DOI: 10.1002/mas.21794

REVIEW ARTICLE

WILEY

What are we imaging? Software tools and experimental strategies for annotation and identification of small molecules in mass spectrometry imaging

Gerard Baquer[1] | **Lluc Sementé**[1] | **Toufik Mahamdi**[1] | **Xavier Correig**[1,2,3] | **Pere Ràfols**[1,2,3] | **María García-Altares**[1,2]

[1]Department of Electronic Engineering, University Rovira I Virgili, Tarragona, Spain

[2]Spanish Biomedical Research Center in Diabetes and Associated Metabolic Disorders (CIBERDEM), Madrid, Spain

[3]Institut D'Investigacio Sanitaria Pere Virgili, Tarragona, Spain

Correspondence
Pere Ràfols, Department of Electronic Engineering, University Rovira i Virgili, Tarragona, Spain.
Email: pere.rafols@urv.cat

María García-Altares, Spanish Biomedical Research Center in Diabetes and Associated Metabolic Disorders (CIBERDEM), 28029 Madrid, Spain.
Email: maria.garcia-altares@urv.cat

Abstract

Mass spectrometry imaging (MSI) has become a widespread analytical technique to perform nonlabeled spatial molecular identification. The Achilles' heel of MSI is the annotation and identification of molecular species due to intrinsic limitations of the technique (lack of chromatographic separation and the difficulty to apply tandem MS). Successful strategies to perform annotation and identification combine extra analytical steps, like using orthogonal analytical techniques to identify compounds; with algorithms that integrate the spectral and spatial information. In this review, we discuss different experimental strategies and bioinformatics tools to annotate and identify compounds in MSI experiments. We target strategies and tools for small molecule applications, such as lipidomics and metabolomics. First, we explain how sample preparation and the acquisition process influences annotation and identification, from sample preservation to the use of orthogonal techniques. Then, we review twelve software tools for annotation and identification in MSI. Finally, we offer perspectives on two current needs of the MSI community: the adaptation of guidelines for communicating confidence levels in identifications; and the creation of a standard format to store and exchange annotations and identifications in MSI.

KEYWORDS

identification confidence levels, mass spectrometry imaging, metabolomics, molecular annotation, molecular identification, software

Abbreviations: CCS, collision cross-section; DDA, data dependent acquisition; DESI, desorption electrospray ionization; ESI, electrospray ionization; FDR, false discovery rate; FFPE, formalin-fixed paraffin-embedded; FTICR, Fourier-transform Ion cyclotron resonance; FT-IR, Fourier-transform infrared; GC-MS, gas chromatography-mass spectrometry; HCD, higher-energy collision-induced dissociation; IMS, ion mobility spectrometry; IT, ion trap; KMD, Kendrick mass defect; LA-ICP, laser ablation inductively coupled plasma; LCM, laser-capture microdissection; LC-MS, liquid chromatography-mass spectrometry; MALDI, matrix-assisted laser desorption/ionization; MS, mass spectrometry; MSI, mass spectrometry imaging; MS/MS, tandem mass spectrometry; m/z, mass to charge; NMR, nuclear magnetic resonance; NP, nanoparticle; ROI, region of interest; RT, retention time; SIL, stable isotope labeling; SIMS, secondary ion mass spectrometry; TOF, time-of-flight; t-MALDI, transmission MALDI.

Gerard Baquer and Lluc Sementé contributed equally to this study.

1 | INTRODUCTION: THE CHALLENGE OF ANNOTATION AND IDENTIFICATION IN MASS SPECTROMETRY IMAGING (MSI)

MSI is an analytical technique capable of spatially resolving the chemical composition of biological tissues (Buchberger et al., 2018). Over recent years, MSI has become a key technique in diverse fields such as biochemistry, pharmaceutics, and medical diagnostics (Patti et al., 2012; Ren et al., 2018; Schulz et al., 2019; Vaysse et al., 2017). Its use in metabolomics, the study of small molecules in biological specimens (Clish, 2015), is of particular interest as metabolites serve a wide variety of biological purposes such as structural, signaling, immune modulators, endogenous toxins, and environmental sensors (Wishart, 2019).

To draw meaningful biological and diagnostic conclusions from MSI experiments, the mass to charge (m/z) ratios obtained need to be traced back to unique compound identifications. This is a nontrivial task considering that spectra in mass spectrometry (MS) are often cluttered with signals from isotopes, adducts, in-source fragments, multiple-charges, matrix, and other exogenous compounds. It is estimated that monoisotopic endogenous peaks only represent 5% of the MS signals in an MSI experiment (Wang et al., 2019). This is particularly challenging in metabolomics since matrix signals and in-source fragments are densely concentrated in the low mass range (Baquer et al., 2020; Janda et al., 2021). The vast amount of MS signals leaves research groups using MSI around the world struggling with the question: "What are we detecting in MSI experiments?"

Workflows for identification of compounds by other MS-based techniques such as gas or liquid chromatography-mass spectrometry (GC-MS and LC-MS) mostly rely on chromatographic separation, followed by MS analysis and often MS/MS experiments. However, these workflows cannot be directly applied to MSI experiments:

(1) MSI lacks chromatographic separation: GC-MS and LC-MS use chromatographic columns to separate compounds by their chemical properties (such as polarity) (Lisec et al., 2006; Pitt, 2009) and use retention times (RT) as complementary information to aid compound identification. This information is not available in MSI experiments (Amstalden van Hove et al., 2010; Buchberger et al., 2018; Yagnik et al., 2013).
(2) Most MSI experiments are only performed in Full MS scan: multiple isobars and isomers with different chemical, physical and functional properties can be associated with a given monoisotopic mass (Kyle et al., 2016). Tandem mass spectrometry (MS/MS) can distinguish them by their fragmentation spectra (McLafferty, 1981). Similarly, ion mobility instruments use ion drift times to facilitate the identification of isomers (Mesa Sanchez et al., 2020). In MSI it is still not routinary to perform MS/MS fragmentation and ion mobility separation on-tissue in an untargeted fashion (Amstalden van Hove et al., 2010; Buchberger et al., 2018; Yagnik et al., 2013).

On the flip side, peak annotation in MSI experiments is statistically more robust given the higher number of data points (each pixel contains a unique spectrum). Spatial correlations between different ion MS signals add statistical confidence to ion annotations (Sementé et al., 2021).

This complex analytical context calls for well-designed experimental strategies and automated software-based solutions to perform robust molecular annotation and identification in MSI metabolomics.

In this review, we explain how each step of the sample preparation and acquisition process influences annotation and identification, from artifacts that may be introduced during sample preservation, to the use of orthogonal techniques like LC-MS/MS with the same tissue. Later, we discuss how different bioinformatics tools annotate and identify compounds in MSI experiments. We specifically target tools for small molecule applications such as lipidomics and metabolomics. This review offers an analytical background for the bioinformatician to understand the influence of each experimental step on annotation and identification. In turn, analytical chemists will discover the possibilities that bioinformatics offers to support compound annotation and identification in MSI. We also point out how the MSI community struggles to communicate confidence levels for identification and lacks a standard format to report annotations and identifications. As a solution, we propose to adopt the five level scheme by Schymanski et al. (2014), and we draft a file format annex to imzML based on mzTab-M (Hoffmann et al., 2019) to report annotations and identifications in MSI.

2 | THE NEED FOR REPORTING STANDARDS IN MSI

2.1 | A word about the terms *annotation* and *identification*

According to the Metabolomics Standards Initiative, a nonnovel molecule is considered "identified" when its experimental data is compared to a standard by at least two types of orthogonal data (for instance, RT and MS/

MS), while a compound would be considered "annotated" if identification is not achieved (Sumner et al., 2007). A common problem in metabolomics (Salek et al., 2013) and MSI scientific articles is that the terms annotation and identification are sometimes used interchangeably, at times even accompanied by the adjectives "putative" or "tentative." This confusion impedes the comparison of different annotation/identification strategies and the interpretation results.

To seize the impact of this problem in the MSI community, we reviewed the usage of the terms "annotation" and "identification" in 58 papers published in the last 5 years (Supporting Information: Table S1) dealing with annotation/identification from several perspectives (bioinformatics, experimental protocol, instrumental, and application).

We found that 52% of the papers use the term "identification" to refer to exact mass matching at least once (when "annotation" should be used). Moreover, the adjectives "putative" and "tentative" are used in 31% of the papers. When they appear, they accompany the terms annotation and identification indistinctly to refer to exact mass matching.

2.2 | Adaptation of identification confidence levels for MSI

Communicating the degree of confidence in compound identification is essential to avoid misinterpretation of the results, and to compare identification strategies. While the MSI community has its own initiative for improving standardization and reproducibility (MALDISTAR, https://www.maldistar.org/), at the moment the aims of this initiative do not include the definition of guidelines for reporting the confidence of compound annotation and identification. Besides, current reporting standards for mass spectrometry imaging (Gustafsson et al., 2018; McDonnell et al., 2015) do not explicitly mention identification confidence levels. The 2015 guideline proposed by McDonnell et al. (2015) defines the minimum reporting standards for identifications as (1) experimental and theoretical *m/z*, (2) mass tolerance, (3) MS/MS on-tissue, and (4) orthogonal measurements (i.e., LC-MS/MS). However, this scheme does not communicate different degrees of confidence in MSI identifications and annotations.

On the other hand, the metabolomics community does have well-accepted guidelines for communicating identification confidence based on the four-level system suggested by the Metabolomics Standards Initiative in 2007 (Sumner et al., 2007). In 2014, Schymanski et al. (2014) proposed a five level system to rank levels of confidence in identification: (Level 1) Confirmed structure matched against a reference standard (MS, MS/MS, and RT); (Level 2) Probable structure matched against literature or library spectrum (MS, MS/MS, and RT); (Level 3) Tentative candidates matched against literature or library spectrum (MS, MS/MS, and RT); (Level 4) Unequivocal molecular formula (MS with adduct and isotope information); (Level 5) Exact mass (MS). Later, Schrimpe-Rutledge et al. (2016) expanded the model by proposing the use of orthogonal techniques, such as nuclear magnetic resonance (NMR) or ion mobility, to reach level 2 and level 3 identifications.

The scheme of five confidence levels used in metabolomics (Schrimpe-Rutledge et al., 2016; Schymanski et al., 2014) shown in Supporting Information: Figure S1 could be adopted to report identification confidences in MSI experiments. As the information obtained by MSI is different from the data collected by common metabolomics techniques (usually based on chromatographic separation), we suggest the adaptation of the five level system to report identification confidence in MSI experiments as described below. The strategies to achieve the different confidence levels mentioned in this section are described in detail in Sections 3 and 4.

Level 1 Confirmed structure: Reporting exact mass, unequivocal molecular formula, and a single confirmed structure. At this level, a unique structure is confirmed by comparing all experimental data from Levels 2–5 to reference standards. The use of reference standards for confirming identifications in MSI may include spotting the standard on the glass slide or substrate, on a replicate tissue, or spiking a homogenized replicated tissue. Alternatively, one can perform LC-MS/MS measurements of tissue homogenates or microdissection of the tissue to compare against standards dissolved in solvents or in tissue extracts (matrix-matched comparison). The discrimination of isobaric and isomeric species is one of the major challenges of MSI, since it cannot rely of the chromatographic separation of these species. As described in a recent review (Bednařík et al., 2022), there are several strategies to discriminate isomers in MSI, including ion activation and chemical derivatization, ion mobility spectroscopy, and tandem MS analysis directly on tissue, among others.

Level 2 Probable structure: Reporting exact mass, unequivocal molecular formula, and a single possible structure. This level is achieved when only one unambiguous possible structure results after following the procedures described in Level 3.

Level 3 Tentative candidates: Reporting exact mass, unequivocal molecular formula, and a list of possible structures. This level requires information complementary to the MS measurement that can be obtained using

orthogonal data, obtained during the MSI experiment (like ion mobility or MS/MS fragmentation) or by orthogonal techniques such as LC-MS/MS on homogenized tissues, or complementary molecular imaging techniques. If MS/MS is used, the obtained experimental spectra are matched against experimental, in silico or literature libraries.

Level 4 Unequivocal molecular formula: Reporting exact mass and unequivocal molecular formula. This requires the integration of MS information such as isotopes, adducts, and/or in-source fragments. In MSI, the annotation of isotopes, adducts, and in-source fragments benefit from the high number of sampling points over the tissue. The spatial correlation of signals (not available in other MS methods) ensures robust Level 4 annotation.

Level 5 Exact mass of interest: Reporting only the exact mass of the compound, together with the mass tolerance of the MSI method. Unable to distinguish between different molecular formulas within the mass tolerance of the method.

3 | INFLUENCE OF THE SAMPLE PREPARATION AND SPECTRA ACQUISITION PROCEDURES FOR MOLECULAR ANNOTATION AND IDENTIFICATION

This section covers the influence of experimental procedures in compound annotation and identification in MSI. It describes experimental strategies regarding sample preparation, instrumental setups, and combinations of MSI with other techniques. It provides a solid analytical background for bioinformaticians working in MSI annotation and identification. For a deeper explanation of MSI experimental procedures, the reader is referred to more extensive reviews (Amstalden van Hove et al., 2010; Buchberger et al., 2018; Chatterji & Pich, 2013; Gode & Volmer, 2013; Norris & Caprioli, 2013). Table 1 contains a compendium of the principal effects in annotation/identification of all the procedures and instruments covered in this section.

3.1 | Effects of the sample preparation in MSI annotations and identifications

Sample preparation is a critical step in any MSI experiment, as it largely influences which compounds will be ionized and detected. Proper sample preparation will also reduce ion suppression, adduct formation, matrix interferences, and in-source fragmentation. Besides, the use of calibrants improves the mass axis calibration and increases the confidence of annotations by exact mass.

3.1.1 | Sample preservation

Sample preservation is the first decision that affects an MSI experiment, as it determines what type of compounds will remain in the tissue. There are three main preservation options: formalin-fixed paraffin-embedded (FFPE) tissues, fresh-frozen tissues, and formalin-fixed frozen tissues.

FFPE tissues have been the gold standard for the fixation and storage of samples for histopathological analyses. FFPE tissues can be preserved at room temperature for years without degradation and are easy to section and transport thanks to the wax embedding. Nevertheless, paraffine induces ion suppression during the ionization process in MSI, and formalin fixation (which cross-links proteins together) hampers the desorption/ionization of proteins and peptides. Moreover, both compounds contaminate the spectra by adding more signals. Thus, the use of FFPE tissues for MSI requires the removal of the paraffine before MSI analysis (by a series of xylene and ethanol washing steps); and the reversal of the cross-linking of proteins (by antigen retrieval protocols). These washing steps lead to the loss of lipids and metabolites, thus FFPE tissues are better suited for peptide and protein analysis by MSI (Hermann et al., 2020; Ly et al., 2016; Wisztorski et al., 2010).

Fresh-frozen tissues have the advantage of stopping postmortem decay (autolysis) without using any chemical agent that may induce changes in the tissue. In principle, this allows the preservation of all the molecular species in the tissue, thus enabling the detection of metabolites, lipids, and proteins. This makes fresh-frozen the standard sample preservation for MSI. Nevertheless, fresh-frozen samples are costly to store, as they require −80°C freezers to avoid the rapid deterioration in room temperature. This makes the sample vulnerable to power outages and mechanical failures in the closing door.

Formalin-fixed frozen tissue is a combination of both previous approaches. In this case, the sample is fixed by formalin, but it is stored as fresh-frozen tissue without paraffin embedding. Heat-induced antigen retrieval protocols can be used to avoid metabolite loss (Groseclose et al., 2008), but formalin may reduce the ionization yield of amine-containing lipids, and generate $[M + HSO_4]-$ adducts (Vos et al., 2019). Using this sample preservation, it is possible to measure compounds in all mass ranges although with lower effectiveness than fresh-frozen tissues for the low mass range (Pietrowska et al., 2016).

TABLE 1 Summary of the principal effects of experimental steps in compound annotation and identification in MSI

Procedure	Effect in annotation/identification
Sample preservation	
FFPE tissue	• Severe contamination of the spectra. • Requires deparaffinization. • Suitable for protein and peptide detection.
Formalin-fixed fresh-frozen tissue	• Formalin may suppress the ionization of amine-containing lipids and introduce $[M + HSO_4]-$ adducts. • Suitable for sampling all families of compounds but less effective than fresh-frozen in the low mass range.
Fresh-frozen tissue	• No chemical changes in the tissue. • Risk of shattering and degradation during transport. • Suitable for sampling all families of compounds.
On-tissue sample treatment	
On-tissue enzymatic digestion	• Proteins are broken down into their peptides, which are easier to ionize and detect than intact proteins. • Peptides are used to elucidate possible proteins. • Enzymes hydrolyze proteins in specific bonds.
On-tissue chemical derivatization	• Added moieties increase ionization efficiency and the mass of targeted compounds.
Matrix application (only for MALDI sources)	
Organic matrices	• Introduce matrix signals in the low mass spectra region and matrix adducts. • Matrix selection influences which ionization polarity should be used.
Reactive matrices	• More selective measurement. • Act as derivatization agents.
Isotopically labeled matrices	• Controlled isotopic pattern used to annotate matrix signals and matrix-endogenous adducts.
Inorganic matrices	• Introduce fewer matrix signals. • In general, produce more fragmentation peaks. • Some inorganic matrix peaks can be used as calibrants.
Spraying matrix deposition	• Small amount of matrix used. • Solvent required for desorption of some molecules (such as proteins). • Higher risk of analyte delocalization.
Sublimation matrix deposition	• More matrix amount required. • More homogenous layer and less analyte delocalization.
Sputtering matrix deposition	• Requires inorganic material. • More homogenous layer and less analyte delocalization.
Stable isotope labeling	
SIL matrices	• Shift matrix signals to uncover endogenous signals. • Distinct isotopic pattern that helps annotation.
Ion source	
MALDI	• Requires matrix, which might contaminate the spectra. • Broad mass range (up to several kDa). • Common spatial resolution range from 100 to 10 μm. • Both ionization polarities (influences type of adducts). • MALDI-2 increases sensitivity. • t-MALDI increases routine spatial resolution to 1 μm and below.
DESI	• Minimal sample preparation (dopants may be added to the spray solvent). • Preference for detecting low molecular weight molecules. • Spatial resolution range from 200 to 20 μm. • Both ionization polarities (influences type of adducts).

(Continues)

TABLE 1 (Continued)

Procedure	Effect in annotation/identification
SIMS	• Minimal sample preparation. • Suitable for detecting low molecular weight molecules (hard ionization). • Highest spatial resolution (sub μm). • Both ionization polarities (influences type of adducts).
LA-ICP	• Used to map atomic composition. • Spatial resolution range from 200 to 10 μm.
Mass analyzer	
TOF	• Theoretically unlimited mass range. • Mass resolution increases as *m/z* increases. • Fastest scan rate.
FTICR	• Ultrahigh mass resolution for low-weight compounds. • Mass resolution decreases linearly as *m/z* increases.
Orbitrap	• Very-high mass resolution for low-weight compounds. • Mass resolution decreases linearly as *m/z* increases.
Combining MSI with other analytical techniques	
MS/MS	• Structural hypothesis using fragments of precursors. • Fragmentation patterns may be poor quality or precursor intensity is too low. • MS/MS libraries mostly contain $[M + H]^+$ fragmentation patterns.
LC-MS and LC-MS/MS	• Most common approach for identification. • Chromatographic separation allows better spectra interpretability. • Can use homogenization of the sample or other related biofluids. • LCM allows the LC-MS analysis of specific tissue regions. • Usually, Electrospray Ionization (ESI), may generate different adducts than MSI.
IMS	• CCS can be used to resolve isomeric species and get structural information.
Multimodal molecular imaging	• Vibrational spectroscopy can determine functional groups. • Fluorescence microscopy enables labeled imaging. • Registration of images is required.
Reference standards	
In-solution	• Easy sample preparation. • Fails to capture matrix effects, ion suppression effects, and endogenous adducts.
On-tissue	• Easy sample preparation. • Captures matrix effects, ion suppression effects, and endogenous adducts. • Low extraction efficiency. The standard only interacts with the surface.
Tissue mimetics	• Complex sample preparation. • Captures matrix effects, ion suppression effects, and endogenous adducts. • High extraction efficiency. • Loses spatial context.

3.1.2 | On-tissue enzymatic digestion of intact proteins

MSI analysis of intact proteins is usually restricted to those molecules below 25 kDa (although some matrix-assisted laser desorption/ionization [MALDI] matrices like ferulic acid can extend this range; Mainini et al., 2013), thus classical top-down proteomic strategies may not be efficient in MSI. Thus, on-tissue enzymatic digestion is included in most protein identification routines, which allow larger proteome coverage identification. This bottom-up approach is based on spraying or spotting enzymes (usually trypsin) over the tissue to cleave the proteins into their peptides, followed by an incubation step (Cillero-Pastor & Heeren, 2014; Diehl et al., 2015). Besides trypsin, other enzymes can be used to digest proteins, such as the enzyme peptide-N-glycosidase F for N-glycan profiling (Drake et al., 2018). Sequencing the detected peptides by common MS/MS approaches can help both identify and spatially locate

proteins directly on the tissue. Previous reviews on protein identification in MSI (Mascini & Heeren, 2012; Ryan et al., 2019) have covered this topic in depth.

Since the reactions for protein digestion are performed in solution, the tissues need to be covered by solvents containing the digestive enzymes, which can lead to the delocalization of the peptide products. Solvent-free solutions can avoid peptide delocalization, for instance by the use of plasmonic thermal decomposition/digestion (Zhou & Basile, 2017). This process uses continuos wave laser excitation and gold nanoparticles to decompose proteins at known locations (C-terminus of aspartic acid and at the N-terminus of cysteine) and since this is a dry technique, product peptides retain their original location on the tissue.

