Mass-spectrometry imaging-based explainable machine learning can be used to reveal biochemical landscapes of the brain

Jacob Gildenblat[1], Jorunn Stamnæs[2], Jens Pahnke[1,3,4,5,6,7*]

[1*]Pahnke Lab, Oslo, Norway, www.pahnkelab.eu.
[2]Proteomics Core Facility, Department of Immunology, Department of Clinical Medicine (KlinMed), Medical Faculty, University of Oslo (UiO) and Division of Laboratory Medicine (KLM), Oslo University Hospital (OUS), Sognsvannsveien 20, Oslo, NO-0372, Norway.
[3*]Translational Neurodegeneration Research and Neuropathology Lab, Department of Clinical Medicine, Medical Faculty, University of Oslo, Sognsvannsveien 20, Oslo, NO-0372, Norway.
[4*]Section of Neuropathology Research, Department of Pathology, Division of Laboratory Medicine, Oslo University Hospital, Sognsvannsveien 20, Oslo, NO-0372, Norway.
[5]Institute of Nutritional Medicine, University of Lübeck and University Medical Center Schleswig-Holstein, Ratzeburger Allee 160, Lübeck, D-23538, Germany.
[6]Department of Neuromedicine and Neuroscience, The Faculty of Medicine and Life Sciences, University of Latvia, Jelgavas iela 3, Riga, LV-1004, Latvia.
[7]Department of Neurobiology, School of Neurobiology, Biochemistry and Biophysics, The Georg S. Wise Faculty of Life Sciences, University of Tel Aviv, G. S. Wise Street, Ramat Aviv, IL-6997801, Israel.

*Corresponding author(s). E-mail(s): jens.pahnke@gmail.com;
Contributing authors: jacob.gildenblat@gmail.com;
jorunn.stamnas@medisin.uio.no;

Abstract

Recent computational advances in mass-spectrometry imaging (MSI) now enable unprecedented insight into organ-wide molecular composition and functional architecture.

Here, we present a first example of a high-resolution, molecular-computational atlas of specific mouse brain lipids - acquired using a NEDC matrix and negative mode MSI - covering 123 anatomically defined regions with 191 polygonal annotations derived solely from the MSI data — no auxiliary imaging required.

To overcome annotation ambiguity and MSI complexity, we introduced the Computational Brain Lipid Atlas (CBLA), a graph-based visual-explainability framework that generates *virtual landscape visualizations* (VLV) of specific lipids' distributions across the brain substructures. The CBLA integrates dimensionality reduction and ensembles of supervised models to (i) refine annotations, (ii) elucidate interregional relationships, (iii) interpret model behavior, and (iv) formulate biologically testable hypotheses.

The CBLA revealed novel lipid distribution patterns, functional integrations, anatomical connections - 'the brain's telephone cables', and region-specific disease signatures - disease networks in the basal ganglia.

A new algorithm decomposes annotated regions into precise m/z features and resolves full-precision m/z values from MSI data, producing a comprehensive high-resolution brain map. It can be used for any MS measurements: metabolites, lipids, and peptides. This resource underpins downstream studies, exemplified here by characterizing the lipid molecular composition of A$\boldsymbol{\beta}$ plaques, their spatial arrangement, and their connections with the surrounding tissue.

Keywords: mass spectrometry imaging, brain atlas, dimensionality reduction, machine learning, data visualization, explainable machine learning, lipidomics, explainability, Alzheimer's disease, MSI-VISUAL, MSI-ATLAS, ABCA7

Highlights

1. Mass Spectrometry Imaging (MSI) data were used to generate truthful visualizations for annotating brain regions with high resolution without the need for other modalities.
2. MSI data were used to generate a computational atlas representation of the annotated brain regions.
3. Pathological structures reveal their origin and their effects on specific brain networks.
4. Anatomical regions and functional networks show specific lipid patterns.
5. Brain stem nuclei and the white matter have a unique composition of lipids revealing their involvement in pathological networks.
6. The atlas' Virtual Landscape Visualizations (VLV) enable visualizing region-specific differences between different mouse models.

Introduction

Studying brain composition is fundamental to advancing our understanding of its structure, function, development, and diseases. Traditional approaches, such as histology, immunohistochemistry, and bulk biochemical analyses, have provided critical insights into neural processes, such as synaptic transmission, energy metabolism, and cell signaling, by probing the distribution of lipids, proteins, and metabolites. However, these methods often lack the spatial resolution or molecular specificity required to fully capture the intricate organization of the brain.

To tackle these problems and as an example, we focused on using Mass Spectrometry Imaging (MSI) to study the brain's lipid composition. We addressed the challenge of generating and visualizing a highly detailed molecular map of functional brain regions using MSI data alone. To achieve this, we employed recently developed computational methods for dimensionality reduction and visualization of MSI images [1, 2]. MSI data are inherently high-dimensional, making visualization and annotation complicated. Supervised machine learning (ML) methods require region-level annotations; however, generating these annotations is challenging without a clear visual delineation of the subregions. This limitation has led to a scarcity of prior studies using supervised annotations *directly* on MSI data. For instance, Huang et al. [3] performed binary annotations (hippocampus and cortex only) over PCA visualizations to compare the lipid distributions. Guo et al. [4] relied on automatic segmentation to identify regions, whereas Sun et al. [5] used hematoxylin-eosin (H&E)-stained tissue to generate annotations, which were later registered to MSI data. More recently, Xie et al. [6] overlaid MRI-defined annotations onto MSI to create a molecular map of 11 region types. In the MSI field, it is often suggested to use immunohistochemical stains (IHC) as overlays for the MSI data [7] or even MALDI IHC [8]. The actual problem here lies in the poor presentation of the MSI data as ion images and the actual lack of morphological expertise. Neuropathologists and pathologists are able to retrieve tissue structure information from panels of truthful visualizations [2] solely based on information-rich MSI data - no additional modalities are needed.

