu, X.; Weaver, E. M.; Hummon, A. B., *Anal. Chem.* **2013,** *85* (13), 6295-302.

(9) Sun, M.; Tian, X.; Yang, Z., *Anal. Chem.* **2017,** *89* (17), 9069-9076.

(10) Zhang, A.; Sun, H.; Wang, P.; Han, Y.; Wang, X., *Analyst* **2012,** *137* (2), 293-300.

(11) Bundy, J. G.; Davey, M. P.; Viant, M. R., *Metabolomics* **2008,** *5* (1), 3-21.

(12) Sreekumar, A.; Poisson, L. M.; Rajendiran, T. M.; Khan, A. P.; Cao, Q.; Yu, J.; Laxman, B.; Mehra, R.; Lonigro, R. J.; Li, Y.; Nyati, M. K.; Ahsan, A.; Kalyana-Sundaram, S.; Han, B.; Cao, X.; Byun, J.; Omenn, G. S.; Ghosh, D.; Pennathur, S.; Alexander, D. C.; Berger, A.; Shuster, J. R.; Wei, J. T.; Varambally, S.; Beecher, C.; Chinnaiyan, A. M., *Nature* **2009,** *457* (7231), 910-4.

(13) Spratlin, J. L.; Serkova, N. J.; Eckhardt, S. G., *Clin. Cancer Res.* **2009,** *15* (2), 431-40.

(14) Griffin, J. L.; Shockcor, J. P., *Nat. Rev. Cancer* **2004,** *4* (7), 551-61.

(15) Roux, A.; Lison, D.; Junot, C.; Heilier, J. F., *Clin. Biochem.* **2011,** *44* (1), 119-35.

(16) Major, H. J.; Williams, R.; Wilson, A. J.; Wilson, I. D., *Rapid Commun. Mass Spectrom.* **2006,** *20* (22), 3295-302.

(17) Weljie, A. M.; Newton, J.; Mercier, P.; Carlson, E.; Slupsky, C. M., *Anal. Chem.* **2006,** *78* (13), 4430-4442.

(18) Werner, E.; Croixmarie, V.; Umbdenstock, T.; Ezan, E.; Chaminade, P.; Tabet, J.-C.; Junot, C., *Anal. Chem.* **2008,** *80* (13), 4918-4932.

(19) Lee, D. Y.; Bowen, B. P.; Northen, T. R., *Biotechniques* **2010,** *49* (2), 557-65.

(20) Miura, D.; Fujimura, Y.; Wariishi, H., *J. Proteomics* **2012,** *75* (16), 5052-60.

(21) Evans, J. D.; Jethwa, K. R.; Ost, P.; Williams, S.; Kwon, E. D.; Lowe, V. J.; Davis, B. J., *Pract. Radiat. Oncol.* **2018,** *8* (1), 28-39.

(22) Beger, R. D., *Metabolites* **2013,** *3* (3), 552-74.

(23) Haddadin, I. S.; McIntosh, A.; Meisamy, S.; Corum, C.; Styczynski Snyder, A. L.; Powell, N. J.; Nelson, M. T.; Yee, D.; Garwood, M.; Bolan, P. J., *NMR Biomed.* **2009,** *22* (1), 65-76.

(24) Judenhofer, M. S.; Wehrl, H. F.; Newport, D. F.; Catana, C.; Siegel, S. B.; Becker, M.; Thielscher, A.; Kneilling, M.; Lichy, M. P.; Eichner, M.; Klingel, K.; Reischl, G.; Widmaier, S.; Rocken, M.; Nutt, R. E.; Machulla, H. J.; Uludag, K.; Cherry, S. R.; Claussen, C. D.; Pichler, B. J., *Nat. Med.* **2008,** *14* (4), 459-65.

(25) Li, Y.; Shrestha, B.; Vertes, A., *Anal. Chem.* **2008,** *80* (2), 407-420.

(26) He, J.; Luo, Z.; Huang, L.; He, J.; Chen, Y.; Rong, X.; Jia, S.; Tang, F.; Wang, X.; Zhang, R.; Zhang, J.; Shi, J.; Abliz, Z., *Anal. Chem.* **2015,** *87* (10), 5372-9.

(27) Wang, X.; Han, J.; Hardie, D. B.; Yang, J.; Pan, J.; Borchers, C. H., *Biochim. Biophys. Acta* **2017,** *1865* (7), 755-767.

(28) Kurreck, A.; Vandergrift, L. A.; Fuss, T. L.; Habbel, P.; Agar, N. Y. R.; Cheng, L. L., *Prostate Cancer Prostatic Dis.* **2018,** *21* (3), 297-305.

(29) Amstalden van Hove, E. R.; Smith, D. F.; Heeren, R. M., *J. Chromatogr. A* **2010,** *1217* (25), 3946-54.

(30) Takats, Z.; Wiseman, J. M.; Cooks, R. G., *J. Mass Spectrom.* **2005,** *40* (10), 1261-75.
(31) Nemes, P.; Vertes, A., *Anal. Chem.* **2007,** *79* (21), 8098-8106.
(32) Wu, C.; Dill, A. L.; Eberlin, L. S.; Cooks, R. G.; Ifa, D. R., *Mass Spectrom. Rev.* **2013,** *32* (3), 218-43.
(33) Knochenmuss, R.; Dubois, F.; Dale, M. J.; Zenobi, R., *Rapid Commun. Mass Spectrom.* **1996,** *10* (8), 871-877.
(34) McCombie, G.; Staab, D.; Stoeckli, M.; Knochenmuss, R., *Anal. Chem.* **2005,** *77* (19), 6118-24.
(35) Sodhi, R. N., *Analyst* **2004,** *129* (6), 483-7.
(36) Dill, A. L.; Eberlin, L. S.; Costa, A. B.; Zheng, C.; Ifa, D. R.; Cheng, L. A.; Masterson, T. A.; Koch, M. O.; Vitek, O.; Cooks, R. G., *Chem. Eur. J.* **2011,** *17* (10), 2897-2902.
(37) Alexandrov, T.; Becker, M.; Deininger, S. O.; Ernst, G.; Wehder, L.; Grasmair, M.; von Eggeling, F.; Thiele, H.; Maass, P., *J. Proteome Res.* **2010,** *9* (12), 6535-6546.
(38) Alexandrov, T., *BMC Bioinformatics* **2012,** *13 Suppl 16*, S11.
(39) Ruckebusch, C.; Blanchet, L., *Anal. Chim. Acta* **2013,** *765*, 28-36.
(40) Tian, X.; Zhang, G.; Shao, Y.; Yang, Z., *Anal. Chim. Acta* **2018,** *1037*, 211-219.
(41) Goodacre, R.; Vaidyanathan, S.; Dunn, W. B.; Harrigan, G. G.; Kell, D. B., *Trends Biotechnol.* **2004,** *22* (5), 245-52.
(42) Inglese, P.; McKenzie, J. S.; Mroz, A.; Kinross, J.; Veselkov, K.; Holmes, E.; Takats, Z.; Nicholson, J. K.; Glen, R. C., *Chem. Sci.* **2017,** *8* (5), 3500-3511.
(43) Deininger, S.-O.; Ebert, M. P.; Fütterer, A.; Gerhard, M.; Röcken, C., *J. Proteome Res.* **2008,** *7* (12), 5230-5236.
(44) Nagpal, A.; Jatain, A.; Gaur, D., *2013 Ieee Conference on Information and Communication Technologies (Ict 2013)* **2013**, 298-303.
(45) Heylman, C.; Datta, R.; Sobrino, A.; George, S.; Gratton, E., *PLoS ONE* **2015,** *10* (12).
(46) Schwamborn, K.; Krieg, R. C.; Reska, M.; Jakse, G.; Knuechel, R.; Wellmann, A., *Int. J. Mol. Med.* **2007,** *20* (2), 155-9.
(47) Hanselmann, M.; Kothe, U.; Kirchner, M.; Renard, B. Y.; Amstalden, E. R.; Glunde, K.; Heeren, R. M. A.; Hamprecht, F. A., *J. Proteome Res.* **2009,** *8* (7), 3558-3567.
(48) Smith, A.; Piga, I.; Galli, M.; Stella, M.; Denti, V.; Del Puppo, M.; Magni, F., *Int. J. Mol. Sci.* **2017,** *18* (12).
(49) Rao, W.; Pan, N.; Tian, X.; Yang, Z., *J. Am. Soc. Mass Spectrom.* **2016,** *27* (1), 124-34.
(50) Rao, W.; Pan, N.; Yang, Z. B., *J. Am. Soc. Mass Spectrom.* **2015,** *26* (6), 986-993.
(51) Pan, N.; Rao, W.; Kothapalli, N. R.; Liu, R.; Burgett, A. W.; Yang, Z., *Anal. Chem.* **2014,** *86* (19), 9376-80.
(52) Pan, N.; Rao, W.; Standke, S. J.; Yang, Z., *Anal. Chem.* **2016,** *88* (13), 6812-9.
(53) Sun, M.; Yang, Z.; Wawrik, B., *Front Plant Sci* **2018,** *9*, 571.
(54) Liu, R.; Zhang, G.; Yang, Z., *Chem. Commun.* **2018**.
(55) Sun, M.; Yang, Z., *Anal. Chem.* **2019,** *91* (3), 2384-2391.
(56) Standke, S. J.; Colby, D. H.; Bensen, R. C.; Burgett, A. W. G.; Yang, Z., *Anal. Chem.* **2019,** *91* (3), 1738-1742.
(57) Lanekoff, I.; Heath, B. S.; Liyu, A.; Thomas, M.; Carson, J. P.; Laskin, J., *Anal. Chem.* **2012,** *84* (19), 8351-8356.
(58) Thomas, M.; Heath, B. S.; Laskin, J.; Li, D. S.; Liu, E.; Hui, K.; Kuprat, A. P.; van Dam, K. K.; Carson, J. P., *2012 Annual International Conference of the Ieee Engineering in Medicine and Biology Society (Embc)* **2012**, 5545-5548.
(59) Li, H.; Hummon, A. B., *Anal. Chem.* **2011,** *83* (22), 8794-801.
(60) Ahlf Wheatcraft, D. R.; Liu, X.; Hummon, A. B., **2014,** (94), e52313.
(61) Race, A. M.; Styles, I. B.; Bunch, J., *J. Proteomics* **2012,** *75* (16), 5111-5112.
(62) Jaumot, J.; Gargallo, R.; de Juan, A.; Tauler, R., *Chemom. Intell. Lab. Syst.* **2005,** *76* (1), 101-110.

(63) Maechler, M., Rousseeuw, P., Struyf, A., Hubert, M., Hornik, K. , *R package version 2.0.7-1.* **2018**.
(64) Liaw, A.; Wiener, M., *R news* **2002,** *2* (3), 18-22.
(65) Tredan, O.; Galmarini, C. M.; Patel, K.; Tannock, I. F., *J. Natl. Cancer Inst.* **2007,** *99* (19), 1441-54.
(66) Mehta, G.; Hsiao, A. Y.; Ingram, M.; Luker, G. D.; Takayama, S., *J. Control. Release* **2012,** *164* (2), 192-204.
(67) Ward, J. P.; King, J. R., *Math. Biosci.* **2003,** *181* (2), 177-207.
(68) Bleicher, R. J.; Cabot, M. C., *Biochim. Biophys. Acta* **2002,** *1585* (2), 172-178.
(69) Testai, F. D.; Kilkus, J. P.; Berdyshev, E.; Gorshkova, I.; Natarajan, V.; Dawson, G., *J. Neurochem.* **2014,** *131* (4), 530-40.
(70) Alvarez-Perez, J.; Ballesteros, P.; Cerdan, S., *MAGMA* **2005,** *18* (6), 293-301.
(71) Smith, C. A.; O'Maille, G.; Want, E. J.; Qin, C.; Trauger, S. A.; Brandon, T. R.; Custodio, D. E.; Abagyan, R.; Siuzdak, G., *Ther. Drug Monit.* **2005,** *27* (6), 747-751.
(72) Lordan, R.; Tsoupras, A.; Zabetakis, I., *Molecules* **2017,** *22* (11).
(73) Cummings, B. S., *Biochem. Pharmacol.* **2007,** *74* (7), 949-59.
(74) Inazu, M., *Biopharm. Drug Dispos.* **2014,** *35* (8), 431-49.
(75) Jung, J. H.; Lee, M. Y.; Choi, D. Y.; Lee, J. W.; You, S.; Lee, K. Y.; Kim, J.; Kim, K. P., *Proteomics* **2015,** *15* (4), 824-35.
(76) Tamada, M.; Nagano, O.; Tateyama, S.; Ohmura, M.; Yae, T.; Ishimoto, T.; Sugihara, E.; Onishi, N.; Yamamoto, T.; Yanagawa, H.; Suematsu, M.; Saya, H., *Cancer Res.* **2012,** *72* (6), 1438-48.
(77) Wang, J.; Hong, F.; Xue, T.; Xu, P.; Zhai, Y., *Recent Advance In Colon Cancer* **2017,** *2*, 2-36.
(78) Moreau, A.; Vilarem, M. J.; Maurel, P.; Pascussi, J. M., *Mol. Pharm.* **2008,** *5* (1), 35-41.
(79) Zhou, Y.; Bollu, L. R.; Tozzi, F.; Ye, X.; Bhattacharya, R.; Gao, G.; Dupre, E.; Xia, L.; Lu, J.; Fan, F.; Bellister, S.; Ellis, L. M.; Weihua, Z., *Mol. Cancer Ther.* **2013,** *12* (12), 2782-91.
(80) Menendez, J. A.; Lupu, R., *Nat. Rev. Cancer* **2007,** *7* (10), 763-77.

Chapter4. Multimodal Imaging of Amyloid Plaques: Fusion of the Single-probe Mass Spectrometry Image and Fluorescence Microscopy Image

This project was a collaborative work that consists of the following authors: Tian, Xiang; Xie, Boer; Zou, Zhu; Jiao, Yun; Lin, Li-En; Chen, Chin-Lin; Hsu, Cheng-Chih.; Peng, Junmin; Yang, Zhibo

Boer Xie and Yun Jiao conducted the animal experiments.

Zhu Zou conducted some data analysis using Molecular Image Fusion software.

Li-En Lin, Chin-Lin Chen, and Cheng-Chih Hsu assisted the data preprocessing of using the Molecular Image Fusion software.

4.1. Abstract

Alzheimer's Disease (AD) is one of the most common neurodegenerative diseases. The formation of amyloid plaques by aggregated amyloid beta (Aβ) peptides is a primary event in AD pathology. Understanding the metabolomic features and related pathways are critical for studying plaque-related pathological events (e.g., cell death and neuron dysfunction). Mass spectrometry imaging (MSI), due to its high sensitivity and ability of obtaining spatial distribution of metabolites, has been applied to AD studies. However, limited studies of metabolites in amyloid plaques were performed due to the drawbacks of commonly used techniques such as matrix-assisted laser desorption/ionization (MALDI) MSI. In the current study, we obtained high spatial resolution (~17 μm) MS images of AD mouse brain using the Single-probe, a microscale sampling and ionization device, coupled to a mass spectrometer under ambient conditions. The adjacent slices were used to obtain a fluorescence microscopy images to locate amyloid plaques. MS image and fluorescence

microscopy image were fused to spatially correlate histological protein hallmarks with metabolomic features. The fused images produced significantly improved spatial resolution (~5 μm), allowing for the determination of fine structures in MS images and metabolomic biomarkers representing amyloid plaques.

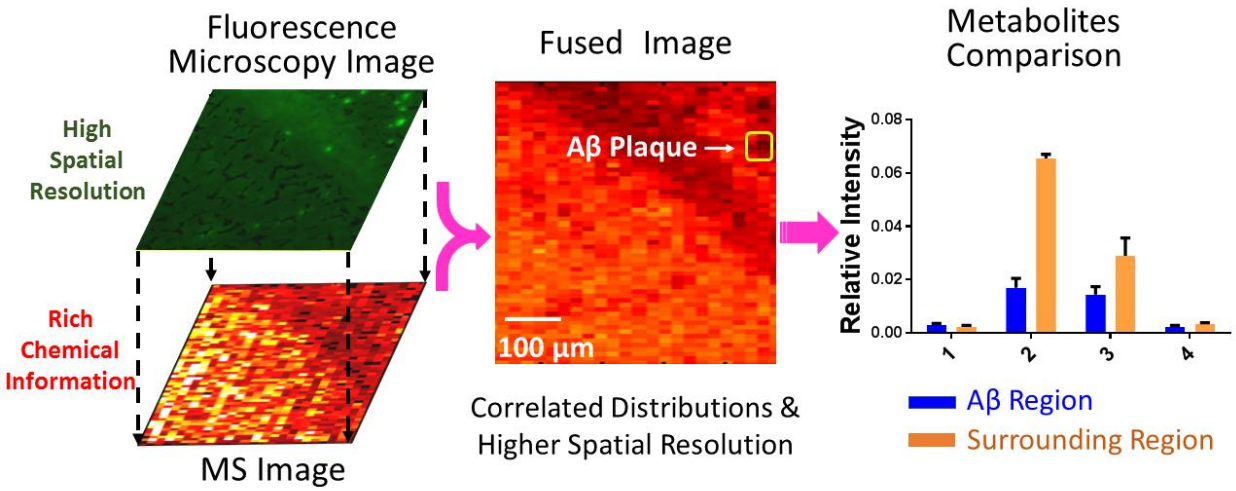


Figure 4-1. Graphic abstract.

4.2. Introduction

Alzheimer's disease (AD) affects more than 5.3 million Americans, and it is the most common neurodegenerative disease causing cognitive impairment.[1] There are two major pathologic features of AD: intracellular neurofibrillary tangles (NFT), which are composed of hyperphosphorylated

tau proteins, and extracellular plaques, which contain the aggregated β-amyloid peptide (Aβ).[2] Although the mechanisms of plaque formation are still unclear, the most commonly accepted hypothesis is that Aβ is secreted into the extracellular space and aggregated to form amyloid plaques.[3] Aβ plaques are believed toxic to neuron cells, both in vitro and in vivo,[4] causing the symptom of AD such as cognitive impairment, oxidative damage, and inflammation.[5, 6] In the past decades, a large number of studies focus on the Aβ plaques using genomics,[7] proteomics,[8] and metabolomics[9-12]. At gene-level, the gene of β-Amyloid precursor protein (APP) on chromosome 21 are highly expressed in Down's syndrome patients (trisomy 21)[13], and the duplication of the APP gene induces early onset of AD.[14] At protein-level, soluble oligomerized Aβ are regarded as the main cause of synaptic dysfunction instead of Aβ monomer or larger aggregates.[8] In addition, studying metabolomics, which directly reflects the status of cells[15], is essential to the research of AD. For example, certain metabolites, such as lipids[9], neurotransmitters,[10] fatty acids[11], and glucose[12], are highly correlated with AD pathology, and they may play important roles in the development of the disease. The metabolomics studies in AD can significantly benefit the understanding of physiopathology, drug discovery, and related metabolite pathways.

High resolution mass spectrometry (MS), with high sensitivity and wide ranges of molecular coverage, is one of the most important analytical methods to detect large numbers of metabolites from samples when coupling with separation technique such as liquid chromatography (LC), gas chromatography (GC), and capillary electrophoresis (CE).[16] However, because lysates need to be prepared from homogenized samples, the spatial distribution of metabolites, which is critical to understand the complex biological process and the pathophysiology, is inevitably lost.[17, 18] Mass spectrometry imaging (MSI) is a powerful technique to achieve the spatial information of metabolites on biological sample surface.[18] Different MSI techniques, such as matrix-assisted laser

desorption/ionization (MALDI),[19] desorption electrospray ionization (DESI),[20] secondary ion MS (SIMS),[21] and nanospray desorption electrospray ionization (nano DESI),[22] have been applied to study the metabolomics. Among these developed MS imaging methods, MALDI and SIMS are non-ambient techniques, which have excellent spatial resolution and sensitivity, whereas DESI and nano-DESI belong to ambient sampling and ionization techniques, which require less or no sample preparation. Particularly, MALDI has been implemented to analyze proteins, peptides, and metabolites in AD studies.[9, 23-28] However, due to the complex background of MALDI mass spectra at low-mass range (< 500 Da),[29] this technique is generally less effective to measure small but important molecules, including neurotransmitters[30], fatty acids,[11] and phospholipids.[24] In addition to MS imaging techniques, laser capture microdissection (LCM) coupled with MS has been used to isolate the region of interest in tissue and to determine the metabolites profiles.[27]

Here, we utilized the Single-probe MSI technique to study the metabolites in both Aβ plaques and their surrounding regions. The Single-probe is a micro-scale sampling and ionization device that can be directly coupled with mass spectrometer to conduct research in multiple fields, including live single cell analysis[31-38], extracellular metabolites in spheroids,[35] and MSI of biological tissues.[39-42] Because the Single-probe MSI technique is based on micro-liquid extraction without using matrix (e.g., molecules involved in MALDI-MSI technique), relatively clean background, particularly at low-mass range, facilitates MSI studies of small molecules such as metabolites. To spatially correlate the metabolites with Aβ plaques, image fusion method developed by Caprioli *et al* for MALDI-TOF MSI[43], which has recently been implemented to DESI and nano-DESI MSI,[44] was applied to fuse the images of the Single-probe MSI and fluorescence microscopy in this study. This combined method takes the advantages of both techniques, i.e., rich chemical information from MSI and high spatial resolution from fluorescence microscopy image, to provide the

metabolomic information correlated to Aβ plaques, and improve our understandings of the pathways and functions of metabolites related to AD.

4.3. Experimental section

4.3.1. Chemicals and materials

Chemicals used in the experiments include methanol, water (Sigma-Aldrich, St. Louis, MO), ethanol (Pharmco-AAPPER, Shelbyville, KY), xylene (Sigma-Aldrich, St. Louis, MO), and thioflavin S (Chem Cruz, Dallas, TX). Materials needed to fabricate the Single-probe include the fused capillary (O.D. 105 µm; I.D. 40 µm, Polymicro Technologies, Phoenix, AZ, USA) and dual-bore quartz tubing (O.D. 500 µm; I.D. 127 µm, Friedrich & Dimmock, Inc., Millville, NJ, USA).

The detailed fabrication protocol of the Single-probe was described in our previous work[32], and only the outlined procedures are provided here. The Single-probe has three components: a laser-pulled dual-bore quartz needle, a fused silica capillary (solvent-providing capillary), and a nano-electrospray (nano-ESI) emitter. A Sutter P-2000 laser micropipette puller (Sutter Instrument, Novato, CA, USA) is used to prepare the dual-bore needle and nano-ESI emitter. A Single-probe is fabricated by embedding within a dual-bore quartz needle with one fused silica capillary and one nano-ESI emitter.