3.1.3 | On-tissue chemical derivatization

Some compounds are difficult to detect using MSI due to their low ionization efficiency, ion suppression, low concentration, and/or small molecular weight. Sample preparation steps (i.e., the proper matrix selection in MALDI MS or solvent selection in desorption electrospray ionization MS [DESI]-MS) might alleviate this concern. On-tissue chemical derivatization applies reagents over a tissue section to modify the chemical structure of specific compounds and enhance their detectability, by adding moieties with specific properties. For instance, adding a charged moiety often counteracts low ionization efficiency problems. Ion suppression due to low molecular weight can also be avoided by the reaction of the target compound and a derivatization molecule, which increases the analyte *m/z* ratio. All these mechanisms alter the detectability of specific compounds and therefore, the capacity of annotating and identifying them. Harkin et al. review concrete examples of these procedures (Harkin et al., 2021). For instance, pyrylium salts react selectively with primary amines in neurotransmitters, thus they can be incorporated into matrices (Shariatgorji et al., 2015) or synthesized as bromopyrylium to introduce a distinctive isotopic pattern only in targeted neurotransmitters (Shariatgorji et al., 2020). Additionally, the induced epoxidation of peracetic acid has been used to localize the C=C bonds in unsaturated fatty acids, allowing the discrimination of isomeric fatty acid (Zhang et al., 2021); and Giard's reagent P has been used to label *N*-glycans in FFPE tissue samples, increasing the sensitivity of the tissue samples characterization (Zhang et al., 2020).

Chemical reagents can also be used to promote a specific adduct of relevant biological molecules that are present in low concentration in tissues. For instance, *Duncan* et al. added silver ions to the solvent for nanospray desorption electrospray ionization MSI to enhance the ionization of prostaglandins as silver adducts, which allowed their monitoring directly on mice tissues (Duncan et al., 2018).

3.1.4 | Matrix selection and deposition in MALDI MSI

In MALDI MSI, matrices are compounds that assist the desorption/ionization of analytes from the tissue. Most common applications use small organic compounds as matrices that are either sprayed or sublimated over the tissue (Gemperline et al., 2014). MALDI matrix application techniques should ensure good homogeneity of the deposited layer and minimize in-tissue compound delocalization to get high-quality images.

The selection of appropriate matrices and optimization of the deposition method greatly affect the outcome of MALDI MSI analysis and the annotation and identification of analytes.

Matrices may introduce undesired effects that clutter the mass spectra and hamper compound annotation, such as matrix clusters, matrix adduct formation, and detector saturation. This is a particular issue in the low mass range where matrix-metabolite adducts can explain a considerable amount of non-annotated peaks (Janda et al., 2021). Lipidomics and metabolomics identification routines are very sensitive to the matrix method used (Fernández et al., 2011; Thomas et al., 2012).

The selection of the matrix will define the ionization polarity mode. For instance, MALDI matrices with an acidic group (like benzoic acid and cinnamic acid derivatives) are mostly used in positive ionization mode, while matrices that are basic and contain amino functions tend to be used in negative ionization mode. The ionization mode will favor the detection of specific compounds, for example, lipids with a polar headgroup like phosphatidylcholines will be detected in positive mode, while glycerophosphoinositol will have better ionization yield in negative mode (Leopold et al., 2018). To increase the coverage of the lipidome, several research groups opt for the use of matrices and acquisition modes that allow dual polarity MALDI MSI analysis on the same sample (Huang et al., 2020; Kaya et al., 2018; Li et al., 2019).

Developing new matrices is a hot research field in MSI. While classical first-generation matrices like alpha-Cyano-4-hydroxycinnamic acid and 2,5-Dihydroxybenzoic acid are still widely used, the design of second-generation and reactive matrices (simultaneously a derivatization reagent and a matrix) allow the selective desorption/ionization of

specific analytes. The analytes of interest are detected with higher signal-to-noise ratios and sometimes present specific spectra features (such as a distinctive isotopic pattern) that facilitate their annotation and identification. Reactive matrices can also aid discriminating between isomeric compounds, such as the reactive matrix benzophenone, that serves both as ionization promoter and as derivatization reagent to selectively functionalize unsaturated phospholipids (Wäldchen et al., 2019). The reviews by Zhou et al. and Calvano et al. provide an excellent reference on selective matrices for MSI metabolomics and lipidomics (Calvano et al., 2018; Zhou et al., 2021).

On the other hand, inorganic nanoparticles (NPs) (of gold and silver, among others), as well as some metal-oxides (TiO_2, CeO_2, etc.), have been proposed as an alternative to organic matrices for the analysis of small molecules by MSI (Abdelhamid, 2018; Basu et al., 2019). They often produce fewer matrix clusters and adducts, leading to a cleaner background spectrum. Additionally, their distinctive carbon-free isotopic pattern and easily identifiable peaks can serve as internal calibrants during data processing (Nizioł & Ruman, 2013; Ràfols & Castillo, 2018; Ràfols et al., 2018).

Matrix deposition is one of the most important sample preparation steps toward the production of high-quality ion images. Researchers use different techniques to apply matrices onto the target tissue, including spray (Khatib-Shahidi et al., 2006; Norris et al., 2007) and sublimation (Hankin et al., 2007; Thomas et al., 2012) for organic matrices, and sputtering for NPs (Dufresne et al., 2013; Ràfols et al., 2018). The spray method is based on applying the matrix solution into the tissue section manually (DeKeyser et al., 2007; Ye et al., 2013) or using automated spray devices allowing controllable solvent flow rate and matrix layers number (Gemperline et al., 2014; Mounfield & Garrett, 2012; Phan et al., 2016). Sublimation is a dry deposition technique (the transition of one chemical substance from the solid phase to the gas phase without passing through the intermediate liquid phase), in which matrices are sublimated and deposited under reduced pressure and specific elevated temperature parameters, leading to the deposition of dry matrix layer on tissue target (Hankin et al., 2007; Nakamura et al., 2017). However, sublimation alone is not sufficient for the ionization of some compound species, such as proteins, therefore a rehydration or recrystallization step is needed to promote the integration of these molecules with the matrix crystals (Yang & Caprioli, 2011).

Sputtering is a thin film deposition process where inorganic NPs or metal-oxide targets (such as gold or silver) are bombarded with high-energy ions in a vacuum chamber resulting in the condensation of the target atoms on the substrate tissue section as thin layers (Hansen et al., 2019; Ogrinc Potočnik et al., 2014).

3.1.5 | Stable isotope labeling (SIL)

SIL consists of the synthesis of compounds containing atoms with artificial isotopic abundances highly dissimilar to the ones that occur in nature. Common isotope labels include ^{13}C, ^{15}N, and deuterium (^{2}H). This technique has many applications in several aspects of MSI (Grey et al., 2021) such as tracing of drugs and metabolites (Eckelmann et al., 2018; Ellis et al., 2021). Additionally, the labeled compounds introduced in the sample can be used as internal standards to normalize signal intensity (Barry et al., 2019; Chumbley et al., 2016) and provide quantitative results (Grey et al., 2019).

For annotation, one of the most relevant applications is SIL MALDI matrices. By isotopically labeling the matrix, their background signals can be shifted and uncover relevant endogenous signals. Additionally, their distinct isotopic pattern can be exploited to develop more robust annotation tools. As an example, Shariatgorji et al. (2012) managed to shift the matrix peaks by using deuterated CHCA to uncover and annotate several neurotransmitters.

3.2 | MSI image acquisition

Mass spectrometers intrinsically affect the annotation and identification procedures, as they determine which species of ions will be generated in the ion source, and the *m/z* resolving power and accuracy. The parts of the mass spectrometer that affect the annotation/identification process are the ion source, responsible for the desorption and ionization of the molecules, and the mass analyzer, responsible for the determination and counting of the *m/z* ratio of the ions. Figure 1 shows a broad comparison between the main ion sources and mass analyzers.

3.2.1 | Ion source

The ion source induces the desorption of the analytes from the tissue, and the ionization of compounds that will be transferred into the mass analyzer. Depending on the polarity of the electrical field applied in the ion source, the ions formed will be positive (usually protonated adducts and adducts with cations, such as Na^+ and K^+) or negative (like deprotonated adducts and adducts with anions, such as Cl^-). The different technologies result in differences in the mass range

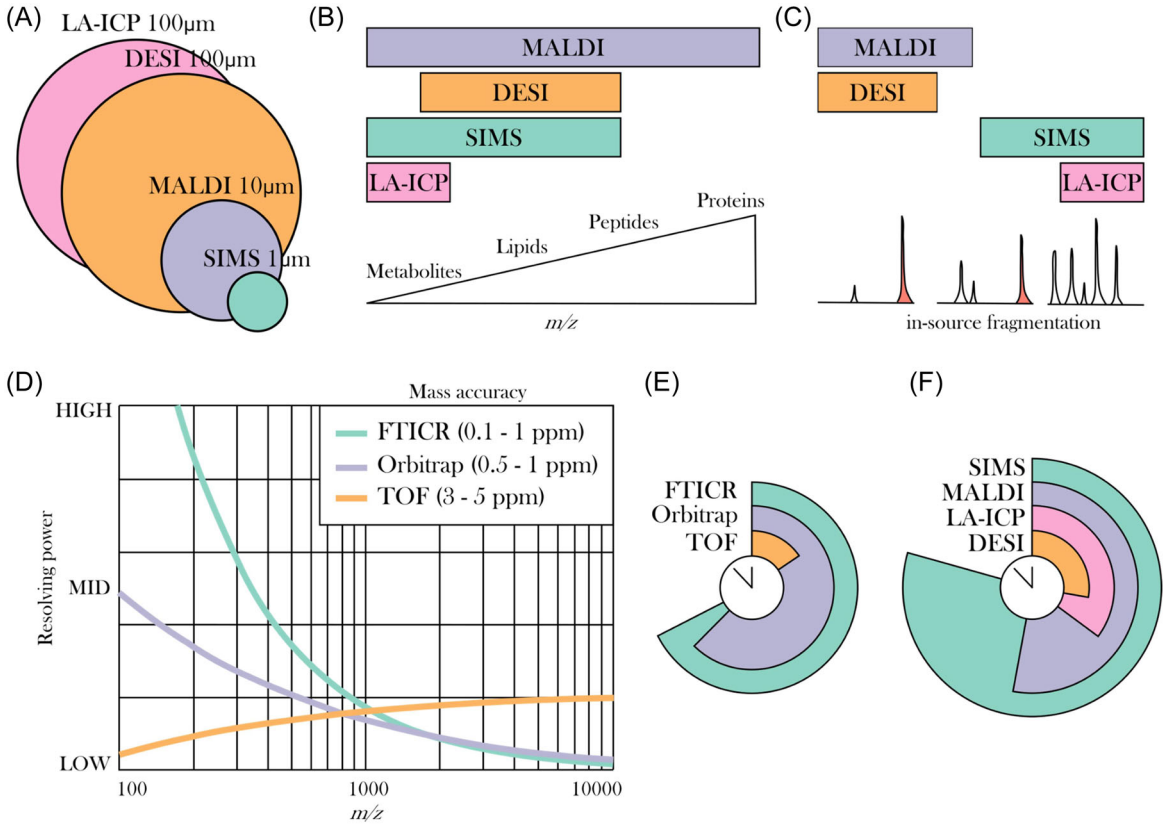


FIGURE 1 General comparison of the most widely used ion sources and mass analyzers for Mass spectrometry imaging. (A) spatial resolution, (B) mass range, and (C) in-source fragmentation of the four most common ion sources. (D) resolving power and mass accuracy and (E) acquisition time of the three most common mass analyzers. (F) acquisition time of the four most common ion sources. Adapted with permission from Evers et al. (2019) (A,B,F), and Ayet San Andrés et al. (2019), Zubarev and Makarov (2013) (D). Copyright 2022 American Chemical Society. CC-BY license https://creativecommons.org/licenses/by/4.0/ [Color figure can be viewed at wileyonlinelibrary.com]

analyzed, the number of charges of the produced ions, the amount of in-source fragments generated, and the sensitivity to detect low concentration compounds. Spatial resolution and sensitivity are related concepts, as increasing the spatial resolution results in decreasing the ablated area and therefore, reduces the sensitivity. In MSI, the most used ion sources are MALDI, DESI, secondary ion mass spectrometry (SIMS) and laser ablation inductively coupled plasma (LA-ICP).

MALDI sources ionize the sample using a pulsating laser (usually UV or IR) inside a vacuum or low-pressure chamber with the assistance of the previous matrix deposition. The laser strikes the sample and generates a plume of charged ions that are directed to the mass analyzer. MALDI sources tend to produce low fragmentation and singly charged ions (Jaskolla & Karas, 2011; Karas et al., 2000), which enable the ionization of metabolites (Stoeckli et al., 2007), lipids (Züllig & Köfeler, 2021), peptides (Phillips et al., 2019) and proteins (Piga et al., 2019), and usually achieve spatial resolutions in the range of 100–10 µm and close to 1 µm with specific setups (Hansen & Lee, 2017; Kompauer et al., 2017; Wäldchen et al., 2020). In recent years, enhanced versions of MALDI sources have been proposed, like MALDI-2 (Heijs et al., 2020; Soltwisch et al., 2015), which increases the sensitivity of the MALDI source by adding a second postionization laser that ionizes the neutral molecules in the ion plume; transmission MALDI (t-MALDI) (Steven et al., 2019; Trimpin et al., 2009; Zavalin et al., 2012, 2015), which increases the lateral resolution up to 1 µm and below by changing the laser focus geometry; and more recently t-MALDI-2 (Bien et al., 2021; Dreisewerd et al., 2022; Niehaus et al., 2019), which combines the benefits of both improved designs.

DESI sources produce ions at atmospheric pressure conditions directing a spray of charged microdroplets directly into the tissue. DESI sources require minimal sample preparation. They are commonly used to analyze small molecules and lipids, but bigger compounds like peptides and proteins can also be analyzed (Towers et al., 2018), although most solvents used with DESI denature proteins, affecting the three-dimensional structure (Hale & Cooper, 2021). Typically, DESI sources

achieve spatial resolutions in the range of 200–20 μm (Claude et al., 2017; Ifa et al., 2007; Nguyen et al., 2018; Towers et al., 2018; Zhang et al., 2020) and are known to produce little fragmentation and singly charged ions (Towers et al., 2018).

SIMS sources bombard samples using an ion beam, ionizing molecules from the sample surface and ejecting them into the vacuum environment but, due to the high energy of the beam, SIMS sources easily cause the fragmentation of the molecular ions (Yoon & Lee, 2018). Currently, SIMS sources provide the greatest spatial resolution for MSI, reaching the nanometer scale (Gamble & Anderton, 2016), but have less sensitivity, as the area ablated is lower than other technologies. Applications of SIMS sources are principally focused on small metabolites and lipids (Touboul & Brunelle, 2016) and require minimal sample preparation.

LA-ICP sources use an inductively heated plasma to atomize molecules ablated from a specific region, generating atomic composition maps over the sample. LA-ICP is generally used to track metals in biological sections with a spatial resolution between 200 and 10 μm (Becker et al., 2011, 2012; Pornwilard et al., 2013; Sabine Becker, 2013). In terms of fragmentation, LA-ICP fragments all the compounds in the sample to their atomic composition, resulting in null preservation of precursor ions.

3.2.2 | Mass analyzer

The mass analyzer detects the ions generated by the source, determines their mass-to-charge ratio, and composes the spectrum at each sample position or pixel of the image. There are three parameters that influence the identification of compounds for each mass analyzer: (1) mass range, the lowest and highest *m/z* that the mass analyzer can detect; (2) mass accuracy, the difference between the measured *m/z* of an ion and the real *m/z* (usually specified in ppm); and (3) mass resolution, the ability to distinguish ions separated by small *m/z* values, often defined as the *m/z* of a peak divided by the peak width at 10% or 50% of peak height. The most common mass analyzers in MSI systems are time-of-flight (TOF), Fourier-transform ion cyclotron resonance (FTICR), and Orbitrap.

TOF mass analyzers are vacuum tubes in which ions travel through an electric field to the detector. The longer the tube, the higher the mass resolution of the spectra, as the ions have more time to gain distance between them during the flight. Despite this, TOF mass analyzers tend to have lower mass resolution compared to other mass analyzers used in MSI, as enhancing it implies an increase in the physical size of the whole MSI system and in the sampling time. With reflectron set-ups, the mass resolution can be increased, but it remains lower than other analyzers. Moreover, TOFs are very susceptible to temperature changes, as the metal tube may suffer expansions and contractions that affect the mass accuracy of sampled ions. On the other hand, TOF analyzers do not have a theoretical upper *m/z* detection limit like other mass analyzers (Xian et al., 2012), and their mass resolution increases along the mass range. TOF mass analyzers are extensively used with MALDI ion sources to image almost any kind of compounds, with a preference for compounds in the high mass range like peptides and proteins, with a typical upper limit of *m/z* 30,000 (Spengler, 2015). Common set-ups of TOF mass analyzers are MALDI-TOF, MALDI-TOF/TOF, MALDI-Q-TOF, and TOF-SIMS.

FTICR mass analyzers use a magnetic field to resonate the ions into cyclotron orbits and transduce the orbiting frequencies into *m/z* using the Fourier Transform. These mass analyzers are built around powerful magnets; the stronger the magnetic field, the greater the mass resolution, reaching values of up to 1,600,000 at *m/z* 400 for a 21T magnet (Bowman et al., 2020) with mass accuracies below 1 ppm. FTICR mass analyzers are used to analyze all families of compounds, but preferably not higher than *m/z* 3000, as the mass resolution decreases as the *m/z* ratio increases (Almeida et al., 2015) and the magnetic field and sampling time required to detect these ions are high. Still, there are examples of high mass protein MSI investigations up to *m/z* 30,000 using a 15T FTICR mass analyzer with a mass accuracy below 10 ppm and transients close to 4 s per pixel (Dilillo et al., 2017). Common set-ups of FTICR mass analyzers are MALDI-FTICR and DESI-FTICR.

Orbitrap mass analyzers use electrically charged ion trap cells to excite the ions into orbits. The longitudinal movement of the orbits contains the information of the cyclotron frequencies of each ion, which can be converted to mass using the Fourier transform. Orbitraps achieve high mass resolution values by increasing the electric field. With Orbitraps it is possible to analyze a wide range of compounds but, as FTICR, high mass compounds are typically excluded as the mass resolution decreases by the square root of the *m/z* ratio and require long sampling times and strong fields to compensate for this (Bielow et al., 2017). Common set-ups of Orbitrap mass analyzers are DESI-Orbitrap and MALDI-Orbitrap.

All mass analyzers are often calibrated before acquisition to obtain accurate *m/z* measurements. The calibration consists in tuning the electronic parameters of the instrument to modify the *m/z* axis according to different calibration curves built upon measured calibration standards. The calibration standards are liquid

mixtures of highly purified molecules designed for positive and/or negative mode in a particular mass range. Different strategies to obtain calibration curves are used depending on the time of the standard application and the mass analyzers (Smith et al., 2012). In terms of standard application, internal calibration consists of mixing calibration standards directly with the sample of interest, while external calibration consists of measuring the standards alone (Muddiman & Oberg, 2005). When FT-ICR or Orbitrap instruments are used, it is convenient to use an abundance dependent calibration curve, as different amount of ions inside the ICR cell produce varying frequency shifts, resulting in different mass errors at each sampling point (Easterling et al., 1999; Gorshkov et al., 2010; Zhang et al., 2005).

3.3 | Combinations of MSI with other analytical techniques (Levels 2–3 identification)

To ensure high levels of confidence in molecular identification with MSI, a common strategy is to examine the tissue with additional or orthogonal techniques (those based on fundamentally different principles). LC-MS and tandem mass spectrometry (MS/MS) are the most used confirmatory techniques. Recently, ion mobility has been included in commercial MSI instruments to provide an additional dimension for metabolite analysis and resolve isomers (Łącki et al., 2021; Meier et al., 2015, 2020). Finally, the combination of different imaging techniques coupled to MSI has been used to improve the identification process. Multimodal imaging combines non-destructive orthogonal analysis like immunohistochemistry, immunofluorescence, or vibrational spectroscopy imaging techniques with MSI (Iakab et al., 2021; Tuck et al., 2021).

3.3.1 | MS/MS

MS/MS uses a combination of ion traps, mass analyzers, and fragmentation chambers to measure fragments of molecules and reveal their structure. The typical setup is two consecutive mass spectrometers separated by a fragmentation chamber. The first mass spectrometer is in charge of recording the ionization product of an ion source that keeps the precursor compounds with low fragmentation. Later, some of the precursor ions are directed to a collision chamber to achieve a controlled fragmentation. The resulting fragments are registered in a second mass analyzer to obtain the fragmentation spectra of all the selected precursors. By knowing the precursor *m/z* value and examining the fragmentation spectrum, it is possible to provide hypotheses about the structure of the compound and hence its identification.

In MSI, MS/MS analysis can be performed in some instruments either by sampling consecutive slides in MS/MS mode (Dueñas et al., 2017) or adjacent regions in the same slide (Zhan et al., 2021), which can be a problem when analyzing very localized compounds or limited sample material. Common set-ups are based on TOF/TOF and Q-TOF devices, commonly used for top-down proteomics (Alam et al., 2012; Xu et al., 2019; Ye et al., 2014).

To overcome these limitations, new methods have been investigated in recent years. Multiplex MSI has achieved to overlap scans of MS and MS/MS in the same location using a spiral pattern and proved successful in 10 μm high-spatial-resolution imaging of maize leaf cross-sections in both the high and low mass ranges for a variety of metabolites (Hansen & Lee, 2017; Perdian & Lee, 2010; Yagnik et al., 2013). Ellis et al. developed an automatic structural identification workflow consisting of parallel acquisition of a MALDI-Orbitrap instrument with an ion trap (IT)-MS/MS (Ellis et al., 2018). Lanekoff et al. coupled a nano-DESI source with a high-resolution Q-Exactive Orbitrap and a higher-energy collision-induced dissociation (HCD) cell to identify and image isobaric and isomeric species combining the MSI and the MS/MS data (Lanekoff et al., 2013). Finally, Fu et al. were able to analyze and image by tandem MS the molecular products of natural biosynthesis of rubrynolide and rubrenolide in Amazonian trees using a TOF-SIMS and a triple ion focusing time-of-flight (TRIFT) analyzer with a precursor selection window of a monoisotopic ion, which allow the parallel and lossless collection of MS and MS/MS data (Fu et al., 2018). Tandem MS on tissue can help discriminating isomers of lipids due to their differential fragmentation. For instance, (Takeo et al., 2019) used tandem MS (MS^3) to discriminate between structural isomers of some steroids, after applying on-tissue chemical derivatization techniques to enhance their ionization efficiency.

Despite all the efforts, MS/MS is rarely used with MSI data as many commercial instruments still do not include this option. Moreover, the concentration of precursors is limited to the area covered by the scans, which might be low for some compounds (unless strategies to promote their ionization are considered).

