Analytica Chimica Acta 1325 (2024) 343124

Contents lists available at ScienceDirect

Analytica Chimica Acta

journal homepage: www.elsevier.com/locate/aca

MassLite: An integrated python platform for single cell mass spectrometry metabolomics data pretreatment with graphical user interface and advanced peak alignment method

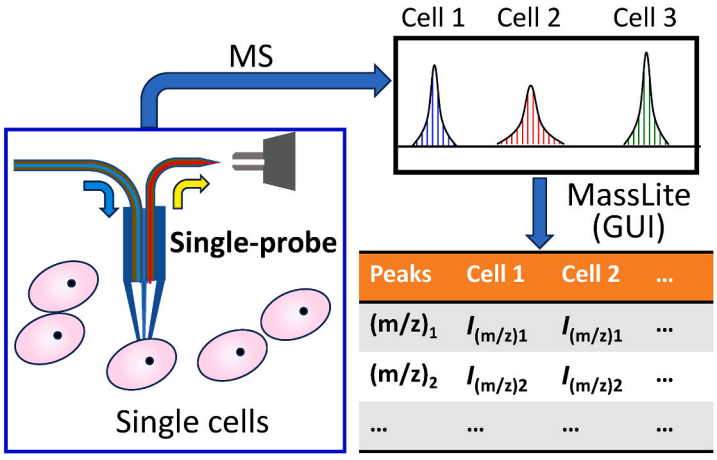
Zhu Zou, Zongkai Peng, Deepti Bhusal, Shakya Wije Munige, Zhibo Yang *

Department of Chemistry and Biochemistry, University of Oklahoma, Norman, OK, 73019, USA

HIGHLIGHTS

- First data pretreatment platform specifically compatible with improvised SCMS data acquisition.
- New algorithm for alignment process with more accurate *m/z* results.
- Graphical user interface for easy use.

GRAPHICAL ABSTRACT



ARTICLE INFO

Handling Editor: L. Liang

ABSTRACT

Mass spectrometry (MS) has been one of the most widely used tools for bioanalytical analysis due to its high sensitivity, capability of quantitative analysis, and compatibility with biomolecules. Among various MS techniques, single cell mass spectrometry (SCMS) is an advanced approach to molecular analysis of cellular contents in individual cells. In tandem with the creation of novel experimental techniques, the development of new SCMS data analysis tools is equally important. As most published software packages are not specifically designed for pretreatment of SCMS data, including peak alignment and background removal, their applicability on processing SCMS data is generally limited. Hereby we introduce a Python platform, MassLite, specifically designed for rapid SCMS metabolomics data pretreatment. This platform is made user-friendly with graphical user interface (GUI) and exports data in the forms of each individual cell for further analysis. A core function of this tool is to use a novel peak alignment method that avoids the intrinsic drawbacks of traditional binning method, allowing for more effective handling of MS data obtained from high resolution mass spectrometers. Other functions, such as

* Corresponding author.
E-mail address: Zhibo.Yang@ou.edu (Z. Yang).

https://doi.org/10.1016/j.aca.2024.343124
Received 14 April 2024; Received in revised form 7 August 2024; Accepted 18 August 2024
Available online 20 August 2024

void scan filtering, dynamic grouping, and advanced background removal, are also implemented in this tool to improve pretreatment efficiency.

1. Introduction

Mass spectrometry (MS) has been playing an increasingly important role in the field of chemistry and bioanalysis since the invention of electrospray ionization (ESI) [1] and matrix-assisted laser desorption/ionization (MALDI) [2]. Assisted with improved sensitivity, resolution, and throughout of mass spectrometers [3] as well as advancement of computing power from hardware and software algorithm, MS has been broadly adopted in applications such as proteomics [4], metabolomics [5], biomarker discovery [6], and drug discovery [7].

Among various experimental methods, liquid chromatography-MS (LC-MS) has especially been widely-applied, excelling in the separation and quantification of complex mixtures and biological samples [8]. However, due to the obligatory sample preparation, some critical information, such as spatial distribution of molecular species in tissues and cell heterogeneity, is inevitably lost from LC-MS measurements of bulk samples. To overcome disadvantages of traditional LC-MS techniques, novel MS methods have been developed. Among them, MS imaging (MSI) is capable of offering insights into the spatial distribution of compounds in tissues, providing knowledge in histopathology and drug distribution [9–11]. These imaging techniques utilize vacuum-based (e.g., matrix-assisted laser desorption/ionization (MALDI) [12], secondary ion mass spectrometry (SIMS) [13], and matrix-free laser desorption/ionization (LDI) [14]) and ambient-based (e.g., desorption electrospray ionization (DESI) [15], nanospray desorption electrospray ionization (nano-DESI) [16], and Single-probe [17]) methods for sampling and ionization. In addition, single cell MS (SCMS) has recently gained increasing popularity due to its capability of reaching cellular and subcellular resolution and performing cell heterogeneity analysis [18–22]. Compared with traditional bulk analysis, SCMS reveals the chemical profiles of individual cells, providing unique understanding of complicated cell activities controlled by numerous intracellular and extracellular factors. The existing SCMS methods use varied techniques for sampling and ionization, with a number of studies using methods similar to MSI with vacuum-based (e.g., MALDI [23], SIMS [24]) and ambient-based (e.g., DESI [25], laser ablation electrospray ionization (LAESI) [26], nano-DESI [27], Single-probe [28]), and other methods using other special sampling probes [29,30] or fluidic-based devices [31] for single-cell isolation.

With the application of newly developed mass spectrometers possessing higher mass resolution, faster scan rate, and better sensitivity, the size of MS data has significantly increased, making development of effective data pretreatment algorithm increasingly important in modern MS bioanalysis. Numerous software packages have been designed to process the experiment data (e.g., peak picking, peak alignment, and intensity normalization) and to extract essential information from data acquired from traditional LC-MS (e.g., *MZmine* [32–34] and *XCMS* [35,36]) and novel MSI (e.g., *Cardinal* [37] and *Metaspace* [38]) experiments. In fact, some of these software tools have been utilized to analyze certain types of SCMS data. For example, single cell proteomics experiments coupled with LC separation prior to MS analysis can be analyzed using conventional LC/MS proteomics data analysis method without major changes [39–41]. Similarly, MALDI-based single cell metabolomics can be acquired using strategies similar to those in high-spatial resolution MALDI techniques [42–44]. However, very few attempts have been made to handle data acquired from ambient SCMS metabolomics [45–48]. In our previous studies, a generalized data analysis workflow was introduced for SCMS metabolomic data analysis. This workflow, which consists of data pretreatment, multivariate analysis, and univariate analysis, was adopted from traditional methods in LC-MS data processing [48]. However, unlike most well-established LC-MS and MSI experiments, which are generally conducted using programmed, pre-loaded sampling and data acquisition process, most ambient SCMS metabolomics studies of single cells are commonly associated with improvised single cell sampling and segmented signal due to experimental conditions and operations, causing incompatibility with existing data processing tools that were designed for LC-MS and MSI [49,50]. Therefore, the previously published workflow with traditional LC-MS data pretreatment approaches still have drawbacks such as lack of operational convenience and accuracy of m/z value determination. As any separation can potentially induce sample loss and dilution, in the SCMS studies of small molecules (e.g., metabolites), analyte separation is generally not performed prior to MS analysis, resulting in mass spectra with large numbers of peaks. Analysis of dense peaks heavily relies on the comparison between measured accurate mass from the spectra with calculated exact mass from known compound structures in the database [51]. Particularly, accurate mass measurements provide crucial information for molecular identification in untargeted metabolomics studies. Therefore, retaining all the valuable information obtained from high resolution mass spectra is a crucial need for SCMS metabolomics data pre-processing.

