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

Andrew Palmer[1], Prasad Phapale[1], Ilya Chernyavsky[2], Regis Lavigne[3,4], Dominik Fay[1], Artem Tarasov[1], Vitaly Kovalev[1], Jens Fuchser[5], Sergey Nikolenko[1,6,7], Charles Pineau[3,4], Michael Becker[5] & Theodore Alexandrov[1,8,9]

High-mass-resolution imaging mass spectrometry promises to localize hundreds of metabolites in tissues, cell cultures, and agar plates with cellular resolution, but it is hampered by the lack of bioinformatics tools for automated metabolite identification. We report pySM, a framework for false discovery rate (FDR)-controlled metabolite annotation at the level of the molecular sum formula, for high-mass-resolution imaging mass spectrometry (https://github.com/alexandrovteam/pySM). We introduce a metabolite-signal match score and a target–decoy FDR estimate for spatial metabolomics.

Metabolomics is promising important advances in cell biology, physiology, and medicine. Metabolomics complements other omics technologies by analyzing the instantaneous state of biochemical processes and revealing the contributions of nongenetic factors. Metabolomics has progressed from cataloging chemical structures to answering complex biomedical questions[1,2], but it generally requires pooling of cell populations. Spatial localization of metabolites *in situ* has been an important challenge.

Arguably, the most promising technique for spatial metabolomics analysis is imaging mass spectrometry (imaging MS)[3–8], in which mass spectra are acquired directly from a section of tissue, cell culture, or agar. The key challenge in imaging MS is the molecular identification[9], particularly because conducting data-dependent tandem MS (MS/MS)-like imaging in liquid chromatography (LC)– or gas chromatography–MS/MS online is not feasible. This puts high demands on mass-based separation. The introduction of high-mass-resolution (HR) MS capable of discriminating metabolites differing by a few millidaltons promises unprecedented reliability of metabolite annotation.

However, to our knowledge, no bioinformatics tools exist for automated metabolite annotation in HR imaging MS. This has restricted this technique mainly to targeted imaging of a few metabolites only[10]. Existing approaches either rely on visual examination or are based on exact mass filtering, which can produce false positives even when ultra-HR MS is used[11]. A challenge has been the development of algorithms that are robust to strong pixel-to-pixel noise and efficient enough to mine 10–100-gigabyte data sets.

An additional obstacle is the lack of a metabolomics-compatible approach for estimating the FDR[12,13]. The FDR is defined as the ratio of false positives to the total number of annotations. The FDR is a cornerstone of the quantification of annotation quality in genomics, transcriptomics, and proteomics[14]. 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 has been proposed[15] but is limited to phytochemical metabolites, has not found widespread use, and cannot be applied to imaging MS because it does not allow for the incorporation of spatial information. An alternative approach to estimating FDR would be the use of a phantom sample with controlled molecular content, but this is inherently complex and restricted to a specific protocol.

We have addressed this double challenge by developing a comprehensive bioinformatics framework for FDR-controlled metabolite annotation for HR imaging MS. Our open-source framework pySM (https://github.com/alexandrovteam/pySM) is based on the three 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, and a novel target–decoy FDR-estimation approach with a decoy set generated via the use of implausible adducts.

Our framework takes the following as input: (i) an HR imaging MS data set in the imzML format, (ii) a database of metabolite sum formulas in a CSV format (for sections from human tissue and similar biological systems we recommend the Human Metabolome Database)[16], and (iii) an adduct of interest (for positive ion mode we recommend $+H^+$, $+Na^+$, and $+K^+$; for negative we recommend $-H^+$ and $+Cl^-$). For a specified FDR level (we recommend 0.1), the framework provides metabolite annotations for 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 formulas[17].

Our MSM score quantifies the likelihood of the presence of a metabolite with a given sum formula in the sample (**Fig. 1**, **Supplementary Fig. 1**, and **Supplementary Note 1**). For an ion

[1]Structural and Computational Biology Unit, European Molecular Biology Laboratory, Heidelberg, Germany. [2]Center for Industrial Mathematics, University of Bremen, Bremen, Germany. [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]Higher Institute of Information Technologies and Information Systems, 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 T.A. (theodore.alexandrov@embl.de).