3.3.2 | LC-MS

LC-MS incorporates chromatographic separation before the mass analyzer. RT allows differentiation of the

compounds based on criteria other than *m/z*, like polarity or compound size. Most LC-MS systems use tandem MS and can provide fragmentation information on the analytes.

The combination of LC-MS with MSI is one of the most common approaches used to identify and spatially visualize a compound in all kinds of metabolomics and lipidomics experiments (Garate et al., 2020). The identification workflow usually consists of homogenizing some of the tissue samples to identify as many compounds as possible with the LC-MS instrument (Baijnath et al., 2016; Ntshangase et al., 2019; Shobo et al., 2016). Later, the identified compounds are searched in the MSI spectra by exact mass matching.

Other approaches combine LC-MS with laser-capture microdissection (LCM), which allows the isolation and compound profiling of specific cells or tissue regions of interest (ROIs) determined by MSI (Marialaura Dewez et al., 2019; Dilillo et al., 2017). This approach ensures that the LC-MS identifications come from the same region in the tissue that was mapped by MSI.

3.3.3 | Ion mobility spectrometry (IMS)

IMS is a technology that separates ions according to their size, shape, and weight by directing and colliding them into a chamber filled with an inert gas. The collision cross-section (CCS) value is computed from the time each ion takes to reach the end of the chamber. In combination with MS, IMS can be used as an additional dimension of information to resolve isomeric species, improve selectivity, and get structural information of compounds, including metabolites (Lapthorn et al., 2013). Sans et al. reviewed an extensive amount of applications and advances combining MSI and IMS for biological applications (Sans et al., 2018).

3.3.4 | Multimodal molecular imaging

Other molecular imaging techniques can provide the orthogonal chemical information needed to provide structural identification of *m/z* features (Porta Siegel et al., 2018).

Vibrational spectroscopy imaging techniques (i.e., Raman and FT-IR) measure the energy scattering and absorption of different lasers to determine functional groups and other chemical features (Harrison & Berry, 2017). This structural information is rarely enough to fully resolve isomers, but it can be used to discard candidates and achieve Level 3 annotation. As an example, Lasch and Noda (2017) applied Raman, FT-IR, and MSI to study the composition of the hamster brain. They could identify and spatially locate several lipids by the spectral correlation between Raman bands (for instance, bands 548 and 703 cm^{-1} for cholesterol) and *m/z* features (*m/z* 369.30 for $[Cholesterol–H_2O + H]^+$).

Fluorescence microscopy techniques enable imaging of specific compounds by labeling them with fluorescent probes (Lichtman & Conchello, 2005). Cyclic or multiplexed immunofluorescence images the same sample with dozens of different fluorescent probes (Lin et al., 2016). Highly selective fluorescent probes (Dong et al., 2020; Li et al., 2011; Uslu et al., 2017) can target specific isomers and enable Levels 3–2 annotation. For instance, Fuchs et al. (2018) monitored the biodistribution of the anticancer drug sunitinib and its metabolites in rabbit liver tissue using fluorescence to measure the total amount of the drug, and MSI to characterize in situ the presence of its metabolites.

3.4 | Validation against reference standards in MSI (Level 1 identification)

According to the system for reporting identification confidence in MSI (Section 2.2), to achieve Level 1 identification (highest level of confidence), the experimental data (MSI and orthogonal technique of choice) has to be matched against a reference standard. One common strategy in MSI experiments is to homogenize the tissue, spike it with the reference standard of the compound of interest, and measure it with LC-MS/MS (Baijnath et al., 2016; Ntshangase et al., 2019; Shobo et al., 2016). Using LCM, the tissue homogenates can be obtained from specific tissue ROIs selected by MSI (Marialaura Dewez et al., 2019; Dilillo et al., 2017). Nevertheless, even when using LCM, homogenizing the tissue leads to the loss of the spatial information provided by MSI. Additionally, due to differences in their ionization, LC-MS/MS and MSI data may not be directly comparable (i.e., the analytes of interest may form different adducts in each system, etc.). An alternative technique to LCM is liquid extraction surface analysis mass spectrometry (LESA-MS), which combines microextraction in the liquid phase directly from the tissue with nano-electrospray MS. While it is considered a low spatial resolution technique (generally aprox. 1 mm), it is a valuable complement to high spatial resolution techniques (like MALDI-MSI), since it provides additional information for identification of compounds in situ without the need to homogenate the sample. This technique has been successfully applied for instance to monitor drugs and drug metabolism on mice organs (Eikel et al., 2011; Swales et al., 2015).

Full confirmation of MSI identifications requires strategies to measure reference standards directly in MSI. Most of the developments in this area have been conducted for the study of synthetic drugs and their metabolites in situ (Buck et al., 2015; Groseclose et al., 2015) but they are largely applicable to endogenous neurotransmitters (Shariatgorji et al., 2014), metabolites (Pirman et al., 2013), lipids (Jadoul et al., 2015), and peptides (Zhang et al., 2013). In general there are three strategies (Rzagalinski & Volmer, 2017; Unsihuay et al., 2021): (1) "in-solution" (2) "on/under tissue" and (3) "mimetic tissue."

The "in-solution" strategy is the most straightforward of the three, as the standard is spotted directly on the substrate next to the sample. This method will inform about isotopic patterns, general adducts, matrix adducts, and in-source fragments that can be formed with the analyte of interest during the MSI experiment. However, it fails to capture endogenous adduct formation and ion suppression effects. As an example, the in-solution strategy was used for identifying the drug Erlotinib and its metabolites in rat tissue sections (Signor et al., 2007).

The "on/under tissue" strategy alleviates these limitations by spotting the standard beneath or on top of the tissue. Normally, this is performed on a control tissue, preferably a consecutive slice. If allowed by the application (i.e., in synthetic drug applications), the control tissue should be blank and not contain the endogenous compound to be compared to the reference standard. As a variation of this approach, some studies apply the standard mixed with the MALDI matrix. As an example, the "on-tissue" approach has been used to identify the drug paclitaxel in the study of pleural tumors (Giordano et al., 2016), glutathione in ovarian tissue (Nazari et al., 2018), and raclopride and SCH 23390 in rat brain tissue (Goodwin et al., 2011).

Finally, the "mimetic tissue" approach relies on homogenizing the tissue and spiking it with the standard. This mixture is then deposited on the MSI slice and treated with the same sample preparation protocol. This approach provides a more realistic scenario on how the analyte behaves during the MSI experiment, as the standard is fully mixed within the sample. One drawback is that it fails to capture differences in matrix and suppression effects across anatomical regions. The mimetic tissue approach has been successfully used for identifying the drugs lapatinib and nevirapine in rat liver (Groseclose & Castellino, 2013), GSH in human ocular lens tissue (Grey et al., 2019), and clozapine and norclozapine in rat liver (Barry et al., 2019).

4 | BIOINFORMATICS STRATEGIES FOR ANNOTATION AND IDENTIFICATION IN MSI

In this section, we discuss automated data processing strategies for annotation and identification in MSI. We first start by discussing the importance of preprocessing to ensure robust annotation and identification. Later, we provide a wide picture of the basic principles in the development of software-based annotation and identification. We close the section with a comprehensive comparison of twelve software tools developed in the last 5 years.

4.1 | Data preprocessing

Good quality MSI data is crucial to conduct successful molecular annotation and identification (Norris et al., 2007). As stated in the previous section, careful analytical design is key, as it will set the boundaries of what is possible in compound identification. But even when the analytical procedure is carefully designed and executed, variability due to experimental factors can worsen data quality. Chemical noise and variations in the intensity and exact mass of each MS feature are some of the examples of unwanted experimental variability. Additionally, when dealing with large samples and high spatial resolution, MS intensities and *m/z* values can drift during the long acquisition (Ràfols et al., 2018). Proper data preprocessing mitigates these negative effects and enhances the chances of correct identification.

The typical preprocessing workflow includes the following steps: baseline correction, noise reduction, spectral alignment, normalization, peak picking, and binning (Ràfols et al., 2018). Depending on the experiment, some steps may be performed in a different order or even be omitted. The resulting processed data can come in two forms: (1) profile data retains the continuous shape of the spectra, as no peak picking is performed, and (2) centroid data only retains certain features of each peak (commonly the *m/z* and maximum intensity value) after peak picking.

Calibration (a form of spectral alignment) is the most relevant step for annotation and identification, as it increases the mass accuracy of the measured *m/z*. In calibration, a list of known *m/z* values is used to compute a warping function that minimizes the *m/z* error in the MSI data set. The calibration *m/z* values can come from reference standards spotted on the plate (phosphorus red) (Paine et al., 2019), the matrix or ionization promoter (Ràfols & Castillo, 2018; Ràfols et al., 2018), or well-characterized endogenous compounds (He et al., 2019). Additionally, label-free alignment can further improve data quality. In this case, a reference

spectrum from within the sample is used to minimize the *m/z* errors between pixels.

All MSI instrumentation vendors provide in-house software capable of performing to some extent this preprocessing pipeline. SCiLS (Trede et al., 2012) by Bruker is one of the most widely used commercial solutions. Several open-access alternatives such as MSIReader (Bokhart et al., 2018; Robichaud et al., 2013), CARDINAL (Bemis et al., 2015), rMSIproc (Ràfols et al., 2020), and MALDIQuant (Gibb & Strimmer, 2012) have gained importance over recent years.

4.2 | Basic software-related principles in annotation and identification of MSI

Figure 2 shows the general workflow of annotation and identification software tools in MSI. Each of the steps increases the level of confidence and relies on different experimental data and libraries.

There are various basic concepts to consider while designing or choosing an annotation tool for MSI data. How input data represent each *m/z* feature, the direction of the flow of information between data and libraries, how to match the information in the libraries, and how to use the annotation or share them. The following section comments on several of these topics.

4.2.1 | Working with profile versus centroided data

Molecular annotation and identification can either be performed on profile or centroided spectra. Profile spectra provide richer information: (1) they keep potentially relevant small and noisy peaks, (2) they

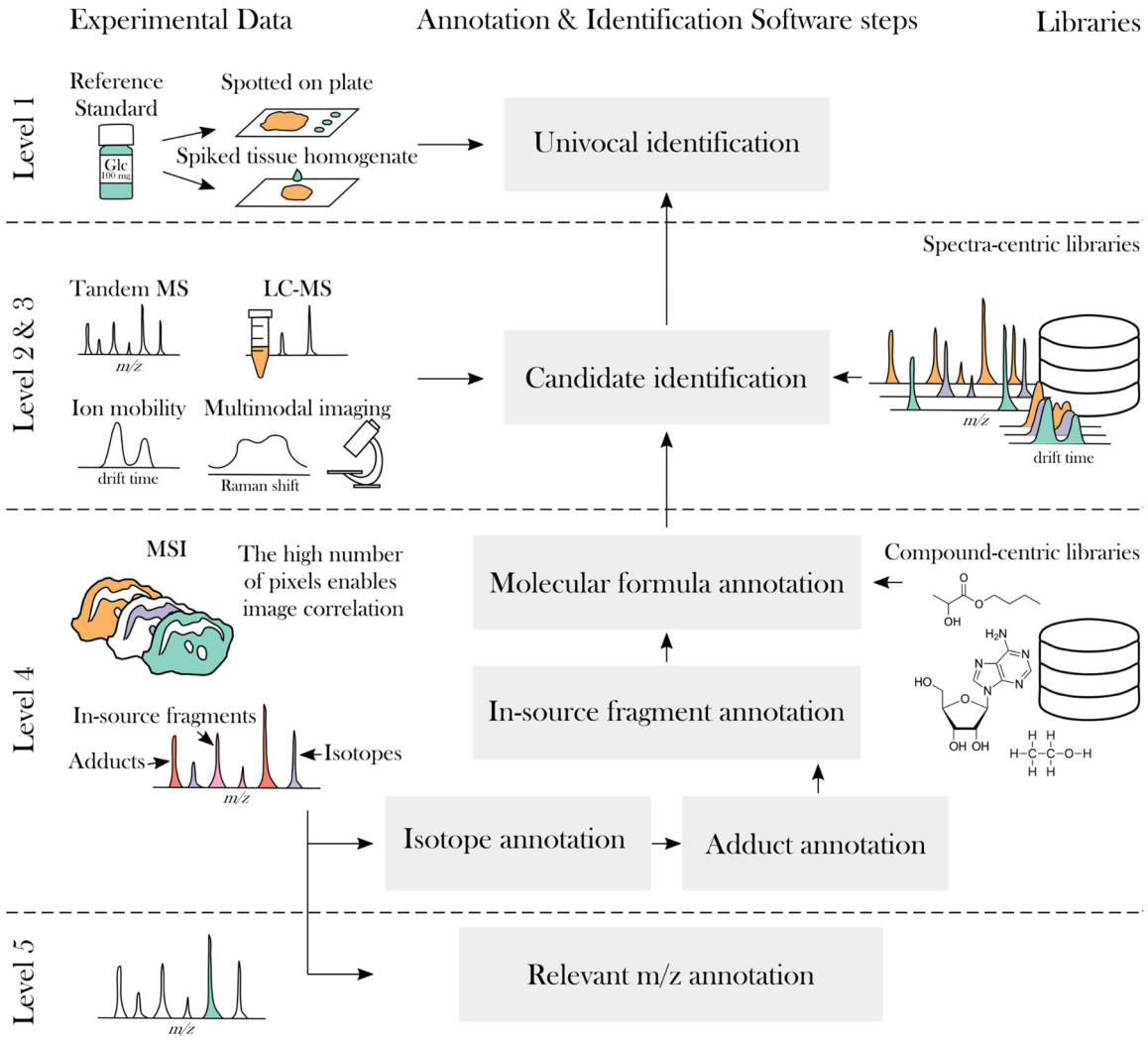


FIGURE 2 General steps in software annotation and identification in mass spectrometry imaging experiments. [Color figure can be viewed at wileyonlinelibrary.com]

retain peak shape, and (3) they enable overlapped peaks to be recognized and eventually deconvoluted (Polanska et al., 2012). The main problem with data in profile mode is the higher computational load, which is oftentimes prohibitive in terms of memory and CPU time requirements. For this reason, most annotation and identification software tools work on centroided spectra. Centroided mode retains only the most relevant features of a peak (*m/z* and maximum intensity or peak area) to dramatically reduce the size of the data set, which leads to relaxed memory and CPU time requirements.

4.2.2 | Library-centric versus feature-centric strategies

There are two general approaches to determine chemical composition in MSI: library-centric or feature-centric. These approaches are applicable to both annotation (using only exact mass matching) and identification (combining MSI with orthogonal techniques and reference standards).

Library-centric approaches match library information to experimental data. For each candidate compound in the library, the algorithm will generate an in silico theoretical spectrum (with isotopes, adducts, or ion fragments) using the molecular formula, and will determine its presence in the sample by matching them against the experimental spectra (usually the mean spectra) (Alexandrov & Bartels, 2013; Novák et al., 2020; Tortorella et al., 2020). These approaches tend to be computationally consuming in terms of time and memory, as the algorithm will try to fit all the compounds in the libraries. Besides, the results are limited to the compounds existing in the libraries (if the compound does not exist in the library, the associated *m/z* signals will not be annotated).

Feature-centric approaches look for patterns in the data (adducts, isotopes, or fragments) to create several networks of related MS signals. In general, this strategy gathers information from the data and tries to construct isotopic patterns of unknown compounds taking into account the spatial correlation, the intensity profile, and the mass error between features (Bond et al., 2017; Janda et al., 2021; Sementé et al., 2021). This approach also includes using the Kendrick mass defect (KMD) to assign families of compounds (Kune et al., 2019). At the end of these procedures, some *m/z* features are confidently annotated as monoisotopic ion candidates, taking into account all the information gathered, and can be searched against libraries of compounds. These approaches tend to be faster to run but require extra steps to assign compounds to the *m/z* features. Additionally, they are less generalizable, as they make certain assumptions about the data that might be specific only to a certain family of compounds, like the shape of the isotopic pattern due to the elemental composition; or about the experimental procedure, like searching for specific adducts or labeled moieties.

4.2.3 | Isotopic pattern generation

Tools that follow the library-centric approach tend to generate the in silico pattern of the compounds in the libraries to compare with the spectra. This can be achieved using in-house algorithms or with enviPat (Loos et al., 2015), an R-package that generates the profile spectrum and the centroids of sum formulas simulating different resolving power; and Rdispo (Böcker et al., 2006), an R-package that generates isotopic patterns and elucidates molecular formulas for a given mass.

4.2.4 | Match scores

Regardless of the approach followed (library-centric or feature-centric) all software tools rely on several match scores to determine the fitness of each hit. The two main metrics are (1) spectral similarity (to compare experimental data against theoretical isotopic ratios, fragmentation spectra, or CCS) and (2) spatial similarity (to determine if isotopes, adducts, and fragments are colocalized).

The most widely used spectral similarity metrics are Pearson's correlation (McDonnell et al., 2008), and cosine similarity. Smets et al. proposed histogram matching as an alternative (Smets et al., 2019). Recently, a new metric inspired by natural language processing algorithms (Spec2Vec) (Huber et al., 2021) has been proposed and compared with cosine similarity, obtaining better results in library matching fragmented molecules.

Spatial similarity can be determined using Pearson's/Spearman's correlation, cosine similarity, hypergeometric similarity measure (Kaddi et al., 2011) or Structural Similarity Index (SSIM) (Ekelöf et al., 2018). Ovchinnikova and Stuart (2020) used 2210 ion images ranked by similarity by 42 MSI experts to quantitatively compare several spatial similarity metrics. One of the machine learning models (Pi-Model) included in their software ColocML obtained the highest performance (0.797 correlation to the gold standard) closely followed by cosine similarity (0.794) and Pearson's correlation (0.788).

The match score can be further refined using other metrics such as mass error (Sementé et al., 2021) or spatial chaos (Palmer et al., 2016; Tortorella et al., 2020). Additionally, the notion of false discovery rate (FDR) has been used to estimate the confidence of annotations using a target-decoy approach in which the resulting molecular formulas are compared with impossible adduct formations (Guo et al., 2021; Palmer et al., 2016).

There is no consensus on the relationship between different scores or how to unify them. Typically, when multiple scores are available, each score is scaled to fit a range of 0–1 and the product of all scores is taken as a single metric (Baquer et al., 2020; Palmer et al., 2016; Sementé et al., 2021; Tortorella et al., 2020).

4.2.5 | Library matching

Both in annotation and identification, it is crucial to compare the MS signals obtained in the experiment to a list of known compounds or references. To obtain the highest degree of confidence in annotation, the experimental data must be matched against a reference standard. Nevertheless, reference standards are not always available or compatible with the experimental workflow of choice. Reference standard matching is particularly challenging in untargeted studies, where tens or even hundreds of compounds are analyzed at the same time. To aid compound annotation in these cases, several libraries compile and index thousands of previous experimental MS and MS/MS measurements of standards from laboratories around the world. Libraries offer a reliable, automatable, and easy-to-use substitute to real standards. They can be considered compound-centric or spectra-centric, depending on their content.

Compound-centric (or metadata-centric) libraries such as HMDB (Wishart et al., 2018), ChEBI (Hastings et al., 2016), PubChem (Kim et al., 2019) include information such as the monoisotopic mass of the compound, molecular formula, SMILES, Inchy Key, molecular structure, and in some cases, other relevant metadata such as compound origin (plant, animal, bacterial, etc.) or even metabolic function. This first type of library is mainly useful for exact mass matching.

Spectra-centric libraries store MS and MS/MS spectra of thousands of compounds. Identification is obtained by matching experimental data to the spectra in the library. Several libraries are available for MS/MS, some examples include METLIN (Smith et al., 2005), NIST (Lemmon et al., 2010), and MassBank (Horai et al., 2010). For ion-mobility, the most ambitious projects include the Online Collision Cross Section Compendium (Picache et al., 2019) and the AllCCS atlas (Zhou et al., 2020). Most databases in this category have been developed with traditional MS technologies in mind (mainly LC-MS and GC-MS) and almost exclusively include fragments of the $[M + H]^+$ and $[M - H]^-$ adducts. There is a lack of databases of experimental spectra acquired by MSI.

4.2.6 | In silico libraries

With the advent of machine learning and cheminformatics techniques, in silico libraries have emerged. They typically generalize from experimental data of pure compounds and rely on advanced algorithms to generate relevant information of unknown or unmeasured compounds. This can include information such as monoisotopic mass, molecular formula, chemical structure and even MS and MS/MS spectra. Some of these in silico tools for tandem MS include LipidBlast (Kind et al., 2013), Sirius (Dührkop et al., 2019), MetFrag (Ruttkies et al., 2016), and CFM-ID (Djoumbou-Feunang et al., 2019). For ion mobility, AllCCS (Zhou et al., 2020) uses machine learning to predict CCS values from SMILES. These tools should be carefully evaluated and used in a case-by-case scenario. Blindly trusting them in untargeted studies can lead to incorrect annotations.

One of the main limitations of these libraries for MSI is that they tend to be trained with experimental data obtained by LC-MS/MS experiments, where most parental ions are fragmented as protonated adducts. Therefore, most common in silico libraries only include fragmentation predictions of protonated molecules, which do not represent how other adducts fragment (Al-Saad et al., 2003). In MSI, it is common to obtain different adducts such as sodium and potassium, depending on the sample preparation and acquisition, and for some species the protonated adduct may not even be detected (Garate et al., 2020).

4.2.7 | Peak filtering

Peak annotation results can be used to filter out redundant or nonbiologically relevant peaks from downstream statistical analyses.

A common peak filtering strategy is deisotoping (Bond et al., 2017; Sementé et al., 2021), which consists of localizing monoisotopic peaks in the spectra to remove all the subsequent isotopic peaks. This eliminates redundancy in the data, as all the isotopic peaks in a pattern are highly correlated, and facilitates the posterior identification of the monoisotopic peaks. In this same line, another peak filtering strategy is de-adducting, which consists in discovering as many adducts as

possible for each compound to combine them as a unique feature. The identification of adducts is mainly based on the mass difference between them; which could lead to the detection of false adducts since there may exist multiple mass differences between ions that match with several possible adducts. Additionally, the images produced by adducts are not necessarily colocalized among them (like in the case of isotopes) because the natural abundance of the adduct-forming elements over the tissue sample (Hankin et al., 2011) (i.e., Na^+ or K^+ ions) may be not homogeneous and dependent on the tissue type.