Moreover, external modalities such as MRI or H&E/IHC are often just unavailable and limit annotation granularity to *only* features visible in those modalities! Many brain substructures may not be distinguishable using pathology stains, such as H&E, which is fully blind to lipids, or MRI alone. While Schede et al. [9] recently demonstrated the feasibility of MSI for profiling approximately 100 hypothetical lipids through lipid database mapping of m/z values in *Danio rerio* fish embryos, their analysis made only limited use of the spatial information. In contrast, our approach enables high-fidelity mapping of lipid distributions across a wide range of anatomically defined brain regions. Similar to Schede et al., we generated mappings of m/z values across regions; however, by focusing specifically on the imaging dimension of MSI and their truthful visualizations in a multi-visualization panel [1, 2, 10], we were able to resolve a far larger number of regions and, crucially, provide new tools for studying their relationships and compositional organization.

Here, we demonstrate that MSI contains rich, untapped information capable of supporting fine-grained expert-generated annotations and supervised machine learning.

We report the development of a customizable pipeline (e.g., for metabolites, lipids, and peptides), the MSI-ATLAS, for integrated annotation and analysis of deep MSI data, which is not reliant on external imaging modalities. To enable structural annotations, we have developed MSI-VISUAL [2], a continuously improved tool that combines multiple MSI visualization methods enabling detailed visualization of high-dimensional MSI data as richly colored images, also including established dimensionality reduction approaches mainly from the genomics field like UMAP [11]. This multi-vizualization approach reveals subtle anatomical features missed by conventional segmentation techniques, which can serve as the basis for human-guided region and subregion polygon annotations. Here, we combine the use of MSI-VISUAL for region annotation with large-scale mapping of m/z values associated with each brain region through two complementary computational methods. To resolve m/z features with high precision, we also employed a custom Gaussian Mixture Model–based approach.

A central challenge of our example is the visual exploration and interpretation of the lipidomic relationships between brain regions. Given the complexity of brain organization and the inherently complex nature of MSI data, regional annotations or categories can be partially overlapping or potentially mislabeled. To enable the evaluation of inter-region molecular relationships, we developed novel visualization tools to interrogate and explore the distribution of features (m/z values) and their relationships within and between annotated regions or categories. These tools also enable the diagnosis and interpretation of machine learning models trained on the data. To enable interactive exploration and interpretation of the lipid landscape, we developed the *Computational Brain Lipid Atlas* (CBLA), an interpretable graph where each annotated region or category is visualized as a node connected by edges that reflect biological or annotation-induced molecular similarity. The CBLA graph can be overlaid by *Virtual Landscape Visualizations* (VLV), which represent the responses of different annotated regions to specific m/z values. These tools allow users to query m/z values and interactively examine their distribution across gross anatomical regions and small subregions, enabling the formulation of biological hypotheses based on MSI data.

We use the MSI-ATLAS to demonstrate as an application example the rich lipid complexity of the brain, annotating 191 polygon annotations and mapping them to 123 brain region types according to the Allen mouse brain atlas [12]. To our knowledge, this represents the first comprehensive molecular mapping of brain substructures using human-labeled MSI data at this level of detail. We observed that anatomically and functionally related regions share characteristic profiles of their lipid composition. Brain connectivity appears mirrored in the molecular structure: functionally connected regions, e.g., the network of extrapyramidal brainstem nuclei, exhibit shared lipid signatures, akin to a colored wiring diagram. Finally, we demonstrate the power of this new pipeline by analyzing Aβ plaques and, for the first time, quantify their molecular relationship with surrounding brain regions, demonstrating that Aβ plaques incorporate specific lipid signatures originating from affected cerebral cortex regions.

Materials and Methods

MSI data generation and extraction

The lipidomics MSI mouse brain dataset. The mouse brain lipidomics MSI data files were generated using frozen hemispheres of four 100-day-old animals in the C57Bl/6J background. Two mice expressed combined human mutated APP and PS1-mutated transgenes (APPtg) for the generation of Aβ plaques [13–15], two mice were ABCA7 knockouts [16]. Frozen mouse hemispheres were cut into 10 µm-thick coronar sections using a cryomicrotome (CM1950, Leica, Nussloch, Germany) and placed on a single Intellislide (Bruker Daltonics, Bremen, Germany). The section was dried in a vacuum chamber and sprayed with N-(1-Naphthyl) ethylenediamine dihydrochloride (NEDC) matrix dissolved to 7 mg/ml in 70% methanol at 35C using an HTX3+ TM Sprayer (HTX technologies, Carrboro, NC, USA) adopted as described by Andersen et al. [17]. Matrix-assisted laser desorption ionization trapped ion mobility mass spectrometry imaging (MALDI-TIMS-MSI) was performed in negative mode using a timsTOF fleX™ mass spectrometer (Bruker Daltonics, Bremen, Germany). Mass calibration was performed using red phosphorus and ion mobility using ESI-L Low Concentration Tuning Mix (Agilent Technologies, Santa Clara, CA, USA) in ESI mode. For MALDI-TIMS-MSI analysis, a mass range of m/z 300–1350 Da and a range of the inverse reduced mobility (1/K0) of 0.7–1.8 Vs/cm^2 was used. The laser was operated at the frequency of 10 kHz with a burst of 200 shots per pixel, with a spatial resolution of $20 \times 20\mu m^2$. We used the machine standard of $20 \times 20\mu m^2$ (4×4 pixels of $5 \times 5\mu m^2$) to generate sufficient signal (AU) also for very low abundant species. $5 \times 5\mu m^2$ generates higher spatial resolution for the cost of the loss of low signals. Four mouse hemispheres were assessed in a single overnight run on a single slide.

