Using biological signals for mass recalibration of mass spectrometry imaging data

Raphaël La Rocca[1], Christopher Kune[1], Mathieu Tiquet[1], Lachlan Stuart[2], Theodore Alexandrov[2,3], Edwin De Pauw[1], Loïc Quinton[1]

[1]Mass Spectrometry Laboratory, MolSys Research Unit, University of Liège, Liège, Belgium
[2]Structural and Computational Biology Unit, European Molecular Biology Laboratory, Heidelberg, Germany.
[3]Skaggs School of Pharmacy and Pharmaceutical Sciences, University of California at San Diego, San Diego, California 92093, United States

Abstract

Mass spectrometry imaging (MSI) is a powerful and convenient method to reveal the spatial chemical composition of different biological samples. The molecular annotation of the detected signals is only possible when high mass accuracy is maintained across the entire image and the *m/z* range. However, the heterogeneous molecular composition of biological samples could result in fluctuations in the detected *m/z*-values, called mass shift. Mass shifts impact the interpretability of the detected signals by decreasing the number of annotations and by affecting the spatial consistency and accuracy of ion images. The use of internal calibration is known to offer the best solution to avoid, or at least to reduce, mass shifts. The selection of internal calibrating signals for a global MSI acquisition is not trivial, prone to false positive detection of calibrating signals and therefore to poor recalibration. To fill this gap, this work describes an algorithm that recalibrates each spectrum individually by estimating its mass shift with the help of a list of internal calibrating ions generated automatically in a data-adaptive manner. The method exploits RANSAC (*Random Sample Consensus*) algorithm, to select, in a robust manner, the experimental signal corresponding to internal calibrating signals by filtering out calibration points with infrequent mass errors and by using the remaining points to estimate a linear model of the mass shifts. We applied the method to a zebrafish whole body section acquired at high mass resolution to demonstrate the impact of mass shift on data analysis and the capacity of our algorithm to recalibrate MSI data. We illustrate the broad applicability of the method by recalibrating 31 different public MSI datasets from METASPACE from various samples and types of MSI and show that our recalibration significantly increases the numbers of METASPACE annotations, especially the high-confident annotations at a low false discovery rate.

Introduction

In biology and medicine, the *in situ* determination of the molecular environment is of prime importance to understand biological processes and pathology evolution [1]. This growing interest pushes the development of analytical methods that correlate spatial distribution with the detection of different biological molecules such as peptides [2], metabolites and lipids [3]. Mass spectrometry imaging (MSI) has been demonstrated particularly powerful as it can rapidly reveal, in an untargeted manner, a wide range of compounds present in small amounts in biological samples as various as whole body sections [4,5], tissue sections, bacteria colonies [6], plants [7] or again single cells [8,9]. MSI is a particular application of Mass spectrometry (MS) in which spectra are recorded, usually thanks to MALDI or to a lesser extent DESI [10], at different positions, called pixel, over a sample forming a 2D image. A pixel is then a spectrum identified by its (x, y) coordinates, containing *m/z* values (channels) and their intensities [11,12].

If high resolution in mass is mandatory to distinguish isobaric and quasi-isobaric compounds in complex mixtures and to exploit isotopic signatures, high accuracy of *m/z* measurement is essential for a confident identification of compounds, leading to a deeper interpretation of the molecular content of samples. Therefore, high resolution mass spectrometry (HRMS), combining both high resolution and mass accuracy, is promoted for annotating molecular signatures of biological samples [13]. However, MSI data acquired with HRMS have shown to suffer from inconsistent variation of measured *m/z* values, from a pixel to another. The analysis of MSI images is based on a so-called average spectrum, representing all the ions detected in the image, i.e., summing each spectrum from every pixel. This pixel-to-pixel fluctuation reaches up to several ppm and strongly influences the global accuracy and resolution of the MSI average spectra [14–16]. It has therefore a crucial impact on MSI interpretation. Indeed, mass shifts affect the quality of the results by (i)

decreasing the number of identifications, (ii) increasing the number of false identifications, (iii) reducing the confidence of each identification and (iv) by impacting the capacity to reconstruct the proper spatial distributions of specific detected species. We have earlier shown that the automated metabolite annotation for MSI data critically depends on the *m/z* accuracy and requires the accuracy of at least +3 ppm in *m/z* [17]. Mass shifts are, in summary, strongly weakening the advantages of HRMS instruments for MSI data analysis.

The mass accuracy depends mainly on the quality of the MS calibration of the instrument while the mass resolving power which is linked to the mass analyzer device [18]. The instrument calibration is performed by locking experimental *m/z* signals on their theoretical *m/z* values with an adapted mathematical function (e.g., linear, quadratic, or cubic functions). In MSI, as in MS in general, internal and external calibration can be considered. In external calibration, mostly used by the community, the signal of a calibrating substance is acquired before the acquisition of the MSI data. The calibration function is determined from this acquisition and then applied to each spectrum (pixel) of the MSI data [19]. Because it is very easy to set up, this calibration procedure is the most exploited in MSI. However, since mass shift is a pixel-dependent effect, the consistency of MSI data across all pixels when using an external calibration is strongly and negatively impacted by the phenomenon [14]. The reasons behind mass shift phenomenon in MSI are multiple and depend on the mass analyzer design. For example, it has been shown that the number of ions in Fourier-transform ion cyclotron resonance cell (FT-ICR) is correlated with mass shifts [20]. Due to the partial or total incapacity of predicting mass shifts in MSI experiments, internal calibration appears as the method of choice. MSI data acquired with an internal calibration are less affected by mass shifts, as they are directly corrected by the calibration made from specific signals present in the same spectrum (pixel) [14,21]. Reference molecules can be added to the matrix or during sample acquisition (by exploiting dual ESI/MALDI ionization sources) [14,16]. A potential drawback may

be due to the ion suppressive effect generated by the reference molecules to the sample signals. Another strategy for internal calibration is to exploit the presence of endogenous molecules [21], avoiding the much-feared suppressive effect. Even pertinent, this approach is hardly achievable in MSI experiments, as a set of identified ions present in every pixel, has to be known in advance and have to be present in sufficient number [15]. This is highly restrictive as MSI samples are heterogeneous by nature and that the complexity of MSI spectra increases the chance to select non-suitable ions for calibrating all the pixels.

Other approaches have been proposed to solve this problem. Alignment methods have been developed to reduce the mass variation from pixels to pixels, by aligning each spectrum of an MSI between themselves. However, despite increasing the consistency of spectra, alignment does not necessarily correct for mass shifts, contrary to the recalibration [22–25]. Others used the signal produced by peptides for recalibration. In LC/MS it was possible to recalibrate the signals based on confidently identified peptides from a database [26]. In MSI, another work shows that the chemical noise produced in MALDI can be used for recalibration [27]. However, those methods require using specific signals which may not be applicable when using MSI for other molecules than peptides.

In this context, this work aims at proposing a post-acquisition data-adaptive recalibration methodology to correct mass shifts in MS data. The idea is based on the automation of the selection of calibrant/reference signals in each spectrum of a MSI dataset, based on the signal of the lipids and metabolites confidently identified in a similar MSI dataset by METASPACE. METASPACE is a tool of choice for MSI users as it is an open and free access platform for annotating a broad range of metabolites and lipids in MSI data in a confident way controlled with a false discovery rate (FDR) [17]. In this work, we take advantage of METASPACE platform at two different levels. First, we generate a list of potential calibrating ions for the MSI data subjects to recalibration by

selecting similar already annotated MSI data publicly available in METASPACE. Second, we use the numbers of the METASPACE annotations to quantify the data improvement after recalibration. Comparing the numbers of molecular annotations across different datasets or after data recalibration represents an advantage of using an FDR-controlled molecular annotation as established in other omics and represents an advantage of METASPACE in contrast to other ways of metabolite or lipid annotation such as *m/z*-matching (see [17] for more details). Therefore, an efficient recalibration method would be expected to increase the number of METASPACE annotations at the same FDR. For validation of our recalibration method and for showing its broad applicability, we considered 31 public MSI sets from METASPACE coming from various laboratories, acquired with different MS analyzers and representing diverse samples.