4.3.2. The Single-probe MSI setup

The experiment setup is largely adopted from our previous MS single cell and MSI study[45] (Figure S4-1). To precisely control the movement of tissue slice, the sample was attached to a XYZ-translational stage system (CONEX-MFACC, Newport Co., Irvine, CA, USA) controlled using a LabView software package[46]. A digital microscope was placed next to the Single-probe to adjust the distance between the Single-probe tip and tissue slice surface, and to monitor the sampling

process. MS spectra were collected using a Thermo LTQ XL mass spectrometer (Thermo Scientific, Waltham, MA, USA) with the following parameters: mass resolution 60,000 (m/Δm), 4.5 kV ionization voltage (both positive and negative ion mode), 1 microscan, 100 ms max injection time, and AGC on (5 X 105). The sampling solvent (i.e., 85% methanol/15%water (v/v)) was continuously delivered (flowrate 200 nL/min) by a syringe pump (PHD ULTRA, Harvard Apparatus, Holliston, MA, USA). Ions of interest were identified using tandem MS (MS2) directly on tissue slices, and results were compared with database METLIN. The MS images of selected ions were generated using MSI QuickView software[46]. More experimental details related to the current study were provided in the Supporting Information.

4.3.3. Animal sample preparation

Animals used in this study were treated in accordance with the NIH Guide for the Care and Use of Laboratory Animals, all protocols were approved by the St Jude Children's Research Hospital ACUC under IACUC protocol. Experiments were carried out in accordance with The Code of Ethics of the World Medical Association (Declaration of Helsinki) for animal experiments.

Two pairs of control (wild type) and 5xFAD mice (Tg (APPSwFlLon, PSEN1*M146L*L286V) 6799Vas) obtained from Jackson Laboratory (Bar Harbor, ME, USA) were used in this study. Mice were housed in a temperature and humidity-controlled room with a 12:12 h reversed light/dark cycle. Food and water were available ad libitum. All mice used in this study were sacrificed at 10–12 months old with weight between 23–25 g by cervical dislocation and decapitation. Brains were rapidly removed from the calvarias, and immediately frozen on dry ice, then stored in 2 ml Eppendorf tubes at –80°C until further experiments. Approximately 5 mins per mouse were required for the brain collection.

Mouse brain was embedded in 10% HPMC ((hydroxypropyl)methyl cellulose, Sigma-Aldrich, St. Louis, MO), frozen on dry ice, sectioned into slices (~15 μm in thickness) using a cryotome (American Optical 845 Cryo-cut Mictorome, Southbridge, MA, USA) at –15°C, and attached onto a microscope slide (VWR, Radnor, PA, USA). The mouse brain slices were dried in air and stored at –80°C before usage. The bright-field optical images of slices were taken using a PathScan Enabler IV histology slide scanner (Meyer Instruments, Inc. Houston, TX, USA) prior to MSI experiment. One tissue slide of each mouse brain sample was reported in each MSI study.

4.3.4. Fluorescence microscopy image.

The mouse brain slice adjacent to the one used for MSI measurement of each brain sample was thawed at room temperature for 20 mins before usage. The slices were fixed (in 75% ethanol for 1 min), stained (in 1% thioflavin-S aqueous solution for 1 min), and differentiated (in 75% ethanol for 1 min) to remove excess fluorochrome. After that, the slides were dehydrated in a series of graded ethanol (95% ethanol twice, 100% ethanol twice, and 3-5 sec for each step), cleared (in xylene for 5 mins), and dried in the air. The stained slices were kept in a dark environment, and fluorescence microscope images were immediately taken using a Nikon Eclipse Ti-S fluorescence microscope (Melville, NY, USA).

4.3.5. Data pre-processing.

Before conducting the image fusion, the MSI data (.raw) need to be converted into an appropriate format that can be utilized in the Molecular Image Fusion software. The data pre-processing method is adopted from our previous publication,[39] and a brief outline is provided here. First, the MSI data format was converted from .raw to .mzML using MSConvert (a tool in ProteoWizard), and further converted into .imzML format using imzML Converter[47]. Second, a built-in function of MATLAB, Bioinformatics Toolbox, was implemented to achieve data pre-processing,

including smoothing, noise removal, peak alignment, peak picking, and insensitivity normalization. Third, a data matrix was generated to represent the MS imaging data. For example, the data matrix of the positive MS image is composed of 2115 X 996, i.e., 2115 pixels (141(scans/line) X 15 lines) with 996 aligned common ions. The number of pixels of each MS image depends on the size and the spatial resolution of the MS image. The details of data pre-processing are provided in the Supporting Information.

4.3.6. Image fusion.

The image fusion was conducted using Molecular Image Fusion software[43, 44]. Six files, including microscope image data, microscope image information (pixel size, pixel number, and spatial resolution), MS image data, MS image information (pixel size, pixel number, and spatial resolution), fusion parameters, and registration information, were needed to conduct the image fusion. Among them, the MSI data matrix was achieved using data pre-processing as mentioned above, whereas the microscopy image data matrix and registration information were generated using in-house developed MATLAB script (details were provided in the Supporting Information).

4.3.7. Averaged MS spectra and *t*-test.

Based on the fused image, the pixels representing Aβ plaques and their surrounding regions were manually selected from MSI dataset, and averaged into two mass spectra, respectively, using in-house developed R script. To obtain the metabolites with significant differences between two regions, a t-test was carried out using GraphPad.

4.4. Results and discussion

4.4.1. Fluorescence microscopy imaging of mouse brain

As a common fluorescence dye binding to Aβ plaques,[48] Thioflavin S was utilized to stain brain slices obtained from 5xFAD mice and in the littermate control group. Although Thioflavin S labels

both amyloid plaques and tau tangles, they can be readily distinguished based on their morphologies. Aβ plaques are extracellular, and they are larger with round structures; tangles are intraneuronal, and they are smaller with flame shape.[49] The fluorescence microscopy image shows the Aβ aggregated areas are clearly illustrated as bright green dots on the slice of 5xFAD mouse brain (Figure 4-2A), but they are not observed in the control sample (Figure 4-2B). Previous studies indicate that the plaque size generally ranges from 10 μm to 80 μm with an average size of around 50−60 μm.[50] In our studies, most plaques are smaller than 50 μm, indicating that high-spatial resolution MSI techniques are needed for conducting experiments

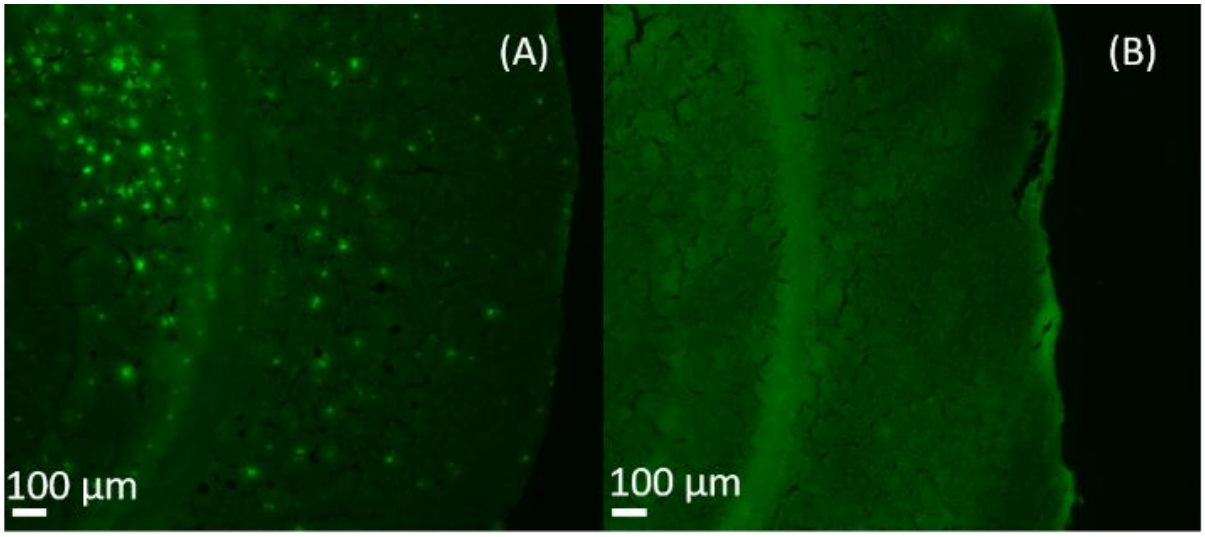


Figure 4-2. Fluorescence microscopy images of brain slices of (A) 5xFAD and (B) control mice.

4.4.2. MSI of mouse brain and image fusion

Previous studies indicate that Aβ plaques have relatively higher abundances in hippocampus because amyloid precursor protein (APP) are transported to the nerve terminal in this region and further processed into Aβ.[51] Thus, we conducted MSI measurements of the selected small areas (e.g., 1 mm X 0.3 mm) on brain slice containing hippocampus. For example, a slice of whole

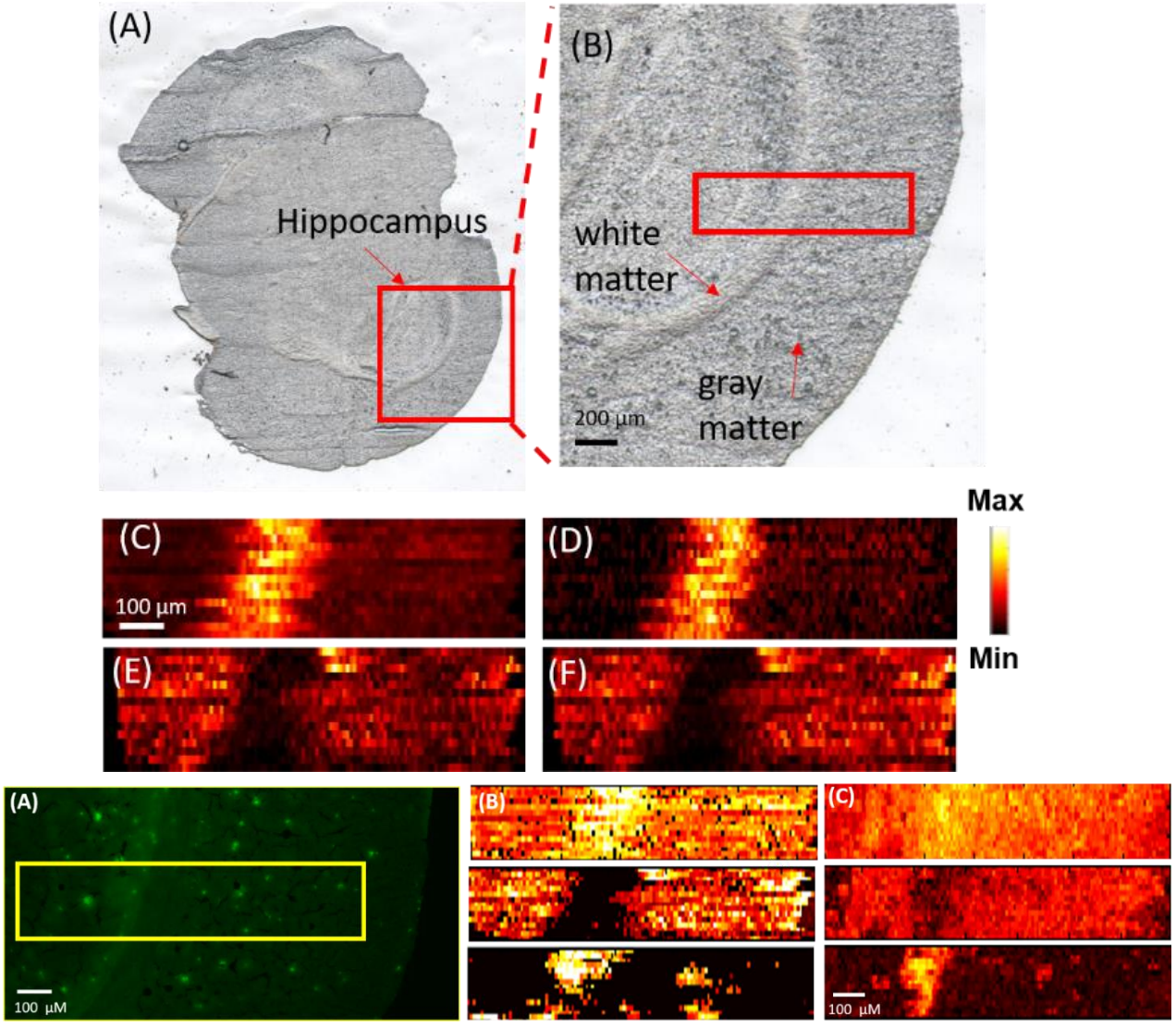
mouse brain with AD (Figure 4-3A) containing the hippocampus, as shown in a zoomed-in picture (Figure 4-3B), was selected for the MSI experiments. The MSI experiments were conducted in both positive and negative ion modes to detect broader ranges of molecules. In the positive ion mode, a large number of lysophospholipids were observed, whereas, in the negative ion mode, organic acids were primarily detected. The MS images of the selected metabolites were then constructed (Figure 4-3) to demonstrate their spatial distributions. For example, $[PC(36:1)+H]^+$ (*m/z* 788.6137) and $[PC(38:1)+H]^+$ (*m/z* 814.6292) were mainly distributed in the white matter, whereas $[PC(38:4)+K]^+$ (*m/z* 848.5542) and $[PC(38:6)+K]^+$ (*m/z* 844.5221) were primarily observed in the gray area. Similar results of a number of lipids were reported in previous studies.[52, 53] For example, Antonio Veloso *et al.*[52] showed PC(38:4) and PC(38:6) are more abundant in the grey matter, whereas PC (36:1) has higher abundance in the white matter. MSI experiments of mice brain samples in the control group were also conducted (positive ion mode), and MS images of representative metabolites in the white and grey matter were provided in Figure S4-10. All

metabolites were identified using MS^2 from tissue slice, and results were compared with METLIN (Figure S2-11).

Figure 4-3. Optical and MS images of FAD mouse brain.

Optical image of (A) a coronal section of the mouse brain (B) the zoomed-in region containing the area for MSI measurement (enclosed in the red rectangle). MS images of (C) $[PC(36:1) + H]^+$ and (D) $[PC(38:1) + H]^+$) representing metabolites primarily distributed in the white matter. MS images of (E) $[PC(38:4) + K]^+$ and (F) $[PC(38:6) + K]^+$) representing metabolites primarily distributed in the gray matter. All metabolites were identified using MS^2 from tissue slice, and results were compared with METLIN (Table S4-4).

To localize the area of the Aβ plaques, the adjacent slice was stained using thioflavin S, and the



fluorescence microscopy image was taken to match the region measured in the MSI experiment

(Figure 4-4A). Molecular Image Fusion software was employed to fuse the MS image and the corresponding fluorescence microscopy image, and the spatial resolution of MS. images (~17 μm), which was determined based previously established method[54], has been increased to ~5 μm in the fused image (Figure 4-4C) (characterization of the spatial resolution is provided in Figure S4-8). For example, the fused images of $[PC(34:1) + H]^+$ (*m/z* 760.5851) and $[PC(38:8) + H]^+$ (*m/z* 834.5983) become sharper compared with their original MS images. High reconstruction rates (76.3-78.4%, Figure S4-2) indicate the image fusion process is reliable and satisfactory (i.e., reconstruction rate > 75% is regarded as satisfied fusion[43]). Particularly, the fused images correlate histological protein hallmarks with metabolomic features. We obtained a number of ions (e.g., $[LPC(18:0) + H]^+$ (*m/z* 524.3693)) accumulated on Aβ plaques (i.e., bright spots on the fluorescence microscopy images).

Figure 4-4. Fusion of fluorescence microscopy image and MS image.

(A) Fluorescence microscopy image of a 5XFAD mouse brain slice stained using Thioflavin S. (B) Original MS images of metabolites ($[PC(34:1) + H]^+$ (760.5851) (upper), $[PC(38:6) + H]^+$ (844.5218) (middle), and $[LPC(18:0) + H]^+$ (*m/z* 524.3693) (lower)) and (C) their fused images. All metabolites were identified using MS^2 from tissue slice, and results were compared with METLIN (Figure S4-3).

4.4.3. Data analysis

To achieve a comprehensive analysis of MSI data representing Aβ plaques and their surrounding regions, a statistical data analysis was carried out. By comparing the fused image with the fluorescence microscope image (Figure 4-4A and 4-5A), multiple Aβ plaques were identified, and three of them were selected to determine the metabolites with significantly different abundances between Aβ plaques and their surrounding regions. Based on the fused image, pixels representing different regions were manually selected to generate the averaged mass spectra. For example, in region 1, the pixels encompassed by the blue rectangle represent the Aβ plaque, whereas those between the yellow and blue rectangles represent the surrounding region of the Aβ plaque (details of pixel numbers selection are provided in the Table S4-1). An averaged mass spectrum of each region (Figure 4-5B) was generated using the corresponding MSI data, and ion intensities were normalized to the TIC (total ion current) of the average mass spectrum. The *t*-test was conducted to determine the ions with significantly different abundances between Aβ plaques and their surrounding regions.

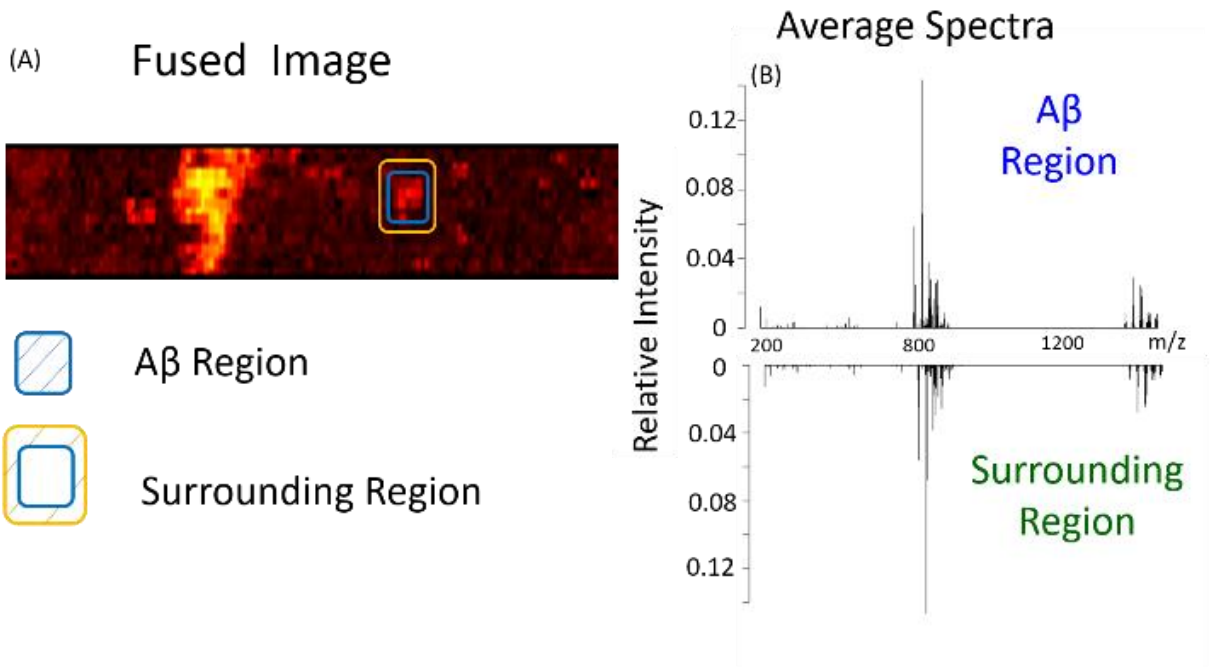


Figure 4-5. Pixel selection and average spectra (MS positive ion mode).

(A) Fused image and the pixels representing Aβ plaques and its surrounding region. (B) The averaged mass spectra of an Aβ plaque and its surrounding region.

Except for the example shown in Figure 4-5, two additional plaques were analyzed (Figure S4-6). Our results indicate that metabolites accumulated in Aβ plaques (Figure 4-6) are primarily lysophospholipids such as LPC (lysophosphatidylcholine) and LPG (lysophosphatidylglycerol). For example, the relative intensities of $[LPC\ (16:0) + H]^+$ and $LPC\ [(18:1) + H]^+$ were about 14 and 11 times higher in Aβ plaques than the surrounding regions, respectively. Previous studies demonstrate that lysophospholipids are mainly generated along with fatty acids through hydrolyzation of membrane phospholipids[55]. As bioactive lipid mediators, lysophospholipids play a variety of roles such as proliferation, differentiation, survival, migration, adhesion, invasion, and morphogenesis[56]. However, lysophospholipids may potentially cause cell membrane disruption and ultimately lead to cell lysis due to their toxic detergent-like properties[57]. Thus, the change of lysophospholipids level may deter fundamental functions of cell membrane such as ion transportation[58], molecule secretion,[59] and membrane related signal reception[60]. According to previous studies of immunostaining of amyloid peptide in 5xFAD mice models, the accumulation of LPCs in Aβ plaque is likely due to the hyperactivity of enzyme PLA_2 (an enzyme directly converts phospholipid to lysophospholipids and fatty acids)[61]. *Kaya et. al*[62] also demonstrated other lysophospholipids (e.g., LPC, lysophosphatidic acid (LPA), and lysophosphatidylethanolamine (LPE)) are aggregated in the plaque of AD mice brain using multimode MALDI-IMS technique. Furthermore, decreasing the overexpressed PLA_2 is regarded as one of the goals in AD treatment. For example, PLA_2 inhibitor has been employed to treat

Alzheimer's disease.[63] In fact, the reduction of PLA_2 can ameliorate the cognitive deficits in a mouse model of AD[64].

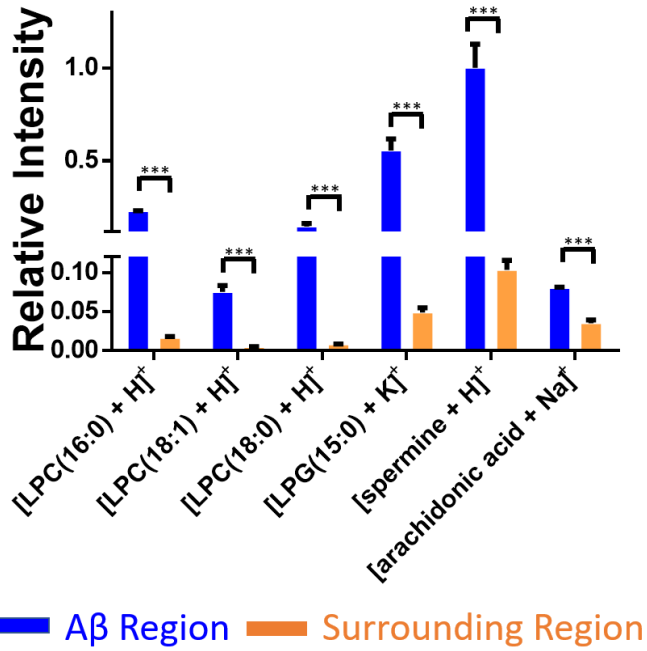


Figure 4-6. Representative metabolites (MS positive ion mode) with significantly different abundances between Aβ plaques and their surrounding regions.