RECEIVED 9 MARCH; ACCEPTED 13 OCTOBER; PUBLISHED ONLINE 14 NOVEMBER 2016; DOI:10.1038/NMETH.4072

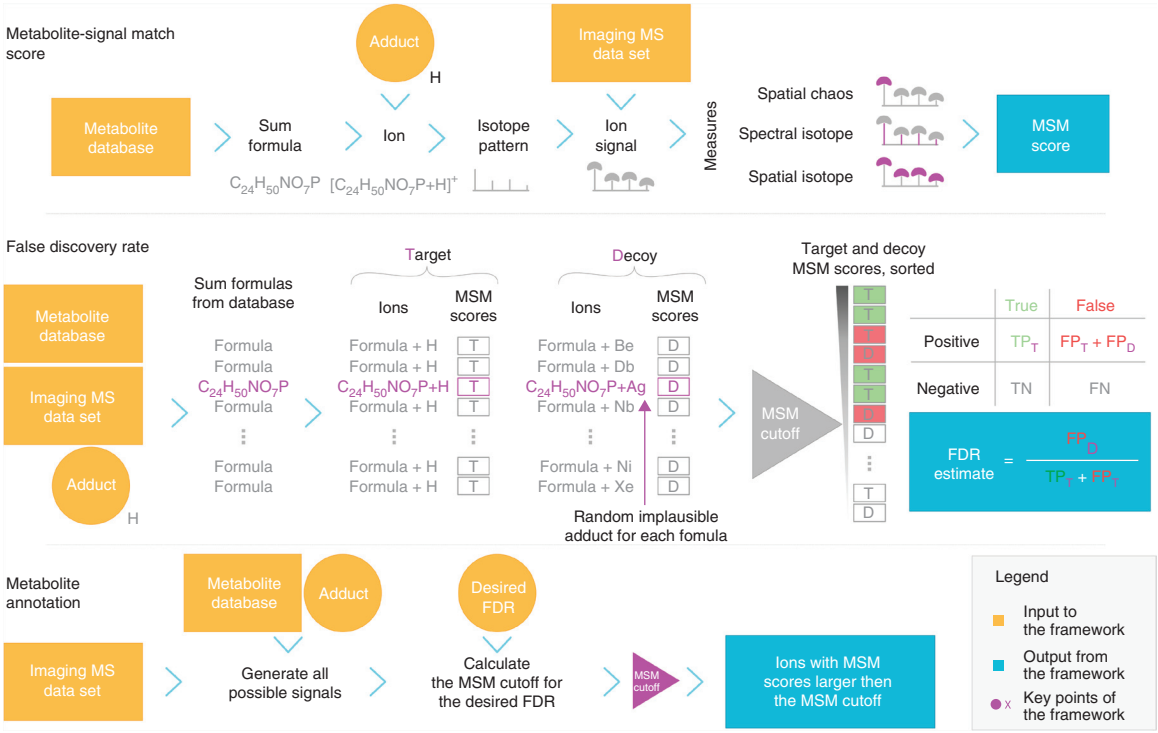

Figure 1 | The proposed bioinformatics framework for metabolite annotation for HR imaging MS. The framework has three essential parts: (top) the MSM score quantifying the likelihood of the presence of a metabolite with a given sum formula in the sample, (middle) an estimate of the FDR among annotations with MSM scores above an MSM cutoff, and (bottom) an FDR-controlled metabolite annotation that delivers annotations with a desired FDR value. For a more detailed scheme, see **Supplementary Figure 1**.

(sum formula plus ion adduct; for example, H^+), we generate its isotopic pattern, accounting for the instrument's resolving power with the isotopic fine structure if resolvable. Then we sample an ion signal from the imaging MS data set, namely, the ion images for all isotopic peaks with predicted intensity greater than 0.01% of the principal peak (**Supplementary Fig. 2** and **Supplementary Note 1**). MSM is computed by multiplication of the following measures. (i) The measure of spatial chaos quantifies spatial informativeness within the image of the principal peak[18]. We introduce an improved measure of spatial chaos (Online Methods) that outperforms earlier proposed measures[18,19] with respect to both speed and accuracy (**Supplementary Fig. 3** and **Supplementary Note 1**). (ii) The spectral isotope measure quantifies the spectral similarity between a theoretical isotopic pattern and relative sampled isotopic intensities. (iii) The spatial isotope measure quantifies spatial colocalization between isotopic ion images. An MSM score of 1 indicates the maximal likelihood of the signal corresponding to the ion.

Our FDR estimate helps experimenters select an MSM cutoff that allows them to be confident in the correspondence between the ions with MSM scores above the cutoff and metabolites from the sample (**Fig. 1**, **Supplementary Fig. 1**, and **Supplementary Note 1**). In accordance with the target–decoy approach[14], we propose the construction of a decoy set as follows. We define a target set as ions from the selected metabolite database with a given ion adduct (for example, 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[20] (for example, B^+, Db^+, or Ag^+), excluding plausible adducts. MSM scores are calculated for target and decoy ions. For any MSM cutoff, the FDR is estimated as the ratio between the numbers of decoy false positives (FP_D; the decoy ions with MSM scores above the cutoff) 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 that the target and decoy sets are similar. The sampling of implausible adducts is repeated, averaging the resulting FDR estimate.

