

SIGNAL PREPROCESSING, MULTIVARIATE ANALYSIS AND SOFTWARE TOOLS FOR MA(LDI)-TOF MASS SPECTROMETRY IMAGING FOR BIOLOGICAL APPLICATIONS

**Pere Ràfols,^{1,2} Ddac Vilalta,^{1,2} Jesús Brezmes,^{1,2}
Nicolau Canellas,^{1,2} Esteban del Castillo,² Oscar Yanes,^{1,2}
Noelia Ramírez,^{1,2*} and Xavier Correig^{1,2}**

¹Spanish Biomedical Research Center in Diabetes and Associated Metabolic Disorders (CIBERDEM), C/Monforte de Lemos 3-5, Madrid 28029, Spain

²Department of Electronic Engineering, Institute of Health Research Pere Virgili, Rovira i Virgili University, IISPV, Avinguda Pàsos Catalans 26, Tarragona 43007, Spain

Received 6 October 2016; accepted 11 October 2016

Published online in Wiley Online Library (wileyonlinelibrary.com). DOI 10.1002/mas.21527

Mass spectrometry imaging (MSI) is a label-free analytical technique capable of molecularly characterizing biological samples, including tissues and cell lines. The constant development of analytical instrumentation and strategies over the previous decade makes MSI a key tool in clinical research. Nevertheless, most MSI studies are limited to targeted analysis or the mere visualization of a few molecular species (proteins, peptides, metabolites, or lipids) in a region of interest without fully exploiting the possibilities inherent in the MSI technique, such as tissue classification and segmentation or the identification of relevant biomarkers from an untargeted approach. MSI data processing is challenging due to several factors. The large volume of mass spectra involved in a MSI experiment makes choosing the correct computational strategies critical. Furthermore, pixel to pixel variation inherent in the technique makes choosing the correct preprocessing steps critical. The primary aim of this review was to provide an overview of the data-processing steps and tools that can be applied to an MSI experiment, from preprocessing the raw data to the more advanced strategies for image visualization and segmentation. This review is particularly aimed at researchers performing MSI experiments and who are interested in incorporating new data-processing features, improving their computational strategy, and/or desire access to data-processing tools currently available. © 2016 Wiley Periodicals, Inc. Mass Spec Rev 9999: XX–XX, 2016.

Keywords: mass spectrometry imaging (MSI); MSI preprocessing; MSI multivariate analysis; MSI computational requirements; MSI software tools

I. INTRODUCTION

In recent years, mass spectrometry imaging (MSI), also called imaging mass spectrometry (IMS), has become a key analytical technique in proteomics, lipidomics, metabolomics (Hillenkamp & Peter-Katalinic, 2014) and related research fields, such as drug discovery and toxicology (Greer, Sturm, & Li, 2011; Prideaux & Stoeckli, 2012). MSI provides molecule-specific images that enable correlation of the spatial occurrence of target molecules and their abundance by direct analysis of biological samples without labeling or staining. To date, hundreds of biological and clinical MSI applications can be found in the literature detailing tissue-based disease classification, discovery of phenotypic intra-tumor heterogeneity, therapy-response prediction and prognosis, and drug development in the fields of oncology, pathology, diagnostics, and surgery (Addie et al., 2015; Aichler & Walch, 2015; Kriegsmann, Kriegsmann, & Casadonte, 2015).

Among analytical strategies used for MSI, matrix-assisted laser desorption/ionization mass spectrometry (MALDI-TOF) (Nilsson et al., 2015; Norris & Caprioli, 2013) is the most commonly used technique due to its simplicity, soft ionization, fast analysis, high throughput, and the versatility and selectivity ensured by a wide range of successfully used organic matrices. Furthermore, the recent developments in MALDI-TOF instrumentation allow for high-throughput acquisition of high-resolution MS images, revealing MSI as a potential tool for diagnostics and clinical applications.

Nevertheless, MSI applications are sometimes limited by the complexity of data processing due, among other factors, to the large amount of raw data generated, peak misalignment during image acquisition, or adduct formation and/or molecule fragmentation produced by the desorption/ionization processes. Therefore, the aim of this review was to provide an overview of the data-processing steps necessary for MS data treatment and visualization of MS images in proteomics, lipidomics, or metabolomics, as well as the processing tools and software currently available. This review consists of seven sections that include an introduction, a brief description of the MSI workflow, data-preprocessing steps, multivariate analysis, data-handling strategies and considerations, currently available software packages, and concluding remarks. In 2012, two reviews concerning

Contract grant sponsor: Spanish Ministry of Economía y Competitividad; Contract grant numbers: TEC2012-31074, TEC2015-69076-P, BES-2013-065572; Contract grant sponsor: Direcció General de Recerca of the Government of Catalonia; Contract grant number: 2014 SGR 01267; Contract grant sponsor: Marie Skłodowska-Curie; Contract grant number: 660034.

*Correspondence to: Noelia Ramírez, Department of Electronic Engineering, Institute of Health Research Pere Virgili, Rovira i Virgili University, IISPV, Avinguda Pàsos Catalans 26, Tarragona 43007, Spain. E-mail: noelia.ramirez@urv.cat

the data processing of MALDI-based MSI were published and mainly focused on proteomics applications (Alexandrov, 2012; Jones et al., 2012a). More recently, another review was published focusing on strategies for data mining and visualization of 3D images (Thiele et al., 2014). Our review extends the information provided in these previous works by both collecting and reporting on the most updated bibliography in this field and specifically addressing aspects not reviewed previously, such as data formats and other computational considerations, as well as the currently available software tools and the specific problems derived from the use of matrix-free methods currently employed in metabolomics applications.

Although, this review focuses on the data-processing challenges of MALDI-MS and matrix-free LDI-MS in proteomics, lipidomics, and metabolomics applications, the computational and statistical strategies discussed here can generally be applied to other MSI approaches. We hope with this review to encourage researchers currently performing MSI experiments to incorporate new data-processing features to either improve their computational strategies or broaden their knowledge regarding the data processing tools currently available.

II. MSI WORKFLOW

A typical MSI workflow has three main steps: sample preparation, MS acquisition, and data processing and visualization. As

an example, Figure 1 shows a typical MALDI-MS experimental workflow. In this section, we review the first two steps of the workflow and their influence on the subsequent data-processing step. Further information regarding MALDI-MS experiments and their basis can be found in recent reviews (Norris & Caprioli, 2013; Hillenkamp & Peter-Katalinic, 2014).

Sample handling and preparation is key to optimizing sensitivity and spatial resolution (Goodwin, 2012; Norris & Caprioli, 2013), and parameters, such as tissue-section thickness (generally 3–20 μm), must be optimized for the analytical platform selected for data acquisition. Biological tissues are usually snap frozen and stored at -80°C immediately after collection. MSI measurement of tissues fixed in paraffin- or alcohol-embedding media is not straightforward, because the molecules of the fixing material interfere and can cause contamination and ion suppression (Chughtai & Heeren, 2010; Goodwin, 2012; Norris & Caprioli, 2013). However, it was recently demonstrated that it is possible to perform MSI experiments from formalin-fixed and paraffin-embedded clinical tissue samples (Buck et al., 2015).

In MALDI-MS-based MSI, an organic matrix is deposited over the tissue to assist in ionization. Standard matrix-deposition techniques consisting of deposition by the spraying of organic matrices (i.e., α -cyano-4-hydroxycinnamic acid, 2,5-dihydroxybenzoic acid, etc.) could lead to metabolite delocalization (compromising the spatial resolution) and the formation of

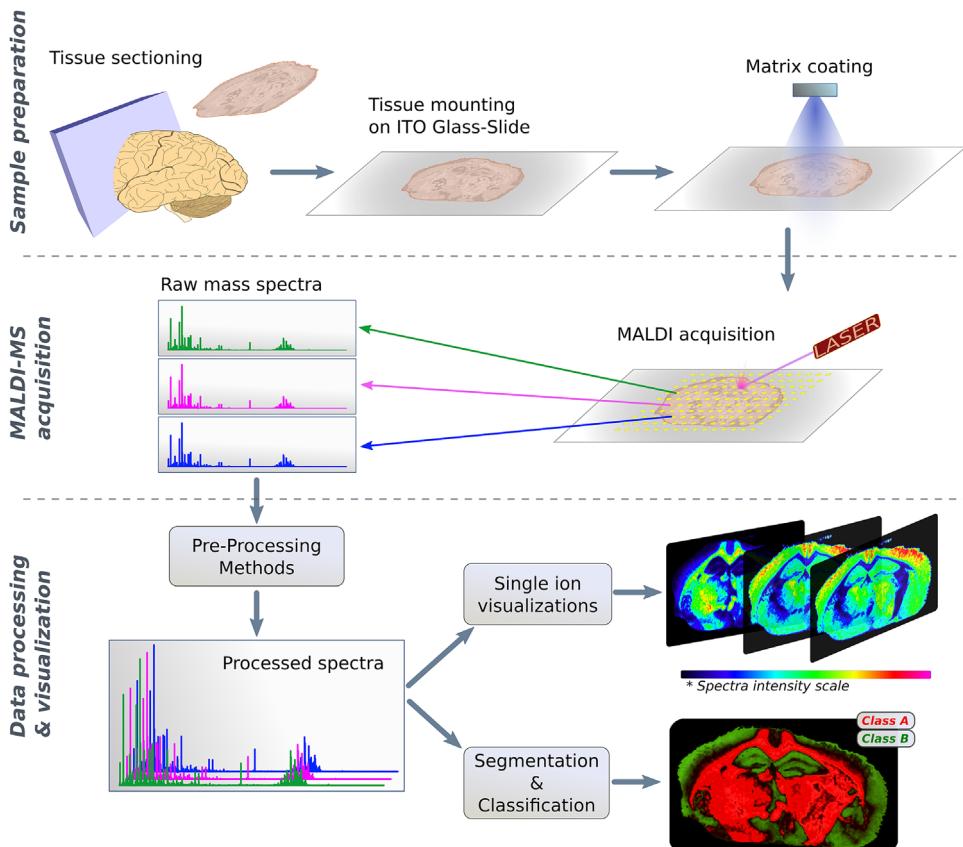


FIGURE 1. A typical MALDI-MS experiment workflow. In the sample-preparation stage, tissue is sectioned, fixed on indium tin oxide glass slides, and coated with matrix. MS spectra are then acquired using a MALDI instrument. Raw spectra are preprocessed, single-ion images are visualized, and a segmented image is displayed.

heterogeneities that cause unexpected variations in signal intensities and background noise. These affect biological interpretation of the results and determine the application of specific data-processing algorithms. Matrix effects resulting from ionization of the matrix compounds are also common in MALDI-MS experiments and interfere and suppress MS signals in the m/z region <1000 Da, which is the common m/z region in metabolomics experiments. Nevertheless, several strategies were recently developed to minimize analyte delocalization and improve sensitivity and imaging spatial resolution (Gemperline, Rawson, & Li, 2014; Lauzon et al., 2015) and overcome interference from matrix peaks (Chen et al., 2012; Bergman, Shevchenko, & Bergquist, 2014). Matrix-free LDI-MS platforms, such as surface-assisted laser/desorption ionization (SALDI) (Chiang, Chen, & Chang, 2011; Silina & Volmer, 2013; Bergman, Shevchenko, & Bergquist, 2014) or nanostructure-initiator mass spectrometry (NIMS) (Northen et al., 2007), have recently emerged as valuable alternatives, especially for the analysis of low-molecular-weight metabolites, offering minimal analyte delocalization and fewer background peaks <1000 Da. Furthermore, the recent application of metal and metal oxide nanoparticles and nanolayers to MSI (frequently called nano-PALDI-MSI) is opening up a wide range of possible approaches in this field. The main advantages of this technique are the few interfering peaks in the low- m/z area of the metal nanolayers, the high homogeneity of the surfaces, and high spatial resolution (down to 5 μm and only limited by the diameter of the laser) (Dufresne et al., 2013). The main drawback is the possible formation of metal and metal oxide adducts of the metabolites with the different isotopic forms of the metals, which can make metabolite identification more difficult. Nevertheless, the characteristic metal peaks and clusters can be used for internal mass calibration throughout the various m/z regions of the obtained spectrum (Dufresne et al., 2013).

Following sample preparation, an ultraviolet (UV) or infrared (IR) laser is used in MA(LDI)-MS to desorb and ionize the molecules. The mechanisms involved in desorption/ionization are still not fully understood and depend upon the LDI approach (McDonnell & Heeren, 2007; Rainer, Qureshi, & Bonn, 2011; Calavia et al., 2012; Silina & Volmer, 2013; Hillenkamp & Peter-Katalinic, 2014). The spatial resolution of the MS image is determined by the matrix-crystal size, the possible lateral compound diffusion occurring along the matrix-deposition process, and the laser-beam diameter of a specific instrument [normally between 10 and 250 μm (Norris & Caprioli, 2013)]. One strategy to reach spatial resolutions below the beam diameter involves use of an oversampling method (Norris & Caprioli, 2013). Although at low spatial resolutions smaller tissue regions can be molecularly characterized, the acquisition time increases and the quality of the MSI worsens due to the abundance of lower MS peaks in the acquired spectra. Furthermore, lower resolutions generate higher volumes of data and, therefore, the need of sophisticated computational strategies. As an example of acquisition time, a laser operating at 2 KHz can perform a simple pixel measurement within 1 sec, enabling acquisition of a 1 cm \times 1 cm tissue sample over 1 hr at a lateral resolution of 100 μm . The increase of the lateral resolution by a factor of two causes a fourfold increase in acquisition time. Nevertheless, it is

worth mentioning that recent developments in MALDI instruments could significantly decrease acquisition time. The recently released Bruker RapifleX MALDI Tissuetyper spectrometer (Bruker Daltonics, Billerica, MA, USA) is capable of acquiring 50 pixels/s, resulting in <2 min data acquisition for a 1 cm \times 1 cm MS image. Therefore, the spatial resolution of each experiment must be fixed as a compromise between the abovementioned factors.

The MS platform most suitable for each application depends upon the sensitivity required, dynamic range (the range of analyte concentration that can be detected), mass accuracy, and resolving power. Time of flight (TOF) analyzers are the most commonly used detectors, especially in MALDI applications for proteomics (Norris & Caprioli, 2013). The most common type of detector is the axial TOF spectrometer, which provides a mass-accuracy error between 10 and 20 ppm due to the initial velocity/drift of the generated ions. The addition of an ion reflector together with delayed ion extraction helps to compensate for this effect, which can result from non-flat-sample morphology. Using this configuration, mass accuracies of 5–10 ppm can be achieved. Modern MALDI spectrometers are equipped with an orthogonal reflector capable of deflecting the ions perpendicular to the original direction of motion, thereby eliminating the high initial axial-velocity distribution of the plume. Mass errors <10 ppm are common with this configuration. If higher mass resolution is needed, Fourier transform orbitrap (FT-orbitrap) and Fourier transform ion cyclotron resonance (FT-ICR) mass spectrometers are available, with mass errors <1 ppm at m/z 300 (Hölscher et al., 2015), which makes it easier to identify compounds by their exact mass. Tandem mass spectrometry (MS/MS), a feature commonly found in MS detectors, also increases selectivity and improves identification power. MS-acquisition ranges differ depending on the MSI application, from masses <1000 Da for metabolomics to thousands of Da for proteomics.

III. IMAGE PREPROCESSING

The preprocessing stage is fundamental for any MSI experiment, because the quality of the MS images depends largely upon the appropriateness of the previous preprocessing operations. The experimental variability in mass spectrometry derives from sample-preparation procedures and MAL(LDI)-MS acquisition. This variability is reflected in the raw data by introducing chemical noise, variations in the intensity and exact mass of each MS peak. In the case of large samples or high-resolution images, the overall MS spectra intensities can drift during acquisition due to instrumental reasons, such as the deposition of debris on the MS-ionization source (Jones et al., 2012a). As reference for the magnitude of this drift, we can observe $\sim 30\%$ intensity reduction during the acquisition of a MS image of >8000 pixels acquired at 500 shots per pixel using a commercial MALDI-TOF spectrometer.