Results were obtained from three Aβ plaques. The error bar indicates the standard deviation of the relative intensities obtained from the selected pixels. All metabolites were identified using MS^2 from tissue slice, and results were compared with METLIN (Figure S4-3). (From *t*-test: ***, < 0.001)

In addition to LPCs, we discovered the abundances of a number of other small molecules (e.g., spermine and arachidonic acid) are significantly different between plaques and their surrounding regions (Table S4-2). Spermine, which has the largest number of amine group among three naturally occurring polyamines[65] (i.e., spermine, spermidine, and putrescine), plays multiple important roles such as controlling gene express, stabilizing chromatin,[66] and protecting replicating DNA from being damaged by singlet oxygen[67]. The upregulation of spermine in plaques can be understood as a protective mechanism: Aβ is toxic to neurons through a free radical dependent oxidative mechanism,[68] whereas increased polyamines can act as radical scavengers to neutralize

free radicals generated by the Aβ and protect neurons.[69, 70] Another small molecule accumulated in Aβ plaque is arachidonic acid (AA), which is an essential omega-6 fatty acid critical for synaptic signaling, long-term potentiation, learning, and memory[71]. Although AA can be generated from other sources, such as hydrolysis of fatty acid amide hydrolase and monoacylglycerol lipase, the level of free AA in brain is primarily determined by a specific form of PLA_2, GIVA-PLA_2[58]. According to previous studies, the stimulation of Aβ results in an increased level of GIVA-PLA2, and further alters the amount of AA[72]. For example, upregulation of AA was previously reported in the APP (amyloid precursor protein) transferred PC12 cells[73] and in the plaque region of mouse using the PET (positron emission tomography)[74].

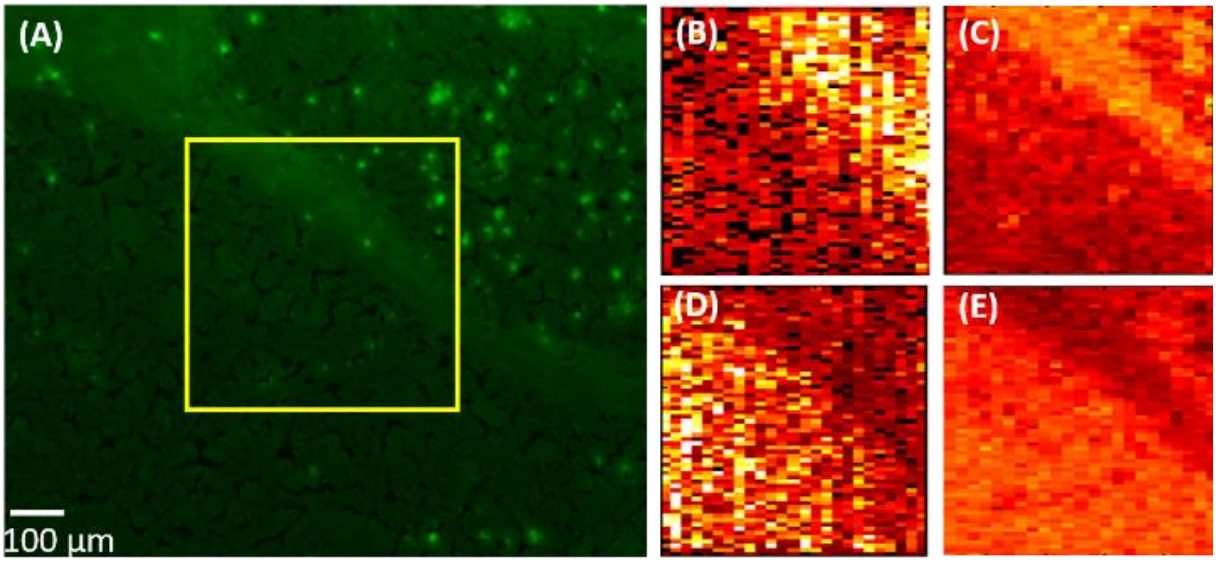


Figure 4-7. Image fusion of fluorescence microscopy and MS images (MS negative ion mode).

(A) Fluorescence microscopy image of 5xFAD mouse brain. (B) MS image of $[PA(O\text{-}32:0)\text{-}2H]^{2-}$ and (C) its fused image. (D) MS image of dodecenoic acid $[M+K-2H]^-$ and (E) its fused image.

The MSI experiments were also conducted in the negative ion mode and the area is 0.5 mm X 0.5 mm. Fluorescence microscopy image (Figure 4-7A) and MS image (Figure 4-7B and 4-7C) were

obtained from AD mouse brain slice, and fused using procedures as described above. Similar to the results acquired from the positive ion mode, metabolites exhibit different distribution features. For example, ([PA(O-32:0) - 2H]$^{2-}$) and dodecenoic acid ([M+K-2H]$^{-}$) primarily present in the white matter and gray matter, respectively (Figure 4-7D and 4-7E). Comparing the fused image with fluorescence microscopy image, multiple Aβ regions were located, and two of them were selected to extract pixels representing Aβ plaques and their surrounding regions (Figure 4-8 and Figure S4-7). A number of metabolites (e.g., malic acid, glutamine, aspartic acid, and docosahexaenoic acid (DHA)) possess significantly different abundances between Aβ plaques and their surrounding regions (Table S4-3). For example, the relative ion intensity of glutamine is two times lower in plaque region compared with its surrounding (Figure 4-8B). Glutamine, which is an important amino acid for multiple processes such as promoting and maintaining cell function.[75] Glutamine is mainly produced by glutamine synthetase (GS), which converts glutamate and ammonium to glutamine.[76] It was reported that intense oxidative stress is one the important symptoms of AD brain[77, 78] and closely related to the Aβ[79, 80]. The oxidative stress leads to alterations of large numbers of molecular processes such as lipid peroxidation[81], DNA oxidation[82], and protein dysfunction[83]. Particularly, the activity of GS is significantly decreased due to the oxidative stress[84, 85] which may further resulting the deactivation of GS in Aβ region and lower the level of glutamine.

The relative intensity of aspartic acid was about four times lower in Aβ plaques compared with the surrounding regions. A previous study revealed that the metabolic profiles of a number of amino acids and relevant molecules, including aspartic acid, alanine, serine, glycine, and N-acetyl aspartic acid, were significantly altered in the mice during aging[86, 87]. For example, the level of aspartic acid is decreased in AD mice compared with control mice brain[87]; this trend is similar to

our results. The potential mechanism is likely related to the dysfunction of mitochondrial and hypometabolism in AD pathogenesis[88, 89]. In addition, our results indicate that the relative abundance of malic acid is about 1.5 times higher in plaque than in surrounding region; the upregulation of malic acid in plaque is also reported in previous studies[90]. As an important intermediate molecule, malic acid is involved in Krebs cycle mainly occurring in mitochondrial[91]. Thus, the alteration of malic acid is likely due to mitochondrial dysfunction and decreased function of key enzymes in AD[92]. In addition, the abundances of other metabolites, such as xanthine and DHA (Table S4-3), are also significantly different between the Aβ and surrounding regions, whereas the relevant mechanisms need to be further studied.

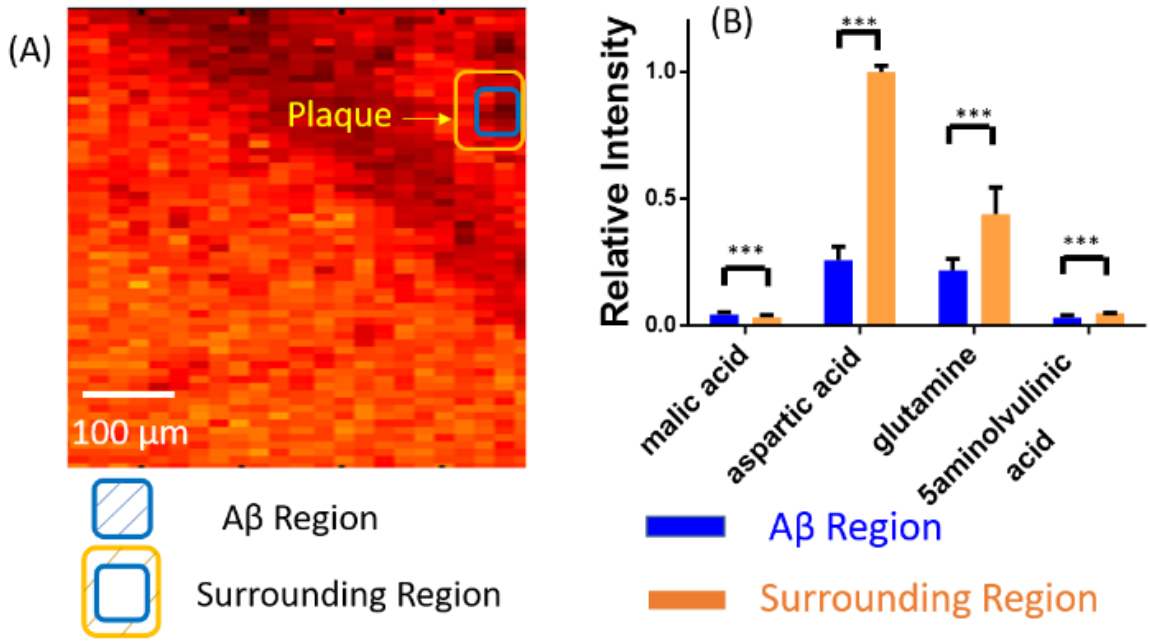


Figure 4-8. Pixel selection and ion abundance comparison (MS negative ion mode).

(A) Pixels representing Aβ plaques and their surrounding areas in the fused images. (B) Representative metabolites possessing significantly different abundances between Aβ plaques and their surrounding regions (From *t*-test: ***, < 0.001). All metabolites were identified using MS^2 from tissue slice, and results were compared with METLIN (Figure S4-4).

4.5. Conclusion

In this study, we utilized the Single-probe MSI technique to achieve high-resolution, ambient MS images of mouse brain with Alzheimer Disease in both negative and positive ion modes. Fluorescence microscopy images were obtained using the adjacent mouse brain slices to locate Aβ plaque positions. Image fusion was carried out to integrate fluorescence microscopy images (with higher spatial resolution) and MS images (with rich chemical information). Particularly, the fused images can provide significantly improved spatial resolution of MS images, allowing for the correlation of histological hallmarks of Aβ plaques and their surrounding regions to their relevant metabolomic profiles. Our results indicate that a number of metabolites (e.g., lysophospholipids, spermine, arachidonic acid, malic acid, glutamine, and aspartic acid) are abnormally expressed in Aβ plaques, and they are very likely related to the development of AD. Our method can be potentially used in other studies to illustrate the spatially resolved correlations between metabolites and potential biomarkers of diseases. The fully established methods may benefit the physiopathology, therapeutic resistance, and biomarker discovery.

The material in chapter 3 is adapted from Tian, X.; Xie, B.; Zou, Z.; Jiao, Y.; Lin, L. E.; Chen, C. L.; Hsu, C. C.; Peng, J.; Yang, Z., *Anal. Chem.* **2019,** *91* (20), 12882-12889. The copyright permission is obtained from ACS.

References

(1) Kumar, A.; Singh, A.; Ekavali, *Pharmacol. Rep.* **2015,** *67* (2), 195-203.

(2) Gamblin, T. C.; Chen, F.; Zambrano, A.; Abraha, A.; Lagalwar, S.; Guillozet, A. L.; Lu, M.; Fu, Y.; Garcia-Sierra, F.; LaPointe, N.; Miller, R.; Berry, R. W.; Binder, L. I.; Cryns, V. L., *Proceedings of the National Academy of Sciences* **2003,** *100* (17), 10032-10037.

(3) Takahashi, R. H.; Nagao, T.; Gouras, G. K., *Pathol. Int.* **2017,** *67* (4), 185-193.

(4) Pike, C. J.; Walencewicz, A. J.; Glabe, C. G.; Cotman, C. W., *Brain Res.* **1991,** *563* (1), 311-314.

(5) Takeda, A.; Aliev, G.; Atwood, C. S.; Perry, G.; Smith, M. A.; Petersen, R. B.; Nunomura, A.; Balraj, E. K.; Hirai, K.; Jones, P. K.; Chiba, S.; Ghanbari, H.; Wataya, T.; Shimohama, S., *J. Neuropathol. Exp. Neurol.* **2001,** *60* (8), 759-767.

(6) Akiyama, H.; Barger, S.; Barnum, S.; Bradt, B.; Bauer, J.; Cole, G. M.; Cooper, N. R.; Eikelenboom, P.; Emmerling, M.; Fiebich, B. L.; Finch, C. E.; Frautschy, S.; Griffin, W. S. T.; Hampel, H.; Hull, M.; Landreth, G.; Lue, L. F.; Mrak, R.; Mackenzie, I. R.; McGeer, P. L.; O'Banion, M. K.; Pachter, J.; Pasinetti, G.; Plata–Salaman, C.; Rogers, J.; Rydel, R.; Shen, Y.; Streit, W.; Strohmeyer, R.; Tooyoma, I.; Van Muiswinkel, F. L.; Veerhuis, R.; Walker, D.; Webster, S.; Wegrzyniak, B.; Wenk, G.; Wyss–Coray, T., *Neurobiol. Aging* **2000,** *21* (3), 383-421.

(7) Selkoe, D. J., *Neurol. Clin.* **2000,** *18* (4), 903-921.

(8) Walsh, D. M.; Townsend, M.; Podlisny, M. B.; Shankar, G. M.; Fadeeva, J. V.; El Agnaf, O.; Hartley, D. M.; Selkoe, D. J., *J. Neurosci.* **2005,** *25* (10), 2455-62.

(9) Michno, W.; Kaya, I.; Nystrom, S.; Guerard, L.; Nilsson, K. P. R.; Hammarstrom, P.; Blennow, K.; Zetterberg, H.; Hanrieder, J., *Anal. Chem.* **2018,** *90* (13), 8130-8138.

(10) Nitsch, R. M., *Neurodegeneration* **1996,** *5* (4), 477-482.

(11) Söderberg, M.; Edlund, C.; Kristensson, K.; Dallner, G., *Lipids* **1991,** *26* (6), 421.

(12) Hoyer, S., *Eur. J. Pharmacol.* **2004,** *490* (1-3), 115-25.

(13) Motte, J.; Williams, R. S., *Acta Neuropathol.* **1989,** *77* (5), 535-546.

(14) Prasher, V. P.; Farrer, M. J.; Kessling, A. M.; Fisher, E. M. C.; West, R. J.; Barber, P. C.; Butler, A. C., *Ann. Neurol.* **1998,** *43* (3), 380-383.

(15) Minchinton, A. I.; Tannock, I. F., *Nat. Rev. Cancer* **2006,** *6* (8), 583-92.

(16) Major, H. J.; Williams, R.; Wilson, A. J.; Wilson, I. D., *Rapid Commun. Mass Spectrom.* **2006,** *20* (22), 3295-302.

(17) Lee, D. Y.; Bowen, B. P.; Northen, T. R., *Biotechniques* **2010,** *49* (2), 557-65.

(18) Miura, D.; Fujimura, Y.; Wariishi, H., *J. Proteomics* **2012,** *75* (16), 5052-60.

(19) Cornett, D. S.; Reyzer, M. L.; Chaurand, P.; Caprioli, R. M., *Nat. Meth.* **2007,** *4*, 828.

(20) Chen, H.; Pan, Z.; Talaty, N.; Raftery, D.; Cooks, R. G., *Rapid Commun. Mass Spectrom.* **2006,** *20* (10), 1577-84.

(21) Passarelli, M. K.; Winograd, N., *Biochim. Biophys. Acta* **2011,** *1811* (11), 976-90.

(22) Roach, P. J.; Laskin, J.; Laskin, A., *Analyst* **2010,** *135* (9), 2233-6.

(23) Carlred, L.; Michno, W.; Kaya, I.; Sjovall, P.; Syvanen, S.; Hanrieder, J., *J. Neurochem.* **2016,** *138* (3), 469-78.

(24) Kaya, I.; Brinet, D.; Michno, W.; Baskurt, M.; Zetterberg, H.; Blenow, K.; Hanrieder, J., *ACS Chem. Neurosci.* **2017,** *8* (12), 2778-2790.

(25) Stoeckli, M.; Staab, D.; Staufenbiel, M.; Wiederhold, K.-H.; Signor, L., *Anal. Biochem.* **2002,** *311* (1), 33-39.

(26) Strnad, S.; Prazienkova, V.; Sykora, D.; Cvacka, J.; Maletinska, L.; Popelova, A.; Vrkoslav, V., *Talanta* **2019,** *201*, 364-372.

(27) Xu, B. J., *Proteomics Clin. Appl.* **2010,** *4* (2), 116-23.

(28) Kaya, I.; Michno, W.; Brinet, D.; Iacone, Y.; Zanni, G.; Blennow, K.; Zetterberg, H.; Hanrieder, J., *Anal. Chem.* **2017,** *89* (8), 4685-4694.
(29) Knochenmuss, R.; Dubois, F.; Dale, M. J.; Zenobi, R., *Rapid Commun. Mass Spectrom.* **1996,** *10* (8), 871-877.
(30) Francis, P. T., *CNS Spectrums* **2014,** *10* (S18), 6-9.
(31) Pan, N.; Rao, W.; Standke, S. J.; Yang, Z., *Anal. Chem.* **2016,** *88* (13), 6812-9.
(32) Pan, N.; Rao, W.; Kothapalli, N. R.; Liu, R.; Burgett, A. W.; Yang, Z., *Anal. Chem.* **2014,** *86* (19), 9376-80.
(33) Sun, M.; Yang, Z.; Wawrik, B., *Front Plant Sci* **2018,** *9*, 571.
(34) Sun, M.; Yang, Z., *Anal. Chem.* **2019,** *91* (3), 2384-2391.
(35) Sun, M.; Tian, X.; Yang, Z., *Anal. Chem.* **2017,** *89* (17), 9069-9076.
(36) Standke, S. J.; Colby, D. H.; Bensen, R. C.; Burgett, A. W. G.; Yang, Z., *Anal. Chem.* **2019,** *91* (3), 1738-1742.
(37) Liu, R.; Zhang, G.; Yang, Z., *Chem. Commun.* **2018**.
(38) Zhu, Y.; Liu, R.; Yang, Z., *Anal. Chim. Acta* **2019**.
(39) Tian, X.; Zhang, G.; Shao, Y.; Yang, Z., *Anal. Chim. Acta* **2018,** *1037*, 211-219.
(40) Rao, W.; Pan, N.; Yang, Z., *J. Vis. Exp.* **2016,** (112).
(41) Rao, W.; Pan, N.; Tian, X.; Yang, Z., *J. Am. Soc. Mass Spectrom.* **2016,** *27* (1), 124-34.
(42) Tian, X.; Zhang, G.; Zou, Z.; Yang, Z., *Anal. Chem.* **2019,** *91* (9), 5802-5809.
(43) Van de Plas, R.; Yang, J.; Spraggins, J.; Caprioli, R. M., *Nat. Methods* **2015,** *12* (4), 366-72.
(44) Chen, C.-L.; Lin, L.-E.; Huang, Y.-C.; Chung, H.-H.; Lin, C.-W.; Chen, K.-C.; Peng, Y.-J.; Ding, S.-T.; Shen, T.-L.; Hsu, C.-C., *bioRxiv* **2019**, 657494.
(45) Rao, W.; Pan, N.; Yang, Z. B., *J. Am. Soc. Mass Spectrom.* **2015,** *26* (6), 986-993.
(46) Lanekoff, I.; Heath, B. S.; Liyu, A.; Thomas, M.; Carson, J. P.; Laskin, J., *Anal. Chem.* **2012,** *84* (19), 8351-8356.
(47) Race, A. M.; Styles, I. B.; Bunch, J., *J. Proteomics* **2012,** *75* (16), 5111-5112.
(48) Urbanc, B.; Cruz, L.; Le, R.; Sanders, J.; Ashe, K. H.; Duff, K.; Stanley, H. E.; Irizarry, M. C.; Hyman, B. T., *Proceedings of the National Academy of Sciences* **2002,** *99* (22), 13990.
(49) Liao, L.; Cheng, D.; Wang, J.; Duong, D. M.; Losik, T. G.; Gearing, M.; Rees, H. D.; Lah, J. J.; Levey, A. I.; Peng, J., *J. Biol. Chem.* **2004,** *279* (35), 37061-8.
(50) Sheikh, A. M.; Nagai, A., *FEBS J.* **2011,** *278* (4), 634-42.
(51) Buxbaum, J. D.; Thinakaran, G.; Koliatsos, V.; O'Callahan, J.; Slunt, H. H.; Price, D. L.; Sisodia, S. S., *The Journal of Neuroscience* **1998,** *18* (23), 9629.
(52) Veloso, A.; Fernandez, R.; Astigarraga, E.; Barreda-Gomez, G.; Manuel, I.; Giralt, M. T.; Ferrer, I.; Ochoa, B.; Rodriguez-Puertas, R.; Fernandez, J. A., *Anal. Bioanal. Chem.* **2011,** *401* (1), 89-101.
(53) Yuki, D.; Sugiura, Y.; Zaima, N.; Akatsu, H.; Takei, S.; Yao, I.; Maesako, M.; Kinoshita, A.; Yamamoto, T.; Kon, R.; Sugiyama, K.; Setou, M., *Sci. Rep.* **2014,** *4*, 7130.
(54) Luxembourg, S. L.; Mize, T. H.; McDonnell, L. A.; Heeren, R. M. A., *Anal. Chem.* **2004,** *76* (18), 5339-5344.
(55) Ross, B. M.; Kish, S. J., *J. Neurochem.* **1994,** *63* (5), 1839-1848.
(56) Ishii, I.; Fukushima, N.; Ye, X.; Chun, J., *Annu. Rev. Biochem.* **2004,** *73* (1), 321-354.
(57) Weltzien, H. U., *Biochimica et Biophysica Acta (BBA) - Reviews on Biomembranes* **1979,** *559* (2), 259-287.
(58) Rustenbeck, I.; Lenzen, S., *Cell Calcium* **1992,** *13* (4), 193-202.
(59) Soga, T.; Ohishi, T.; Matsui, T.; Saito, T.; Matsumoto, M.; Takasaki, J.; Matsumoto, S.-i.; Kamohara, M.; Hiyama, H.; Yoshida, S.; Momose, K.; Ueda, Y.; Matsushime, H.; Kobori, M.; Furuichi, K., *Biochem. Biophys. Res. Commun.* **2005,** *326* (4), 744-751.
(60) Gräler, M. H.; Goetzl, E. J., *Biochim. Biophys. Acta* **2002,** *1582* (1), 168-174.