For FDR-controlled metabolite annotation, one specifies the desired value of the FDR (for example, 0.1) and chooses the smallest MSM cutoff that provides the desired FDR (**Fig. 1**, **Supplementary Fig. 1**, and **Supplementary Note 1**). This provides annotations of a given level of confidence, independent of the MSM cutoff, data set, MS settings, and operator, and can be used for comparative and interlab 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 (**Fig. 2a**) even though the decoy ions had higher relative intensities for heavier isotopic peaks owing to their more complex isotopic patterns. The target and decoy ions were also found to have similar mass-to-charge ratios (m/z) and mass defect spaces (**Fig. 2b**), with a positive offset in m/z for decoy adducts, which typically have heavier elements. Second, we compared the estimated and true FDRs for a simulated data set with a known ground truth (**Fig. 2c** and **Supplementary Note 1**). Although there were some differences in the low-value region, the estimated FDR followed the true FDR overall. Finally, negative control experiments using each of the implausible adducts as the target showed that FDR

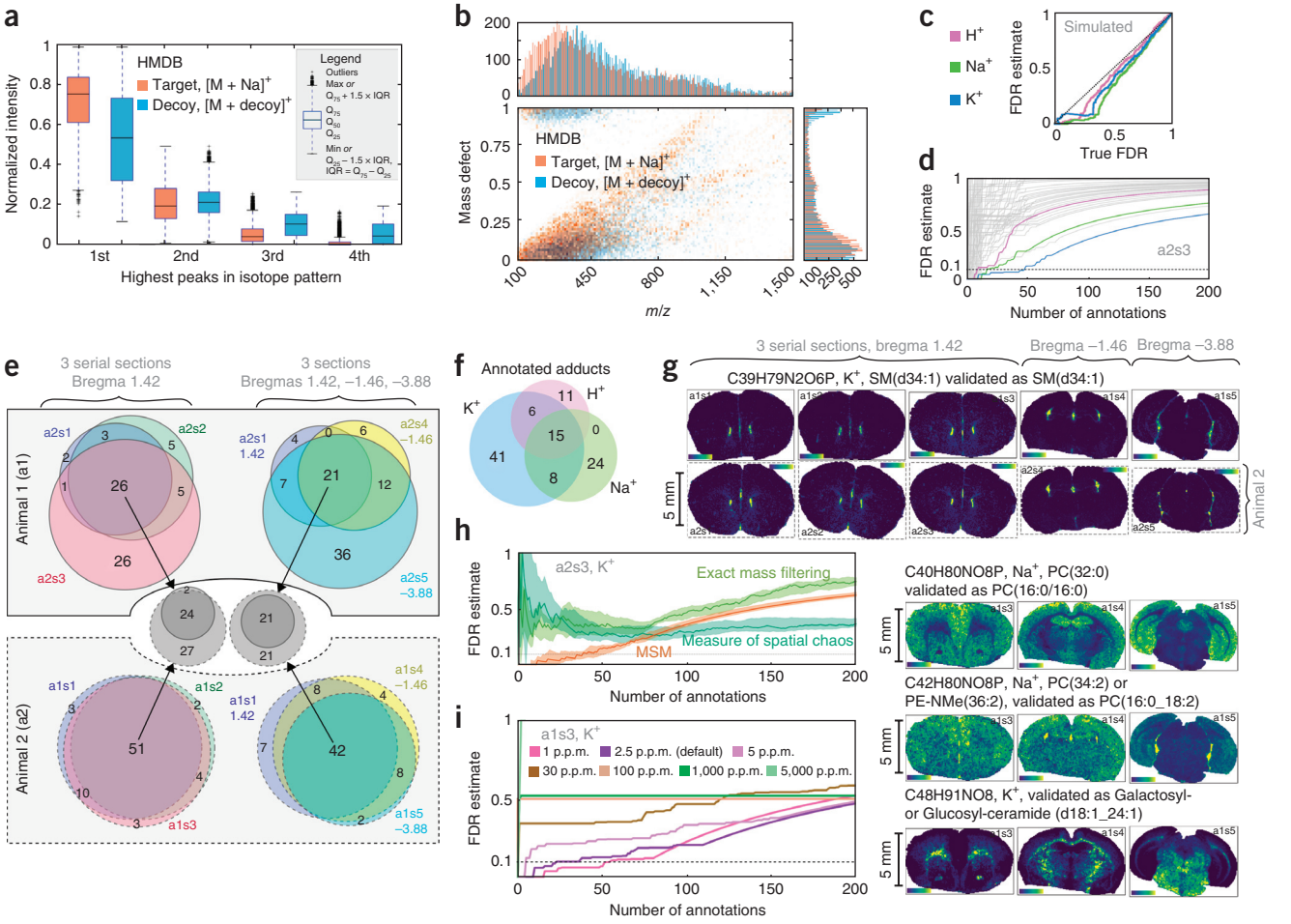

Figure 2 | Evaluation of the proposed framework. (**a**–**d**) Evaluation of the FDR approach. (**a**) Intensities of the four highest peaks in the isotope patterns for the ions from the Human Metabolomics Database (HMDB) with Na^+ as the target adduct versus one randomly sampled decoy set. (**b**) Kendrick plot of mass defect for the same target and decoy as in **a** (see **Supplementary Fig. 6** and **Supplementary Note 1** for other adducts). (**c**) Estimated FDR versus true FDR for a simulated data set (see **Supplementary Fig. 7** and **Supplementary Note 1**). (**d**) Negative control experiment showing the FDR curves for implausible adducts (gray; see **Supplementary Fig. 8** and **Supplementary Note 1**). (**e**–**i**) FDR-controlled molecular annotation. (**e**) FDR-controlled metabolite annotation for ten MALDI-FTICR imaging MS data sets from mouse brain coronal sections (HMDB; $FP_D = 0.1$). The Venn diagrams show overlaps in metabolite annotations (sum formulas) between the data sets from two animals (see Table SN1.2 in **Supplementary Note 1** for a breakdown of these annotations). (**f**) Overlap between adducts of the annotations. (**g**) Selected examples of molecular ion images validated using LC-MS/MS (see **Supplementary Note 2** for the full set of molecular images and **Supplementary Data 1** for the detailed LC-MS/MS validation). (**h**) FDR curves illustrating the application of MSM as compared to individual measures for a2s3, K^+ adduct (see **Supplementary Fig. 4** for other data sets and adducts). The curve is the median, and the shaded region shows 95% confidence interval. (**i**) Numbers of annotations when simulating lower mass resolution and/or lower accuracy for a1s3, K^+ adduct (see **Supplementary Fig. 5** and **Supplementary Note 1**).

values for implausible adducts were characteristically higher (**Fig. 2d** and **Supplementary Note 1**).