The purpose of preprocessing is to improve image reconstruction by reducing the unwanted effects introduced by experimental variation and sample preparation. A carefully designed preprocessing workflow also helps the peak-picking procedure, the process of converting a mass spectrum into a list of relevant features for further data analysis and biological interpretation, making the statistical analysis more robust and reliable. In a typical MSI-preprocessing pipeline, the common algorithms are as follows:

baseline correction, noise reduction, spectral alignment, normalization, peak picking, binning, and removal of matrix peaks. The order of the preprocessing steps is not a fixed sequence and should be adapted to accommodate the requirements of each application. Figure 2 illustrates each of the preprocessing steps described below using simulated data. Some of these steps may be omitted or computed in a different order, depending on the experiment. Table 1 summarizes the preprocessing methods used to date for MSI. Notably, most data-processing methods are focused on single pixel/spectrum processing and, therefore, can be also used in other MS applications.

A. Baseline Correction

In MS, the baseline is the smooth curve offsetting the actual compound peaks throughout the spectrum. This signal is clearly identifiable and interferes in the base of the MS peaks, especially on those with low intensity. The effect of baseline can be observed by comparing the RAW spectra in Figure 2A with the baseline-corrected spectra in Figure 2B. Several baseline-correction algorithms can be used to correct this effect. One of the most common algorithms is Top-Hat

(van Herk, 1992), which applies a moving minimum (called an erosion filter) and subsequently a moving maximum (called a dilation filter) to the intensity values. Another generic method consists of fitting the baseline with a monotonic decay function and subtracting it from the spectra. Källback et al. (2012) compared three methods for baseline estimation based on sliding windows (simple moving first quartile [SMQ1], simple moving average [SMA], and simple moving median [SMM]), concluding that SMQ1 provided a better baseline correction with minimal peak deformation. Another approach based on peak detection was introduced for baseline correction in the LIMPIC software package (Mantini et al., 2007). In this package, MS peaks obtained from sample acquisition are detected and removed from the spectrum. The resulting function is an approximation of the baseline profile. In this method, the peak-detection threshold must be adjusted to obtain accurate baseline estimations. Another strategy for baseline correction consists of standardizing the intensity of each spectrum in a defined mass range using statistical methods. This complex and powerful method provides an estimated baseline minimally affected by higher peaks (Satten et al., 2004).

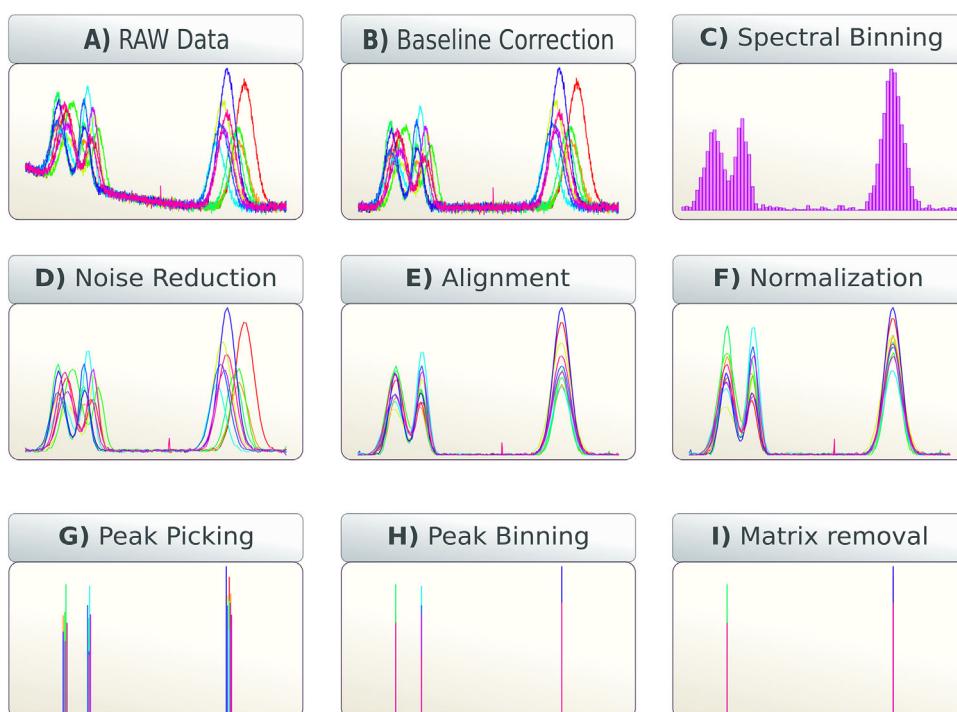


FIGURE 2. Representation of MSI-preprocessing steps (figures created with simulated data). (A) RAW MS spectra before processing. (B) MS spectra after baseline correction. The baseline was estimated and then subtracted using the Top-Hat method. (C) Spectral binning used to reduce the number of data points. This use of binning is generally performed as one of the first preprocessing steps to take advantage of data reduction. Here each data point are mapped to nearest mass bin represented as a bar in the graphic. (D) Noise reduction using a Savitzky–Golay smoothing routine. The spectral random noise is drastically reduced and the peaks shape is retained. (E) Mass alignment and calibration that allow correction of possible mass drifts. Here, some peaks were identified as reference compounds and used to calculate m/z shifts and minimize the drift. (F) Intensity normalization applied to reduce variability. All spectra are mapped to a similar intensity scale using TIC normalization. (G) Peak picking that reduces each spectrum to a list of MS peaks. Each peak list is represented here as a color line pointing to the original peak location. Each detected peak retains information of: m/z , intensity and signal to noise ratio (SNR). (H) Peak binning applied immediately after peak picking. Binning is used after peak peaking to eliminate slight mass shifts between the same detected compounds throughout all spectra. (I) Matrix-peaks removal. Peaks known as matrix peaks are removed from the binned peak list.

TABLE 1. Preprocessing methods summary.

Ref.	Baseline Correction	Background noise removal	Noise reduction	Normalization	Alignment	Peak Picking	Binning	Matrix peak removal
(Norris et al., 2007)	DataExplorer, FlexAnalysis, BioMap			Cube root, Logarithmic, TIC, Noise scaled (wavelet based)	By a subset of peaks common in 90% of pixels, Iterative cross-correlation			
(Källback et al., 2012)	SMQ1, SMST, SMA, SMM		Savitsky-Golay	TIC, RMS, Median, Labeled peak	Labeled peaks	1st derivative interpolated centroid		
(Mantini et al., 2007)	Kaiser filter	Kaiser filter	Kaiser filter					
(Fonville et al., 2012)				Reference molecule, TIC "matrix peaks", TIC "all data", TIC "informative peaks", PQN, Histogram matching, Median intensity of the informative peaks			Fixed mass binning to 0.2 Da bins prior to peak picking.	Matrix is removed using information extracted from out of tissue acquired regions.
(Mccombe et al., 2005)		Multivariate PCA and DA identification matrix peaks			Linear shift based on Cross correlations to a reference			PCA based
(Savitzky & Golay, 1964)			Savitzky-Golay					
(Jones et al., 2012a)				Isotopically labeled standard, TIC, Mean intensity				
(Deininger et al., 2011)				TIC, RMS, Median noise, SQRT-TIC, Log-TIC				
(Hamm et al., 2012)				TEC factor				
(Satten et al., 2004)	Custom method		Noise estimation threshold					
(Tracy et al., 2008)					Label-Free	Gaussian fitting	Binning after peak picking comparing all reported masses.	
(He et al., 2011)					Self-Calibrated Warping			
(Alexandrov et al., 2010)						OMP		

Baseline correction is generally performed prior to any other preprocessing step, because most preprocessing stages take advantage of baseline-corrected spectra; however, this could be generally avoided in metabolomics studies, because the baseline curve is very low (< 1000 Da). A visual inspection of the corrected spectrum to find the flattest resulting baseline is recommended for choosing the correct baseline-compensation procedure. Therefore, the selection of a specific method depends upon the characteristics of the acquired spectra (Norris, Cornett, & Mobley, 2007).

B. Noise Reduction

Noise is mixed into the spectra due to the random experimental variability associated with many factors,

including biological noise, matrix or surface inhomogeneity, electronic fluctuations, or ionization effects. Figure 2D shows that the application of a noise-reduction algorithm attenuates the small random variations in the spectra, thereby increasing spectra quality. The application of a noise-reduction step is always recommended, because noise can interfere in most of the subsequent data-analysis steps and, therefore, must generally be performed as soon as possible in the preprocessing chain. The most common noise-reduction technique is smoothing that removes the random variations in intensity from the spectra without significant alteration of the actual signal peaks. There are numerous smoothing algorithms, each of which has their own adjustable parameters. Common smoothing techniques include moving-average windowing and low-pass filtering.

As previously stated, these methods are also useful for baseline estimation, but are used for noise reduction with a different parameterization that does not completely smooth signal peaks. A more sophisticated smoothing method is the Savitzky–Golay polynomial approach, which preserves data shape (Savitzky & Golay, 1964). Another completely different noise-reduction technique is based on a hard threshold adjusted at a noise-level estimation (Satten et al., 2004). This method is useful when spectrum shape must be maintained for high-intensity peaks, with possible loss of low-intensity signals under the threshold.

A more robust alternative to smoothing is the application of de-noising methods using neighboring pixels instead of simply processing isolated spectra (as is done in common smoothing algorithms). These strategies may be especially useful to reduce noise in MS images reconstructed from spectra with low-intensity peaks. Based on the assumption that peak intensity should not largely change in a local domain, vector-valued median filtering and Markov random fields are also valid strategies for noise reduction (Hanselmann et al., 2009).

Spectra-smoothing methods are generally efficient to denoise the signal for most applications. Single-spectrum-smoothing algorithms require less computational cost than approaches using information from neighbor pixels. Consequently, simple smoothing techniques are often preferred, except for some specific applications as mentioned.

C. Spectral Alignment and Mass Calibration

Tissue-surface irregularities (or sample topography) in conjunction with spectrometer drift originate through small dilatations/contractions of the flight tube in the case of TOF detectors, with slight variations in high-voltage power sources producing small and random mass shifts in the spectra (Norris et al., 2007; Tracy et al., 2008). Pixel-to-pixel mass shifts can degrade the reconstructed image and reduce the performance of subsequent data-analysis methods; consequently, direct data analysis can easily lead to erroneous peak detection due to mass variations at all raster positions. To overcome these problems, a spectral-alignment algorithm is typically used in the preprocessing pipeline. The spectral alignment consists of equalizing the mass axis of each raster spot to obtain an internal coherency when peaks are compared pixel to pixel. Figure 2E illustrates spectra following an alignment stage. Alignment algorithms work by comparing the peak distribution of an unaligned spectrum, known as the test spectrum $t(x)$, with a reference spectrum, $r(x)$, that contains the correct m/z information. The algorithm is designed to find the warping function, $w(x)$, that minimizes the mass error of known peaks in $t(x)$ after applying the mass-axis transformation $t(x + w(x))$. The reference spectrum, $r(x)$, can be built using two main approaches: using actual m/z values from known compounds (calibration) or calculated from the MS image itself (label-free).

Calibration involves the use of known peaks homogeneously distributed over the tissue surface or using well-known endogenous molecules. To obtain an accurate mass calibration, various known peaks must be selected as references covering the m/z range of interest, because spectra misalignment often varies in a nonlinear way. For example, in low-weight-compound studies using matrix-free approaches, substrate background peaks can also be used to align the masses. This

strategy is accurate in the case of MSI using metal nanolayers, such as silver nanolayers, where metal peaks are distributed throughout the spectrum (Dufresne et al., 2013; Guinan et al., 2015).

Label- or reference-free alignment strategies are based on the use of cross-correlations between pixels to align biologically similar spectra. Label-free methods are based on algorithms designed to detect repeated peaks throughout the dataset and use this information to minimize the mass shifts. All spectra can be aligned to a reference spectrum calculated as the average of various spectra or the spectrum with the highest correlation coefficient of all the spectra in the dataset (He et al., 2011). It is also possible to align spectra without using any reference spectrum (Tracy et al., 2008). These strategies are useful in cases where it is difficult to correctly detect the calibration compounds at every raster position or where no reference compounds are used (Mccombie et al., 2005). Label-free alignment can also be performed prior to mass calibration, enabling the same calibration function to be applied to all pixels independently, regardless of whether the calibrated compounds are found in a given pixel. This alignment method is represented in Figure 3, where raw spectra (Fig. 3A) are aligned to the same mass axis, but the masses are still not calibrated to their references (Fig. 3B). In Figure 3C, a mass calibration method is applied to previously label-free-aligned spectra.

D. Normalization

Normalization is defined as the process of transforming the spectral intensity of every pixel to a common intensity scale (Norris et al., 2007). Normalization is a crucial step in overcoming pixel-to-pixel intensity variability due to substrate inhomogeneity and/or experimental drifts during acquisition. Figure 2F shows the changes in relative intensities of the different spectra when normalizing by the total ion count (TIC). Despite using an appropriate normalization approach, artifacts can still be introduced. Therefore, normalization might alter pixel relative-intensity distributions in an undesired way. The most common and simplest method is normalization by TIC, which assumes that an overall variation in the spectral intensity is associated with the matrix distribution throughout the sample. However, this assumption is not always true, because the concentrations of the tissue-detected compounds vary according to the biological composition of every pixel. Therefore, in tissues with clearly differentiated areas, such as brain samples, TIC normalization tends to equalize the intensities of the biological regions, which leads to inaccurate image reconstruction (Deininger et al., 2011; Fonville et al., 2012). A useful alternative could be to scale the intensities in accordance with the TIC computed using the selected peaks after peak picking or using a set of peaks relevant to the study (Fonville et al., 2012). In general, TIC normalization should be preferred for untargeted analysis due to its implementation simplicity and wide availability. However, in situations where tissue holes or “hot spots” are present (Jones et al., 2012a), TIC normalization can introduce side effects for further data analysis, subsequently requiring exploration of other normalization strategies. Another normalization method consists of replacing each peak intensity based on the signal-to-noise ratio (SNR) estimated around a window (Deininger et al., 2011). This strategy assumes that the analyte concentration is proportional to the SNR of the peak and not

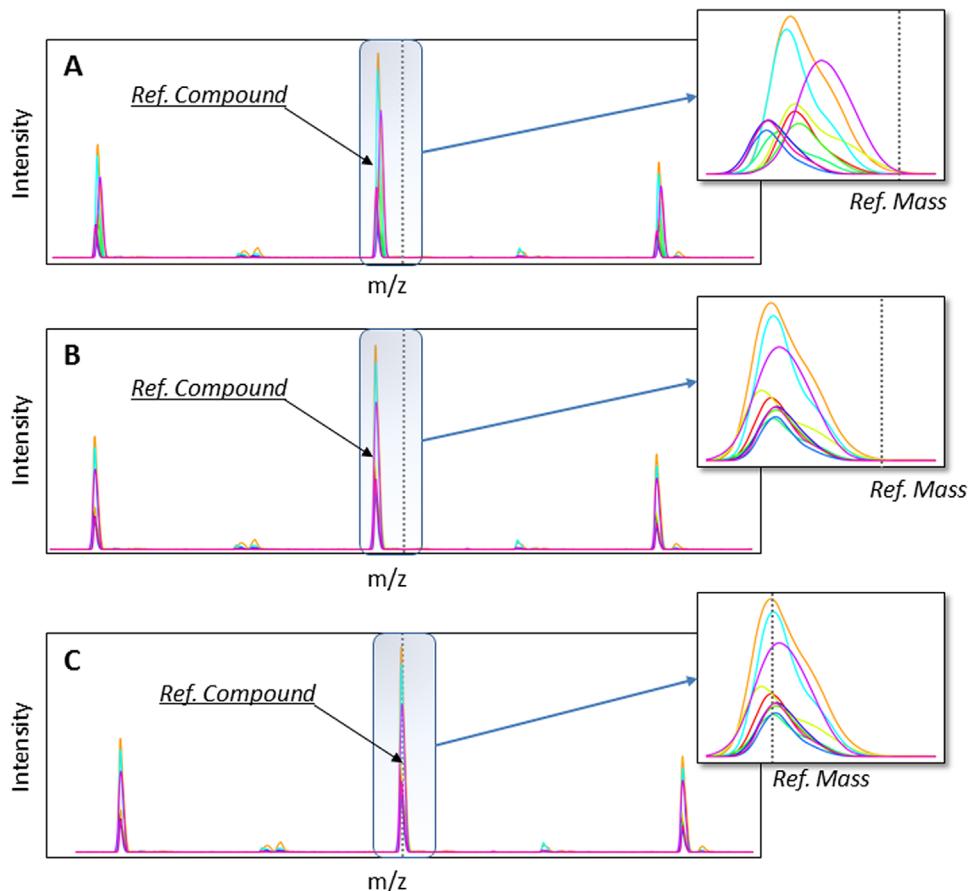


FIGURE 3. Example of two alignment approaches using simulated data. **(A)** Raw spectra without alignment. **(B)** Spectra aligned with a label-free technique. Here, all spectra share the same mass axis, but the peak masses are inaccurate. **(C)** Spectra aligned to internal standards (calibration). Reference-compound peaks and its theoretical mass are represented as dashed lines.

only standardizes the intensity axis, but also compensates for the baseline noise. Nevertheless, SNR does not take into account that the ionization efficiency is not homogeneous throughout the tissue slice.