Material

In this work, two types of datasets have been analyzed: a zebrafish cryo-section and a set 31 public MSI datasets from METASPACE. First, a 12µm thick slices of one-month-old zebrafish embedded in gelatin were realized on the Cryostar NX70 (Thermo Scientific) then placed on an Indium Tin Oxide glass slide (ITO slide, Bruker, Bremen, Germany). After 15 minutes of desiccation, tissue slices were covered by CHCA matrix (97% purity, Sigma-Aldrich, Taufkirchen, Germany) using an automatic sprayer SunCollect System (SunChrom). MSI acquisition was performed on a SolariX XR 9.4T (Bruker) using the automation software FlexImaging 5.0 (Bruker, Bremen, Germany). Acquisition method consists of 400 laser shots per pixel fired at 1000 Hz with the laser power fixed at 70%. The minimum laser focus was employed with a raster width of 60µm leading to images ranging from 10 to 15k pixels. All MSI data were converted to imzML format using FlexImaging 5.0. On the other hand, 31 public MSI datasets were selected in METASPACE as representative

for different ion sources (MALDI/DESI), analyzers (Orbitrap/FTICR), polarities (positive/negative) and MALDI matrices (DHB for positive mode, DAN for negative mode, CHCA/Norharmane for both modes). To represent a maximum of the METASPACE samples, we have selected human/mouse samples when possible. In addition, 6 datasets from Waters were included to cover several TOF-based analyzers. All the MSI were downloaded as centroided imzML (see SI_2 for more details). The metadata about the sample preparation was, however, only partially provided (see SI_2) and the parameters used to convert the raw image into centroided imzML are not known.

Methods

General approach

In this work, the term "hit" is used to depict a match between experimental signal *m/z* and internal calibrating ion *m/z* within a given mass tolerance from the internal calibrating ion *m/z*. We will call "mass error" the difference in mass between the experimental signal *m/z* and its linked calibrating ion *m/z*. A hit is considered as true if the detected experimental signal corresponds to the calibrating ion. Otherwise, an erroneous match between the experimental signal and the internal calibrating ion is a false hit. In MSI, the mass spectra are information-rich and often contain 10e4 peaks per spectrum. Therefore, it's particularly difficult to discriminate true hits from false hits as multiple signals can be found in the mass range of the *m/z* value of an internal calibrating ion. This difficulty is amplified by the presence of mass shifts since higher mass tolerance is required to capture the internal calibrating signal. However, increasing the mass tolerance inevitably increases the number of false hits, decreasing the recalibration performances. Therefore, the algorithm presented in this work optimizes the selection of true hits for recalibration for each pixel. We assume that true hits

have mass errors directly correlated to the mass shifts. Therefore, by increasing the similarities between the sample signal and the list of internal calibrants, it is expected that the hits with the most frequent mass errors over all the hits predominantly correspond to true hits. The hits with the most frequent mass errors are thus selected for fitting a linear model of the mass errors according to *m/z*. Finally, the recalibration is performed by removing the estimated errors in every detected *m/z* values.

This algorithm is divided into 5 steps (Figure 1). (1) The generation of a list of internal calibrating ions for the whole MSI data according to similar public MSI datasets from METASPACE (i.e., representing the same kind of biological samples) since we assume them to share metabolites with the sample of interest. (2) Centroid MS spectra are extracted from each pixel. (3) The calibrant hits are generated by computing the mass errors between the list of potential internal calibrating ions and the spectrum signals for each pixel. (4) The preferential calibrant hits are selected as those with the most frequent mass errors with the aim to select true hits. (5) A linear model for predicting the mass shifts based on the preferential hits is then constructed and applied to all spectra for their recalibration.

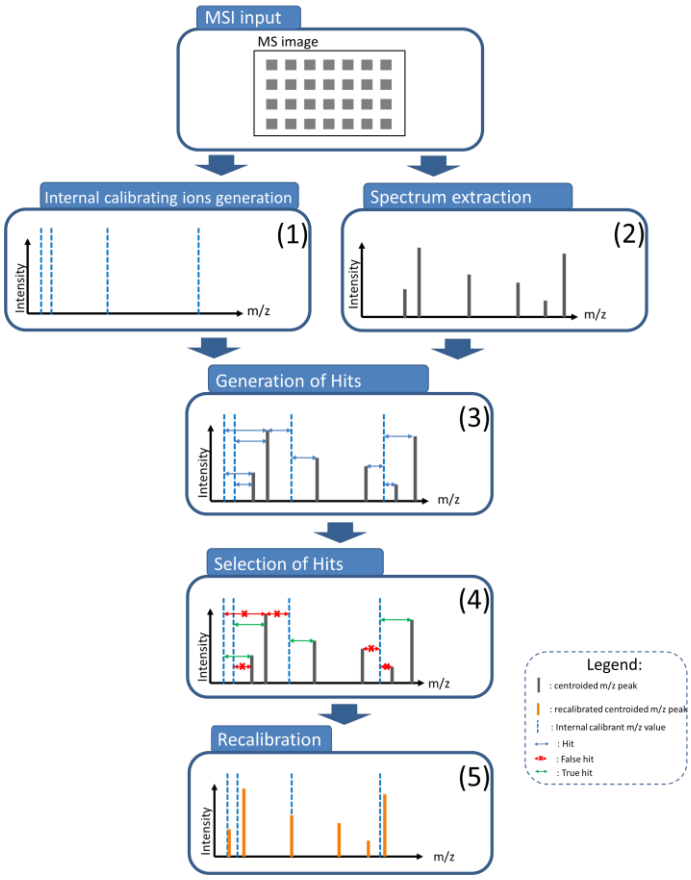


Figure 1: Overview of the recalibration algorithm.

Generation of the list of internal calibrating ions

Generating a list of internal calibrating ions covering the signals from the sample of interest is a critical step for accurate estimation of mass shifts as it directly influences the collection of true hits. Therefore, for each MSI data subject to recalibration, a list of annotations from similar public METASPACE datasets is generated. Similar METASPACE MSI data are selected by their metadata which entails the sample type and experimental parameters of the MSI experiment. We search for data with similar acquisition mode, organisms and organs (regardless of the molecular

database used in METASPACE for annotation). Once selected, the combined list of annotations is reduced to the annotations annotated with an FDR ≤ 10% and detected in at least 10% of considered public datasets. This aims at removing any atypical compound identified only in a few METASPACE MSI datasets, which would have a low chance to be present in the sample of interest. To increase the number of potential calibrating ions we include the two most intense isotopes for each selected calibrating ion. Once a meaningful list of internal calibrating ions generated, the algorithm was applied to each spectrum (pixel), in centroid profile, for hits generation, hits selection, errors estimation and recalibration.