We showcased our framework on HR imaging MS data sets from fresh-frozen brain coronal sections of two (a1 and a2) female adult wild-type mice (**Supplementary Tables 1** and **2**). Five coronal sections were collected from each brain: three serial sections at the bregma +1.42 mm (s1–s3, which can be considered technical replicates), s4 at −1.46 mm, and s5 at −3.88 mm. The sections were imaged using an HR matrix-assisted laser desorption/ionization (MALDI)–Fourier transform ion cyclotron resonance (FTICR) imaging mass spectrometer in the positive mode. FDR-controlled annotation was carried out with an m/z tolerance of 2.5 p.p.m. and the desired level of FDR = 0.1 for metabolites from HMDB with H^+, Na^+, and K^+ adducts (**Fig. 2e–i**). Venn diagrams of annotated metabolites (**Fig. 2e**) showed good reproducibility between sections from the same animal (especially between the serial sections from a2, in which 51 of 73 sum formulas were annotated in all three sections) and between the animals (only two sum formulas were annotated in animal a1 only). The numbers of detected adducts were similar (**Fig. 2f**). Exemplary molecular images of annotations illustrated good reproducibility between technical replicates and animals (**Fig. 2g**); the full set of molecular images is shown in **Supplementary Note 2**. Most of the metabolites detected were phospholipids (phosphatidylcholine (PC), phosphatidylethanolamine (PE), sphingomyelin (SM), and phosphatidic acid (PA), typical for MALDI imaging MS of brain tissue using an α-cyano-4-hydroxycinnamic acid (HCCA) matrix)[21]; however, we did detect some other small molecules. From 103 overall annotations, we validated 16 representative ones with LC-MS/MS by either using authentic standards or assigning fragment structures to MS/MS data (**Supplementary Data 1**).

We demonstrated the potential of using FDR curves in two examples. First, we showed that MSM outperformed the individual measures (**Fig. 2h**, **Supplementary Fig. 4**, and **Supplementary Note 1**). The exact mass filtering performed significantly worse, achieving a lowest FDR of 0.25 for ten annotations (versus an FDR of 0 for the same number of annotations with MSM). Second, we demonstrated that the number of FDR-controlled annotations decreased with decreasing mass-resolving power (**Fig. 2i**, **Supplementary Fig. 5**, and **Supplementary Note 1**). For this, we artificially reduced the mass-resolving power by using different m/z tolerances when sampling m/z signals: 1, 2.5 (default), 5, 30, 100, 1,000, and 5,000 p.p.m. This indicated that high mass accuracy and resolution were essential for confident metabolite annotation. In **Supplementary Data 2**, we present additional results for a rat brain data set, and in **Supplementary Figure 9**, we compare results when using the SwissLipids lipids database instead of HMDB, demonstrating the general applicability of our approach.

Our framework is directly applicable to other types of HR imaging MS data collected using FTICR or orbitrap analyzers (MALDI, desorption electrospray ionization, secondary ion MS, infrared matrix-assisted laser desorption electrospray ionization, etc., with proper adducts to be selected for each source) and other types of samples (plant tissue, cell culture, agar plate, etc.) for which proper metabolite databases are available.

METHODS

Methods, including statements of data availability and any associated accession codes and references, are available in the online version of the paper.

Accession codes. The data are publicly available at the MetaboLights repository under accession codes MTBLS313, MTBLS378, and MTBLS317.