Data extraction from MSI files. The MSI-VISUAL toolbox [2] was used to extract the data at the measured m/z range of 300–1350 Da with a fixed bin resolution of 5 and 20 bins per m/z (resolution of 0.2 and 0.05 Da, respectively). For each image pixel, the detected m/z values were quantized to their closest bin, and their intensities were summed into this bin. The 5-bin extracted files were used for visualizations, and the 20-bin extraction files were used for initial statistical comparisons before high-precision resolution with Gaussian Mixture Models (GMM). Determination of full-precision m/z values was performed using the *raw* MSI files (see next paragraph). For each extracted 5-bin image file (n = 4), we created a multi-visualization panel using MSI-VISUAL [2]. Multiple visualizations were generated for each hemisphere, using the Saliency Optimization method (SALO), the TOP3 and Percentile Ratio (PR3D) methods, and dimensionality reduction methods, such as UMAP [11], PCA [18], and Non-Negative-Matrix Factorization [19]. For the following annotation of brain regions, an overlay of the multi-visualization was generated of SALO, TOP3, and PR3D visualizations, which reveal greater details inside the different brain regions [2], and a combination of segmentation methods (e.g., Kmeans [20]) with specific visualizations within the segments. All visualizations were created with total ion current (TIC) normalization [21].

Resolving full precision m/z values. The brain region m/z values are initially determined in a 0.05 Da resolution, created from a fixed 20-bin representation of the MS images. The goal was to obtain a higher precision measurement of the m/z values. We used a four-stage workflow:

1. We re-extracted the original raw MSI spectra, focusing on the narrowed-down m/z values, recording intensities within a range of up to 0.01 m/z from every m/z value. In this stage, we did not obtain an image but rather recorded all m/z values within this range from each detected m/z value. The goal here was to capture the m/z distribution.
2. For each m/z value, we fitted a Gaussian Mixture Model (GMM) [22] using scikit-learn [23] with three components. We discarded minor components that receive less than 5% of the data. The component centers represent the distributions of m/z values.
3. We then re-extracted MSI images using these components and summed m/z values within a range of 0.01 to them.
4. At the end, we recomputed the m/z statistics to re-verify the full precision m/z values.

Annotating

MSI-based annotations of brain regions and subregions. Polygon annotations were performed with the *VGG image annotator* [24] by a specialist in neuropathology (J.P.) according to the Allen mouse brain atlas (AMBA) [12], carefully comparing each annotated region across different visualizations in the multi-visualization panel (Fig. 1. There were four iterative rounds of annotations, during which the polygon annotations were refined based on the assessment with the generated Computational Lipid Brain Atlas (CBLA). Additional sub-annotations were added to the AMBA regions when detectable in the visualizations and were labeled with 'sub' followed by an integer number. Each annotation was categorized according to the AMBA nomenclature into main anatomical categories (gross brain regions), such as cortex (CTX), cerebral nuclei (CNU), brain stem (BST), white matter tracts (WM), choroid plexus (PLX), or pathological changes, e.g., amyloid-β plaque category (PLQ) and a background (BG) category. These main categories were used to color-code the resulting data (e.g., nodes) in the CBLA diagrams.

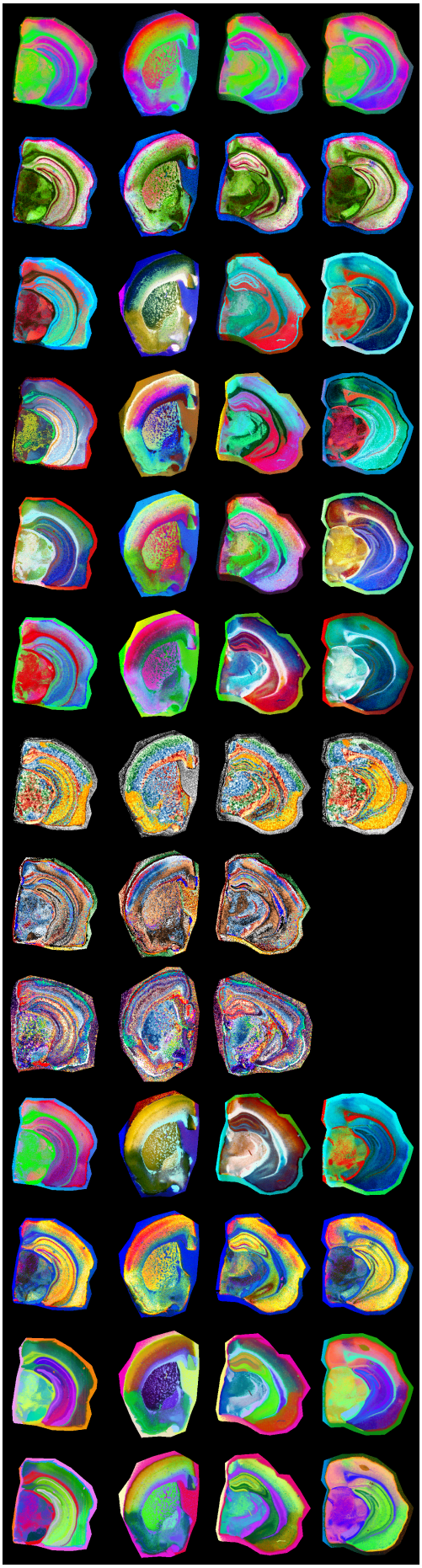

Fig. 1 Overview of a multi-visualization panel for annotating brain regions. We used custom generated panels of different locations respective to bregma to enable high-resolution annotations of brain regions as annotated in the Allen mouse brain atlas.