In actual MS experiments, the accurate m/z (mass-to-charge ratio) value of an ion is determined by the mean value of individual measurements containing the corresponding peaks. For MS analysis of unknown substances, such as in untargeted SCMS metabolomics studies, accuracy of m/z measurement cannot be defined without knowing the "correct" reference, making precision a more important factor in the process of data analysis. Among all data-processing steps (e.g., peak picking, peak alignment, intensity normalization, and mass correction), peak alignment is the step designed to correct the random variation in the measurement of the same peak, rendering the measured m/z values of the peaks for further advanced analysis (e.g., multivariate analysis, data visualization, and structure identification). Without peak alignment, ion signals from the same substance can be mistakenly split into multiple peaks, while a poor alignment might merge ions from different substance into the same peak; both types of outcomes lead to a misinterpretation of MS data. In addition, complexity of datasets induced by improper peak alignment not only alters the output of advanced data analysis methods, but also significantly increases the cost for further processing.

Among all developed peak alignment algorithms, binning is a method commonly used in numerous studies due to its simplicity. Briefly, for the convenience of data analysis, the entire range of the m/z values of a mass spectrum is divided into a large number of equidistant small chunks (i.e., bins) through a histogram-based method [52,53]. Although binning can significantly reduce the computational cost, this method possesses multiple intrinsic drawbacks [53–55]. First, the outcome of data processing is influenced by the parameters of bins, including bin width and bin position. Peaks could be artificially merged, split, or shifted due to unideal bin parameters, resulting in a loss of information. Second, using linear equidistant bins can lead to unequal mass error (i.e., ppm) of MS measurement. For example, as a commonly used bin width, 0.01 Da mass difference corresponds to 100 ppm at 100 Da, but 5 ppm at 2000 Da. Thus, binning cannot take full advantage of the capabilities of the high resolutions of mass spectrometers, which provide advantages of accurate measurements (i.e., m/z values) of numerous species in complex samples. Due to its intrinsic drawbacks, binning method cannot effectively extract molecular information from complex SCMS metabolomics data, which heavily rely on accurate mass measurement.

In this study, we introduce MassLite, a user-friendly, Python-based platform with graphical user interface (GUI) specifically designed for

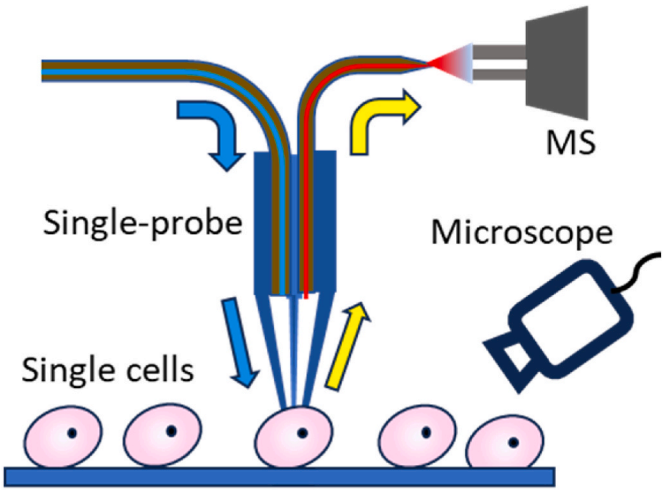


Fig. 1. Setup of the Single-probe SCMS experiment.

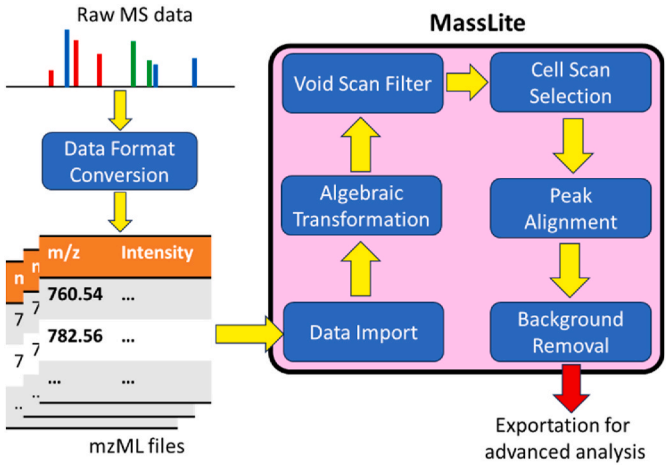


Fig. 2. Schematic data processing workflow of MassLite.

SCMS metabolomics data pretreatment. Compared with the existing SCMS metabolomics analysis tools, this new software package possesses multiple advantages. First, this platform is robust to handle SCMS data acquired from intermittent acquisition processes, in which ion signals from individual cells are sequentially segmented. Second, MassLite can take full advantage of high-resolution mass spectrometers by maintaining high mass-resolution of peaks during peak alignment process. Third, automatic cell region selection is used to replace the existing labor-intensive, manual process to increase processing throughput. Fourth, the algorithms of peak alignment and background removal have been improved to be specifically compatible with SCMS metabolomics data. Last, the computational cost was significantly reduced with our purposed dynamic grouping method. Although the capability of this tool was demonstrated using the data generated from the Single-probe SCMS method performed using a Thermo Orbitrap XL mass spectrometer, data produced from other SCMS techniques and platforms can be converted to standard.mzML format and then processed by MassLite.

2. Method

2.1. Cell culture

In this study, a human colorectal carcinoma cell line (HCT-116) (ATCC, Rockville, MD, USA) was used as a model. Cells were cultured in McCoy's 5A Medium (Fisher Scientific Company LLC, IL, USA). The medium was supplemented with 10 % fetal bovine serum (GE Healthcare Bio-science Corp, Marlborough, MA, USA) and 1 % penicillin-streptomycin (Life Technologies Corporation, Grand Island, NY, USA). The cells were incubated at 37 °C in the presence of 5 % CO_2. Once the confluence of the HCT-116 cells reached 80 %, the cells were passaged. Cells were transferred into 12-well plates, and a glass coverslip (18 mm in diameter) was placed in each well. 2 mL of diluted cell suspension ($1x10^5$ cells/well) were added to each well in the 12-wells plate, followed by overnight incubation to enable cell attachment.

2.2. SCMS experiment

SCMS experiments were performed using the established Single-probe SCMS technique as reported in our previously studies [28,46,56,57]. A Single-probe was fabricated by embedding a solvent-providing fused silica capillary (O.D. 105 μm; I.D. 40 μm, Polymicro Technologies, Phoenix, AZ), a nano-ESI emitter (pulled from the same fused silica capillary using a butane micro torch) into a dual-bore quartz needle (produced from dual-bore quartz tubing (O.D. 500 μm; I.D. 127 μm, Friedrich & Dimmock, Millville, NJ) using a laser micropipette puller (Sutter P-2000, Sutter Instrument, Novato, CA)). The Single-probe device was coupled to a LTQ Orbitrap XL mass spectrometer (Thermofisher Scientific, San Jose, CA) (Fig. 1). The sampling solvent (acetonitrile with 0.1 % formic acid) was continuously delivered (flowrate 200 nL/min) to the solvent-providing capillary. A glass coverslip containing cells was rinsed by fresh 0.9% (w/w) ammonium formate, and then placed onto the XYZ-translational stage (step size = 0.1 μm). Guided by a digital microscope (Shenzhen D&F Co., China), a target cell was selected and sampled by gradually moving the stage. Cellular metabolites were extracted by the liquid junction formed at the tip of the Single-probe, and immediately ionized and analyzed. MS analysis parameters are listed as follows: ionization voltage +4.5 kV, mass range 200–1500, mass resolution 60,000 at *m/z* 400, 1 microscan, 500 ms max injection time, and automatic gain control (AGC, 1E6) on.

2.3. Data pretreatment

MS data obtained from the experiment must undergo pretreatment prior to further advanced analysis. For SCMS metabolomics data handled by our platform in this study, the data pretreatment includes format conversion, algebraic transformation, void scan selection, cell scan selection, peak alignment, background peak removal, and data exportation (Fig. 2). The entire data processing was performed using MassLite, except that data format conversion (i.e., from *.raw* to *.mzML*) was conducted through other existing tools. The converted data was imported into our platform, and algebraic transformation, which allowed us to use relative mass difference to describe the original *m/z* difference among peaks, was performed. Next, a void scan filter was applied to distinguish intermittent scans during the data acquisition process. Then, the filtered scans of each cell were grouped based upon the extracted ion chromatogram of selected cell markers. Afterwards, peak alignment was performed with dynamic grouping to correct the mass shift of peaks. Last, background peaks can be selected and removed prior to data exportation.