Note: Any Supplementary Information and Source Data files are available in the online version of the paper.

ACKNOWLEDGMENTS

We thank O. Vitek, A. Makarov, and M. Savitski for discussions on FDR, and D. Feichtner-Kozlov for discussions on computational topology. We acknowledge funding from the European Union's Horizon2020 and FP7 programs under grant agreements 634402 (A.P., R.L., A.T., V.K., S.N., C.P., and T.A.) and 305259 (I.C., R.L., and C.P.), and from the Russian Government Program of Competitive Growth of Kazan Federal University (S.N.). We thank EMBL Metabolomics Core Facility for instrumentation for LC-MS/MS analysis. T.A. thanks P. Dorrestein (UCSD) and P. 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

A.P. and T.A. conceived the study; A.P., I.C., D.F., A.T., S.N. and V.K. implemented the algorithms; R.L., J.F., C.P., and M.B. provided imaging data; A.P. and T.A. analyzed imaging data; P.P. collected LC-MS/MS data; P.P. and T.A. performed LC-MS/MS validation; A.P. and T.A. wrote manuscript with feedback from all other coauthors; and T.A. coordinated the project.

COMPETING FINANCIAL INTERESTS

The authors declare competing financial interests: details are available in the online version of the paper.

1. Baker, M. *Nat. Methods* **7**, 157–161 (2010).
2. Johnson, C.H., Ivanisevic, J. & Siuzdak, G. *Nat. Rev. Mol. Cell Biol.* **17**, 451–459 (2016).
3. Watrous, J.D., Alexandrov, T. & Dorrestein, P.C. *J. Mass Spectrom.* **46**, 209–222 (2011).
4. Watrous, J.D. & Dorrestein, P.C. *Nat. Rev. Microbiol.* **9**, 683–694 (2011).
5. Schwamborn, K. & Caprioli, R.M. *Nat. Rev. Cancer* **10**, 639–646 (2010).
6. Attia, A.S. *et al. Cell Host Microbe* **11**, 664–673 (2012).
7. Johnson, C.H. *et al. Cell Metab.* **21**, 891–897 (2015).
8. Soltwisch, J. *et al. Science* **348**, 211–215 (2015).
9. Palmer, A., Trede, D. & Alexandrov, T. *Metabolomics* **12**, 107 (2016).
10. Spengler, B. *Anal. Chem.* **87**, 64–82 (2015).
11. Kind, T. & Fiehn, O. *BMC Bioinformatics* **7**, 234 (2006).
12. Benjamini, Y. & Hochberg, Y. *J. R. Stat. Soc. Series B Stat. Methodol.* **57**, 289–300 (1995).
13. Storey, J.D. *J. R. Stat. Soc. Series B Stat. Methodol.* **64**, 479–498 (2002).
14. Käll, L., Storey, J.D., MacCoss, M.J. & Noble, W.S. *J. Proteome Res.* **7**, 29–34 (2008).
15. Matsuda, F. *et al. PLoS One* **4**, e7490 (2009).
16. Wishart, D.S. *et al. Nucleic Acids Res.* **41**, D801–D807 (2013).
17. Sumner, L.W. *et al. Metabolomics* **3**, 211–221 (2007).
18. Alexandrov, T. & Bartels, A. *Bioinformatics* **29**, 2335–2342 (2013).
19. Wijetunge, C.D. *et al. Bioinformatics* **31**, 3198–3206 (2015).
20. Wieser, M.E. & Coplen, T.B. *Pure Appl. Chem.* **83**, 359–396 (2011).
21. Gode, D. & Volmer, D.A. *Analyst (Lond.)* **138**, 1289–1315 (2013).

ONLINE METHODS

Imaging MS data from mouse brain samples. *Samples.* Two female adult wild-type C57 mice (a1 and a2) were obtained from Inserm U1085, IRSET Research Institute (University of Rennes 1, France). Animals were age 60 d and were reared under *ad libitum* conditions. Care and handling of all animals complied with EU directive 2010/63/EU on the protection of animals used for scientific purposes, and protocols were approved by the University of Rennes 1 Animal Experimentation Ethics Committee. The whole brain was excised from each animal immediately post-mortem, loosely wrapped rapidly in aluminum foil to preserve its 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 a distance of +1.42 mm from bregma (sections s1, s2, and s3) and two further sections were acquired at −1.46 and −3.88 mm from bregma, respectively (data sets s4 and s5, respectively). The sections were thaw-mounted onto indium tin oxide (ITO)-coated glass slides (Bruker Daltonics, Bremen, Germany) and immediately desiccated. An HCCA MALDI matrix was applied using the ImagePrep matrix deposition device (Bruker Daltonics). The method for matrix deposition was as follows: after an initialization step consisting f 10–15 cycles with a spray power at 15%, an incubation time of 15 s, and a drying time of 65 s, we carried out three cycles under sensor control with a final voltage difference of 0.07 V, a spray power at 25%, an incubation time of 30 s, a drying time under sensor control at 20%, and a safe dry of 10 s; six cycles under sensor control with a final voltage difference of 0.07 V, a spray power at 25%, an incubation time of 30 s, a drying time under sensor control at 20%, and a safe dry of 15 s; nine cycles under sensor control with a final voltage difference of 0.2 V, a spray power at 15%, an incubation time of 30 s, a drying time under sensor control at 20%, and a safe dry of 50 s; and finally 20 cycles under sensor control with a final voltage difference of 0.6 V (± 0.5 V), a spray power at 25%, an incubation time of 30 s, a drying time under sensor control at 40%, and a safe dry of 30 s.