Advanced normalization algorithms based on statistical data analysis can improve the results of untargeted data analysis. Normalization based on statistics aims to compensate for the effect of experimental variance and minimize the influence of biological information on the intensity scaling factor. Here, the rationale is that the spatial distribution of variations in intensity associated with experimental variance tends to be uncorrelated with biological-sample morphology. The most simple statistical-normalization factor is probably the median of the intensities of the peaks of interest (Fonville et al., 2012). This method calculates the normalization factor using the selected peaks as input to compute the median. More complex normalization approaches include histogram matching or probabilistic quotient normalization (PQN), where spectra are scaled by a coefficient associated with the distance of the median spectrum from each TIC spectrum. These methods are reportedly more robust due to their compensation for acquisition artifacts and presentation of better noise separation when multivariate methods are used (Fonville et al.,

2012). As a successful example of the application of statistics on normalization, Veselkov et al. (2014) introduced variance-stabilizing normalization (VSN), a logarithmic normalization method that decreases much of the variance in high-intensity peaks and allows for hyperspectral profiling of lipid signatures in colorectal cancer tissues.

Another side effect from normalization is the change in intensity of each ion, which might exert a strong impact on reconstructed images, because different normalization strategies can produce very different results and alter final image-intensity distribution. Figure 4 shows the effects of various normalization algorithms on image reconstruction of three different ions. This MS image was acquired using a sputtered-gold nanolayer over the tissue to promote ionization, but does not provide any MS signal in the absence of tissue. Figure 4A shows the images associated with the data without any processing (RAW data). Figure 4B shows the effect of TIC normalization calculated as the sum of all intensities of each RAW spectrum. In Figure 4C, maximum-intensity normalization is used. Here, the peak with the maximum-intensity value was used as the normalization factor, and the most intense peak was assumed to be representative of the rest of the spectrum intensities. Figure 4D introduces a TIC-based normalization approach designed

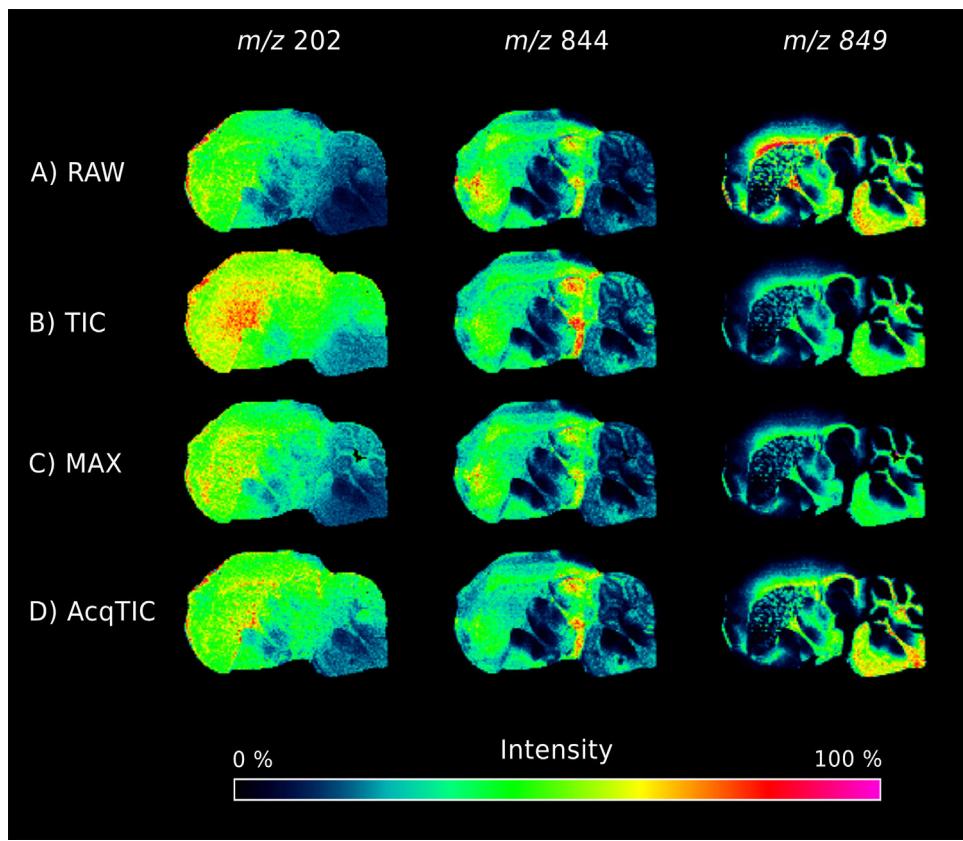


FIGURE 4. Comparison of the intensity maps for three ions (m/z 202, 844, and 849) from a sagittal mouse brain section using various normalization approaches. The MS image has been acquired using sputtered-gold nanoparticles to promote ionization and analyzed in a MALDI TOF/TOF UltraflexXtreme instrument from Bruker Daltonics in a mass range from m/z 80–1000 using a raster size of $80\text{ }\mu\text{m}$. (A) Raw data without performing any normalization. (B) TIC normalization was computed as the sum of all intensities in each spectrum. Pixels acquired outside of the tissue are removed from the normalized image. Such pixels are detected, because they have a very low TIC (<the mean TIC minus one standard deviation). (C) Maximum normalization was calculated by dividing each spectrum by the intensity of its maximum peak. Here, pixels acquired outside of the tissue are discarded using the same criterion as that used for TIC normalization. (D) AcqTIC is used to compensate for MALDI instrument ionization-source degradation during acquisition. AcqTIC was calculated as TIC smoothed by the TIC of neighboring pixels using a sliding window.

to compensate for ionization-source degradation during MS acquisition. Here, the produced images were very similar to the RAW version (Fig. 4A), but exhibited a flat overall intensity variation across the entire image. Each normalization strategy presented in Figure 4 produced different spatial distributions of the ions, thereby confirming the relevance of selecting the appropriate normalization strategy depending on experimental design.

When target-compound concentration requires quantification, normalization is performed relative to the peak intensity of a reference molecule deposited homogeneously over the sample. Usually an isotope-labeled compound is deposited on tissue, and normalization is performed by dividing each spectrum by the labeled-standard peak intensity (Jones et al., 2012b). Additionally, the tissue-extinction coefficient (TEC) factor was introduced as a quotient of the intensity of a standard deposited with the matrix either on or off of the tissue (Hamm et al., 2012). This coefficient evaluates the signals lost due to ionization effects for a given molecule and can be used as a normalization factor.

Accurate quantification results were reported using TEC normalization without using isotope-labeled compounds.

E. Peak Picking and Peak Selection

Peak picking allows for the detection of peaks in a mass spectrum and provides information about peak m/z , intensity, and quality. This process reduces a mass spectrum to a list of characteristics where only peak information is retained. Figure 2G illustrates the peak-picking process, which involves retaining only peaks positions instead of the entire spectrum. The simplest approach for a peak-detection algorithm consists of locating the zero crossings in the first derivative of the spectrum. However, this will result in significant mass errors due to factors that include limitations of spectrometer resolution and noise in the data. More accurate peak m/z values can be obtained using the peak shape to predict the actual m/z instead of using only the most intense MS peak. Various methods were proposed to accurately determine the peak shape. Källback et al. (2012) detected the approximate peak locations using the zero crossing

of the spectrum first derivative. A cubic interpolation was also applied around each peak area to determine the peak mass more accurately. Alexandrov et al. (2010) used a sequence of different algorithms. First, they modeled each mass spectrum as a sequence of Dirac delta peaks convolved with a Gaussian kernel, followed by using the orthogonal matching-pursuit (OMP) algorithm (Denis, Lorenz, & Trede, 2009) to de-convolve the peaks. They then applied the maximum-likelihood method consisting of fitting the spectrum contained in a sliding window to a Gaussian shape (Tracy et al., 2008).

The result of peak picking is an array-like data structure summarizing all of the relevant features of the entire MS image. In this data array, each peak can be considered as a variable and each pixel as an observation to perform further processing.

F. Binning

Binning describes the process of reducing the number of points in the spectrum by mapping neighbor m/z values into the same mass bin. Binning can be performed in two different ways. In one case, binning is performed prior to peak picking, and the mass bin size should be defined according the desired mass error. Using this method, the data size can be dramatically reduced in order to successfully execute demanding data-analysis algorithms. This binning approach is illustrated in Figure 2C, where the number of points on the spectra has been reduced. Fonville et al. (2012) used binning to test different normalization techniques under principal components analysis (PCA). The drawback of this method is that some close peaks derived from different compounds can be merged together, resulting in degradation of the results from further data analysis.

A second type of binning is performed after peak picking. Here, the m/z of each peak is slightly adjusted to report exactly the same m/z for each detected compound in all pixels (Fig. 2H). To achieve this, each peak mass is compared with its neighbors through all pixels in a defined tolerance. The most representative mass is then used for all of the peaks expected to derive from the same compound. This technique was successfully used to enhance mass resolution in MALDI experiments (Tracy et al., 2008). However, the main drawback of this binning technique is that spectra must be well aligned in order to enable selection of a tight bin tolerance, resulting in the requirement for complex implementation processes.

G. Matrix-Peak Removal

In MALDI-MSI or LDI-MS spectra, the organic matrix, metal ions, or surface compounds ionize with the molecules of the biological sample. Here, we discuss the different strategies used to eliminate these matrix peaks. Such non-informative peaks do not appear in all MSI applications, because these matrix signals are commonly more intense in the low m/z range, and, therefore, this step is not always required. In MSI applications where unwanted signals are strong, matrix-peak removal may be beneficial, especially in the case of untargeted data analysis and metabolomics studies where matrix signals have the most impact on low-mass ranges.

Determining which peaks correspond to a matrix or have a biological origin can be challenging. Some methods were developed for robust and automatic selection. Fonville et al. (2012) described two approaches for obtaining background-related peaks. First, the signal acquired outside of the tissue (containing only matrix peaks) is correlated using the signal acquired on the tissue (containing the matrix plus biological peaks). These correlation factors are then used to retain uncorrelated variables that are defined as biologically relevant. Second, the variance explained (VE) is used to determine which peaks constitute background signals, because matrix-related signals should be homogeneously distributed over the entire surface, leading to lower VE values. In another study, (McCombie et al., 2005) identified background peaks using multivariate analysis tools to manually draw regions of interest (ROIs) inside and outside of tissue regions. In this respect, algorithms, such as PCA, can determine which masses are associated with the tissue. Once matrix-related peaks have been determined, they can be removed by deleting their corresponding variables in the peak list. To illustrate this, Figure 2I shows removal of one of the spectral peaks due to its origination for the matrix.

IV. MULTIVARIATE ANALYSIS OF IMAGES

The most common and direct strategy in MSI consists of spatial visualization of one ion or a small group of ions, each of which is assigned to a color code. This strategy is especially useful for targeted analysis, with most commercial and open-source programs including many functions for plotting images of the ions of interest. However, this simple visualization strategy does not exploit the full potential of MSI, such as biomarker discovery and identification, image clustering (or segmentation), histology driven image reconstruction, tissue classification, and 3D-image reconstruction. To achieve all these objectives, multivariate methods that consider the full MS spectrum of each pixel as an intrinsic multivariate problem are introduced here.

We have divided the discussion here into three sections. The first section corresponds to the multivariate analysis of images and it is also divided in to three different approaches: supervised, using histological or microscopy images as a reference; unsupervised, which does not require previous information about the samples; and unsupervised strategies with further expert evaluation that combine the information given by the histological images with unsupervised algorithms. The second section focuses on 3D-image-reconstruction strategies, and the final section describes the different uses of PCA, which is the most used multivariate algorithm in MSI.

The up-to-date bibliography discussed here is also reviewed in Table 2. Notably, few papers attempt to identify the key ions involved in cluster differentiation. Although this is an essential task in biomarker discovery in proteomics, lipidomics, and metabolomics applications, the problems associated with the generation of adduct ions, possible fragmentation of the molecular ions, poor mass resolution of the TOF detectors (the most commonly used), and low sensitivity of the MS/MS working mode makes this task difficult. To overcome this, a common strategy is to identify the metabolites detected in the MSI experiment by performing high-performance liquid chromatography MS analysis using the same tissue sample (Lee et al., 2012).

TABLE 2. Overview of the literature about multivariate analysis applied to MSI.

Author (Ref)	Study	Kind of samples	Multivariate algorithm used	Main Results
<i>Supervised algorithms</i>				
(McCombie et al., 2005)	Spatial and Spectral Correlations in MALDI Mass Spectrometry Images by Clustering and Multivariate Analysis	Brain from mouse with Alzheimer	PCA, hierarchical clustering (HC), k-means and ISODATA	Differentiation of tissue regions according to pixel spectral similarities
(Schwamborn et al., 2007)	Identifying prostate carcinoma by MALDI-Imaging	Human prostatic samples with or without cancer	SVM and GA	Differentiation of prostatic tissues with and without cancer
(McDonnell et al., 2008)	Mass Spectrometry Image Correlation: Quantifying Colocalization	Rat brain	Pearson coefficient correlations	Determination of the similarity of the distributions of specific molecules within and between MS Images
(Hanselmann et al., 2009)	Toward Digital Staining using Imaging Mass Spectrometry and Random Forests research articles	Animal models of human breast cancer	Random Forest classifier	Differentiation of different regions in a tumor sample and comparison between different tumor samples
(Rauser et al., 2010)	Classification of HER2 Receptor Status in Breast Cancer Tissues by MALDI Imaging Mass Spectrometry	Human breast cancer tissues	Supervised clustering of tissue regions using ANN i SVM	Definition of HER+ and HER- tissues
(Veselkov et al., 2014)	Chemo-informatics strategy for imaging mass spectrometry-based hyperspectral profiling of lipid signatures in colorectal cancer	Human colorectal cancer tissues	Images co-registration, PLS-DA for tissue region differentiation	In house H&E and MSLDI Determination of region-specific lipid signatures in colorectal cancer tissues
(Van de Plas et al., 2015)	Image fusion of mass spectrometry and microscopy: a multimodality paradigm for molecular tissue mapping	Mouse brain	Fusion image by multivariate regression	'fusing' two distinct technologies: mass spectrometry and microscopy images
<i>Unsupervised algorithms</i>				
(Hanselmann et al., 2008)	Concise representation of Mass Spectrometry Images by Probabilistic Latent Semantic Analysis	Highly and weakly metastatic tumor	pLSA, PCA, ICA and non-negative PARAFAC	Image reconstruction by pLSA