Generation of Hits

Only the 300 most intense peaks are taken into account for each pixel (Figure 1.2). This number is set in advance for all MSI data and is optimized to maximize true signals over noisy signals. This selection is not necessary if denoising is already applied but mandatory when no information is known concerning the generation of the centroid MSI (see material). The 300 *m/z* values are then compared to the list of the internal calibrating signals, within a mass tolerance of ± 0.01 Da (Figure 1.3). This mass tolerance window is considered large enough to encompass most extreme mass shifts in the data, and small enough to discard the contribution of non-relevant peaks and isotopes. The mass errors in Da, used for calculating the regression and the error distribution, are calculated for each hit using Equation 1 where M_{error}, M_{exact} and $M_{experiment}$ are, respectively, the mass error, the exact mass (from the internal calibration list) and the experimental mass.

$$M_{error} = M_{experiment} - M_{exact} \qquad \text{Equation 1}$$

M_{error} values are expressed in Da instead of ppm to ensure a linear evolution of M_{error} errors along the m/z axis (See Figure 2A).

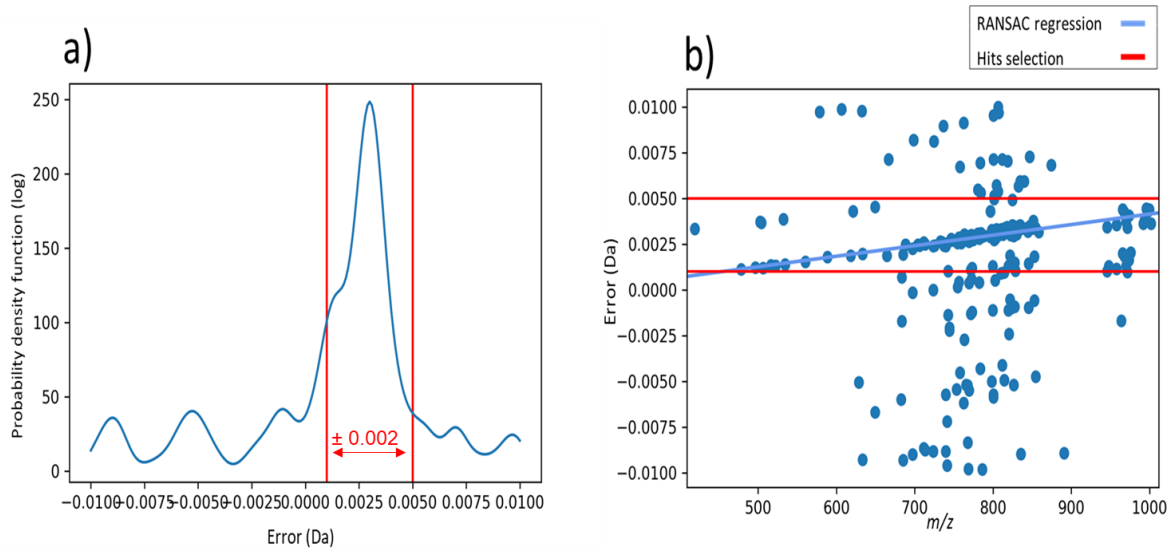
Selection of Hits

Since the mass tolerance window for finding true hits can be large, a preselection of hits is necessary before fitting any model (Figure 1.4). The calculated mass errors of the hits (with Equation 1) are used for discarding false hits. As discussed previously, the assumption is that true hits display similar mass errors and the most populated errors should therefore contain a maximum of true hits. A kernel density estimation of the errors is estimated with the Python library SciPy by a Gaussian kernel [28], a bandwidth of 0.002 divided by the standard deviation of the mass errors was used as illustrated in Figure 2.a. The hits of interest are finally selected within a certain range from the maximum of the density distribution. This range was set at ±0.002 Da for all the MSI data, which was chosen according to the data subject to recalibration.

Mass errors estimation and recalibration

Hits selection reduces the probability to select false hits. Unfortunately, the application of simple linear regression function such as OLS (*Ordinary Least Squares*) leads to poor results since false hits are still present due to the complexity of biological signals. Therefore, RANSAC (*Random Sample Consensus*) algorithm from Python library [29], scikit-learn, is used for regression as it is more robust to outliers compared to OLS. A minimum of 10 hit (i.e., calibration point) and an upper limit of 300 trials for the random points selection are imposed to increase the chance of performing a correct recalibration. If this number of hits is met and if RANSAC algorithm is able to perform a linear regression (Figure 2.b), the recalibration is performed by removing the mass errors estimated by the linear fit, from the original spectrum. This step is made for each detected *m/z* value. (Figure 1.5)

236 **Single spectrum example of application**



237

238 ***Figure 2: "Selection of hits" (a) and errors estimation (b) from the hits of a pixel from the MSI***

239 ***of the zebrafish slice.***

240 The Figure 2 highlights the application of our algorithm to a single pixel of the zebrafish MSI data.

241 The selection of hits based on the density estimation of the mass errors is shown in Figure 2.a, the

242 selected hits are the points located within the red lines. The goal of this step is to filter out as many

243 as possible of those hits which have too high error deviations from true hits, to maximize the chance

244 of not including false hits during the model estimation. RANSAC linear model is then estimated

245 on the selected hits. As many outliers are still present in the selected hits, the use of a robust linear

246 estimator is necessary for detecting only true hits (Figure 2b). In this example, two limitations of

247 this method can already be highlighted. The efficiency of the selection of hits will decrease

248 according to the amplitude of the slope of the linear model (impact of the increase in mass on the

249 error). This effect will disperse true hits over a larger region of mass errors. Therefore, values of

250 the selection of hits tolerance and the density bandwidth must increase, which will decrease the

251 efficiency of the approach as more false hits will be introduced in the calibration points. Secondly,

the errors must follow a linear trend according to the *m/z* due to the model estimation. Non-linear error trends may lead to non-uniform recalibration efficiency across the *m/z* range of the spectrum.

Summary of the Pipeline

The only input needed for the recalibration program is an imzML file [30] composed of centroid spectra. The metadata of this input is exploited to select similar MSI data from METASPACE and generate a suitable list of internal calibrating ions. In each spectrum (pixel), the most 300 intense peaks are selected and matched with a certain mass tolerance (0.01 Da in this work) against the list of internal calibrating ions. A selection of the hits according to the most frequent mass errors (+-0.002 Da in this work) is made. If a minimum10 hits are maintained after selection, a recalibration function is learned from the resulting hits by their errors and *m/z* values. The *m/z* of the initial spectrum are recalibrated. Finally, a new imzML is generated from the recalibrated MS spectra.

Discussion and Results

Recalibration impact on data analysis

To evaluate the efficiency of our internal calibration algorithm, a zebrafish slice has been imaged with a MALDI FT-ICR. Mass shift is the result of the highly heterogeneous molecular composition of the organs and the tissues of this sample leading to different amounts of generated ions in the FT-ICR cell [20] . The settings of the recalibration are detailed in the method section. In this case, too few public MSI datasets of zebrafish in positive mode were found on METASPACE. Therefore, we decide to generate the list of internal calibrating ions from Lipidmaps database [31] with a selection of glycerophospholipids and sphingolipids ions (Na^+, K^+ and H^+ adducts). The Figure 3 shows a comparison of the mean spectra of the image before and after recalibration for two different

