



HAL
open science

FDR-controlled metabolite annotation for high-resolution imaging mass spectrometry

A. Palmer, P. Phapale, I. Chernyavsky, Régis Lavigne, D. Fay, A. Tarasov, V. Kovalev, J. Fuchser, S. Nikolenko, Charles Pineau, et al.

► **To cite this version:**

A. Palmer, P. Phapale, I. Chernyavsky, Régis Lavigne, D. Fay, et al.. FDR-controlled metabolite annotation for high-resolution imaging mass spectrometry. *Nature Methods*, 2017, 14 (1), pp.57-60. 10.1038/nmeth.4072 . hal-01467553

HAL Id: hal-01467553

<https://univ-rennes.hal.science/hal-01467553v1>

Submitted on 13 Apr 2018

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

FDR-controlled metabolite annotation for high-resolution imaging mass spectrometry

Andrew Palmer¹, Prasad Phapale¹, Ilya Chernyavsky²⁺, Regis Lavigne^{3,4}, Dominik Fay¹, Artem Tarasov¹, Vitaly Kovalev¹, Jens Fuchser⁵, Sergey Nikolenko^{6,1,7}, Charles Pineau^{3,4}, Michael Becker⁵, Theodore Alexandrov^{1,2+,8,9}

1: Structural and Computational Biology Unit, European Molecular Biology Laboratory, Heidelberg, Germany

2: Center for Industrial Mathematics, University of Bremen, Bremen, Germany

+: ex-affiliation of IC and TA

3: Protim, Inserm U1085 - Irset, Rennes, France

4: Université de Rennes I, Rennes, France

5: Bruker Daltonik GmbH, Bremen, Germany

6: National Research University Higher School of Economics, St. Petersburg, Russia

7: Kazan Federal University, Kazan, Russia

8: SCiLS GmbH, Bremen, Germany

9: Skaggs School of Pharmacy and Pharmaceutical Sciences, University of California San Diego, La Jolla, California, USA

Correspondence should be addressed to TA (theodore.alexandrov@embl.de)

Abstract

High-mass-resolution imaging mass spectrometry promises to localize hundreds of metabolites directly from tissues, cell cultures, and agar plates with cellular resolution, but is hampered by the lack of bioinformatics for automated metabolite identification. We developed the first bioinformatics framework for False Discovery Rate (FDR)-controlled metabolite annotation for high-mass-resolution imaging mass spectrometry (<https://github.com/alexandrovteam/pySM>) introducing a Metabolite-Signal Match (MSM) score and a target-decoy FDR-estimate for spatial metabolomics.

Main text

Metabolomics is promising important advances in cell biology, physiology, and medicine. Metabolomics complements other -omics by analyzing the instantaneous state of biochemical processes and revealing contributions of non-genetic factors. Metabolomics has progressed from cataloguing chemical structures to answering complex biomedical questions^{1,2} and generally requires pooling cell populations. The next grand challenge is to spatially localize metabolites *in situ*.

Spatial metabolomics is emerging enabled by novel analytical techniques. Arguably the most promising technique that demonstrated its potential and attracts wide attention is imaging mass spectrometry (imaging MS^{3,4-8}), where mass spectra are acquired directly from a section of tissue, cell culture, or agar. The key challenge in imaging MS is the molecular identification⁹, particularly because online data-dependent MS/MS like in LC- or GC-MS/MS is not feasible. This puts high demands on mass-based separation. The introduction of high-

mass-resolution (HR) MS that discriminates metabolites differing by a few *mDa* promises to achieve unprecedented reliability of metabolite annotation.

However, no bioinformatics exists for automated metabolite annotation in HR imaging MS. This has restricted this powerful technique mainly to targeted imaging of a few metabolites only¹⁰. Existing approaches either need visual examination or are based on the exact mass filtering known to produce false positives even for ultra-HR MS¹¹. This gap can be explained by the field novelty and high requirements to the algorithms which should be robust to strong pixel-to-pixel noise and efficient enough to mine 10-100 gigabyte datasets.

An additional obstacle is the lack of a metabolomics-compatible approach for estimating False Discovery Rate (FDR)^{12,13}. FDR is defined as the ratio of false positives in a set of annotations. FDR is a cornerstone of quantifying quality of annotations in genomics, transcriptomics, and proteomics¹⁴. The proteomics target-decoy FDR-estimation is not directly applicable in metabolomics where there is no equivalent of a decoy database of implausible peptide sequences. An FDR-estimate in metabolomics was proposed earlier¹⁵ but is limited to phytochemical metabolites, has not found widespread use and cannot be applied to imaging MS as it does not allow incorporating spatial information. An alternative approach to estimate FDR is to use a phantom sample with controlled molecular content but it is inherently complex and narrowed to a specific protocol.

We have addressed this double challenge and developed a comprehensive bioinformatics framework for FDR-controlled metabolite annotation for HR imaging MS. Our open-source framework (<https://github.com/alexandrovteam/pySM>) is based on the following principles: database-driven annotation by screening for metabolites with known sum formulas, an original Metabolite-Signal Match (MSM) score combining spectral and spatial measures, a novel target-decoy FDR-estimation approach with a decoy set generated by using implausible adducts.

Our framework takes as input: 1) an HR imaging MS dataset in the imzML format, 2) a database of metabolite sum formulas in a CSV format (e.g., HMDB¹⁶), 3) an adduct of interest (e.g., +H, +Na, +K). For a specified FDR level (e.g., 0.1), the framework provides metabolite annotations: metabolites from the database detected as present in the sample. The framework cannot resolve isomeric metabolites; the provided putative molecular annotations are on the level of sum formulas¹⁷.

Our novel MSM score quantifies the likelihood of the presence of a metabolite with a given sum formula in the sample (Figure 1; Supplementary Note 1, Figure S2). For an ion (sum formula plus ion adduct, e.g., +H), we generate its isotopic pattern accounting for the instrument resolving power with isotopic fine structure if resolvable. Then, we sample from the imaging MS dataset an ion signal, namely, the ion images for all isotopic peaks with predicted intensity greater than 0.01% of the principal peak (Supplementary Note 1, Figure S1). MSM is computed by multiplying the following measures. (1) *Measure of spatial chaos* quantifies spatial informativeness within the image of the principal peak¹⁸. We introduce an improved measure of spatial chaos (Algorithm OM1) which outperforms earlier proposed measures^{18,19} in both speed and accuracy (Supplementary Note 1). (2) *Spectral isotope measure* quantifies spectral similarity between a theoretical isotopic pattern and relative sampled isotopic intensities. (3) *Spatial isotope measure* quantifies spatial co-localization

between isotopic ion images. The MSM score of 1 indicates the maximal likelihood of the signal to correspond to the ion.