Imaging MS. For MALDI-MS measurements, the prepared slides were mounted on a slide adaptor (Bruker Daltonics) and loaded into the dual source of a 7T FTICR mass spectrometer solariX XR (Bruker Daltonics) equipped with a Paracell at resolving power R = 130,000 at m/z = 400. The *x*-*y* raster width was set to 50 µm using Smartbeam-II laser optics with the laser focus set to 'small' (20–30 µm). For each pixel, a spectrum was accumulated from 10 laser shots. The laser was run at 1,000 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 $C_{42}H_{82}NO_8P+K^+$ at m/z = 798.5410. For animal a2, every spectrum was internally calibrated by multipoint correction using a matrix cluster of HCCA ($C_{20}H_{14}N_2O_6+H^+$, m/z = 379.0925) if present plus known phospholipids present in the mouse brain, $C_{40}H_{80}NO_8P+H^+$ at m/z = 734.5694 and $C_{42}H_{82}NO_8P+K^+$ at m/z = 798.5410. Data were acquired for the mass range 100 < m/z < 1,200 followed by a single zero filling and a sin-apodization. Online feature reduction was performed in ftmsControl software version 2.1.0 (Bruker Daltonics) to return only the peak centroids and intensities.

Signal processing. Centroid data were exported into the imzML format using SCiLS Lab software version 2016a (SCiLS, Bremen, Germany). Ion images were generated with a tolerance of ±2.5 p.p.m. We performed hot-spot removal independently for each image by setting the value of the 1% highest-intensity pixels to the value of the 99th percentile and then applying an edge-preserving denoising using a filter with a median 3 × 3 window. Individual measures used in the 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 + adduct) at the mass resolution of the data set, and peak centroids were detected.

Statistics. No statistical analyses or tests were used because no comparative group analysis was performed. Imaging MS was performed on each tissue section individually.

Simulated imaging MS data. We simulated an imaging MS data set that contained 300 sum formulas from the HMDB metabolite database (version 2.5) and 300 randomly generated formulas not contained in HMDB. An H^+, Na^+, or K^+ adduct was randomly assigned to each formula. 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 were the same as in all formulas from HMDB. Isotope patterns were generated for each formula at a resolving power of R = 140,000 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 and 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 3,500 ± 127 peaks. The resulting line spectra were then convolved with a Gaussian function with σ = 0.015.

Measure of spatial chaos. The measure of spatial chaos quantifies whether the principal ion image is informative (structured) or noninformative (noise). We have proposed this approach for image-based peak picking in the past[18], but here we developed an improved measure based on the concept of level sets previously applied for image segmentation[22]. For an ion image, its range of intensities is split into a number of levels, n_{levels}. For each level, a level set is calculated as an indicator set with a value of 0 or 1, set at 1 for pixels with intensities above the level. Then, the number of closed 1-valued objects (connected areas of 1-valued pixels) in the resulting level set is computed. Images with structure tend to exhibit a small number of objects, which shrink in size as the threshold increases, whereas images with noisy distributions produce a great number of objects because the pixels above the threshold level are randomly spatially distributed. The algorithm was inspired by a concept of computational topology called 'persistent homology'[23]and implemented using morphological image operators[24]. The proposed measure of spatial chaos returns a value ρ_{chaos} between 0 and 1, which is high for spatially structured images and low for noisy images. For an algorithmic description, please see **Supplementary Note 3**.

The computational complexity of the level-sets algorithm is $O(N \times n_{levels})$ where N is the number of pixels. The n_{levels} parameter controls the smoothness of the curve seen in **Supplementary Figure 3b**, and above a certain granularity the value of ρ_{chaos}

stabilizes 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 ref. 18 and random noise (data not shown).

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, …, 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 ρ_{chaos} between zero and one; higher values imply a better match.

$$\rho_{spatial} = \frac{1}{\sum_{i=1}^{p} a_i} \sum_{i=2}^{p} \alpha a_i \mathrm{corr}(b_1, b_i) \quad (1)$$

where α = 1 if corr(b_1,b_i) > 0; otherwise α = 0. Equation (1) is the spatial isotope measure quantifying the spatial similarity of each isotope peak to the principal peak where corr(·) returns a Pearson's correlation coefficient and b_i is a vector of intensities from ion image B_i of the ith isotope peak.