Finally, spectra contain exogenous peaks, (coming from the substrate, the matrix, the embedding medium, etc.) that may be desirable to exclude from the analysis. In the ideal case, these peaks should be annotated and discarded, although sometimes they could be useful for calibration purposes, like some inorganic matrix peaks (Ràfols et al., 2018). Most of the strategies behind annotating these off-sample ion peaks are based on exact mass matches by knowing which compounds are expected to appear in the sample preparation (Baquer et al., 2020; Niedermeyer & Strohalm, 2012), but there are also software programs that rely on machine learning methods to annotate them (Ovchinnikova & Kovalev, 2020).

4.2.8 | Data sharing and repositories

Data repositories are an essential tool for data sharing. The vast amount of experimental data available allow two main benefits to the community: experimental results are easily accessible to the whole community, and the data can be used to validate and develop software tools.

METASPACE (Alexandrov et al., 2019) is the main repository available in MSI. To date, METASPACE holds over 6000 experimental studies. Both the experimental data (in.imzML; Schramm et al., 2012; format and centroid mode) and resulting annotations using PySM (Palmer et al., 2016) can be freely downloaded.

More generic repositories include Metabolights (Haug et al., 2013) or Metabolomics Workbench (Sud et al., 2016). Nevertheless, their coverage of MSI experiments is rather limited. Only 50 (0.2% of all entries) and 4 (0.25% of all entries) of their respective entries correspond to MSI experiments.

4.3 | Specific software packages

The MSI community has dedicated their efforts to developing several software tools for compound annotation/identification of MSI data. In this section, we review 12 current software tools to guide the readers in selecting the most suitable ones for their application. Table 2 contains a summary of the main characteristics of each tool including the confidence levels of the annotations/identifications they can provide, the target features, the output, and the general type of annotation. We have defined three types of annotation: (1) "general annotation" if all the peaks in the spectra are targeted; (2) "specific annotation" if specific peaks (e.g., matrix) are annotated; and (3) "identification" if MSI is combined with MS/MS or other orthogonal techniques.

4.3.1 | Alex[123]

Alex[123] (Ellis et al., 2018) is a software for the automated identification of lipids. It relies on a unique experimental setup multiplexing an FTMS Orbitrap for high-mass resolution MSI and an IT-MS/MS for data-dependent acquisition (DDA) on-tissue fragmentation of almost every detected m/z value. By alternating the two acquisitions in 20 μm steps, they are able to effectively determine high-mass MSI and structural information in situ. This tool achieves Levels 3 and 2 identification confidence.

They rely on an in-house library that contains more than 430k molecular lipid species and their adduct-specific fragments. They use different adducts based on the lipid family. To annotate a sum-composition lipid species from the FTMS data, the peak must be present in all three replicates and at least one fragment must be detected by IT-MS/MS. To identify the lipid species, three conditions must be met: (1) at least 50% of the fragments must be detected, (2) two complementary pairs of fragments (adding to the parental ion) must be detected, and (3) the parental ion must be found by FTMS.

Using the MS data, they managed to annotate 165 unique sum-composition lipid species in rat brain tissue. From these sum-compositions, they managed to structurally identify 113 lipid species using the parallel IT-MS/MS run. A total of 92% of the identified lipids could be validated with HPLC-MS/MS.

4.3.2 | CycloBranch 2

CycloBranch 2 (Novák et al., 2020) is a standalone software package implemented in C++ that can annotate LC-MS, MSI, and MS/MS data independently or combine all of them. CycloBranch 2 generates molecular formulas from an input list of chemical elements to form a database of compounds, optimized for peptides and some small molecules. Later, all the molecules in the database

TABLE 2 Summary of software tools for annotation and identification

Name	Confidence level	– Annotation type – Target features	– Approach – Library	– Input data format – Output – Ranking scores	– Programming language – Installation – License	References
Alex[123]	2–3	– Specific to lipids. – On-tissue MS/MS fragments	– Library-centric – In-house	– RAW (Thermo Fisher Scientific) – List of identified lipids in 2 levels (annotated by exact mass matching and identified by MS/MS) – (1) Matched fragments percentage	– Python – Install from repository – GNU GPL v3.0	Ellis et al. (2018)
CycloBranch 2	2–4	– General identification and annotation. – Isotopes, adducts and molecular formulas.	– Library-centric – In silico library of molecular formulas tunable by the user.	– Profile in imzML, mzML and some proprietary formats. – Interactive windows with tables and spectra. – (1) Matched fragment count (2) Sum of fragment intensities	– C++ – Download and install a standalone package. – GNU GPL v3.0	Novák et al. (2020)
HIT-MAP	4	– Specific to peptides and proteins. – Full isotopic pattern	– Library-centric – In silico library from a protein sequence file in FASTA format.	– imzML – Two subfolders: one with identification data and the other with containing peptide and protein lists as well as the corresponding ion images – (1) Matched peaks percentage (2) RMS spectral error (3) Mass error (4) FDR (Target-Decoy)	– R – Install from github or docker image – GNU GPL v3.0	Guo et al. (2021)
LipostarMSI	2–4	– Specific to lipids. – Lipids, metabolites, and drug metabolites	– Library-centric – HMDB, LIPID MAPS and in-house.	– imzML (MSI) and CSV(MS/MS) – Interactive windows with tables and spectra – (1) Matched fragment percentage (2) TIC ratio (3) Spatial chaos (4) Mass error	– Not specified – Download and install a standalone package. – Private Software	Tortorella et al. (2020)
mass2adduct	4	– Specific to metabolite-matrix adducts. – Isotopes, adducts, and matrix adducts	– Feature-centric	– imzML – List of adduct masses in the R environment – (1) Pearson correlation	– R – Install from github – GNU GPL v3.0	Janda et al. (2021)

TABLE 2 (Continued)

Name	Confidence level	– Annotation type – Target features	– Approach – Library	– Input data format – Output – Ranking scores	– Programming language – Installation – License	References
MassPix	4	– General annotation. Specific molecular formulas of lipids. – Isotopes, adducts and molecular formulas.	– Feature-centric	– Centroid in imzML – Tables in CSV format – None	– R – Install from github – GNU GPL v3.0	Bond et al. (2017)
MSKendrickFilter	5	– Specific compound family annotation. – Suggests compound family based on KMD	– Feature-centric	– imzML – Images of MS signals classified as a user defined compound family. – (1) KMD	– Python – Available under request to the authors – Unlicensed	Kune et al. (2019)
OffsampleAI	5	– Specific to compounds outside of the sample. – MS signals outside of the sample	– Feature-centric	– imzML – Indication of of-sample ion in data – None	– Python – Install from github/ Built-in functionality in METASPACE – Apache 2.0	Ovchinnikova and Stuart (2020)
pySM (METASPACE)	4	– General annotation. – Metabolites high-resolution imaging	– Library-centric – In-house, HMDB, LipidMaps and SwissLipids.	– imzML – Tables in CSV format with annotations and FDR level of confidence. – (1) Weighted Pearson Correlation (2) Average difference of normalized intensities (3) Spatial chaos (4) FDR (Target-Decoy)	– Python – Install from github/ Built-in functionality in METASPACE – Apache 2.0	Palmer et al. (2016)
ReSCORE METASPACE	4	– General annotation – Improve sensitivity of annotation of metabolites with pySM	– Feature-centric	– Annotations from pySM – CSV table with annotations and q values – (1) q values	– Python – Install from github – Apache 2.0	C. Silva et al. (2018)

(Continues)

TABLE 2 (Continued)

Name	Confidence level	– Annotation type – Target features	– Approach – Library	– Input data format – Output – Ranking scores	– Programming language – Installation – License	References
rMSIannotation	4	– General annotation. – Isotopes and adducts of metabolites and peptides.	– Feature centric – Modeled after HMDB and Peptide Atlas	– imzML – R objects containing isotopes and adducts. – (1) Linear Regression R^2 (2) M + 0/M + 1 ratio difference (3) Mass error	– R/C++ – Install from github – GNU GPL v3.0	Sementé et al. (2021)
rMSIcleanup	4	– Specific to matrix peaks – Matrix-related MS signals	– Library-centric – In-house	– imzML – R object containing matrix clusters and PDF with spectra, ion images and matrix clusters – (1) Weighted Pearson's Correlation (2) Exponential of Euclidean distance	– R – Install from github – GNU GPL v3.0	Baquer et al. (2020)

are tested using various rules like the nitrogen to oxygen ratio, the Senior's rules (Kind & Fiehn, 2007) and matching the *m/z* in an experimental input spectrum. Additionally, CycloBranch 2 supports fine isotope structure annotation, being able to resolve $^{34}S/^{13}C_2$ and $^{41}K/^{13}C_2$ peaks. Moreover, CycloBranch 2 includes a tool to visualize the annotations over the MSI image combined with multiple microscopy or histology images, which can be shifted and adjusted manually to increase the overlap between them. The output of the software consists of a list of interactive tables that show the annotations over the spectra and images.

The tool was used to annotate an MSI data set consisting of a mixture of three commercial siderophores standards of bis-methylthio gliotoxin, ferrioxamine, and triacetylfusarinine C ferriform over an ITO glass. Cyclobranch 2 predicted elemental compositions of all three compounds reported in at least 50 spectra from a total of 1215. Later, the peaks were searched in a library of 709 siderophores and secondary metabolites as a positive control.

4.3.3 | HIT-MAP

HIT-MAP (Guo et al., 2021) is an R package that annotates peptides and proteins in high mass resolution MSI data sets using peptide mass fingerprint analysis and a scoring system. To annotate, HIT-MAP generates a customized local database of digested proteolytic peptides in silico from a protein sequence file in FASTA format containing the proteome of the species under investigation and a complete in silico digestion framework. Moreover, HIT-MAP generates a decoy database to produce FDR-controlled annotations.

To match the reference database with the experimental data, three principal scores are used. First, the number of peaks in the experimental isotopic pattern found in the theoretical pattern, discarding those peaks below 2.5% of the most intense isotopic peak; second, the intensity profile of the patterns; and third the mass error between peaks. Once a list of annotated peptides is generated, protein annotation is achieved by grouping peptides into the target proteins computing an FDR. The output of HIT-MAP consists of two subfolders, one containing all the identification data and the other a summary with peptide and protein lists as well as ion images.

4.3.4 | LipostarMSI

LipostarMSI (Tortorella et al., 2020) is a commercial software for targeted and untargeted MSI data analysis

with automated annotation of lipids, metabolites, and drug metabolites. It annotates by accurate *m/z* ratio matching within user-defined tolerances in libraries of compounds like the HMDB or LIPID MAPS. In-house libraries are also supported. Each hit to the database is ranked based on a mass score (proximity to the theoretical mass), an isotopic pattern score (compliance to theoretical intensity ratios and mass distances), and chaos score (spatial distribution of the *m/z* density image).

The software also allows the inclusion of MS/MS data to reach higher levels of confidence in identification. Each experimental MS/MS spectra can be compared to fragmentation libraries or to in silico fragments produced by a set of proprietary lipid fragmentation rules. In addition to the scores used in annotation, a new fragment score is introduced. This score is based on (1) the percentage of theoretical fragments found in the experimental data, and (2) the ratio between experimental and theoretical fragment intensities. Each theoretical fragment can be labeled as "mandatory" or "recommended" either manually or based on user-defined intensity thresholds. This allows the fragment score to only focus on relevant fragments.

The output of the software consists of a list of compounds assigned to each *m/z* ratio and ranked by the LipostarMSI score. Each annotation/identification is color-coded based on the confidence of annotation. Green indicates successful structural identification, orange indicates the presence of conflicts that need to be manually reviewed and approved; and red indicates unsuccessful identification. Finally, after approving correct identifications, all adducts assigned to the same compound are merged in a unique identification.

4.3.5 | Mass2adduct

Mass2adduct (Janda et al., 2021) is an R tool that follows a feature-centric approach to automatically annotate common alkali metal adducts, matrix adducts, and isotopes. The tool computes the mass difference between all *m/z* feature pairs available in the data set and plots them in a histogram. The most common mass differences are matched against a list of common adducts to determine their identity. Finally, the Pearson's correlation of each candidate adduct to their parental ion is used to discard unlikely adducts. Bonferroni correction and false-discovery rate analysis based on *q*-value cutoff are applied to Pearson's correlation values.

To validate their approach, they conducted on-tissue tandem MS on mouse brain tissue using DHB as the matrix. They focused on four pairs of *m/z* values with a mass difference of 136.016 Da (DHB-H_2O) and found that they showed identical MS/MS fragments.

They showcase their annotation tool on several tissue types, sources, mass analyzers and two matrices (DHB and CHCA). Comparable $[M + Na]^+$ and $[M + K]^+$ adduct frequencies were found across tissue types and experimental setups. Abundant matrix peaks were found for DHB (up to 30% of the total amount of features). CHCA was less abundant (up to 10% of all features).

As a final validation, they compare their results to METASPACE (Alexandrov et al., 2019). Out of the 604 *m/z* features annotated as matrix adducts by Mass2adduct for a mussel data set, a total of 103 were annotated as metabolites by METASPACE. This highlights that matrix adducts can cause false-positive annotations and they should be taken into account for library searches. They also conclude that exact mass matching is not enough for identification and the use of orthogonal techniques is required.

4.3.6 | massPix

massPix (Bond et al., 2017) is an R package that combines data analysis functionalities with deisotoping and exact mass matching against generated lipid libraries. The deisotoping algorithm finds monoisotopic ions (M + 0) and removes the first and second isotopes (M + 1 and M + 2) which are within a calculated proportion of M + 0. To achieve lipid annotation, first, a library of lipids is generated by combining common fatty acids, lipid head-groups and adducts; and second, the M + 0 ions previously found are matched against the generated library. The output consists of various CSV files with annotations.

4.3.7 | MSKendrickFilter

MSKendrickFilter (Kune et al., 2019) is a python software capable of exploiting the benefits of KMD analysis to classify chemically related compounds in their corresponding families. It is based on the conversion of exact mass measurements to a Kendrick Mass (KM) scale (linear conversion factor computed with the nominal and exact mass of a reference molecule of choice). With this transformation, the mass of the reference molecule—which is usually a repeating block in a bigger structure like CH_2 in lipids—does not contribute to the decimal part of the KM, which contains only information of the other elements of the molecular structure. The KMD is later obtained by subtracting the rounded KM from the KM (Kune et al., 2019). Their results show how

using CH_2 as a reference molecule, different tetraalkylammonium, lipids, and lipopeptides families can be identified. When using C_2H_4O as a reference molecule, different polymers groups could be separated. Their results were validated on bacteria cocultures and brain tissue sections.

4.3.8 | OffsampleAI

OffsampleAI (Ovchinnikova & Kovalev, 2020) is an artificial intelligence approach to recognize ion images localized outside of the sample (off-sample). The authors initially compiled a database of 23,238 ion images from 87 public MSI data sets manually labeled as on-sample and off-sample by five experts (using a custom web app). This database is used as a validation for the three algorithms proposed. The two first methods proposed, the "Spatio-molecular biclustering method" and the "molecular colocalization method" rely on the spatial correlation between ions and clustering of pixels to identify off-sample ions. The top-performing method is based on a deep residual learning approach trained on part of the gold standard.

4.3.9 | pySM (METASPACE)

Palmer et al. (2016) proposed a novel approach to annotate metabolite data in MSI with a confidence estimation approach. Using the compound-specific databases selected by the user, as well as a list of possible adducts, a list of all possible monoisotopic molecular matches is compiled. These molecular matches are then ranked based on the so-called metabolite-signal match score (MSM score), a composite score that relies on three metrics: (1) the "spatial chaos metric" quantifies the informativeness of the monoisotopic peak (2) the "spectral isotope metric" indicates the degree of similarity between the theoretical isotopic pattern and the experimental one and (3) the "spatial isotope metric" indicates the degree of similarity between the ionic images for all isotopes.

The MSM score values will depend largely on the sample at hand, making it difficult to specify a stable MSM cutoff. This is addressed using an FDR value estimation using a Target-Decoy approach. The main database with normal adducts is referred to as the Target database and it is extended with a Decoy database of the same size. In this case, the decoy is composed by randomly selecting implausible adducts. For each search in the Target database (using plausible adducts) a search in the Decoy database is conducted (using implausible adducts). All hits, from both the target and the decoy databases, are ranked based on MSM. The number of Decoy hits and Target hits above a certain MSM cutoff is used to estimate the FDR. This allows converting an MSM cutoff to a much more easily interpretable FDR cutoff.

pySM is currently integrated in the online annotation platform METASPACE (Alexandrov et al., 2019), which allows users to submit high-resolution data sets to be annotated using four libraries: CoreMetabolome (an in-house library), HMDB (Wishart et al., 2018), LipidMaps (Sud et al., 2007) and SwissLipids (Aimo et al., 2015). Moreover, METASPACE allows sharing the results online by storing all data online, both the MSI data and the annotations, and includes options for privacy and teamwork. METASPACE contains nowadays close to 6000 downloadable MSI data sets, being one of the biggest MSI data repositories in the world.

4.3.10 | ReSCORE METASPACE

Some strategies try to extract more information from the annotations and identifications rather than only speculating with the identity of peaks for MSI data sets. One of them is annotation rescoring, which implies a verification step after the initial annotation to increase the precision of the workflow. In this line, C. Silva et al. (2018) applied this strategy with METASPACE (Alexandrov et al., 2019) to increase the FDR of the target-decoy approach. The strategy consists of various recursive iterations of selecting some of the annotations with higher scores from the target set and some annotations from the decoy set to train a linear classifier using a collection of 34 features extracted for each annotation. At each iteration, the annotations are rescored using the linear classifier until a certain number of iterations is reached. The result of this procedure increases the number of annotated compounds for a given FDR in METASPACE.

4.3.11 | rMSIannotation

rMSIannotation (Sementé et al., 2021) is an annotation workflow integrated into the MSI processing R package rMSIproc (Ràfols et al., 2020) and implemented in C++. The algorithm annotates monoisotopic ions from metabolites and peptides by directly searching in the spectra peaks that accomplish three rules: spatial correlation, isotopic mass distance, and intensity profile of the isotopic pattern, which can be extracted with confidence thanks to the great number of sampling points in an MSI experiment. To avoid direct searches in libraries,

rMSIannotation uses a previous modelization of the intensity profiles of different compounds found in the HMDB (Wishart et al., 2018) and the Peptide Atlas (Desiere, 2006), which allows the prediction of variations in the intensity profile along the *m/z* axis. After detecting monoisotopic peaks, the algorithm groups them creating networks of adducts using spatial correlation as a criterion. The output of the algorithm consists of different R structures containing all the annotations in tables, information about the isotopic patterns, and the adduct networks. Moreover, it retrieves structures to facilitate the inclusion or exclusion of monoisotopic and isotopic peaks from the data analysis and there are visualization options included in rMSIproc.

4.3.12 | rMSIcleanup

rMSIcleanup (Baquer et al., 2020) is an R package that annotates matrix-related signals in MSI data sets. It annotates them by computing all the theoretical isotopic patterns related to the matrix clusters and matching them to the spectra using cluster spectral similarity and intra-cluster morphological similarity. Moreover, it detects overlapped peaks in the isotopic pattern using the clustering algorithm bisecting k-means and based on the correlation of their spatial distribution. The output of rMSIcleanup is an R data frame that can be exported in Rdata or CSV formats. Additionally, the package produces an informative visual report in PDF with all the patterns detected, ion images, and matrix-related annotations.

5 | EXTENDING THE IMZML FORMAT TO INCLUDE ANNOTATIONS AND IDENTIFICATIONS

The imzML is a data format (Schramm et al., 2012) created to enable the exchange of MSI data between different software and instruments. It uses two files linked by a universally unique identifier (UUID): (1) an XML file that stores experimental metadata that expands on the HUPO-PSI mzML standard format, and (2) a binary file to store spectral data efficiently. The spectral data can be stored in continuous mode, where all pixel MS measurements share the same *m/z* values, or in processed mode, where each pixel has its *m/z* values.

The imzML format is currently the gold standard for MSI data storage and sharing. Nevertheless, it does not contemplate a standard way of including molecular annotations and identifications.

The MS community has recognized the importance of storing annotations and identifications in a reproducible manner to stimulate data sharing and accountability. This interest promoted the creation of several file formats that complement the popular mzML file format (Martens et al., 2011), a standard format developed by the HUPO Proteomics Standards Initiative (Hermjakob, 2006) to "capture the use of a mass spectrometer, the data generated, and the initial processing of that data (to the level of the peak list)." Although these file formats are not compatible with MSI experiments, the current and most relevant formats to store annotations and identifications in MS are mzTab, mzTab-M, and mzIdentML.

mzTab (Griss et al., 2014) was first released in 2014 and it is intended to store only the final reported results of an MS proteomics experiment and to provide a simple way to share data with MS proteomics repositories. It can contain protein, peptide, and small molecule identifications with basic quantitative information. Using the same core as mzTab, a new format to better support small molecule experiments was developed by the end of 2019 as the 20th version of mzTab, the mzTab-M (Hoffmann et al., 2019). This file format is intended to extend the concept of mzTab to include more details for quantification, including different charge states or adducts, and was developed specifically for experiments on small molecules like metabolites and lipids. In the future, mzTab-M might be adopted to create a specific version of mzTab for proteomics only (mzTab-P; Salek, 2019), but at the moment, mzTab version 1.0 remains active for proteomics. Both standard file formats are structured as tab-delimited text files and are intended to share part of the results of an experiment (not all the MS data), which make them suitable for searches in libraries and to be the output of library searches. The files are structured as big tables of compound identifications with fields like database identifier, chemical formula, theoretical neutral mass, adduct ions, and various study variables that can be defined by the user. A heading containing metadata and some defining words are also included.

mzIdentML (Jones et al., 2012) is an XML-based format that was first released in August 2009 and reached the current version 1.2 in March 2017. It is intended for the systematic description of polypeptide identification and characterization based upon MS. The format was originally named AnalysisXML to encapsulate different computational analyses on proteomics performed with mass spectra, but it was decided to split the development into two branches: mzIdentML for peptide and protein identification, and mzQuantML (Walzer et al., 2013), to describe quantification experiments. mzIdentML can store MS data by itself, but it is expected to be accompanied by an mzML file (there is an mzML unique

identifier camp inside mzIdentML) containing the complete data set, as mzIdentML is best suited for results and not the complete experiment. Polypeptides identifications can be stored in different ways depending on the identification procedure, but the information usually consists of the sequence accession, the length of the sequence, information about the enzyme used, and fragmentation information among many others.

