Accepted Manuscript

MassImager: A software for interactive and in-depth analysis of mass spectrometry imaging data

Jiuming He, Luojiao Huang, Runtao Tian, Tiegang Li, Chenglong Sun, Xiaowei Song, Yiwei Lv, Zhigang Luo, Xin Li, Zeper Abliz

PII: S0003-2670(18)30245-9

DOI: 10.1016/j.aca.2018.02.030

Reference: ACA 235746

To appear in: *Analytica Chimica Acta*

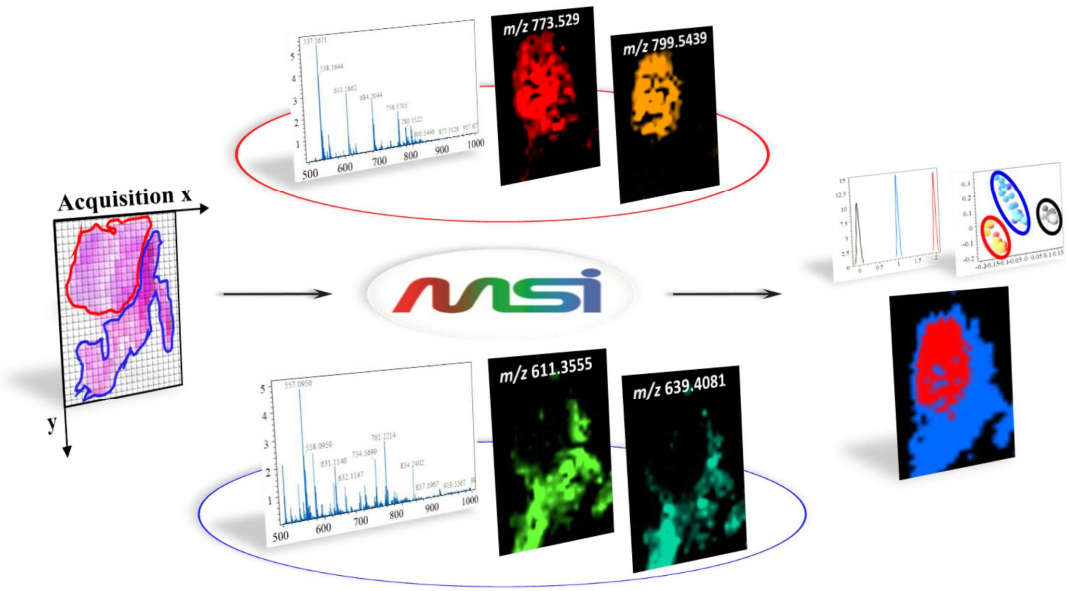
Received Date: 23 November 2017

Revised Date: 4 February 2018

Accepted Date: 13 February 2018

Please cite this article as: J. He, L. Huang, R. Tian, T. Li, C. Sun, X. Song, Y. Lv, Z. Luo, X. Li, Z. Abliz, MassImager: A software for interactive and in-depth analysis of mass spectrometry imaging data, *Analytica Chimica Acta* (2018), doi: 10.1016/j.aca.2018.02.030.

Graphical abstract



MassImager: A software for interactive and in-depth analysis of mass spectrometry imaging data

Jiuming He,[a,†] Luojiao Huang,[a,†] Runtao Tian,[b] Tiegang Li,[a] Chenglong Sun,[a] Xiaowei Song,[a] Yiwei Lv, [a] Zhigang Luo,[a] Xin Li,[a] and Zeper Abliz[*,a,c]

[a] State Key Laboratory of Bioactive Substance and Function of Natural Medicines, Institute of Materia Medica, Chinese Academy of Medical Sciences and Peking Union Medical College, Beijing 100050, China

[b] Chemmind Technologies Co., Ltd., Beijing 100085, China

[c] Center for Imaging and Systems Biology, Minzu University of China, Beijing 100081, China

[†] These authors contributed equally to this work.

[*] Corresponding author: Zeper Abliz

Address: No.1 Xian Nong Tan Street, Beijing 100050, China

Tel.: 86-010-63165218

Fax: 86-010-63165218

E-mail: zeper@imm.ac.cn; zeper@muc.edu.cn

Abstract

Mass spectrometry imaging (MSI) has become a powerful tool to probe molecule events in biological tissue. However, it is a widely held viewpoint that one of the biggest challenges is an easy-to-use data processing software for discovering the underlying biological information from complicated and huge MSI dataset. Here, a user-friendly and full-featured MSI software including three subsystems, Solution, Visualization and Intelligence, named MassImager, is developed focusing on interactive visualization, *in-situ* biomarker discovery and artificial intelligent pathological diagnosis. Simplified data preprocessing and high-throughput MSI data exchange, serialization jointly guarantee the quick reconstruction of ion image and rapid analysis of dozens of gigabytes datasets. It also offers diverse self-defined operations for visual processing, including multiple ion visualization, multiple channel superposition, image normalization, visual resolution enhancement and image filter. Regions-of-interest analysis can be performed precisely through the interactive visualization between the ion images and mass spectra, also the overlaid optical image guide, to directly find out the region-specific biomarkers. Moreover, automatic pattern recognition can be achieved immediately upon the supervised or unsupervised multivariate statistical modeling. Clear discrimination between cancer tissue and adjacent tissue within a MSI dataset can be seen in the generated pattern image, which shows great potential in visually *in-situ* biomarker discovery and artificial intelligent pathological diagnosis of cancer. All the features are integrated together in MassImager to provide a deep MSI processing solution at the *in-situ* metabolomics

level for biomarker discovery and future clinical pathological diagnosis.

Keywords: mass spectrometry imaging, data processing software, interactive visualization, *in-situ* biomarker discovery, artificial intelligent pathological diagnosis

1. Introduction

Mass spectrometry imaging (MSI) was developing fast with the invention of various *in-situ* ionization sources and their combinations with different mass analyzers for specific analytical demands [1-5]. The greatest power of MSI lies in its ability to simultaneously acquire the molecule features and their morphological distributions in one single MSI experiment [6]. Nowadays, it has become the most promising analytical tool in the field of life science [7], metabolomics [8-10], proteomics [11-13], drug development [14, 15], clinical diagnosis [16-19] and molecular mechanism research [20, 21]. Newly developed ambient ionization techniques allow MSI to be performed in an ambient environment with minimal sample preparation. Samples can be analyzed in nearly native states and compounds with molecular weights under 2000 Da are mainly studied [22].

Data processing is a major part of an integral MSI experiment. Currently, more and more MSI instruments pursue higher spatial and mass resolution. For example, the spatial resolution of near-infrared femtosecond-laser induced ionization based on the ambient technique can reach 10 μm at the cellular level [23]. The maximum mass resolving power of mass spectrometry was reported to be more than 100000. These two factors both influence the amount of MSI data which can be considerably large to dozens of gigabytes. Besides, other characteristics of MSI data including complexity, compatibility and high dimensionality jointly compose great challenges to data processing. Although, it is easy to obtain the distributions of ions of interest in biological tissue due to the acquisition of thousands of ion images in one scan.

Uncovering the underlying biological information from complicated and huge MSI dataset may have more practical significance. More emphasis should be placed on the associations between different ions [24]. So that these inherent bio-information could be useful in clinical diagnosis as an assistant histological tool in discovering disease molecular mechanisms or targets of drug action [25]. Automatic high-throughput data processing and intelligent data mining techniques are quite necessary for deep biological analysis.