(Van de Plas et al., 2008)	Discrete wavelet transform-based multivariate exploration of tissue via imaging mass spectrometry	Mouse brain in the context of a resource-hungry study	DWT and PCA	Analysis of a sagittal section of mouse brain
(Bonnel et al., 2011)	Multivariate analyses for biomarkers hunting and validation through on-tissue bottom-up or in-source decay in MALDI-MSI: Application to prostate cancer	Prostate cancer samples	PCA-SDA and HC	identification and localization of specific markers in histological samples
(Cho et al., 2012)	Combining MALDI-TOF and molecular imaging with principal component analysis for biomarker discovery and clinical diagnosis of cancer	Rat renal samples	PCA	differentiation of the different biologically regions in a tissue comparing lipid, peptide and protein profiles
(Lee et al., 2012)	Resolving brain regions using nanostructure initiator mass spectrometry imaging of phospholipids	Mouse brain	NMFA	Resolving of neuronal and glial reach brain regions
(Franceschi & Wehrens, 2014)	Self-organizing maps : A versatile tool for the automatic analysis of untargeted imaging datasets	Apple slices	SOM	spatial distribution of ions associated with the regions
<i>Unsupervised strategies with further expert evaluation</i>				
(Deininger et al., 2008)	MALDI imaging combined with hierarchical clustering as a new tool for the interpretation of complex human cancers	Gastric cancer and non-neoplastic mucosa tissues	HA coupled to PCA	Determination of regions within and between tissues
(Alexandrov et al., 2010)	Spatial segmentation of imaging mass spectrometry data with edge-preserving image denoising and clustering	Rat brain and neuroendocrine tumor samples	image clustering process, using (HDDC) method	New procedure for spatial segmentation of MALDI-imaging data sets, that clusters all spectra into different groups based on their similarity.
(Alexandrov & Kobarg, 2011)	Efficient spatial segmentation of large imaging mass spectrometry datasets with spatially aware clustering	Rat brain and neuroendocrine tumor samples	Novel strategies for spatial segmentation that incorporates spatial relations between pixels into cluster	Tumor heterogeneity exploration

			regions.	
(Bruand et al., 2011)	AMASS: algorithm for MSI analysis by semi-supervised segmentation	Rat brain	AMASS method: Automatic segmentation of maps according to patterns of co-expression of individual molecules	Discovery of novel molecular signatures
(Jones et al., 2011)	Multiple Statistical Analysis Techniques Corroborate Intratumor Heterogeneity in Imaging Mass Spectrometry Datasets of Myxofibrosarcoma	Myxofibrosarcoma	PCA, ICA, non-negative matrix factorization, pLSA, k-means clustering and hierarchical clustering).	Automatically localization of clusters in datasets
(Alexandrov et al., 2013)	Analysis and interpretation of imaging mass spectrometry data by clustering mass-to-charge images according to their spatial similarity	Rat brain coronal section	Clustering method according to the spectral similarity of the pixels	Segmented rat brain with 10 regions and the spectrum with the ions associated to every segment
(Fonville et al., 2013)	Hyperspectral Visualization of Mass Spectrometry Imaging Data	Rat brain	The visualization strategy was applied to results of PCA, SOM and t-distributed stochastic neighbor embedding.	The image color-coding is based on spectral characteristics, such that pixels with similar molecular profiles are displayed with similar colors

A. MS Image Multivariate Processing

1. Supervised Strategies

Many studies have described MSI techniques involving hematoxylin and eosin (H&E) staining or immunohistochemistry images (Groseclose et al., 2008). These strategies are used for tissue recognition and classification and for biomarker discovery. Many different algorithms, including random forest, support vector machines (SVM) (Wang, 2005), PCA-discriminant analysis (Hair et al., 2013), recursive maximum-margin criterion (RMMC), or artificial neural networks (ANN) (Samarasinghe, 2006), have been used to compare MSI and histological images.

McCombie et al. (2005) used compression algorithms (PCA, hierarchical clustering, *k*-means, and iterative self-organizing data analysis technique) in combination with a DA algorithm to maximize the spectral differences between two ROIs in a brain section from an Alzheimer's disease rat model. Results showed that the multivariate methods were capable of extracting complex information from a tissue section and that it was much easier to identify contrasting regions in an image taken from a complete rat head. Genetic algorithms and SVMs

were used by (Schwamborn et al., 2007) to differentiate prostatic tissues with and without cancer. Additionally, an SVM was able to identify four distinctively overexpressed peaks, with overall cross-validation, sensitivity, and specificity >85%.

McDonnell et al. (2008) presented a new methodology for analyzing MSI datasets. The (Pearson) correlation coefficient was calculated between images acquired in an experiment with rat brain tissues to determine the correlation between ions. As a result, an interesting correlation-map matrix was obtained that described distribution similarities between 28 biomolecular ions. One important problem encountered in this study was that the method was highly sensitive to background noise.

The output of random forest algorithms was used as a class-probability estimate for classifying human breast cancer in mice models (Hanselmann et al., 2009). Using this approach, various regions (separate necrotic tissue, viable tumor, gelatin, tumor interface, and glass/hole) were differentiated within and between samples with high sensitivity rates (~90%) and positive predictive values (~85%).

ANN and SVM algorithms were used to differentiate HR2+ and HR2- regions in breast cancer tissues (Rauser et al., 2010). The area under the curve calculated by receiver operating

characteristic analysis exhibited high sensitivity (83%) and specificity (92%), and an overall accuracy of 89%. Furthermore, they discovered specific changes in protein/peptide expression (ion m/z 8404, identified as cysteine-rich intestinal protein 1) that were strongly correlated with HER2 overexpression.

Recently, Veselkov et al. (2014) reported interesting advances in MSI, including automated algorithms for co-registration of histology and molecular images to aid correlation of histological and biochemical features. In the same study, partial least squares-discriminant analysis (PLS-DA) was used to extract tissue-specific molecular patterns and maximize the variance between regions and minimize variance within regions. This enabled characterization of lipid signatures in tissue regions surrounding colorectal cancer tissue. Figure 5 shows the regions selected from H&E-stained high-resolution optical images used to guide a segmentation process in a desorption electrospray ionization (DESI)-MSI image. This strategy is also widely used to compare case/control samples. For example, it was used to compare brain samples from an Alzheimer's disease mouse model from those of controls (Hong et al., 2016). In this study, ROIs from equivalent histological regions in both samples were selected, and PCA combining the pixels of the ROIs was used to identify the metabolites differentially expressed between both samples as a consequence of disease progression.

Recently, Caprioli and collaborators (Van de Plas et al., 2015) established a new paradigm consisting of the fusion of

histological and MS images. They calculated a correlation function, q , between microscopy and MS-image patterns in an attempt to create new images that combined the high-spectral resolution of microscopy with the high molecular-specificity of MSI.

2. Unsupervised Strategies

Unsupervised strategies were introduced in MS-image processing to disclose new molecular and morphological information independently from classical histology. These unsupervised strategies do not require prior information for clustering and are capable of revealing several molecular fingerprints, making it ideal for analyzing heterogeneous tissues and discovering biomarkers. Clinical research requires independent methods for tissue evaluation beyond classical histology. In this sense, the number of clinical studies attempting to correlate MS images with biological and clinical variables increases annually, highlighting the need for clinically validated "molecular histology." Clustering techniques, such as PCA, self-organizing maps (SOMs) (Kohonen et al., 2000), probabilistic latent semantic analysis (pLSA), and k -means, can be used in this respect.

The segmentation of an image is the only technique that allows for visualization of regions with similar molecular compositions that is essential for image comparison and tissue characterization and recognition. One intrinsic problem in

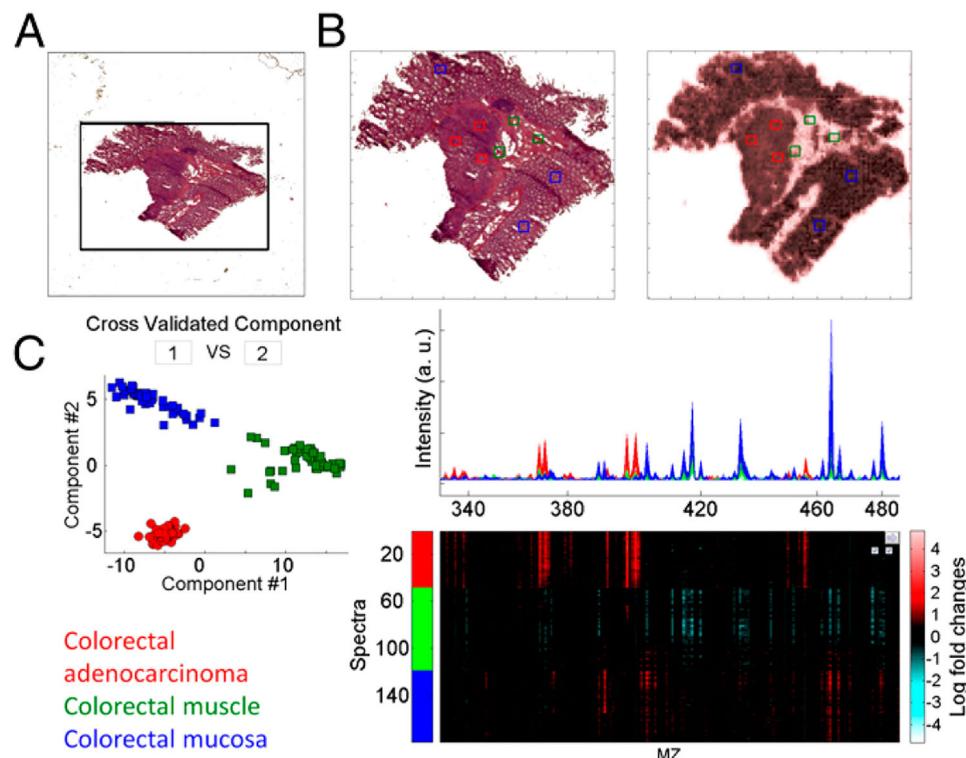


FIGURE 5. Image co-registration, feature co-selection, and multivariate analysis. (A) Automatic image transformation for accurate co-registration of biochemical and histological features. (B) High-resolution optical image of an H&E tissue section with regions of tumor (red boxes), muscle (green boxes), and healthy mucosa (blue boxes) selected. Shown is aligned DESI-MSI image with automated co-selection of pixels corresponding to defined regions of interest. (C) Discriminatory analysis using the RMMC method with leave region-out cross-validation for enhanced separation of tissue classes based on biochemistry (taken from [Veselkov et al., 2014] and reprinted with permission of the National Academy of Sciences).

unsupervised clustering comprises the difficulty in the determination of the optimum number of clusters, the setting of parameters values for pixel clustering, and validation of the results (Jones et al., 2012a).

Cho et al. (2012) used PCA to compare the lipid, peptide, and protein profiles of various biological matrices, including MS images of tissue sections. They used PCA loadings to select the ions differentiating biological regions in a tissue sample. However, an important drawback of PCA is the negative and positive distributions of the scores, making it difficult to interpret the results when applied to MSI. To overcome this limitation, classical techniques, such as the discrete wavelet transform (DWT) (Figueiredo & Nowak, 2003) algorithm for data compression and de-noising, was used to solve information-technology problems based on its generating reduced sets of wavelet coefficients. DWT was used for the MSI analysis of sagittal sections of mouse brain (Van de Plas, De Moor, & Waelkens, 2008). In this study, the results of DWT application were proven to be more compact than those obtained using PCA. Another advantage of using DWT is that it also retains mass-spectral information by means of the inverse DWT.

Another useful approach is pLSA, a statistical technique that allows for a low-dimensional representation of the observed variables in terms of their affinity to certain hidden variables. pLSA has been used to analyze MALDI-TOF images (Hanselmann et al., 2008) and provide better physical interpretations relative to those provided by PCA, independent component analysis (ICA), and non-negative PARAFAC (Bro, 1997), because the decomposed components can be directly interpreted as peak-intensity lists.

A SOM is a type of ANN that is trained using unsupervised learning to produce a low-dimensional map (typically 2D) as a discretized representation of the input space of the training samples. SOMs use a neighborhood function to preserve the topological properties of the input space and are very attractive in MSI analysis. Franceschi & Wehrens, 2014 used SOMs to illustrate the spatial distribution of ions associated with the regions generated for a dataset of apple slices, retaining the key ions for further analysis and metabolite identification.

PCA-symbolic discriminant analysis based on hierarchical analysis was used by (Bonnel et al., 2011) in a study of prostate cancer and was suitable for identifying and localizing specific markers in human prostatic tissues.

Non-negative matrix factorization analysis (NMFA) was used to resolve glial and neuronal cell-enriched brain regions (Lee et al., 2012). Based on potassium adducts from a set of 18 selected lipids, NMFA provided six components representing spectral patterns associated with brain morphology. A method for hyperspectral visualization was recently proposed (Fonville et al., 2013), consisting of a RGB color-coding based on the spectral characteristics of every pixel. The application of this strategy to various data-reduction models (PCA, SOM, and *t*-distributed stochastic neighbor embedding [a neural-network-based manifold-learning technique]) revealed its capability for unsupervised creation of images exhibiting good correspondence between molecular and anatomical information.

3. Unsupervised Strategies With Further Expert Evaluation

Unsupervised strategies with further expert evaluation strategies assess the results of unsupervised clustering by comparing them with histological images, even though these images do not take part in the clustering process. (Deininger et al., 2008) used hierarchical analysis coupled with PCA to identify several gastric cancer and non-neoplastic mucosa tissues. Using this semi-supervised approach, classifications were based on pathological information about healthy and cancerous regions, thus opening avenues for the discovery of new cancer biomarkers.

Jones et al. (2011) used various statistical methods (PCA, ICA, NMFA, pLSA, *k*-means clustering, and hierarchical clustering) to automatically determine clusters in datasets of intermediate-grade myxofibrosarcoma. Results showed that the MS images generated by the different methods exhibited similar distributions, confirming the ability to discover different nodules in identical histology tumor sections and suggesting its usefulness as a “molecular histology” technology.

In the field of multivariate approaches to clustering of MS images, it is worth mentioning the work of T. Alexandrov's research group. They used high-dimensional discriminant clustering to analyze and interpret a larynx carcinoma section and compared the automatic spatial-segmentation image obtained by MALDI-TOF with H&E-stained microscopic images (Alexandrov et al., 2010). The molecular image enabled exploration of tumor heterogeneity and pharmaceutical metabolism. The same research group also proposed novel strategies for spatial segmentation that incorporated spatial relationships between pixels into cluster regions, enabling pixels to be clustered together with their neighbors (Alexandrov and Kobarg, 2011). Additionally, they evaluated the segmentation method in a rat brain section and a neuroendocrine tumor section and identified various tumor regions by discovering the anatomical structure and identifying functionally similar regions. In 2011, they created an algorithm to increase the spatial resolution of the segmentation map by resizing the map by splitting the pixels (Alexandrov et al., 2011).

The algorithm for MSI analysis by semi-supervised segmentation (AMASS) method was created to match pathological and segmented MS images (Bruand et al., 2011) in order to determine correspondences between the two images in a semi-supervised way. The AMASS method has helped distinguish between anatomical regions in slices of rat brain and enabled the discovery of peptide masses that are differentially expressed between segmented regions. Recently, the same group published a new segmentation method where *m/z* images are clustered on the basis of spectral similarity in the pixels (Alexandrov et al., 2013), enabling pixels exhibiting common ion patterns in the spectra to be clustered together. Figure 6 shows a rat brain segmented in 10 regions (Fig. 6C), a spectrum with the ions associated with every segment (Fig. 6B), and the spatial pattern of every segment (Fig. 6A).