2.3.1. Data import

The SCMS raw data is complex as it contains a variety of different ion signals of cellular analytes and non-analytes (Fig. S1). To make MassLite compatible with SCMS metabolomics data acquired from all types of mass spectrometers, the algorithm in our platform was designed on basis of a universal MS data format, *mzML* [58]. For the Single-probe SCMS data tested here, the original file generated using a Thermo Orbitrap LTQ XL mass spectrometer was in *.raw* format, which was converted into the widely-used *.mzML* format using *MSConvert* incorporated in *ProteoWizard* [58–60]. The converted data was read using *pymzML* package in our Python platform to extract the *m/z* values and intensities of the

peaks. Peak picking was then conducted to obtain centroid peaks for each MS scan prior to further processing.

2.3.2. Algebraic transformation

As most MS studies utilize mass accuracy or mass measurement error (i.e., the difference between an individual measurement and the true value) in the unit of ppm, relative mass difference is likely a more straightforward way to describe the difference between two *m/z* values. In order to perform simpler peak comparison during the pretreatment process, we performed a scaled, dynamic logarithmic transformation to intuitively describe the relative mass difference in the unit of ppm. In addition, this algorithm reflects mass accuracy with respect to *m/z* values, minimizing the influence of mass range on peak comparison.

For two peaks at $\left(\frac{m}{z}\right)_1$ and $\left(\frac{m}{z}\right)_2$ (assuming $\left(\frac{m}{z}\right)_1 > \left(\frac{m}{z}\right)_2$), their relative mass difference can be described as:

$$\frac{\Delta ppm}{10^6} = \left[\left(\frac{m}{z}\right)_1 - \left(\frac{m}{z}\right)_2\right] \Big/ \left\{\frac{1}{2} \times \left[\left(\frac{m}{z}\right)_1 + \left(\frac{m}{z}\right)_2\right]\right\} \quad \text{(Eq. 1)}$$

When these two peaks are very close to each other, as generally observed in MS analysis with slight mass shifts from scan to scan, the absolute difference is significantly smaller than their *m/z* values. $\left(\frac{m}{z}\right)_1 + \left(\frac{m}{z}\right)_2$ can be redeemed as $2 \times \left(\frac{m}{z}\right)_2$, and we have

$$\frac{\Delta ppm}{10^6} = \left[\left(\frac{m}{z}\right)_1 - \left(\frac{m}{z}\right)_2\right] \Big/ \left(\frac{m}{z}\right)_2 \quad \text{(Eq. 2)}$$

By reorganizing the formula, we have

$$1 + \frac{\Delta ppm}{10^6} = \left(\frac{m}{z}\right)_1 \Big/ \left(\frac{m}{z}\right)_2 \quad \text{(Eq. 3)}$$

Taking logarithmic transformation on both sides, we have

$$\ln\left(1 + \frac{\Delta ppm}{10^6}\right) = ln\left(\frac{m}{z}\right)_1 - ln\left(\frac{m}{z}\right)_2 \quad \text{(Eq. 4)}$$

When these two peaks are close enough to each other, $\Delta ppm \to 0$. Given that $\lim_{x \to 0} \ln(1+x) = x$ according to Taylor expansion, we have the following representation:

$$\frac{\Delta ppm}{10^6} = ln\left(\frac{m}{z}\right)_1 - ln\left(\frac{m}{z}\right)_2 \quad \text{(Eq. 5)}$$

Thus, when transformation $f\left(\frac{m}{z}\right) = \ln\left(\frac{m}{z}\right) \times 10^6$ is applied on two close neighboring peaks, we have

$$f\left(\left(\frac{m}{z}\right)_1\right) - f\left(\left(\frac{m}{z}\right)_2\right) = \left[ln\left(\frac{m}{z}\right)_1 - ln\left(\frac{m}{z}\right)_2\right] \times 10^6 = \Delta ppm \quad \text{(Eq. 6)}$$

Pairwise Euclidean distance between transformed *m/z* values, i.e. $f\left(\left(\frac{m}{z}\right)_1\right) - f\left(\left(\frac{m}{z}\right)_2\right)$, can reflect the relative mass difference of the original *m/z* values in the unit of ppm, enabling fast processing and peak matching in the subsequent steps. In practical applications, a linear shift was included according to the lower limit of the mass range being detected.

2.3.3. Void scan removal

A typical ambient SCMS metabolomics dataset consists of informative scans (i.e., signals of cellular analytes along with coexisting solvent background and culture media) and void scans (i.e., scans containing only instrument noise without identifiable species from cell analyte, solvent background, or culture media) (Fig. S1). The void scans are commonly included in data acquisition processes, mostly due to certain operations during experiments (e.g., cell sampling is paused or interrupted while data acquisition is continuously running). To automatically identify the void scans within the file, K-means, an unsupervised clustering method, was used to analyze intensity histogram of MS spectra for each scan. Because the intensity histogram reflects the general profiles of detected ions, significant changes of global pattern are expected between informative and void scan signals. For the actual K-means input, options of Uniform Manifold Approximation and Projection (UMAP) and logarithmic scaling are provided for transformation of the intensity histogram to enhance the discrimination between void scans and other scans. TIC (total ion current) of the clusters generated by the unsupervised K-means method can be visualized in the GUI for inspection, and clusters matching the definition of void scans can be dropped to reduce workload for the subsequent processes.

2.3.4. Selection of MS scans of single cells

To further increase the throughput of SCMS data processing, we developed an algorithm to automatically differentiate scans representing single cells from those from background such as solvent or cell culture media. First, a chromatogram was generated based upon the intensity of cell markers selected by users. For example, *m/z* 782.58 and 760.56 are commonly detected ions in cells, and they were selected as default indicators of single cell detection (i.e., marker signals). Second, an initial Gaussian smoothing was performed for extracted ion chromatogram (EIC) to avoid unideal splits of signals from each single cell due to signal fluctuation during the data acquisition process. Third, MS scans of cells and background were defined. After the maxima and minima of ion intensities of the selected markers in the MS scans were primarily found, a finer global search across the whole chromatogram was conducted, minimizing the generation of artificial peaks due to ion intensity fluctuation. A stricter intensity requirement for peak search within maxima found in the previous search was applied to account for possible peak splitting issue due to signal fluctuation in the EIC. In the current study, the regions containing marker signals $\geq$20 % (default value) of the local maxima were defined as cell regions, whereas region containing marker signals <5 % (default value) of the local maxima were regarded as background regions unless otherwise defined.