Our novel FDR-estimate helps select an MSM cutoff so that the ions with MSM scores above the cutoff will confidently correspond to metabolites from the sample (Figure 1; Supplementary Note 1, Figure S2). According to the target-decoy approach¹⁴, we propose to construct a decoy set as follows. We define a target set as ions from a metabolite database with a given ion adduct (e.g., +H). We define the decoy set as ions for the same sum formulas but with the following implausible adducts. For each sum formula, we randomly select an implausible adduct from the CIAAW 2009 list of the elements (e.g., +B, +Db, +Ag) excluding plausible adducts. MSM scores are calculated for target and decoy ions. For any MSM cutoff, FDR is estimated as the ratio between the numbers of decoy false positives (the decoy ions with MSM scores above the cutoff, FP_D) and target positives (the target ions with MSM scores above the cutoff). Here, we approximate the number of target false positives (FP_T) by FP_D assuming the target and decoy sets to be similar. The sampling of implausible adducts is repeated, averaging the resulting FDR-estimate.

FDR-controlled metabolite annotation is performed by specifying the desired value of FDR (e.g., 0.1) and choosing the smallest MSM cutoff providing the desired FDR (Figure 1; Supplementary Note 1, Figure S2). FDR-controlling provides annotations of a given confidence independently on the MSM cutoff, dataset, MS settings and operator, and can be used for comparative and inter-lab studies.

We evaluated the proposed FDR-estimation (Supplementary Note 1). First, we studied the similarity between the decoy and target ions required to fulfill $FP_D \approx FP_T$. Relative intensities of isotopic patterns for target and decoy ions were found to be similar (Figure 2a) despite the decoy ions have higher relative intensities for heavier isotopic peaks due to more complex isotopic patterns. The target and decoy ions were also found to be similar in the m/z - and mass defect-space (Figure 2b), with a positive offset in m/z for decoy adducts which typically have heavier elements. Second, we compared the estimated and true FDR for a simulated dataset with a known ground truth (Figure 2c; Supplementary Note 1). Although there is some difference in the low-values region, estimated FDR follows the true FDR overall. Finally, negative control experiments using each of the implausible adducts as a target one showed that FDR values for implausible adducts are characteristically higher (Figure 2d; Supplementary Note 1).

We showcased our framework on HR imaging MS datasets from two (a1 and a2) female adult wild-type mice (Supplementary Note 1). The brains were extracted, snap-frozen, and sectioned using a cryostat. Five coronal sections were collected from each brain: 3 serial sections (s1-s3) at the Bregma 1.42 mm, s4 at -1.46 mm and s5 at -3.88 mm. The sections were imaged using a 7T MALDI-FTICR mass spectrometer solariX XR (Bruker Daltonics) in the positive mode with 50 μm raster size. The datasets were of 20-35 gigabytes in size each. FDR-controlled annotation was performed with the desired level of FDR=0.1 for metabolites from HMDB with +H, +Na, +K adducts, and m/z -tolerance of 2.5 ppm (Figure 2e-i). Venn diagrams of annotated metabolites (Figure 2e) show a high reproducibility between sections from the same animal (especially between the serial sections from a2 where 51 of 73 sum formulas were annotated in all three sections), and between the animals (only two sum formulas were annotated in the animal a1 only). The numbers of detected adducts were

similar (Figure 1f). Exemplary molecular images of annotations illustrate a high reproducibility between technical replicates and animals (Figure 1g). Phospholipids were detected mostly (PCs, PEs, SMs, PAs; Supplementary Note 1, Table S5 and Figure S10) that is typical for MALDI imaging MS of brain tissue using the HCCA matrix²⁰. From overall 103 annotations, 16 representative ones were validated with LC-MS/MS by either using authentic standards or assigning fragment structures to MS/MS data (Supplementary Note 3).

We demonstrated the potential of using FDR curves in two examples. First, we showed that MSM outperforms the individual measures (Figure 2h; Supplementary Note 1, Figure S8). The exact mass filtering performs significantly worse, achieving the lowest FDR=0.25 for 10 annotations (vs. FDR=0 for the same number of annotations when using MSM). Second, we demonstrated that the number of FDR-controlled annotations decreases with the decreasing mass resolving power (Figure 2i; Supplementary Note 1, Figure S9). For this, we artificially reduced mass resolving power by using different *m/z*-tolerances when sampling *m/z*-signals: 1, 2.5 (default), 5, 30, 100, 1000, and 5000 *ppm*. This indicates that a high mass accuracy and resolution are essential for confident metabolite annotation.

Our framework is directly applicable to other types of HR imaging MS with FTICR or Orbitrap analyzers (MALDI-, DESI-, SIMS-, IR-MALDESI-, etc.; with proper adducts to be selected for each source) and other types of samples (plant tissue, cell culture, agar plate, etc.) for which a proper metabolite database can be selected.

Accession Codes

MTBLS313: imaging mass spectrometry data from mouse and rat brain samples,
MTBLS317: simulated imaging mass spectrometry data and MTBLS378: LC-MS/MS data from mouse brain samples.

Data Availability Statement

The data is publicly available at the MetaboLights repository under the following accession numbers: MTBLS313: imaging mass spectrometry data from mouse and rat brain samples, MTBLS378: LC-MS/MS data from mouse brain samples, and MTBLS317: simulated imaging mass spectrometry data.

Acknowledgements

We thank Olga Vitek (Northeastern University), Alexander Makarov (ThermoFisher Scientific) and Mikhail Savitski (EMBL) for discussions on FDR and Dmitry Feichtner-Kozlov (University of Bremen) for discussions on computational topology. We acknowledge funding from the European Union's Horizon2020 and FP7 programmes under the grant agreements No. 634402 (AP, RL, AT, VK, SN, CP, TA), 305259 (IC, RL, CP), and from the Russian Government Program of Competitive Growth of Kazan Federal University (SN). We thank EMBL Core Facilities for instrumentation for LC-MS/MS analysis. TA thanks Pieter

Dorrestein (UCSD) and Peter Maass (University of Bremen) for providing a stimulating environment as well as for discussions on mass spectrometry and image analysis during the years of this work.

Author Contributions

AP and TA conceived the study, AP, IC, DF, AT, VK implemented the algorithms, RL, JF, CP, MB provided imaging data, AP and TA analyzed imaging data, PP collected LC-MS/MS data, PP and TA performed LC-MS/MS validation, AP and TA wrote manuscript, with feedback from all other coauthors, TA coordinated the project.

Competing Financial Interest Statements

Theodore Alexandrov is the scientific director and a shareholder of SCiLS GmbH, a company providing software for imaging mass spectrometry. During the work presented in the paper, Michael Becker was an employee of Bruker Daltonik GmbH, a company providing instrumentation and software for imaging mass spectrometry.