Resolving annotation problems using the CBLA. In each annotation iteration, we searched for nodes that were very close to each other or overlapping, and connected them with a strong edge. These nodes were further investigated for their m/z profiles to determine unique m/z that could verify the edge presentation and the node locations in the CBLA.

Modeling

Machine learning-based classification of pixel categories. For feature-based classification of pixels, we employed a simplified Bayesian linear model for classification, implemented with Pyro [25] using logistic regression. Each feature weight was assigned a *Gaussian prior* to encode uncertainty over its contribution to the predictive signal. During inference, we approximated the posterior distribution over the weights using variational inference. To generate predictions, we sampled weights from the learned posterior distribution and averaged the resulting predictions, enabling a probabilistic interpretation of model confidence. Since the model is linear, we can use the model weights for interpretability and identifying significant m/z values per region, as described in the next sections.

Ensemble-based evaluation of category confusion. To quantify the worst-case ambiguity between annotated categories, we evaluated category confusion using an ensemble-based strategy. We trained an ensemble of 10 independently initialized models with the Bayesian architecture described above. For each trained model, we estimated the soft confusion matrix by averaging predicted probabilities over 100 stochastic forward passes (sampling from the variational posterior). Each entry in the confusion matrix (i, j) represents the average probability that a sample from the true category i is assigned to the category j. To assess the worst-case confusion across the ensemble, we computed the maximum confusion per pair of categories:

$$\text{Confusion}_{i,j}^{\max} = \max_{m=1,\ldots,10} \text{Confusion}_{i,j}^{(m)}$$

This approach highlights the strongest disagreement between categories observed across model variations and posterior samples, serving as a conservative estimate of label ambiguity and potential annotation or biological overlap.

Mapping m/z values for brain regions. To identify m/z values associated with specific categories (brain regions or subregions), we employed a combination of two approaches with different behaviors: i) a model-based approach that trains an ensemble of models and measures feature importance for the m/z values, and ii) a model-free approach comparing pairs of brain region categories to mine for m/z values that are significant in those categories, denoted as *mPAUC*.

Both approaches have different selection behaviours. mPAUC tends to select m/z values that are less associated with other categories, while the model-based selection can also select m/z values that are shared with other brain region categories, as long as it is still predictive. Fig. 2 shows the effect of the selection methods for the plaques (PLQ) category. We measure the log_2 fold-change for the PLQ annotations from the rest of the non-background annotations. Notice that m/z values that are associated

with the PLQ may have a negative log fold-change, for example, because they are highly abundant in the various cortex categories. These m/z values may still be associated with the plaques, and we would want to select them. The intersection of both selection methods results in m/z values that are highly selective for the PLQ category.

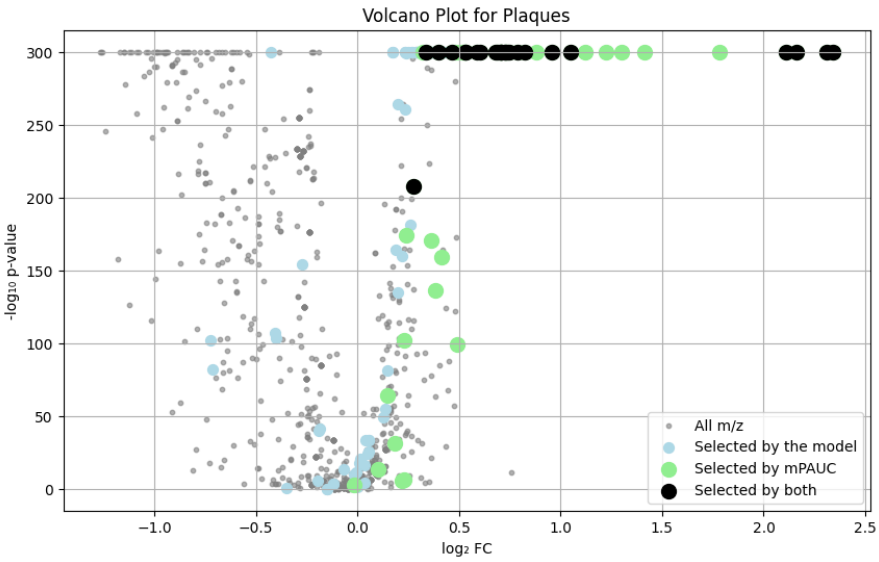

Fig. 2 Different m/z selection strategies applied on the Aβ plaque (PLQ) category. The model-based approach also selects m/zs that are shared with other categories, and therefore have a negative log_2 fold-change. The mPAUC method is selective for m/z values that are specific for PLQ due to its nature in pairwise comparing the values. The intersection of both methods results in highly selective masses for the PLQ, while their union may help gather all potential m/z values associated with PLQ.