To meet the urgent need for MSI data processing, lots of software packages are being continuously developed. Most commercial software such as FlexImaging (Bruker), ImageQuest (Thermo Scientific) and High Definition Imaging (Waters) support their own proprietary data format, which limits the flexibility of data analysis. Whereas, freeware and open source software have higher levels of compatibility by introducing the open-data format imzML [26-28]. Other formats, such as xml, mzXML, Analyze 7.5, ASCII and NetCDF, are also welcome [28-31]. However, open source software packages, such as OmniSpect, Cardinal, are executed under MATLAB or R platforms, which are difficult for beginners without programming background to perform data analysis efficiently [31, 32]. Freeware software are more practical in direct image visualization with additional processing functions. Recently, MSiReader v1.0 was updated as a standalone version with many new improvements, but without the associations between different ions and multivariate statistical analysis [33]. In this way, another freeware software, SpectralAnalysis, is superior to the others. It shows great performance in spectrally differentiating micro-regions from raw data through

preprocessing to multivariate analysis, for data sets acquired from single experiments to large multi-instrument, multimodality, and multicenter studies [34].

Our team developed a new commercial software focusing on MSI data analysis called MassImager. It contains three subsystems, namely, 'MassImager Solution', 'MassImager Visualization' and 'MassImager Intelligence'. The core design is aimed at user-friendly design, high throughput, instrument-independence, interactive visualization, various multivariate statistical analysis and pattern recognition of large MSI dataset. Until now, many ambient MSI datasets from bio-tissue have been processed using MassImager in our lab [35-38]. On the one hand, due to its high-throughput and high-speed of data processing, it plays a crucial role in accurately visualizing the process of drug absorption, distribution, metabolism, excretion and toxicity, which brings great convenience to drug discovery and research. On the other hand, interactive ROIs analysis and deep multivariate statistical data-mining for biomarker discovery makes it an important tool for *in-situ* metabolomics study. The following AFADESI-MSI experiments are provided as examples to illustrate the powerful functionalities and usage of the MassImager software.

2. Material and methods

2.1. Solvents and Reagents

The HPLC-grade organic solvent methanol was purchased from Merck (Muskegon, MI). Purified water was obtained from Wahaha (Hangzhou, China). Formic acid was

purchased from Sigma-Aldrich (St. Louis, MO). The Wistar rats were purchased from the Institute of Medical Laboratory Animals, Chinese Academy of Medical Sciences (Beijing, China).

2.2. Sample preparation

'MSI' writing sample: Letters of 'MSI' were written on a clean slide respectively in red, blue and black marker (silica, 25×75×1mm). The sample can be analyzed immediately without extra preparation.

Tissue section: According to the relevant protocol and requirements, tissue sections of the rat brain were collected from the Institute of Medical Laboratory Animals, Chinese Academy of Medical Sciences. The post-operative tissue sample of case '983322T' was collected in the Peking Union Medical College Hospital. Study protocols were approved by the Ethics Review Committee of the Peking Union Medical College Hospital and the informed consent form was also signed by patient involved in the study. Tissue sample was fresh without been soaked in formalin. It was snap-frozen in liquid nitrogen after operation and soon transferred to -80□ refrigerator. Then, this frozen tissue was cut into 8 μm sections at −20°C in a cryomicrotome (CM 3050S, Leica Microsystems, Wetzlar, Germany) and thaw-mounted onto microscope glass slides (Superfrost Plus slides, Thermo Fisher Scientific, USA). Next, pathological diagnosis was achieved by H&E staining on adjacent cryosections. Tissue sections were stored at −80°C until they were analyzed under ambient environment. Prior to MSI analysis, these sections were dried in a

vacuum desiccator till room temperature for approximately 1-2 h.

2.3. AFADESI-MSI experiment

Experiments were performed on an air-flow-assisted desorption electrospray ionization (AFADESI)-MSI platform, which is illustrated in our previous work [36, 39]. The AFADESI-MSI platform is mainly equipped with a home-built AFADESI ion source and a Q Orbitrap mass spectrometer (Q Exactive, Thermo Scientific, Bremen, Germany). The former can achieve spatial resolution up to 100 μm, meanwhile the latter's mass resolution sets at 70000. All the MSI data were acquired with full MS scan mode ranging from *m/z* 100 to 1000 in the positive mode. Other details of experiment parameters can be seen in the Table S1.

2.4. Software development

MassImager is defined as a standalone software which provides an automatic processing pipeline for MSI data. It is compatible with the Microsoft Windows XP/Vista/7/8/10 operating system and has two editions for Win32 and Win64 platforms. The recommended CPUs are quadcore Intel® Core™ processors or equivalent and the minimum memory is 2 GB. For faster calculations and data caching, 16 GB is a better choice. Computers with discrete graphics and at least 10 GB of disk space is necessary for running the software smoothly. The key technical points of MassImager are listed as follows: (1) Optimization based on C++ programing language together with multi-threading acceleration calculation for MSI big-data processing. (2) High performance MSI image reconstruction and

158 visualization powered by MSI processing algorithms. (3) Various chemometric
159 algorithms have been introduced to achieve the mass image artificial intelligence
160 technique in terms of pattern recognition modeling.
161 The demo software is freely available for download at the project webpage:
162 http://www.chemmind.com/en/support_download.html. Three demo MSI datasets are
163 also provided for user experience, including a handwriting 'MSI' ink sample, a human
164 thyroid cancer tissue section and a rat brain tissue section (sagittal plane).
165

166 ## 3. Results and discussion

167 ### 3.1. MassImager Solution: data loading and image reconstruction

168 A solution is created especially for each case. In an ambient mass spectrometry
169 imaging experiment, the sample on the slide is scanned line by line, and then, the
170 acquired profile data of each line are saved as a raw data file (Figure 1A). Raw data
171 files can be con-verted into the appropriate formats before being loaded into a new
172 solution. Several data formats are supported in the software, including ANDI,
173 mzXML, Matlab and ASCII (Figure 1B) [40, 41]. Based on these formats,
174 MassImager is compatible with current commercial mass spectrometer manufacturers,
175 such as AB SCIEX, Thermo Scientific, Bruker, Waters and Agilent. For each
176 manufacturer, the suitable data formats are listed in Table S2.
177 Upon the conversion of consecutive raw data files, an importing sequence must be
178 created (Figure 1C). During this process, the first step is to choose the storage path of
179 the initial line file and confirm the total number of line files. Moreover, the physical
180 length and height of a rectangular region covered by the tissue sample also need to be

entered to reflect the actual size of the image. Then, the analytical scan numbers and sampling time can be checked automatically at the end of line sequence creation. It is also necessary to transform continuous profile data into centroid data in consideration of the speed of reconstruction and convenience of subsequent data analysis. Meanwhile, peak detection and peak integration are implemented in the part of integration parameter setting, and the integration result can be previewed in real time for appropriate adjustments (Figure 1D). Four types of parameters are available for optimum reconstruction (see supporting information). By completing all the settings, the consecutive batch file can be imported into the database from the initial file to the last file at one time. Image reconstruction is immediately completed after the processing of sequence file. The final generated image presents the map of ions covering the whole mass range in the MSI experiment (Figure 1E).