276 well-characterized ions (Figure 3A and 3B), as well as the consequence on the reconstructed images
277 of the ions for different tolerance (Figure 3C to 3J). The two peaks chosen for this evaluation are
278 the phosphatidylcholine lipids $[C_{42}H_{82}NO_8P+Na]^+$ and $[C_{40}H_{80}NO_8P+Na]^+$(*m/z*: 782.5670 and
279 756.5514 respectively).
280 Figure 3A and 3B clearly show a large distribution of the experimental *m/z* values (in red) scattered
281 over 7.5 mDa and centered quite far from the exact theoretical *m/z* (2.1% and 3.3% of the pixels
282 are located within 1 ppm from the exact mass of $[C_{40}H_{80}NO_8P+Na]^+$ and $[C_{42}H_{82}NO_8P+Na]^+$,
283 respectively). After recalibration (in green), the distribution of the *m/z* is thinner (spread over ± 0.
284 5 mDa) and more accurate (97.35 % and 98.74 % of the pixels are located within 1 ppm from the
285 exact mass of $[C_{40}H_{80}NO_8P+Na]^+$ and $[C_{42}H_{82}NO_8P+Na]^+$ respectively). The 2D distribution of
286 these two compounds extracted at +- 1 ppm and +- 5 ppm around their theoretical *m/z* value are
287 represented before calibration (Figure 3C and 3G for 1 ppm and 3D and 3H for 5 ppm) and after
288 recalibration (Figure 3E and 3I for 1 ppm and 3F and 3G for 5 ppm).
289 Before recalibration, a mass tolerance window of 5 ppm was necessary to reconstruct the ion
290 distribution as no image was obtained with 1 ppm. However, a large selection window increases
291 the risk to get other ions included in the selected window and to create composite images. After
292 recalibration, the vast majority of the signals of the investigated ions are included in the mass
293 selection window of ± 1ppm (green distributions). It results that these images are much more
294 contrasted and more detailed, enhancing molecular description and interpretation. This highlight
295 the impact of mass shift on the reconstruction of *m/z* image. The comparison of MSI data before
296 and after recalibration supports that our recalibration procedure avoids the loss in mass accuracy
297 in the average MS signal due to mass shifts.
298

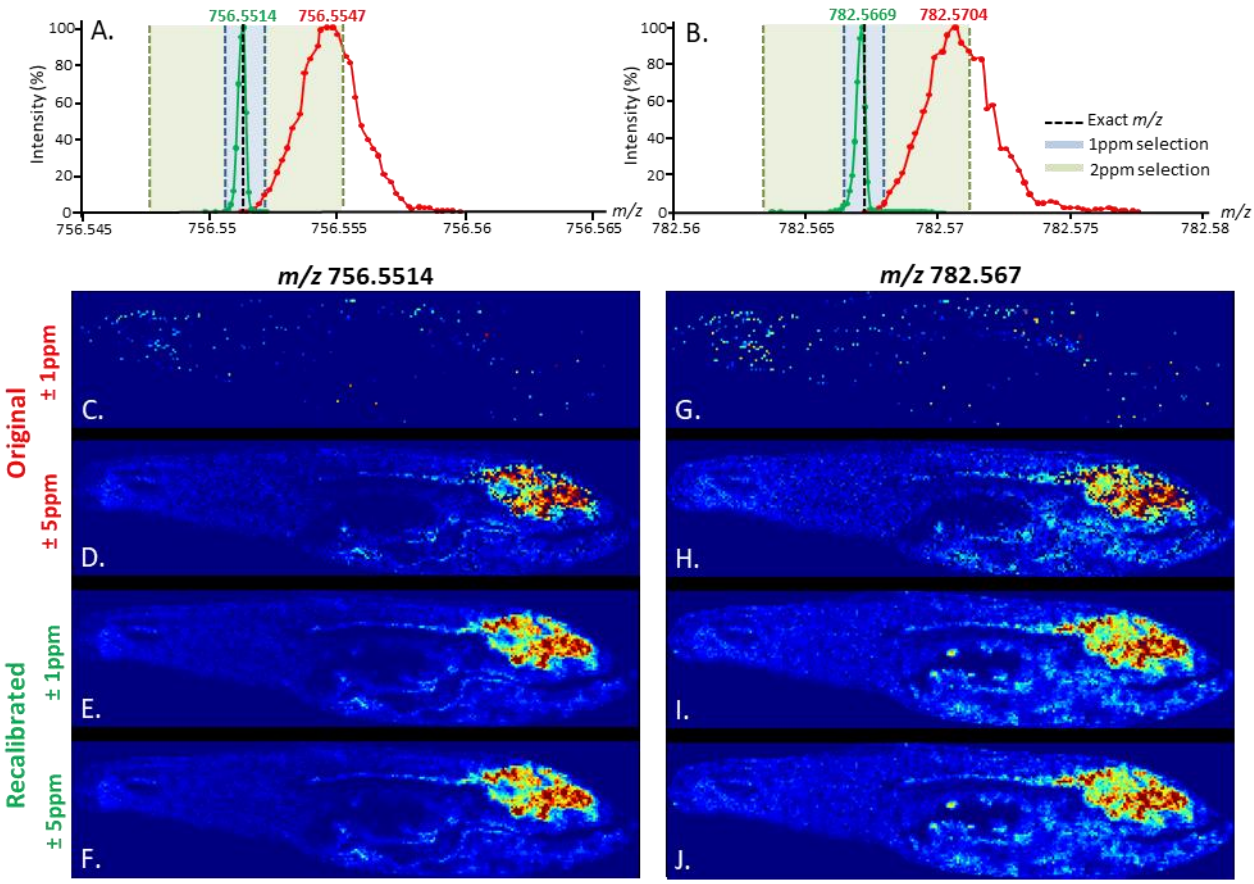


Figure 3: From a zebrafish MSI data. Comparison of the mean spectra before and after recalibration for two different well-characterized ions $[C_{40}H_{80}NO_8P+Na]^+$ and $[C_{42}H_{82}NO_8P+Na]^+$, with a theoretical m/z of 756.5514 m/z and 782.5670 m/z respectively (Figure 3A and 3B), as well as the consequence on the reconstructed images at +- 1 ppm (Figure 3C to 3G and 3E to 3I) and at +- 5 ppm (Figure 3D to 3H and 3F to 3J). The images are reconstructed from the theoretical mass of the two ions.

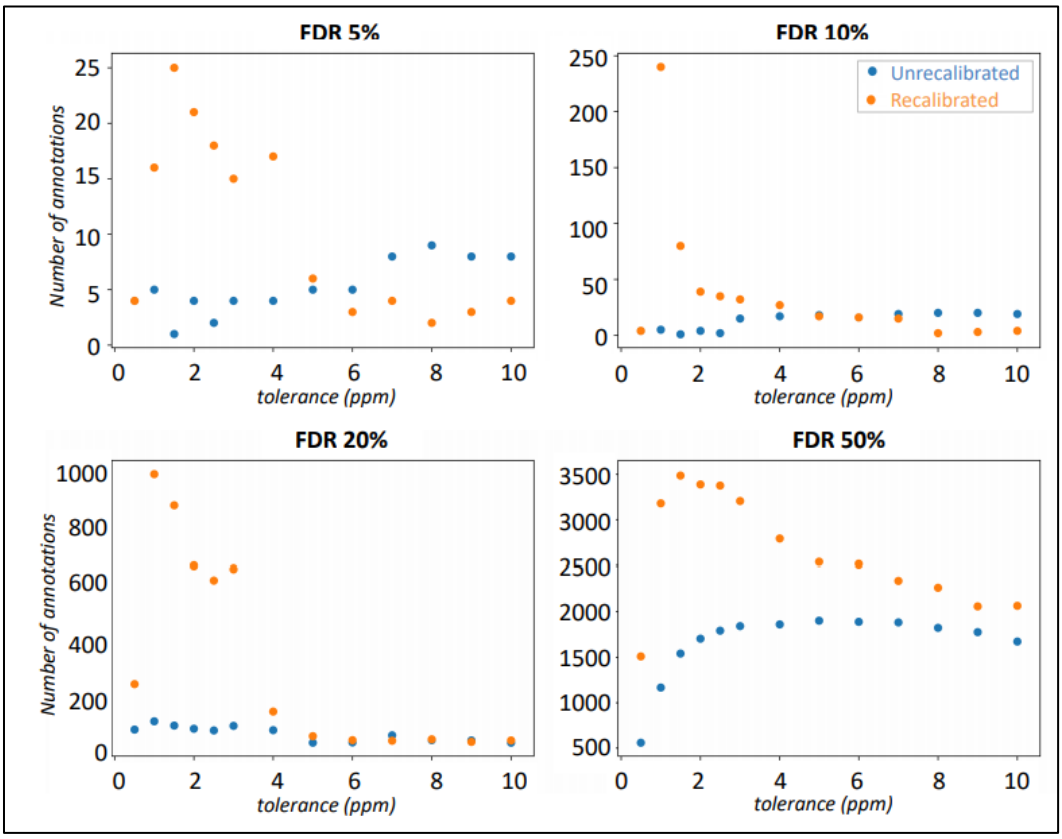
Impact of Recalibration on the Numbers of Molecular Annotations