2.3.5. Peak alignment with dynamic grouping

Due to multiple factors (e.g., the intrinsic performance of instrument and fluctuation of ion signals and instrument conditions), mass shift generally occurs during MS analysis [61–63]. Because accurate *m/z* values provide important information for molecular identification, mass shift correction is critical in high resolution MS studies, in which multiple ions with similar *m/z* values can be simultaneously detected. Inappropriate handling of mass shift may result in artifacts such as peak splitting, loss of peaks, or inaccurate *m/z* assignment. To compensate for the mass shift across different scans, peak alignment must be performed. For the ease of processing and precise *m/z* value description, centroiding on all peaks was performed, keeping only one *m/z* value of peak center and one intensity value for each peak. All centroid peaks along with their transformed *m/z* values from all imported scans were included for peak alignment. Hierarchical clustering was performed for observed peaks to find internal matching among themselves. The cluster size was set as double of the desired mass shift tolerance for hierarchical clustering, ensuring the coverage for each aligned peak is within the threshold. For example, if the maximum mass shift tolerance is less than 5 ppm, a 10-ppm cluster size is adopted to guarantee that only peaks within ±5 ppm shift from the center will be included. Thus, the mass accuracy of peak alignment was guaranteed (e.g., within 5 ppm in the above example). With the algebraic transformation performed in earlier steps, simple one-dimensional Euclidean distance can be redeemed as the relative mass difference between the *m/z* values of different peaks. To reduce the cost of pairwise distance calculations in the hierarchical clustering process, we utilized "divide and conquer" strategy. In general, this strategy decomposes a given problem into multiple smaller

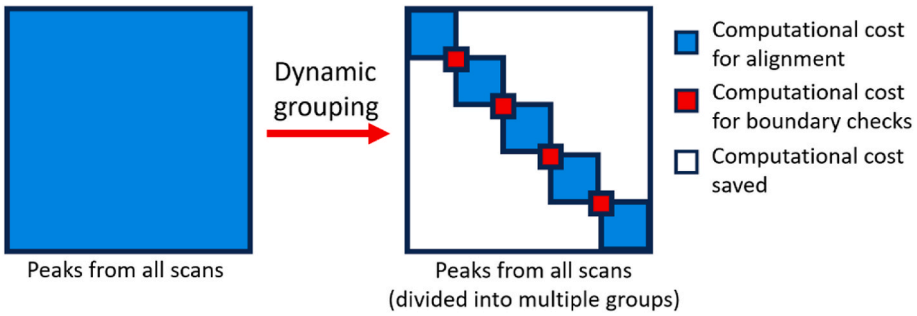


Fig. 3. Computational cost reduction by dynamic grouping. Cost of hieriarchical clustering is $O(n^2)$ due to pairwise comparison and distance matrix update. With dynamic grouping strategy, computational cost is largely saved by eliminating uncessary comparison between peaks from different chunks (white space off the diagonal line).

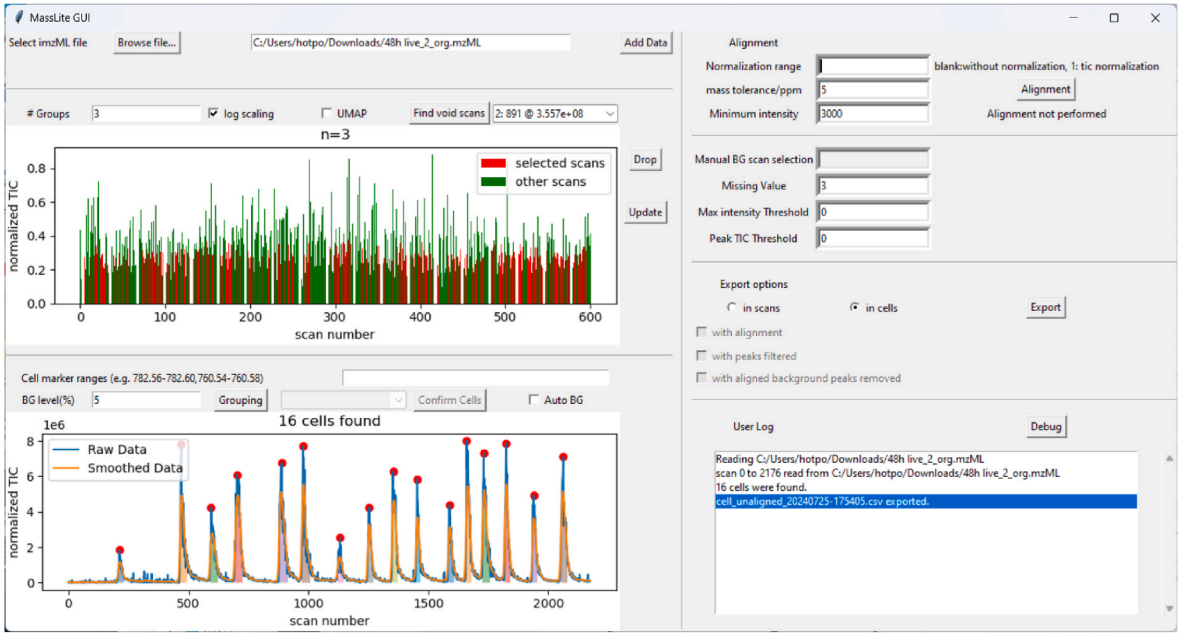


Fig. 4. MassLite graphical user interface. Six main modules are included.

subproblems, and solutions to subproblems are then combined to solve the given problem. We developed a so-called "dynamic grouping" method to split the data in chunks, eliminating unnecessary comparison of peaks from different data chunks which accounted for most of the cost from direct comparison (Fig. 3). The ranked peaks are divided into multiple different groups and processed individually, reducing the cost as a function of the total number of groups (see "Cost reduction of dynamic grouping" in the SI). To address potential peak splitting issues due to this dividing strategy, boundary checks, which compared data at the boundary between two adjacent chunks, were added to merge split peaks due to chunk division. This binning-free method can maintain higher mass resolution from the original data.

2.3.6. Background removal

During ambient SCMS measurement, particularly for live cell analysis, interfering ions generated from impurities in solvent or species in cell culture media are generally detected along with cellular contents. To eliminate these artifacts in analysis, interfering ions should be treated as background and excluded. Thus, aligned peaks with their highest intensities in one of the background scan regions, which could be automatically determined in the cell scan selection step, were regarded as the background substance and subsequentially filtered from the data. Compared with the traditional binning method for background removal, our algorithm is capable of distinguishing peaks from background substances and cell analytes, which possess similar *m/z* values, without prior knowledge of the cell systems.

3. Result and discussion

3.1. Graphical user interface

The graphical user interface (GUI) (Fig. 4) was built using *tkinter* package in Python. In our current design, the GUI has six major parts: data read-in, void scan filter, cell sorting, peak alignment, exportation filter, and debugging modules. Detailed description of each part is provided in the following context, and the user manual is included in the Supplementary Material.

3.2. Parameter optimization for void scan filter

Because the global spectral features of the void scans are significantly different from those of informative scans, which contain signal from cellular analytes and solvent background, the clustering was based on the intensity histogram, which can describe the overall feature of the entire MS spectrum. A series of different parameters needed for the generation of the intensity histogram were tested, along with two

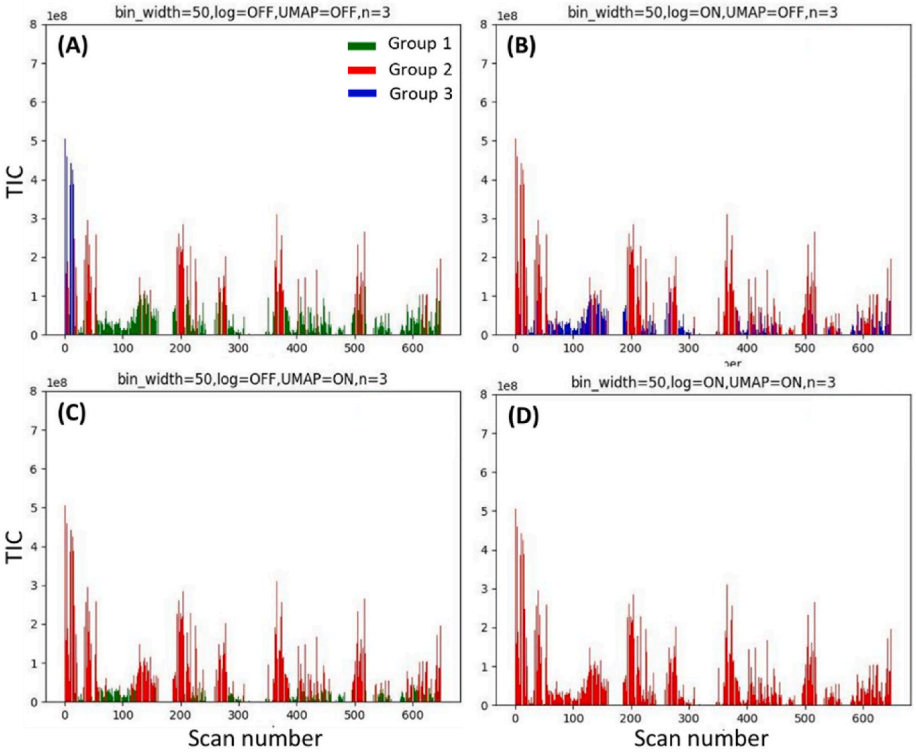


Fig. 5. Test of descrimination enhancement techniques (logarithm scaling and UMAP) after histogram generation. Unsupervised clustering results in each plot are labeled in different colors. The interval is 50 Da and cluster number in K-means is n = 3 (i.e., Groups 1, 2, and 3). Results were obtained using (A) no scaling, (B) only logarithm scaling, (C) only UMAP, or (D) both logarithm scaling and UMAP. The clusters with the lowest ion intensities (Group 1 in B and Group 3 in C) are labeled as void scans. (For interpretation of the references to color in this figure legend, the reader is referred to the Web version of this article.)

different techniques (i.e., logarithmic scaling and UMAP dimensionality reduction) aiming at enhancement of discrimination between void scans and other scans.