The performance of MassImager in loading and processing a large dataset is remarkable. For one thing, a sequence can be automatically imported after checking the parameters. Files are continuously parsed and serialized to a local database and then stored in a cache file. Since the database file can be shared freely, as long as the software is correctly installed, the data can be recorded and shared on multiple computers. For another, the average speed for initial import and construction of a large data cache is tested around 0.058 GB/s with the following computer specification: Windows 7 Professional configuration and an Intel® Core™ i3-3220U CPU, RAM 6.00GB and 32-bit operating system. (Table S3) Then, reloading the data directly from local database takes less than 10 seconds. This makes it possible for

high throughput and quick MSI data analysis.

3.2. MassImager Visualization: visual processing and image analysis

MassImager provides powerful tools for image visualization along with a user-friendly MS office-like interface. These practical tools can be classified into two categories: visual processing and image analysis, in terms of different operating purposes.

For the best image quality, a series of visual processing settings is available in MassImager (Figure S1). First, the software provides multiple forms of MSI view. Users can choose a single view, horizontal views, vertical views or 2×2 views. Each view can be independently displayed and processed with MSI data in the same solution, making comparisons among multiple ions easier to do. Both a single *m/z* and the mass range can be manually or mouse-click entered into the channel. In total, nine channels can be displayed independently or simultaneously. It is worth mentioning that the superposition of multiple ion channels is called ‘Overlap EII’ in MassImager. The overlapping mode includes layer overlapping and layer blending. Layer blending enables different layers are fused together based on their own RGB values or color bars (Figure S2). This special feature makes it possible for paralleled presentation of different molecular distributions correlated with the specific sample region. Several basic functions, such as normalization of each channel, channel opacity, background color, and a color bar, are provided for user modification. Optical image overlay is another outstanding feature to facilitate the understanding of the distribution of

characteristic ions in the micro-regions and helps to extract information in the specific region more precisely as well. (see supporting information).

Background subtraction is the essential part of visual processing, considering the serious background interferences. The illustration of background subtraction is shown in Figure S3. Two methods of deduction are optional: one is complete deduction another is proportional deduction. There are also parameters for user-defined visualization. Normalization is a unified standardization for every mass pixel intensity to reduce the variation caused by the various properties of tissue structure [42]. Three types of normalization methods are proposed in MassImager (Figure S4). They are the global image, current selected region and intensity threshold. The intensity threshold is based on the customized threshold as the highest intensity of the full image. It particularly benefits the visualization of ions with low intensity among large intensity ranges. The tolerance for each *m/z* in the ion extraction channel determines the display of the corresponding ion distribution image. The minimum value can be set to 0.001. These self-defined visual processing parameters can be saved as a template to be applied to a new case for the next experiment.

To acquire the biochemical information contained in a particular area, the region of interest (ROI) must be indicated for image analysis. MassImager offers three types of predefined regions with different pixel areas (1*1, 3*3, 5*5) and also a free-hand drawing option for arbitrary regions. Six ROIs with the mark of different colors can be displayed at the same time. The total or average mass spectrum for the marked region is generated at once after the ROI is defined. The marked color of the ROI and

the number of pixels can also be recorded in the title of the relevant mass spectrum.

The abundant information behind an ROI can be processed in many ways. If the ROI and its mass spectrum are required for the next analysis, users can save them to the local mass spectra library in each solution. They can also be added to the source data for multivariate statistical analysis. Each ROI generates a peak list that contains the information of every peak, such as centroid mass and peak intensity. This list data then can be exported into a new spreadsheet of Excel or a text file for further analysis in third-party software. One of the most attractive features of MassImager is the interactive processing between ion images and mass spectra (Figure 2). On the one hand, as mentioned above, one can immediately obtain the corresponding mass spectrum by marking the ROI. On the other hand, by double-clicking the peak in the mass spectrum, the associated ion image can be displayed at once. This allows rapid screening for characteristic ions in the qualitative analysis. Significant ions or specific *m/z* ranges can be recorded in the ion channels for direct display the next time.

3.3. MassImager Intelligence: multivariate statistical analysis and pattern recognition

The module 'MassImager Intelligence' is a more advanced feature of the software that is designed to perform multivariate statistical analysis and mass image pattern recognition. With its emphasis on screening *in-situ* metabolomic features or biomarkers and automatic image recognition, it exceptionally facilitates intelligent diagnostic visualization of a tissue sample. For a better understanding, the following

discussion will be divided into two sections. The first section covers deep multivariate statistical analysis of ion maps and identification of biomarkers. The second section focuses on the technique of automatic image recognition for a tissue sample. A tissue section of lung cancer was taken as the example in the following sections.

3.3.1. Multivariate Statistical Analysis

Multivariate statistical analysis is recognized as a classical method to realize efficient mining of information and discovery of *in-situ* markers. MassImager provides three classical multivariate statistical methods, which include principle component analysis (PCA), partial least squares discriminant analysis (PLS-DA) and orthogonal partial least squares discriminant analysis (OPLS-DA). Among these, PCA belongs to unsupervised recognition, which does not require previous information about the analyzed samples, while the other two belong to supervised recognition, using histological or optical images as the reference information [43, 44]. A mass of mass spectrometric data can be projected down on a few latent variables (LV) by these projection methods. Multiple regression is then constructed between each factor X (LVs) and result Y. Finally, a score will be calculated for each latent variable with a value representing its contribution to the overall classification model.

In the analysis of MSI data of a lung cancer tissue sample, *in-situ* extraction was completed first based on the superimposition of the H&E staining image and the ion image. Several ROIs of tumor tissue (group I) and adjacent normal tissue (group II) are outlined and then added to the data source. Due to the premise of knowing the histological information, PLS-DA was chosen as the multivariate statistical method.

After filtration of ions, the data matrix can be reduced for fast computation. Given the large intensity ranges, we prefer to use the scaling method of log transformation since it pays more attention to the low abundance ions which have the same discriminant ability as the high abundance ions. Finally, five types of resulting figures are achievable for model interpretation (Figure 3B). The score plot directly shows the first two latent variables (score of LV1: 5.95%, score of LV2: 5.14%) in two dimensions could achieve good clustering between tumor tissue and normal tissue. Also, the degree of model discrimination can be described intuitively in the classification plot. Its horizontal axis represents class prediction value of each ROI in this model, while the vertical axis indicates how well each classification group meets a gaussian distribution, in which smaller variance means a more concentrated data distribution in each group, thus leading to a thinner bell curve. The loading plot shares a similar principle with the weighted (log-transform) loading plot. Both can present the significant contributing ions. It is more obvious in the weighted loading plot that the loading dots on the left part refer to ions contributing to tumor tissue, while those on the right part refer to ions contributing to normal tissue. Additionally, MassImager efficiently integrates the information of loading value, *m/z* and its intensity together in a mass loading plot. Each *m/z* has a loading value. The absolute value of loading represents the discriminant ability, and the positive or negative value refer to which class the ion contributes more to. The intensity of the ion is reflected in the form of the color bar. As we can see, most ions below *m/z* 850 are in high abundance, while almost all ions above *m/z* 850 are in low abundance. Here, we can rapidly screen for

tumor biomarkers by clicking on each *m/z* and observing its relevant ion map just as in the mass spectrum. We can also export the data list of the mass loading plot for intensive analysis such as t-Test and fold-change analysis.