6 | PERSPECTIVES

6.1 | Identification confidence levels for MSI

As MSI matures into an analytical technique frequently used in untargeted metabolomic studies, the scientific community expects the same level of accuracy and accountability in MSI experiments as in studies with LC or GC coupled to MS or NMR. Thus, we propose the adoption of the identification confidence levels used in LC-MS metabolomics [19] to the field of MSI as described in Section II.B. and Supplementary Figure 1.

MSI lacks the chromatographic separation available in LC and GC metabolomics, which impedes the acquisition of orthogonal information (i.e., RT). Nevertheless, the high number of pixels enables image and peak intensity correlations to reliably annotate isotopes, adducts, and in-source fragments.

We are confident that the MSI community, especially in the field of software development for annotation and identification, would benefit from this proposal. First, we encourage the community to be consistent with the terms annotation and identification. As shown in Supporting Information: Table S1, more than 50% of the papers reviewed used the term identification to refer to exact mass matching. Assignments based on only exact mass matching (Levels 4–5) should be referred to as annotation. "Annotation" should still be used even when using orthogonal information to distinguish isomers and isobars (Levels 2–3). The term "identification" should only be used when all experimental data is matched against a reference standard (Level 1).

Second, we claim that users of software tools would appreciate a clear indication of the level of confidence the tool provides. The list of annotations and identifications produced by the software should include a field indicating the level of confidence (Levels 1–5). Furthermore, we consider that they should also be specified in any accompanying publication.

The adoption of these guidelines will provide a clear framework to communicate confidence in annotation and identification and ensure correct biological interpretation of the results. This initiative will also encourage the community to strive for higher identification confidence in their studies by adjusting their experimental and software workflows.

6.2 | Incorporation of annotations and identifications to the imzML format

Table 1 shows that imZML (Schramm et al., 2012) is the default input format in the overwhelming majority of software tools for annotation and identification in MSI. This indicates the full commitment of the community to the idea of cross-instrument, open protocol, and standardized data sharing. The imZML format has been a clear success. At the same time, Table 2 also shows a clear disparity of output formats (.csv, .xlsx, .Rdata, .pdf...). The resulting annotations and identifications are usually reported in loosely defined in-house formats with different fields that impede data sharing, integration, and reusability. Thus, we identified an imperative need for a standard format to report MSI annotations and identifications easily integrable with imZML.

We have observed that most data formats for MS that contain identifications (mzIdentML, Jones et al., 2012; mzTab, Griss et al., 2014; and mzTab-M, Hoffmann et al., 2019) were not designed to contain all the spectral data but as an annex to the mzML (Martens et al., 2011) data storing format.

We propose adopting this same strategy to define a new file format to include annotations and identifications as an annex to the imzML standard. In particular, we consider that in the field of metabolomics the format mzTab-M should be used as a reference. Each data set would now be described by three key files: the common.ibd and .imzML files containing the spectral data and a new.mzTab-M file containing annotations, identifications, and supporting evidence. All these files would be linked using the same Universally Unique Identifier (UUID). The .mzTab-M file could contain multiple UUIDs in studies with multiple.imzML files. Figure 3 shows a high-level abstraction of the imzML format, the mzTab-M format, and their integration by a list of UUIDs.

mzTab-M is the result of years of collaborative work between the Metabolomics Standards Initiative, Proteomics Standards Initiative, and Metabolomics Society. It relies on a well-defined structure and controlled vocabulary and it can be read, written, and validated using mzTab-M (Hoffmann, Hartler, et al., 2019). It has successfully been adopted by some of the main MS annotation software such as Lipid Data Analyzer (Hartler

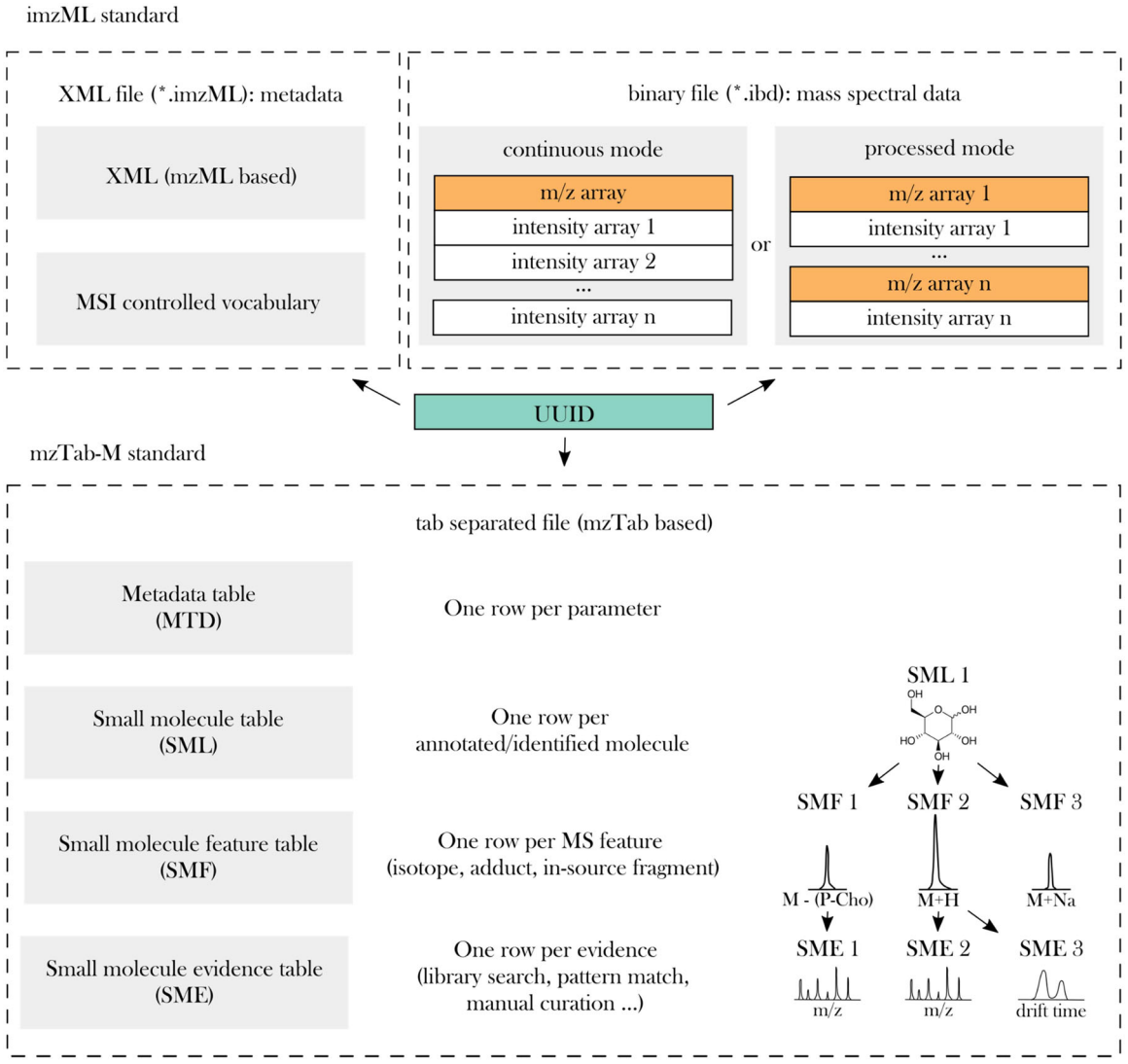


FIGURE 3 Adaptation of mzTab-M format to be compatible with imzML. A list of unique universally identifiers (UUIDs) would link multiple imzML files from the same study to a single mzTab-M file containing annotations and identifications. Adapted with permission from Schramm et al. (2012) and Hoffmann, Hartler, et al. (2019). Copyright 2022 Elsevier. CC-BY license https://creativecommons.org/licenses/by/4.0/ [Color figure can be viewed at wileyonlinelibrary.com]

et al., 2011), GNPS (Nothias et al., 2020), MS-Dial (Tsugawa et al., 2015), and MetaboAnalyst (Chong et al., 2018).

mzTab-M is in plain text, making it visually easy to read and understand. Additionally, its tab-separated format, similar to the CSV format, is natively supported in Excel and other spreadsheet software. It is therefore a viable alternative to excel and CSV files used in publications and statistical programming languages like R.

The main drawback of mzTab-M is that it relies on a custom structure defined by its own specification. We consider that using extensible markup language (XML), a ubiquitous file format in all fields of computer science, would offer several advantages. All major programming languages and platforms have plenty of reliable tools to read, write, and validate XML, and its well-defined structure makes it extensible. In the long run, adapting mzTab-M to XML ensures a robust adoption by more developers and easier maintenance. We consider that one of the priorities when adopting mzTab-M for MSI applications is to redefine it in XML format. To ensure ease of access to the annotations and identifications by researchers with a lack of coding background, the community should develop a converter to the original mzTab-M tab-separated format.

Additionally, to adapt it to the field of MSI, part of the controlled vocabulary and fields defined by the mzTab-M format would need to be updated or removed. New fields would also need to be defined. As an example, all columns regarding RT in the small

molecule feature table (SMF) should be removed. The general structure of metadata, small molecule table (SML), SMF, and small molecule evidence table would remain unchanged.

Finally, the most crucial point to take into account is how to include the spatial information of the identified compounds. The same MS signal can correspond to different molecules in different areas of the tissue, especially when working with low-resolution MS analyzers, like peptides with the same *m/z* belonging to different proteins (Guo et al., 2021). Accounting for this phenomenon is a nontrivial task. We suggest including a column to specify the ROI of a specific MS feature. The representation and storage of ROIs are not properly solved in MSI and multiple vendors and software tools use their own custom-built formats.

6.3 | The future of automatic annotation and identification in MSI

We have extensively reviewed 12 software tools available between 2016 and 2022 to perform automatic identification and annotation of MSI data. Tools specialize in different target molecules (i.e., metabolites, lipids, peptides, or proteins), different experimental data (i.e., MS, tandem MS, ion mobility or other orthogonal techniques), and different approaches (i.e., library-centric or feature centric). Most of the tools available to date only focus on annotation and only reach identification Level 4 as they rely on exact mass matching. ALEX^{123} (Ellis et al., 2018), CycloBranch 2 (Novák et al., 2020) and Lipostar (Tortorella et al., 2020) are the only tools that can consistently provide Level 3 or Level 2 identifications. There is a clear need for automatic tools that can provide identifications with a confidence level over 3. Combining structural information obtained from orthogonal techniques is an important area of research that needs to be further explored.

For a confident identification, it is important to highlight the importance of proper mass calibration (Ràfols et al., 2018), using internal standard compounds or matrix peaks, and the use of high-resolution mass analyzers with mass accuracy below 5 ppm.

The future of automated annotation and identification in MSI relies not only on instrumental development but also on creativity in the application of strategies inspired by more established MS-based techniques such as LC-MS and GC-MS. We have identified the following challenges where software developers have an opportunity to make an impact in the field of annotation and identification by MSI:

- *In-source fragmentation*

  To date, there is no automatic tool that directly addresses the annotation of in-source fragments (fragments generated naturally during ionization or desorption) in MSI. Their correct annotation is key, as in-source fragments clutter the spectra and can be wrongfully annotated as other parental ions (Garate et al., 2020). This is particularly problematic in ion sources like SIMS and LA-ICP, but it is still a problem in soft-ionization sources like DESI or MALDI. At the same time, if properly dealt with, in-source fragments promise to increase confidence in annotation as they can provide insights into the structure of a molecule (much like tandem MS). In a recent LC-MS study, Xue et al. (2020) proposed adjusting the ESI source to produce in-source fragmentation patterns comparable to the MS/MS spectra available in METLIN (Smith et al., 2005). They found that 90% of 50 mixed metabolites showed in-source fragmentation patterns consistent with METLIN. This could lead to potentially high levels of confidence (above Level 3) only using MS1 data.
- *Exogenous compounds*

  Similarly, although several efforts have been presented in recent years (Baquer et al., 2020; Janda et al., 2021; Ovchinnikova & Kovalev, 2020), a comprehensive and reliable tool for the annotation of matrix-related signals of all widely used matrices is still missing. The use of inorganic matrices limits the presence of matrix fragments in the low range of the spectrum, but its use is far from being widespread.

  Another area needing further research is the annotation of exogenous compounds. Various MSI workflows contemplate the use of FFPE slides as sampling material, but identifying all the peaks that originated during the sample processing is still an open issue. Here we see an opportunity for researchers to develop software tools dealing with the identification and removal of all the peaks related to FFPE, OCT, or other cutting materials, which would require an in-depth analysis of the chemical processes produced by the sample processing.
- *SIL annotation*

  Following this line, SIL methods for MSI would benefit from the development of annotation tools specially designed for targeting different compounds with distinct or artificial isotopic patterns. There are various annotation tools for LC-MS data that are able to target SIL compounds (Capellades et al., 2016; Dange et al., 2020; Neumann et al., 2014). These tools could be used to inspire the development of new software for MSI. Even better, contributing to the development of this software to include MSI data

would allow combining both LC-MS and MSI SIL methods, which would benefit both disciplines and open the door for more collaboration between techniques in the SIL field.

- *Pathways in LC or GC-MS and how to apply them in MSI*

  Metabolic pathway analysis (a.k.a. metabolic pathway enrichment analysis) compares two sample classes (i.e., control vs. treatment or condition vs. wildtype) to produce a list of dysregulated (upregulated or downregulated) metabolic pathways. Data about metabolic pathways is obtained from databases such as KEGG (Kanehisa & Goto, 2000), HMDB (Wishart et al., 2018), or BioCyc (Caspi et al., 2014). For each pathway found, the coverage percentage is given (the percentage of metabolites in the pathway annotated). For each feature annotation, the dysregulation (up or down), fold-change and *p*-value are given. Additionally, an overview of all pathways can be represented in a variety of plots showing overall significance (*p* value) or metabolite overlap percentage. This process is typically performed on the list of annotations, but using the mummichog algorithm it can be applied directly to MS features. XCMS (Forsberg et al., 2018) and MetaboAnalyst (Chong et al., 2018), two major MS metabolomics processing platforms, implement pathway analysis.

  Additionally, to facilitate the generation of hypotheses, several software tools also include interactive network explorers. Metaboanalyst (Chong et al., 2018), for example, allows the user to show the metabolite annotations on the KEGG (Kanehisa & Goto, 2000) global metabolic network and other networks.

  To date, there is no automatic tool that can provide pathway analysis in MSI. Currently, pathway analysis in MSI is typically done by (1) running annotation/identification, (2) exporting a list of significant metabolites when comparing two ROIs, and (3) conducting pathway analysis using non-MSI targeted tools data such as XCMS or MetaboAnalyst. As an example, Sun et al. (2018) followed this approach (using KEGG and MetaboAnalyst) to metabolically compare the cortex and medulla in human adult adrenal gland samples. Among other pathways, the purine metabolism pathway was upregulated in the medulla while the biosynthesis of unsaturated fatty acids was upregulated in the cortex.
- *The role of AI and deep learning (DL) in annotation*

Finally, we conclude the review by addressing the hot topic on every researcher's lips: DL. DL has already achieved science-fiction-like results in a wide range of fields such as robotics (Sünderhauf et al., 2018), natural language processing (Otter et al., 2021), and medical image processing (Minaee et al., 2021). In recent years, MSI has seen some developments in machine learning (ML) and DL in applications such as tumor classification (Behrmann et al., 2018), clustering (Zhang et al., 2021), image registration (Race et al., 2021), and peak picking (Abdelmoula et al., 2021). In the field of molecular annotation and identification, OffSample AI (Ovchinnikova & Kovalev, 2020) used several DL models for the annotation of matrix-related and off-sample MS features. Nevertheless, the adoption of these technologies for MSI metabolomics is slow and we seem to be missing out on this Artificial Intelligence revolution ("Why the metabolism field risks missing out on the AI revolution," 2019). The two main drawbacks that are holding the community back are (1) the lack of result transparency and accountability, and (2) the lack of big data for training.

MSI is used in fields such as biochemistry, pharmaceutics, and medical diagnostics where reliable annotations and identifications are crucial. Since their inception, ML and DL have struggled with their inability to transparently justify their learning-based nonlinear results (black-box problem) (Castelvecchi, 2016). This inherent problem leaves scientists and funding bodies unable to fully interpret and trust DL results (von Eschenbach, 2021). There are three strategies to open the black box (Azodi et al., 2020). In the field of MSI molecular annotation and identification, the black-box problem could be mitigated by coupling DL models with more traditional score-based methods (i.e., spectral similarity, spatial similarity, spectral chaos, FDR estimates, etc.). Only annotations and identifications ranking high in both approaches would be accepted automatically, while mismatching annotations and identifications would be manually curated by the user.

The second bottleneck limiting the adoption of DL is the lack of big, labeled and curated sets of MSI data ("ground truth") needed to train the models (Alexandrov, 2020). Ideally, for training DL models, in the task of annotation and identification, we would need thousands of MSI data sets with a complete list of Level 1 identifications. Additionally, for the DL model to generalize, it should be exposed to enough sample types (specimen, condition, and tissue) and instrumental setups (ion source, ion mode, and mass analyzer). METASPACE (Ovchinnikova & Kovalev, 2020) includes thousands of publicly available data sets, but it does not include a complete list of confident annotations. The creation of this ground truth could follow two approaches (Alexandrov, 2020). The first approach relies on expert crowdsourcing to manually annotate MSI data sets and has successfully been used in MSI to estimate quality (Palmer et al., 2015), off-sample signals (Ovchinnikova &

Kovalev, 2020), and colocalization (Ovchinnikova & Stuart, 2020). Nevertheless, expert annotation could prove unfeasible and unreliable in the task of molecular annotation and identification. Following the success of MS/MS libraries like METLIN (Smith et al., 2005), NIST (Lemmon et al., 2010), or MassBank (Horai et al., 2010), the second approach involves the creation of an MSI metabolite spectral library using tissue mimetics (or alternative approaches described in Section 3). This is certainly one of the biggest challenges ahead for our community, but DL promises to give birth to the next generation of automated tools to more reliably answer the question "what are we imaging?".

ACKNOWLEDGMENTS

The authors acknowledge the financial support of the Spanish Ministry of Economy and Competitivity through project RTI2018-096061-B-100. GB acknowledges the financial support of the European Union's Horizon 2020 research and innovation program under the Marie Skłodowska-Curie grant agreement No. 713679 and the Universitat Rovira i Virgili (URV). LS acknowledges the financial support of Universitat Rovira i Virgili through the predoctoral grant 2017PMF-PIPF-60. TM acknowledges the financial support of the Universitat Rovira i Virgili through the predoctoral grant ref. PRE2019-089374. MGA acknowledges the financial support from the Agency for Management of University and Research Grants of the Generalitat de Catalunya (AGAUR) through the postdoctoral grant 2018 BP 00188. CRUE/CSUC hybrid - Universitat Rovira i Virgili.

ORCID

Gerard Baquer http://orcid.org/0000-0002-4433-4972
Lluc Sementé http://orcid.org/0000-0001-9956-6570
Toufik Mahamdi http://orcid.org/0000-0001-7132-0106
Xavier Correig http://orcid.org/0000-0002-6902-3054
Pere Ràfols http://orcid.org/0000-0002-9240-4058
María García-Altares http://orcid.org/0000-0003-4255-1487

REFERENCES

Abdelhamid, H.N. (2018) 'Nanoparticle assisted laser desorption/ionization mass spectrometry for small molecule analytes'. https://doi.org/10.1007/s00604-018-2687-8

Abdelmoula, W.M. et al. (2021) 'Peak learning of mass spectrometry imaging data using artificial neural networks', *Nature communications*, *12*(1), p. 5544. https://doi.org/10.1038/s41467-021-25744-8

Aimo, L. et al. (2015) 'The SwissLipids knowledgebase for lipid biology', *Bioinformatics*, *31*(17), pp. 2860–2866. https://doi.org/10.1093/bioinformatics/btv285

Alam, S.I., Kumar, B. and Kamboj, D.V. (2012) 'Multiplex detection of protein toxins using MALDI-TOF-TOF tandem mass spectrometry: application in unambiguous toxin detection from bioaerosol', *Analytical chemistry*, *84*(23), pp. 10500–10507. https://doi.org/10.1021/ac3028678

Alexandrov, T. et al. (2019) 'METASPACE: a community-populated knowledge base of spatial metabolomes in health and disease', *bioRxiv* [Preprint], pp. 1–22. https://doi.org/10.1101/539478

Alexandrov, T. (2020) 'Spatial metabolomics and imaging mass spectrometry in the age of artificial intelligence', *Annual review of biomedical data science*, *3*, pp. 61–87. https://doi.org/10.1146/annurev-biodatasci-011420-031537

Alexandrov, T. and Bartels, A. (2013) 'Testing for presence of known and unknown molecules in imaging mass spectrometry', *Bioinforcmatics*, *29*(18), pp. 2335–2342. https://doi.org/10.1093/bioinformatics/btt388

Almeida, R. et al. (2015) 'Comprehensive lipidome analysis by shotgun lipidomics on a hybrid quadrupole-orbitrap-linear ion trap mass spectrometer', *Journal of the american society for mass spectrometry*, *26*, pp. 133–148. https://doi.org/10.1007/s13361-014-1013-x

Al-Saad, K.A. et al. (2003) 'Structural analysis of phosphatidylcholines by post-source decay matrix-assisted laser desorption/ionization time-of-flight mass spectrometry', *Journal of the american society for mass spectrometry*, *14*, pp. 373–382. https://doi.org/10.1016/s1044-0305(03)00068-0

Amstalden van Hove, E.R., Smith, D.F. and Heeren, R.M.A. (2010) 'A concise review of mass spectrometry imaging', *Journal of chromatography. A*, *1217*(25), pp. 3946–3954. https://doi.org/10.1016/j.chroma.2010.01.033

Ayet San Andrés, S. et al. (2019) 'High-resolution, accurate multiple-reflection time-of-flight mass spectrometry for short-lived, exotic nuclei of a few events in their ground and low-lying isomeric states', *Physical review C: Nuclear physics*, *99*(6), p. 064313. https://doi.org/10.1103/PhysRevC.99.064313

Azodi, C.B., Tang, J. and Shiu, S.-H. (2020) 'Opening the black box: interpretable machine learning for geneticists', *Trends in genetics*, *36*(6), pp. 442–455. https://doi.org/10.1016/j.tig.2020.03.005

Baijnath, S. et al. (2016) 'Small molecule distribution in rat lung: a comparison of various cryoprotectants as inflation media and their applicability to MSI', *Journal of molecular histology*, *47*(2), pp. 213–219. https://doi.org/10.1007/s10735-016-9658-3

Baquer, G. et al. (2020) 'RMSIcleanup: an open-source tool for matrix-related peak annotation in mass spectrometry imaging and its application to silver-assisted laser desorption/ionization', *Journal of cheminformatics*, *12*(1), p. 45. https://doi.org/10.1186/s13321-020-00449-0

Barry, J.A. et al. (2019) 'Multicenter validation study of quantitative imaging mass spectrometry', *Analytical chemistry*, *91*(9), pp. 6266–6274. https://doi.org/10.1021/acs.analchem.9b01016

Basu, S.S. et al. (2019) 'Metal oxide laser ionization mass spectrometry imaging (MOLI MSI) using cerium (IV) oxide', *Analytical chemistry*, *91*(10), pp. 6800–6807. https://doi.org/10.1021/acs.analchem.9b00894

Becker, J.S. et al. (2011) 'Mass spectrometric imaging (MSI) of metals using advanced BrainMet techniques for biomedical research', *International journal of mass spectrometry*, *307*(1–3), pp. 3–15. https://doi.org/10.1016/j.ijms.2011.01.015

Becker, J.S. et al. (2012) 'Mass spectrometry imaging (MSI) of metals in mouse spinal cord by laser ablation ICP-MS', *Metallomics: Integrated biometal science*, *4*(3), pp. 284–288. https://doi.org/10.1039/c2mt00166g

Bednařík, A. et al. (2022) 'Mass spectrometry imaging techniques enabling visualization of lipid isomers in biological tissues', *Analytical chemistry*, *94*(12), pp. 4889–4900. https://doi.org/10.1021/acs.analchem.1c05108

Behrmann, J. et al. (2018) 'Deep learning for tumor classification in imaging mass spectrometry', *Bioinformatics*, *34*(7), pp. 1215–1223. https://doi.org/10.1093/bioinformatics/btx724

Bemis, K.D. et al. (2015) 'Cardinal: an R package for statistical analysis of mass spectrometry-based imaging experiments', *Bioinformatics*, *31*(14), pp. 2418–2420. https://doi.org/10.1093/bioinformatics/btv146

Bielow, C. et al. (2017) 'On mass ambiguities in high-resolution shotgun lipidomics', *Analytical chemistry*, *89*(5), pp. 2986–2994. https://doi.org/10.1021/acs.analchem.6b04456.