Model-based selection of m/z markers for each brain region. To identify category-specific m/z features and interpret the weights of our linear model, we developed a custom feature importance scheme inspired by concepts from *boruta* [26] and *STABL* [27]. *Boruta* is a wrapper-based feature selection method designed for tree-based models, which compares the importance of real features to permuted noise features using random forests. *STABL*, by contrast, is built on sparse linear models and stability selection, combining resampling with noise-informed thresholds to control false discoveries. While both methods share the use of permuted features and repeated subsampling, they are structurally distinct. Given the computational demands of *STABL* and the tree-based nature of *boruta*, we implemented a lightweight alternative tailored to our linear setting. Specifically, we trained 200 logistic regression models, each on a random 50% subsample of the data. To calibrate feature selection, we introduced N=50 noise features by randomly selecting and permuting existing m/z values. For each model and category, we computed the maximum absolute weight assigned to any noise feature and selected real m/z features whose positive weights exceeded this

threshold. Finally, following the stability selection principle, we retained only those features that were selected in at least 80% of the models.

m/z-Pairwise-AUC (mPAUC) – ranking m/z values per regions by comparing region pairs. This approach assigns a category score for every m/z value based on AUC ranking comparisons. For every m/z value, we score each category by comparing its separation from other categories with the AUC score by summing AUC values with a low p-value. This score shows that the m/z value distinguishes this category from other categories. Then, we fetch the top 10 categories that score highest for the m/z value. During this comparison, categories that are not relevant will also participate and may skew the result. We repeat this to promote competition between relevant categories, considering only the top 20 categories for each m/z. In other words, we wanted to identify categories that scored high for each m/z value and check how they compared without irrelevant categories from the first round. For each m/z value m and each category i, we computed a category score by accumulating the AUC scores between that category and other categories that have a p-value below 0.05:

$$S_{im} = \sum_{i=1}^{N} AUC_{ijm} \quad \text{where} \quad pval_{ijm} < 0.05$$

For each m/z value m, we retrieve the categories with the highest k=20 scores and recompute S_{im} with this set of categories.

Construction of the Computational Brain Lipid Atlas

The Computational Brain Lipid Atlas (CBLA) is a graph-based framework designed to enhance the interpretability of MSI data, annotations, and trained models (Fig. 4). It provides a unified visual representation that supports the exploration of biological structure, annotation reliability, and model behavior.

Defining molecular relationships between brain regions. To understand the molecular relationships between the annotated brain regions, we first computed the average m/z intensity profile for each region. These region-level molecular profiles are embedded into two dimensions using UMAP [11]. Each region was visualized as a node positioned according to its 2D embedding. Edges between nodes are weighted using the maximum confusion values from an ensemble of 10 Bayesian classifiers, with each model's normalized confusion matrix averaged over 100 posterior samples by taking the mean + $2 \cdot std$ of the confusion. The maximum pairwise confusion between models was used to define the edge weight between nodes (brain regions). Edges between nodes within the same anatomical group (e.g., cortex layers or white matter tracts) were colored red, suggesting likely sources of annotation ambiguity or shared signal, such as biological or functional similarities. Edges between distinct anatomical systems were colored bright green, indicating potential biological links or molecular similarities warranting hypothesis generation.

Explaining and refining the annotations. CBLA also provides a lens for evaluating annotation quality. Edges with high confusion between regions expected to

be distinct (e.g., between functionally unrelated areas) suggest potential annotation errors, class imbalances, or poorly defined categories. Conversely, high similarity between nominally different but biologically similar subregions may suggest duplicated or unnecessary labels. This allows researchers to iteratively refine annotations and resolve inconsistencies in the dataset structure.

Explaining the model. To visualize the model's internal understanding of brain regions, we constructed a parallel version of CBLA using the learned prediction weights for each category instead of average m/z profiles. These region-level weight vectors represent the molecular features the model relies on for classification. Dimensionality reduction and edge construction followed the same process as for the data-based atlas (Fig. 5). By aligning the structure of the data, annotation signals, and model decisions, CBLA enables transparent inspection and quality assessment of the MSI-based brain map, which supports both error diagnosis and biological discovery.

Visualisations

Colorizing the atlas - Virtual Landscape Visualizations. For a selected m/z value, an intensity value is obtained for every category node. To explain the data, the value stands for the average intensity of that category for the m/z. To explain the model, the intensity stands for the model category weights for that category. For every pixel, we measure the coordinate Euclidean distances d_i from all the nodes in the graph (i.e., the node 2D coordinates after dimensionality reduction), and take a weighted average of the node representations with a softmax [28]:

$v_i = \frac{e^{T \cdot d_i}}{\sum_j e^{T \cdot d_j}}$

The relationship of the selected m/z value with the CBLA network is visualized as a topographic overlay color scheme coded to visualize the 0-1 normalized total ion current of the m/z

Visualizing mouse model and region-specific differences with the CBLA. We grouped the mice into two cohorts (APPtg+ vs. APPtg– (wild-type), or A7+ (wild-type) vs. A7– (ABCA7 knockout) and computed the average value within each group for every annotated brain category. However, not all brain categories are annotated in every scan, which may prevent the computation of group averages for those categories. One way to address this problem would be to require annotations for all brain categories in each scan. Instead, we adopted a different approach: predicting the brain categories using machine learning-based segmentation. Specifically, we use an ensemble of 100 linear models and assign categories based on the majority vote across the ensemble.

Virtual Pathology Stain (VPS). To simulate a commonly used immunohistochemical (IHC) stain, the grayscale ion intensity images were mapped to a brown color scale. Total ion current (TIC) intensities of fully resolved m/zs were normalized for each m/z to 0-1 for all pixels, and a linear interpolation was performed for each pixel between white (RGB: 255, 255, 255) and dark brown (RGB: 139, 69, 19). This mapping ensures that low-intensity pixels remain light, while high-intensity regions are rendered

in progressively deeper shades of brown, visually resembling 3,3'-diaminobenzidine (DAB) staining used in histopathology (e.g., Fig. 7c).