Through this processing flow chart, potential biomarkers of lung cancer can be initially acquired. In our experiment, the ions of *m/z* 773.5290 and *m/z* 799.5439 are unique in tumor tissue, while the ions of *m/z* 611.3555 and *m/z* 639.4081 are particular to normal tissue. The image of the layer mixture explains the good complementarity between these representative ion markers (Figure 3D). From this perspective, representative ion markers that are unique to tumor or normal tissue can respectively form a marker set. The combination of both marker sets may facilitate the initial identification of clinical tumor. If needed, potential biomarkers can be automatically annotated by utilizing the package "pySM", a new bioinformatics tools of false discovery rate-controlled metabolite annotation for high-resolution MSI [45].

3.3.2. Automatic Pattern Recognition

Pattern recognition is a way to interpret tissue region with similar biochemical profile. A good model based on multivariate statistical methods is the core precondition of automatic pattern recognition. As an example for illustration, another new solution of the same lung cancer case was created, which added the glass slide group as a third background class. With PLS-DA chosen as the multivariate statistical method, a new model involving three regions of tumor, adjacent normal tissue and background was built according to the same workflow (Figure 3). The score plot and classification plot both show the good clustering between the three groups (Figure 4). Extra cross

validations were completed for internal model quality evaluation (see supporting information). Since the excellent model was tested, users can apply it to the whole tissue sample so that the characteristics of other regions in this tumor can be acquired. MassImager possesses the prominent capability of performing automatic image recognition analysis. Once this pretrained MSI model file based on metabolomic profiling was loaded, a pattern image of the whole tissue section was immediately generated in the mass image window after activating the MSI classification function. Three different representations of pattern images based on model loading #1, model loading #2 and the class prediction value are shown in Figure 4. Each representation focuses on different contributing factors. As can be seen, model loading #1 underlines the differentiation between tissue sample and glass slide background. Model loading #2 places an emphasis on the normal tissue. The class prediction value clearly describes the discrimination between all three groups. This representation is almost consistent with the H&E staining image shown above (Figure 3 A). The results indicate that as long as a good model is initially built, it is possible to perform pattern recognition of the whole tissue based on the representative ROIs selected. Therefore, to some extent, the technique of automatic image recognition may serve as a complementary and artificial intelligent visualization tool for pathological diagnosis.

4. Conclusion

MassImager was designed as a user-friendly software that enables researchers without intensive training of professional programing knowledge to perform the entire

workflow of MSI data processing easily. Through close integration of the three subsystems ‘MassImager Solution’, ‘MassImager Visualization’ and ‘MassImager Intelligence’, high-performance MSI data analysis can be carried out in real time when processing a large MSI dataset, including ion image reconstruction, visualization, ROI, multivariate statistical analysis and pattern recognition. The excellent capability makes it an immediate imaging tool.

Meanwhile, one remarkable innovation of MassImager is the interactive visualization between the ion images and mass spectra. With the combination of regional spectra calculation, it is beneficial for thorough analysis of ROIs and quick screening for direct region-specific biomarkers. Pattern recognition based on several multivariate statistical methods is another distinctive feature in MassImager. Our study indicated that, apart from deep mining for biomarkers, artificial intelligent recognition for pathological diagnosis can be achieved instantly within a tissue MSI dataset despite the existence of tumor heterogeneity.

Therefore, all the results demonstrated that MassImager is a generic, flexible, powerful tool and also an easy-to-use data processing software for discovering the underlying biological information from complicated and huge MSI dataset. These features are promising to provide a deep MSI processing solution at the *in-situ* metabolomics level for biomarker discovery and future clinical pathological diagnosis.

Acknowledgements

379 This work was supported by the National Natural Science Foundation of China (Grant
380 No. 81373370, 81773678), PUMC Youth Fund and the Fundamental Research Funds
381 for the Central Universities (Grant No. 3332015177), the National Instrumentation
382 Program (Grant No. 2016YFF0100304) , and the CAMS Innovation Fund for Medical
383 Sciences of Peking Union Medical College (Grant No. 2016-I2M-1-009).

384
385

386 **Reference**

387 [1] J.C. Vickerman, Molecular imaging and depth profiling by mass
388 spectrometry--SIMS, MALDI or DESI?, Analyst, 136 (2011) 2199-2217.
389 [2] T. Greer, R. Sturm, L. Li, Mass spectrometry imaging for drugs and metabolites, J.
390 Proteomics, 74 (2011) 2617-2631.
391 [3] B. Feng, J. Zhang, C. Chang, L. Li, M. Li, X. Xiong, C. Guo, F. Tang, Y. Bai, H.
392 Liu, Ambient Mass Spectrometry Imaging: Plasma Assisted Laser Desorption
393 Ionization Mass Spectrometry Imaging and Its Applications, Anal. Chem., 86 (2014)
394 4164-4169.
395 [4] X. Gong, Y. Zhao, S. Cai, S. Fu, C. Yang, S. Zhang, X. Zhang, Single Cell
396 Analysis with Probe ESI-Mass Spectrometry: Detection of Metabolites at Cellular and
397 Subcellular Levels, Anal. Chem., 86 (2014) 3809-3816.
398 [5] M.M. Nudnova, J. Sigg, P. Wallimann, R. Zenobi, Plasma ionization source for
399 atmospheric pressure mass spectrometry imaging using near-field optical laser
400 ablation, Anal. Chem., 87 (2015) 1323-1329.
401 [6] K. Chughtai, R.M.A. Heeren, Mass spectrometric imaging for biomedical tissue
402 analysis, Chem. Rev., 110 (2010) 3237-3277.
403 [7] J. Pól, M. Strohalm, V. Havlíček, M. Volný, Molecular mass spectrometry
404 imaging in biomedical and life science research, Histochem. Cell Biol., 134 (2010)
405 423-443.
406 [8] B. Rocha, B. Cillero-Pastor, G. Eijkel, A.L. Bruinen, C. Ruiz-Romero, R.M.A.
407 Heeren, F.J. Blanco, Characterization of lipidic markers of chondrogenic

differentiation using mass spectrometry imaging, Proteomics, 15 (2015) 702-713.

[9] D. Miura, Y. Fujimura, H. Wariishi, In situ metabolomic mass spectrometry imaging: recent advances and difficulties, J. Proteomics, 75 (2012) 5052-5060.

[10] M.S. Ferreira, D.N. de Oliveira, R.F. Gonçalves, R.R. Catharino, Lipid characterization of embryo zones by silica plate laser desorption ionization mass spectrometry imaging (SP-LDI-MSI), Anal. Chim. Acta, 807 (2014) 96-102.

[11] C. Lodeiro, J.L.C. Martínez, H.M. Santos, Tissue proteomics and imaging mass spectrometry, Proteomics, 14 (2014) 805-806.

[12] J.L. Moore, K.W. Becker, J.J. Nicklay, K.L. Boyd, E.P. Skaar, R.M. Caprioli, Imaging mass spectrometry for assessing temporal proteomics: analysis of calprotectin in Acinetobacter baumannii pulmonary infection, Proteomics, 14 (2014) 820-828.