Bien, T. et al. (2021) 'Transmission-mode MALDI mass spectrometry imaging of single cells: Optimizing sample preparation protocols', *Analytical chemistry*, *93*(10), pp. 4513–4520. https://doi.org/10.1021/acs.analchem.0c04905

Böcker, S. et al. (2006) 'Decomposing metabolomic isotope patterns', *Lecture Notes in Computer Science*, *4175*, pp. 12–23. https://doi.org/10.1007/11851561_2

Bokhart, M.T. et al. (2018) 'MSiReader v1.0: Evolving open-source mass spectrometry imaging software for targeted and untargeted analyses', *Journal of the American Society for Mass Spectrometry*, *29*(1), pp. 8–16. https://doi.org/10.1007/s13361-017-1809-6

Bond, N.J. et al. (2017) 'massPix: an R package for annotation and interpretation of mass spectrometry imaging data for lipidomics', *Metabolomics: Official journal of the Metabolomic Society*, *13*(11), p. 128. https://doi.org/10.1007/s11306-017-1252-5

Bowman, A.P. et al. (2020) 'Ultra-high mass resolving power, mass accuracy, and dynamic range MALDI mass spectrometry imaging by 21-T FT-ICR MS', *Analytical chemistry*, *92*(4), pp. 3133–3142. https://doi.org/10.1021/acs.analchem.9b04768

Buchberger, A.R. et al. (2018) 'Mass spectrometry imaging: A review of emerging advancements and future insights', *Analytical Chemistry*, *90* pp. 240–265. https://doi.org/10.1021/acs.analchem.7b04733

Buck, A. et al. (2015) 'Distribution and quantification of irinotecan and its active metabolite SN-38 in colon cancer murine model systems using MALDI MSI', *Analytical and bioanalytical chemistry*, *407*(8), pp. 2107–2116. https://doi.org/10.1007/s00216-014-8237-2

Calvano, C.D. et al. (2018) 'MALDI matrices for low molecular weight compounds: An endless story?', *Analytical and bioanalytical chemistry*, *410*(17), pp. 4015–4038. https://doi.org/10.1007/s00216-018-1014-x

Capellades, J. et al. (2016) 'GeoRge: A computational tool to detect the presence of stable isotope labeling in LC/MS-based untargeted metabolomics', *Analytical chemistry*, *88*(1), pp. 621–628. https://doi.org/10.1021/acs.analchem.5b03628

Caspi, R. et al. (2014) 'The MetaCyc database of metabolic pathways and enzymes and the BioCyc collection of Pathway/Genome Databases', *Nucleic acids research*, *42*(Data-Database issue), pp. D459–D471. https://doi.org/10.1093/nar/gkt1103

Castelvecchi, D. (2016) 'Can we open the black box of AI?', *Nature*, *538*(7623), pp. 20–23. https://doi.org/10.1038/538020a

Chatterji, B. and Pich, A. (2013) 'MALDI imaging mass spectrometry and analysis of endogenous peptides', *Expert review of proteomics*, *10*(4), pp. 381–388. https://doi.org/10.1586/14789450.2013.814939

Chong, J. et al. (2018) 'MetaboAnalyst 4.0: towards more transparent and integrative metabolomics analysis', *Nucleic acids research*, *46*(W1), pp. W486–W494. https://doi.org/10.1093/nar/gky310

Chumbley, C.W. et al. (2016) 'Absolute quantitative MALDI imaging mass spectrometry: a case of rifampicin in liver tissues', *Analytical chemistry*, *88*(4), pp. 2392–2398. https://doi.org/10.1021/acs.analchem.5b04409

Cillero-Pastor, B. and Heeren, R.M.A. (2014) 'Matrix-assisted laser desorption ionization mass spectrometry imaging for peptide and protein analyses: a critical review of on-tissue digestion', *Journal of proteome research*, *13*(2), pp. 325–335. https://doi.org/10.1021/pr400743a

Claude, E., Jones, E.A. and Pringle, S.D. (2017) 'DESI mass spectrometry imaging (MSI)', *Methods in molecular biology*, *1618*, pp. 65–75. https://doi.org/10.1007/978-1-4939-7051-3_7

Clish, C.B. (2015) 'Metabolomics: an emerging but powerful tool for precision medicine', *Cold Spring Harbor molecular case studies*, *1*(1), p. a000588. https://doi.org/10.1101/mcs.a000588

C. Silva, A.S. et al. (2018) 'Data-driven rescoring of metabolite annotations significantly improves sensitivity', *Analytical chemistry*, *90*(19), pp. 11636–11642. https://doi.org/10.1021/acs.analchem.8b03224

Dange M.C. et al. (2020) 'Evaluation of freely available software tools for untargeted quantification of 13C isotopic enrichment in cellular metabolome from HR-LC/MS data', *Metabolic Engineering Communications*, *10*, p. e00120. https://doi.org/10.1016/j.mec.2019.e00120

DeKeyser, S.S. et al. (2007) 'Imaging mass spectrometry of neuropeptides in decapod crustacean neuronal tissues', *Journal of proteome research*, *6*(5), pp. 1782–1791. https://doi.org/10.1021/pr060603v

Desiere, F. (2006) 'The PeptideAtlas project', *Nucleic Acids Research*, pp. D655–D658. https://doi.org/10.1093/nar/gkj040

Dewez, F. et al. (2019) 'Precise co-registration of mass spectrometry imaging, histology, and laser microdissection-based omics', *Analytical and bioanalytical chemistry*, *411*(22), pp. 5647–5653. https://doi.org/10.1007/s00216-019-01983-z

Diehl, H.C. et al. (2015) 'The challenge of on-tissue digestion for MALDI MSI- a comparison of different protocols to improve imaging experiments', *Analytical and bioanalytical chemistry*, *407*(8), pp. 2223–2243. https://doi.org/10.1007/s00216-014-8345-z

Dilillo, M. et al. (2017) 'Mass spectrometry imaging, laser capture microdissection, and LC-MS/MS of the same tissue section', *Journal of proteome research*, *16*(8), pp. 2993–3001. https://doi.org/10.1021/acs.jproteome.7b00284

Dilillo, M. et al. (2017) 'Ultra-high mass resolution MALDI imaging mass spectrometry of proteins and metabolites in a mouse model of glioblastoma', *Scientific reports*, *7*(1), p. 603. https://doi.org/10.1038/s41598-017-00703-w

Djoumbou-Feunang, Y. et al. (2019) 'metabolites CFM-ID 3.0: Significantly Improved ESI-MS/MS Prediction and Compound Identification'. https://doi.org/10.3390/metabo9040072

Dong, F. et al. (2020) 'Highly selective isomer fluorescent probes for distinguishing homo-/cysteine from glutathione based on AIE', *Talanta*, *206*, p. 120177. https://doi.org/10.1016/j.talanta.2019.120177

Drake, R.R. et al. (2018) 'MALDI mass spectrometry imaging of N-linked glycans in tissues', *Advances in experimental medicine and biology*, *1104*, pp. 59–76. https://doi.org/10.1007/978-981-13-2158-0_4

Dreisewerd, K., Bien, T. and Soltwisch, J. (2022) 'MALDI-2 and t-MALDI-2 mass spectrometry imaging', *Methods in molecular biology*, *2437*, pp. 21–40. https://doi.org/10.1007/978-1-0716-2030-4_2

Dueñas, M.E. et al. (2017) 'High spatial resolution mass spectrometry imaging reveals the genetically programmed, developmental modification of the distribution of thylakoid membrane lipids among individual cells of maize leaf', *The plant journal: For cell and molecular biology*, *89*(4), pp. 825–838. https://doi.org/10.1111/tpj.13422

Dufresne, M. et al. (2013) 'Silver-assisted laser desorption ionization for high spatial resolution imaging mass spectrometry of olefins from thin tissue sections', *Analytical chemistry*, *85*(6), pp. 3318–3324. https://doi.org/10.1021/ac3037415

Dührkop, K. et al. (2019) 'SIRIUS 4: a rapid tool for turning tandem mass spectra into metabolite structure information', *Nature methods*, *16*(4), pp. 299–302. https://doi.org/10.1038/s41592-019-0344-8

Duncan, K.D. et al. (2018) 'Quantitative mass spectrometry imaging of prostaglandins as silver ion adducts with nanospray desorption electrospray ionization', *Analytical chemistry*, *90*(12), pp. 7246–7252. https://doi.org/10.1021/acs.analchem.8b00350

Easterling, M.L., Mize, T.H. and Amster, I.J. (1999) 'Routine part-per-million mass accuracy for high- mass ions: space-charge effects in MALDI FT-ICR', *Analytical chemistry*, *71*(3), pp. 624–632. https://doi.org/10.1021/ac980690d

Eckelmann, D., Kusari, S. and Spiteller, M. (2018) 'Stable isotope labeling of prodiginines and serratamolides produced by Serratia marcescens directly on agar and simultaneous visualization by matrix-assisted laser desorption/ionization imaging high-resolution mass spectrometry', *Analytical chemistry*, *90*(22), pp. 13167–13172. https://doi.org/10.1021/acs.analchem.8b03633

Eikel, D. et al. (2011) 'Liquid extraction surface analysis mass spectrometry (LESA-MS) as a novel profiling tool for drug distribution and metabolism analysis: the terfenadine example', *Rapid communications in mass spectrometry: RCM*, *25*(23), pp. 3587–3596. https://doi.org/10.1002/rcm.5274

Ekelöf, M. et al. (2018) 'Evaluation of digital image recognition methods for mass spectrometry imaging data analysis', *Journal of the american society for mass spectrometry*, *29*(12), pp. 2467–2470. https://doi.org/10.1007/s13361-018-2073-0

Ellis, S.R. et al. (2018) 'Automated, parallel mass spectrometry imaging and structural identification of lipids', *Nature methods*, *15*(7), pp. 515–518. https://doi.org/10.1038/s41592-018-0010-6

Ellis, S.R. et al. (2021) 'Mass spectrometry imaging of phosphatidylcholine metabolism in lungs administered with therapeutic surfactants and isotopic tracers', *Journal of lipid research*, *62*, p. 100023. https://doi.org/10.1016/j.jlr.2021.100023

von Eschenbach, W.J. (2021) 'Transparency and the black box problem: why we do not trust AI', *Philosophy & technology*, *34*(4), pp. 1607–1622. https://doi.org/10.1007/s13347-021-00477-0

Evers, T.M.J. et al. (2019) 'Deciphering metabolic heterogeneity by single-cell analysis', *Analytical chemistry*, *91*(21), pp. 13314–13323. https://doi.org/10.1021/acs.analchem.9b02410

Fernández, J.A. et al. (2011) 'Matrix-assisted laser desorption ionization imaging mass spectrometry in lipidomics', *Analytical and bioanalytical chemistry*, *401*(1), pp. 29–51. https://doi.org/10.1007/s00216-011-4696-x

Forsberg, E.M. et al. (2018) 'Data processing, multi-omic pathway mapping, and metabolite activity analysis using XCMS Online', *Nature protocols*, *13*(4), pp. 633–651. https://doi.org/10.1038/nprot.2017.151

Fuchs, K. et al. (2018) 'Mapping of drug distribution in the rabbit liver tumor model by complementary fluorescence and mass spectrometry imaging', *Journal of controlled release: Official journal of the Controlled Release Society*, *269*, pp. 128–135. https://doi.org/10.1016/j.jconrel.2017.10.042

Fu, T. et al. (2018) 'Tandem mass spectrometry imaging and in situ characterization of bioactive wood metabolites in Amazonian tree species Sextonia rubra', *Analytical chemistry*, *90*(12), pp. 7535–7543. https://doi.org/10.1021/acs.analchem.8b01157

Gamble, L.J. and Anderton, C.R. (2016) 'Secondary ion mass spectrometry imaging of tissues, cells, and microbial systems', *Microscopy today*, *24*(2), pp. 24–31. https://doi.org/10.1017/S1551929516000018

Garate, J. et al. (2020) 'Influence of lipid fragmentation in the data analysis of imaging mass spectrometry experiments', *Journal of the american society for mass spectrometry*, *31*(3), pp. 517–526. https://doi.org/10.1021/jasms.9b00090

Gemperline, E., Rawson, S. and Li, L. (2014) 'Optimization and comparison of multiple MALDI matrix application methods for small molecule mass spectrometric imaging', *Analytical chemistry*, *86*(20), pp. 10030–10035. https://doi.org/10.1021/ac5028534

Gibb, S. and Strimmer, K. (2012) 'MALDIquant: a versatile R package for the analysis of mass spectrometry data', *Bioinformatics*, *28*(17), pp. 2270–2271. https://doi.org/10.1093/bioinformatics/bts447

Giordano, S. et al. (2016) '3D mass spectrometry imaging reveals a very heterogeneous drug distribution in tumors', *Scientific reports*, *6*, p. 37027. https://doi.org/10.1038/srep37027

Gode, D. and Volmer, D.A. (2013) 'Lipid imaging by mass spectrometry--a review', *The analyst*, *138*(5), pp. 1289–1315. https://pubs.rsc.org/en/content/articlehtml/2013/an/c2an36337b

Goodwin, R.J.A. et al. (2011) 'Qualitative and quantitative MALDI imaging of the positron emission tomography ligands raclopride (a D2 dopamine antagonist) and SCH 23390 (a D1 dopamine antagonist) in rat brain tissue sections using a solvent-free dry matrix application method', *Analytical chemistry*, *83*(24), pp. 9694–9701. https://doi.org/10.1021/ac202630t

Gorshkov, M.V. et al. (2010) 'Calibration function for the Orbitrap FTMS accounting for the space charge effect', *Journal of the*

american society for mass spectrometry, *21*(11), pp. 1846–1851. https://doi.org/10.1016/j.jasms.2010.06.021

Grey, A.C. et al. (2019) 'A quantitative map of glutathione in the aging human lens', *International journal of mass spectrometry*, *437*, pp. 58–68. https://doi.org/10.1016/j.ijms.2017.10.008

Grey, A.C. et al. (2021) 'Applications of stable isotopes in MALDI imaging: current approaches and an eye on the future', *Analytical and bioanalytical chemistry*, *413*(10), pp. 2637–2653. https://doi.org/10.1007/s00216-021-03189-8

Griss, J. et al. (2014) 'The mzTab data exchange format: communicating mass-spectrometry-based proteomics and metabolomics experimental results to a wider audience', *Molecular & cellular proteomics: MCP*, *13*(10), pp. 2765–2775. https://doi.org/10.1074/mcp.O113.036681

Groseclose, M.R. et al. (2008) 'High-throughput proteomic analysis of formalin-fixed paraffin-embedded tissue microarrays using MALDI imaging mass spectrometry', *Proteomics*, *8*, pp. 3715–3724. https://doi.org/10.1002/pmic.200800495

Groseclose, M.R. et al. (2015) 'Imaging MS in toxicology: an investigation of juvenile rat nephrotoxicity associated with dabrafenib administration', *Journal of the american society for mass spectrometry*, *26*(6), pp. 887–898. https://doi.org/10.1007/s13361-015-1103-4

Groseclose, M.R. and Castellino, S. (2013) 'A mimetic tissue model for the quantification of drug distributions by MALDI imaging mass spectrometry', *Analytical chemistry*, *85*(21), pp. 10099–10106. https://doi.org/10.1021/ac400892z

Guo, G. et al. (2021) 'Automated annotation and visualisation of high-resolution spatial proteomic mass spectrometry imaging data using HIT-MAP', *Nature communications*, *12*(1), p. 3241. https://doi.org/10.1038/s41467-021-23461-w

Gustafsson, O.J.R. et al. (2018) 'Balancing sufficiency and impact in reporting standards for mass spectrometry imaging experiments', *GigaScience*, *7*(10). https://doi.org/10.1093/gigascience/giy102

Hale, O.J. and Cooper, H.J. (2021) 'Native mass spectrometry imaging of proteins and protein complexes by Nano-DESI', *Analytical chemistry*, *93*(10), pp. 4619–4627. https://doi.org/10.1021/acs.analchem.0c05277

Hankin, J.A. et al. (2011) 'MALDI mass spectrometric imaging of lipids in rat brain injury models', *Journal of the american society for mass spectrometry*, *22*(6), pp. 1014–1021. https://doi.org/10.1007/s13361-011-0122-z

Hankin, J.A., Barkley, R.M. and Murphy, R.C. (2007) 'Sublimation as a method of matrix application for mass spectrometric imaging', *Journal of the american society for mass spectrometry*, *18*(9), pp. 1646–1652. https://doi.org/10.1016/j.jasms.2007.06.010

Hansen, R.L., Dueñas, M.E. and Lee, Y.J. (2019) 'Sputter-coated metal screening for small molecule analysis and high-spatial resolution imaging in laser desorption ionization mass spectrometry', *Journal of the american society for mass spectrometry*, *30*(2), pp. 299–308. https://doi.org/10.1007/s13361-018-2081-0

Hansen, R.L. and Lee, Y.J. (2017) 'Overlapping MALDI-mass spectrometry imaging for in-parallel MS and MS/MS data acquisition without sacrificing spatial resolution', *Journal of the american society for mass spectrometry*, pp. 1910–1918. https://doi.org/10.1007/s13361-017-1699-7

Harkin, C. et al. (2021) 'On-tissue chemical derivatization in mass spectrometry imaging', *Mass spectrometry reviews*, pp. 1–33. https://doi.org/10.1002/mas.21680

Harrison, J.P. and Berry, D. (2017) 'Vibrational spectroscopy for imaging single microbial cells in complex biological samples', *Frontiers in microbiology*, *8*, p. 675. https://doi.org/10.3389/fmicb.2017.00675

Hartler, J. et al. (2011) 'Lipid data analyzer: unattended identification and quantitation of lipids in LC-MS data', *Bioinformatics*, *27*(4), pp. 572–577. https://doi.org/10.1093/bioinformatics/btq699

Hastings, J. et al. (2016) 'ChEBI in 2016: improved services and an expanding collection of metabolites', *Nucleic acids research*, *44*(D1), pp. D1214–D1219. https://doi.org/10.1093/nar/gkv1031

Haug, K. et al. (2013) 'MetaboLights—an open-access general-purpose repository for metabolomics studies and associated meta-data', *Nucleic acids research*, *41*(D1), D781. https://doi.org/10.1093/nar/gks1004

He, H. et al. (2019) '3,4-dimethoxycinnamic acid as a novel matrix for enhanced in situ detection and imaging of low-molecular-weight compounds in biological tissues by MALDI-MSI', *Analytical chemistry*, *91*(4), pp. 2634–2643. https://doi.org/10.1021/acs.analchem.8b03522

Heijs, B. et al. (2020) 'MALDI-2 for the enhanced analysis of -linked glycans by mass spectrometry imaging', *Analytical chemistry*, *92*(20), pp. 13904–13911. https://doi.org/10.1021/acs.analchem.0c02732

Hermann, J. et al. (2020) 'Sample preparation of formalin-fixed paraffin-embedded tissue sections for MALDI-mass spectrometry imaging', *Analytical and bioanalytical chemistry*, *412*(6), pp. 1263–1275. https://doi.org/10.1007/s00216-019-02296-x

Hermjakob, H. (2006) 'The HUPO proteomics standards initiative —overcoming the fragmentation of proteomics data', *Proteomics*, pp. 34–38. https://doi.org/10.1002/pmic.200600537

Hoffmann, N. et al. (2019) 'mzTab-M: a data standard for sharing quantitative results in mass spectrometry metabolomics', *Analytical chemistry*, *91*(5), pp. 3302–3310. https://doi.org/10.1021/acs.analchem.8b04310