Results

A computational workflow for highly detailed exploration and mapping with MSI

A schematic depiction of the MSI-ATLAS workflow is shown in Fig. 3. MSI-ATLAS was established to generate functional and anatomical representations of organ-specific multiomics networks. While demonstrated here using brain lipidomics MSI data, the workflow is **generalizable to other omics modalities**, such as metabolomics and proteomics.

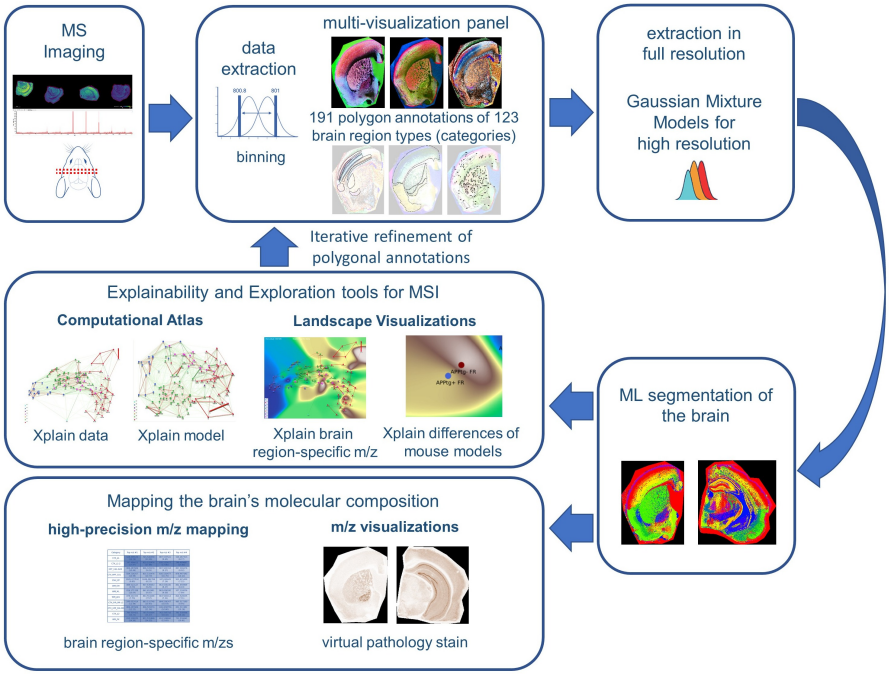

Fig. 3 Overview of the MSI-ATLAS workflow. The workflow of *MSI-ATLAS* involves a comprehensive analysis of mouse brain tissue utilizing Mass Spectrometry Imaging (MSI). The resulting high-dimensional datasets were converted to fixed-bin representations and visualized through advanced visualization methods provided by *MSI-VISUAL* [2]. These reliable visualizations enabled 191 highly detailed polygon annotations based on the Allen mouse brain atlas nomenclature [12], as well as identifying newly detected subregions. The resulting dataset comprised 123 brain region types (categories) along with an additional background category, all of which were mapped against the binned m/z values. Following this mapping, a full-resolution extraction from the raw mass spectrometric data was performed, and Gaussian Mixture Models were employed to derive high-resolution m/z values. Using the full resolution data, we create the Computational Brain Lipidomics Atlas (CBLA), where the spatial arrangement of the nodes is determined using UMAP for dimensionality reduction of the average category spectra (brain region type) to two dimensions. This computational atlas was subsequently overlaid with a *Virtual Landscape Visualization* (VLV) to elucidate the distribution of selected m/z values across various brain regions, or specifically tailored to mouse models. We utilize ensembles of Bayesian machine learning models, which we sample from, to compute confusion between categories that correspond to the graph edges. We also utilize the models to map stable m/z values associated with each brain region type. The mapping was visualized for verification using a *Virtual Pathology Staining* (VPS) technique designed to mimic traditional immunostained slides. The CBLA and VLVs were used to iteratively refine the annotations.

To establish and evaluate MSI-ATLAS workflow, we generated and analyzed an example of a MSI lipidomic dataset of four full brain hemispheres from control and APP-transgenic mice representing aged, healthy, and diseased brains, respectively. The dataset was comprised of 20020 features (m/z values) across four spectra with a pixel resolution of $20 \times 20\mu m^2$.

We enabled an expert neuropathologist (J.P.) to create region and subregion annotations directly on MSI data by leveraging a multi-visualization panel of complementary visualizations (Fig. 1), each optimized for highlighting different molecular

patterns [2, 10]. Using these polygonal annotations, we mapped the molecular composition of each annotation and created a database of m/z values of 123 brain region types. We employed a novel method based on ranking specific m/z values between categories, m/z-Pairwise-AUC (mPAUC), and a machine learning (ML) model-based approach that used feature selection of stable features to associate m/z values with brain regions. The resulting region-specific m/z associations form a molecular atlas that may serve as a resource for future studies.

To effectively explore and interpret the lipid landscape of the brain, we developed an *explainability and visualization tool*: the *Computational Brain Lipid Atlas* (CBLA). With 191 annotations for 123 brain region types, we constructed high-dimensional representations in a graph for each brain region and embedded them as **nodes**. In this graph, **edges** reflect molecular similarity, whether biological or annotation-induced, based on the ability of ensembles of ML models to distinguish between the brain regions. Background colors, the *virtual landscape visualization*, represent the responses of different categories to specific m/z values. CBLA allows users to query m/z values and interactively examine their distribution across gross anatomical regions and small subregions. Using CBLA, we observed that anatomically and functionally related regions share characteristic profiles of their lipid composition. Brain connectivity appears mirrored in the molecular structure: functionally connected brain regions exhibit shared lipid signatures, akin to a colored wiring diagram. The CBLA also supports iterative annotation refinement by identifying regions that may benefit from subdivision into smaller subregions. This feedback loop helped us improve the quality and granularity of our initial annotations, as well as detect previously unknown subnuclei, such as those in the Substantia nigra and the Globus pallidus.