High mass accuracy is a crucial parameter for obtaining accurate annotations of molecular ions in MSI in absence of MS/MS (fragmentation information). As shown above, the recalibration increases the accuracy of the MS measurement (mass error from 4.3 ppm to 0.12 ppm). Errors below 1 ppm are totally in line with the SolariX FT-ICR mass analyzer precision for single spectrum (considering adequate MS calibration). To quantify the impact of the recalibration on the annotation quality, both unprocessed (original) and recalibrated MSI data have been submitted to

314 METASPACE for automatic annotation. The numbers of annotations at various FDR levels are
315 considered to evaluate the performance of the recalibration algorithm. Moreover, any
316 METASPACE annotation is provided with its MSM score that quantifies the likelihood of the
317 measured signal to match the signal predicted for the molecule from a target database. The MSM
318 score is computed by integrating (i) measure of spatial chaos of the ion image at the selected *m/z*,
319 (ii) the similarity between the experimental and theoretical isotopic patterns and (iii) the spatial
320 correlation between the reconstructed images of the isotopes. The estimation of false positives is
321 made by employing a target-decoy approach where the decoy database contains implausible ions.
322 From METASPACE, the target database can be selected among different popular options such as
323 ChEBI (Chemical Entities of Biological Interest), HMDB (Human Metabolome Database), and
324 LipidsMaps. The FDR is estimated as the proportion of signals that matches the decoy database for
325 that score against the signal that matches the target database for the same score. Therefore, mass
326 shifts should decrease the number of annotations for a given FDR by decreasing the true positive
327 matches in the target database, by increasing the possibility of matching decoy signals and by
328 decreasing the structure of the spatial localization of an ion (Figure 2). The number of
329 METASPACE annotations for a given FDR appears then as an adequate criterion to evaluate the
330 performance of a recalibration strategy. Original and recalibrated MSI are annotated by
331 METASPACE, using the Lipidmaps database and considering the following adducts: $[M+H]^+$,
332 $[M+Na]^+$ and $[M+K]^+$. Mass tolerances used for the identification are 0.5, 1.0, 1.5, 2.0, 2.5 and
333 from 3.0 to 10 ppm with a step of 1 ppm. The performance of the recalibration is assessed by
334 comparing the number of annotations of the original and recalibrated MSI (Figure 4). Classically,
335 annotations with a FDR of 10% or lower are kept for analysis, as lower FDR corresponds to better
336 annotation quality [17]. The most important increase would be at 1 ppm with an FDR of 10 %
337 where more than 200 additional compounds were identified. This increase of annotation strongly

338 supports the effectiveness of our recalibration strategy for reducing mass shift effect and for
339 increasing the accuracy of MSI data.
340
341 **FDR and Tolerance Selection**
342 The evolution of annotation numbers with the tolerance for any FDR is due to two main factors.
343 First, the chances of matching decoy signals increase when the tolerance value increases, reducing
344 the number of annotations. The second factor concern the signal in the MSI data. When the
345 tolerance value increases, the chance of matching isobaric or quasi-isobaric species instead of the
346 expected signal also increases. If the isotopic pattern of those species is close to the suspected
347 annotations, then it will falsely increase the number of annotations.
348 It results that the variation in the number of annotations according to the mass tolerance depends
349 on the predominance of these two factors which is unpredictable for unknown signal (Figure 4).
350 However, the value of tolerance in ppm should be low enough to optimize the number of
351 annotations as it decreases the number of false positives. Therefore, the most interesting tolerance
352 is the minimum value of tolerance giving the highest number of annotations.
353
354 FDR at 5% or 10% are commonly used for the interpretation of the sample composition. FDR
355 below 10% are indicative of the quality of the sample's signals, keeping only the most relevant
356 identifications. Higher FDR such 20% and 50% can, however, be useful for considering
357 annotations with low intensities of signals. Indeed, as annotation scoring also depends on spatial
358 structures and isotope distribution, low intensity signals impacts the scores and appears at higher
359 FDR values. Since the recalibration shows an important increase in annotations at lower FDR
360 (Figure 4), the majority of the signal is probably low in the initial data. In the following part of this
361 paper, only highest quality annotations (5% FDR and tolerance of 1 ppm) will be considered. Even

362 if all compared MSI's analyzer cannot all reach 1 ppm precision, comparing the number of obtained
363 annotations at this value is still interesting as the recalibration should also have an impact, even if
364 not all the pixels of the MSI data are used for annotations.
365
366
367



368

369 ***Figure 4 Number of METASPACE annotations of the unprocessed (blue dots) and recalibrated***
370 ***(orange dots) zebrafish MSI, for different mass tolerances (ppm). The number of annotations is***
371 ***shown for FDR ≤ 5%, ≤ 10%, ≤ 20% and ≤ 50%. The recalibration shows an increase in the***
372 ***number of annotations at low mass tolerance.***
373
374

Multiple MSI Testing

The algorithm is evaluated with 31 different MSI datasets, to assess its robustness against various experimental conditions. The recalibration of these datasets is made thanks to the parameters detailed in the Material and Methods section. Each of the original and recalibrated MSI are annotated by METASPACE. For MSI data in negative mode, only $[M-H]^-$ and $[M+Cl]^-$ ions are considered, whereas for MSI data in positive mode, $[M+H]^+$, $[M+Na]^+$ and $[M+K]^+$ ions were taken into account. The mass tolerances set for the identification are 0.5 and every unit from 1.0 to 8.0 ppm. The performance of the recalibration is assessed by comparing the number of annotations of the original and recalibrated MSI as discussed above. Every possible database available on METASPACE platform is involved as a target database. Among them, the database leading to the most annotation on the unprocessed MSI data at 3 ppm and FDR of 10% has been chosen. Therefore, the annotations may come from different databases in function of the dataset.