References

1. Baker, M. Mass spectrometry for biologists. *Nat. Methods* **7**, 157–161 (2010).
2. Johnson, C. H., Ivanisevic, J. & Siuzdak, G. Metabolomics: beyond biomarkers and towards mechanisms. *Nat. Rev. Mol. Cell Biol.* (2016). doi:10.1038/nrm.2016.25
3. Watrous, J. D., Alexandrov, T. & Dorrestein, P. C. The evolving field of imaging mass spectrometry and its impact on future biological research. *J. Mass Spectrom.* **46**, 209–222 (2011).
4. Watrous, J. D. & Dorrestein, P. C. Imaging mass spectrometry in microbiology. *Nat. Rev. Microbiol.* **9**, 683–694 (2011).
5. Schwamborn, K. & Caprioli, R. M. Molecular imaging by mass spectrometry — looking beyond classical histology. *Nat. Rev. Cancer* **10**, 639–646 (2010).
6. Attia, A. S. *et al.* Monitoring the inflammatory response to infection through the integration of MALDI IMS and MRI. *Cell Host Microbe* **11**, 664–673 (2012).
7. Johnson, C. H. *et al.* Metabolism links bacterial biofilms and colon carcinogenesis. *Cell Metab.* **21**, 891–897 (2015).
8. Soltwisch, J. *et al.* Mass spectrometry imaging with laser-induced postionization.

- Science* **348**, 211–215 (2015).
9. Palmer, A., Trede, D. & Alexandrov, T. Where imaging mass spectrometry stands: here are the numbers. *Metabolomics* **12**, 1–3 (2016).
 10. Spengler, B. Mass spectrometry imaging of biomolecular information. *Anal. Chem.* **87**, 64–82 (2015).
 11. Kind, T. & Fiehn, O. Metabolomic database annotations via query of elemental compositions: mass accuracy is insufficient even at less than 1 ppm. *BMC Bioinformatics* **7**, 234 (2006).
 12. Benjamini, Y. & Hochberg, Y. Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing. *J. R. Stat. Soc. Series B Stat. Methodol.* **57**, 289–300 (1995).
 13. Storey, J. D. A direct approach to false discovery rates. *J. R. Stat. Soc. Series B Stat. Methodol.* **64**, 479–498 (2002).
 14. Käll, L., Storey, J. D., MacCoss, M. J. & Noble, W. S. Assigning significance to peptides identified by tandem mass spectrometry using decoy databases. *J. Proteome Res.* **7**, 29–34 (2008).
 15. Matsuda, F. *et al.* Assessment of metabolome annotation quality: a method for evaluating the false discovery rate of elemental composition searches. *PLoS One* **4**, e7490 (2009).
 16. Wishart, D. S. *et al.* HMDB 3.0--The Human Metabolome Database in 2013. *Nucleic Acids Res.* **41**, D801–7 (2013).
 17. Sumner, L. W. *et al.* Proposed minimum reporting standards for chemical analysis
Chemical Analysis Working Group (CAWG) Metabolomics Standards Initiative (MSI). *Metabolomics* **3**, 211–221 (2007).
 18. Alexandrov, T. & Bartels, A. Testing for presence of known and unknown molecules in imaging mass spectrometry. *Bioinformatics* **29**, 2335–2342 (2013).
 19. Wijetunge, C. D. *et al.* EXIMS: an improved data analysis pipeline based on a new peak picking method for EXploring Imaging Mass Spectrometry data. *Bioinformatics* **31**,

3198–3206 (2015).

20. Gode, D. & Volmer, D. A. Lipid imaging by mass spectrometry -- a review. *Analyst* **138**, 1289 (2013).

Figures

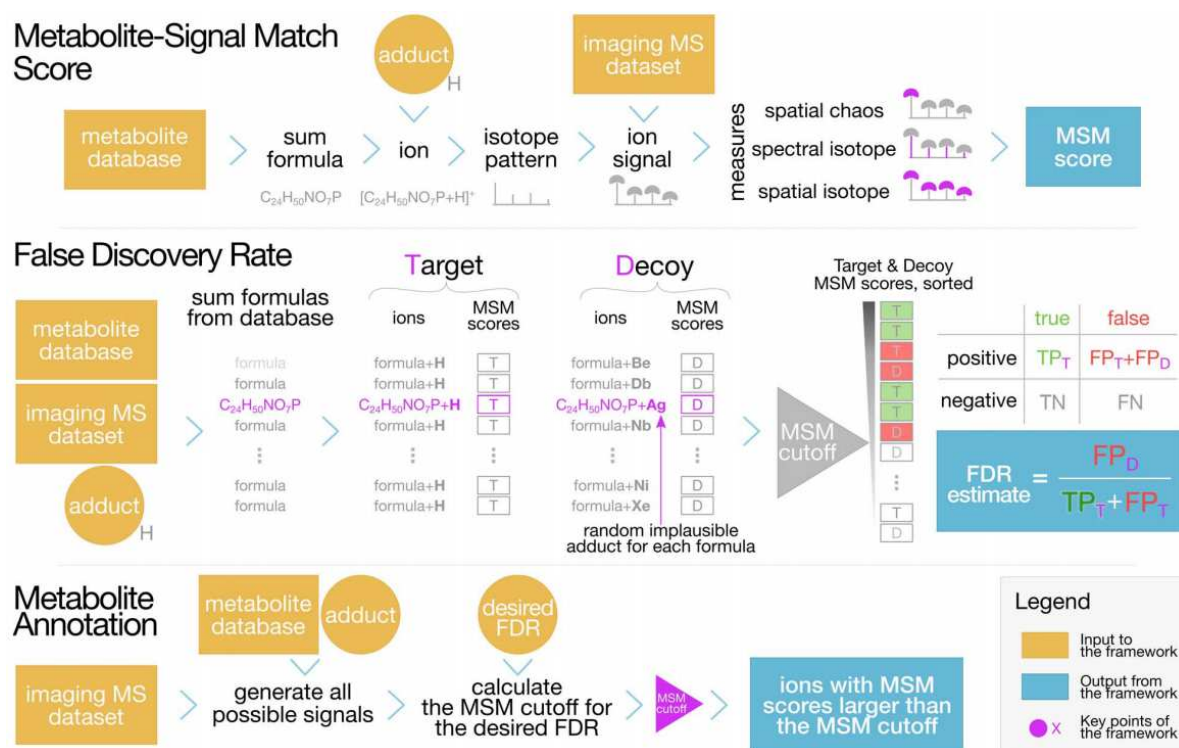


Figure 1. The proposed framework for metabolite annotation for HR imaging MS; for a more detailed scheme, see Supplementary Note 1, Figure S2.