To address the practical challenge of fixed-bin m/z representations used in MSI visualizations, we developed a workflow to recover high-precision m/z values. Fixed binning can obscure distinct molecular species, especially at high resolutions. We modeled the m/z distributions of significant m/zs with Gaussian Mixture Models [22], allowing us to identify and refine high-precision molecular features. Finally, the workflow includes tools for spatial m/z visualization by Virtual Pathological Stains (VPS) that mimic the commonly used DAB staining for immunohistochemistry, and a novel approach to extract high-accuracy m/z values from annotated regions of interest (Fig. 3, Methods and Suppl. Data).

We demonstrate how the CBLA can be used to gain several biological insights. We employed the CBLA to study the spatial distribution coding of lipids in Aβ plaques and relationships with other brain regions, and how it can be used to visualize region-specific differences in mouse models. We also identify and visualize distinct molecular signatures of ABCA7 knockout mice [16] demonstrating the potential of our approach for studying neurodegenerative disease models [29].

Construction of the Computational Brain Lipid Atlas

The *Computational Brain Lipid Atlas* (CBLA) is a visual explainability tool built for iterative refinement and exploration (Fig. 4). CBLA was generated after each round of annotation, helping to guide a total of three refinement iterations. As described in detail in the Methods section, the CBLA aggregates MSI data by brain region, applies

dimensionality reduction to reveal global organization, and constructs a region-to-region graph based on confusion scores derived from machine learning classification models. Edge strengths in the graph reflect model uncertainty or misclassification rates, offering insight into potentially overlapping biology or ill-defined regions. Larger nodes represent a higher confusion with other categories.

In addition, we introduce a *Virtual Landscape Visualization* (VLV) of the data (Fig. 4d) and of the model (Fig. 5). The VLV enables spatial interpretation of responses to individual or grouped m/z values and evaluation of model certainty. The virtual landscape is colored by the predicted m/z response intensity of the different categories and interpolated to form a smooth, easy-to-interpretable visual map.

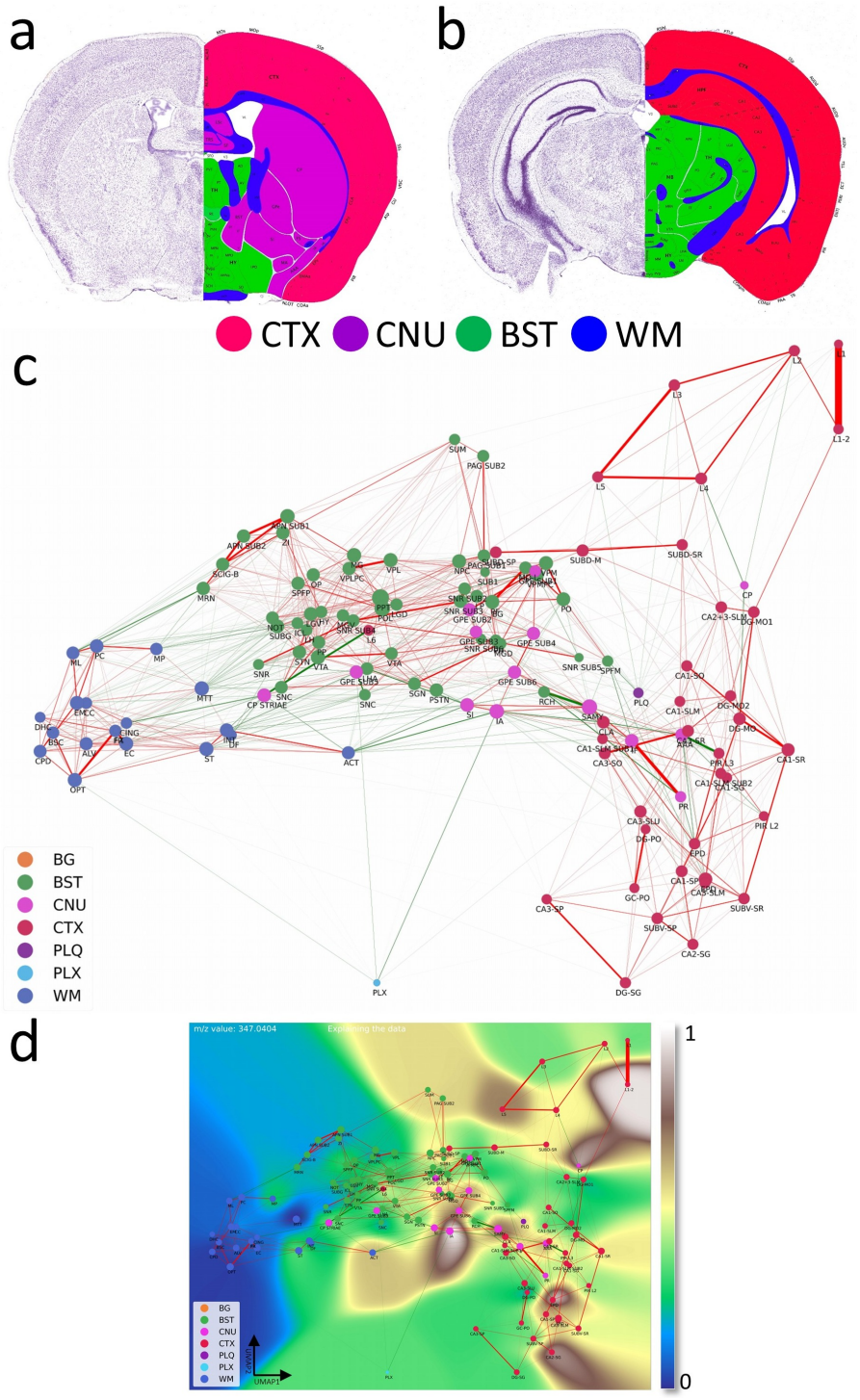
a
b
CTX
CNU
BST
WM
c
BG
BST
CNU
CTX
PLQ
PLX
WM
d
m/z value: 347.0404
Explaining the data
UMAP2
UMAP1
1
0