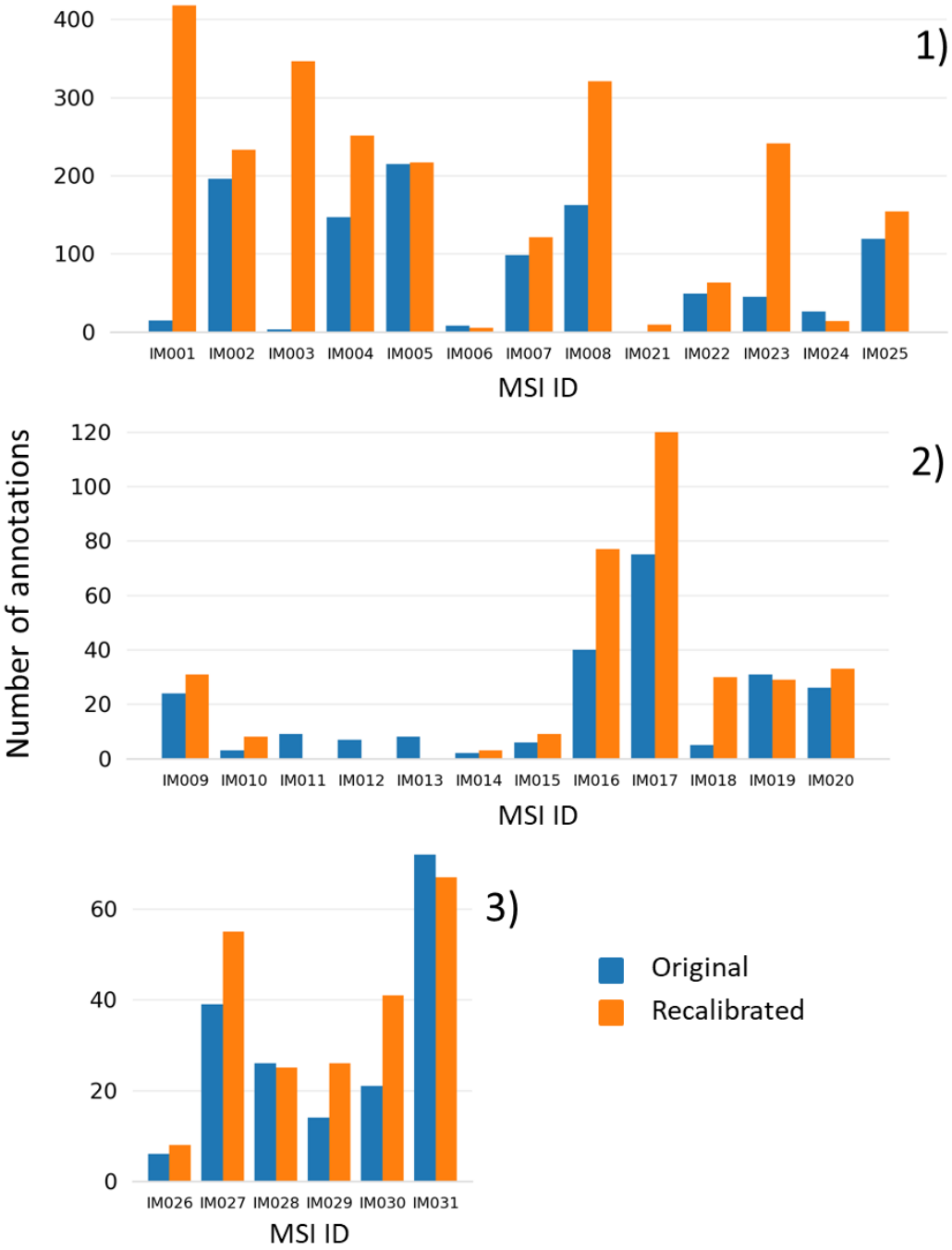


Figure 5: Number of annotations between the recalibrated and original MSI for 31 dataset group by mass analyzer (1) Orbitrap, (2) FT-ICR and (3) TOF including qTOF, SYNAPT-XS, SYNAPT. The annotations were performed on METASPACE for 1 ppm tolerance and an FDR of 5%.

394 As an illustration, Figure 5 compares the number of annotations at 5% FDR and 1 ppm tolerance
395 for the 31 MSI datasets in function of the MS analyzers. The algorithm is able to enhance the
396 number of annotations for 75% of the datasets, by keeping the recalibration parameters identical
397 for each of the MSI data and having no previous knowledge on the samples. Twelve of the
398 recalibrated MSI have more than 20 additional annotations at 1 ppm tolerance and 5% FDR which
399 is not negligible as they are among the most relevant signals of the images. Consequently, our
400 methodology appears robust and independent of the experimental condition of MSI data
401 acquisition. Thus, the same results can be expected on other HRMS MSI data.
402
403 Recalibration of Orbitrap MSI Data shows the most impressive results, gaining from 20 to 400
404 additional annotations. The other analyzers display variable increases. One aspect that could
405 explain this difference between Orbitraps and other mass analyzers is that the MSI data acquire
406 with Orbitraps have more signals compared to the other analyzers (*data not shown*). Therefore,
407 better recalibrations are observed as well as a higher number of annotations.
408
409 The analysis of the 31 recalibration plots (see SI_1) shows that 23 images get a higher number of
410 annotations after recalibration (1 ppm, 5% FDR). However, 6 images (depicted as IM005, IM010,
411 IM015, IM019, IM021, IM024, and IM031 in Figure 5 and SI_1) show a more significant number
412 of annotations only for higher FDR (20% and 50%), which could be related to a low intensity
413 signal. No annotation can be obtained from 3 images (IM011, IM012, and IM013 in Figure 5 and
414 SI_1), as the number of hits is under the applied threshold of 10 hits (see methods). For some cases,
415 the limitation of 10 hits seems to be insufficient to reach a correct recalibration (IM014 and IM026,
416 Figure 5 and SI_1). The low number of hits for the images IM011, IM012, IM013, IM014 and
417 IM026 might be a combination of low sample signals and unidentified METASPACE signals.

However, this limitation of the number of hits will lessen in the future as the number of submitted datasets will increase, covering different sample types and origins. The sample signals will also much probably be enhanced in the next months and years, by increasing the ionization efficiency with new ionization method such as laser-induced post-ionization (MALDI-2) [32] or improving the transmission between the ion source and the MS analyzer. Finally, fewer annotations are observed after recalibration without clear apparent reasons for the IM028 data (see SI_1). These results show that the use of those data and plots can be considered as a quality control to assess the reliability of the recalibration.

In a more general context, the quality of the recalibration is related to the error trend (i.e., mass error vs m/z) and the number of hits in each pixel. As discussed in the method, the amplitude of the slope of the linear model and non-linear mass errors will impact the selection of hits and mass errors model regression. However, taking care of this effect is challenging since the error trend may change from pixel-to-pixel due to the heterogeneous signals in MSI data. A sufficient number of true hits must also be reached to properly determine the error trend over the mass range. The other essential parameter for robust recalibration is the ratio of true hits over false hits, which is dependent on the applied mass tolerance (i.e., mass selection window) for the generation of hits, the sample signals and the list of potential calibrating ions involved for the recalibration. Other mass tolerances were tested on the previous MSI data but didn't show better results (*data not shown*). The list of internal calibrating ions generated to recalibrate each MSI depends only on a subpart of the metadata (see M&M section) where the ionization source, the nature of the MALDI matrix and the tissue preparation method are, for example, not taken into account. The consideration of these metadata could be considered to enhance the generation of calibrating ions list.

Application to Other MS Datasets

HRMS is not the only type of MS instrument suffering from mass shift as a lower-resolution mass spectrometer can also be impacted by it [33]. In this case, the mass tolerance for selection of hits (see M&M section) should be increased which is not a problem as our method is robust to large mass tolerance for the generation of hits. However, limitations concerning the error trend and the range of mass shifts for these instruments should be investigated.

More generally, mass spectrometers are commonly used in direct infusion or hyphenated with separation techniques. Imaging can be considered as a particular case of separative technique, where the spatial distribution of the molecules is used for signal characterization, instead of retention or migration time. Although, the developed method for recalibration is essential for MSI as demonstrated by this work, this is also true for other MS-based approaches involving separative method where the MS signal is expected to change during the acquisition (HPLC-MS, CE-MS…). As for MSI acquisition, the fluctuation of ions during the acquisition requires internal calibration to effectively reduce the mass shift effect. The reported recalibration strategy can be applied to every MS data type, provided each acquisition scan contains enough peaks to properly estimate the true mass error.