In our experience of analyzing SCMS metabolomics data of mammalian cells acquired using the Single-probe SCMS techniques, lipid signals are significantly increased when cellular contents are extracted and detected, especially in the range of *m/z* 700–800 Da [28, 64]. On the other hand, appearance of ions of cell analytes suppresses the base peak intensities in background scan, usually in the range of *m/z* 350–550 Da. This trend is also expected in studies using other SCMS platforms. To ensure important features in the MS spectra, including both cellular analytes and background species, can be captured in the intensity histogram, we tested both 50-Da and 100-Da intervals to generate histograms from data in *m/z* 50–2000 Da. Our results indicate that although histograms with smaller intervals may retain more details of the spectra, the extra amount of information decreased the efficiency for machine learning classification, deviating from our purpose for quick detection. In contrast, larger intervals could possibly fail to capture changes in spectra features if the intensity of ions fluctuates within the same interval.

To perform clustering and acquire efficient identification of void scans, different strategies have been used to enhance discrimination

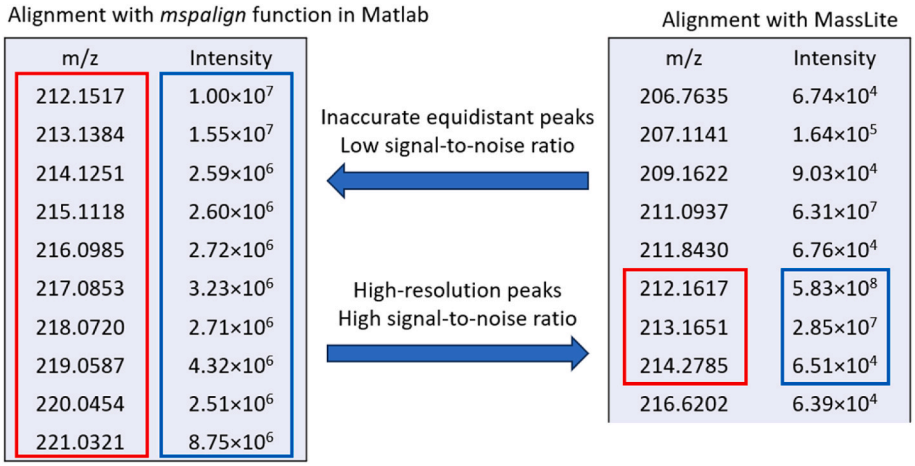


Fig. 6. Comparison between *mspalign* (MATLAB) and MassLite results. Peaks generated from *mspalign* function (by binning) possess inaccurate *m/z* values (due to the artifact of equidistant peaks ($\Delta m/z = 0.9867$)) and low S/N ratios (due to accumulated noise). MassLite provides aligned peaks with higher mass accuracy and higher S/N ratios.

between void scans and other scans. Based upon previous observations in our Single-probe SCMS data, void scans usually contain lower signal intensities compared with other informative scans. Logarithmic scaling is likely a quick, feasible strategy to identify void scans. Although the original ion intensities can better reflect relative ion abundances, logarithmic transformations can reduce signal intensities' differences for ions with significantly different abundances, enhancing the detection of low intensity ions. However, if low-intensity ions are not observed in void scans, dimensionality reduction tools provide alternative options. UMAP, a powerful technique with relatively low computational cost compared with other nonlinear dimensionality reduction methods, has been adopted as an example and tested. In addition, the effect of logarithmic scaling and dimensionality reduction using UMAP were tested both individually and jointly.

Although the SCMS data is not labeled beforehand for unsupervised clustering, certain criteria must be defined to match the goal of quick identification of void scans through clustering. Given that the variance between scans of the same type of signal can hardly be estimated due to the heterogeneity among individual cells, the total number of clusters would be a more practical parameter to guide the process compared with cluster variation. In a typical ambient SCMS measurement of live cells, ion signals are primarily attributed to three types of sources: cellular analytes, solvent background, and cell culture media. With possible subpopulations existing within each type, a total number of cluster $n \geq 3$ would be a reasonable blind guess suitable for different types of SCMS experiments. In the following discussion, the default cluster number ($n = 3$) was used in unsupervised clustering (by K-means), resulting in three groups (Group 1, 2, and 3) of ions (Fig. 5). Although these cluster numbers are not associated with any specific biological features, the cluster with the lowest ion intensities was regarded as the void scans. We used a total of 24 combinations of different approaches (i.e., UMAP, logarithm scaling, cluster number, and bin width) to test the same dataset (Fig. S2), and part of the results are shown in Fig. 6.

When directly using data from the previous step (algebra transformation) as the K-means input, the difference between void scans and low intensity scans was much less significant, leading to insufficient discrimination between void scans and other types of scans in the clustering result (Fig. 5A). To address this issue, logarithmic and UMAP transformation were tested for their capabilities to enhance the separation of scans with low intensities, both individually and synergistically. Logarithmic scaling was adopted because void scans tend to possess considerably lower ion signals compared with informative scans. Alternatively, nonlinear dimensionality reduction can catch the similarities within each group of scans to differentiate void scans from other scans, and therefore UMAP was adopted as an example of nonlinear dimensionality reduction to treat the data. When working individually, either logarithmic or UMAP transformation provided satisfactory clustering output for the purpose of identifying void scans. One of the three clusters matched our definition for void scans (i.e., the one with the lowest ion intensities among all clusters), leaving two clusters representing informative scans. For example, Group 1 is regarded as the void scans when logarithmic scaling is on and UMAP is off (Fig. 5B), whereas Group 3 consists of the void scans when logarithmic scaling is off and UMAP is on (Fig. 5C). However, simultaneously implementing both logarithmic scaling and UMAP tends to lead to undesired results, in which only one group (Group 2) was shown. To effectively sort out void scans, either logarithmic or UMAP transformation is adequate without causing artificial split. In addition, different intervals used for histogram generation seemed to be the least sensitive parameter because either 100 or 50 Da interval in the range of m/z 100–2000 provided enough features for the K-means clustering.

3.3. Alignment result

To evaluate the performance of our peak alignment algorithm, a dataset collected from 16 cells, which consists of 2176 MS scans, was tested using MassLite, MATLAB (i.e., *mspalign* function), *Geena 2* [65], and *MZmine 3* [34]. The existing platforms (e.g., MATLAB, *Geena 2*, and *MZmine 3*) have been widely used for MS data processing. Due to intrinsic drawbacks of binning, equidistant peaks (i.e., $\Delta m/z = 0.9867$ between two neighboring peaks) were produced from *mspalign* in MATLAB, and the aligned results with this artifact cannot accurately represent peak locations in the original MS spectra (Fig. 5). In addition, the binning method in MATLAB resulted in low signal-to-noise (S/N) ratios, primarily due to the accumulated noise in the binning process, as well as increased computational costs. In contrast, MassLite successfully filtered such noise and provided improved S/N ratios of the aligned data with reduced computational resources.