Evaluation of the FDR approach

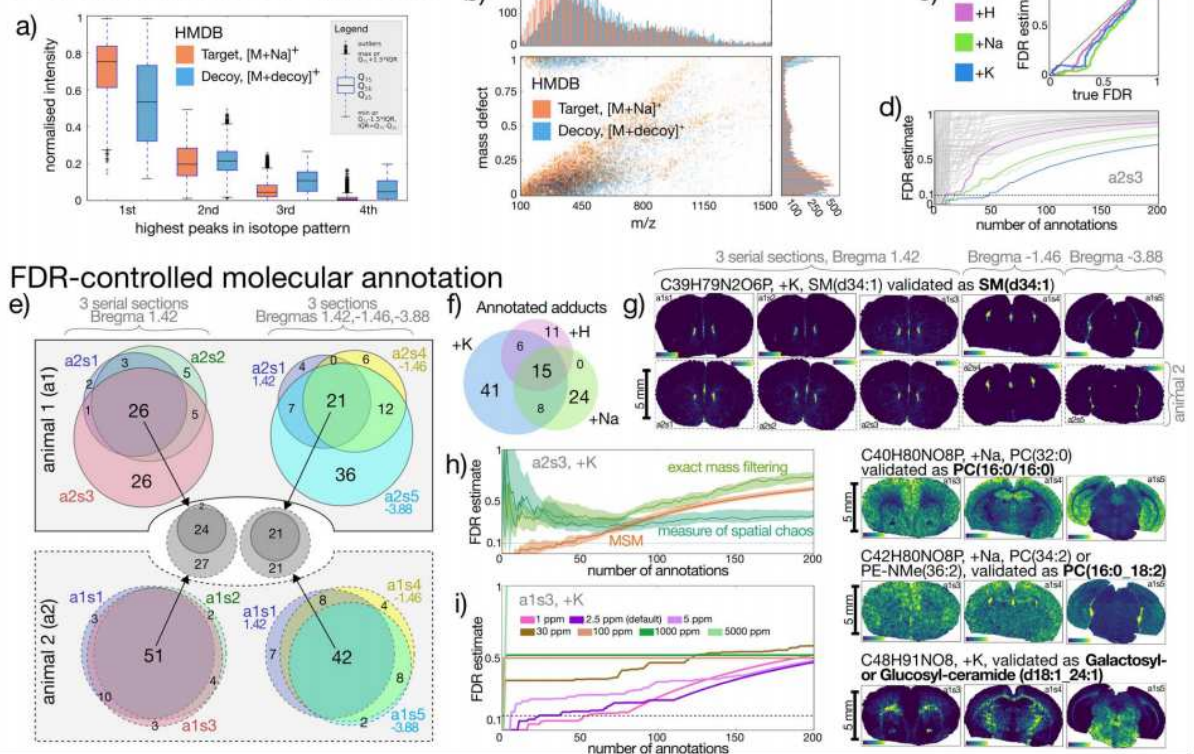


Figure 2. Evaluation of the proposed framework: a) intensities of highest four peaks in the isotope patterns for the ions from HMDB with +Na as the target adduct vs. one randomly sampled decoy set, b) a Kendrick plot for the same target and decoy as in Figure 2a (see Supplementary Note 1, Figure S5 for other adducts), c) the estimated FDR vs. true FDR for a simulated dataset (cf. Supplementary Note 1, Figure S6), and d) negative control experiment showing in grey the FDR curves for implausible adducts (cf. Supplementary Note 1, Figure S6). Showcasing the FDR-controlled metabolite annotation for ten MALDI-FTICR imaging MS datasets from mouse brain coronal sections (HMDB, $FDR_{desired}=0.1$): e) Venn diagrams showing overlaps in metabolite annotations (sum formulas) between the datasets from two animals (see Supplementary Note 1, Table S5 for breakdown about the annotations), f) overlaps between adducts of the annotations, g) examples of molecular ion images for annotations validated using LC-MS/MS (cf. Supplementary Note 2, Figures S11 and S12; Supplementary Note 3), as well as FDR curves illustrating h) superiority of MSM as compared to individual measures for a2s3, +K (see Supplementary Note 1, Figure S8 for other datasets and adducts), and g) decrease of number of annotations when simulating

lower mass resolution/accuracy for a1s3, +K (cf. Supplementary Note 1, Figure S9).

Tables

N/A

Online Methods

1. Imaging mass spectrometry

1.1 Imaging mass spectrometry data from mouse brain samples

Samples

Two female adult wild-type C57 mice (a1, a2) were obtained from Inserm U1085 - Irset Research Institute (University of Rennes1, France). Animals were age 60 days and were reared under ad-lib conditions. Care and handling of all animals complied with EU directive 2010/63/EU on the protection of animals used for scientific purposes. The whole brain was excised from each animal immediately post-mortem and are loosely wrapped rapidly in an aluminum foil to preserve their morphology and snap-frozen in liquid nitrogen. Frozen tissues were stored at -80 °C until use to avoid degradation.

Sample preparation

For each animal, five coronal 12 µm-thick brain sections were collected on a cryomicrotome CM3050S (Leica, Wetzlar, Germany) as follows. Three consecutive sections were acquired at the Bregma distance of 1.42 mm (sections s1, s2, s3) and two further sections were acquired at the Bregma distances of -1.46 and -3.88 mm (datasets s4 and s5). The sections were thaw-mounted onto indium tin oxide (ITO) coated glass slides (Bruker Daltonics, Bremen, Germany) and immediately desiccated. Alpha-Cyano-4-hydroxycinnamic acid (HCCA) MALDI-matrix was applied using the ImagePrep matrix deposition device (Bruker Daltonics). The method for matrix deposition was set as described: after an initialization step consisting in between 10-15 cycles with a spray power at 15%, an incubation time of 15 s and a drying time of 65 s, 3 cycles were performed under sensor control with a final voltage difference at 0.07 V, a spray power at 25%, an incubation time of 30 s and a drying time under sensor control at 20% and a safe dry of 10 s; then 6 cycles were performed under sensor control with a final voltage difference at 0.07 V, a spray power at 25%, an incubation time of 30 s and a drying time under sensor control at 20% and a safe dry of 15 s; 9 cycles were performed under sensor control with a final voltage difference at 0.2 V, a spray power at 15%, an incubation time of 30 s and a drying time under sensor control at 20% and a safe dry of 50 s; finally 20 cycles were performed under sensor control with a final voltage difference at 0.6 V (+/- 0.5 V), a spray power at 25%, an incubation time of 30 s and a drying time under sensor control at 40% and a safe dry of 30 s.