Moreover, more and more instruments integrate ion-mobility facility with mass spectrometry, as mobility can also be used as an additional molecular descriptor (related to the tridimensional structure) for improving the annotations. Mobility shifts have, however, also been identified but the recalibration of mobility can be performed, especially as collision cross section (CCS) or mobility value databases (collision cross section) for different biological metabolites and lipids are available in open-source [34,35]. Similarly, to the proposed MS recalibration strategy, those databases can be used as internal references to calibrate the ion mobility values (CCS or mobility). Therefore, the application of this method to mobility has yet to be investigated.

466

467

468 **Conclusion**

469

470 Pixel dependent mass shift is decreasing the interpretability of MSI by affecting the image
471 reconstruction of identified ions and by decreasing the quality and the number of annotations in
472 HRMS. Those have thus a crucial impact on biological interpretation, reducing the capacity to
473 locate and annotate biomarkers on biological samples.

474 Here, we report a new algorithm to recalibrate MSI data, pixel by pixel without preconceptions of
475 the ion composition in the sample. The method is based on the use of publicly available MS
476 databases and annotated MSI data on METASPACE for internal calibrating ions generation. Data
477 recalibration is performed by proper matches between theoretical masses (of the calibrating
478 compounds list) and uncalibrated signals with a robust algorithm requiring very poor human input.
479 Moreover, the calibrating compounds list is adapted from pixel to pixel allowing reaching the better
480 calibration in function of the regions of the MSI data. The automation of the procedure is a
481 prerequisite since ones cannot adjust the mass list for every spectrum of an MSI where an MSI
482 contains thousands of spectra.

483 The comparison of the number of annotations obtained on the METASPACE platform for original
484 and recalibrated MSI data platform are considered as an indicator of the recalibration performances.
485 Moreover, the plots of the number of annotations according to the FDR and the mass tolerance was
486 used as quality control for the image signal quality.

487 The performance and robustness of our recalibration algorithm has been evaluated on 31 different
488 MSI data, acquired from various samples and different MS analyzers, representing the different
489 MSI on the METASPACE repository. An increase of the number of annotations is observed after

data recalibration for most of the investigated data (75%). The different levels of performance of the method according to the different MSI was discussed and mainly attribute to the initial low sample signals and in the METASPACE annotations coverage. In the future, we can expect that the importance of those limitations will be overcome as the quality of the detected signal will increase due to different advancement in the ionization efficiency, MS instrumentation, and in METASPACE annotations coverage.

Another limitation was hypothetically cited such as the error trend and the influence of the mass on the error variation. Even if those parameters are not predominant in this analysis further investigation must be led to understand in which cases those can influence the recalibration performances and how the algorithm could be improved for these particular cases.

The integration of this data post-processing in METASPACE is currently considered since it is using numerous features already available in METASPACE. Moreover, it will enable the further testing of the method on more samples, which will highlight the best parameters to use with this algorithm.

The effectiveness of our recalibration strategy has been shown on tissues MSI data but this approach can be considered for all MS data whereas MS signals are heterogeneous during the acquisition scans (e.g., HPLC, CE, IMS, TLC). The only condition is that enough peak is detected to ensure a good estimation of the mass shift. Moreover, the reported method can be considered for internal calibration of other ion descriptor than *m/z* ratios such as mobility or collision cross section values using adequate databases.

Acknowledgments

We thank the METASAPCE submitters who shared their data for this work: James McKenzie (ICL), Zoltan Takats (ICL), Dhaka Bhandari (JLU Giessen), Mario Kompauer (JLU Giessen), Shane Ellis (M4I), Dusan Velickovic (PNNL), Jessica Lukowski (PNNL), Lavigne Régis (U Rennes 1), Annapurna Pamreddy (UTHSA), Guanshi Zhang (UTHSA), Bindesh Shrestha (Waters), Emrys Jones (Waters).

We acknowledge the funding from Eurlipids, EU-FT-ICR, the European Research Council (agreement 773089) and the European Horizon2020 ICT project CloudButton (agreement 825184).

References

[1] P.-M. Vaysse, R.M.A. Heeren, T. Porta, B. Balluff, Mass spectrometry imaging for clinical research – latest developments, applications, and current limitations, Analyst. 142 (2017) 2690–2712. https://doi.org/10.1039/C7AN00565B.

[2] R.M. Caprioli, T.B. Farmer, J. Gile, Molecular Imaging of Biological Samples: Localization of Peptides and Proteins Using MALDI-TOF MS, Anal. Chem. 69 (1997) 4751–4760. https://doi.org/10.1021/ac970888i.

[3] D. Gode, D.A. Volmer, Lipid imaging by mass spectrometry – a review, Analyst. 138 (2013) 1289. https://doi.org/10.1039/c2an36337b.

[4] S. Khatib-Shahidi, M. Andersson, J.L. Herman, T.A. Gillespie, R.M. Caprioli, Direct Molecular Analysis of Whole-Body Animal Tissue Sections by Imaging MALDI Mass Spectrometry, Anal. Chem. 78 (2006) 6448–6456. https://doi.org/10.1021/ac060788p.

[5] M. Stoeckli, D. Staab, A. Schweitzer, Compound and metabolite distribution measured by MALDI mass spectrometric imaging in whole-body tissue sections, International Journal of Mass Spectrometry. 260 (2007) 195–202. https://doi.org/10.1016/j.ijms.2006.10.007.

[6] J.Y. Yang, V.V. Phelan, R. Simkovsky, J.D. Watrous, R.M. Trial, T.C. Fleming, R. Wenter, B.S. Moore, S.S. Golden, K. Pogliano, P.C. Dorrestein, Primer on Agar-Based Microbial Imaging Mass Spectrometry, Journal of Bacteriology. 194 (2012) 6023–6028. https://doi.org/10.1128/JB.00823-12.

[7] N. Zaima, N. Goto-Inoue, T. Hayasaka, M. Setou, Application of imaging mass spectrometry for the analysis of Oryza sativa rice, Rapid Commun. Mass Spectrom. 24 (2010) 2723–2729. https://doi.org/10.1002/rcm.4693.

[8] E.K. Neumann, J.F. Ellis, A.E. Triplett, S.S. Rubakhin, J.V. Sweedler, Lipid Analysis of 30 000 Individual Rodent Cerebellar Cells Using High-Resolution Mass Spectrometry, Anal. Chem. 91 (2019) 7871–7878. https://doi.org/10.1021/acs.analchem.9b01689.

[9] K.J. Boggio, E. Obasuyi, K. Sugino, S.B. Nelson, N.Y. Agar, J.N. Agar, Recent advances in single-cell MALDI mass spectrometry imaging and potential clinical impact, Expert Review of Proteomics. 8 (2011) 591–604. https://doi.org/10.1586/epr.11.53.

[10] R.G. Cooks, Ambient Mass Spectrometry, Science. 311 (2006) 1566–1570. https://doi.org/10.1126/science.1119426.

[11] T. Alexandrov, MALDI imaging mass spectrometry: statistical data analysis and current computational challenges, BMC Bioinformatics. 13 (2012) S11. https://doi.org/10.1186/1471-2105-13-S16-S11.

[12] A.R. Buchberger, K. DeLaney, J. Johnson, L. Li, Mass Spectrometry Imaging: A Review of Emerging Advancements and Future Insights, Anal. Chem. 90 (2018) 240–265. https://doi.org/10.1021/acs.analchem.7b04733.