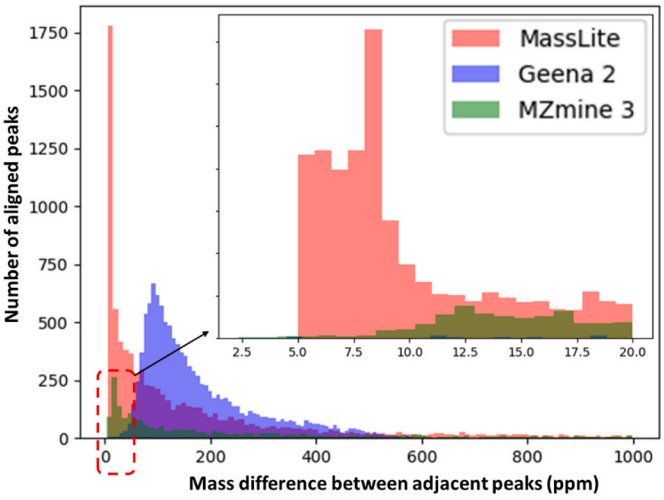


Fig. 7. Histogram showing the relationship between relative mass difference (ppm) of adjacent peaks and the number of aligned peaks acquired from MassLite, *Geena 2*, and *MZmine 3*. The inset illustrates the details in the zoomed-in region.

Geena 2 was also tested in this work. However, this online platform could not handle this entire dataset from 16 cells (with 2176 MS scans) with 5 ppm mass shift tolerance because the data size is over the memory limit of *Geena 2*. Alternatively, a truncated dataset of 4 cells (with 830 MS scans) at default 0.1 Da mass shift tolerance was submitted and processed by *Geena 2*. Both *MZmine 3* and MassLite were able to handle the original entire dataset. To investigate the mass accuracy maintained by each platform, all aligned peaks were re-ordered in ascending order, and the relative mass difference between adjacent peaks were calculated using our algebraic transformation. The relative mass difference between adjacent peaks can reflect the ability of data processing platforms on resolving peaks with similar m/z values. For an intuitive view, a histogram showing the distribution of relative mass differences between adjacent peaks was generated (Fig. 7). As illustrated in the zoomed-in region of the histogram, compared to *Geena 2* and *MZmine 3*, MassLite was able to better differentiate more signals within 5–10 ppm apart from each other, demonstrating its superior capability of aligning peaks from MS spectra at a higher resolution. In fact, the 5-ppm cut-off was used in the current study, whereas users can determine the suitable values according to the specific studies. Lower cut-off values can be potentially used to treat MS data acquired using mass spectrometers with higher resolving power. The examples of raw data as well as alignment results obtained using Geena 2, MZmine 3, and MassLite are provided in the Supplementary Material (Tables S1–S4).

3.4. Computational cost for alignment

The computational cost, including both CPU time and memory usage, for peak alignment was evaluated with or without dynamic

Table 1
Time cost for peak alignment.

16.2 MB dataset			2.83 GB dataset		
Group size*	Alignment cost (s)	Normalized cost	Group size	Alignment cost (h)	Normalized cost
2 × n	3.17	34 %	1.5 × n	5.4	11 %
3 × n	3.13	34 %	2 × n	7.1	15 %
10 × n	4.53	49 %	3 × n	10.4	22 %
Without grouping	9.21	100 %	Without grouping	>48	100 %

n: number of total scans in the file.

grouping. The computational cost depends on both the total number of scans and the total number of peaks in each scan. Because of the variance among the MS profiles in each particular scan, the number of peaks in each scan is subject to change. The total number of peaks is positively correlated, but in a non-linear fashion, with the number of scans. Among all data pretreatment steps, peak alignment without binning is the most expensive part due to the pairwise distance calculation and distance matrix update during hierarchical clustering. Regular pairwise comparison (i.e., without grouping) between peaks requires computational cost to the second power of total number of peaks. Although binning can reduce the computational cost, the loss of mass accuracy in peak alignment step limited its applicability on SCMS metabolomics data. To overcome these challenges, we proposed a dynamic grouping method (Fig. 3). Dynamic grouping reduces the computational cost for peak alignment using a "divide and conquer" strategy. When the whole dataset was divided into multiple chunks, the number of unnecessary comparisons between peaks was largely reduced. Particularly, this strategy eliminated the comparisons between peaks from different chunks, which could theoretically reduce the cost by second power to the number of chunks. To overcome the potential peak splitting issue due to the boundaries of the chunks, we implemented an automatic check at the boundary of adjacent neighboring chunks. This automatic check method can merge artificially split peaks due to the chunk division, which slightly increased the cost by the first power to the number of chunks. Because computational cost reduction using dynamic grouping depends on multiple factors (e.g., dataset size, total number of peaks, and total number of scans and number of chunks), we tested datasets with a small size (16.2 MB *imzML* file) and a large size (2.83 GB *imzML* file) (Table 1). We discovered that, compared results without grouping, the time used for peak alignment using dynamic grouping (with optimized group sizes) was reduced to ~1/3 and ~1/10 for the small and large datasets, respectively. In addition to time cost, memory usage is another major concern because storing all pairwise distances (i.e., without grouping) for millions of peaks, which lead to trillions of distances, can occupy several TBs of memory, potentially resulting in a breakdown of the program. Dynamic grouping significantly reduced both CPU time and memory usage while providing reasonable results, allowing for customizable studies using a local computer. Additional details can be found in the Supplementary Material. The method can be potentially improved when multiple cores are available for parallel processing.

4. Conclusion

We developed MassLite, a Python-based GUI platform, for the pretreatment of SCMS metabolomics data, including void scan filter, cell scan grouping, peak alignment, and background removal. Experimental data can be converted into a standard MS data format *.mzML* and then processed by MassLite. An algebraic transformation has been introduced to describe relative m/z difference in an intuitive manner, enabling faster processing in the following steps. A novel peak alignment has been implemented into MassLite, allowing for extraction of ion signals with more accurate m/z values of peaks, including those with low abundances. This function is especially important for untargeted chromatography-free SCMS metabolomics studies, in which accurate m/z values provide critical information for molecular identification. Because all results can be stored prior to exportation, the trade-off between 'keeping more low-abundance signal' and 'removing more noise' can be tuned by users using different parameters. The automatic algorithms, which were used for void scan filtering and cell scan selection, allowed higher throughput and more robust analysis outcome. This platform can effectively remove background signal and noise, eliminating artifacts in the follow-up analysis with significantly reduced computational cost. Importantly, MassLite is capable of retaining low-intensity peaks among complex signals, providing better chances to find more molecules from limited analytes in single cells. We expect MassLite to be smoothly adopted to analyze SCMS data collected using other types of experimental setups.

Code and data availability

Source code of MassLite is available on GitHub: https://github.com/chemzzchem/MassLite/blob/main/published%20versions/. Raw data of the Single-probe SCMS experiments can be obtained from the MassIVE database (MSV000095500).

CRediT authorship contribution statement

Zhu Zou: Writing – original draft, Visualization, Validation, Software, Investigation, Formal analysis, Conceptualization. **Zongkai Peng:** Formal analysis, Data curation. **Deepti Bhusal:** Formal analysis, Data curation. **Shakya Wije Munige:** Formal analysis, Data curation. **Zhibo Yang:** Writing – review & editing, Supervision, Resources, Project administration, Methodology, Investigation, Funding acquisition.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

Data will be made available on request.

Acknowledgement

This work was supported by funds from National Institutes of Health (1R01AI177469), National Science Foundation (2305182), and Chan Zuckerberg Initiative.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.aca.2024.343124.

References

[1] J.B. Fenn, M. Mann, C.K. Meng, S.F. Wong, C.M. Whitehouse, Electrospray ionization for mass spectrometry of large biomolecules, Science 246 (4926) (1989) 64–71.