B. 3D-Image Reconstruction

3D-image reconstruction is performed using combinations of images obtained from consecutive tissue slices. One of the main analytical challenges of 3D-image reconstruction is that the extended period required to acquire all tissue sections needed

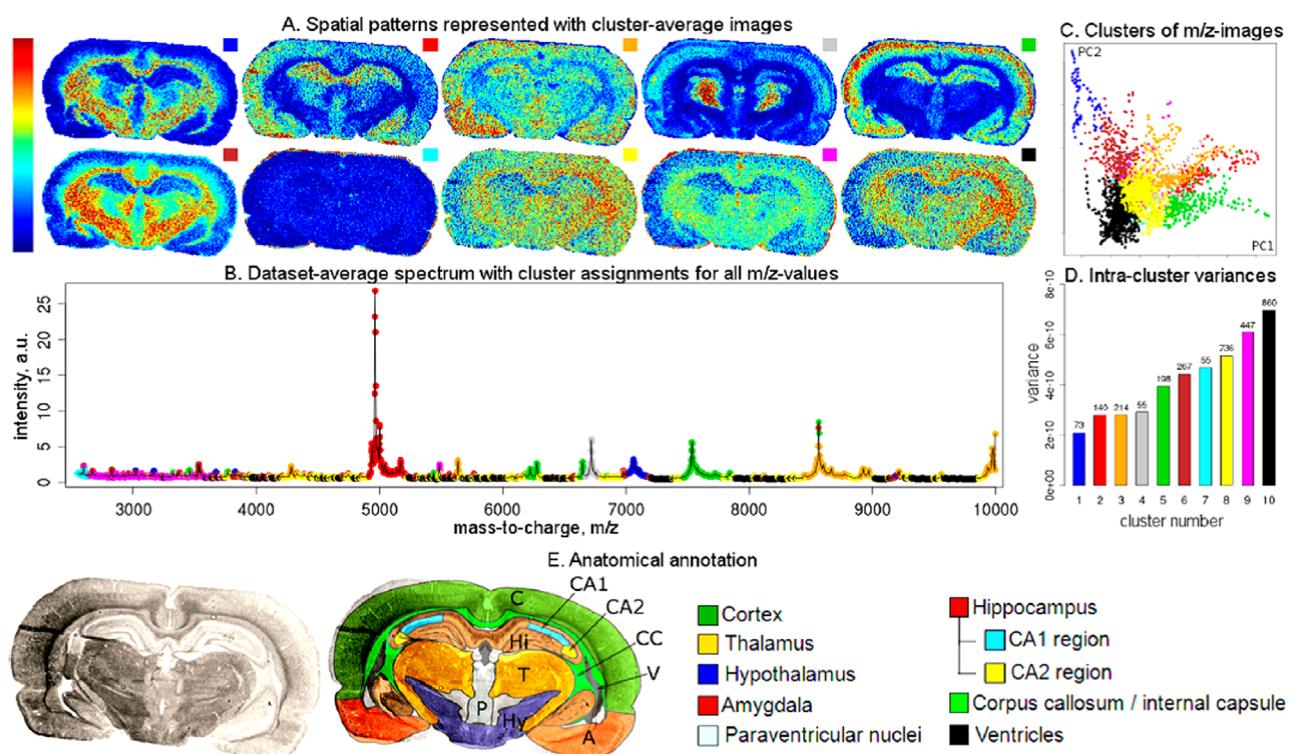


FIGURE 6. Results of the analysis of a MALDI imaging mass spectrometry dataset of a rat brain coronal section, following the proposed approach based on clustering m/z images into 10 clusters according to their spatial similarity. (A) Cluster-averaged images represent detected spatial patterns. (B) Dataset-averaged spectrum with assignments of m/z values to the clusters. (C) Visualization of m/z images in the space of their two first principal components; one dot represents an m/z image, and dots are colored according to their cluster assignments. (D) Intracluster variances, where the numbers on the top of the bars represent cluster sizes. (E) Optical image of the section with anatomical annotation provided. Plots A and B show the variety of the spatial patterns among m/z images and help understand how each m/z image looks. Plots C and D help evaluate clustering results [taken from (Alexandrov et al., 2013) and reprinted with permission from ACS Publications].

for constructing a full 3D image makes tissue degradation a critical issue. The challenges, approaches, and future research directions associated with 3D images obtained by serial sectioning and MALDI-MS have been extensively reviewed (Palmer & Alexandrov, 2015).

In 3D-image segmentation and reconstruction, the lack of efficient computational algorithms for data reduction, processing, and visualization of large 3D datasets constitutes a bottleneck. Xiong et al. (2012) developed many algorithms for 3D MSI, including data reduction, 2D data alignment, 3D visualization, and statistical analysis for clustering. The morphological features of brain-tissue sections were revealed using a self-organizing feature map ANN on MS images obtained by DESI-MS. Of particular interest was the ability of this method to directly compare 3D images acquired by MALDI-MS and magnetic resonance imaging (MRI), making it possible to match information from morphological and molecular datasets.

A new data-processing pipeline for analyzing and interpreting 3D MALDI-MS images was proposed by Trede et al. (2012), which was based on the edge-preserving, de-noising methods developed for 2D-image segmentation, that implements a hierarchical-clustering method called bisecting k -means. The reconstructed 3D images consisted of 33 serial sections of mouse kidney at 3.5 μm thickness acquired at a resolution of 50 μm . More than half a million spectra were acquired,

representing >50 GB of data. The computational pipeline showed the anatomical structure of the kidney following correct alignment of the 2D sections, as well as molecular-mass co-localization at major anatomical regions. The same group used the PAXgene tissue container (Oetjen et al., 2013) and paraffin embedding to preserve tissues, with results similar to those obtained on frozen samples. The same publication compared the MRI images of a mouse kidney with 3D MS images, enabling reconstruction of the anatomical structure.

C. On the Uses of PCA in MSI

PCA is likely the most often used algorithm in multivariate analysis and MSI applications. There are four primary uses of PCA in MSI analysis: exploratory analysis, data compression, clustering-performance assessment in unsupervised strategies, and biomarker identification. These four uses are illustrated in real examples in Figure 7.

A Exploratory analysis. Exploratory analysis of MS images constitutes the most frequent use of PCA. PCA can be used for the assessment of ionization-source drift as shown in Figure 7A. Another application of PCA could be the detection of outlier pixels denoting possible hot-spots or holes in a tissue section.

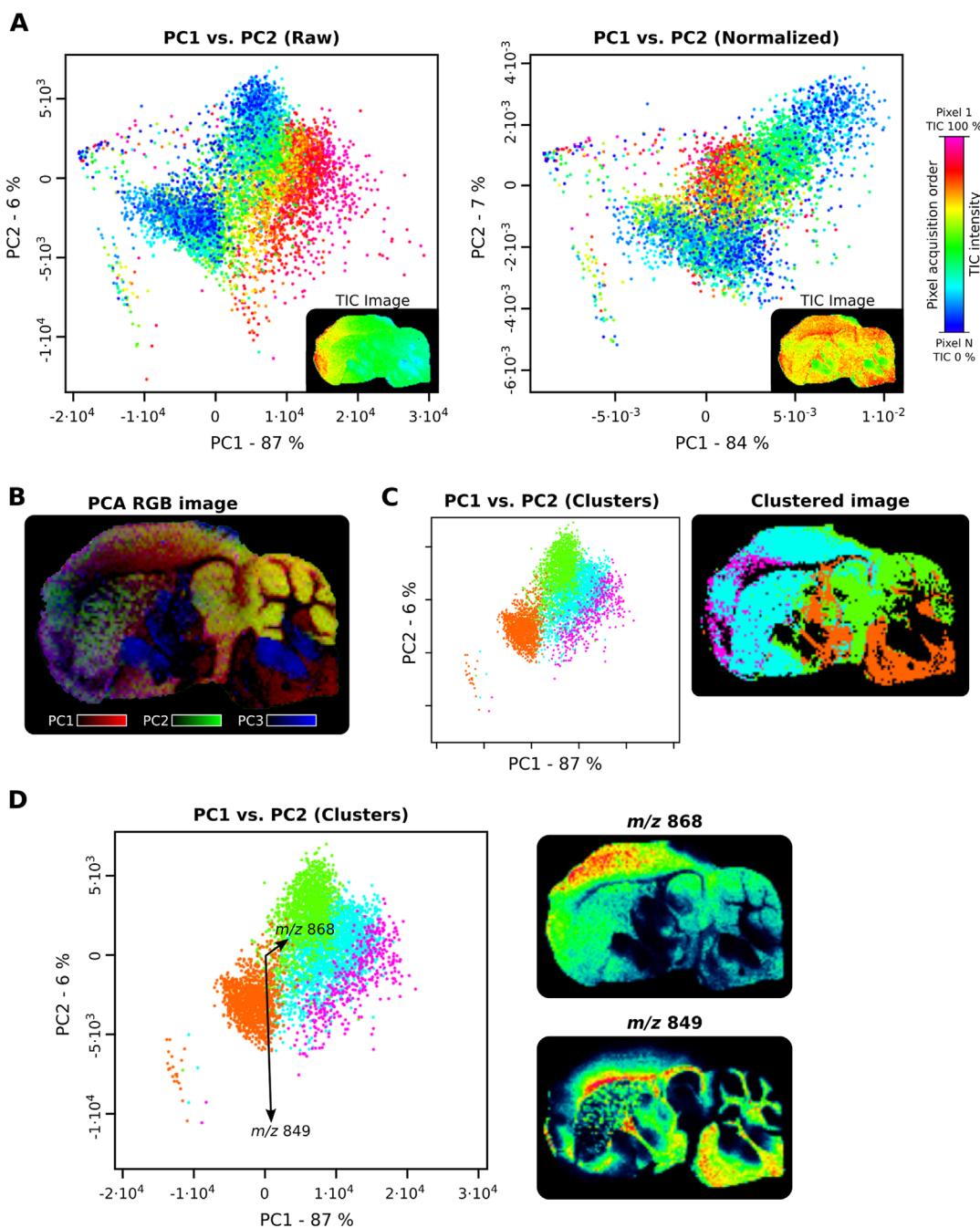


FIGURE 7. Several uses of PCA in MSI experiments. PCA is computed using data from the same experiment shown in Figure 4. PCA was calculated from a peak list generated using alignment, calibration, peak picking, and peak binning as preprocessing steps. (A) PCA as a tool for exploratory MSI analysis. Comparison of PCA performed before (left) and after (right) spectra normalization. We used AcqTIC (described in Fig. 4D) to compensate for ionization-source degradation. In both cases, PC1 versus PC2 is plotted, coloring each data point according to its order during acquisition. In PCA with no normalization, PC1 is affected by intensity degradation during acquisition. In PCA after normalization, data points do not follow the acquisition pattern compensating for intensity degradation effects. (B) PCA as a compression tool. An RGB image is built using PC1, PC2, and PC3 to encode red, green, and blue colors, respectively. This RGB image shows how PCA is able to compress almost all of the information using only three components, facilitating tissue-region localization. (C) PCA used for image-segmentation evaluation. In the clustered image (upper), the four larger clusters determined by an in-house image-segmentation method are shown. Pixels are colored according their cluster. The PCA plot shows an excellent separation between the pixels of different clusters, confirming the good performance of the image-segmentation technique. (D) PCA for biomarker-compound discovery. Two loadings contributing to group separation are selected for illustration (ions m/z 868 and m/z 849) of the PCA plot. The corresponding ion-intensity maps reveal a high degree of complementarity between the intensity of both ions, indicating its influence in the identification of different morphological areas in brain tissue.

B Data compression. The high dimensionality of the pixel spectra results in files with large dimensionality in MSI experiments. A direct method for data compression consists of transformation of the original variables into the principal components of the PCA. In general, most of the variance can be retained in five principal components. Figure 7B depicts a RGB brain MSI image considering the three principal components. The main problem associated with this data-compression approach is that the information concerning individual MS ions is lost.

C PCA for clustering-performance assessment in unsupervised strategies. The validation of unsupervised clustering results is difficult due to the lack of alternative methods for their comparison. The representation of the pixels in a PCA, labeled according to the cluster to which they belong, offers an estimation of clustering-algorithm efficacy. Generally, the higher the pixel separation between clusters, the better the performance of the clustering process. Figure 7C shows an example of PCA used to validate an in-house-developed clustering method.

D Biomarker identification. In targeted strategies that use references of histological (or microscopy) images, it is common to compare pixels between different regions of interest (i.e., healthy and tumoros regions). The analysis of the PCA loading is a powerful technique enabling identification of the most influential ions in the pixel separation (Fig. 7D).

V. DATA HANDLING STRATEGIES AND CONSIDERATIONS

Because MSI data consists of a large collection of mass spectra corresponding to the spatial location of each pixel of the tissue acquired, the amount of data produced in this kind of experiment tends to be very large, and, therefore, the computational strategies required to handle data processing are complex. In this section, we have divided these strategies into three sections: data formatting, processing requirements, and data-reduction strategies, including peak-picking, feature-selection, and data-compression strategies.

A. Data Formatting

As previously mentioned, for a single imaging experiment, a MALDI-MS instrument generates a large amount of raw data. In most cases, the format used to store the acquired data is determined by the instrument manufacturer and is only compatible with a few supported software tools generally provided by the same manufacturer. Such proprietary formats usually force the end user to adopt the data-processing workflow defined by the software producer, limiting flexibility in the data-analysis process. Fortunately, the main MSI open-data format imzML (Schramm et al., 2012), has started to become a standard used throughout all software platforms, with almost every manufacturer currently offering some level of compatibility.

Public-domain data formats, such as plain-text files (ASCII), mzML (Martens et al., 2011), or Analyze7.5 (included in the BioMap software), can be used to exchange MSI data. However, these formats have not been developed specifically for MSI and, therefore, present some limitations. In the case of plain

text files (ASCII), the MS image is converted into a collection of text files where each file contains a spectrum associated with a raster spot. This strategy is very straightforward, but requires much more hard-drive space as compared to that required for binary data formats, and spectra take longer to parse. Additionally, the portability of MSI experiments to ASCII files is not efficient, because this format does not support storage of metadata (i.e., the type of experiment) and other significant information, such as pixel coordinates associated with each spectrum. The mzML (Martens et al., 2011) format improves the situation by adding binary data formatting and enabling the storage of standardized metadata. However, mzML was not created for MSI, and some important information, such as raster positions, are not supported. Finally, the Analyze7.5 format was initially designed to store medical 3D images, but is also an export option in some MSI-software packages. Analyze7.5 stores 2D information with the raster positions, and 3D fields are populated with mass spectra. Analyze7.5 is optimal for storing MSI data in a compact way, but it presents limitations in metadata storage, because it was not designed for MS applications. Moreover, mass spectra intensities are often encoded in 16-bit integers when exporting to Analyze7.5 in some software tools, reducing the accuracy of the original intensity axis encoded in 32-bit integers by the instrument detector.

To overcome these limitations, an open standard has been developed under the name imzML (Schramm et al., 2012), which aims to become the global MSI reference-exchange format. The openness refers to the data-formatting specification used, which is fully detailed and available in open-access format that enable everyone to access and implement imzML support. Recently, imzML began being incorporated as a data-export option in many proprietary software packages, with some developers creating third-party tools to facilitate data conversion to imzML. An example of this effort to promote imzML is the “imzMLConverter” tool, which converts files from mzML format to imzML (Race, Styles, & Bunch, 2012).

B. Processing Requirements

Data acquired from an imaging experiment represents a collection of spectra, the size of which depends upon the scanned area, spatial resolution, and mass-spectra range and resolution. Each spectrum consists of a vector of intensities generally encoded in 32-bit integer numbers, and the number of data points in each spectrum depends upon the resolution of the spectrometer. Furthermore, the raw data size is proportional to the number of pixels in the image multiplied by the number of points in each spectrum. As an example of TOF-acquired data, an MS image at 100 pixels × 100 pixels with 50,000 points per spectrum consumes ~1.86 GB (100 × 100 pixels × 50,000 points × four bytes) when it is fully loaded in computer memory. Because pixels are arranged in a bi-dimensional space, an increase in image size is expressed as a scan-area expansion leading to a dramatic amount of memory usage. For instance, doubling the image size in X and Y dimensions as in the previous example produces ~7.45 GB of data. This memory requirement indicates the physical limitations of computer memory, making it difficult to efficiently handle such volumes of data. In addition to memory requirements, processing time and CPU use must also be taken into consideration. The large amounts of data produced by MSI experiments require heavy processing resources,

especially when complex multivariate statistical algorithms are used.

Most mathematical and statistical software packages use a load-and-process approach where all data is first loaded into random access memory (RAM) and then processed with the desired algorithms. Although this approach makes it easy to handle the data, it requires large quantities of RAM. Moreover, most of these packages are based on interpreted languages that work in a mono-task approach. Due to these processing requirements, an interesting approach may be to adopt parallel-processing approaches to benefit from modern multicore-processor systems. However, many statistical libraries have not been designed to support multithreading; therefore, parallelization is often not an out-of-the-box solution. A possible strategy to overcome this could be splitting the data into fragments and processing each one in a different instance of the mono-task program. This approach was demonstrated as being effective when a processing platform with many processors is available (Smith et al., 2015). Nevertheless, some packages are available for parallel processing, including those of R and MATLAB. These packages make the multithreaded implementation of algorithms more straightforward. Processing infrastructures that use graphics processing units have been tested with multithreaded algorithms, resulting in reduced computation times (Jones et al., 2012b). However, this reduction occurs at the expense of flexibility and simplicity.