(61) Hong, J. H.; Kang, J. W.; Kim, D. K.; Baik, S. H.; Kim, K. H.; Shanta, S. R.; Jung, J. H.; Mook-Jung, I.; Kim, K. P., *J. Lipid Res.* **2016,** *57* (1), 36-45.
(62) Kaya, I.; Zetterberg, H.; Blennow, K.; Hanrieder, J., *ACS Chem. Neurosci.* **2018,** *9* (7), 1802-1817.
(63) Clemens, J. A.; Sofia, M. J.; Stephenson, D. T., Use of PLA2 inhibitors as treatment for alzheimer's disease. Google Patents: 1995.
(64) Sanchez-Mejia, R. O.; Newman, J. W.; Toh, S.; Yu, G.-Q.; Zhou, Y.; Halabisky, B.; Cissé, M.; Scearce-Levie, K.; Cheng, I. H.; Gan, L.; Palop, J. J.; Bonventre, J. V.; Mucke, L., *Nat. Neurosci.* **2008,** *11*, 1311.
(65) Muscari, C.; Guarnieri, C.; Stefanelli, C.; Giaccari, A.; Caldarera, C. M., *Mol. Cell. Biochem.* **1995,** *144* (2), 125-129.
(66) Yatin, S. M.; Yatin, M.; Varadarajan, S.; Ain, K. B.; Butterfield, D. A., *J. Neurosci. Res.* **2001,** *63* (5), 395-401.
(67) Khan, A. U.; Mei, Y. H.; Wilson, T., *Proceedings of the National Academy of Sciences* **1992,** *89* (23), 11426.
(68) Yatin, S. M.; Yatin, M.; Aulick, T.; Ain, K. B.; Butterfield, D. A., *Neurosci. Lett.* **1999,** *263* (1), 17-20.
(69) Ha, H. C.; Sirisoma, N. S.; Kuppusamy, P.; Zweier, J. L.; Woster, P. M.; Casero, R. A., *Proceedings of the National Academy of Sciences* **1998,** *95* (19), 11140.
(70) Gilad, G. M.; Gilad, V. H., *Exp. Neurol.* **1991,** *111* (3), 349-355.
(71) Sanchez-Mejia, R. O.; Mucke, L., *Biochim. Biophys. Acta* **2010,** *1801* (8), 784-90.
(72) Malaplate-Armand, C.; Florent-Bechard, S.; Youssef, I.; Koziel, V.; Sponne, I.; Kriem, B.; Leininger-Muller, B.; Olivier, J. L.; Oster, T.; Pillot, T., *Neurobiol. Dis.* **2006,** *23* (1), 178-89.
(73) Chalimoniuk, M.; Stolecka, A.; Cakała, M.; Hauptmann, S.; Schulz, K.; Lipka, U.; Leuner, K.; Eckert, A.; Muller, W. E.; Strosznajder, J. B., *Acta Biochim. Pol.* **2007,** *54* (3), 611-623.
(74) Esposito, G.; Giovacchini, G.; Liow, J. S.; Bhattacharjee, A. K.; Greenstein, D.; Schapiro, M.; Hallett, M.; Herscovitch, P.; Eckelman, W. C.; Carson, R. E.; Rapoport, S. I., *J. Nucl. Med.* **2008,** *49* (9), 1414-21.
(75) Newsholme, P.; Procopio, J.; Lima, M. M.; Pithon-Curi, T. C.; Curi, R., *Cell Biochem. Funct.* **2003,** *21* (1), 1-9.
(76) Robinson, S. R., *J. Neurosci. Res.* **2001,** *66* (5), 972-980.
(77) Markesbery, W. R., *Free Radic. Biol. Med.* **1997,** *23* (1), 134-147.
(78) Nunomura, A.; Castellani, R. J.; Perry, G.; Zhu, X.; Smith, M. A.; Moreira, P. I., *J. Neuropathol. Exp. Neurol.* **2006,** *65* (7), 631-641.
(79) Swomley, A. M.; Forster, S.; Keeney, J. T.; Triplett, J.; Zhang, Z.; Sultana, R.; Butterfield, D. A., *Biochim. Biophys. Acta* **2014,** *1842* (8), 1248-57.
(80) Christen, Y., *The American Journal of Clinical Nutrition* **2000,** *71* (2), 621S-629S.
(81) Butterfield, D. A.; Lauderback, C. M., *Free Radic. Biol. Med.* **2002,** *32* (11), 1050-1060.
(82) Gabbita, S. P.; Lovell, M. A.; Markesbery, W. R., *J. Neurochem.* **1998,** *71* (5), 2034-2040.
(83) Smith, C. D.; Carney, J. M.; Starke-Reed, P. E.; Oliver, C. N.; Stadtman, E. R.; Floyd, R. A.; Markesbery, W. R., *Proceedings of the National Academy of Sciences* **1991,** *88* (23), 10540.
(84) Emerit, J.; Edeas, M.; Bricaire, F., *Biomed. Pharmacother.* **2004,** *58* (1), 39-46.
(85) Castegna, A.; Aksenov, M.; Aksenova, M.; Thongboonkerd, V.; Klein, J. B.; Pierce, W. M.; Booze, R.; Markesbery, W. R.; Butterfield, D. A., *Free Radic. Biol. Med.* **2002,** *33* (4), 562-571.
(86) Wang, H.; Lian, K.; Han, B.; Wang, Y.; Kuo, S. H.; Geng, Y.; Qiang, J.; Sun, M.; Wang, M., *J. Alzheimers Dis.* **2014,** *39* (4), 841-8.
(87) Oberg, J.; Spenger, C.; Wang, F. H.; Andersson, A.; Westman, E.; Skoglund, P.; Sunnemark, D.; Norinder, U.; Klason, T.; Wahlund, L. O.; Lindberg, M., *Neurobiol. Aging* **2008,** *29* (9), 1423-33.
(88) Bubber, P.; Haroutunian, V.; Fisch, G.; Blass, J. P.; Gibson, G. E., *Ann. Neurol.* **2005,** *57* (5), 695-703.
(89) Yao, J.; Brinton, R. D., *Adv. Pharmacol.* **2012,** *64*, 327-71.
(90) Xu, J.; Begley, P.; Church, S. J.; Patassini, S.; Hollywood, K. A.; Jullig, M.; Curtis, M. A.; Waldvogel, H. J.; Faull, R. L.; Unwin, R. D.; Cooper, G. J., *Biochim. Biophys. Acta* **2016,** *1862* (6), 1084-92.

(91) Krebs, H., *Perspect. Biol. Med.* **1970,** *14* (1), 154-172.
(92) Gibson, G. E.; Starkov, A.; Blass, J. P.; Ratan, R. R.; Beal, M. F., *Biochim. Biophys. Acta* **2010,** *1802* (1), 122-34.

Appendix 1: Chapter 2 Supplemental

Figure S2-1. MSI data analysis using MCR-ALS method.

The spatial distribution patterns (left) and the grouped molecules (right) present in (A) inner medulla and (B) cortex of mouse kidney slice. (C) and (D) are from the other two components.

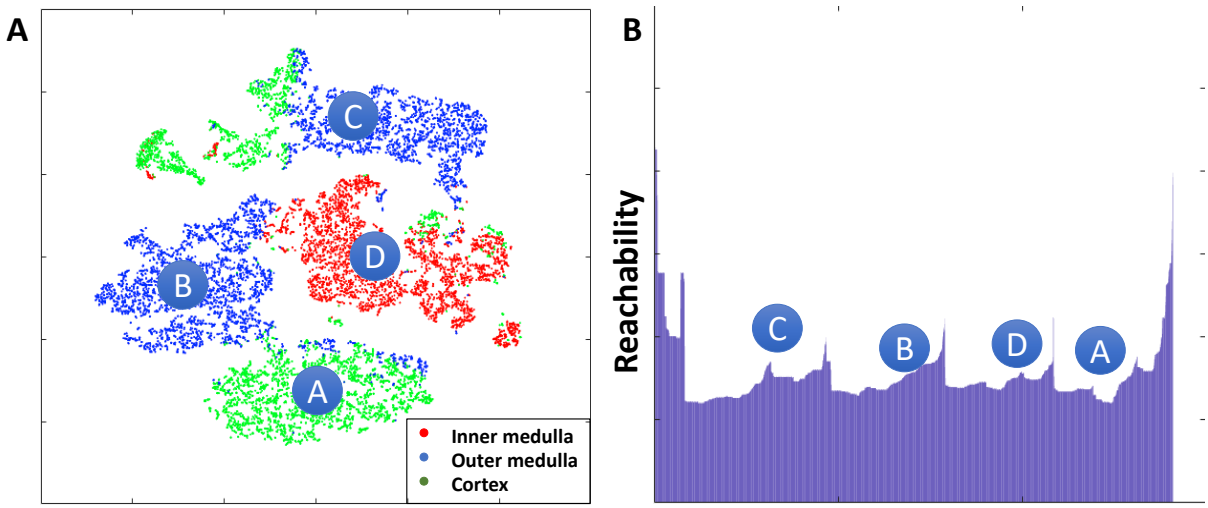


Figure S2-2. Validation of results obtained from the supervised ML.

(A) Visualization of supervised ML results using *t*-SNE. Tissue labels are color-coded as red (inner medulla), blue (outer medulla), and green (cortex) shown in separated groups. (B) Bar plot of point reachability from OPTICS analysis. Similar to the *t*-SNE result, four major clusters (A, B, C, and D) were achieved. Although the outer medulla was reduced into two separated clusters (B and C), they have shorter reachability distance (i.e., the highest bar between B and C) than that between A (inner medulla) and D (cortex). In fact, the relative shorter reachability distance between B and C is likely due to their similar chemical compositions. It follows that OPTICS reachability plot supported the *t*-SNE distribution map. Overall, the dimensionality reduction using *t*-SNE provided satisfactory results that are suitable for the flowing unsupervised ML approaches.

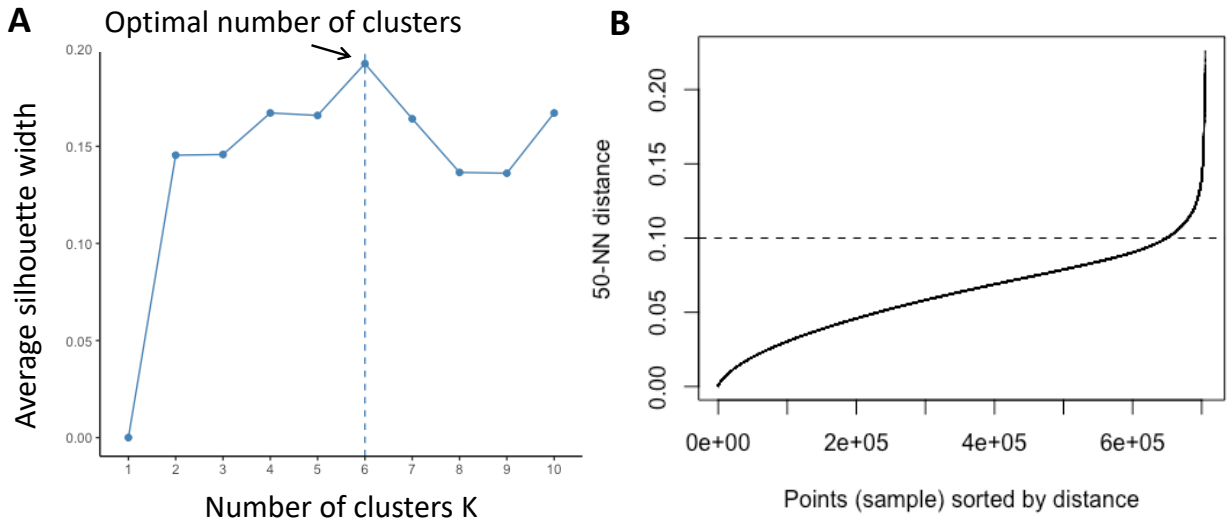


Figure S2-3. Optimization of unsupervised ML methods.

(A) Optimization of the number of clusters in CLARA. (B) Optimization of the epsilon value in DBSCAN.

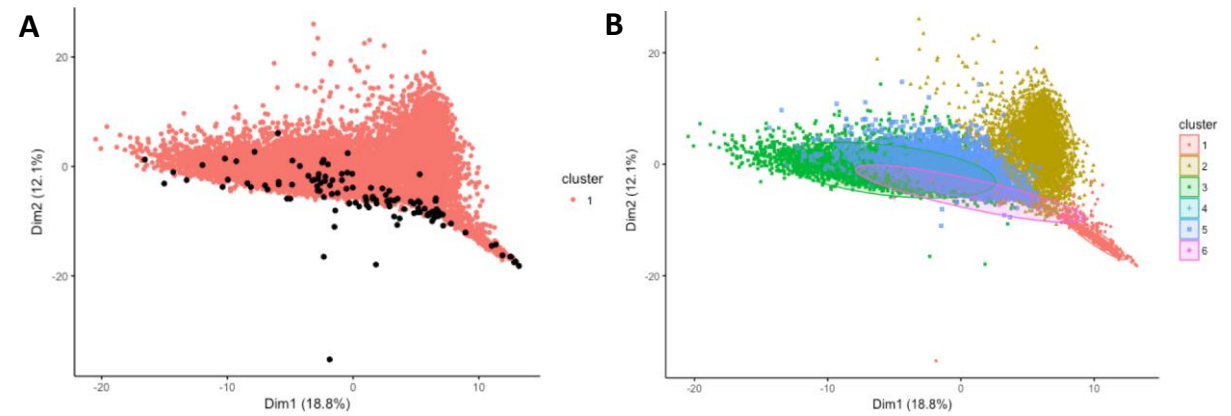


Figure S2-4. Unsupervised ML analyses of the high-dimensional MSI dataset without performing t-SNE dimensionality reduction.

Unsatisfactory results were obtained from (A) CLARA and (B) DBSCAN without conducting *t*-SNE dimensionality reduction in advance.

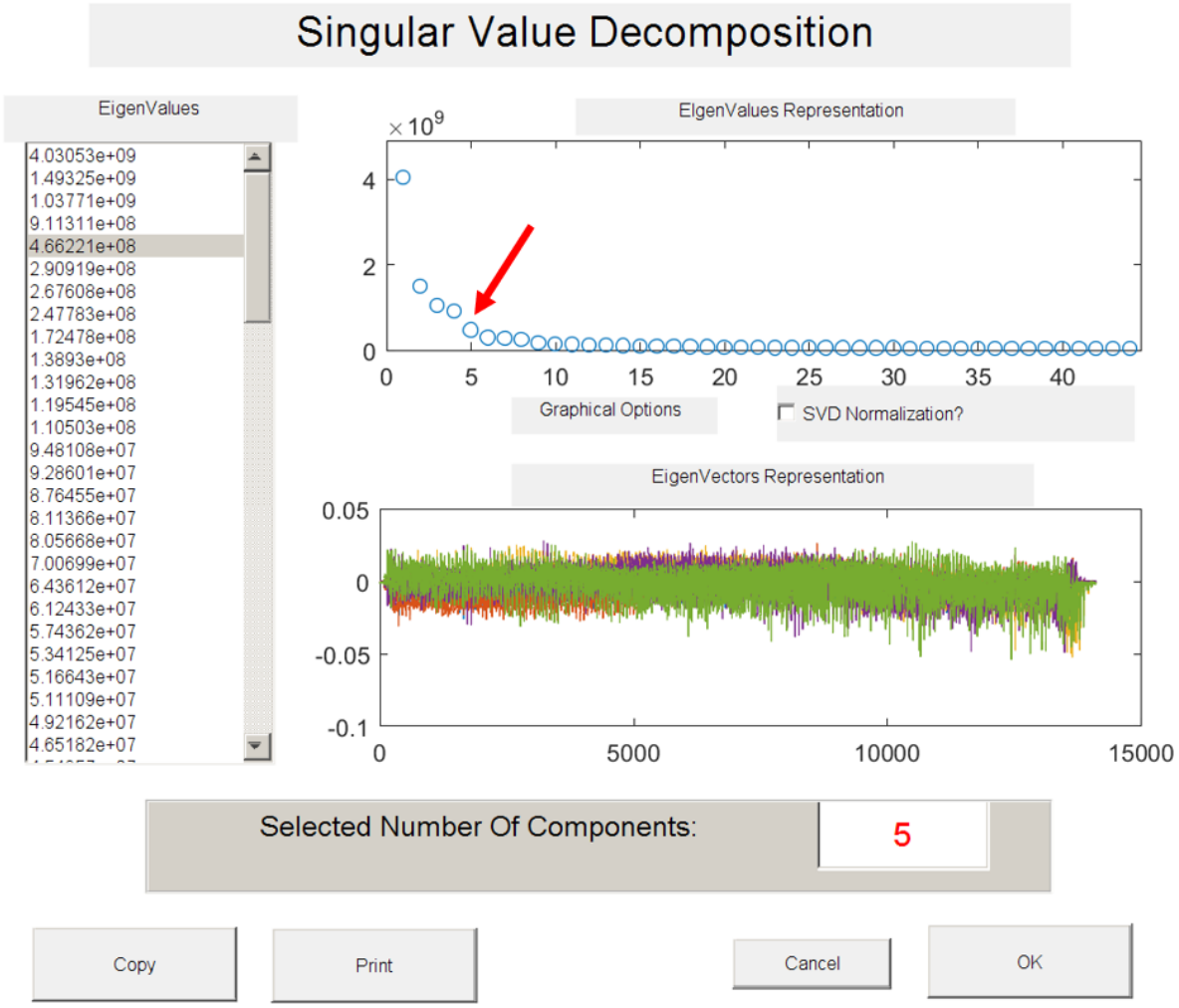


Figure S2-5. Determining the number of components in MCR analysis using the Singular Value Decomposition (SVD).

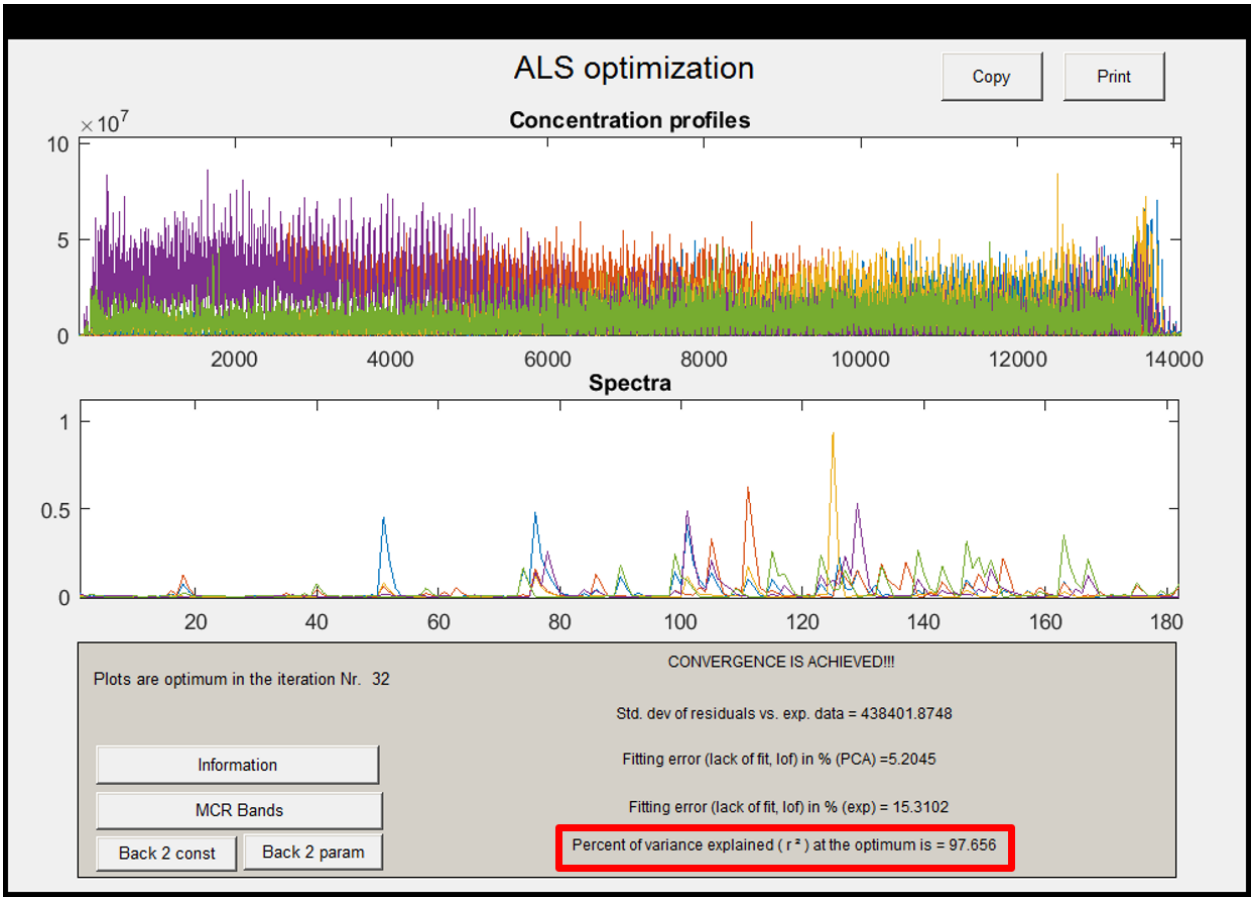


Figure S2-6. The percent of variance explained (97.656%) with the number of component (5).