Imaging mass spectrometry

For MALDI-MS measurements the prepared slides were mounted into a slide adapter (Bruker Daltonics) and loaded into the dual source of a 7T FTICR mass spectrometer solariX XR (Bruker Daltonics) equipped with a Paracell, at the resolving power $R=130000$ at m/z 400. The x-y raster width was set to $50\mu\text{m}$ using smartbeam II laser optics with the laser focus setting 'small' ($20\text{-}30\ \mu\text{m}$). For a pixel, a spectrum was accumulated from 10 laser shots. The laser was running at 1000 Hz and the ions were accumulated externally (hexapole) before being transferred into the ICR cell for a single scan. For animal a1, each spectrum was internally calibrated by one-point correction using a known phospholipid with the ion $\text{C}_{42}\text{H}_{82}\text{NO}_8\text{P}+\text{K}^+$, at the m/z 798.540963. For animal a2, every spectrum was internally calibrated by several point correction using: matrix cluster of HCCA [$\text{C}_{20}\text{H}_{14}\text{N}_2\text{O}_6+\text{H}^+$, m/z 379.092462] if present and known phospholipids present in the mouse brain [$\text{C}_{40}\text{H}_{80}\text{NO}_8\text{P}+\text{H}^+$, m/z 734.569432] and [$\text{C}_{42}\text{H}_{82}\text{NO}_8\text{P}+\text{K}^+$, m/z 798.540963]. Data was acquired for the mass range $100 < m/z < 1200$ followed by a single zero filling and a sin-apodization. Online feature reduction was performed in the ftmsControl software, version 2.1.0 (Bruker Daltonics) to return only the peak centroids and intensities.

Signal processing

Centroid data was exported into the imzML format by using the SCiLS Lab software, version 2016a (SCiLS, Bremen, Germany). Ion images were generated with the tolerance ± 2.5 ppm. A hot-spot removal was performed for each image independently by setting the value of 1% highest-intensity pixels to the value of the 99th percentile followed by an edge-preserving denoising using a median 3x3-window filter.

Data availability

The imaging mass spectrometry data is publicly available at the MetaboLights repository under the accession numbers MTBLS313.

1.2 Simulated imaging mass spectrometry data

An imaging MS dataset was simulated that contained 300 sum formulas from the HMDB metabolite database, version 2.5, and 300 randomly generated formulas not contained in HMDB. To each sum formula, either +H, +Na, or +K adduct was randomly assigned. Random sum formulas were generated such that the probability distributions of the number of CHNOPS atoms, the C-H ratio, and the C-O ratio are the same as all formulas from HMDB. Isotope patterns were generated for each formula at a resolving power of $R=140000$ at m/z 400. Each isotope pattern was multiplied by a random intensity in the range [0.2-1.0]. The patterns were assigned to one of two partially overlapping square regions: one with sum formulas from HMDB, the other with sum formulas not from HMDB. Additionally 700 peaks at randomly selected m/z -values were added independently to each spectrum so that a spectrum inside one of the squares would have 3500 ± 127 peaks. The resulted line spectra were then convolved with a Gaussian function with the sigma equal to 0.015.

Data availability

The simulated imaging mass spectrometry data is publicly available at the MetaboLights repository under the accession numbers MTBLS317.

2. Metabolite-Signal Match score

Individual measures used in the Metabolite-Signal Match (MSM) score were defined based on the ion images generated from each peak within the isotope pattern for a particular sum formula and adduct. Isotope envelopes were predicted for an ion (sum formula plus adduct) at the mass resolution of the dataset and peak centroids were detected.

Code availability

The reference implementation of the developed framework is freely available at <https://github.com/alexandrovteam/pySM> as open source under the permissive license Apache 2.0.

2.1 Measure of spatial chaos

The measure of spatial chaos (Algorithm OM1) quantifies whether the principal ion image is informative (structured) or non-informative (noise). This approach was previously proposed by us for image-based peak picking²¹ but here we developed an improved measure based on the concept of level sets earlier applied for image segmentation²². For an ion image, its range of intensities is split into a number n_{levels} of levels. For each level, a level set is calculated as an 0-or-1-valued indicator set having 1-values for pixels with intensities above the level. Then, the number of closed 1-valued objects (connected areas of 1-valued pixels) in the produced level set is computed. Images with structure tend to exhibit a small number of objects that simply shrink in size as the threshold increases whilst images with a noisy distribution produce a great number of objects as the pixels above the threshold level are randomly spatially distributed (see Figure S3a). The algorithm was inspired by a concept of computational topology called persistent homology²³. The proposed measure of spatial chaos returns a value ρ_{chaos} between zero and one which is high for spatially-structured images and low for noisy images.

Input: Real-valued image B , number of levels n_{levels}

Output: measure of spatial chaos ρ_{chaos}

Algorithm:

- ```
// scale image intensity range to [0 1]
1. $B' = \frac{B - \min(B)}{\max(B) - \min(B)}$
// main part
2. For n in $\{[0, 1, \dots, n]_{levels}\}$:
// threshold image at a current level
3. $t_n = \frac{n}{n_{levels}}$
4. $B_{bw} = B' > t_n$
// fill single-pixel holes
5. $\widehat{B}_{bw} = \psi_s(B_{bw})$
// count separate objects with 4-connectivity
```

6.  $n_{objects_n} = \text{imlabel}(\widehat{B}_{bw})$
7.  $\rho_{chaos} = 1 - \frac{\sum_{n=1}^{n_{levels}} n_{objects_n}}{n_{pixels} * n_{levels}}$
8. return  $m_c$

**Algorithm OM1.** The level-sets based algorithm for calculating the measure of spatial chaos of an ion image.  $\psi_s(A)$  is a hole-filling operation to ‘fill in’ isolated missing pixels that can happen in HR imaging MS (and to avoid overestimating the number of objects). It consists of a sequence of morphological operation:  $\psi_s(A) = (A \ominus S) \oplus S'$  with structuring elements  $S' = [[010], [111], [010]]$ ,  $S = [[111], [111], [111]]$ <sup>24</sup>. `imlabel` uses the label function from `scipy`<sup>25</sup> with 4-connectivity and returns the number of disconnected objects in an image.

The computational complexity of the level-sets algorithm is  $O(N * n_{levels})$  where  $N$  is the number of pixels. The  $n_{levels}$  parameters controls the smoothness of the curve seen in Figure S3b and above a certain granularity the value of  $\rho_{chaos}$  stabilises to a constant for a particular image. A value of  $n_{levels} = 30$  was found to be sufficient to provide stable results for both the test images from<sup>2</sup> and random noise (data not shown).