Hoffmann, N., Hartler, J. and Ahrends, R. (2019) 'jmzTab-M: a reference parser, writer, and validator for the proteomics standards initiative mzTab 2.0 metabolomics standard', *Analytical chemistry*, *91*(20), pp. 12615–12618. https://doi.org/10.1021/acs.analchem.9b01987

Horai, H. et al. (2010) 'MassBank: a public repository for sharing mass spectral data for life sciences', *Journal of mass spectrometry*, *45*(7), pp. 703–714. https://doi.org/10.1002/jms.1777

Huang, P. et al. (2020) 'Toward the Rational Design of Universal Dual Polarity Matrix for MALDI Mass Spectrometry', *Analytical Chemistry*, *92* pp. 7139–7145. https://doi.org/10.1021/acs.analchem.0c00570

Huber, F. et al. (2021) 'Spec2Vec: improved mass spectral similarity scoring through learning of structural relationships', *PLoS computational biology*, *17*(2), p. e1008724. https://doi.org/10.1371/journal.pcbi.1008724

Iakab, S.A. et al. (2021) 'Perspective on multimodal imaging techniques coupling mass spectrometry and vibrational spectroscopy: picturing the best of both worlds', *Analytical*

chemistry, *93*(16), pp. 6301–6310. https://doi.org/10.1021/acs.analchem.0c04986

Ifa, D.R. et al. (2007) 'Development of capabilities for imaging mass spectrometry under ambient conditions with desorption electrospray ionization (DESI)', *International journal of mass spectrometry*, *259*(1–3), pp. 8–15. https://doi.org/10.1016/j.ijms.2006.08.003

Jadoul, L. et al. (2015) 'A spiked tissue-based approach for quantification of phosphatidylcholines in brain section by MALDI mass spectrometry imaging', *Analytical and bioanalytical chemistry*, *407*(8), pp. 2095–2106. https://doi.org/10.1007/s00216-014-8232-7

Janda, M. et al. (2021) 'Determination of abundant metabolite matrix adducts illuminates the dark metabolome of MALDI-mass spectrometry imaging datasets', *Analytical chemistry*, *93*(24), pp. 8399–8407. https://doi.org/10.1021/acs.analchem.0c04720

Jaskolla, T.W. and Karas, M. (2011) 'Compelling evidence for Lucky Survivor and gas phase protonation: the unified MALDI analyte protonation mechanism', *Journal of the American Society for Mass Spectrometry*, *22*(6), pp. 976–988. https://doi.org/10.1007/s13361-011-0093-0

Jones, A.R. et al. (2012) 'The mzIdentML data standard for mass spectrometry-based proteomics results', *Molecular & cellular proteomics*, *11*(7), p. M111.014381. https://doi.org/10.1074/mcp.M111.014381

Kaddi, C., Parry, R.M. and Wang, M.D. (2011) 'Hypergeometric similarity measure for spatial analysis in tissue imaging mass spectrometry', *Proceedings of IEEE international conference on bioinformatics and biomedicine*, *2011*, pp. 604–607. https://doi.org/10.1109/BIBM.2011.113

Kanehisa, M. and Goto, S. (2000) 'KEGG: kyoto encyclopedia of genes and genomes', *Nucleic acids research*, *28*(1), pp. 27–30. https://doi.org/10.1093/nar/28.1.27

Karas, M., Glückmann, M. and Schäfer, J. (2000) 'Ionization in matrix-assisted laser desorption/ionization: singly charged molecular ions are the lucky survivors', *Journal of mass spectrometry*, *35*(1) 1–12. https://doi.org/10.1002/(SICI)1096-9888(200001)35:1%3C1::AID-JMS904%3E3.0.CO;2-0

Kaya, I. et al. (2018) 'Dual polarity MALDI imaging mass spectrometry on the same pixel points reveals spatial lipid localizations at high-spatial resolutions in rat small intestine', *Analytical Methods*, *10*, pp. 2428–2435. https://doi.org/10.1039/c8ay00645h

Khatib-Shahidi, S. et al. (2006) 'Direct molecular analysis of whole-body animal tissue sections by imaging MALDI mass spectrometry', *Analytical chemistry*, *78*(18), pp. 6448–6456. https://doi.org/10.1021/ac060788p

Kim, S. et al. (2019) 'PubChem 2019 update: Improved access to chemical data', *Nucleic acids research*, *47*(D1), pp. D1102–D1109. https://doi.org/10.1093/nar/gky1033

Kind, T. et al. (2013) 'LipidBlast in silico tandem mass spectrometry database for lipid identification', *Nature methods*, *10*(8), pp. 755–758. https://doi.org/10.1038/nmeth.2551

Kind, T. and Fiehn, O. (2007) 'Seven golden rules for heuristic filtering of molecular formulas obtained by accurate mass spectrometry', *BMC bioinformatics*, *8*, p. 105. https://doi.org/10.1186/1471-2105-8-105

Kompauer, M., Heiles, S. and Spengler, B. (2017) 'Atmospheric pressure MALDI mass spectrometry imaging of tissues and cells at 1.4-μm lateral resolution', *Nature methods*, *14*(1), pp. 90–96. https://doi.org/10.1038/nmeth.4071

Kune, C. et al. (2019) 'Rapid visualization of chemically related compounds using kendrick mass defect as a filter in mass spectrometry imaging', *Analytical chemistry*, *91*(20), pp. 13112–13118. https://doi.org/10.1021/acs.analchem.9b03333

Kyle, J.E. et al. (2016) 'Uncovering biologically significant lipid isomers with liquid chromatography, ion mobility spectrometry and mass spectrometry', *The analyst*, *141*(5), pp. 1649–1659. https://doi.org/10.1039/c5an02062j

Łącki, M.K. et al. (2021) 'OpenTIMS, TimsPy, and TimsR: open and easy access to tims TOF raw data', *Journal of proteome research*, *20*(4), pp. 2122–2129. https://doi.org/10.1021/acs.jproteome.0c00962

Lanekoff, I. et al. (2013) 'High-speed tandem mass spectrometric in situ imaging by nanospray desorption electrospray ionization mass spectrometry', *Analytical chemistry*, *85*(20), pp. 9596–9603. https://doi.org/10.1021/ac401760s

Lapthorn, C., Pullen, F. and Chowdhry, B.Z. (2013) 'Ion mobility spectrometry-mass spectrometry (IMS-MS) of small molecules: separating and assigning structures to ions', *Mass spectrometry reviews*, *32*(1), pp. 43–71. https://doi.org/10.1002/mas.21349

Lasch, P. and Noda, I. (2017) 'Two-dimensional correlation spectroscopy for multimodal analysis of FT-IR, Raman, and MALDI-TOF MS hyperspectral images with Hamster brain tissue', *Analytical chemistry*, *89*(9), pp. 5008–5016. https://doi.org/10.1021/acs.analchem.7b00332

Lemmon, E.W. et al. (2010) 'NIST standard reference database 23', Reference fluid thermodynamic and transport properties (REFPROP), version, 9. Available at: https://www.nist.gov/system/files/documents/srd/REFPROP8_manua3.htm

Leopold, J. et al. (2018) 'Recent developments of useful MALDI matrices for the mass spectrometric characterization of lipids', *Biomolecules*, *8* p. 173. https://doi.org/10.3390/biom8040173

Li, B. et al. (2019) '3-Aminophthalhydrazide (Luminol) as a matrix for dual-polarity MALDI MS imaging', *Analytical chemistry*, *91*(13), pp. 8221–8228. https://doi.org/10.1021/acs.analchem.9b00803

Lichtman, J.W. and Conchello, J.-A. (2005) 'Fluorescence microscopy', *Nature methods*, *2*(12), pp. 910–919. https://doi.org/10.1038/nmeth817

Lin, J.-R. et al. (2016) 'Cyclic immunofluorescence (CycIF), a highly multiplexed method for single-cell imaging', *Current protocols in chemical biology*, *8*(4), pp. 251–264. https://doi.org/10.1002/cpch.14

Lisec, J. et al. (2006) 'Gas chromatography mass spectrometry–based metabolite profiling in plants', *Nature protocols*, *1*(1), pp. 387–396. https://doi.org/10.1038/nprot.2006.59

Li, X., Liu, D. and Wang, Z. (2011) 'Highly selective recognition of naphthol isomers based on the fluorescence dye-incorporated SH-β-cyclodextrin functionalized gold nanoparticles', *Biosensors & bioelectronics*, *26*(5), pp. 2329–2333. https://doi.org/10.1016/j.bios.2010.10.005

Loos, M. et al. (2015) 'Accelerated isotope fine structure calculation using pruned transition trees', *Analytical chemistry*, *87*(11), pp. 5738–5744. https://doi.org/10.1021/acs.analchem.5b00941

Ly, A. et al. (2016) 'High-mass-resolution MALDI mass spectrometry imaging of metabolites from formalin-fixed paraffin-embedded tissue', *Nature protocols*, *11*(8), pp. 1428–1443. https://doi.org/10.1038/nprot.2016.081

Mainini, V. et al. (2013) 'Detection of high molecular weight proteins by MALDI imaging mass spectrometry', *Molecular bioSystems*, *9*(6), pp. 1101–1107. https://doi.org/10.1039/c2mb25296a

Martens, L. et al. (2011) 'mzML—a community standard for mass spectrometry data', *Molecular & cellular proteomics: MCP*, *10*(1), pp. 1–7. https://doi.org/10.1074/mcp.R110.000133

Mascini, N.E. and Heeren, R.M.A. (2012) 'Protein identification in mass-spectrometry imaging', *Trends in analytical chemistry: TRAC*, *40*, pp. 28–37. https://doi.org/10.1016/j.trac.2012.06.008

McDonnell, L.A. et al. (2008) 'Mass spectrometry image correlation: quantifying colocalization', *Journal of proteome research*, *7*(8), pp. 3619–3627. https://doi.org/10.1021/pr800214d

McDonnell, L.A. et al. (2015) 'Discussion point: reporting guidelines for mass spectrometry imaging', *Analytical and bioanalytical chemistry*, *407*(8), pp. 2035–2045. https://doi.org/10.1007/s00216-014-8322-6

McLafferty, F.W. (1981) 'Tandem mass spectrometry', *Science*, *214*(4518), pp. 280–287. https://doi.org/10.1126/science.7280693

Meier, F. et al. (2015) 'Parallel accumulation-serial fragmentation (PASEF): multiplying sequencing speed and sensitivity by synchronized scans in a trapped ion mobility device', *Journal of proteome research*, *14*(12), pp. 5378–5387. https://doi.org/10.1021/acs.jproteome.5b00932

Meier, F. et al. (2020) 'diaPASEF: parallel accumulation-serial fragmentation combined with data-independent acquisition', *Nature methods*, *17*(12), pp. 1229–1236. https://doi.org/10.1038/s41592-020-00998-0

Mesa Sanchez, D. et al. (2020) 'Ion mobility-mass spectrometry imaging workflow', *Journal of the american society for mass spectrometry*, *31*(12), pp. 2437–2442. https://doi.org/10.1021/jasms.0c00142

Minaee, S. et al. (2021) 'Image segmentation using deep learning: a survey', *IEEE transactions on pattern analysis and machine intelligence*. https://doi.org/10.1109/TPAMI.2021.3059968

Mounfield III, W.P. and Garrett, T.J. (2012) 'Automated MALDI matrix coating system for multiple tissue samples for imaging mass spectrometry', *Journal of the american society for mass spectrometry*, *23*(3), pp. 563–569. https://doi.org/10.1007/s13361-011-0324-4

Muddiman, D.C. and Oberg, A.L. (2005) 'Statistical evaluation of internal and external mass calibration laws utilized in fourier transform ion cyclotron resonance mass spectrometry', *Analytical chemistry*, *77*(8), pp. 2406–2414. https://doi.org/10.1021/ac048258l

Nakamura, J. et al. (2017) 'Spatially resolved metabolic distribution for unraveling the physiological change and responses in tomato fruit using matrix-assisted laser desorption/ionization-mass spectrometry imaging (MALDI-MSI)', *Analytical and bioanalytical chemistry*, *409*(6), pp. 1697–1706. https://doi.org/10.1007/s00216-016-0118-4

Nazari, M. et al. (2018) 'Quantitative mass spectrometry imaging of glutathione in healthy and cancerous hen ovarian tissue sections by infrared matrix-assisted laser desorption electrospray ionization (IR-MALDESI)', *The Analyst*, *143*, pp. 654–661. https://doi.org/10.1039/c7an01828b

Neumann, N.K.N. et al. (2014) 'Automated LC-HRMS(/MS) approach for the annotation of fragment ions derived from stable isotope labeling-assisted untargeted metabolomics', *Analytical chemistry*, *86*(15), pp. 7320–7327. https://doi.org/10.1021/ac501358z

Nguyen, S.N. et al. (2018) 'Towards high-resolution tissue imaging using nanospray desorption electrospray ionization mass spectrometry coupled to shear force microscopy', *Journal of the American Society for Mass Spectrometry*, *29*(2), pp. 316–322. https://doi.org/10.1007/s13361-017-1750-8

Niedermeyer, T.H.J. and Strohalm, M. (2012) 'mMass as a software tool for the annotation of cyclic peptide tandem mass spectra', *PloS one*, *7*(9), p. e44913. https://doi.org/10.1371/journal.pone.0044913

Niehaus, M. et al. (2019) 'Transmission-mode MALDI-2 mass spectrometry imaging of cells and tissues at subcellular resolution', *Nature methods*, *16*(9), pp. 925–931. https://doi.org/10.1038/s41592-019-0536-2

Nizioł, J. and Ruman, T. (2013) 'Surface-transfer mass spectrometry imaging on a monoisotopic silver nanoparticle enhanced target', *Analytical chemistry*, *85*(24), pp. 12070–12076. https://doi.org/10.1021/ac4031658

Norris, J.L. et al. (2007) 'Processing MALDI mass spectra to improve mass spectral direct tissue analysis', *International journal of mass spectrometry*, *260*(2–3), pp. 212–221. https://doi.org/10.1016/j.ijms.2006.10.005

Norris, J.L. and Caprioli, R.M. (2013) 'Analysis of tissue specimens by matrix-assisted laser desorption/ionization imaging mass spectrometry in biological and clinical research', *Chemical reviews*, *113*(4), pp. 2309–2342. https://doi.org/10.1021/cr3004295

Nothias, L.-F. et al. (2020) 'Feature-based molecular networking in the GNPS analysis environment', *Nature methods*, *17*(9), pp. 905–908. https://doi.org/10.1038/s41592-020-0933-6

Novák, J., Škríba, A. and Havlíček, V. (2020) 'CycloBranch 2: molecular formula annotations applied to imzML data sets in bimodal fusion and LC-MS data files', *Analytical chemistry*, *92*(10), pp. 6844–6849. https://doi.org/10.1021/acs.analchem.0c00170

Ntshangase, S. et al. (2019) 'Spatial distribution of elvitegravir and tenofovir in rat brain tissue: Application of matrix-assisted laser desorption/ionization mass spectrometry imaging and liquid chromatography/tandem mass spectrometry', *Rapid communications in mass spectrometry: RCM*, *33*(21), pp. 1643–1651. https://doi.org/10.1002/rcm.8510

Ogrinc Potočnik, N. et al. (2014) 'Gold sputtered fiducial markers for combined secondary ion mass spectrometry and MALDI imaging of tissue samples', *Analytical chemistry*, *86*(14), pp. 6781–6785. https://doi.org/10.1021/ac500308s

Otter, D.W., Medina, J.R. and Kalita, J.K. (2021) 'A survey of the usages of deep learning for natural language processing', *IEEE transactions on neural networks and learning systems*, *32*(2), pp. 604–624. https://doi.org/10.1109/TNNLS.2020.2979670

Ovchinnikova, K., Stuart, L., et al. (2020) 'ColocML: machine learning quantifies co-localization between mass spectrometry images', *Bioinformatics*, *36*(10), pp. 3215–3224. https://doi.org/10.1093/bioinformatics/btaa085

Ovchinnikova, K., Kovalev, V., et al. (2020) 'OffsampleAI: artificial intelligence approach to recognize off-sample mass spectrometry images', *BMC bioinformatics*, *21*(1), p. 129. https://doi.org/10.1186/s12859-020-3425-x

Paine, M.R.L. et al. (2019) 'Three-dimensional mass spectrometry imaging identifies lipid markers of medulloblastoma metastasis', *Scientific reports*, *9*(1), p. 2205. https://doi.org/10.1038/s41598-018-38257-0

Palmer, A. et al. (2015) 'Using collective expert judgements to evaluate quality measures of mass spectrometry images', Bioinformatics, 31, pp. i375–i384. https://doi.org/10.1093/bioinformatics/btv266

Palmer, A. et al. (2016) 'FDR-controlled metabolite annotation for high-resolution imaging mass spectrometry', *Nature methods*, *14*(1), pp. 57–60. https://doi.org/10.1038/nmeth.4072

Patti, G.J., Yanes, O. and Siuzdak, G. (2012) 'Innovation: metabolomics: the apogee of the omics trilogy', *Nature reviews. Molecular cell biology*, *13*(4), pp. 263–269. https://doi.org/10.1038/nrm3314

Perdian, D.C. and Lee, Y.J. (2010) 'Imaging MS methodology for more chemical information in less data acquisition time utilizing a hybrid linear ion trap–orbitrap mass spectrometer', *Analytical Chemistry*, *82*, pp. 9393–9400. https://doi.org/10.1021/ac102017q

Phan, N.T.N. et al. (2016) 'Laser desorption ionization mass spectrometry imaging of Drosophila brain using matrix sublimation versus modification with nanoparticles', *Analytical chemistry*, *88*(3), pp. 1734–1741. https://doi.org/10.1021/acs.analchem.5b03942

Phillips, L., Gill, A.J. and Baxter, R.C. (2019) 'Novel prognostic markers in triple-negative breast cancer discovered by MALDI-mass spectrometry imaging', *Frontiers in oncology*, *9*, p. 379. https://doi.org/10.3389/fonc.2019.00379

Picache, J.A. et al. (2019) 'Collision cross section compendium to annotate and predict multi-omic compound identities', *Chemical science*, *10*(4), pp. 983–993. https://doi.org/10.1039/c8sc04396e

Pietrowska, M. et al. (2016) 'Tissue fixed with formalin and processed without paraffin embedding is suitable for imaging of both peptides and lipids by MALDI-IMS', *Proteomics*, *16*(11–12), pp. 1670–1677. https://doi.org/10.1002/pmic.201500424

Piga I. et al. (2019) 'Ultra-high resolution MALDI-FTICR-MSI analysis of intact proteins in mouse and human pancreas tissue'. *International journal of mass spectrometry*, *437*, pp. 10–16. https://doi.org/10.1016/j.ijms.2017.11.001

Pirman, D.A. et al. (2013) 'Identifying tissue-specific signal variation in MALDI mass spectrometric imaging by use of an internal standard', *Analytical chemistry*, *85*(2), pp. 1090–1096. https://doi.org/10.1021/ac3029618

Pitt, J.J. (2009) 'Principles and applications of liquid chromatography-mass spectrometry in clinical biochemistry', *The Clinical biochemist. Reviews/Australian Association of Clinical Biochemists*, *30*(1), pp. 19–34. Available at: https://www.ncbi.nlm.nih.gov/pubmed/19224008.