[2] M. Karas, F. Hillenkamp, Laser desorption ionization of proteins with molecular masses exceeding 10,000 daltons, Anal. Chem. 60 (20) (1988) 2299–2301.
[3] C.A. Crutchfield, W. Clarke, Present and future applications of high resolution mass spectrometry in the clinic, Discoveries 2 (2) (2014) e17.
[4] A.B. Baxi, L.R. Pade, P. Nemes, Mass spectrometry based proteomics for developmental neurobiology in the amphibian Xenopus laevis, Curr. Top. Dev. Biol. 145 (2021) 205–231.
[5] P. Vangeenderhuysen, J. Van Arnhem, B. Pomian, M. De Graeve, L. De Commer, G. Falony, J. Raes, A. Zhernakova, J. Fu, L.Y. Hemeryck, et al., Dual UHPLC-HRMS metabolomics and lipidomics and automated data processing workflow for comprehensive high-throughput gut phenotyping, Anal. Chem. 95 (22) (2023) 8461–8468.
[6] A. Martín-Blázquez, C. Díaz, E. González-Flores, D. Franco-Rivas, C. Jiménez-Luna, C. Melguizo, J. Prados, O. Genilloud, F. Vicente, O. Caba, et al., Untargeted LC-HRMS-based metabolomics to identify novel biomarkers of metastatic colorectal cancer, Sci. Rep. 9 (1) (2019) 20198.
[7] L.H.J. Richter, C.M. Jacobs, F. Mahfoud, I. Kindermann, M. Böhm, M.R. Meyer, Development and application of a LC-HRMS/MS method for analyzing antihypertensive drugs in oral fluid for monitoring drug adherence, Anal. Chim. Acta 1070 (2019) 69–79.
[8] J.J. Pitt, Principles and applications of liquid chromatography-mass spectrometry in clinical biochemistry, Clin. Biochem. Rev. 30 (1) (2009) 19–34.
[9] M.R.L. Paine, J. Liu, D. Huang, S.R. Ellis, D. Trede, J.H. Kobarg, R.M.A. Heeren, F. M. Fernández, T.J. MacDonald, Three-dimensional mass spectrometry imaging identifies lipid markers of medulloblastoma metastasis, Sci. Rep. 9 (1) (2019) 2205.
[10] X. Tian, B. Xie, Z. Zou, Y. Jiao, L.-E. Lin, C.-L. Chen, C.-C. Hsu, J. Peng, Z. Yang, Multimodal imaging of amyloid plaques: fusion of the single-probe mass spectrometry image and fluorescence microscopy image, Anal. Chem. 91 (20) (2019) 12882–12889.
[11] P. Xie, H. Zhang, P. Wu, Y. Chen, Z. Cai, Three-dimensional mass spectrometry imaging reveals distributions of lipids and the drug metabolite associated with the enhanced growth of colon cancer cell spheroids treated with triclosan, Anal. Chem. 94 (40) (2022) 13667–13675.
[12] M. Aichler, A. Walch, MALDI Imaging mass spectrometry: current frontiers and perspectives in pathology research and practice, Lab. Invest. 95 (4) (2015) 422–431.
[13] L.J. Gamble, C.R. Anderton, Secondary ion mass spectrometry imaging of tissues, cells, and microbial systems, Micros Today 24 (2) (2016) 24–31.
[14] L.Z. Samarah, A. Vertes, Mass spectrometry imaging based on laser desorption ionization from inorganic and nanophotonic platforms, View 1 (4) (2020) 20200063.
[15] E. Claude, E.A. Jones, S.D. Pringle, DESI mass spectrometry imaging (MSI), Methods Mol. Biol. 1618 (2017) 65–75.
[16] X. Li, H. Hu, R. Yin, Y. Li, X. Sun, S.K. Dey, J. Laskin, High-throughput nano-DESI mass spectrometry imaging of biological tissues using an integrated microfluidic probe, Anal. Chem. 94 (27) (2022) 9690–9696.
[17] W. Rao, N. Pan, Z. Yang, High resolution tissue imaging using the single-probe mass spectrometry under ambient conditions, J. Am. Soc. Mass Spectrom. 26 (6) (2015) 986–993.
[18] M. Tajik, M. Baharfar, W.A. Donald, Single-cell mass spectrometry, Trends Biotechnol. 40 (11) (2022) 1374–1392.
[19] L. Zhang, A. Vertes, Single-cell mass spectrometry approaches to explore cellular heterogeneity, Angew Chem. Int. Ed. Engl. 57 (17) (2018) 4466–4477.
[20] S. Lee, H.M. Vu, J.H. Lee, H. Lim, M.S. Kim, Advances in mass spectrometry-based single cell analysis, Biology 12 (3) (2023).
[21] H.M. Bennett, W. Stephenson, C.M. Rose, S. Darmanis, Single-cell proteomics enabled by next-generation sequencing or mass spectrometry, Nat. Methods 20 (3) (2023) 363–374.
[22] Single-cell proteomics: challenges and prospects, Nat. Methods 20 (3) (2023) 317–318.
[23] D.C. Castro, Y.R. Xie, S.S. Rubakhin, E.V. Romanova, J.V. Sweedler, Image-guided MALDI mass spectrometry for high-throughput single-organelle characterization, Nat. Methods 18 (10) (2021) 1233–1238.
[24] J. Nuñez, R. Renslow, J.B. Cliff 3rd, C.R. Anderton, NanoSIMS for biological applications: current practices and analyses, Biointerphases 13 (3) (2017) 03b301.
[25] C.R. Ferreira, V. Pirro, A.K. Jarmusch, C.M. Alfaro, R.G. Cooks, Ambient lipidomic analysis of single mammalian oocytes and preimplantation embryos using desorption electrospray ionization (DESI) mass spectrometry, Methods Mol. Biol. 2064 (2020) 159–179.
[26] M.J. Taylor, J.K. Lukowski, C.R. Anderton, Spatially resolved mass spectrometry at the single cell: recent innovations in proteomics and metabolomics, J. Am. Soc. Mass Spectrom. 32 (4) (2021) 872–894.
[27] H.-M. Bergman, I. Lanekoff, Profiling and quantifying endogenous molecules in single cells using nano-DESI MS, Analyst 142 (19) (2017) 3639–3647.
[28] N. Pan, W. Rao, N.R. Kothapalli, R. Liu, A.W. Burgett, Z. Yang, The single-probe: a miniaturized multifunctional device for single cell mass spectrometry analysis, Anal. Chem. 86 (19) (2014) 9376–9380.
[29] T. Nakashima, H. Wada, S. Morita, R. Erra-Balsells, K. Hiraoka, H. Nonami, Single-cell metabolite profiling of stalk and glandular cells of intact trichomes with internal electrode capillary pressure probe electrospray ionization mass spectrometry, Anal. Chem. 88 (6) (2016) 3049–3057.
[30] R. Yin, V. Prabhakaran, J. Laskin, Quantitative extraction and mass spectrometry analysis at a single-cell level, Anal. Chem. 90 (13) (2018) 7937–7945.
[31] J.F. Cahill, J. Riba, V. Rapid Kertesz, Untargeted chemical profiling of single cells in their native environment, Anal. Chem. 91 (9) (2019) 6118–6126.
[32] M. Katajamaa, J. Miettinen, M. MZmine Orešič, Toolbox for processing and visualization of mass spectrometry based molecular profile data, Bioinformatics 22 (5) (2006) 634–636.
[33] T. Pluskal, S. Castillo, A. Villar-Briones, M. Orešič, MZmine 2: modular framework for processing, visualizing, and analyzing mass spectrometry-based molecular profile data, BMC Bioinf. 11 (1) (2010) 395.
[34] R. Schmid, S. Heuckeroth, A. Korf, A. Smirnov, O. Myers, T.S. Dyrlund, R. Bushuiev, K.J. Murray, N. Hoffmann, M. Lu, et al., Integrative analysis of multimodal mass spectrometry data in MZmine 3, Nat. Biotechnol. 41 (4) (2023) 447–449.
[35] C.A. Smith, E.J. Want, G. O'Maille, R. Abagyan, G. Siuzdak, XCMS: processing mass spectrometry data for metabolite profiling using nonlinear peak alignment, matching, and identification, Anal. Chem. 78 (3) (2006) 779–787.
[36] E.M. Forsberg, T. Huan, D. Rinehart, H.P. Benton, B. Warth, B. Hilmers, G. Siuzdak, Data processing, multi-omic pathway mapping, and metabolite activity analysis using XCMS Online, Nat. Protoc. 13 (4) (2018) 633–651.
[37] 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 (14) (2015) 2418–2420.
[38] A. Palmer, P. Phapale, I. Chernyavsky, R. Lavigne, D. Fay, A. Tarasov, V. Kovalev, J. Fuchser, S. Nikolenko, C. Pineau, et al., FDR-controlled metabolite annotation for high-resolution imaging mass spectrometry, Nat. Methods 14 (1) (2017) 57–60.
[39] S.M. Williams, A.V. Liyu, C.-F. Tsai, R.J. Moore, D.J. Orton, W.B. Chrisler, M. J. Gaffrey, T. Liu, R.D. Smith, R.T. Kelly, et al., Automated coupling of nanodroplet sample preparation with liquid chromatography–mass spectrometry for high-throughput single-cell proteomics, Anal. Chem. 92 (15) (2020) 10588–10596.
[40] Y. Liang, T. Truong, Y. Zhu, R.T. Kelly, In-depth mass spectrometry-based single-cell and nanoscale proteomics, Methods Mol. Biol. 2185 (2021) 159–179.
[41] Y. Cong, Y. Liang, K. Motamedchaboki, R. Huguet, T. Truong, R. Zhao, Y. Shen, D. Lopez-Ferrer, Y. Zhu, R.T. Kelly, Improved single-cell proteome coverage using narrow-bore packed NanoLC columns and ultrasensitive mass spectrometry, Anal. Chem. 92 (3) (2020) 2665–2671.
[42] C. Lombard-Banek, S.A. Moody, P. Nemes, Single-cell mass spectrometry for discovery proteomics: quantifying translational cell heterogeneity in the 16-cell frog (Xenopus) embryo, Angew. Chem. Int. Ed. 55 (7) (2016) 2454–2458.
[43] I. Virant-Klun, S. Leicht, C. Hughes, J. Krijgsveld, Identification of maturation-specific proteins by single-cell proteomics of human oocytes, Mol. Cell. Proteomics 15 (8) (2016) 2616–2627.
[44] E.K. Neumann, T.J. Comi, S.S. Rubakhin, J.V. Sweedler, Lipid heterogeneity between astrocytes and neurons revealed by single-cell MALDI-MS combined with immunocytochemical classification, Angew. Chem. Int. Ed. 58 (18) (2019) 5910–5914.
[45] Y. Shao, Y. Zhou, Y. Liu, W. Zhang, G. Zhu, Y. Zhao, Q. Zhang, H. Yao, H. Zhao, G. Guo, et al., Intact living-cell electrolaunching ionization mass spectrometry for single-cell metabolomics, Chem. Sci. 13 (27) (2022) 8065–8073.
[46] R. Liu, G. Zhang, Z. Yang, Towards rapid prediction of drug-resistant cancer cell phenotypes: single cell mass spectrometry combined with machine learning, Chem. Commun. 55 (5) (2019) 616–619.
[47] R. Liu, Z. Yang, Single cell metabolomics using mass spectrometry: techniques and data analysis, Anal. Chim. Acta 1143 (2021) 124–134.
[48] R. Liu, G. Zhang, M. Sun, X. Pan, Z. Yang, Integrating a generalized data analysis workflow with the Single-probe mass spectrometry experiment for single cell metabolomics, Anal. Chim. Acta 1064 (2019) 71–79.
[49] N. Pan, W. Rao, Z. Yang, Single-probe mass spectrometry analysis of metabolites in single cells, Methods Mol. Biol. 2064 (2020) 61–71.
[50] Y. Gholipour, R. Erra-Balsells, K. Hiraoka, H. Nonami, Living cell manipulation, manageable sampling, and shotgun picoliter electrospray mass spectrometry for profiling metabolites, Anal. Biochem. 433 (1) (2013) 70–78.
[51] A.G. Brenton, A.R. Godfrey, Accurate mass measurement: terminology and treatment of data, J. Am. Soc. Mass Spectrom. 21 (11) (2010) 1821–1835.
[52] J.P. Finch, T. Wilson, L. Lyons, H. Phillips, M. Beckmann, J. Draper, Spectral binning as an approach to post-acquisition processing of high resolution FIE-MS metabolome fingerprinting data, Metabolomics 18 (8) (2022) 64.
[53] X. Feng, W. Zhang, F. Kuipers, I. Kema, A. Barcaru, P. Horvatovich, Dynamic binning peak detection and assessment of various lipidomics liquid chromatography-mass spectrometry pre-processing platforms, Anal. Chim. Acta 1173 (2021) 338674.
[54] S. Krishnan, J.T.W.E. Vogels, L. Coulier, R.C. Bas, M.W.B. Hendriks, T. Hankemeier, U. Thissen, Instrument and process independent binning and baseline correction methods for liquid chromatography–high resolution-mass spectrometry deconvolution, Anal. Chim. Acta 740 (2012) 12–19.
[55] J. Urban, Resolution, precision, and entropy as binning problem in mass spectrometry, in: I. Rojas, F. Ortuño (Eds.), Bioinformatics and Biomedical Engineering, Springer International Publishing, Cham, 2018//, 2018, pp. 118–128.
[56] X. Chen, M. Sun, Z. Yang, Single cell mass spectrometry analysis of drug-resistant cancer cells: metabolomics studies of synergetic effect of combinational treatment, Anal. Chim. Acta 1201 (2022) 339621.
[57] M. Sun, X. Tian, Z. Yang, Microscale mass spectrometry analysis of extracellular metabolites in live multicellular tumor spheroids, Anal. Chem. 89 (17) (2017) 9069–9076.
[58] J.D. Holman, D.L. Tabb, P. Mallick, Employing ProteoWizard to convert raw mass spectrometry data, Curr Protoc Bioinformatics 46 (2014) 13.24.11–13.24.19.
[59] M.C. Chambers, B. Maclean, R. Burke, D. Amodei, D.L. Ruderman, S. Neumann, L. Gatto, B. Fischer, B. Pratt, J. Egertson, et al., A cross-platform toolkit for mass spectrometry and proteomics, Nat. Biotechnol. 30 (10) (2012) 918–920.