[13] Z. Zhang, H. Ye, J. Wang, L. Hui, L. Li, Pressure-Assisted Capillary Electrophoresis Coupling with Matrix-Assisted Laser Desorption/Ionization-Mass Spectrometric Imaging for Quantitative Analysis of Complex Peptide Mixtures, Anal. Chem., 84 (2012) 7684-7691.

[14] D.F. Cobice, R.J.A. Goodwin, P.E. Andren, A. Nilsson, C.L. Mackay, R. Andrew, Future technology insight: mass spectrometry imaging as a tool in drug research and development, Br. J. Pharmacol., 172 (2015) 3266-3283.

[15] L.R. Rodrigues, D.N. de Oliveira, M.S. Ferreira, R.R. Catharino, *In situ* assessment of atorvastatin impurity using MALDI mass spectrometry imaging (MALDI-MSI), Anal. Chim. Acta, 818 (2014) 32-38.

[16] Y. Mohri, T. Mohri, W. Wei, Y.J. Qi, A. Martin, C. Miki, M. Kusunoki, D.G. Ward, P.J. Johnson, Identification of macrophage migration inhibitory factor and human neutrophil peptides 1-3 as potential biomarkers for gastric cancer, Br. J. Cancer, 101 (2009) 295-302.

[17] N. Goto-Inoue, T. Hayasaka, N. Zaima, K. Nakajima, W.M. Holleran, S. Sano, Y. Uchida, M. Setou, Imaging mass spectrometry visualizes ceramides and the pathogenesis of dorfman-chanarin syndrome due to ceramide metabolic abnormality in the skin, PLoS One, 7 (2012) e49519-49529.

[18] L.S. Eberlin, I. Norton, D. Orringer, I.F. Dunn, X. Liu, J.L. Ide, A.K. Jarmusch, K.L. Ligon, F.A. Jolesz, A.J. Golby, S. Santagata, N.Y.R. Agar, R.G. Cooks, Ambient mass spectrometry for the intraoperative molecular diagnosis of human brain tumors, Proc. Natl. Acad. Sci. U. S. A., 110 (2013) 1611-1616.

[19] A.L. Dill, L.S. Eberlin, D.R. Ifa, R.G. Cooks, Perspectives in imaging using mass spectrometry, Chem. Commun., 47 (2011) 2741-2746.

[20] Y. Takahashi, R. Sano, T. Nakajima, Y. Kominato, R. Kubo, K. Takahashi, N. Ohshima, T. Hirano, S. Kobayashi, T. Shimada, H. Tokue, S. Awata, S. Hirasawa, T. Ishige, Combination of postmortem mass spectrometry imaging and genetic analysis reveals very long-chain acyl-CoA dehydrogenase deficiency in a case of infant death with liver steatosis, Forensic Sci. Int., 244 (2014) e34-37.

[21] Y. Masaki, Y. Shimizu, T. Yoshioka, Y. Tanaka, K.-I. Nishijima, S. Zhao, K. Higashino, S. Sakamoto, Y. Numata, Y. Yamaguchi, N. Tamaki, Y. Kuge, The accumulation mechanism of the hypoxia imaging probe "FMISO" by imaging mass

spectrometry: possible involvement of low-molecular metabolites, Sci. Rep., 5 (2015) 16802-16811.

[22] C. Wu, A.L. Dill, L.S. Eberlin, R.G. Cooks, D.R. Ifa, Mass spectrometry imaging under ambient conditions, Mass Spectrom. Rev., 32 (2013) 218-243.

[23] Y. Coello, A.D. Jones, T.C. Gunaratne, M. Dantus, Atmospheric Pressure Femtosecond Laser Imaging Mass Spectrometry, Anal. Chem., 82 (2010) 2753-2758.

[24] J.M. Fonville, C.L. Carter, L. Pizarro, R.T. Steven, A.D. Palmer, R.L. Griffiths, P.F. Lalor, J.C. Lindon, J.K. Nicholson, E. Holmes, J. Bunch, Hyperspectral visualization of mass spectrometry imaging data, Anal. Chem., 85 (2013) 1415-1423.

[25] E.A. Jones, A. van Remoortere, R.J.M. van Zeijl, P.C.W. Hogendoorn, J.V.M.G. Bovée, A.M. Deelder, L.A. McDonnell, Multiple statistical analysis techniques corroborate intratumor heterogeneity in imaging mass spectrometry datasets of myxofibrosarcoma, PLoS One, 6 (2011) e24913-24927.

[26] T. Schramm, A. Hester, I. Klinkert, J.-P. Both, R.M.A. Heeren, A. Brunelle, O. Laprévote, N. Desbenoit, M.-F. Robbe, M. Stoeckli, B. Spengler, A. Römpp, imzML — A common data format for the flexible exchange and processing of mass spectrometry imaging data, J. Proteomics, 75 (2012) 5106-5110.

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

[28] M.-F. Robbe, J.-P. Both, B. Prideaux, I. Klinkert, V. Picaud, T. Schramm, A. Hester, V. Guevara, M. Stoeckli, A. Roempp, R.M.A. Heeren, B. Spengler, O. Gala, S.

Haan, Software tools of the Computis European project to process mass spectrometry images, Eur. J. Mass Spectrom., 20 (2014) 351-360.

[29] O. Jardin-Mathé, D. Bonnel, J. Franck, M. Wisztorski, E. Macagno, I. Fournier, M. Salzet, MITICS (MALDI Imaging Team Imaging Computing System): a new open source mass spectrometry imaging software, J. Proteomics, 71 (2008) 332-345.

[30] P. Ràfols, D. Vilalta, J. Brezmes, N. Cañellas, E. del Castillo, O. Yanes, N. Ramírez, X. Correig, Signal preprocessing, multivariate analysis and software tools for MA(LDI)-TOF mass spectrometry imaging for biological applications, Mass Spectrom. Rev., 9999 (2016) 1-26.

[31] K.D. Bemis, A. Harry, L.S. Eberlin, C. Ferreira, S.M. van de Ven, P. Mallick, M. Stolowitz, O. Vitek, Cardinal: an R package for statistical analysis of mass spectrometry-based imaging experiments, Bioinformatics, 31 (2015) 2418-2420.

[32] R.M. Parry, A.S. Galhena, C.M. Gamage, R.V. Bennett, M.D. Wang, F.M. Fernández, omniSpect: an open MATLAB-based tool for visualization and analysis of matrix-assisted laser desorption/ionization and desorption electrospray ionization mass spectrometry images, J. Am. Soc. Mass Spectrom., 24 (2013) 646-649.

[33] M.T. Bokhart, M. Nazari, K.P. Garrard, D.C. Muddiman, MSiReader v1.0: Evolving Open-Source Mass Spectrometry Imaging Software for Targeted and Untargeted Analyses, J. Am. Soc. Mass Spectrom., (2017) 1-9. DOI: 10.1007/s13361-017-1809-6.

[34] A.M. Race, B. Josephine, D. Alex, P.A. D., S.R. T., S.I. B., SpectralAnalysis: Software for the Masses, Anal. Chem., 88 (2016) 9451-9458.

[35] T. Li, J. He, X. Mao, Y. Bi, Z. Luo, C. Guo, F. Tang, X. Xu, X. Wang, M. Wang, J. Chen, Z. Abliz, In situ biomarker discovery and label-free molecular histopathological diagnosis of lung cancer by ambient mass spectrometry imaging, Sci. Rep., 5 (2015) 14089-14101.