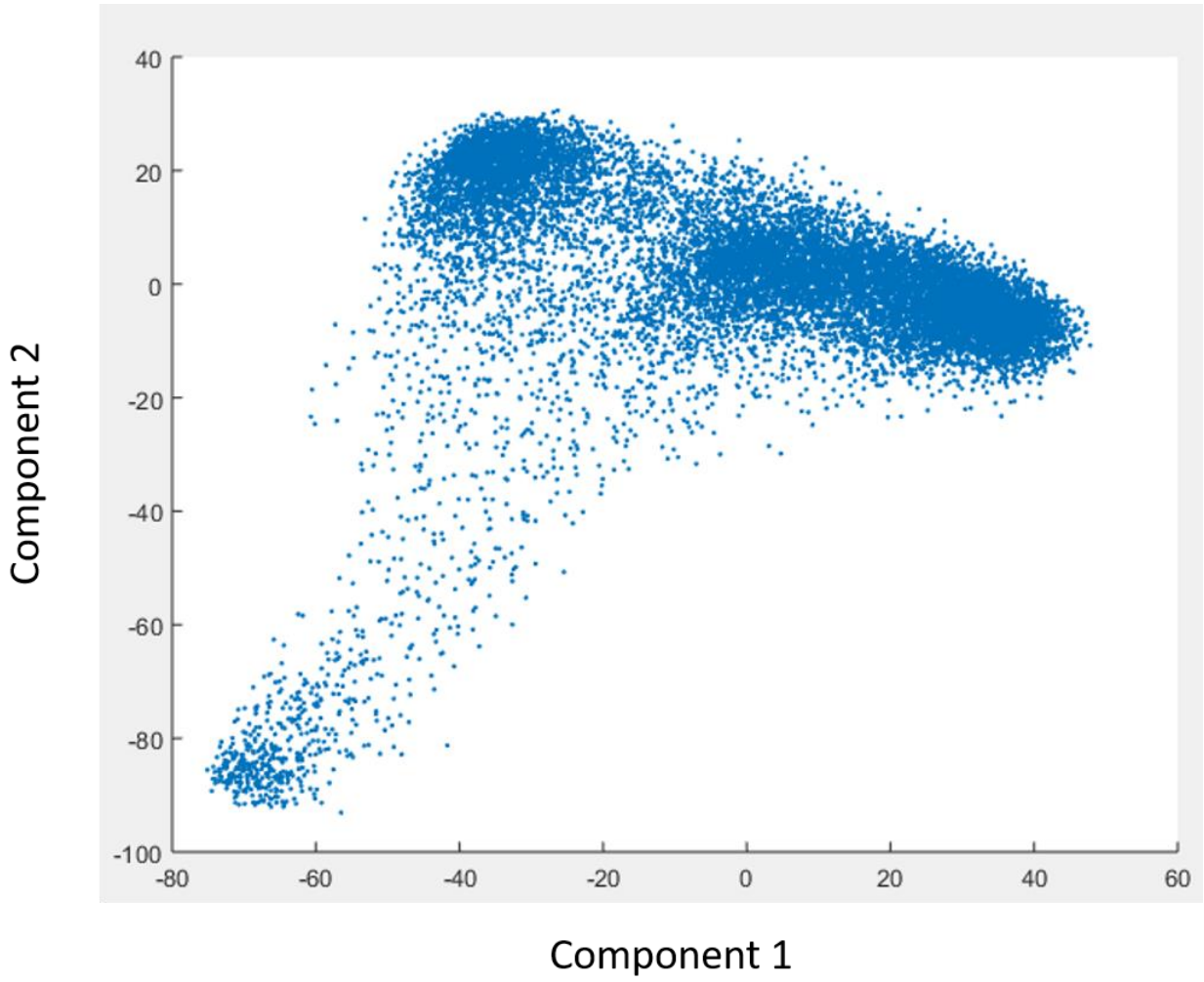


Figure S2-7. Unsatisfactory results were obtained from dimensionality reduction using PCA.

Table S2-1. Data selection for Random Forest model training and testing.

	Rows/pixel	Columns/(m/z)
Selected dataset	2900	182
Train data	2610	182
Test data	290	182

Table S2-2. Validation of supervised ML model using testing data obtained from MCR results.

Trail number	Inner medulla	Outer medulla	Cortex
1	99.01%	100%	100%
2	99.10%	98.95%	100%
3	99.09%	100%	98.78%
Average	99.07%	99.65%	99.59%

Table S2-3. Top 15 most abundant ions grouped in component 2.

MCR (m/z)	Experiment(m/z)	Tentative Labeling*	Exact (m/z)	ppm	Loading Score
810.6423	810.6019	PC(20:3/18:1)	810.6007	1	0.5329
782.5432	782.5701	PC(18:3/18:1)	782.5694	0	0.4913
811.6421	811.6054	SM(17:1/22:0)	811.6090	4	0.2716
760.5466	760.5870	PC(22:1/12:0)	760.5851	2	0.2555
783.5430	783.5739	SM(18:1/19:0)	783.5777	4	0.2319
808.5428	808.5878	PC(20:4/18:1)	808.5851	3	0.2281
786.6442	786.6017	PC(18:1/18:1)	786.6007	1	0.2045
832.5390	832.5841	PC(18:1/20:3)	832.5827	1	0.1538
758.5451	758.5702	PE(22:2/15:0)	758.5694	1	0.1376
804.5418	804.5525	PC(14:0/22:4)	804.5514	1	0.1185
848.5390	848.5534	PC(18:3/20:1)	848.5566	3	0.1178
820.5419	820.5266	PC(18:0/18:4)	820.5253	1	0.0965
806.5413	806.5704	PC(20:2/18:4)	806.5694	1	0.0923
788.6437	788.6201	PC(17:0/19:1)	788.6164	4	0.0704
784.5447	784.5871	PC(19:1/17:2)	784.5851	2	0.0639

*PC: Phosphatidylcholines, PE: Phosphatidylethanolamine, SM: Sphingomyelin.

Table S2-4. Top 15 most abundant ions grouped in component 3.

MCR (m/z)	Experiment(m/z)	Tentative Labeling*	Exact (m/z)	ppm	Loading Score
758.5451	758.5707	PE(22:2/15:0)	785.5694	1	0.4787
734.5470	734.5704	PC(14:0/18:0)	734.5694	1	0.4519
782.5432	782.5703	PC(18:3/18:1)	782.5694	1	0.4080
783.5430	783.5736	SM(18:1/19:0)	783.5777	5	0.1937
756.5456	756.5527	PE(19:0/16:0)	756.5513	1	0.1534
810.6423	810.6024	PC(20:3/18:1)	810.6002	2	0.1444
780.5437	780.5535	PC(20:5/16:0)	780.5538	0	0.1386
760.5466	760.5875	PE(15:0/22:1)	760.5851	3	0.1349
786.6442	786.6027	PC(18:1/18:1)	786.6007	2	0.1332
772.5457	772.5273	PC(13:0/19:0)	772.5253	2	0.1121
792.5428	792.5916	PC(16:0/22:5)	792.5902	1	0.0982
796.5438	796.5272	PC(18:2/16:0)	796.5253	2	0.0981
828.5399	828.5533	PC(22:5/18:4)	828.5538	0	0.0946
784.5447	784.5884	PC(19:1/17:2)	782.5851	4	0.0946
844.5380	844.5272	PC(18:4/20:2)	844.5253	2	0.0816

*PC: Phosphatidylcholines, PE: Phosphatidylethanolamine, SM: Sphingomyelin.

Table S2-5. Top 15 most abundant ions grouped in component 4.

MCR (m/z)	Experiment(m/z)	Tentative Labeling*	Exact (m/z)	ppm	Loading Score
806.5413	806.5709	PC(20:2/18:4)	806.5694	1	0.9384
792.5428	792.5916	PC(16:0/22:5)	792.5902	1	0.1713
758.5451	758.5707	PE(22:2/15:0)	758.5694	1	0.1157
782.5432	782.5703	PC(18:3/18:1)	782.5694	1	0.1151
793.6425	793.5951	PG(19:0/18:0)	793.5953	0	0.0844
734.5470	734.5710	PC(14:0/18:0)	734.5694	2	0.0767
834.6404	834.6027	PC(20:4/20:2)	834.6007	2	0.0468
828.5399	828.5533	PC(22:5/18:4)	828.5538	0	0.0370
844.5380	844.5272	PC(18:4/20:2)	844.5353	2	0.0257
822.2415	822.5658	PS(20:0/20:5)	822.5643	1	0.0256
838.5195	838.5608	PS(20:4/20:1)	838.5593	1	0.0241
760.5466	760.5875	PE(15:0/22:1)	760.5853	3	0.0227
804.5418	804.5526	PC(14:0/22:4)	804.5514	1	0.0144
798.5433	798.5427	PC(18:0/16:1)	798.5410	2	0.0605
849.5388	849.5611	PG(20:1/20:3)	849.5616	0	0.0582

*PC: Phosphatidylcholines, PE: Phosphatidylethanolamine, PG: Phosphatidylglycerol.

PS: Phosphatidylserine

Table S2-6. Top 15 most abundant ions grouped in component 5

MCR (m/z)	Experiment(m/z)	Tentative Labeling*	Exact (m/z)	ppm	Loading Score
844.5380615	844.5272	PC(18:4/20:2)	844.5253	2	0.34957
828.5399867	828.5533	PC(22:5/18:4)	828.5538	0	0.31601
820.5419491	820.5272	PC(18:0/18:4)	820.5253	2	0.26439
796.5438371	796.5366	PS(16:0/20:1)	796.5463	12	0.25983
780.5437628	780.5529	PC(16:0/18:2)	780.5514	1	0.24061
804.5418748	804.5531	PC(20:4/18:3)	804.5538	0	0.23871
830.5394961	830.5468	PC(22:5/16:0)	830.546	0	0.22432
848.5390798	848.5584	PC(18:3/20:1)	848.5566	2	0.21048
832.5390055	832.5853	PC(20:5/20:2)	832.5851	0	0.20511
772.5457252	772.5266	PC(13:0/19:0)	772.5253	1	0.17865
814.5414213	814.5736	PC(13:0/22:0)	814.5723	1	0.17369
824.5409679	824.5583	PC(18:2/18:0)	824.5566	2	0.1725
845.5398157	845.53	PG(18:1/22:5)	845.5303	0	0.17228
756.5456509	756.5528	PE(19:0/16:0)	756.5513	1	0.16383
808.5428931	808.5867	PC(20:4/18:1)	808.5851	2	0.15259

*PC: Phosphatidylcholines, PE: Phosphatidylethanolamine, PG: Phosphatidylglycerol, PS: Phosphatidylserine

Table S2-7. Percent of variance explained when choosing different number of components

Number of component	Percent of variance explained
3	92.80%
4	96.65%
5	97.66%
6	98.18%
7	98.50%

Appendix 2: Chapter 3 Supplemental

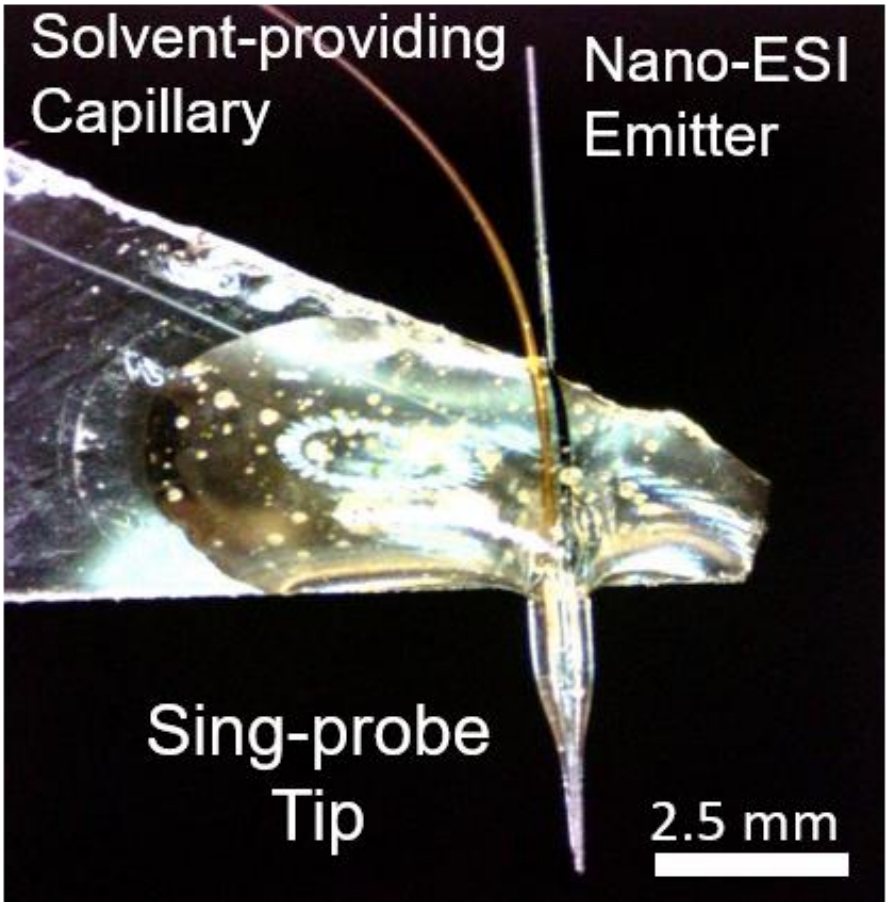


Figure S3-1. The photo of a Single-probe, which was assembled using three parts: a Nano ESI (nano-ESI) emitter, a dual-bore needle, and a solvent-providing capillary.

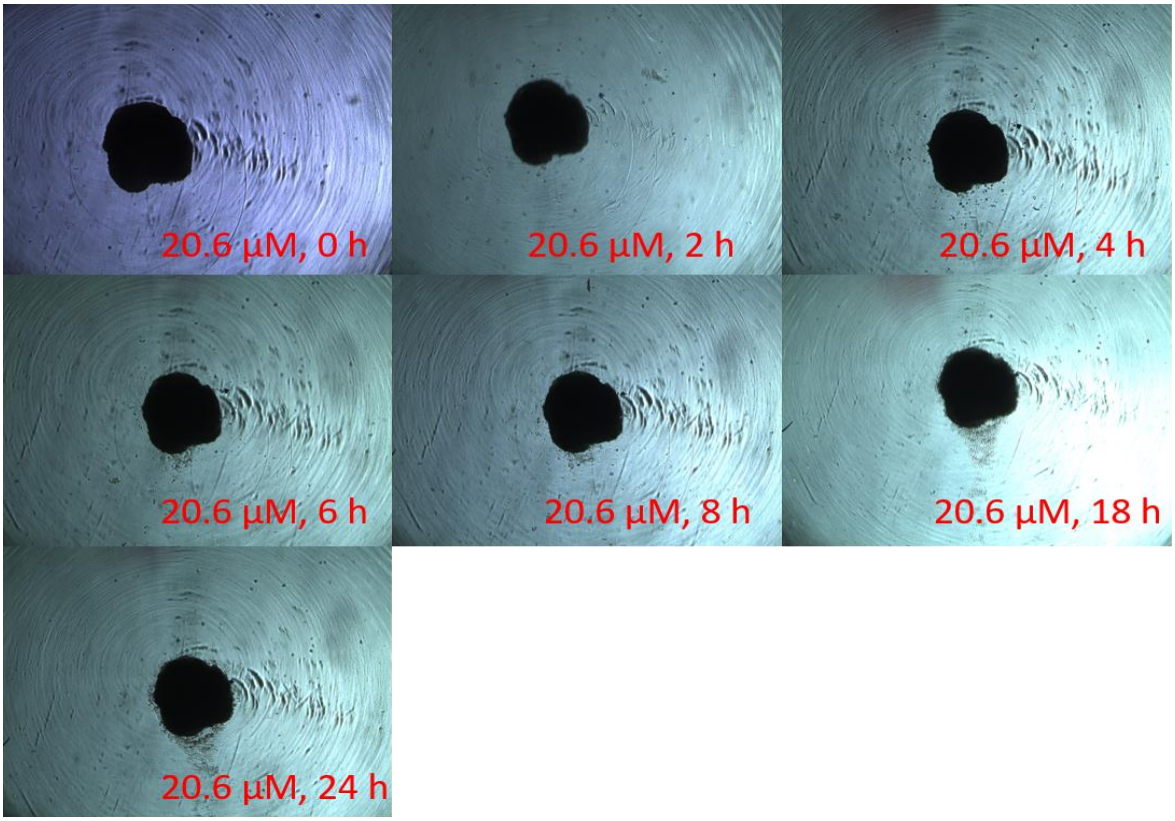


Figure S3-2. Microscope photos of spheroid under the treatment of Irinotecan (20.6 µM) obtained at different time points. The morphology of spheroid has no significant change within 24 hours.

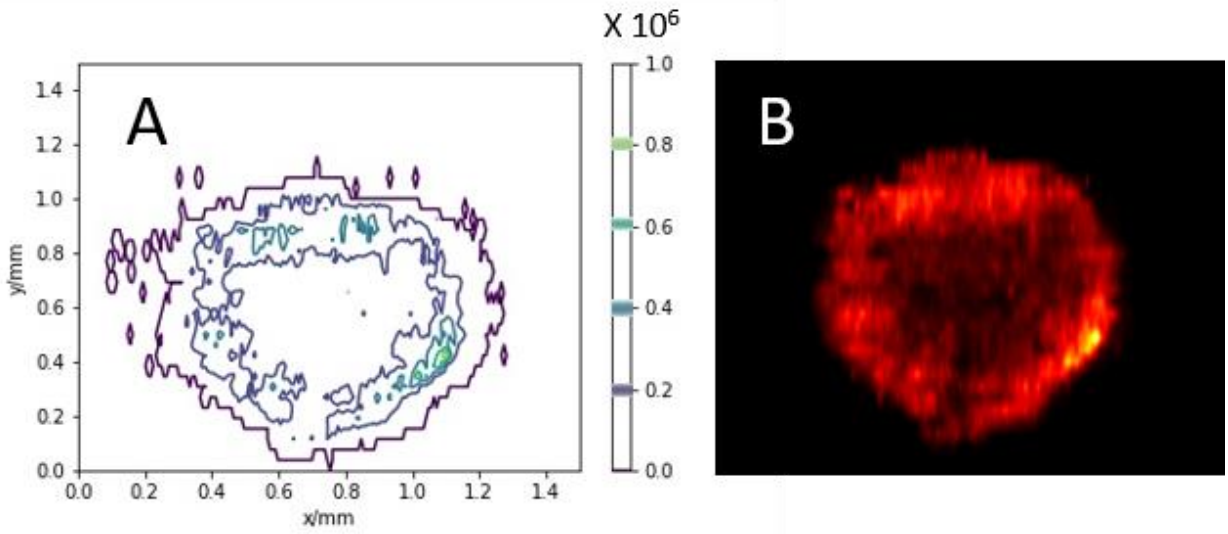


Figure S3-3. The relative abundances of Irinotecan on spheroid (20.6 µM, 10 h) illustrated using (A) contour plot and (B) MSI heatmap.

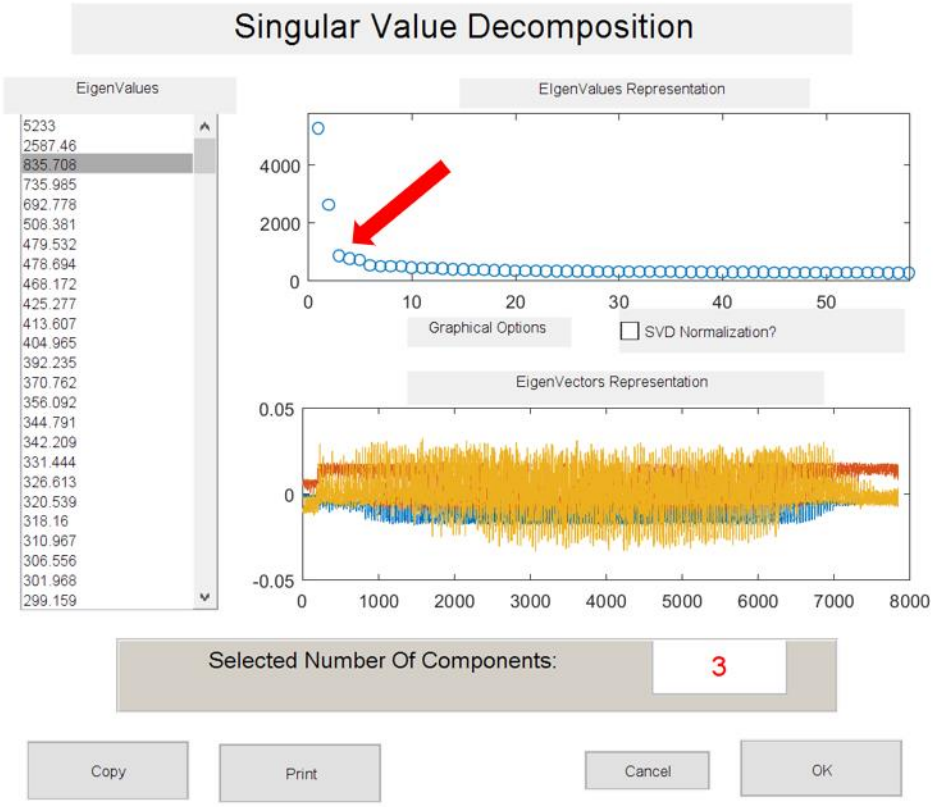


Figure S3-4. Determining the number of components in MCR analysis using the Singular Value Decomposition (SVD).

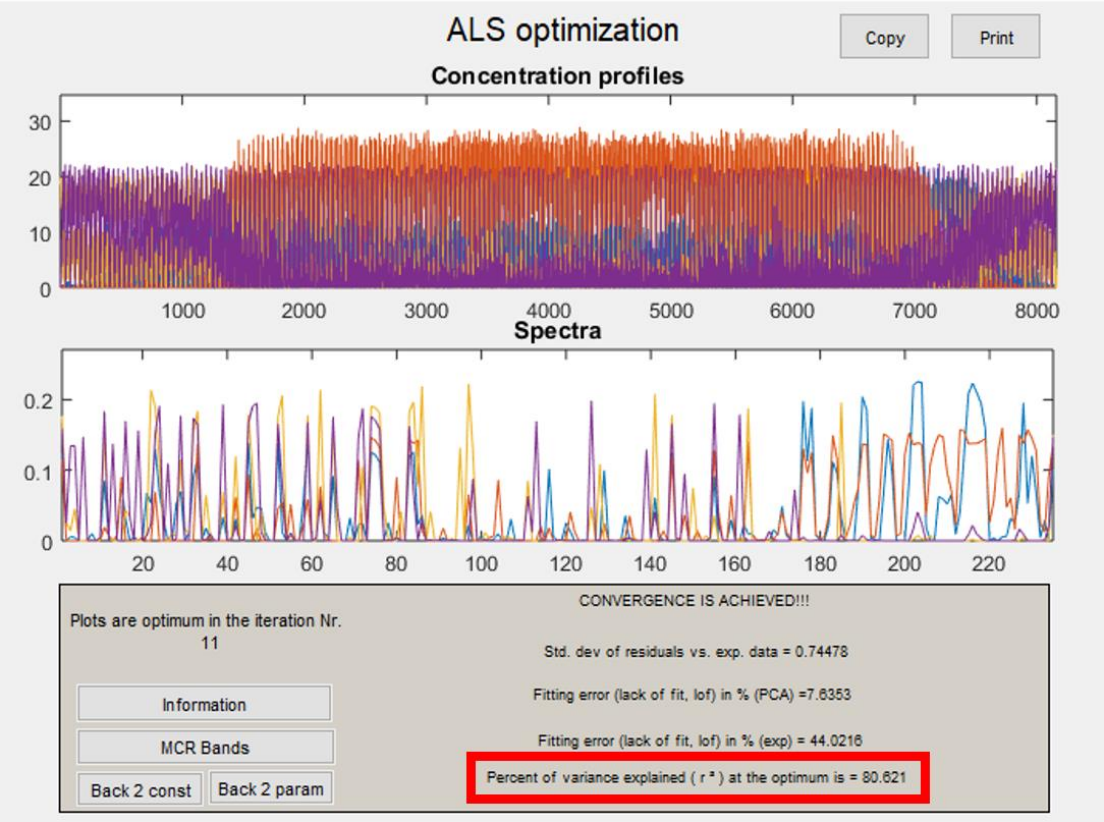


Figure S3-5. Percent of variance explained (80.621%) in the MSI data obtained from the control spheroid. Analyses were conducted using Multivariate Curve Resolution Toolbox.