## 2.2 Spatial isotope measure

The spatial isotope measure quantifies the spatial similarity between the ion images of isotopic peaks, composing a signal for a sum formula. It is calculated as a weighted average linear correlation between the ion image from the most intense isotope peak ( $i = 1$ ) and all others ( $i = 2, \dots, p$ ) where  $p$  is the number of theoretically predicted isotope peak centroids for a particular sum formula and adduct with an intensity greater than 1% of the principal (largest) peak. Each image  $i$  is weighted by the relative abundance of the theoretical isotope peak height  $a_i$ . Negative values are set to zero so the spatial isotope measure returns a value  $\rho_{spatial}$  between zero and one; the higher values imply a better match.

$$\rho_{spatial} = \frac{1}{\sum_{i=1}^p a_i} \sum_{i=2}^p a_i \text{corr}(b_1, b_i), \quad \text{where } \alpha = 1 \text{ if } \text{corr}(b_1, b_i) > 0, \quad \text{otherwise } \alpha$$

**Equation OM1.** Spatial isotope measure quantifying the spatial similarity of each isotope peak to the principal peak where `corr()` returns Pearson’s correlation coefficient and where  $b_i$  is a vector of intensities from ion image  $B_i$  of the  $i$ ’th isotope peak.

## 2.3 Spectral isotope measure

The spectral isotope measure quantifies the spectral similarity between a predicted isotope pattern and measured spatial intensities. It is calculated as the average difference between normalised predicted isotope ratios and normalised measured intensities, reported so that larger values imply a better match.

$$P_{\text{spectral}} = 1 - \frac{1}{p} \sum_{i=1}^p \left\| \frac{b'_i}{\sum b'_i} - \frac{a_i}{\sum a_i} \right\|_1$$

**Equation OM2.** Spectral isotope measure quantifying the spectral similarity between a predicted isotope pattern and the measured intensities of a signal.

In Equation 2,  $b' = \frac{\sum b_1}{n} \dots \frac{\sum b_i}{n}$  is a vector containing the mean image intensity from the ion images  $B = B_1, \dots, B_p$  for the  $n$  pixels in  $B_1$  with non-zero intensity values and  $\frac{x}{\|x\|_2}$ , where  $\|\cdot\|_p = (\sum_{i=1}^n |x_i|^p)^{1/p}$ . This can be considered as projecting both theoretical and empirical isotope patterns onto a sphere and then calculating one minus the average coordinate difference.

## 2.4 Metabolite-Signal Match score

The Metabolite-Signal Match (MSM) score quantifies the similarity between the theoretical signal of a sum formula and its measured counterpart, with the higher value corresponding to higher similarity. It is calculated according to Equation OM3, as a product of the individual measures: measure of spatial chaos, spatial isotope measure and spectral isotope measure). This puts an equal weighting on all measures whilst penalizing any annotation that gets low value for any of the measures.

$$MSM = P_{\text{chaos}} * P_{\text{spatial}} * P_{\text{spectral}}$$

**Equation OM3.** Metabolite-Signal Match (MSM) score quantifying similarity between a theoretical signal of a sum formula and its counterpart sampled from the dataset.

## Section OM3. False Discovery Rate-controlled metabolite annotation

### 3.1 Molecular annotation

First, we consider all unique sum formulas from a metabolite database of interest. We used the Human Metabolome Database (HMDB), v. 2.5, considering only 7708 carbon-containing sum formulas<sup>26</sup>. Then, we select a list of potential ion adducts. The adducts +H, +Na and +K were used as the adducts commonly detected during tissue MALDI imaging MS in the positive mode<sup>27</sup>. Then, we perform molecular annotation of an imaging MS dataset for each ion (combination of a sum formula and an adduct) independently as described in Algorithm OM2. Note that in this algorithm the MSM threshold  $t_{MSM}$  needs to be specified; for the updated algorithm selecting the MSM threshold in an FDR-controlled way, please see Algorithm OM3.

### Code availability

The reference implementation of the developed framework is freely available at <https://github.com/alexandrovteam/pySM> as open source under the permissive license Apache 2.0.

**Input:** Metabolite sum formula, adduct, charge, resolving power of the spectra, imaging MS dataset, MSM threshold  $t_{MSM}$

**Output:** Decision whether the ion  $[sum\ formula + adduct]^{charge}$  is present in the dataset

**Algorithm:**

*// Predict isotopic patterns*

1. Predict the isotope envelope\* for  $[sum\ formula + adduct]^{charge}$  at the resolving power

2. Detect centroids of the isotope envelope\*, exact  $m/z$ 's and relative intensities ( $m, a$ )

*// Generate and score signals from the dataset*

3. For  $m_i$  in  $m$  :

4. Generate an ion image  $B_i$  for the  $i$ 'th isotopic peak at  $m/z\ m_i$

5. Calculate  $P_{chaos}$  from  $B_1$  and  $P_{spatial}, P_{spectral}$  from  $B$  and  $a$  according to Algorithm 1,

Equation OM1, and Equation OM2, respectively

6. Calculate the  $MSM$  score according to Equation OM3

*// Annotate the data*

7. If  $MSM > t_{MSM}$  :

8. the ion  $[sum\ formula + adduct]^{charge}$  is annotated as being present in the dataset

**Algorithm OM2.** MSM-based molecular annotation determining whether a metabolite ion is present in an imaging MS dataset.

## 3.2 Calculation of the False Discovery Rate

To calculate the False Discovery Rate among the molecular annotations provided using Algorithm OM2 with an MSM threshold  $t_{MSM}$ , we developed a target-decoy approach similar to (Elias and Gygi 2007)<sup>28</sup>. The innovative part of his development is in applying the target-decoy approach in the spatial metabolomics context by defining a decoy set appropriate for metabolomics.

A target set was defined as a set of molecular ions for the sum formulas from a metabolite database (e.g. HMDB), with a given ion adduct type (e.g. +H, +Na, +K). A decoy search was defined as a set of implausible ions for the same sum formulas but with implausible ion adduct types. For each sum formula, an implausible elemental adduct is randomly chosen from the CIAAW 2009 list of isotopic compositions of the elements<sup>29</sup> excluding the plausible adducts, namely from He, Li, Be, B, C, N, O, F, Ne, Mg, Al, Si, P, S, Cl, Ar, Ca, Sc, Ti, V, Cr, Mn, Fe, Co, Ni, Cu, Zn, Ga, Ge, As, Se, Br, Kr, Rb, Sr, Y, Zr, Nb, Mo, Ru, Rh, Pd, Ag, Cd, In,

Sn, Sb, Te, I, Xe, Cs, Ba, La, Ce, Pr, Nd, Sm, Eu, Gd, Tb, Dy, Ho, Ir, Th, Pt, Pu, Os, Yb, Lu, Bi, Pb, Re, Tl, Tm, U, W, Au, Er, Hf, Hg, Ta. Once the target and decoy sets are defined, the MSM scores are calculated for all target and decoy ions.