[36] Z. Luo, J. He, Y. Chen, J. He, T. Gong, F. Tang, X. Wang, R. Zhang, L. Huang, L. Zhang, H. Lv, S. Ma, Z. Fu, X. Chen, S. Yu, Z. Abliz, Air flow-assisted ionization imaging mass spectrometry method for easy whole-body molecular imaging under ambient conditions, Anal. Chem., 85 (2013) 2977-2982.

[37] J. He, Z. Luo, L. Huang, J. He, Y. Chen, X. Rong, S. Jia, F. Tang, X. Wang, R. Zhang, J. Zhang, J. Shi, Z. Abliz, Ambient Mass Spectrometry Imaging Metabolomics Method Provides Novel Insights into the Action Mechanism of Drug Candidates, Anal. Chem., 87 (2015) 5372-5379.

[38] X. Song, Z. Luo, X. Li, T. Li, Z. Wang, C. Sun, L. Huang, P. Xie, X. Liu, J. He, Z. Abliz, In Situ Hydrogel Conditioning of Tissue Samples To Enhance the Drug's Sensitivity in Ambient Mass Spectrometry Imaging, Anal. Chem., 89 (2017) 6318-6323.

[39] J. He, F. Tang, Z. Luo, Y. Chen, J. Xu, R. Zhang, X. Wang, Z. Abliz, Air flow assisted ionization for remote sampling of ambient mass spectrometry and its application, Rapid Commun. Mass Spectrom., 25 (2011) 843-850.

[40] M. Turewicz, E. Deutsch, Spectra, chromatograms, Metadata: mzML-the standard data format for mass spectrometer output, Methods Mol. Biol., 696 (2011) 179-203.

518 [41] P.G.A. Pedrioli, J.K. Eng, R. Hubley, M. Vogelzang, E.W. Deutsch, B. Raught,
519 B. Pratt, E. Nilsson, R.H. Angeletti, R. Apweiler, K. Cheung, C.E. Costello, H.
520 Hermjakob, S. Huang, R.K. Julian Jr, E. Kapp, M.E. McComb, S.G. Oliver, G.
521 Omenn, N.W. Paton, R. Simpson, R. Smith, C.F. Taylor, W. Zhu, R. Aebersold, A
522 common open representation of mass spectrometry data and its application to
523 proteomics research, Nat. Biotechnol., 22 (2004) 1459-1466.
524 [42] P. Källback, M. Shariatgorji, A. Nilsson, P.E. Andrén, Novel mass spectrometry
525 imaging software assisting labeled normalization and quantitation of drugs and
526 neuropeptides directly in tissue sections, J. Proteomics, 75 (2012) 4941-4951.
527 [43] J.L. Norris, R.M. Caprioli, Analysis of tissue specimens by matrix-assisted laser
528 desorption/ionization imaging mass spectrometry in biological and clinical research,
529 Chem. Rev., 113 (2013) 2309-2342.
530 [44] P.S. Gromski, H. Muhamadali, D.I. Ellis, Y. Xu, E. Correa, M.L. Turner, R.
531 Goodacre, A tutorial review: Metabolomics and partial least squares-discriminant
532 analysis – a marriage of convenience or a shotgun wedding, Anal. Chim. Acta, 879
533 (2015) 10-23.
534 [45] A. Palmer, P. Phapale, I. Chernyavsky, R. Lavigne, D. Fay, A. Tarasov, V.
535 Kovalev, J. Fuchser, S. Nikolenko, C. Pineau, M. Becker, T. Alexandrov,
536 FDR-controlled metabolite annotation for high-resolution imaging mass spectrometry,
537 Nat Meth, 14 (2017) 57-60.
538
539

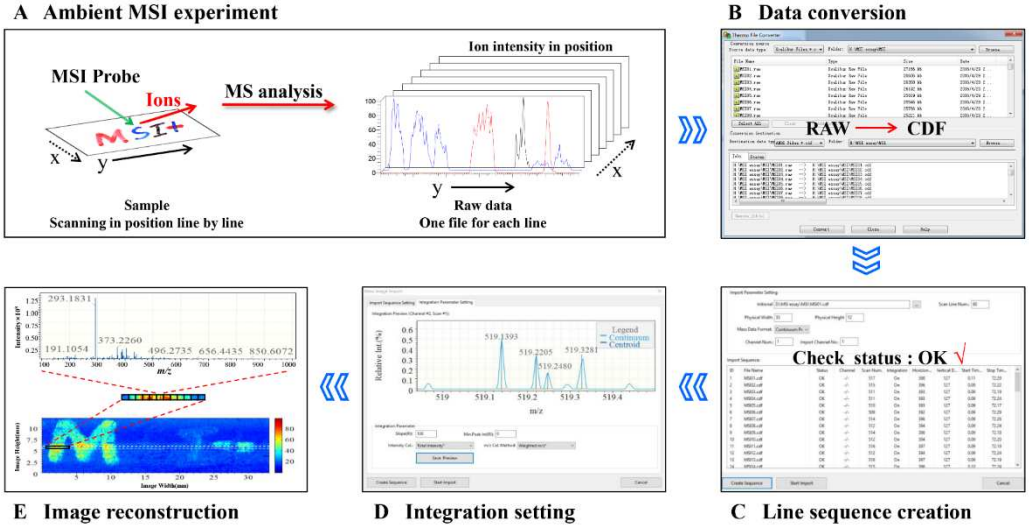


Figure 1. The basic work flow of creating a new solution after a MSI experiment for one case.

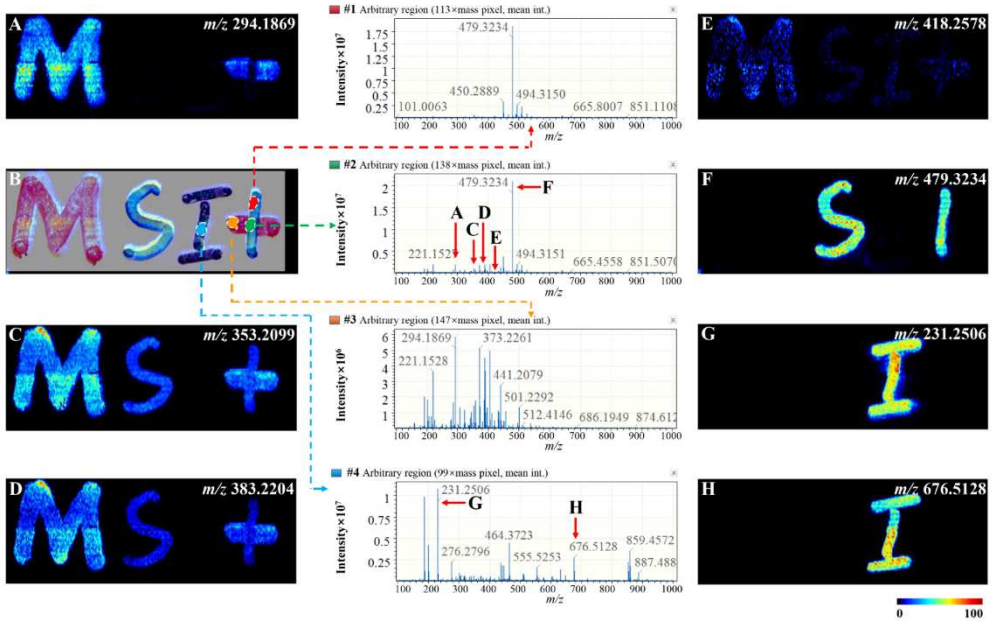


Figure 2. The interactive process between ion images and mass spectra. (B) Optical image and ion image overlay; (A, C-H) Corresponding ion maps of different ions in mass spectrum #2 and #4.