C. Data-Reduction Strategies

Due to the high computational resources required for processing large MS-image datasets, data-reduction techniques play an important role. The goal of data reduction is to extract relevant information from the dataset, while minimizing both the memory footprint and information loss. A common approach involves peak picking (commented in section III.E.), which stores only spectral peaks in a reduced memory space. Following peak picking, a feature-selection routine selects the most informative peaks, which helps to reduce the data requiring further processing. However, peak picking is the final step in the preprocessing chain; therefore, preprocessing actions performed prior to this one will not benefit from this first-stage data reduction. Another strategy involves peak binning after peak picking (Norris et al., 2007). Once the mass resolution of every peak is determined, all peaks under the mass tolerance are grouped into the same bin. This represents an important reduction in the number of variables.

Each spectrum of an MS image contains peaks and a large collection of zeros and noise; therefore, MSI data can be considered as very sparse. Leendert A. Klerk's research group took advantage of this to develop a method to handle MSI data more efficiently using Harwell–Boeing-formatted matrices (Klerk et al., 2007), where the data matrix was stored in a minimal-memory layout that discarded empty values. However, storing sparse matrices does not degrade the information, because data is retained, and null data points are prevented from being stored. In this scenario, the memory footprint is reduced further by associating each peak to a mass bin in the TOF domain. This method is useful for reducing memory requirements and computation time, although most of the algorithms found in commonly used libraries are not designed to handle sparse matrices. Consequently, the main drawback of these data

structures is that alternative algorithm implementations must be written.

Methods of data compression based on raw-data transformations make the algorithms currently used for MSI experiments more efficient (in terms of computational resources). Using these methods, the spectral data is transformed by reducing dimensionality, but keeping the fundamental information. Processing is then executed in this transformed space. The results must be transformed back to the original space if understandable information is to be obtained. Using this workflow, Van de Plas, De Moor, & Waelkens (2008) demonstrated that performing a DWT for each spectrum effectively reduced the computational requirements when only larger wavelet coefficients were retained. Furthermore, when a PCA was computed in this reduced DWT-transformed space instead of using spectral data directly, the results were much more accurate than those acquired in a native data space due to the inherent dimensionality reduction achieved by DWT.

Other strategies rely on dimensionality reduction algorithms, such as PCA (Wold, Esbensen, & Geladi, 1987). These algorithms transform the original variables (i.e., ions) into new ones using linear combinations of the original ions to maximize the variance using as few variables as possible. Here, data reduction are accomplished by removing less-significant variables from the dataset. An associated problem is that the new variables are not directly associated with a particular molecule, making biomarker discovery difficult.

Due to the complexity of processing data for a full MS image, manual segmentation is often chosen. In this case, some ROIs are drawn following the manually discovered patterns in an image. Spectra are then extracted from these images and used as input for data processing. Despite the simplicity and reduced processing time associated with this workflow, results are only obtained for some parts of the image, rendering the rest of the dataset meaningless. However, this approach may be useful for rapidly profiling well-known regions (Hamm et al., 2012; Källback et al., 2012).

VI. MSI SOFTWARE PACKAGES

Recently, various software tools were developed to explore the data produced by MSI instruments and obtain biological tissue information. MALDI-MS equipment usually comes with dedicated software used to control acquisition and perform common imaging tasks. However, in some cases, the software supplied by the manufacturer might not fulfill all of the processing requirements. In such cases, functionality can be increased by including extra software packages in the processing chain. These packages can be obtained from the instrument manufacturer or third-party providers. In this section, we classified the available software packages according to their licensing agreements and features: commercial, freeware, and open source. Commercial tools are private software packages generally developed by companies and can only be used if a license is purchased. Freeware tools are also private, but are under a licensing agreement that allows their use free of charge in some situations. In contrast, an open-source tool provides access to the source code and is very often free of charge.

Below, we discuss the more common software tools that can work with MSI datasets. Each software-licensing group is introduced, describing its weakness and strengths. We hope that

TABLE 3. Summary of software tools for MS imaging.

Software tool	MS Image visualization options	Main preprocessing methods	Main data Analysis methods	Operating systems or platform	Supported input data formats	Supported output data formats
Commercial software tools						
flexImaging (Bruker) www.bruker.com	- Multiple <i>m/z</i> ion visualization using user-selectable colors - Optical image overlay	- Baseline correction - Noise reduction - Alignment - Normalization - Peak Picking (using FlexAnalysis)	- Univariate statistics - Genetic Algorithm - SVM - Supervised Neural Network - QuickClassifier - K-Nearest Neighbor - PCA (using ClinProTools)	Windows	Bruker instruments acquisition	ASCII imzML (since version 4.1) Analyze7.5
SCiLS Lab (SCiLS) scils.de	- Multiple <i>m/z</i> ion visualization using user-selectable colors - Optical image overlay - 3D volume reconstruction	- Baseline correction - Noise reduction - Alignment - Normalization - Peak Picking	- spatial segmentation with edge-preserving spatial denoising - pLSA - pixel classification	Windows	FlexImaging format	FlexImaging format
MALDIVision (PREMIER Biosoft) www.premierbiosoft.com	- Multiple <i>m/z</i> ion from various images visualization using user-selectable colors - Optical image overlay	- Normalization to reference compound	- Histogram - Cumulative Probability Graph	Windows	imzML Analyze7.5	
TissueView (Sciex)	- Single <i>m/z</i> ion representation - Three <i>m/z</i> ions visualization in RGB - Optical image overlay	- Alignment - Normalization - Peak Picking (using MarkerView)	- PCA - PCA-DA (discriminant analysis) - PCVG (Principal Component Variable Grouping) (using MarkerView)	Windows	Analyze7.5	
ImageQuest (Thermo Scientific) www.thermofisher.com	- Single <i>m/z</i> ion representation - Side by side presentation of optical image - <i>m/z</i> ion scroll animation	- Normalization		Windows	Thermo Scientific instruments acquisition	
High Definition Imaging (Waters) www.waters.com	- Multiple <i>m/z</i> ion visualization using user-selectable colors - Three <i>m/z</i> ions visualization in RGB	- Peak picking	- PCA - PLS-DA - S-plots - hierarchical clustering.	Windows	Waters instruments acquisition	ASCII imzML
Quantinetix (ImaBiotech) www.imabiotech.com	- Multiple <i>m/z</i> ion visualization using user-selectable colors - Optical image overlay	- Normalization to reference compound	- Quantification	Windows	FlexImaging format Thermo format Waters format imzML Analyze7.5	
Freeware software tools						
msiQuant (Uppsala University) www.maldi-msi.org	- Single <i>m/z</i> ion representation - Optical image overlay	- Normalization	- Quantification	Windows	FlexImaging imzML	
BioMap (Novartis) www.maldi-msi.org	- Multiple <i>m/z</i> ions using various images - Side by side presentation of optical image			Windows, OSX, Linux	Analyze7.5	
Datacube Explorer (AMOLF) amolf.nl/download/datacubeexplorer	- Single <i>m/z</i> ion representation - 3D volume reconstruction		- Kohonen map clustering	Windows	imzML Analyze7.5	
Mirion (Justus Liebig University) (Paschke et al., 2013)	- Three <i>m/z</i> ions visualization in RGB - Optical image overlay			Windows	imzML	
OpenMSI (Lawrence Berkeley National Laboratory) openmsi.nersc.gov	- Three <i>m/z</i> ions visualization in RGB			any web browser	Author must be contacted for data uploading.	OpenMSI data file based on HDF5
Open source software tools						
MSIReader (NC State University) http://www4.ncsu.edu/~dc/muddim/msireader.html	- Single <i>m/z</i> ion representation - Three <i>m/z</i> ions visualization in RGB - Optical image overlay	- Baseline correction - Noise reduction - Alignment - Normalization - Peak Picking		Matlab (Windows, OSX, Linux)	ASCII mzXML imzml Analyze 7.5	
OmniSpect (Emory University) cs.appstate.edu/omnispect	- Single <i>m/z</i> ion representation - Three <i>m/z</i> ions visualization in RGB		- NMF Algorithm	Matlab (Windows, OSX, Linux)	mzXML imzML Analyze 7.5 NetCDF	
Cardinal (Purdue University) cardinalmsi.org	- Command line: interface, images are generated using R instructions	- Baseline correction - Alignment - Normalization - Peak Picking	- PCA - PLS-DA - Classification based on regularized nearest shrunken centroids	R (Windows, OSX, Linux)	imzML Analyze 7.5	

this discussion helps decide whether a given tool will be useful for a specific application. The main differences between these tools are summarized in Table 3, including input/output data format, build-in processing, and supported platforms.

A. Commercial Software Tools

Usually, commercial software tools come with the MSI instrument and provide the necessary functions to control acquisition and visualization of raw data. Despite its cost, commercial software tools are often the most user-friendly solution, enabling anyone without in-depth knowledge of MSI data to visualize the results.

1. FlexImaging

FlexImaging is the software portion of Bruker's imaging platform. FlexImaging provides a graphical front-end for user-friendly control of data acquisition and visualization. Images of various ions can be represented within a defined tolerance and combined with an optical image of the sample.

FlexImaging delegates the processing of the raw spectra to FlexAnalysis. By doing so, Bruker takes advantage of a long list of well-known algorithms implemented in FlexAnalysis, including baseline correction, normalization, and calibration. For statistical analysis, FlexImaging is designed to easily interface with Bruker ClinProTools, which can perform multivariate calculations, such as PCA, SVM, or hierarchical clustering, as well as univariate statistical tests. The results generated with ClinProTools can be plotted using FlexImaging. These results can then be mapped over the image to view the spatial distribution of the processed data. FlexImaging performs all computations using its own proprietary data format and allows the export of data using open formats, including ASCII, Analyze7.5, and imzML (since v4.1).

2. SCiLS Lab

SCiLS lab is designed for use with the Bruker platform and is part of Bruker's MALDI Molecular Imager solution. SCiLS imports data from FlexImaging in Bruker's native imaging format and can export results to an Excel spreadsheet and also back to FlexImaging. Because SCiLS is able to exchange data only with Bruker's platforms, its use is limited and should be considered as an extension of the Bruker imaging platform.

SCiLS Lab is a full-featured integrated solution for straightforwardly visualizing and statistically analyzing MALDI MSI data in order to make it more readily interpretable. It is able to perform common preprocessing steps, as well as univariate and multivariate statistical analyses. Additionally, it can spatially cluster biologically different tissue regions using supervised or unsupervised approaches. For supervised clustering, the algorithm learns patterns from user-defined tissue regions to obtain a segmentation map.

3. MALDIVision

MALDIVision is a platform-independent tool that is particularly strong in data visualization. In order to be compatible with most MSI instruments in the market, it uses standard file formats (Analyze7.5 and imzML) to import data. Multiple images can

be overlaid and mapped to different colors to enable comparison of spatial distributions of selected ions or combined with optical images to perform histological validation. The images can also be displayed at an intensity normalized to that of a standard compound assumed to be homogenously distributed in tissue.

This software can calculate such typical statistical parameters as mean, median, or standard deviation from user-defined areas, and can also perform more advanced tasks, including the production of histograms and cumulative-probability graphs, to visualize ion-intensity distribution. Many features provided by MALDIVision are also available in some freeware and open-source tools, making MALDIVision an effective solution for simple MS-data exploration.

4. TissueView

TissueView is an MSI tool from Sciex (Framingham, MA, USA). The program can handle imaging data directly from instruments made by the same manufacturer or from Analyze7.5 files. The tool focuses on image visualization and can represent a single mass-ion bin by mapping the intensity onto a color scale. Additionally, up to three ions can be co-localized with each intensity being coded in a RGB-color channel and can also import optical images that can be overlaid with MS images. For data-processing purposes, TissueView can calculate the average spectrum and provide the ion distribution in a particular tissue region. However, for more advanced data analysis, spectra can be interfaced with Sciex MakerView to perform statistics with tools, such as PCA. The software can also load the data in Sciex Data Explorer to identify the proteins using a Mascot server (<http://www.matrixscience.com/server.html>).

6. ImageQuest

ImageQuest is the MSI-visualization tool used by Thermo Fisher Scientific (Waltham, MA, USA) MALDI instruments and reads data in the raw data format used by the manufacturer. This program provides various image-reconstruction alternatives for representing spatially mapped ion intensity. To aid navigation, the optical image used during acquisition is also displayed, but not overlaid, with the MS image.

To rapidly identify which raster positions contain relevant information, ImageQuest introduces a plot window named "chromatogram", where the overall intensity of each pixel is represented versus each scan. Clicking on a chromatogram peak prompts ImageQuest to show where each scan is located on the 2D image, as well as its spectrum. The visualization can also be animated to find unknown peaks. In this mode, ImageQuest will scroll automatically through the defined mass range, enabling the user to observe how the image evolves for each selected mass.

7. High-Definition Imaging (HDI)

HDI is the integrated MSI software solution by Waters Corporation (Milford, MA, USA). It is designed to interface with Waters mass spectrometry instruments in a unified way, from data acquisition to processing and visualization, and is capable of exporting to standard formats, such as imzML and ASCII. Various images from different ions can be represented simultaneously, and images can be reconstructed from a given mass

range, a peak selected from an automatically generated list, or a combination of three overlaid images mapped onto an RGB-color space.

In addition to image visualization, HDI also focuses on discovering meaningful information behind the data. In this regard, the typical preprocessing algorithms are included, as well as a set of statistical tools, such as PCA, PLS-DA, S-plots, and hierarchical clustering.

8. Quantinetix

Quantinetix is suited for specific molecule quantification in MALDI MSI experiments. It was designed to support a wide range of formats, enabling its integration into almost any instrument workflow. The data formats supported range from standard imzML and Analyze7.5 to native proprietary formats, such as those used by Bruker, Sciex, Thermo Fisher Scientific, and Waters Corporation.

To accurately quantify a compound, three normalization techniques enable users to choose which one best fits their needs. These algorithms include on-tissue dilution, isotopic labeling, and Ion suppression. MS images can be overlaid with optical images and are generated from single-ion intensity distributions or with multiple ions assigned to various colors. In addition to the image representation, the tool also provides plot windows showing information about the quantification and normalization algorithms.

B. Freeware Software Tools

Freeware software tools have been widely used as a zero-cost solution for data visualization, providing a frontend for exchanging MSI data through various collaborators. However, freeware tools are limited in MS-data processing, and, consequently in some situations, they are not a viable alternatives to commercial tools.

1. msiQuant

msiQuant (Källback et al., 2012, 2016) is a tool for assisting the labeled normalization and quantitation of drugs and neuropeptides directly in tissue sections. It includes a data-processing chain carefully designed to minimize peak alterations. The baseline correction is implemented with its novel sorted mass spectrum transform algorithm, and, depending on the features of the sample, the normalization approach can be chosen from four algorithms. The software is designed to work within user-defined ROIs and discard meaningless data, thereby saving computer resources.

In terms of data-importing facilities, it can load images from the original Bruker file format and imzML. This software should be considered as an alternative to Quantinetix and is a low-cost solution; however, it also introduces some novel algorithms than can improve quantification in some cases.

2. BioMap

BioMap provides a visualization platform that supports image modalities, such as optical, positron emission tomography, computed tomography, near-infrared fluorescence, and MSI. This makes it possible to combine images generated from

several experimental techniques; however, because BioMap is not a MALDI-MS-specific software package, it lacks typical MS-processing algorithms. It can be extended by adding modules written in interactive data language (IDL) and capable of analyzing specific data. Because BioMap is a general imaging solution that does not specifically target MSI, the file format used for data storage is Analyze7.5.

Despite the frequent use of BioMap in MSI analysis, it also results in frequent memory errors during the processing of large MS datasets. This suggests that all processing is performed in RAM, making this a sub-optimal solution for use with current high-resolution MS images. Moreover, its execution in an IDL environment could complicate the installation procedure for the average user.