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 normalized predicted isotope ratios and normalized measured intensities, reported so that larger values imply a better match.

$$\rho_{spectral} = 1 - \frac{1}{p} \sum_{i=1}^{p} \left\| \frac{\bar{b}'_i}{\sum \bar{b}'_i} - \frac{\bar{a}'_i}{\sum \bar{a}'_i} \right\| 1 \quad (2)$$

In equation (2),

$$b' = \sum \bar{b}_1 / n, \ldots, \sum \bar{b}_i / n$$

is a vector containing the mean image intensity from the ion images $B = B_1, \ldots, B_p$ for the n pixels in B_1 with nonzero intensity values and

$$\bar{x} = x / \|x\|_2, \text{ where } \|\bullet\|_p = \left(\sum_{i=1}^{n} |x_i|^p \right)^{1/p}$$

This can be considered as projecting both theoretical and empirical isotope patterns onto a sphere and then calculating 1 minus the average coordinate difference.

MSM score. The MSM score quantifies the similarity between the theoretical signal of a sum formula and its measured counterpart sampled from the data set, with a higher value corresponding to higher similarity. It is calculated according to equation (3) as a product of the individual measures (measure of spatial chaos, spatial isotope measure, and spectral isotope measure). This puts an equal weighting on each measure while penalizing any annotation that returns a low value for any of the measures.

$$\mathrm{MSM} = \rho_{chaos} \times \rho_{spatial} \times \rho_{spectral} \quad (3)$$

FDR-controlled metabolite annotation. *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 7,708 carbon-containing sum formulas[16]. Then we select a list of potential ion adducts. Adducts H^+, Na^+, and K^+ were used as the adducts commonly detected during MALDI-MS tissue imaging in the positive mode[25]. Then we performed molecular annotation of an imaging-MS data set for each ion (combination of a sum formula and an adduct) independently as described in Algorithm 2 in **Supplementary Note 3**. 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 **Supplementary Note 3**.

Calculation of the FDR. To calculate the FDR among the molecular annotations provided using Algorithm 2 with an MSM threshold t_{MSM}, we developed a target–decoy approach similar to that of Elias and Gygi[26]. The innovative part of this 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 (for example, HMBD) with a given ion adduct type (for example, H^+, Na^+, and 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 was randomly chosen from the CIAAW 2009 list of isotopic compositions of the elements[27], 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, Er, Tm, Yb, Lu, Hf, Ta, W, Re, Os, Ir, Pt, Au, Hg, Tl, Pb, Bi, Th, U and Pu. Once the target and decoy sets were defined, the MSM scores were 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. Because 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 (4):

$$\mathrm{FDR} = \frac{\mathrm{FP}_{target}}{\mathrm{TP}_{target} + \mathrm{FP}_{target}}; \ \mathrm{FDR}' = \frac{\mathrm{FP}_{decoy}}{\mathrm{TP}_{target} + \mathrm{FP}_{target}} = \frac{n_{decoy}}{n_{target}} \quad (4)$$

where FP and TP are false positives and true positives, respectively, and n_{target} and n_{decoy} are the numbers of annotations from the target and decoy sets for the MSM cutoff t_{MSM}.

Similarly to the approach to FDR calculation in genome-wide studies proposed by Storey and Tibshirani[28] and picked up later in proteomics, equation (4) proposes an approximation of the true FDR defined as FP_{target}/P_{target}. This approach relies on having a

doi:10.1038/nmeth.4072

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 the measures used in the annotation. If these assumptions are satisfied, 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 favor median over mean for its general robustness to outliers and for providing integer values that can be translated into the numbers of annotations.

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 FDR. This is the most widely used approach in proteomics for choosing parameters of molecular identification[29]. We used this approach to select a key parameter of the molecular annotation, the MSM cutoff t_{MSM}. We did this similarly to the procedure described by Zhang *et al.*[30], by simultaneously sorting the MSM values for the target and decoy ions, decreasing the MSM cutoff, and thus one by one increasing the number of target ions annotated, recalculating the FDR after every new ion was annotated and selecting the maximal number of annotations that provided an FDR below the desired value (**Fig. 1** and **Supplementary Fig. 1**). This process was repeated 20 times with the decoy adducts each time randomly sampled from the set of all considered implausible adducts and an observed t_{MSM} recorded. After 20 repetitions, the final MSM cutoff value t_{MSM} was set at the median of the observed t_{MSM} values. The final set of molecular annotations was a set of target ions with the MSM scores above the median cutoff value. For an algorithmic description, please see Algorithm 3 in **Supplementary Note 3**.