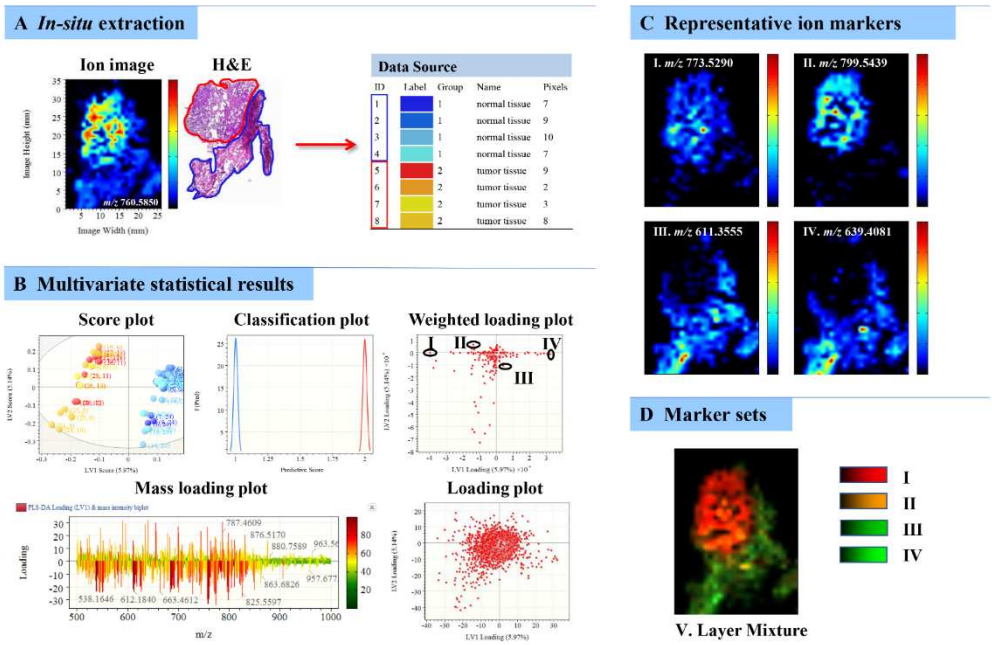


Figure 3. Multivariate statistical analysis for biomarker discovery of lung cancer pathological section in MassImager.