[13] D. Schwudke, K. Schuhmann, R. Herzog, S.R. Bornstein, A. Shevchenko, Shotgun Lipidomics on High Resolution Mass Spectrometers, Cold Spring Harbor Perspectives in Biology. 3 (2011) a004614–a004614. https://doi.org/10.1101/cshperspect.a004614.

[14] D.F. Smith, A. Kharchenko, M. Konijnenburg, I. Klinkert, L. Paša-Tolić, R.M.A. Heeren, Advanced Mass Calibration and Visualization for FT-ICR Mass Spectrometry Imaging, J. Am. Soc. Mass Spectrom. 23 (2012) 1865–1872. https://doi.org/10.1007/s13361-012-0464-1.

[15] A.N. Kozhinov, K.O. Zhurov, Y.O. Tsybin, Iterative Method for Mass Spectra Recalibration via Empirical Estimation of the Mass Calibration Function for Fourier Transform Mass Spectrometry-Based Petroleomics, Anal. Chem. 85 (2013) 6437–6445. https://doi.org/10.1021/ac400972y.

[16] R.D. Burton, K.P. Matuszak, C.H. Watson, J.R. Eyler, Exact mass measurements using a 7 tesla fourier transform ion cyclotron resonance mass spectrometer in a good laboratory practices-regulated environment, J Am Soc Mass Spectrom. 10 (1999) 1291–1297. https://doi.org/10.1016/S1044-0305(99)00106-3.

[17] A. Palmer, P. Phapale, I. Chernyavsky, R. Lavigne, D. Fay, A. Tarasov, V. Kovalev, J. Fuchser, S. Nikolenko, C. Pineau, M. Becker, T. Alexandrov, FDR-controlled metabolite annotation for high-resolution imaging mass spectrometry, Nat Methods. 14 (2017) 57–60. https://doi.org/10.1038/nmeth.4072.

[18] T. Alexandrov, Spatial Metabolomics and Imaging Mass Spectrometry in the Age of Artificial Intelligence, Annu. Rev. Biomed. Data Sci. 3 (2020) 61–87. https://doi.org/10.1146/annurev-biodatasci-011420-031537.

[19] J.H. Gross, Mass Spectrometry, Springer Berlin Heidelberg, Berlin, Heidelberg, 2011. https://doi.org/10.1007/978-3-642-10711-5.

[20] E.N. Nikolaev, Y.I. Kostyukevich, G.N. Vladimirov, Fourier transform ion cyclotron resonance (FT ICR) mass spectrometry: Theory and simulations: FT ICR MS, Mass Spec Rev. 35 (2016) 219–258. https://doi.org/10.1002/mas.21422.

[21] P.C. Kooijman, K.O. Nagornov, A.N. Kozhinov, D.P.A. Kilgour, Y.O. Tsybin, R.M.A. Heeren, S.R. Ellis, Increased throughput and ultra-high mass resolution in DESI FT-ICR MS imaging through new-generation external data acquisition system and advanced data processing approaches, Sci Rep. 9 (2019) 8. https://doi.org/10.1038/s41598-018-36957-1.

[22] M.B. Tracy, H. Chen, D.M. Weaver, D.I. Malyarenko, M. Sasinowski, L.H. Cazares, R.R. Drake, O.J. Semmes, E.R. Tracy, W.E. Cooke, Precision enhancement of MALDI-TOF MS using high resolution peak detection and label-free alignment, Proteomics. 8 (2008) 1530–1538. https://doi.org/10.1002/pmic.200701146.

[23] F. Brochu, P.-L. Plante, A. Drouin, D. Gagnon, D. Richard, F. Durocher, C. Diorio, M. Marchand, J. Corbeil, F. Laviolette, Mass spectra alignment using virtual lock-masses, Sci Rep. 9 (2019) 8469. https://doi.org/10.1038/s41598-019-44923-8.

[24] P. Ràfols, E. del Castillo, O. Yanes, J. Brezmes, X. Correig, Novel automated workflow for spectral alignment and mass calibration in MS imaging using a sputtered Ag nanolayer, Analytica Chimica Acta. 1022 (2018) 61–69. https://doi.org/10.1016/j.aca.2018.03.031.

[25] P. Kulkarni, F. Kaftan, P. Kynast, A. Svatoš, S. Böcker, Correcting mass shifts: A lock mass-free recalibration procedure for mass spectrometry imaging data, Anal Bioanal Chem. 407 (2015) 7603–7613. https://doi.org/10.1007/s00216-015-8935-4.

[26] J. Cox, A. Michalski, M. Mann, Software Lock Mass by Two-Dimensional Minimization of Peptide Mass Errors, J. Am. Soc. Mass Spectrom. 22 (2011) 1373–1380. https://doi.org/10.1007/s13361-011-0142-8.

[27] T. Boskamp, D. Lachmund, R. Casadonte, L. Hauberg-Lotte, J.H. Kobarg, J. Kriegsmann, P. Maass, Using the Chemical Noise Background in MALDI Mass Spectrometry Imaging for Mass Alignment and Calibration, Anal. Chem. 92 (2020) 1301–1308. https://doi.org/10.1021/acs.analchem.9b04473.

[28] T. Ledl, Kernel Density Estimation: Theory and Application in Discriminant Analysis, AJS. 33 (2016) 267–279. https://doi.org/10.17713/ajs.v33i3.441.

[29] M.A. Fischler, R.C. Bolles, Random sample consensus: a paradigm for model fitting with applications to image analysis and automated cartography, Commun. ACM. 24 (1981) 381–395. https://doi.org/10.1145/358669.358692.

[30] T. Schramm, Z. Hester, I. Klinkert, J.-P. Both, R.M.A. Heeren, A. Brunelle, O. Laprévote, N. Desbenoit, M.-F. Robbe, M. Stoeckli, B. Spengler, A. Römpp, imzML — A common data format for the flexible exchange and processing of mass spectrometry imaging data, Journal of Proteomics. 75 (2012) 5106–5110. https://doi.org/10.1016/j.jprot.2012.07.026.

[31] M. Sud, E. Fahy, D. Cotter, A. Brown, E.A. Dennis, C.K. Glass, A.H. Merrill, R.C. Murphy, C.R.H. Raetz, D.W. Russell, S. Subramaniam, LMSD: LIPID MAPS structure database, Nucleic Acids Research. 35 (2007) D527–D532. https://doi.org/10.1093/nar/gkl838.

[32] J. Soltwisch, H. Kettling, S. Vens-Cappell, M. Wiegelmann, J. Muthing, K. Dreisewerd, Mass spectrometry imaging with laser-induced postionization, Science. 348 (2015) 211–215. https://doi.org/10.1126/science.aaa1051.

[33] J.O.R. Gustafsson, J.S. Eddes, S. Meding, T. Koudelka, M.K. Oehler, S.R. McColl, P. Hoffmann, Internal calibrants allow high accuracy peptide matching between MALDI imaging MS and LC-MS/MS, Journal of Proteomics. 75 (2012) 5093–5105. https://doi.org/10.1016/j.jprot.2012.04.054.

[34] K.L. Leaptrot, J.C. May, J.N. Dodds, J.A. McLean, Ion mobility conformational lipid atlas for high confidence lipidomics, Nat Commun. 10 (2019) 985. https://doi.org/10.1038/s41467-019-08897-5.

[35] Z. Zhou, X. Shen, J. Tu, Z.-J. Zhu, Large-Scale Prediction of Collision Cross-Section Values for Metabolites in Ion Mobility-Mass Spectrometry, Anal. Chem. 88 (2016) 11084–11091. https://doi.org/10.1021/acs.analchem.6b03091.