LC-MS/MS validation of annotations. *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 libitum* 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, and protocols were approved by European Molecular Biology Laboratory Institutional Animal Care and Use Committee. The whole brain was excised from the animal immediately post-mortem and rapidly cryo-frozen in CO_2-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, Missouri, USA) and Avanti Polar Lipids (Alabaster, Louisiana, USA). The LC-MS-grade buffers and other reagents were purchased from Sigma Chemical. MS-grade solvents and MiliQ-grade water were used throughout the analysis.

Sample preparation. We extracted 20 mg of brain tissue using the Bligh and Dyer extraction method[31]. The dried extract was reconstituted with 100 µL of methanol and isopropanol (1:1), and 10 µL of this sample solution was injected into the LC-MS system for each run. Lipid standards were prepared in the same solvent with a concentration of 100 ng/mL each.

LC-MS/MS methods. The separation of lipids was carried out on an Agilent 1260 LC system with an Ascentis Express C_{18} column (100 × 2.1 mm; 2.7 µM) and detected with HR MS (Q Exactive Plus from Thermo Fisher Scientific, Bremen, Germany).

Three LC-MS/MS methods were used: (i) positive, an electrospray ionization (ESI) positive mode using 'buffer 1'; (ii) negative 1, an ESI negative mode method using buffer 1; and (iii) negative 2, an ESI negative mode method using 'buffer 2'. LC was run at a flow rate of 0.25 ml/min, with solvent A consisting of acetonitrile–water (6:4) and solvent B consisting of isopropyl alcohol–acetonitrile (9:1), which were buffered either with 10 mM ammonium formate and 0.1% formic acid (buffer 1) or with 10 mM ammonium acetate (buffer 2). MS parameters in the Tune software (Thermo Fisher Scientific) were set at a spray voltage of 4 kV, sheath gas 30 and auxiliary gas 10 units, S-Lens 65 eV, capillary temperature 280 °C, and vaporization temperature of auxiliary gas 280 °C.

Data were acquired in full scan mode in the mass range of 150–900 m/z (resolving power R = 70,000) and data-dependent MS/MS spectra were obtained for all precursors from an inclusion list (resolving power R = 35,000). Tandem mass spectra 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, and s3 at +1.42 mm from bregma) for either of the two animals. We considered adducts relevant for LC-MS (H^+, NH_4^+, and Na^+ for the positive method; H^- and $H+HCOOH^-$ for the negative methods).

LC-MS/MS validation strategy. LC-MS/MS validation of lipid annotations was performed differently for annotations for which lipid standards were available and for annotations for which lipid standards are not available. When lipid standards were available, LC-MS/MS information—in particular the LC retention time (RT), MS, and MS/MS (MS2)—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. For the tolerance values for XICs, for data with standards we used a tolerance value of 5 p.p.m.; for data with no standards we selected the best-fitting tolerance value from 2, 3 and 5 p.p.m. We considered possible adducts for each metabolite (H^+, Na^+, and NH_4^+ for the positive method; H^- and $[FA - H]^+$ for the negative methods, where FA is formic acid) and selected the best matching adduct as follows. The precursor $\Delta m/z$ was calculated for the sample in both MS1 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 accuracies of <6 p.p.m. were considered for structural interpretation to identify possible lipid species. The lipid class was confirmed by the presence of

head-group fragments or their neutral loss (for example, an MS/MS fragment with $m/z = 184.0735$ corresponds to the phosphocholine head-group). Because 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 to rule out the possibility of the sum formula corresponding to a lipid from any class other than the one annotated by our framework. To confirm the fatty acid side chains, the negative LC-MS methods were used (for example, 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 done with the help of the HighChem MassFrontier software (Thermo Fisher Scientific).

Data availability statement. Code is available at https://github.com/alexandrovteam/pySM.

22. Vese, L.A. & Chan, T.F. *Int. J. Comput. Vis.* **50**, 271 (2002).
23. Edelsbrunner, H. & Harer, J. *Contemp. Math.* **453**, 257–282 (2008).
24. Glasbey, C.A. & Horgan, G.W. *Image Analysis for the Biological Sciences* (Wiley Chichester, 1995).
25. Carter, C.L., McLeod, C.W. & Bunch, J. *J. Am. Soc. Mass Spectrom.* **22**, 1991–1998 (2011).
26. Elias, J.E. & Gygi, S.P. *Nat. Methods* **4**, 207–214 (2007).
27. Berglund, M. & Wieser, M.E. *Pure Appl. Chem.* **83**, 397–410 (2011).
28. Storey, J.D. & Tibshirani, R. *Proc. Natl. Acad. Sci. USA* **100**, 9440–9445 (2003).
29. Choi, H., Fermin, D. & Nesvizhskii, A.I. *Mol. Cell. Proteomics* **7**, 2373–2385 (2008).
30. Zhang, J. *et al. Mol. Cell. Proteomics* **11**, M111.010587 (2012).
31. Bligh, E.G. & Dyer, W.J. *Can. J. Biochem. Physiol.* **37**, 911–917 (1959).

doi:10.1038/nmeth.4072