The MSM cutoff ( $t_{MSM}$ ) is a key parameter of the molecular annotation. Setting the MSM cutoff changes the number of molecular annotations made. For any MSM cutoff, we define *positives* as the ions with MSM scores above the cutoff and *negatives* as the ions with MSM scores below the cutoff. We define  $FP_{decoy}$  as *positive hits* from the decoy. Since any decoy ion is constructed to be implausible, all decoy ions detected as positive are false positives. Then, we estimate FDR with FDR' according to Equation OM4.

$$FDR = \frac{FP_{target}}{TP_{target} + FP_{target}}, \quad FDR' = \frac{FP_{decoy}}{TP_{target} + FP_{target}} = \frac{n_{decoy}}{n_{target}}$$

**Equation OM4.** Definition of FDR and the proposed estimate of FDR (FDR').  $FP$ ,  $TP$  are False Positive and respectively True Positive,  $n_{target}$  and  $n_{decoy}$  are the numbers of annotations from the target and decoy sets for the MSM cutoff  $t_{MSM}$ .

Similar to the approach of FDR calculation in genome-wide studies proposed by (Storey & Tibshirani, 2003)<sup>30</sup> and picked up later in proteomics, Equation OM4 proposes an

approximation of the true FDR defined as  $\frac{FP_{target}}{P_{target}}$ . This approach relies on having a high similarity between false-positives in the target set and the decoy set. The decoy set must be the same size as the target set and share the same statistical distributions as used by the measures used in the annotation. If these assumptions are satisfied then the number of false positives from the decoy ( $FP_{decoy}$ ) approximates the number of false positives from the target ( $FP_{target}$ ) while the denominator ( $P_{target}$ ) is equal between FDR and FDR'.

As the decoy generation is a randomized process, with one decoy search formed by a sampling of implausible adducts from all possible implausible adducts, FDR calculation is a repeated sampling process. We propose to repeat it (20 times for the presented results) and calculate the median of the observed FDR values. We favored median over mean for its general robustness to outliers and for providing integer values that can be translated into the numbers of annotations.

## 4. FDR-controlled molecular annotation

The term *FDR-controlled molecular annotation* means that parameters of molecular annotation are optimized so that the set of provided annotations has a desired level of FDR. This is the most widely used approach in proteomics for choosing parameters of molecular identification<sup>31</sup>. We employed this approach to develop in Algorithm OM3 for selecting a key parameter of the molecular annotation, the MSM cutoff  $t_{MSM}$ . This was performed similarly to (Zhang et al., 2012)<sup>32</sup> by simultaneously sorting the MSM values for the target and decoy ions, decreasing the MSM cutoff thus one-by-one increasing the number of target ions annotated, recalculating the FDR after every new ion is annotated, and selecting the maximal number of annotations that provide FDR below the desired value (see Figure 1 in



the main text). This process is repeated 20 times with the decoy adducts every time randomly sampled from the set of all considered implausible adducts and an observed  $t_{MSM}$  recorded. After repetitions, the final MSM cutoff value  $t_{MSM}$  is set at the median of the observed  $t_{MSM}$  values. The final set of molecular annotations is a set of target ions with the MSM scores above the median cutoff value.

**Input:** Metabolite database, resolving power of the mass spectrometer used, imaging MS dataset, ion charge, target adduct, decoy adducts, desired FDR level  $FDR_{desired}$ , number of decoy samplings

**Output:** A set of molecular annotations (ions from the metabolite database detected as present in the dataset)

**Algorithm:**

- // Predict and score all metabolite signals*
1. For *sum formula* in *database* :
  2.  $ion_{target} = [sum\ formula + target\ adduct]^{charge}$
  3. Calculate  $MSM_{ion_{target}}$  according to Algorithm OM2.(1-3)
  4.  $ion_{decoy} = [sum\ formula + decoy\ adduct]^{charge}$ , where decoy adduct is randomly chosen from the list of decoy adducts
  5. Calculate  $MSM_{ion_{decoy}}$  according to Algorithm OM2.(1-3)
- // Calculate the MSM cutoff corresponding to the desired FDR level*
6. Form a combined vector of *MSM* values  $s = MSM_{ion_{target}} \cup MSM_{ion_{decoy}}$
  - // Find the maximal number of annotations providing FDR below  $FDR_{desired}$*
  7. Sort  $s$  in descending order.
  8.  $i = 2 * n_{molecules}$
  9. While  $FDR > FDR_{desired}$  :
  10.
$$n_{target} = \sum_{j=0}^{n_{molecules}} p_j \quad \text{where } p_j = 1 \text{ if } MSM_{ion_{target}j} < s_i, \quad \text{otherwise } p_j = 0$$
  11.
$$n_{decoy} = \sum_{j=0}^{n_{molecules}} p_j \quad \text{where } p_j = 1 \text{ if } MSM_{ion_{decoy}j} < s_i, \quad \text{otherwise } p_j = 0$$
  12. Calculate  $FDR$  according to Equation OM4
  13.  $i := i - 1$
  14.  $t_{MSM} = s_i$
  15. Repeat steps 1-11 according to the number of decoy samplings,  $n$   
 $t = [t_{MSM_1} t_{MSM_2} \dots t_{MSM_n}]$
  16.  $t_{MSM} = median(t)$
- // Perform the MSM-based molecular annotation with the calculated cutoff*
17. For *sum formula* in *database* :
    - a. If  $MSM_{ion_{target}} > t_{MSM}$  then add  $ion_{target}$  into the list of molecular annotations

**Algorithm OM3.** FDR-controlled molecular annotation that screens for metabolite ions present in an imaging MS dataset, with the desired FDR level.

## 5. LC-MS/MS validation of annotations

### 5.1 Samples

#### *Mouse brain sample*

One female adult wild-type C57 mouse age 10 weeks was obtained from the European Molecular Biology Laboratory animal resource (EMBL-LAR, Heidelberg, Germany). The animal was reared under ad-lib conditions within the specific pathogen free facility. Care and handling of the animal complied with EU directive 2010/63/EU on the protection of animals used for scientific purposes. The whole brain was excised from each animal immediately post-mortem and rapidly cryo-frozen in CO<sub>2</sub> cooled isopentane. Tissue was stored at -80 °C until use.

#### *Authentic lipid standards and chemicals*

All lipid standards used for validation of annotations were purchased from Sigma Chemicals (Sigma-Aldrich Co., St. Louis, MO) and Avanti Polar Lipids (Alabaster, LA, USA). The LC-MS grade buffers and other reagents were purchased from Sigma Chemical. All mass spectrometry grade solvents and MiliQ grade water was used throughout the analysis.

#### *Sample preparation*

20 mg of brain tissue was extracted using Bligh and Dyer extraction method<sup>33</sup>. The dried extract was reconstituted with 100 µL of methanol and isopropanol (1:1) and 10µL of this sample solution was injected LC-MS system for each run. Lipid standards were prepared in same solvent with concentration of 100 ng/mL each.