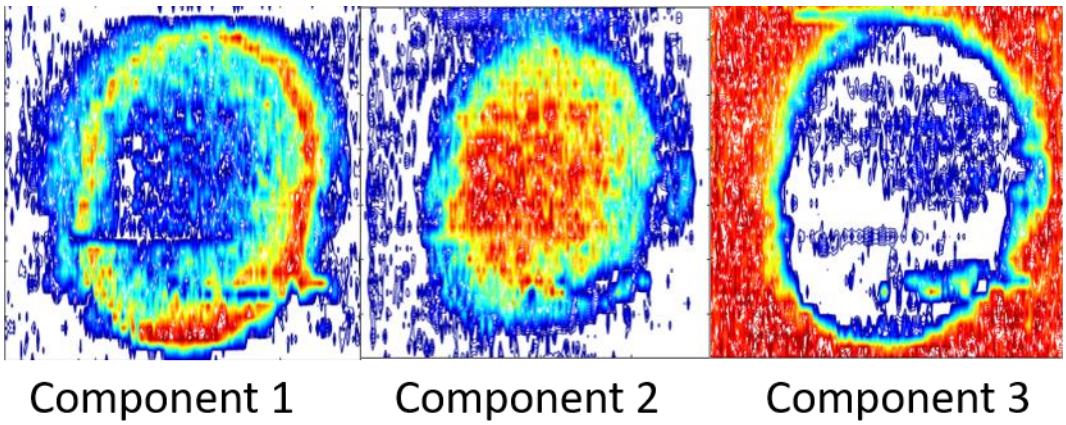


Figure S3-6. Alternative contour plots of Figure 3-5. Plots are constructed based on MCR analysis results of MS images of an Irinotecan treated spheroid (5 µM, 1 h).

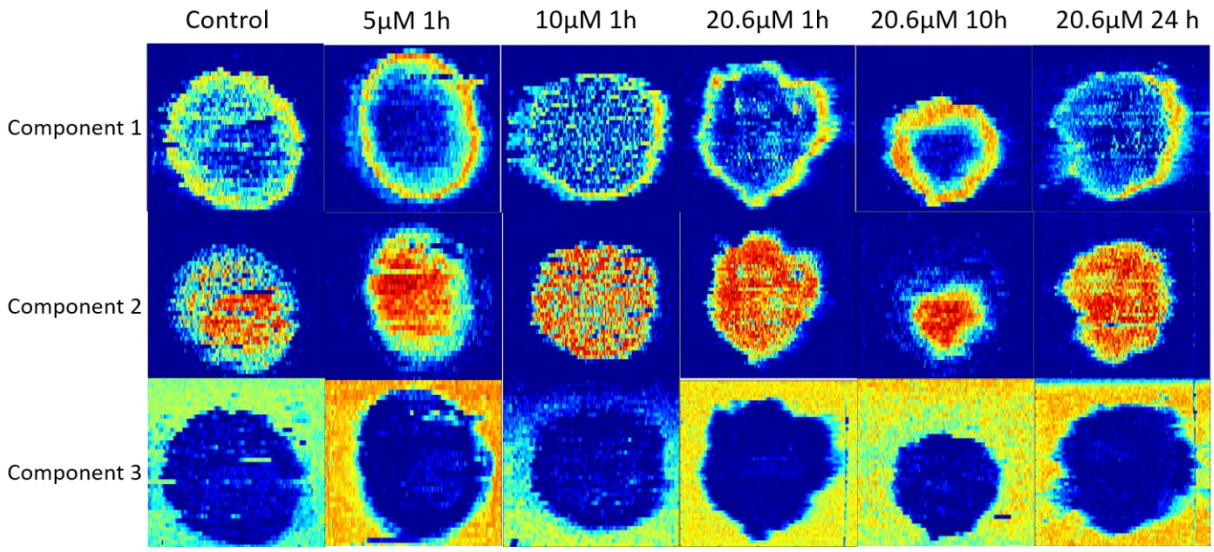


Figure S3-7. MCR-ALS analysis results of MSI data obtained from spheroids in the control and drug-treated groups. All MSI datasets were decomposed into three major components: outer region (component 1), inner region (component 2), and background (component 3).

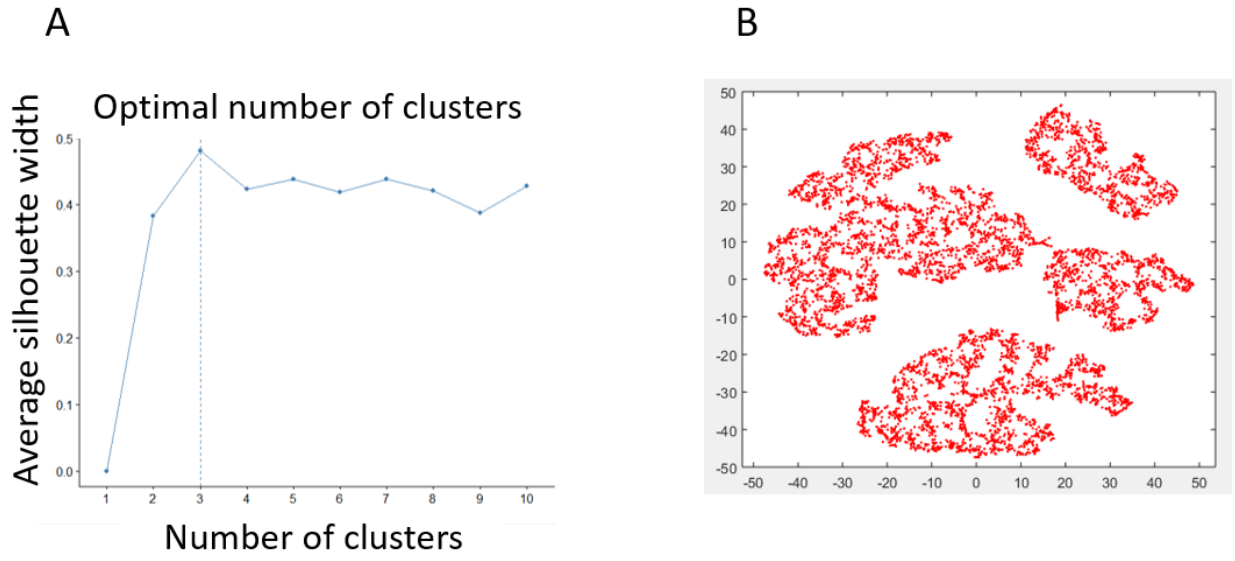


Figure S3-8. Optimization of unsupervised ML methods.

(A) Optimization of the number of clusters in CLARA model. (B) Dimensionality reduction using *t*-SNE prior to CLARA analysis.

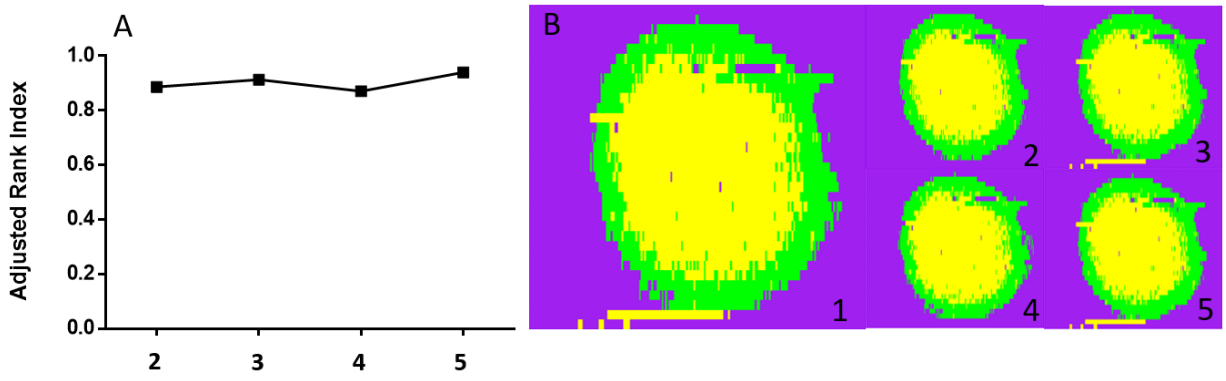


Figure S3-9. Evaluation of the stability of t-SNE for dimensionality reduction. The t-SNE calculations were conducted five times using the same MSI data (treatment condition: 5 µM, 1 h).

(A) Adjusted Rand Index was obtained for each *t*-SNE calculation, and consistently high values (0.85 ─ 0.91) indicate that *t*-SNE has a very stable performance through all five calculations. (Adjusted Rand Index ranges from 0 to 1, and higher values indicate higher similarity). (B) The comparison of five reconstructed CLARA results.

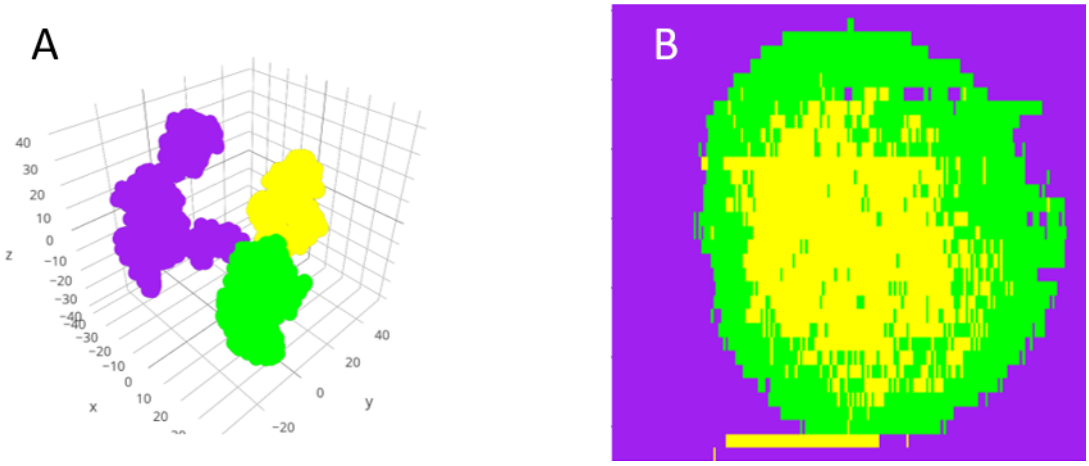


Figure S3-10. Evaluation the influence of 3D t-SNE dimensionality reduction on CLARA results.

(A) CLARA result based on 3D *t*-SNE dimensionality reduction. (B) The reconstructed spatial distribution based on CLARA result shows similar features as those based on 2D *t*-SNE method.

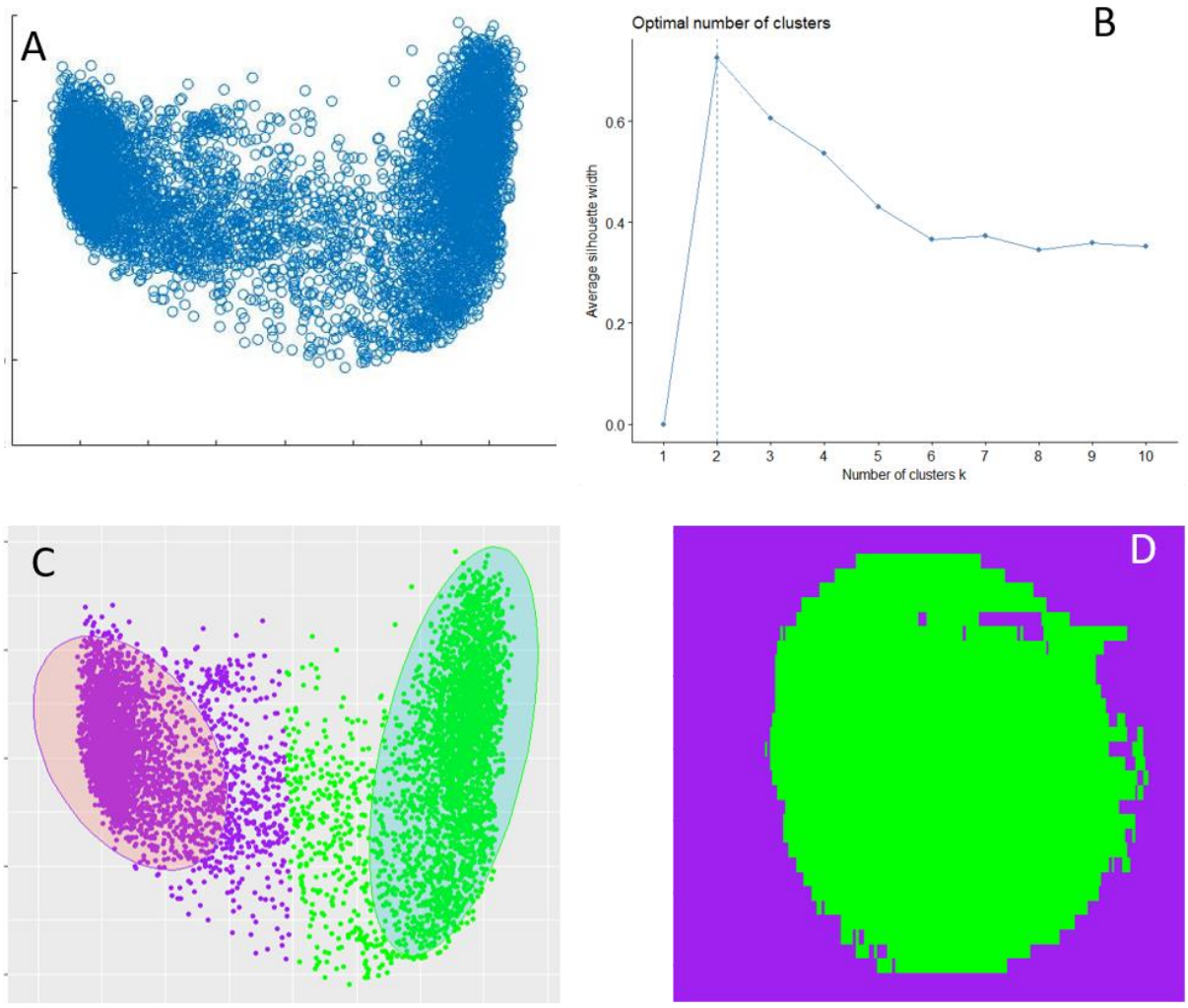


Figure S3-11. Evaluation of PCA for dimensionality reduction.

(A) PCA result of a drug treated spheroid slice (5 μM, 1 h). (B) Optimal number of clusters (i.e., 2) determined from CLARA. (C) CLARA result of with MSI data with lower dimensions (2D). (D) Reconstruction of CLARA result showing the spatial distribution of grouped data points.

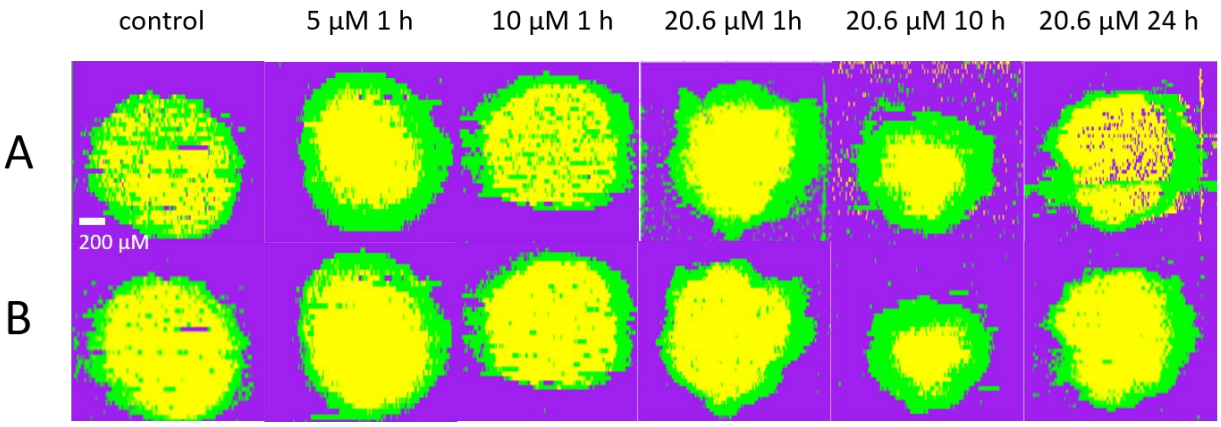


Figure S3-12. Results obtained from (A) the unsupervised and (B) supervised ML analyses of MS images obtained from both control and drug treated spheroids.

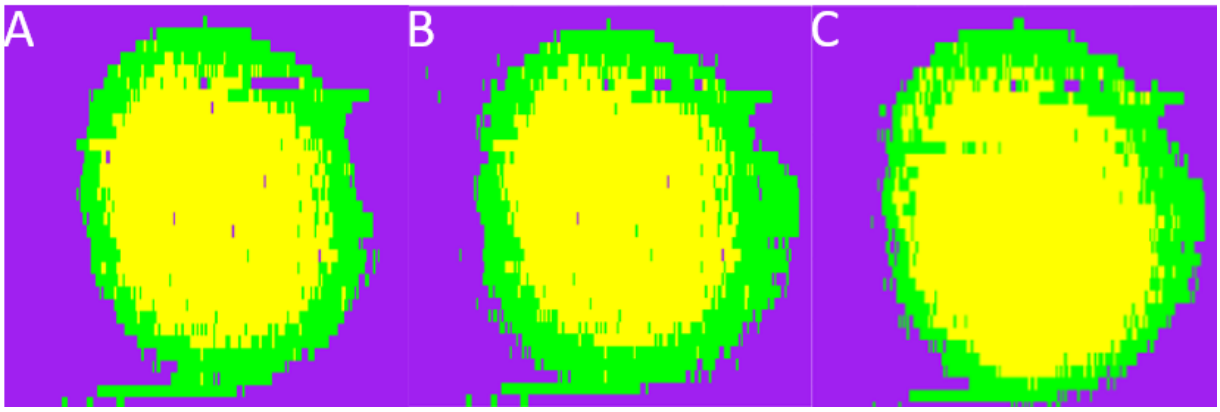


Figure S3-13. Comparison of results obtained from supervised ML (Random Forest) analysis of a drug treated spheroid slice (5 µM, 1 h) using different methods for training data selection: (A) k-means, (B) Hierarchical Clustering, and (C) MCR.

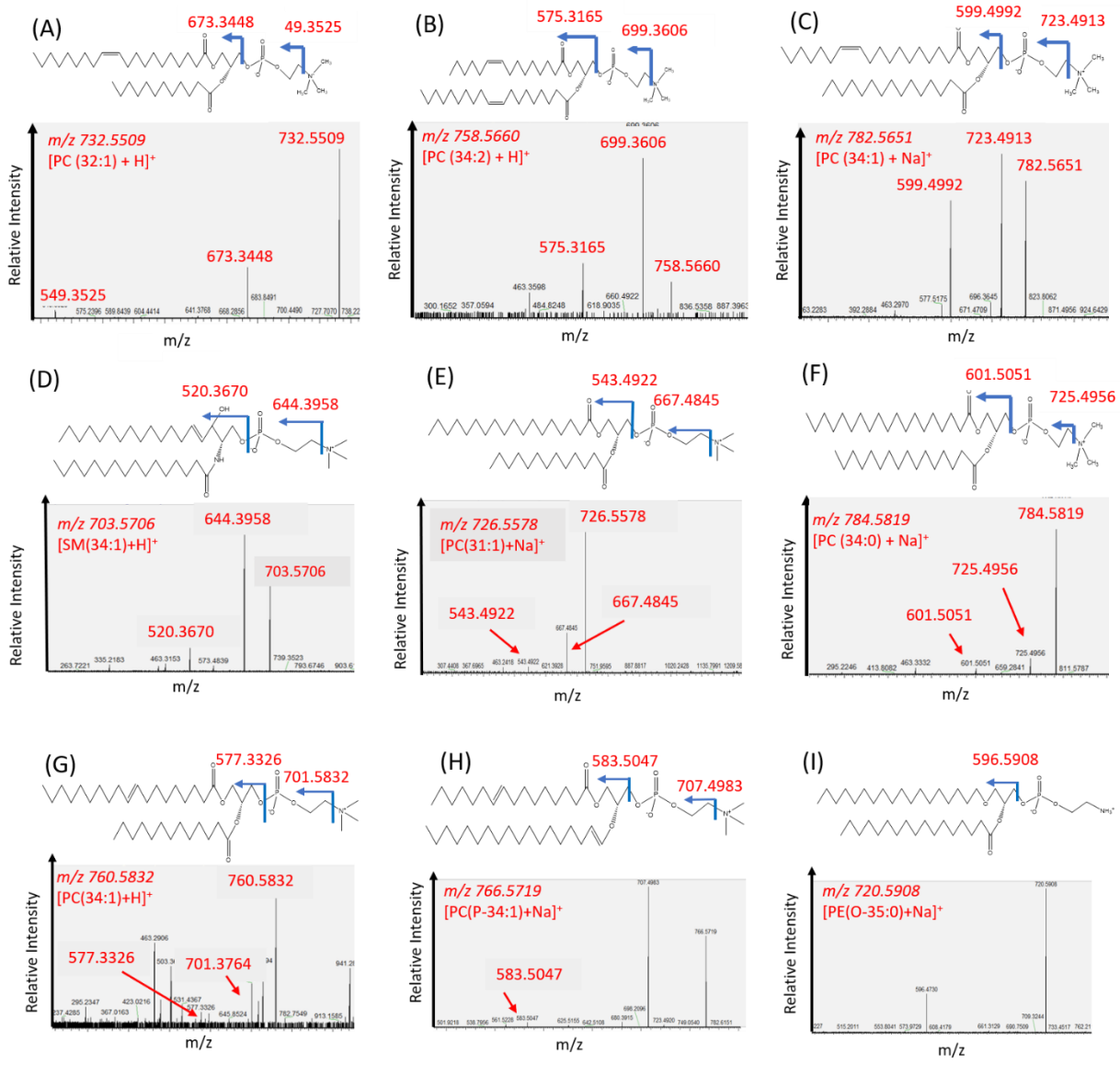


Figure S3-14. MS/MS analysis of selected metabolites.

(A) [PC(32:1) + H]$^+$, (B) [PC(34:2) + H]$^+$, (C) [PC(34:1) + Na]$^+$, (D) [SM(34:1) + H]$^+$, (E) [PC(31:1) + Na]$^+$, (F) [PC(34:0) + Na]$^+$, (G) [PC(34:1) + H]$^+$, (H) [PC(P-34:1) + Na]$^+$, and (I) [PE(O-35:0) + H]$^+$. All MS/MS spectra were compared with online databases Metlin.