3. Datacube Explorer

Datacube Explorer (Klinkert et al., 2014) is an MSI-visualization tool that also includes the capability to perform clustering on images using Kohonen map algorithms. Despite the possibility of the Kohonen map-segmentation feature not being useful for many users, this package is compatible for the future integration of other algorithms. It is also capable of reconstructing 2D images by selecting a particular ion, although 3D reconstruction is also possible when a dataset contains a collection of consecutive tissue slices.

Datacube Explorer supports standard open formats (ImzML and Analyze7.5), but also includes its own format optimized for better handling of large datasets. Datacube Explorer is an optimal low-cost alternative for simple exploration of MS data.

4. Mirion

Mirion (Paschke et al., 2013) is an image-exploration tool that supports importing data from proprietary formats (XCalibur; Thermo Fisher Scientific) and the imzML format. Images can be generated from manual ion selection or automatically using an embedded algorithm based on the selection of the most repetitive peak. This automatic feature uses a mass histogram generated from the full dataset to select the more dominant peaks. Mirion also enables ion images to be combined with optical images, forming a multilayer image that can be represented using different color channels.

Currently, Mirion is limited to run under 32 bits thus its memory is limited at 2 GB. Such limitations prevent exploration of many MS images acquired by modern instruments with high mass and/or spatial resolution. In such situations, Mirion offers the possibility of loading only a part of the data in order to explore it.

5. OpenMSI

OpenMSI (Rübel et al., 2013) is a web platform that provides an application program interface plus an interface to retrieve and explore MSI data. Their website can be used to upload MSI data and explore it anywhere using a computer connected to the internet and without the requirement of specific software tools. Uploading MSI data to a web server overcomes storage problems derived from performing many acquisitions. Moreover, sharing data using a web browser through their OpenMSI interface drastically simplifies manual MS-data exploration in a

large research groups. However, OpenMSI currently lacks the processing tools required for MSI-data analysis.

OpenMSI provides a file format based on the HDF5 format that is highly optimized for efficient storage and access of MSI information. Data are stored in chunks to improve input/output performance, with these chunks compressed into the lossless GZIP format to reduce network bandwidth required for file transfer and provide for efficient storage. Furthermore, to enable rapid access of individual spectra, data are replicated to overcome linearized binary format limitations. Despite data replication, the final stored data size are still compressed to reduce raw data size.

C. Open-Source Software Tools

Open-source software tools are a great option for low-budget MSI-data analysis. Due the fact that anyone can read and modify the code of open-source software, advanced users can adapt them to their specific requirements, enabling anyone with a programming background to expand an open-source software tool to satisfy their processing pipeline. Most open-source tools used for MSI analysis execute under platforms, such as MATLAB or R, making some knowledge of these environments necessary. Because many users are unfamiliar with programming, some freeware tools may be a better choice in situations where raw data needs to be visualized. However, the experienced user will likely discover the most flexible and powerful solutions using open-source tools.

1. MSiReader

MSiReader (Robichaud et al., 2013) is a tool developed in MATLAB and provides a full-featured graphical user interface for the loading and visualization of MSI data from various file formats, including mzXML, imzML, Analyze7.5, and ASCII. Data can be represented in a user-friendly way by selecting a representative ion and a mass tolerance. It also has processing capabilities supporting baseline correction, normalization to a specific peak or by TIC, peak picking, and background subtraction. Features can also be automatically extracted by selecting the most abundant peaks in a selected ROI. Despite its processing tools, MSiReader also enables individual spectra to be exported and custom processing algorithms to be integrated into the MATLAB environment. Despite MSiReader being open source, it has been implemented into the MATLAB platform, which is neither open nor free.

2. OmniSpect

OmniSpect (Parry et al., 2013) performs computationally intensive functions on a remote server, with the functions divided into data-converting tools and multivariate-analysis algorithms for MSI datasets. Similar to MSiReader, OmniSpect makes intensive use of the MATLAB environment to perform calculations; therefore, its code is open-source, but the runtime execution requires a proprietary backend. This tool can import data from most common imaging formats, including NetCDF, mzXML, imzML, and Analyze7.5, and convert it into a MATLAB representation, after which the user can select up to three ions to represent each image. Moreover, OmniSpect can perform multivariate analysis using the NMFA algorithm to

detect similarities in spatial-ion distributions. OmniSpect provides a web interface to represent information and facilitate control, with such remote-processing features very useful when various users require analysis of MSI data. However, enabling a server to utilize MATLAB in a web interface may be more complicated than running a simple standalone program on a personal computer. Given this potential limitation, each user must consider whether remote processing will be beneficial in each situation.

3. Cardinal

Cardinal (Bemis et al., 2015) is a software package for the R environment that enables data import using two standard formats: imzML and Analyze7.5. This toolset is built based on the R language and is distributed using the Bioconductor website (<https://www.bioconductor.org/>). It does not provide a unified graphical user interface to manipulate visualizations; however, many functions are available to enable direct execution in an R session or use in an R script file. Such functions include MSI-data-loading routines, preprocessing tools, segmentation and classification algorithms, and image visualization. Despite lack of a user-friendly command interface, the package provides adequate documentation. Moreover, the availability to create script files and mix Cardinal code with other R packages provides a powerful platform for MSI-data processing. Comparing Cardinal with MATLAB-based solutions, the open-source nature of the program and the language used to create it makes Cardinal the most cost-efficient solution. Furthermore, integration in a growing R environment with plenty of free packages containing multiple algorithm implementations makes Cardinal a suitable choice for advanced users.

VII. FINAL CONCLUSIONS

In the previous decade, MSI became a key technique used for molecular analysis of biological tissues due to its ability to locate ions in space (drugs, metabolites peptides, or proteins). Greater sophistication in sample preparation and improvements in MALDI-MS instruments have resulted in acquisition of high-quality MS images with resolutions ranging from 2 to 200 μm , making this technique useful for clinical diagnosis. The processing of MSI data remains challenging, with researchers confronted with alterations in the distribution of peak intensities caused by possible inhomogeneity in the organic matrix distribution between pixels (in the case of MALDI applications) or tissue inhomogeneity, effects of ion suppression, or reductions in ionization efficiency throughout extended imaging experiments. In such complex scenarios, the use of bioinformatics strategies for MSI analysis are mandatory, with the main conclusions from this review as follows:

- *The relevance of preprocessing steps.* Preprocessing stages compensate for variations and noise in raw data. Currently, there are a wide variety of available preprocessing algorithms whose suitability depends upon the purpose of each MSI experiment. Peak alignment across all pixels is a crucial step necessary for well-resolved spectral images. When internal calibration signals are not available, peak-alignment strategies are based on cross-correlation of image pixels. If internal-calibration peaks are available (as

in the case of matrix-free LDI), they can be used both for peak alignment and mass calibration. An accurate mass-calibration operation is essential for compound identification. Another critical preprocessing step is intensity spectra normalization. The TIC algorithm is effective at compensating for matrix inhomogeneity, but can lead to distorted images when there are different biological areas present in tissue. If an internal-calibration signal is available, it can be used for intensity calibration. Statistical algorithms can provide accurate results, but they are more difficult to implement and have additional computational requirements.

- *Fusion of images in supervised-classification algorithms.* Sections Multivariate analysis of images reported numerous studies using classical microscopy images as referenced for training multivariate models for the segmentation and classification of molecular images. The simultaneous interpretation of the two kinds of images, with the former providing high spatial resolution and pathological interpretation and the latter molecular information, could lead to a new generation of “fused imaging” strategies or techniques.
- *Molecular images for clinical diagnosis.* The ability to rapidly acquire and characterize MS images of tissues (i.e., <1 hr) could enable a myriad of new applications for clinical diagnosis. The automation of matrix-deposition techniques, together with the increase in the frequent use of UV-pulsed lasers in MALDI-MS instruments, opens up many possibilities, including tumor recognition in clinical practices. However, different tissue samples have been compared with limited success, and real-time multivariate algorithms for tissue-image segmentation and classification need to be developed in the future.
- *Computational strategies.* Researchers developing bioinformatics tools for MSI analysis need to design and implement smart and powerful computational strategies due to the high-dimensionality of MSI-datasets, especially when images are taken at high spatial resolution. Currently, almost any high-level programming platforms and languages, such as R, MATLAB, and Python, include libraries supporting parallel programming. Such resources would be beneficial at reducing the computation time necessary for MSI analyses. Moreover, data processing should be optimized to enable a smaller memory footprint. Compiled programming languages, including C, C++, C#, or Java, enable memory to be controlled more carefully as compared with interpreted platforms, such as R, Phyton, or MATLAB. However, interpreted languages usually provide larger algorithm libraries and higher abstraction layers. Data-reduction algorithms (i.e., binning, peak picking, sparse-matrix storing, etc.) are also desirable based on their reductions of computational load.
- *Software tools.* Many software packages suitable for MSI processing and visualization are currently available. In general, proprietary software tools are user-friendly and provide adequate features; however, open-source tools enable scalability and flexibility, with some providing unique processing methods. The primary bottleneck is the lack of compatibility between different software packages, which complicates the exchange of data. Nevertheless,

several LDI-instrument software tools currently include imzML-exporting features as an attempt to overcome this lack of compatibility.

ACKNOWLEDGMENTS

The authors want to acknowledge the financial support of the Spanish Ministry of Economía y Competitividad through project TEC2012-31074, project TEC2015-69076-P, and PRs predoc-toral grant No. BES-2013-065572 and the Direcció General de Recerca of the Government of Catalonia through project 2014 SGR 01267. NR and XC also acknowledge the financial support of the European Union's Horizon 2020 research and innovation programme under the Marie Skłodowska-Curie grant agreement No. 660034. All authors acknowledge Prof. Martins-Green at the Cell Biology Department of the University of California Riverside for kindly provide the mice brain samples used in the acquisition of the MS images showed in Figure 4.

ABBREVIATIONS

AMASS	Algorithm for MSI Analysis by semi-supervised Segmentation
ANN	artificial neural network
ASCII	American Standard Code for Information Inter-change
AUC	area under the curve
CPU	central processing unit
DA	discriminant analysis
DESI-MSI	desorption electrospray ionization mass spectro-metry imaging
DWT	discrete wavelet transform
FT	fourier transform
FT-ICR	fourier transformed ion cyclotron resonance
GA	genetic algorithm
GPU	graphical processing unit
H&E	hematoxylin and eosin
HC	hierarchical clustering
HDDC	high dimensional discriminant clustering
HPLC-MS	high-performance liquid chromatography mass spectrometry
ICA	independent component analysis
ISODAT	iterative self-organizing data analysis technique algorithm
LDI	laser desorption ionization
MALDI	matrix assisted laser/desorption ionization
MRI	magnetic resonance imaging
MS	mass spectrometry
MSI	mass spectrometry imaging
NIMS	nanostructure initiator mass spectrometry
NMFA	nonnegative matrix factorization analysis
OMP	orthogonal matching pursuit
PALDI	particle assisted laser/desorption ionization
PARAFAC	parallel factor analysis
PCA	principal component analysis
PCA-DA	principal component analysis discriminant anal-yisis
PCA-SDA	principal component analysis-symbolic discri-minant analysis
pLSA	probabilistic latent semantic analysis

PLS-DA	partial least squares discriminant analysis
PQN	probabilistic quotient normalization
RAM	random access memory
RGB	red green blue
RMMC	recursive maximum margin criterion
ROI	region of interest
SALDI	surface assisted laser/desorption ionization
SMA	simple moving average
SMM	simple moving median
SMST	sorted mass spectrum transform
SNR	signal to noise ratio
SOFM	self-organizing feature map
SOM	self-organizing maps
SQM1	simple moving first quartile
SVM	support vector machine
TEC	tissue excitation coefficient
TIC	total ion count
TOF	time of flight
TV	total variation
VE	variance explained
VSN	variance stabilizing normalization

REFERENCES

- Addie RD, Balluff B, Bovée JVMG, Morreau H, McDonnell LA. 2015. Current state and future challenges of mass spectrometry imaging for clinical research. *Anal Chem* 87:6426–6433.
- Aichler M, Walch A. 2015. MALDI Imaging mass spectrometry: Current frontiers and perspectives in pathology research and practice. *Lab Investig* 95:422–431.
- Alexandrov T. 2012. MALDI imaging mass spectrometry: Statistical data analysis and current computational challenges. *BMC Bioinformatics* 13:S11.
- Alexandrov T, Becker M, Deininger S-O, Ernst G, Wehder L, Grasmair M, von Eggeling F, Thiele H, Maass P. 2010. Spatial segmentation of imaging mass spectrometry data with edge-preserving image denoising and clustering. *J Proteome Res* 9:6535–6546.
- Alexandrov T, Chernyavsky I, Becker M, Von Eggeling F, Nikolenko S. 2013. Analysis and interpretation of imaging mass spectrometry data by clustering mass-to-charge images according to their spatial similarity. *Anal Chem* 85:11189–11195.
- Alexandrov T, Kobarg JH. 2011. Efficient spatial segmentation of large imaging mass spectrometry datasets with spatially aware clustering. *Bioinformatics* 27:i230–i238.
- Alexandrov T, Meding S, Trede D, Kobarg JH, Balluff B, Walch A, Thiele H, Maass P. 2011. Super-resolution segmentation of imaging mass spectrometry data: Solving the issue of low lateral resolution. *J Proteomics* 75:237–245.
- Bemis KD, Harry A, Eberlin LS, Ferreira C, van de Ven SM, Mallick P, Stolowitz M, Vitek O. 2015. Cardinal: An R package for statistical analysis of mass spectrometry-based imaging experiments. *Bioinformatics* 31:2418–2420.
- Bergman N, Shevchenko D, Bergquist J. 2014. Approaches for the analysis of low molecular weight compounds with laser desorption/ionization techniques and mass spectrometry. *Anal Bioanal Chem* 406:49–61.
- Bonnel D, Longuespee R, Franck J, Roudbaraki M, Gosset P, Day R, Salzet M, Fournier I. 2011. Multivariate analyses for biomarkers hunting and validation through on-tissue bottom-up or in-source decay in MALDI-MSI: Application to prostate cancer. *Anal Bioanal Chem* 401:149–165.
- Bro R. 1997. PARAFAC. tutorial and applications. *Chemom Intell Lab Syst* 38:149–171.
- Bruand J, Alexandrov T, Sistla S, Wisztorski M, Meriaux C, Becker M, Salzet M, Fournier I, Macagno E, Bafna V. 2011. AMASS: algorithm for MSI analysis by semi-supervised segmentation. *J Proteome Res* 10:4734–4743.
- Buck A, Ly A, Balluff B, Sun N, Gorzolka K, Feuchtinger A, Janssen K, Kuppen PJK, van de Velde CJH, Weirich G, Erlmeier F, Langer R, Aubele M, Zitzelsberger H, Aichler M, Walch A. 2015. High-resolution MALDI-FT-ICR MS imaging for the analysis of metabolites from formalin-fixed, paraffin-embedded clinical tissue samples. *J Pathol* 237:123–132.
- Calavia R, Annanouch F, Correig X, Yanes O. 2012. Nanostructure initiator mass spectrometry for tissue imaging in metabolomics: Future prospects and perspectives. *J Proteomics* 75:5061–5068.
- Chen S, Chen L, Wang J, Hou J, He Q, Liu J, Wang J, Xiong S, Yang G, Nie Z. 2012. 2,3,4,5-Tetrakis(3',4'-dihydroxyphenyl)thiophene: A new matrix for the selective analysis of low molecular weight amines and direct determination of creatinine in urine by MALDI-TOF MS. *Anal Chem* 84:10291–10297.
- Chiang C-K, Chen W-T, Chang H-T. 2011. Nanoparticle-based mass spectrometry for the analysis of biomolecules. *Chem Soc Rev* 40: 1269–1281.
- Cho Y-T, Chiang Y-Y, Shiea J, Hou M-F. 2012. Combining MALDI-TOF and molecular imaging with principal component analysis for biomarker discovery and clinical diagnosis of cancer. *Genomic Med Biomarkers, Heal Sci* 4:3–6.
- Chughtai K, Heeren RMA. 2010. Mass spectrometric imaging for biomedical tissue analysis. *Chem Rev* 110:3237–3277.
- Deininger S-O, Cornett DS, Paape R, Becker M, Pineau C, Rauser S, Walch A, Wolski E. 2011. Normalization in MALDI-TOF imaging datasets of proteins: Practical considerations. *Anal Bioanal Chem* 401:167–181.
- Deininger S-O, Ebert MP, Fütterer A, Gerhard M, Röcken C. 2008. MALDI imaging combined with hierarchical clustering as a new tool for the interpretation of complex human cancers. *J Proteome Res* 7: 5230–5236.
- Denis L, Lorenz DA, Trede D. 2009. Greedy solution of Ill-posed problems: Error bounds and exact inversion. *Inverse Probl* 25:115017.
- Dufresne M, Thomas A, Breault-Turcot J, Masson J-F, Chaurand P. 2013. Silver-assisted laser desorption ionization for high spatial resolution imaging mass spectrometry of olefins from thin tissue sections. *Anal Chem* 85:3318–3324.
- Figueiredo MAT, Nowak RD. 2003. An EM algorithm for wavelet-based image restoration. *IEEE Trans Image Process* 12:906–916.
- Fonville J, Carter C, Cloarec O, Nicholson JK, Lindon JC, Bunch J, Holmes E. 2012. . *Anal Chem* 84:1310–1319.
- Fonville J, Carter CL, Pizarro L, Steven RT, Palmer AD, Griffiths RL, Lalor PF, Lindon JC, Nicholson JK, Holmes E, Bunch J. 2013. . *Anal Chem* 85:1415–1423.
- Franceschi P, Wehrens R. 2014. Self-organizing maps: A versatile tool for the automatic analysis of untargeted imaging datasets. *Proteomics* 14:853–861.
- Gemperline E, Rawson S, Li L. 2014. Optimization and comparison of multiple MALDI matrix application methods for small molecule mass spectrometric imaging. *Anal Biochem* 86:10030–10035.
- Goodwin RJA. 2012. Sample preparation for mass spectrometry imaging: Small mistakes can lead to big consequences. *J Proteomics* 75: 4893–4911.
- Greer T, Sturm R, Li L. 2011. Mass spectrometry imaging for drugs and metabolites. *J Proteomics* 74:2617–2631.
- Groseclose MR, Massion PP, Chaurand P, Caprioli RM. 2008. High-throughput proteomic analysis of formalin-fixed paraffin-embedded tissue microarrays using MALDI imaging mass spectrometry. *Proteomics* 8:3715–3724.
- Guinan TM, Gustafsson OJR, McPhee G, Kobus H, Voelcker NH. 2015. Silver coating for high-mass-accuracy imaging mass spectrometry of fingerprints on nanostructured silicon. *Anal Chem* 87:11195–11202.
- Hair JF, Black CW, Babin BJ, Anderson RE. 2013. Multivariate data analysis. Pearson Education Limited. ISBN: 129202190X, 9781292021904.