Polanska, J. et al. (2012) 'Gaussian mixture decomposition in the analysis of MALDI-TOF spectra', *Expert Systems*, *29*(3), pp. 216–231. https://doi.org/10.1111/j.1468-0394.2011.00582.x

Pornwilard, M.-M. et al. (2013) 'Bioimaging of copper deposition in Wilson's diseases mouse liver by laser ablation inductively coupled plasma mass spectrometry imaging (LA-ICP-MSI)', *International journal of mass spectrometry*, *354–355*, pp. 281–287. https://doi.org/10.1016/j.ijms.2013.07.006

Porta Siegel, T. et al. (2018) 'Mass spectrometry imaging and integration with other imaging modalities for greater molecular understanding of biological tissues', *Molecular imaging and biology: The official publication of the academy of molecular imaging*, *20*(6), pp. 888–901. https://doi.org/10.1007/s11307-018-1267-y

Race, A.M. et al. (2021) 'Deep learning-based annotation transfer between molecular imaging modalities: an automated workflow for multimodal data integration', *Analytical chemistry*, *93*, pp. 3061–3071. https://doi.org/10.1021/acs.analchem.0c02726

Ràfols, P., Vilalta, D., Torres, S., et al. (2018) 'Assessing the potential of sputtered gold nanolayers in mass spectrometry imaging for metabolomics applications', *PLoS One*, *13*(12), p. e0208908. https://doi.org/10.1371/journal.pone.0208908

Ràfols, P., Castillo, E. del, et al. (2018) 'Novel automated workflow for spectral alignment and mass calibration in MS imaging using a sputtered Ag nanolayer', *Analytica chimica acta*, *1022*, pp. 61–69. https://doi.org/10.1016/j.aca.2018.03.031

Ràfols, P., Vilalta, D., Brezmes, J., et al. (2018) 'Signal preprocessing, multivariate analysis and software tools for MA(LDI)-TOF mass spectrometry imaging for biological applications', *Mass Spectrometry Reviews*, *37*, pp. 281–306. https://doi.org/10.1002/mas.21527

Ràfols, P. et al. (2020) 'RMSIproc: An R package for mass spectrometry imaging data processing', *Bioinformatics*, *36*(11), pp. 3618–3619. https://doi.org/10.1093/bioinformatics/btaa142

Ren, J.-L. et al. (2018) 'Advances in mass spectrometry-based metabolomics for investigation of metabolites', *RSC advances*, *8*(40), pp. 22335–22350. https://doi.org/10.1039/C8RA01574K

Robichaud, G. et al. (2013) 'MSiReader: an open-source interface to view and analyze high resolving power MS imaging files on Matlab platform', *Journal of the American Society for Mass Spectrometry*, *24*(5), pp. 718–721. https://doi.org/10.1007/s13361-013-0607-z

Ruttkies, C. et al. (2016) 'MetFrag relaunched: incorporating strategies beyond in silico fragmentation', *Journal of cheminformatics*, *8*, p. 3. https://doi.org/10.1186/s13321-016-0115-9

Ryan, D.J., Spraggins, J.M. and Caprioli, R.M. (2019) 'Protein identification strategies in MALDI imaging mass spectrometry: a brief review', *Current opinion in chemical biology*, *48*, pp. 64–72. https://doi.org/10.1016/j.cbpa.2018.10.023

Rzagalinski, I. and Volmer, D.A. (2017) 'Quantification of low molecular weight compounds by MALDI imaging mass spectrometry—a tutorial review', *Biochimica et Biophysica Acta: Proteins and Proteomics*, *1865*(7), pp. 726–739. https://doi.org/10.1016/j.bbapap.2016.12.011

Sabine Becker, J. (2013) 'Imaging of metals in biological tissue by laser ablation inductively coupled plasma mass spectrometry (LA-ICP-MS): state of the art and future developments', *Journal of mass spectrometry: JMS*, *48*(2), pp. 255–268. https://doi.org/10.1002/jms.3168

Salek, R. (2019) 'Data Sharing and Standards', in *Metabolomics*. Chapman and Hall/CRC, pp. 235–252. Available at: https://

www.taylorfrancis.com/chapters/edit/10.1201/9781315370583-10/data-sharing-standards-reza-salek

Salek, R.M. et al. (2013) 'The role of reporting standards for metabolite annotation and identification in metabolomic studies', *GigaScience*. *2*(1), p. 13. https://doi.org/10.1186/2047-217x-2-13

Sans, M., Feider, C.L. and Eberlin, L.S. (2018) 'Advances in mass spectrometry imaging coupled to ion mobility spectrometry for enhanced imaging of biological tissues', *Current opinion in chemical biology*, *42*, pp. 138–146. https://doi.org/10.1016/j.cbpa.2017.12.005

Schramm, T. et al. (2012) 'ImzML—a common data format for the flexible exchange and processing of mass spectrometry imaging data', *Journal of proteomics*, *75*(16), pp. 5106–5110. https://doi.org/10.1016/j.jprot.2012.07.026

Schrimpe-Rutledge, A.C. et al. (2016) 'Untargeted metabolomics strategies—challenges and emerging directions', *Journal of the american society for mass spectrometry*, *27*(12), pp. 1897–1905. https://doi.org/10.1007/s13361-016-1469-y

Schulz, S. et al. (2019) 'Advanced MALDI mass spectrometry imaging in pharmaceutical research and drug development', *Current opinion in biotechnology*, *55*, pp. 51–59. https://doi.org/10.1016/j.copbio.2018.08.003

Schymanski, E.L. et al. (2014) 'Identifying small molecules via high resolution mass spectrometry: communicating confidence', *Environmental science & technology*, *48*(4), pp. 2097–2098. https://doi.org/10.1021/es5002105

Sementé, L. et al. (2021) 'rMSIannotation: A peak annotation tool for mass spectrometry imaging based on the analysis of isotopic intensity ratios', *Analytica chimica acta*, *1171*, p. 338669. https://doi.org/10.1016/j.aca.2021.338669

Shariatgorji, M. et al. (2012) 'Deuterated matrix-assisted laser desorption ionization matrix uncovers masked mass spectrometry imaging signals of small molecules', *Analytical chemistry*, *84*(16), pp. 7152–7157. https://doi.org/10.1021/ac301498m

Shariatgorji, M. et al. (2014) 'Direct targeted quantitative molecular imaging of neurotransmitters in brain tissue sections', *Neuron*, *84*(4), pp. 697–707. https://doi.org/10.1016/j.neuron.2014.10.011

Shariatgorji, M. et al. (2015) 'Pyrylium salts as reactive matrices for MALDI-MS imaging of biologically active primary amines', *Journal of the american society for mass spectrometry*, *26*(6), pp. 934–939. https://doi.org/10.1007/s13361-015-1119-9

Shariatgorji, R. et al. (2020) 'Bromopyrylium derivatization facilitates identification by mass spectrometry imaging of monoamine neurotransmitters and small molecule neuroactive compounds', *Journal of the american society for mass spectrometry*, *31*(12), pp. 2553–2557. https://doi.org/10.1021/jasms.0c00166

Shobo, A. et al. (2016) 'MALDI MSI and LC-MS/MS: towards preclinical determination of the neurotoxic potential of fluoroquinolones', *Drug testing and analysis*, *8*(8), pp. 832–838. https://doi.org/10.1002/dta.1862

Signor, L. et al. (2007) 'Analysis of erlotinib and its metabolites in rat tissue sections by MALDI quadrupole time-of-flight mass spectrometry', *Journal of mass spectrometry*, *42*(7), pp. 900–909. https://doi.org/10.1002/jms.1225

Smets, T. et al. (2019) 'Evaluation of distance metrics and spatial autocorrelation in uniform manifold approximation and projection applied to mass spectrometry imaging data', *Analytical chemistry*, *91*(9), pp. 5706–5714. https://doi.org/10.1021/acs.analchem.8b05827

Smith, C.A. et al. (2005) 'METLIN: a metabolite mass spectral database', *Therapeutic drug monitoring*, *27*(6), pp. 747–751. https://doi.org/10.1097/01.ftd.0000179845.53213.39

Smith, D.F. et al. (2012) 'Advanced mass calibration and visualization for FT-ICR mass spectrometry imaging', *Journal of the american society for mass spectrometry*, *23*(11), pp. 1865–1872. https://doi.org/10.1007/s13361-012-0464-1

Soltwisch, J. et al. (2015) 'Mass spectrometry imaging with laser-induced postionization', *Science*, *348*(6231), pp. 211–215. https://doi.org/10.1126/science.aaa1051

Spengler, B. (2015) 'Mass spectrometry imaging of biomolecular information', *Analytical chemistry*, *87*(1), pp. 64–82. https://doi.org/10.1021/ac504543v

Steven, R.T. et al. (2019) 'Construction and testing of an atmospheric-pressure transmission-mode matrix assisted laser desorption ionisation mass spectrometry imaging ion source with plasma ionisation enhancement', *Analytica chimica acta*, *1051*, pp. 110–119. https://doi.org/10.1016/j.aca.2018.11.003

Stoeckli M., Staab D., Schweitzer A. (2007) 'Compound and metabolite distribution measured by MALDI mass spectrometric imaging in whole-body tissue sections'. *International journal of mass spectrometry*, *260*(2–3), pp. 195–202. https://doi.org/10.1016/j.ijms.2006.10.007

Sud, M. et al. (2007) 'LMSD: LIPID MAPS structure database', *Nucleic acids research*, *35*(Database issue), pp. D527–32. https://doi.org/10.1093/nar/gkl838

Sud, M. et al. (2016) 'Metabolomics workbench: an international repository for metabolomics data and metadata, metabolite standards, protocols, tutorials and training, and analysis tools', *Nucleic acids research*, *44*(D1), pp. D463–70. https://doi.org/10.1093/nar/gkv1042

Sumner, L.W. et al. (2007) 'Proposed minimum reporting standards for chemical analysis', *Metabolomics: Official journal of the metabolomic society*, *3*(3), pp. 211–221. https://doi.org/10.1007/s11306-007-0082-2

Sünderhauf, N. et al. (2018) 'The limits and potentials of deep learning for robotics', *The international journal of robotics research*, *37*(4–5), pp. 405–420. https://doi.org/10.1177/0278364918770733

Sun, N. et al. (2018) 'High-resolution tissue mass spectrometry imaging reveals a refined functional anatomy of the human adult adrenal gland', *Endocrinology*, *159*(3), pp. 1511–1524. https://doi.org/10.1210/en.2018-00064

Swales, J.G. et al. (2015) 'Mapping drug distribution in brain tissue using liquid extraction surface analysis mass spectrometry imaging', *Analytical chemistry*, *87*(19), pp. 10146–10152. https://doi.org/10.1021/acs.analchem.5b02998

Takeo, E. et al. (2019) 'Tandem mass spectrometry imaging reveals distinct accumulation patterns of steroid structural isomers in human adrenal glands', *Analytical chemistry*, *91*(14), pp. 8918–8925. https://doi.org/10.1021/acs.analchem.9b00619

Thomas, A. et al. (2012) 'Sublimation of new matrix candidates for high spatial resolution imaging mass spectrometry of lipids: enhanced information in both positive and negative polarities after 1,5-diaminonapthalene deposition', *Analytical*

chemistry, *84*(4), pp. 2048–2054. https://doi.org/10.1021/ac2033547

Tortorella, S. et al. (2020) 'Lipostar MSI: comprehensive, vendor-neutral software for visualization, data analysis, and automated molecular identification in mass spectrometry imaging', *Journal of the american society for mass spectrometry*, *31*(1), pp. 155–163. https://doi.org/10.1021/jasms.9b00034

Touboul, D. and Brunelle, A. (2016) 'What more can TOF-SIMS bring than other MS imaging methods?', *Bioanalysis*, *8*(5), pp. 367–369. https://doi.org/10.4155/bio.16.11

Towers, M.W. et al. (2018) 'Optimised desorption electrospray ionisation mass spectrometry imaging (DESI-MSI) for the analysis of proteins/peptides directly from tissue sections on a travelling wave ion mobility Q-ToF', *Journal of the american society for mass spectrometry*, pp. 2456–2466. https://doi.org/10.1007/s13361-018-2049-0

Trede, D. et al. (2012) 'O5. SCiLS Lab: software for analysis and interpretation of large MALDI-IMS datasets', *OurCon 2012*, p. 50. Available at: https://orbi.uliege.be/bitstream/2268/131796/1/Book%20of%20abstractsOurCon2012_v1.3%20with%20covers.pdf#page=51

Trimpin, S. et al. (2009) 'Field-free transmission geometry atmospheric pressure matrix-assisted laser desorption/ionization for rapid analysis of unadulterated tissue samples', *Rapid communications in mass spectrometry*, *23*(18), pp. 3023–3027. https://doi.org/10.1002/rcm.4213

Tsugawa, H. et al. (2015) 'MS-DIAL: data-independent MS/MS deconvolution for comprehensive metabolome analysis', *Nature methods*, *12*(6), pp. 523–526. https://doi.org/10.1038/nmeth.3393

Tuck, M. et al. (2021) 'Multimodal imaging based on vibrational spectroscopies and mass spectrometry imaging applied to biological tissue: a multiscale and multiomics review', *Analytical chemistry*, *93*(1), pp. 445–477. https://doi.org/10.1021/acs.analchem.0c04595

Unsihuay, D., Mesa Sanchez, D. and Laskin, J. (2021) 'Quantitative mass spectrometry imaging of biological systems', *Annual review of physical chemistry*, *72*, pp. 307–329. https://doi.org/10.1146/annurev-physchem-061020-053416

Uslu, A. et al. (2017) 'Imidazole/benzimidazole-modified cyclotriphosphazenes as highly selective fluorescent probes for Cu 2+: synthesis, configurational isomers, and crystal structures', *Dalton transactions: A journal of inorganic chemistry*, *46*(28), pp. 9140–9156. Available at: https://pubs.rsc.org/en/content/articlehtml/2017/dt/c7dt01134b

Vaysse, P.-M. et al. (2017) 'Mass spectrometry imaging for clinical research—latest developments, applications, and current limitations', *The Analyst*, pp. 2690–2712. https://doi.org/10.1039/c7an00565b

Vos, D.R.N. et al. (2019) 'Class-specific depletion of lipid ion signals in tissues upon formalin fixation', *International journal of mass spectrometry*, *446*, p. 116212. https://doi.org/10.1016/j.ijms.2019.116212

Wäldchen, F. et al. (2020) 'Multifunctional reactive MALDI matrix enabling high-lateral resolution dual polarity MS imaging and lipid C=C position-resolved MS imaging', *Analytical chemistry*, *92*(20), pp. 14130–14138. https://doi.org/10.1021/acs.analchem.0c03150

Wäldchen, F., Spengler, B. and Heiles, S. (2019) 'Reactive matrix-assisted laser desorption/ionization mass spectrometry imaging using an intrinsically photoreactive paternò-büchi matrix for double-bond localization in isomeric phospholipids', *Journal of the american chemical society*, *141*(30), pp. 11816–11820. https://doi.org/10.1021/jacs.9b05868

Walzer, M. et al. (2013) 'The mzquantml data standard for mass spectrometry--based quantitative studies in proteomics', *Molecular & cellular proteomics*, *12*(8), pp. 2332–2340. Available at: https://www.mcponline.org/article/S1535-9476(20)32541-X/abstract

Wang, L. et al. (2019) 'Peak annotation and verification engine for untargeted LC-MS metabolomics', *Analytical chemistry*, *91*(3), pp. 1838–1846. https://doi.org/10.1021/acs.analchem.8b03132

Why the metabolism field risks missing out on the AI revolution' (2019) *Nature metabolism*, *1*(10), pp. 929–930. https://doi.org/10.1038/s42255-019-0133-9

Wishart, D.S. et al. (2018) 'HMDB 4.0: the human metabolome database for 2018', *Nucleic acids research*, *46*(D1), pp. D608–D617. https://doi.org/10.1093/nar/gkx1089

Wishart, D.S. (2019) 'Metabolomics for investigating physiological and pathophysiological processes', *Physiological reviews*, *99*(4), pp. 1819–1875. https://doi.org/10.1152/physrev.00035.2018

Wisztorski, M. et al. (2010) 'MALDI direct analysis and imaging of frozen versus FFPE tissues: what strategy for which sample?', *Methods in molecular biology*, *656*, pp. 303–322. https://doi.org/10.1007/978-1-60761-746-4_18

Xian, F., Hendrickson, C.L. and Marshall, A.G. (2012) 'High resolution mass spectrometry', *Analytical chemistry*, *84*(2), pp. 708–719. https://doi.org/10.1021/ac203191t

Xue, J. et al. (2020) 'Enhanced in-source fragmentation annotation enables novel data independent acquisition and autonomous METLIN molecular identification', *Analytical chemistry*, *92*(8), pp. 6051–6059. https://doi.org/10.1021/acs.analchem.0c00409

Xu, J. et al. (2019) 'Integrated UPLC-Q/TOF-MS technique and MALDI-MS to study of the efficacy of YiXinshu capsules against heart failure in a rat model', *Frontiers in pharmacology*, *10*, p. 1474. https://doi.org/10.3389/fphar.2019.01474

Yagnik, G.B., Korte, A.R. and Lee, Y.J. (2013) 'Multiplex mass spectrometry imaging for latent fingerprints', *Journal of mass spectrometry*, *48*(1), pp. 100–104. https://doi.org/10.1002/jms.3134

Yang, J. and Caprioli, R.M. (2011) 'Matrix sublimation/recrystallization for imaging proteins by mass spectrometry at high spatial resolution', *Analytical chemistry*, *83*(14), pp. 5728–5734. https://doi.org/10.1021/ac200998a

Ye, H. et al. (2013) 'MALDI mass spectrometry-assisted molecular imaging of metabolites during nitrogen fixation in the Medicago truncatula-Sinorhizobium meliloti symbiosis', *The plant journal: For cell and molecular biology*, *75*(1), pp. 130–145. https://doi.org/10.1111/tpj.12191

Ye, H. et al. (2014) 'Top-down proteomics with mass spectrometry imaging: a pilot study towards discovery of biomarkers for neurodevelopmental disorders', *PLoS One*, *9*(4), p. e92831. https://doi.org/10.1371/journal.pone.0092831

Yoon, S. and Lee, T.G. (2018) 'Biological tissue sample preparation for time-of-flight secondary ion mass spectrometry (ToF–SIMS) imaging', *Nano Convergence*. *5*, 24. https://doi.org/10.1186/s40580-018-0157-y

Zavalin, A. et al. (2012) 'Direct imaging of single cells and tissue at sub-cellular spatial resolution using transmission geometry MALDI MS', *Journal of mass spectrometry*, *47*(11), p. i. https://doi.org/10.1002/jms.3132

Zavalin, A. et al. (2015) 'Tissue protein imaging at 1 μm laser spot diameter for high spatial resolution and high imaging speed using transmission geometry MALDI TOF MS', *Analytical and bioanalytical chemistry*, *407*(8), pp. 2337–2342. https://doi.org/10.1007/s00216-015-8532-6

Zhang, G. et al. (2020) 'DESI-MSI and METASPACE indicates lipid abnormalities and altered mitochondrial membrane components in diabetic renal proximal tubules', *Metabolomics: Official journal of the metabolomic society*, *16*(1), p. 11. https://doi.org/10.1007/s11306-020-1637-8

Zhang, H. et al. (2020) 'On-tissue derivatization with Girard's reagent P enhances n-glycan signals for formalin-fixed paraffin-embedded tissue sections in MALDI mass spectrometry imaging', *Analytical chemistry*, *92*(19), pp. 13361–13368. https://doi.org/10.1021/acs.analchem.0c02704

Zhang, H. et al. (2021) 'Quantification and molecular imaging of fatty acid isomers from complex biological samples by mass spectrometry', *Chemical science*, *12*(23), pp. 8115–8122. https://doi.org/10.1039/d1sc01614h

Zhang, L.-K. et al. (2005) 'Accurate mass measurements by Fourier transform mass spectrometry', *Mass spectrometry reviews*, *24*(2), pp. 286–309. https://doi.org/10.1002/mas.20013

Zhang, W. et al. (2021) 'Spatially aware clustering of ion images in mass spectrometry imaging data using deep learning', *Analytical and bioanalytical chemistry*, pp. 2803–2819. https://doi.org/10.1007/s00216-021-03179-w

Zhang, Z., Kuang, J. and Li, L. (2013) 'Liquid chromatography-matrix-assisted laser desorption/ionization mass spectrometric imaging with sprayed matrix for improved sensitivity, reproducibility and quantitation', *The analyst*, *138*(21), pp. 6600–6606. https://doi.org/10.1039/c3an01225e

Zhan, L. et al. (2021) 'MALDI-TOF/TOF tandem mass spectrometry imaging reveals non-uniform distribution of disaccharide isomers in plant tissues', *Food chemistry*, *338*, p. 127984. https://doi.org/10.1016/j.foodchem.2020.127984

Zhou, Q., Fülöp, A. and Hopf, C. (2021) 'Recent developments of novel matrices and on-tissue chemical derivatization reagents for MALDI-MSI', *Analytical and bioanalytical chemistry*, *413*(10), pp. 2599–2617. https://doi.org/10.1007/s00216-020-03023-7

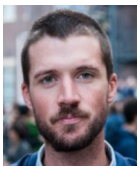
Zhou, R. and Basile, F. (2017) 'Plasmonic thermal decomposition/digestion of proteins: a rapid on-surface protein digestion technique for mass spectrometry imaging', *Analytical chemistry*, *89*(17), pp. 8704–8712. https://doi.org/10.1021/acs.analchem.7b00430

Zhou, Z. et al. (2020) 'Ion mobility collision cross-section atlas for known and unknown metabolite annotation in untargeted metabolomics', *Nature communications*, *11*(1), p. 4334. https://doi.org/10.1038/s41467-020-18171-8

Zubarev, R.A. and Makarov, A. (2013) 'Orbitrap mass spectrometry', *Analytical chemistry*, *85*(11), pp. 5288–5296. https://doi.org/10.1021/ac4001223

Züllig, T. and Köfeler, H.C. (2021) 'High resolution mass spectrometry in lipidomics', *Mass spectrometry reviews*, *40*(3), pp. 162–176. https://doi.org/10.1002/mas.21627

AUTHOR BIOGRAPHIES



Gerard Baquer received his BSc in Electronics Engineering from Universitat Rovira i Virgili (URV) and MSc in Biomedical Electronics from Technical University of Delft (TU Delft). PhD candidate in the Department of Electronic Engineering at the URV since 2018 involved in software development for mass spectrometry imaging. His main research interests are peak annotation (i.e., matrix-related signals, in-source fragments) and the registration of MSI to other imaging modalities (i.e., vibrational spectroscopy imaging, histopathology, immunofluorescence and magnetic resonance imaging).

Lluc Sementé received his BSc in Electronics Engineering and MSc in Electronic Systems from the Universitat Rovira i Virgili (URV). PhD candidate in the Department of Electronic Engineering at the URV since 2018 involved in software development for mass spectrometry imaging (MSI). Currently interested in assessing the potential of soft clustering algorithms in MSI workflows and developing peak annotation tools for MSI.

Toufik Mahamdi received his BSc in Pharmacy from the University of Batna (Algeria) and MSc in Bio and Pharmaceutical Analysis from the University of applied sciences Fresenius (Germany). PhD candidate in the Department of Electronic Engineering at the Universitat Rovira i Virgili (URV). His current research work is focused on analytical method development and sample preparation based on dry deposition for MALDI mass spectrometry imaging of biological samples.

Xavier Correig is a PhD in Telecommunication Engineering from the Polytechnic University of Catalonia. He is a Professor at the Department of Electronic Engineering, Universitat Rovira i Virgili (URV). He directs the Metabolomics Platform, a scientific infrastructure owned by URV and CIBERDEM, whose mission is to provide research in metabolomics to the clinical and biomedicine research groups. The present research lines are the development of new solid-state surfaces and signal processing algorithms for LDI-MS Imaging. He has published more than 200

papers in indexed journals, several invited papers at national conferences and has participated in more than 80 competitive projects.

Pere Ràfols is a PhD in technologies to apply mass spectrometry imaging (MSI) at metabolomics studies from Universitat Rovira i Virgili (URV). He contributes to the development of methods targeted to the detection of low-weight compounds in MSI with very high lateral resolution and non-interfering MS signals. He is the main developer of two well-known R packages targeted to MSI data-analysis: rMSI and rMSIproc. He also focuses on MSI data handling, spectral preprocessing, automated peak annotation and statistical analysis to allow a fast and robust data interpretation.

María García-Altares is a PhD in Analytical Chemistry from Universitat Rovira i Virgili (URV) with a strong track record of international collaborations as an independent researcher. She has received several competitive grants to carry out her research, including an EU Marie Skłodowska-Curie IF. She has applied her broad expertise in mass spectrometry to characterize chemical mediators that define complex biosystems, like the interactions among microbes in the environment and the relationships among cells and tissues in our body. Her current projects aim to develop novel methods for sample preparation, acquisition, and data analysis by Mass Spectrometry Imaging to investigate complex biochemical processes in biomedicine, ecology, and the environment.

SUPPORTING INFORMATION

Additional supporting information can be found online in the Supporting Information section at the end of this article.

How to cite this article: Baquer G, Sementé L, Mahamdi T, Correig X, Ràfols P, García-Altares M. What are we imaging? Software tools and experimental strategies for annotation and identification of small molecules in mass spectrometry imaging. *Mass Spectrometry Reviews*, (2022);e21794. https://doi.org/10.1002/mas.21794