### 5.2 LC-MS/MS methods

The separation of lipids was carried out on Agilent 1260 liquid chromatography (LC) system with Ascentis® Express C<sub>18</sub> column (100 x 2.1 mm; 2.7µM) and detected with high resolution mass spectrometry (Q Exactive Plus MS, Thermo Scientific).

Three LC-MS/MS methods were used: **Positive:** ESI positive mode using 'buffer 1'. **Negative 1:** ESI negative mode method using 'buffer 1'. **Negative 2:** ESI negative mode method used 'buffer 2'. LC was run with flow rate of 0.25 ml/min with solvent A consisted of acetonitrile–water (6:4) and solvent B of isopropyl alcohol–acetonitrile (9:1), which are buffered with either 10 mM ammonium formate and 0.1% formic acid (buffer 1) or 10 mM ammonium acetate (buffer 2). MS parameters (Tune ,Thermo Scientific) were set as: spray voltage of 4 kV, sheath gas 30 and auxiliary gas 10 units, S-Lens 65 eV, capillary temperature 280 °C and vaporisation temperature of auxiliary gas was 280 °C.

Data was acquired in full scan mode in mass range of 150-900  $m/z$  (resolving power  $R=70000$ ) and data dependent tandem mass spectra (MS/MS) were obtained for all precursors from an inclusion list (resolving power  $R=35000$ ). Tandem mass spectra (MS/MS) were acquired using higher energy collisional dissociation (HCD) with normalized collision energies of 10, 20 and 30 units at the mass. The inclusion list was composed of all annotations provided from imaging MS analysis and detected in all three serial sections (s1, s2, s3 at the Bregma 1.42) for either of two animals. We considered adducts relevant for LC-MS (+H, +NH<sub>4</sub>, +Na for the Positive method; -H, -H+HCOOH for the Negative methods).

#### *Data availability*

The LC-MS/MS data from mouse brain samples is publicly available at the MetaboLights repository under the accession numbers MTBLS378.

### 5.3 LC-MS/MS validation strategy

LC-MS/MS validation of lipid annotations was performed differently for annotations when lipid standards are available and for other annotations. When lipid standards were available, LC-MS/MS information in particular the LC retention time (RT), MS and MS/MS (MS<sub>2</sub>) was used to compare the data from a standard with the data from a sample (both acquired using exactly the same LC-MS method and precursor selection range). First, extracted ion chromatograms (XICs) were evaluated for all possible adducts to confirm the presence of the ion of the sum formula obtained from imaging data. As for the tolerance value for XICs: for data with standards we used the 5 *ppm*; for data with no standards we selected the best fitting tolerance value from 2, 3, and 5 *ppm*. We considered possible adducts for each metabolite (+H, +Na, +NH<sub>4</sub> for the 'Positive' method; -H, +FA-H for the 'Negative' methods, FA stands for the formic acid) and selected the best matching adduct as follows. The precursor delta  $m/z$  was calculated for the sample both in MS<sub>1</sub> and MS/MS data. The matching MS/MS spectrum was searched within the elution profile and manually interpreted for fragments corresponding to head-group and fatty acid side chains. Only precursor and fragments with accuracy <6 ppm were considered for structural interpretation to identify possible lipid species. The lipid class was confirmed by the presence of head-group fragment or its neutral loss (e.g. MS/MS fragment with  $m/z$  184.0735 corresponds to the phosphocholine head-group). Since lipids from the classes of phosphatidylcholines (PC) and sphingomyelins (SM) have the same head-group ( $m/z$  184.0735), given a sum formula, we searched in HMDB and SwissLipids to rule out a possibility of the sum formula to correspond to a lipid from another class other than annotated by our framework. Further to confirm the fatty acid side chains, the 'Negative' LC-MS methods were used (e.g. fatty acid fragments for phosphocholines were obtained after fragmentation of formate ion precursors using the 'Negative' LC-MS method). The collision energy was selected as best representing the precursor and the expected fragments. When standards were available, the RT, precursor  $m/z$  and MS/MS fragments corresponding to head-groups and fatty acid chains from the sample were matched with spectra from the corresponding standard. When standards were not available the fragments were manually interpreted. Finally, structural annotation of the matching peaks in the MS/MS spectra was performed with the help of the HighChem MassFrontier software (Thermo Scientific). The MS, MS/MS and RT (for standards) data is presented in Supplementary Note 3 and summarized in Table S5.

## Methods-only References

21. Alexandrov, T. & Bartels, A. Testing for presence of known and unknown molecules in imaging mass spectrometry. *Bioinformatics* **29**, 2335–2342 (2013).
22. Vese, L. A. & Chan, T. F. A Multiphase Level Set Framework for Image Segmentation Using the Mumford and Shah Model. *Int. J. Comput. Vis.* **50**, 271–293
23. Edelsbrunner, H. & Harer, J. Persistent homology-a survey. *Contemp. Math.* **453**, 257–282 (2008).
24. Glasbey, C. A. & Horgan, G. W. *Image analysis for the biological sciences*. **1**, (Wiley Chichester, 1995).
25. SciPy.org — SciPy.org. Available at: <http://www.scipy.org/>. (Accessed: 12th January 2016)
26. Wishart, D. S. *et al.* HMDB 3.0--The Human Metabolome Database in 2013. *Nucleic Acids Res.* **41**, D801–7 (2013).
27. Carter, C. L., McLeod, C. W. & Bunch, J. Imaging of phospholipids in formalin fixed rat brain sections by matrix assisted laser desorption/ionization mass spectrometry. *J. Am. Soc. Mass Spectrom.* **22**, 1991–1998 (2011).
27. Elias, J. E. & Gygi, S. P. Target-decoy search strategy for increased confidence in large-scale protein identifications by mass spectrometry. *Nat. Methods* **4**, 207–214 (2007).
29. Berglund, M., Michael, B. & Wieser, M. E. Isotopic compositions of the elements 2009 (IUPAC Technical Report). *J. Macromol. Sci. Part A Pure Appl. Chem.* **83**, (2011).
30. Storey, J. D. & Tibshirani, R. Statistical significance for genomewide studies. *Proc. Natl. Acad. Sci. U. S. A.* **100**, 9440–9445 (2003).
31. Choi, H., Fermin, D. & Nesvizhskii, A. I. Significance analysis of spectral count data in label-free shotgun proteomics. *Mol. Cell. Proteomics* **7**, 2373–2385 (2008).
32. Zhang, J. *et al.* PEAKS DB: De Novo Sequencing Assisted Database Search for Sensitive and Accurate Peptide Identification. *Mol. Cell. Proteomics* **11**, (2012).

33. Bligh, E. G. & Dyer, W. J. A rapid method of total lipid extraction and purification. *Can. J. Biochem. Physiol.* **37**, 911–917 (1959).