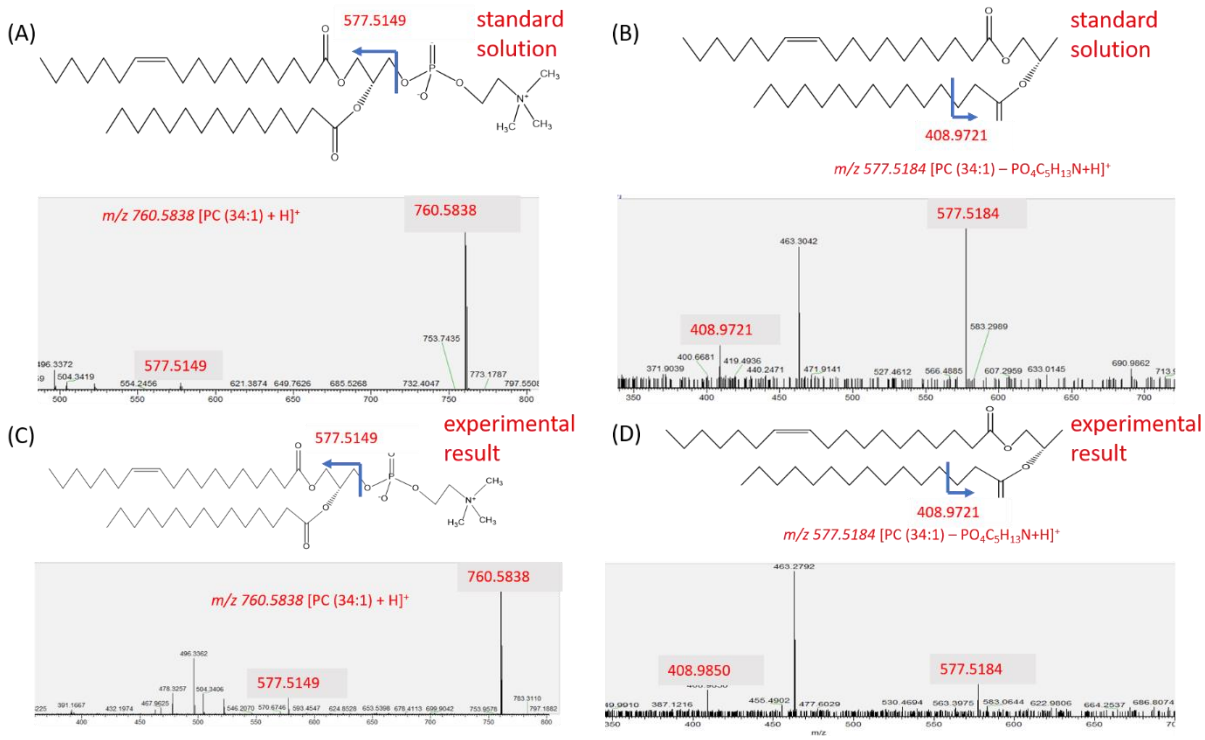


Figure S3-15. Comparison of MSn obtained from spheroid slices and standard compound for structure confirmation of metabolites.

Both MS^2 and MS^3 were conducted for $[PC(34:1) + H]^+$ (A and C) and ($[PC(34:1) - PO_4C_5H_{13}N + H]^+$ (B and D).

Table S3-1. Parameters of data matrix obtained from MSI datasets upon finishing pre-process.

Groups	Pixels	Number of common ions
Control	7480	307
5 μM, 1 h	6369	348
10 μM, 1 h	8160	235
20.6 μM, 1 h	11400	224
20.6 μM, 10 h	7760	346
20.6 μM, 24 h	9700	174

Table S3-2. The prediction accuracy of Random Forest models for MSI dataset.

Group	Background (%)	Inner Region (%)	Outer Region (%)
Control	96.3	91.21	95.50
5 μM 1 h	100	93.12	100
10 μM 1 h	100	98.12	94.50
20.6 μM 1 h	100	100	100
20.6 μM 10 h	94.22	95.15	98.61
20.6 μM 24 h	96.67	100	94.73

Table S3-3. Top-10 most abundant ions in the component 1.

m/z	Labels[a]	Loading Score
760.5832	PC(34:1)*	0.195995
761.5960	PA(40:0)	0.18903
786.6084	PC(36:2)	0.187542
758.5769	PC(34:2)*	0.181174
788.6243	PC(36:1)	0.173391
787.6120	PA(42:1)	0.173062
784.5929	PC(34:0)*	0.170212
759.5804	PA(40:1)	0.159413
703.5816	SM(34:1)*	0.156258
789.6277	PA(42:0)	0.154923

[a]PC: Phosphatidylcholines, PA: Phosphatidic Acid, SM: Sphingomyelin. *Structure was confirmed by MS/MS analysis.

Table S3-4. Top-10 most abundant ions in the component 2.

m/z	Labels[a]	Loading Score
767.2340	PA(P-40:0)	0.146215
770.3107	PC(O-36:3)	0.144249
773.3875	PA(P-42:0)	0.142898
719.1610	PA(O-38:0)	0.141826
746.6176	PE(P-37:0)	0.14169
783.6178	PA(40:0)	0.139972
782.6178	PC(34:1)*	0.139558
744.6177	PC(P-34:1)	0.139281
748.1178	PE(O-37:0)	0.13771
743.1927	PA(P-40:1)	0.136859

[a]PE: Phosphatidylethanolamine *Structure was confirmed by MS/MS

Table S3-5. Supervised machine learning (Random forest) data selection.

	control	5 µM, 1 h	10 µM, 1 h	20.6 µM, 1 h	20.6 µM, 10 h	20.6 µM, 24 h
Selected date (pixels)	600	400	600	500	400	500
Training data (Pixels)	540	360	540	450	360	450
Testing data (pixels)	60	40	60	50	40	50

Table S3-6. Metabolites significantly altered by Irinotecan treatment (concentration-dependent).

m/z	Assignment	Method	Relative Intensity				Region
			Control	5 μM, 1 h	10 μM, 1 h	20.6 μM, 1 h	
726.5446	$[PC (31:1) + Na]^+$	Supervised	1.13E-04	3.80E-04	9.30E-04	1.43E-02	Inner
		Unsupervised	2.17E-03	2.87E-03	2.78E-03	1.03E-02	
758.5685	$[PC (34:2) + H]^+$	Supervised	2.92E-02	3.64E-02	6.52E-02	1.42E-01	Inner
		Unsupervised	6.07E-02	4.50E-01	3.52E-01	2.59E-01	
782.5692	$[PC (34:1) + Na]^+$	Supervised	5.24E-03	2.71E-02	3.96E-02	1.87E-01	Inner
		Unsupervised	1.06E-02	2.38E-01	2.01E-01	1.40E-01	
784.5865	$[PC (34:0) + Na]^+$	Supervised	1.51E-02	1.68E-02	1.13E-01	1.13E-01	Inner
		Unsupervised	3.19E-02	2.31E-01	2.12E-01	1.25E-01	
766.5719	$[PC(P-34:1)+ Na]^+$	Supervised	3.97E-04	1.36E-03	1.66E-03	5.78E-03	Outer
		Unsupervised	3.99E-04	1.06E-03	1.02E-03	2.29E-03	

Table S3-7. Metabolites significantly altered by Irinotecan treatment (time-dependent).

m/z	Assignment	ML Method	Relative Intensities				Region
			Control	20.6 μM, 1 h	20.6 μM, 10 h	20.6 μM, 24 h	
703.5706	$[SM (34:1+H)]^+$	Supervised	4.55E-03	1.32E-01	1.25E-01	1.74E-01	Inner
		Unsupervised	9.69E-03	1.31E-01	1.61E-01	1.94E-01	
732.5438	$[PC (32:1) + H]^+$	Supervised	1.24E-02	1.42E-01	1.50E-01	1.04E-01	Inner
		Unsupervised	2.65E-02	1.67E-01	1.68E-01	1.31E-01	
760.5832	$[PC (34:1) + H]^+$	Supervised	7.82E-02	6.65E-01	1.51E-01	4.63E-01	Inner
		Unsupervised	1.61E-02	6.10E-01	2.10E-01	5.11E-01	
720.5912	$[PE(O-35:0) + Na]^+$	Supervised	2.98E-03	1.23E-02	6.85E-03	9.68E-03	Outer
		Unsupervised	3.26E-03	1.27E-02	1.04E-02	8.36E-03	

Appendix 3: Chapter 4 Supplemental

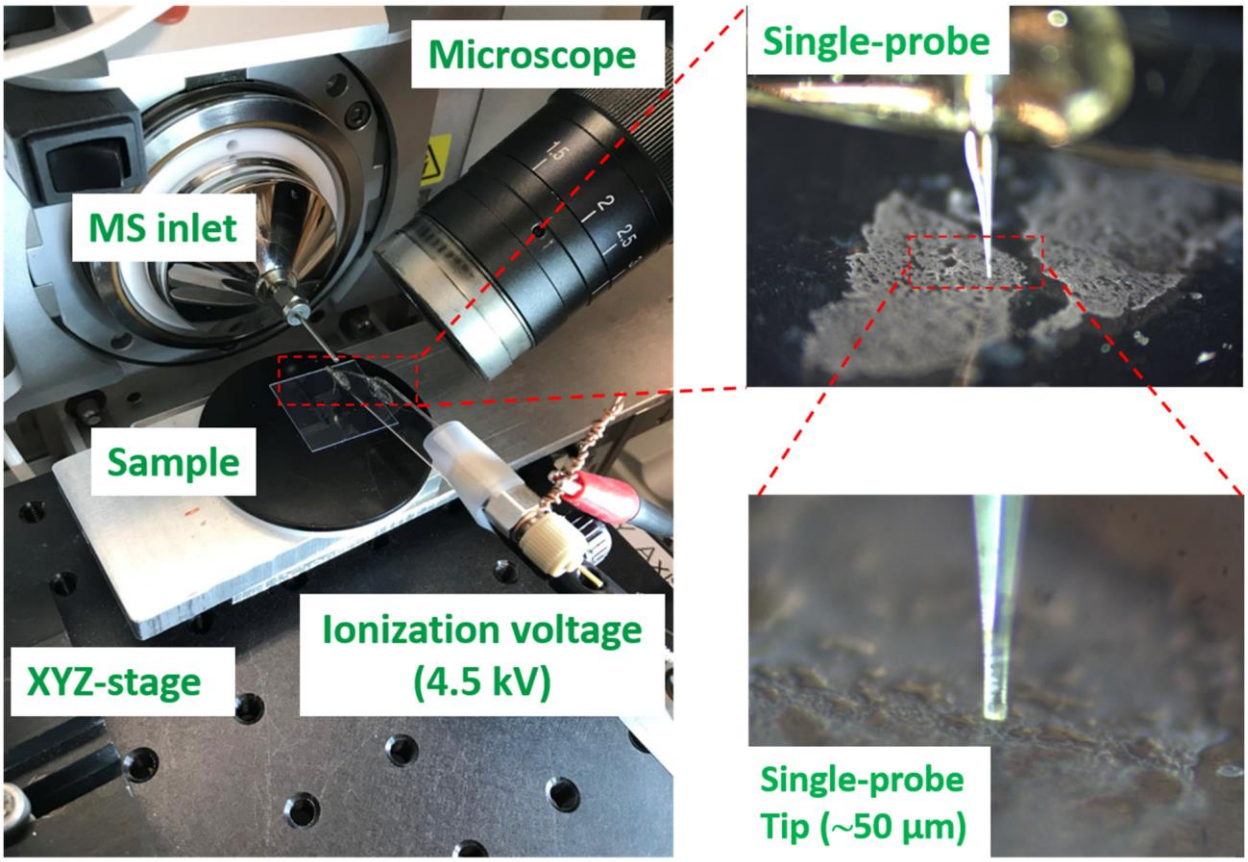


Figure S4-1. The Single-probe MSI experimental set-up with key components labeled.

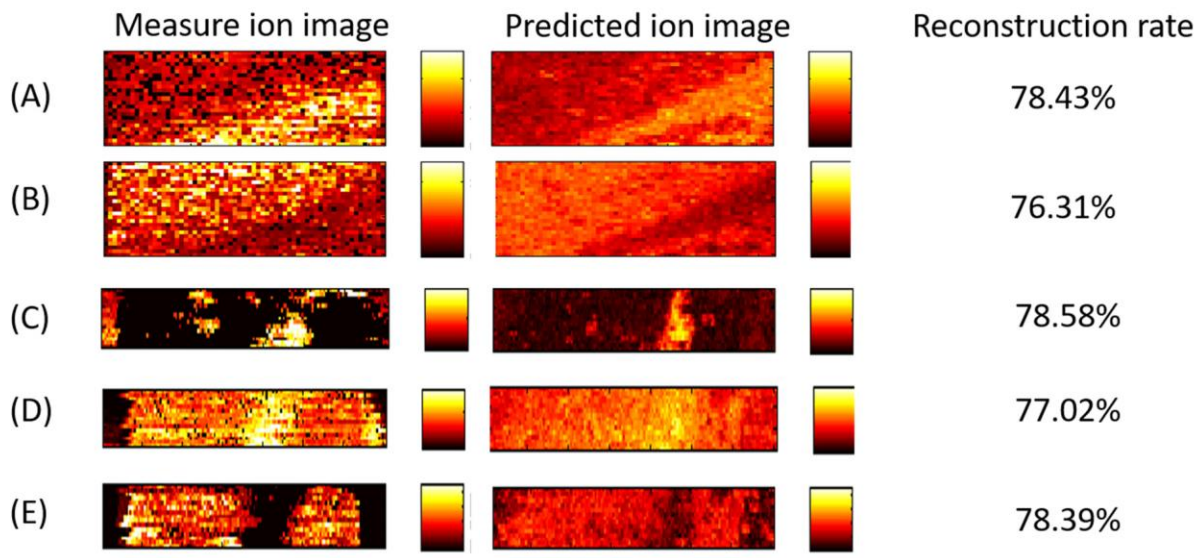


Figure S4-2. The reconstruction rate obtained from measured MS images (left) and predicted images (right).

(A) $[PA(O\text{-}32:0) - 2H]^{2-}$ (m/z 309.2514), (B) $[Dodecenoic\ acid + K - 2H]^{-}$ (m/z 251.1053), (C) $[LPC(18:0) + H]^{+}$ (m/z 524.3696), (D) $[PC(34:1) + H]^{+}$ (m/z 760.5851), and (E) $[PC(38:6) + K]^{+}$ (m/z 844.5218). All metabolites were identified using MS/MS from tissue slice, and results were compared with METLIN.

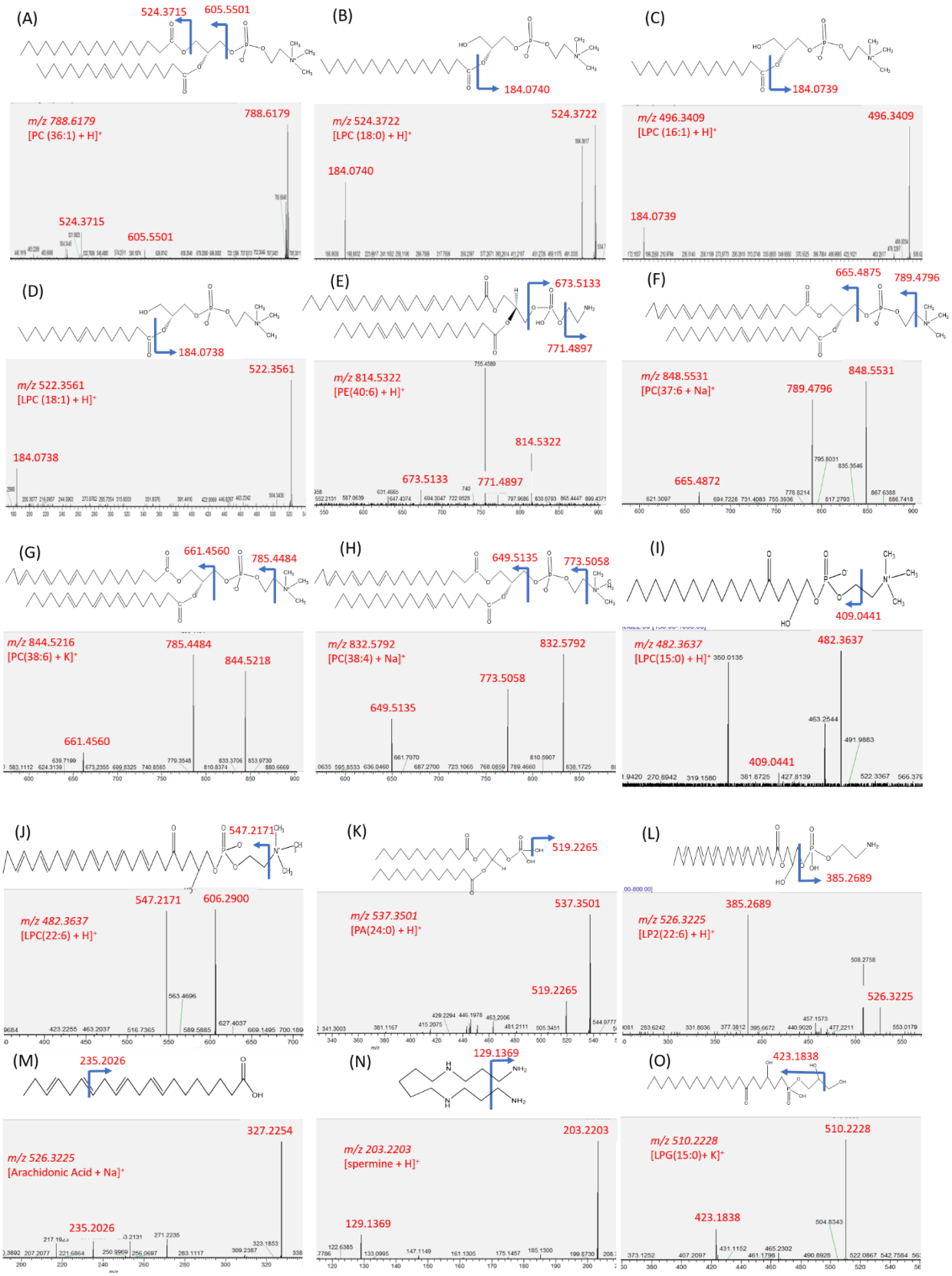


Figure S4-3. MS/MS analysis of selected metabolites (in the positive ion mode).

(A) $[PC(36:1) + H]^+$, (B) $[LPC(18:0) + H]^+$, (C) $[LPC(16:1) + H]^+$, (D) $[LPC(18:1) + H]^+$, (E) $[PE(40:0) + H]^+$, (F) $[PC(37:6) + Na]^+$, (G) $[PC(38:6) + K]^+$, (H) $[PC(38:4) + Na]^+$, (I) $[LPC(15:0 + H)]^+$, (J) $[LPC(22:6) + H]^+$, (K) $[PA(24:0) + H]^+$, (M) [Arachidonic acid + $Na]^+$, (N) [Spermine + $H]^+$, and (O) $[LPG(15:0) + K]^+$(All MS/MS spectra were compared with online databases METLIN).

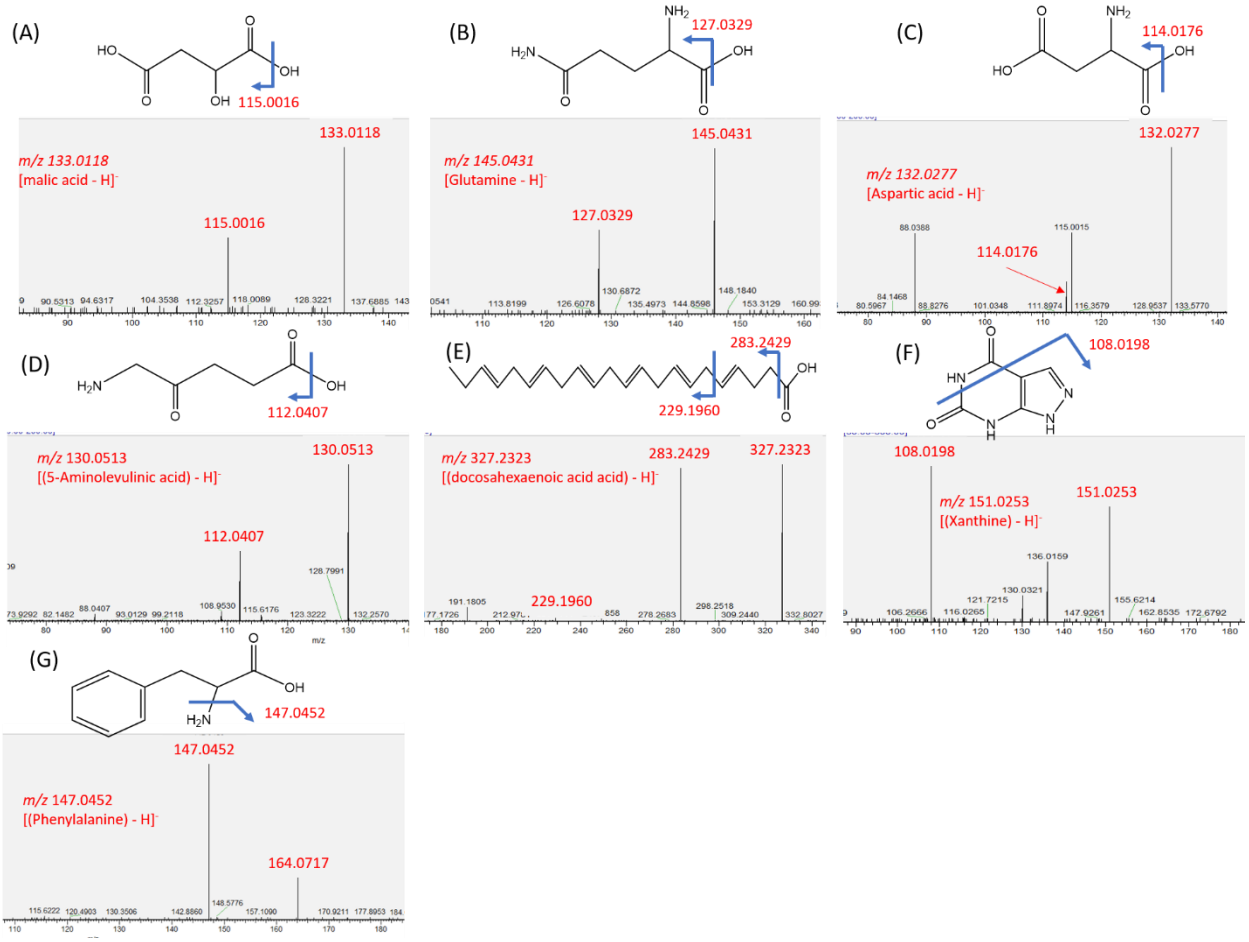


Figure S4-4. MS/MS analysis of selected metabolites (in negative ion mode).

(A) [malic acid - $H]^-$, (B) [glutamine - $H]^-$, (C) [asparic acid - $H]^-$, (D) [5-aminolevulinic acid - $H]^-$, (E) [docosahexaenoic acid - $H]^-$, (F) [Xanthine - $H]^-$, and (G) [Phenylalanine - $H]^-$. (All MS/MS spectra were compared with online databases METLIN).