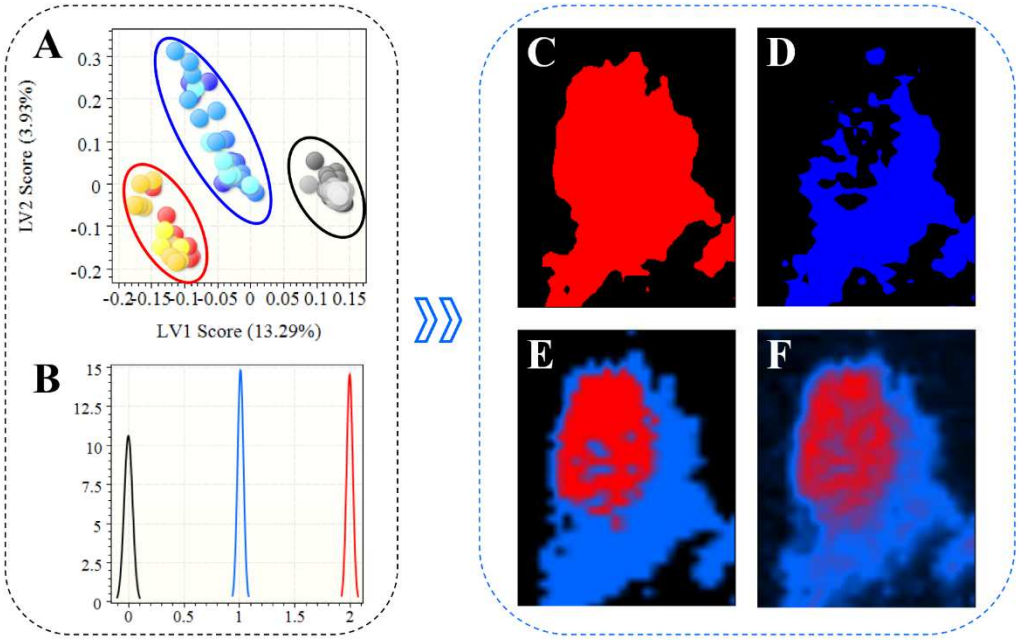
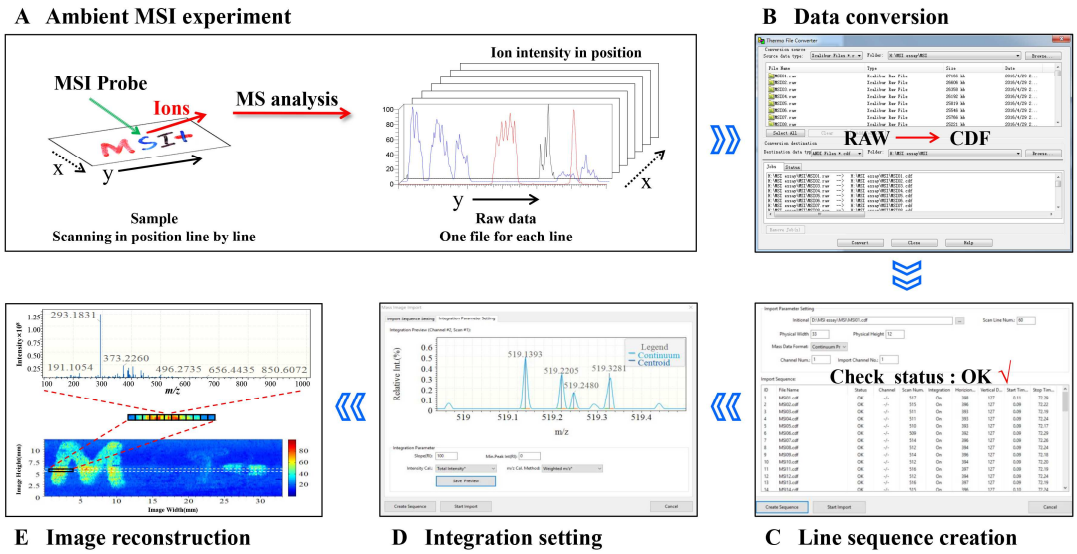
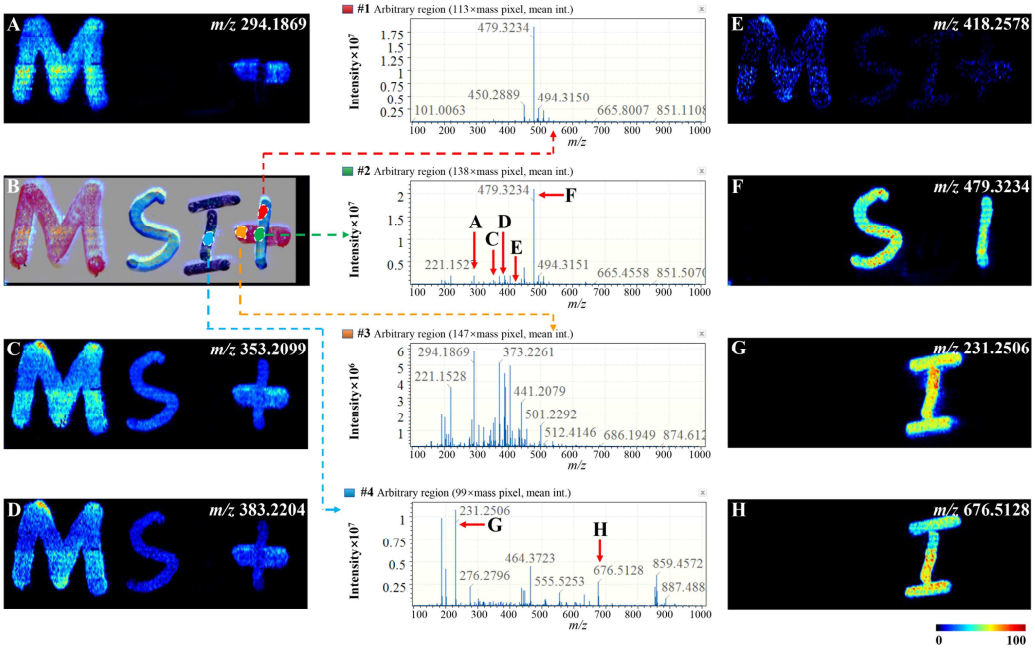
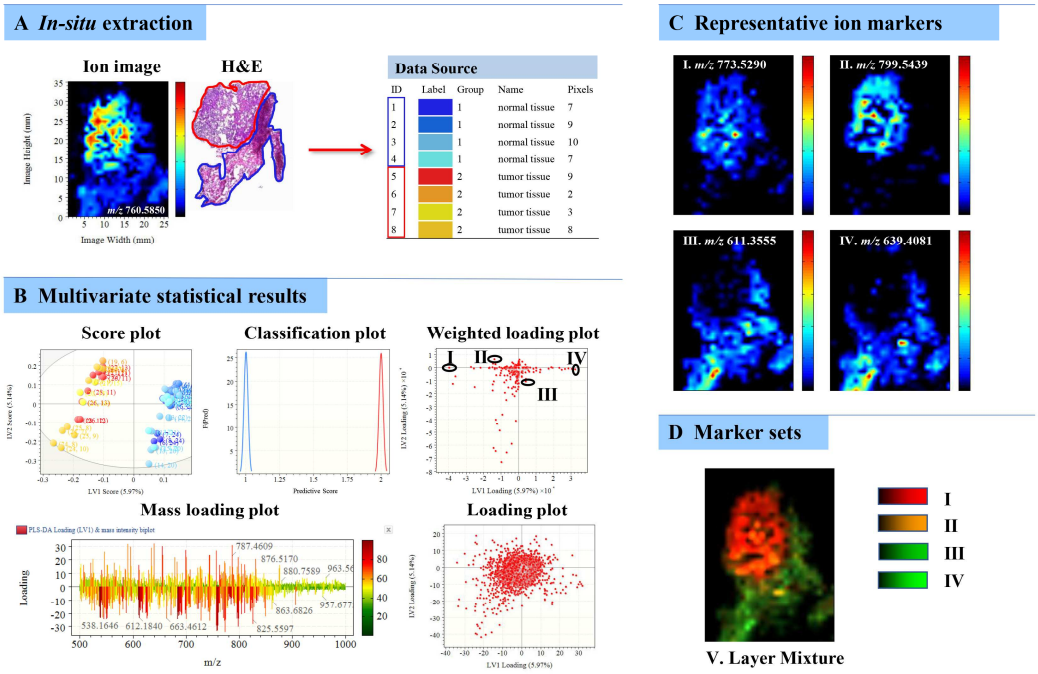
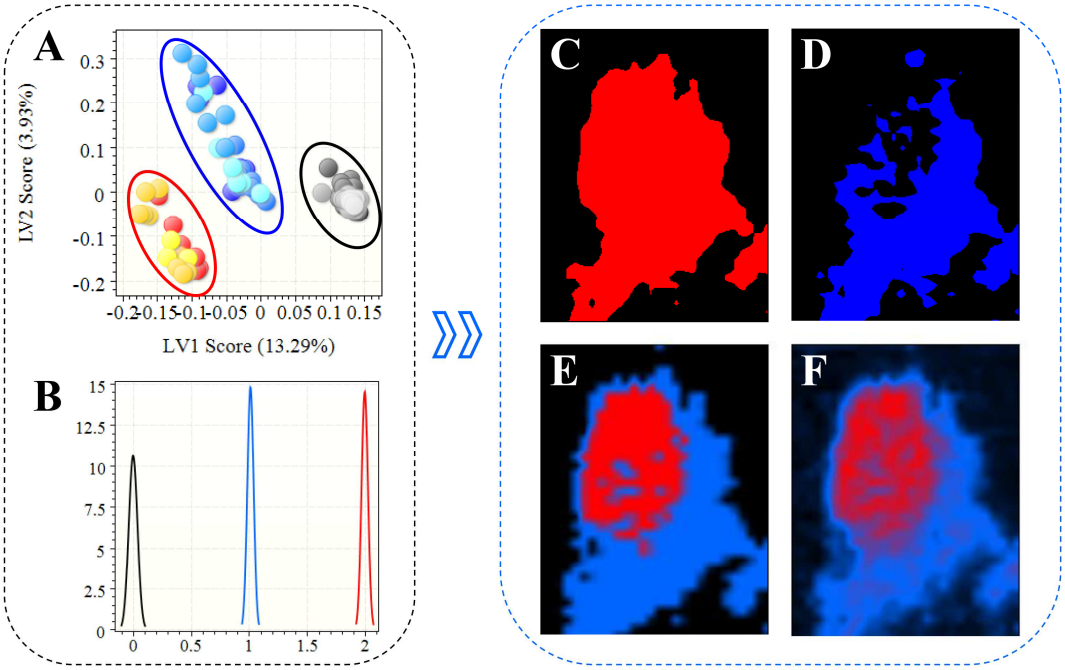


Figure 4. Automatic image recognition of lung cancer sample '983322T' based on PLS-DA model. (A) Score plot; (B) Classification plot; (C) Model loading #1; (D) Model loading #2; (E) Class prediction value (with binarization); (F) Class prediction value (without binarization).









HIGHLIGHTS

- A powerful and easy-to-use software to discover the underlying biological information from complicated and huge mass spectrometry imaging (MSI) dataset.
- High-performance MSI data analysis including quick ion image reconstruction, multi-mode visualization, region-of-interest analysis, multivariate statistical analysis and pattern recognition.
- Quick screening for direct region-specific biomarkers.
- Artificial intelligent pathological diagnosis of bio-tissue based on multivariate statistical analysis at the *in-situ* metabolomics level.