[60] D. Kessner, M. Chambers, R. Burke, D. Agus, P. Mallick, ProteoWizard: open source software for rapid proteomics tools development, Bioinformatics 24 (21) (2008) 2534–2536.
[61] A.G. Brenton, A.R. Godfrey, Accurate mass measurement: terminology and treatment of data, J. Am. Soc. Mass Spectrom. 21 (11) (2010) 1821–1835.
[62] Y. Wang, M. Gu, The concept of spectral accuracy for MS, Anal. Chem. 82 (17) (2010) 7055–7062.
[63] A. McCann, S. Rappe, R. La Rocca, M. Tiquet, L. Quinton, G. Eppe, J. Far, E. De Pauw, C. Kune, Mass shift in mass spectrometry imaging: comprehensive analysis and practical corrective workflow, Anal. Bioanal. Chem. 413 (10) (2021) 2831–2844.
[64] T.D. Nguyen, Y. Lan, S.S. Kane, J.J. Haffner, R. Liu, L.I. McCall, Z. Yang, Single-cell mass spectrometry enables insight into heterogeneity in infectious disease, Anal. Chem. 94 (30) (2022) 10567–10572.
[65] P. Romano, A. Profumo, M. Rocco, R. Mangerini, F. Ferri, A. Facchiano, Geena 2, improved automated analysis of MALDI/TOF mass spectra, BMC Bioinf. 17 (4) (2016) 61.