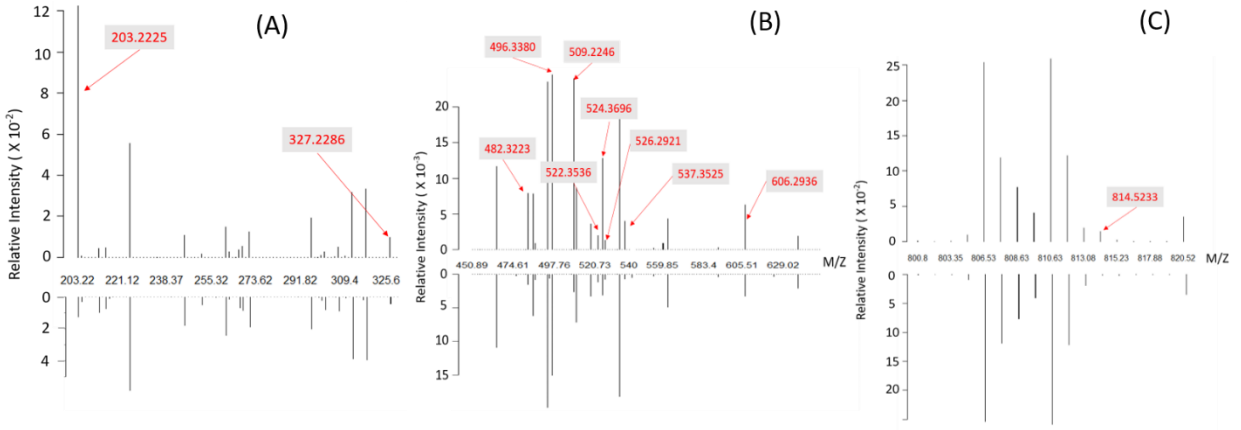


Figure S4-5. Zoomed in averaged spectra of Figure 4-5B and metabolites which have significant differences between plaque and its surrounding region in plaque 1.

(A) Zoomed spectra from 200 m/z to 330 m/z. (B)zoomed spectra from 450 m/z to 630 m/z. (C) Zoomed spectra from 800 m/z to 820 m/z.

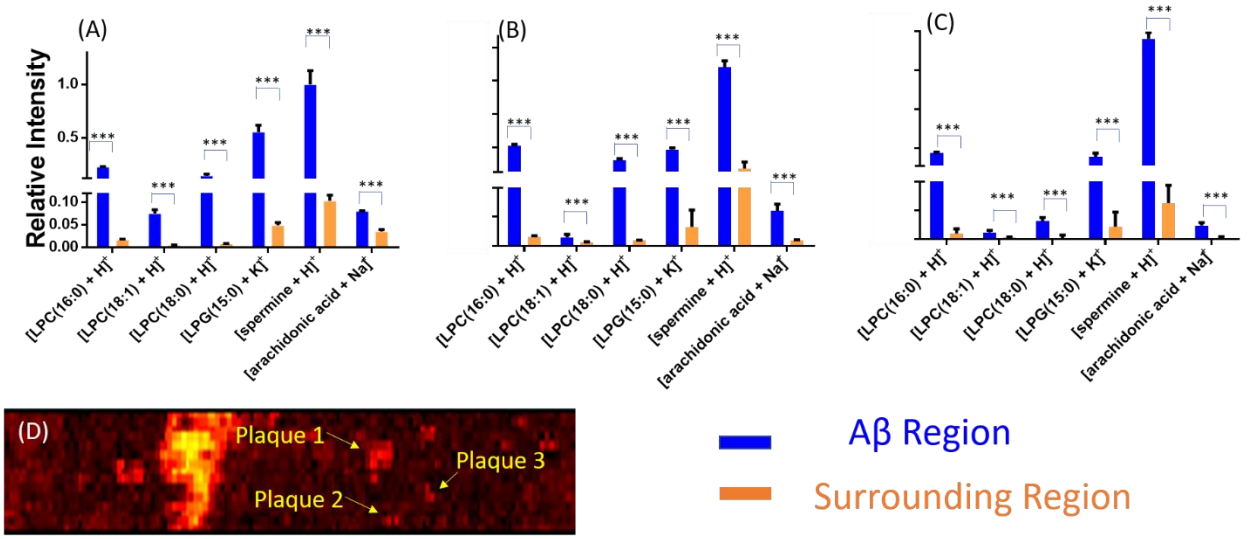


Figure S4-6. Metabolites significantly different abundances between Aβ plaques and their surrounding regions.

Results were obtained from (A) plaque 1, (B) plaque 2, and (C) plaque 3 as shown in (D). (D) Fused image and three selected plaques. All metabolites were identified using the tandem mass spectra on tissue slice, and results were compared with database METLIN. (From t-test: ***, $P < 0.001$)

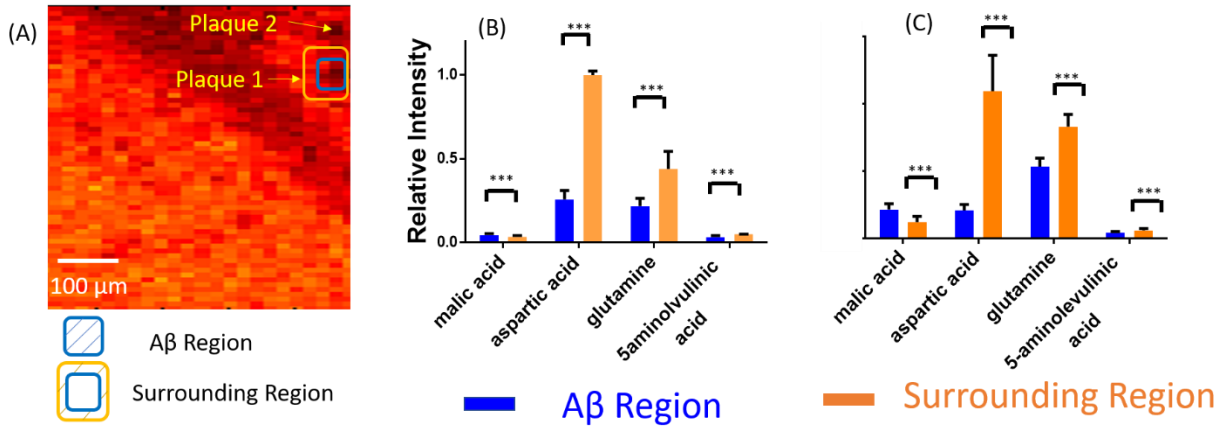


Figure S4-7. Pixels selection and ion relative intensity comparison.

(A) Pixels selection of Aβ plaques and their surrounding areas using the fused images. (B) and (C) Metabolites possessing significantly different abundances between Aβ plaques and their surrounding regions measured from plauqe1 and plaque 2. All metabolites were identified using the tandem mass spectra on tissue slice, and results were compared with database METLIN. (From t-test: ***, < 0.001)

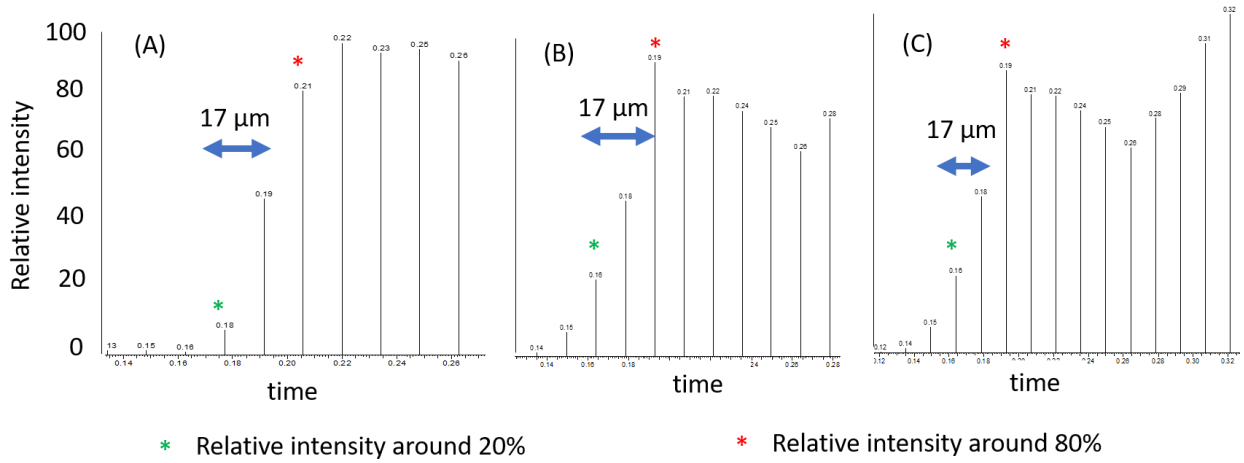


Figure S4-8. Characterization of the spatial resolution of MS images.

The relative ion intensities of ion $[PC(36:1) + H]^+$ in three representative line scans in a MS image (Figure 4-3). The spatial resolution was determined on the distance with the relative ion intensities change from 20% to 80%.

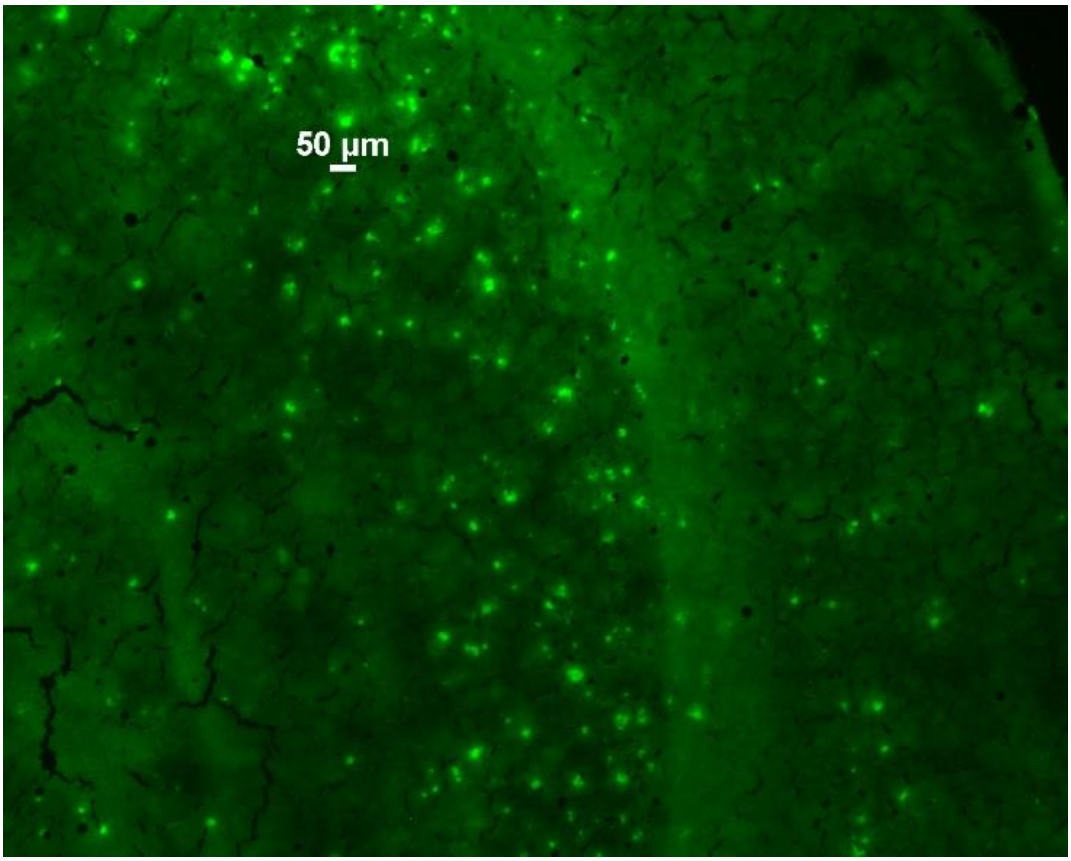


Figure S4-9. Stained slice of mouse brain with Alzheimer Disease.

The sizes of most plaques (green bright dots) are smaller than 50µm.

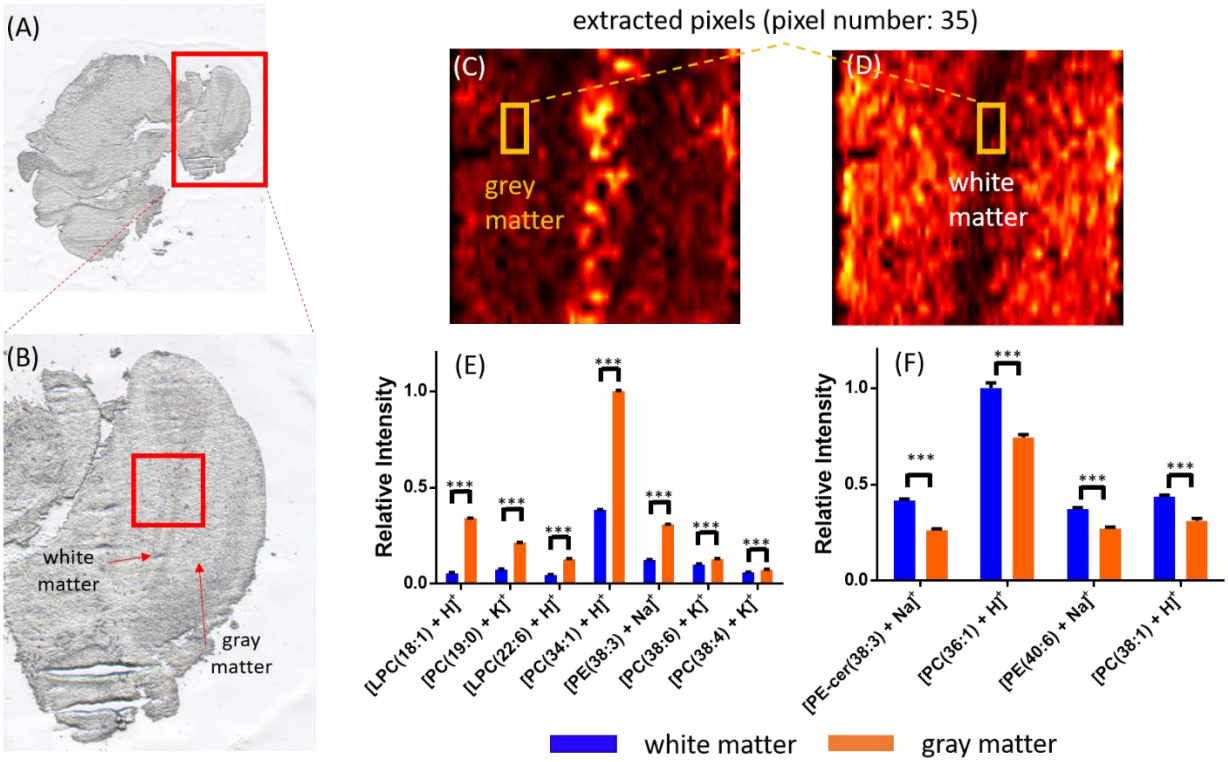


Figure S4-10. Experimental results from the control group.

Bright-field microscopy image of (A) a control mouse brain and the zoomed-in region. MS images of (C) $[PC(36:1) + H]^+$ and (D) $[PC(34:1) + H]^+$. Metabolites with relatively higher abundances in (E) the gray matter and (F) white matter.

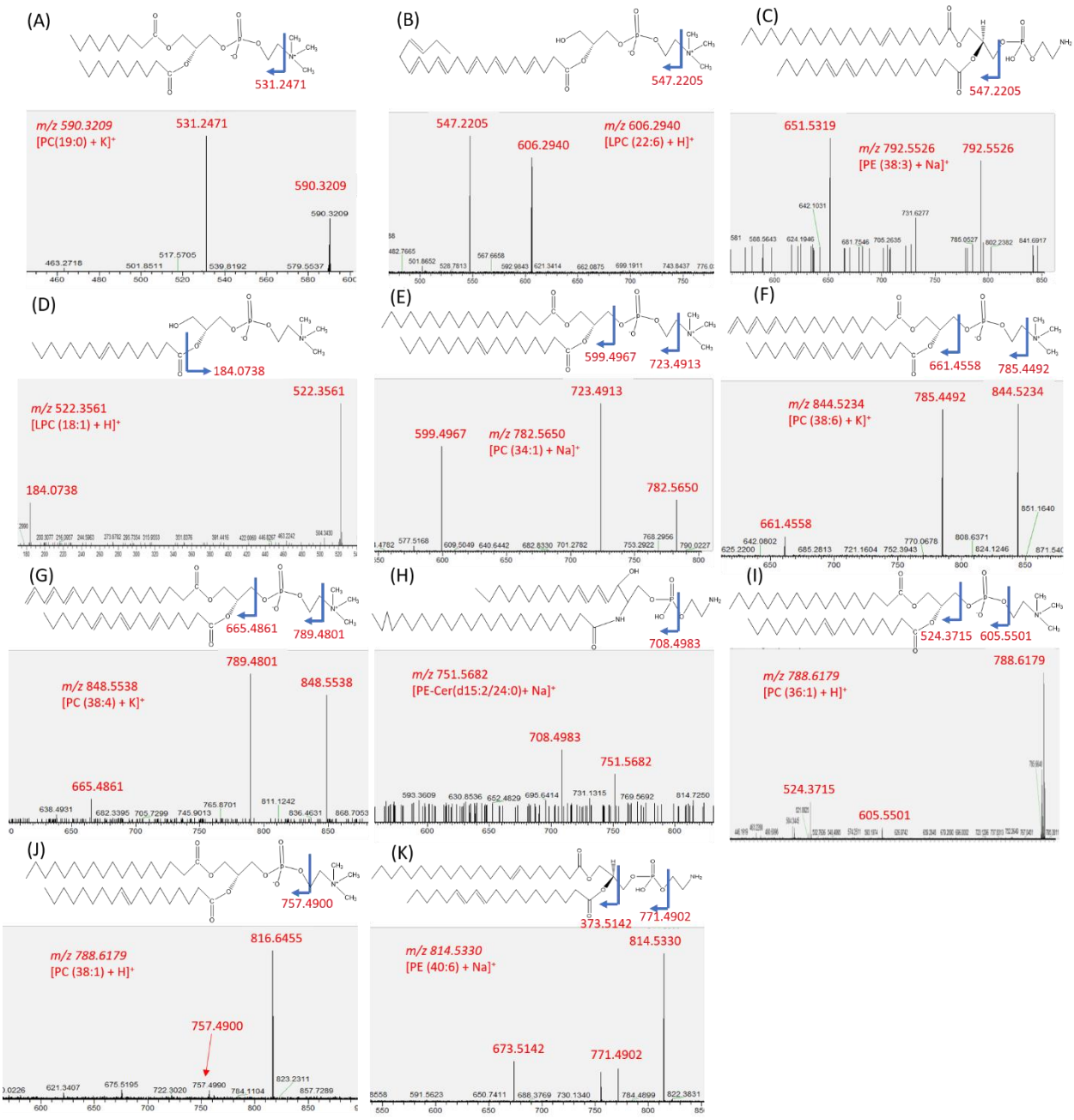


Figure S4-11. MS/MS analysis of selected metabolites in Figure S4-10 (positive ion mode).

(A) $[PC(19:0) + k]^+$, (B) $[LPC(22:6) + H]^+$, (C) $[PE(38:3) + Na]^+$, (D) $[LPC(18:1) +H]^+$, (E) $[PC(34:1) + Na]^+$, (F) $[PC(38:6) + K]^+$, (G) $[PC(38:4) + K]^+$, (H) $[PE\text{-}Cer(d15:2/24:0) + Na]^+$, (I) $[PC(36:1 + H)]^+$, (J) $[LPC(38:1) + H]^+$, (K) $[PE(40:6) + Na]^+$ (All MS/MS spectra were compared with online databases METLIN).

Table S4-1. Pixel selection in Aβ plaques and their surrounding regions from fused images.

Regions	Positive Ion Mode			Negative Ion Mode	
	1	2	3	1	2
Aβ	15	15	15	15	15
Surrounding	30	30	30	20	20

Table S4-2. Metabolites with significantly different abundances in Aβ plaques and their surrounding regions (positive ion mode).

m/z	Assignment*	Relative Intensity (plaque)	Relative Intensity (surrounding)
203.2225	$[Spermine + H]^+$	1.2×10^{-1}	1.2×10^{-2}
327.2286	$[arachidonic\ acid + H]^+$	9.8×10^{-2}	4.3×10^{-2}
482.3241	$[LPC(15:0) + H]^+$	7.8×10^{-3}	1.5×10^{-3}
496.3381	$[LPC(16:0) + H]^+$	2.4×10^{-2}	1.5×10^{-2}
509.2242	$[LPG(15:0) + H]^+$	2.3×10^{-2}	2.6×10^{-3}
522.3536	$[LPC(18:1) + H]^+$	1.9×10^{-3}	1.2×10^{-3}
524.3696	$[LPC(18:0) + H]^+$	1.2×10^{-2}	3.1×10^{-3}
526.2912	$[LPE(22:6) + H]^+$	1.3×10^{-2}	8.0×10^{-3}
537.3532	$[PA(24:0) + H]^+$	3.9×10^{-3}	7.2×10^{-4}
606.2936	$[LPC(22:6) + H]^+$	6.2×10^{-3}	3.3×10^{-3}
814.5233	$[PE(40:0) + H]^+$	1.2×10^{-2}	1.4×10^{-3}

(LPC: Lysophosphatidylcholines, LPG: lysophosphatidylglycerol, LPE: lysophosphatidylethanol, PA: Phosphatidic acids, PE: phosphatidylethanol)

Table S4-3. Metabolites with significantly different abundances in Aβ plaques and their surrounding regions (negative ion mode).

m/z	Assignment*	Relative Intensity (plaque)	Relative Intensity (surrounding)
133.0118	$[malic\ acid - H]^-$	8.5×10^{-3}	1.08×10^{-2}
145.0431	$[Glutamine - H]^-$	8.4×10^{-4}	1.2×10^{-3}
132.0277	$[Aspartic\ acid - H]^-$	1.6×10^{-2}	6.5×10^{-2}
130.0513	$[(5\text{-}Aminolevulinic\ acid) - H]^-$	1.6×10^{-3}	2.2×10^{-3}
327.2323	$[(docosahexaenoic\ acid) - H]^-$	1.4×10^{-2}	2.2×10^{-2}
151.0253	$[(Xanthine) - H]^-$	4.0×10^{-3}	1.2×10^{-2}
147.0452	$[(Phenylalanine) - H]^-$	2.3×10^{-3}	2.8×10^{-3}

*The assignment was based on the tandem mass spectra (MS/MS) and compared with online database METLIN

Table S4-4. Metabolites illustrated in Figure 4-3 and Figure4-7.

m/z	Assignment*	Ion mode
788.6197	$[PC(36:1) + H]^+$	Positive
848.5534	$[PC(37:6) + H]^+$	Positive
844.5218	$[PC(38:6) + K]^+$	Positive
832.5792	$[PC(38:4) + H]^+$	Positive
251.1065	$[Dodecenoic\ acid + K - 2H]^-$	Negative
309.2514	$[PA(O\text{-}32:0) - 2H]^{2-}$	Negative

PC: phosphatidylcholines