- Hamm G, Bonnel D, Legouffe R, Pamelard F, Delbos J-M, Bouzom F, Stauber J. 2012. Quantitative mass spectrometry imaging of propranolol and olanzapine using tissue extinction calculation as normalization factor. *J Proteomics* 75:4952–4961.
- Hanselmann M, Kirchner M, Renard BY, Amstalden ER, Glunde K, Heeren RMA, Hamprecht FA. 2008. Concise representation of mass spectrometry images by probabilistic latent semantic analysis. *Anal Chem* 80:9649–9658.
- Hanselmann M, Köthe U, Kirchner M, Renard BY, Amstalden ER, Glunde K, Heeren RM, Hamprecht FA. 2009. Toward digital staining using imaging mass spectrometry and random forests. *J Proteome Res* 8:3558–3567.
- He QP, Wang J, Mobley JA, Richman J, Grizzle WE. 2011. Self-calibrated warping for mass spectra alignment. *Cancer Inform* 10:65–82.
- Hillenkamp F, Peter-Katalinic J. 2014. MALDI MS: A practical guide to instrumentation, methods and applications (2nd ed.). Weinheim, Germany: Wiley-Blackwell. ISBN: 978-3-527-33331-8.
- Hölscher D, Fuchser J, Knop K, Menezes RC, Buerkert A, Svatoš A, Schubert US, Schneider B. 2015. High resolution mass spectrometry imaging reveals the occurrence of phenylphenalenone-type compounds in red paracytic stomata and red epidermis tissue of *Musa acuminata* ssp. zebrina cv. "Rowe Red." *Phytochemistry* 116:239–245.
- Hong JH, Kang JW, Kim DK, Baik SH, Kim KH, Shanta SR, Jung JH, Mook-Jung I, Kim KP. 2016. Global changes of phospholipids identified by MALDI imaging mass spectrometry in a mouse model of Alzheimer's disease. *J Lipid Res* 57:36–45.
- Jones E, Deininger S-O, Hogendoorn P, Deelder A, McDonnell L. 2012a. Imaging mass spectrometry statistical analysis. *J Proteomics* 75:4962–4989.
- Jones E, van Zeijl RJ, Andrén P, Deelder A, Wolters L, McDonnell LA. 2012b. High speed data processing for imaging MS-based molecular histology using graphical processing units. *J Am Soc Mass Spectrom* 23:745–752.
- Jones EA, van Remoortere A, van Zeijl RJM, Hogendoorn PCW, JVMG Bovée, Deelder AM, McDonnell LA. 2011. Multiple statistical analysis techniques corroborate intratumor heterogeneity in imaging mass spectrometry datasets of myxofibrosarcoma. *PLoS ONE* 6: e24913.
- Källback P, Nilsson A, Sharifgorji M, Andrén PE. 2016. msIQuant—Quantitation software for mass spectrometry imaging enabling fast access, visualization, and analysis of large data sets. *Anal Chem* 88:4346–4353.
- Källback P, Sharifgorji M, Nilsson A, Andrén PE. 2012. Novel mass spectrometry imaging software assisting labeled normalization and quantitation of drugs and neuropeptides directly in tissue sections. *J Proteomics* 75:4941–4951.
- Klerk LA, Broersen A, Fletcher IW, van Liere R, Heeren RMA. 2007. Extended data analysis strategies for high resolution imaging MS: New methods to deal with extremely large image hyperspectral datasets. *Int J Mass Spectrom* 260:222–236.
- Klinkert I, Chughtai K, Ellis SR, Heeren RMA. 2014. Methods for full resolution data exploration and visualization for large 2D and 3D mass spectrometry imaging datasets. *Int J Mass Spectrom* 362:40–47.
- Kohonen T, Kaski S, Lagus K, Salojarvi J, Honkela J, Paatero V, Saarela A. 2000. Self organization of a massive document collection. *IEEE Trans Neural Networks* 11:574–585.
- Kriegsmann J, Kriegsmann M, Casadonte R. 2015. MALDI TOF imaging mass spectrometry in clinical pathology: A valuable tool for cancer diagnostics (Review). *Int J Oncol* 46:893–906.
- Lauzon N, Dufresne M, Chauhan V, Chaurand P. 2015. Development of laser desorption imaging mass spectrometry methods to investigate the molecular composition of latent fingermarks. *J Am Soc Mass Spectrom* 26:878–886.
- Lee DY, Platt V, Bowen B, Louie K, Canaria C a, McMurray CT, Northen T. 2012. Resolving brain regions using nanostructure initiator mass spectrometry imaging of phospholipids. *Integr Biol (Camb)* 4: 693–699.
- Mantini D, Petrucci F, Pieragostino D, Del Boccio P, Di Nicola M, Di Illo C, Federici G, Sacchetta P, Comani S, Urbani A. 2007. LIMPIC: A computational method for the separation of protein MALDI-TOF-MS signals from noise. *BMC Bioinformatics* 8:101.
- Martens L, Chambers M, Sturm M, Kessner D, Levander F, Shofstahl J, Tang WH, Rompp A, Neumann S, Pizarro AD, Montecchi-Palazzi L, Tasman N, Coleman M, Reisinger F, Souda P, Hermjakob H, Binz PA, Deutsch EW. 2011. mzML—A community standard for mass spectrometry data. *Mol Cell Proteomics* 10: R110.000133–R110.000133.
- McCombie G, Staab D, Stoeckli M, Knochenmuss R. 2005. Spatial and spectral correlations in MALDI mass spectrometry images by clustering and multivariate analysis. *Anal Chem* 77:6118–6124.
- McCombie G, Staab D, Stoeckli M, Knochenmuss R. 2005. Spatial and spectral correlations in MALDI mass spectrometry images by clustering and multivariate analysis. *Anal Chem* 77:6118–6124.
- McDonnell LA, Heeren RMA. 2007. Imaging mass spectrometry. *Mass Spectrom Rev* 26:606–643.
- McDonnell LA, van Remoortere A, van Zeijl RJM, Deelder AM. 2008. Mass spectrometry image correlation: Quantifying colocalization. *J Proteome Res* 7:3619–3627.
- Nilsson A, Goodwin RJA, Sharifgorji M, Vallianatou T, Webborn PJH, Andrén PE. 2015. Mass spectrometry imaging in drug development. *Anal Chem* 87:1437–1455.
- Norris J, Cornett D, Mobley J. 2007. Processing MALDI mass spectra to improve mass spectral direct tissue analysis. *Int J Mass Spectrom* 260:212–221.
- Norris JL, Caprioli RM. 2013. Analysis of tissue specimens by matrix-assisted laser desorption/ionization imaging mass spectrometry in biological and clinical research. *Chem Rev* 113:2309–2342.
- Northen TR, Yanes O, Northen MT, Marrinucci D, Uritboonthai W, Apon J, Golledge SL, Nordström A, Siuzdak G. 2007. Clathrate nanostructures for mass spectrometry. *Nature* 449:1033–1036.
- Oetjen J, Aichler M, Trede D, Strehlow J, Berger J, Heldmann S, Becker M, Gottschalk M, Kobarg JH, Wirtz S, Schiffler S, Thiele H, Walch A, Maass P, Alexandrov T. 2013. MRI-compatible pipeline for three-dimensional MALDI imaging mass spectrometry using PAXgene fixation. *J Proteomics* 90:52–60.
- Palmer AD, Alexandrov T. 2015. Serial 3D imaging mass spectrometry at its tipping point. *Anal Chem* 87:4055–4062.
- Parry RM, Galhena AS, Gamage CM, Bennett RV, Wang MD, Fernández FM. 2013. 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* 24:646–649.
- Paschke C, Leisner A, Hester A, Maass K, Guenther S, Bouschen W, Spengler B. 2013. Mirion—A software package for automatic processing of mass spectrometric images. *J Am Soc Mass Spectrom* 24: 1296–1306.
- Prideaux B, Stoeckli M. 2012. Mass spectrometry imaging for drug distribution studies. *J Proteomics* 75:4999–5013.
- Race AM, Styles IB, Bunch J. 2012. Inclusive sharing of mass spectrometry imaging data requires a converter for all. *J Proteomics* 75:5111–5112.
- Rainer M, Qureshi MN, Bonn GK. 2011. Matrix-free and material-enhanced laser desorption/ionization mass spectrometry for the analysis of low molecular weight compounds. *Anal Bioanal Chem* 400:2281–2288.
- Rausch S, Marquardt C, Balluff B, Deininger S-O, Albers C, Belau E, Hartmer R, Suckau D, Specht K, Ebert MP, Schmitt M, Aubele M, Höfler H, Walch A. 2010. Classification of HER2 receptor status in breast cancer tissues by MALDI imaging mass spectrometry. *J Proteome Res* 9:1854–1863.
- Robichaud G, Garrard KP, Barry JA, Muddiman DC. 2013. MSIReader: An open-source interface to view and analyze high resolving power MS imaging files on matlab platform. *J Am Soc Mass Spectrom* 24:718–721.

- Rübel O, Greiner A, Cholia S, Louie K, Bethel EW, Northen TR, Bowen BP. 2013. OpenMSI: A high-performance web-based platform for mass spectrometry imaging. *Anal Chem* 85:10354–10361.
- Samarasinghe S. 2006. Neural networks for applied sciences and engineering: From fundamentals to complex pattern recognition. CRC Press. ISBN: 9780849333750.
- Satten GA, Datta S, Moura H, Woolfitt AR, Carvalho MDG, Carlone GM, De BK, Pavlopoulos A, Barr JR. 2004. Standardization and denoising algorithms for mass spectra to classify whole-organism bacterial specimens. *Bioinformatics* 20:3128–3136.
- Savitzky A, Golay MJE. 1964. Smoothing and differentiation of data by simplified least squares procedures. *Anal Chem* 36:1627–1639.
- Schramm T, Hester A, Klinkert I, Both J-P, Heeren RM, Brunelle A, Stolowitz M, Laprévote O, Desbenoit N, Robbe MF, Stoeckli M, Spengler B, Römpf A. 2012. ImzML—A common data format for the flexible exchange and processing of mass spectrometry imaging data. *J Proteomics* 75:5106–5110.
- Schwamborn K, Krieg RC, Reska M, Jakse G, Knuechel R, Wellmann A. 2007. Identifying prostate carcinoma by MALDI-imaging. *Int J Mol Med* 20:155–159. Retrieved from <http://www.scopus.com/inward/record.url?eid=2-s2.0-35448978414&partnerID=tZOTx3y1>
- Silina YE, Volmer DA. 2013. Nanostructured solid substrates for efficient laser desorption/ionization mass spectrometry (LDI-MS) of low molecular weight compounds. *Analyst* 138:7053.
- Smith DF, Schulz C, Konijnenburg M, Kilic M, Heeren RMA. 2015. Distributed computing strategies for processing of FT-ICR MS imaging datasets for continuous mode data visualization. *Anal Bioanal Chem* 407:2321–2327.
- Thiele H, Heldmann S, Trede D, Strehlow J, Wirtz S, Dreher W, Berger J, Oetjen J, Kobarg JH, Fischer B, Maass P. 2014. 2D and 3D MALDI-imaging: Conceptual strategies for visualization and data mining. *Biochim Biophys Acta – Proteins Proteomics* 1844:117–137.
- Tracy MB, Chen H, Weaver DM, Malyarenko DI, Sasinowski M, Cazares LH, Drake RR, Semmes OJ, Tracy ER, Cooke WE. 2008. Precision enhancement of MALDI-TOF MS using high resolution peak detection and label-free alignment. *Proteomics* 8:1530–1538.
- Trede D, Schiffler S, Becker M, Wirtz S, Steinhorst K, Strehlow J, Aichler M, Kobarg JH, Oetjen J, Dyatlov A, Heldmann S, Walch A, Thiele H, Maass P, Alexandrov T. 2012. Exploring three-dimensional matrix-assisted laser desorption/ionization imaging mass spectrometry data: Three-dimensional spatial segmentation of mouse kidney. *Anal Chem* 84:6079–6087.
- Van de Plas R, De Moor B, Waelkens E. 2008. Discrete wavelet transform-based multivariate exploration of tissue via imaging mass spectrometry. In: Proceedings of the 2008 ACM symposium on applied computing—SAC '08. New York, New York: ACM Press. (p. 1307).
- Van de Plas R, Yang J, Spraggins J, Caprioli RM. 2015. Image fusion of mass spectrometry and microscopy: A multimodality paradigm for molecular tissue mapping. *Nat Methods* 12:366–372.
- van Herk M. 1992. A fast algorithm for local minimum and maximum filters on rectangular and octagonal kernels. *Pattern Recognit Lett* 13: 517–521.
- Veselkov KA, Mirnezami R, Strittmatter N, Goldin RD, Kinross J, Speller AVM, Abramov T, Jones EA, Darzi A, Holmes E, Nicholson JK, Takats Z. 2014. Chemo-informatic strategy for imaging mass spectrometry-based hyperspectral profiling of lipid signatures in colorectal cancer. *Proc Natl Acad Sci* 111:1216–1221.
- Wang L. 2005. Support vector machines: Theory and applications. Springer Science & Business Media. ISBN 978-3-540-24388-5.
- Wold S, Esbensen K, Geladi P. 1987. Principal component analysis. *Chemom Intell Lab Syst* 2:37–52.
- Xiong X, Xu W, Eberlin LS, Wiseman JM, Fang X, Jiang Y, Huang Z, Zhang Y, Cooks RG, Ouyang Z. 2012. Data processing for 3D mass spectrometry imaging. *J Am Soc Mass Spectrom* 23:1147–1156.