Fig. 4 Cerebral Brain Lipid Atlas (CBLA) and Virtual Landscape Visualizations (VLV) - an enhanced analytical framework for molecular profiling. a, b, show coronary sections at two different fronto-occipital locations with colorisation of the gross anatomical regions CTX (red), CNU (violet), BST (green), and WM (blue). This color scheme was used to color-code the nodes in the CBLA graph. The left side of the slides represents a Nissl stain of the neuron distribution. The images were adopted from the Allen mouse brain atlas [12]). **c**, The CBLA provides visualizations and analyses of lipid species within the brain (dataset PXD056609), employing Virtual Landscape Visualizations (VLV) to elucidate the underlying data structures and ML models. The graph nodes are constructed by averaging m/z spectra across distinct brain region categories of all measured mice, followed by applying dimensionality reduction techniques utilizing UMAP [11]. The spatial proximity between these nodes reflects their molecular similarity, facilitating comparative analysis. The size of the nodes corresponds to the number of connected edges (confusion). Edges in the graph denote relationships between the identified categories, generated through the training of various ML models for classification purposes. These edges are informed by an analysis of the confusion matrix, specifically capturing the instances of poor classification performance. Edges colored green indicate connections across different gross anatomical brain regions, suggesting potential biological associations. In contrast, red edges linking categories within the same gross anatomical region, e.g., the cortex layers CTX_L2-L5, may denote possible annotation discrepancies or inherent biological similarities. The node sizes are proportional to the overall confusion they have with other categories. **d**, The VLV is employed to elucidate the dataset's response to the specific m/z value of 347.04044. Here, nodes represent average spectral values for each category, effectively capturing the essence of the data (dataset PXD056609). The response of each category to this m/z is visually represented using the 'terrain' color scheme from Matplotlib [30], allowing for enhanced interpretation of molecular distributions across brain regions. The 0-1 normalized expression of the m/z value 347.04044 can be localized to distinct regions, including the cerebral cortex (CTX; red regions corresponding to layers L1/2, CA2-SG, and EPD) and cerebral nuclei (CNU; pink regions representing the caudate-putamen [CP], external globus pallidus [GPE], and intercalated amygdaloid [IA] nuclei). Legend: BG, background; BST, brain stem; CNU, cerebral nuclei; CTX, cerebral cortex; PLQ, Aβ plaques; PLX, choroid plexus; WM, white matter.

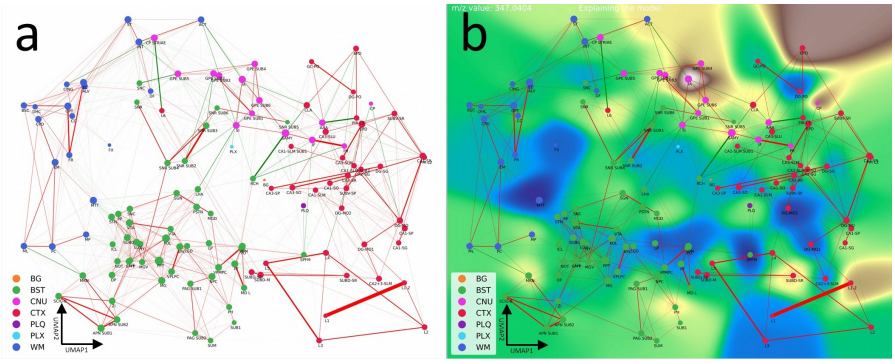

Fig. 5 Explaining the model. a, Graph of the CBLA used for explaining the model weights, where we used a Bayesian logistic regression model for classifying the categories and using the category weights for each node before applying dimensionality reduction. **b**, VLV for explaining the model weights. Here, the nodes represent the learned model weights for each category. The model weight coefficients for the m/z 347.0404 are overlaid with the 'terrain' Matplotlib color scheme. Notice that, like for the data VLV (Fig. 6), the IA and EPD categories show a strong response, confirming the model has found this m/z predictive for these categories. However, unlike the data VLV, the model does not show a strong signal for the CA2-SG or the L1-2 categories, potentially showing that the model could be improved. Legend: BG, background; BST, brain stem; CNU, cerebral nuclei; CTX, cerebral cortex; PLQ, Aβ plaques; PLX, choroid plexus; WM, white matter.

The brain consists of anatomical regions with unique lipid signatures

First, we used the CBLA to understand the global organization of brain lipids within three gross anatomical regions: the cortex (CTX), the white matter (WM), and the brainstem (BST). Fig. 6 shows the color-coded organization of the 123 mapped brain categories and background as nodes in the CBLA. The nodes from each gross anatomical region cluster together at different graph locations, which agree with their spatial and functional separation in the brain and their resulting m/z signatures. Thus, the CBLA accurately condenses and visualizes functionally and anatomically distinct regions solely based on MSI data. Nodes that do not cluster with their gross anatomical region represent interesting categories with variant anatomical or functional behavior, e.g., the cortex layer 6 (CTX_L6), which is located within the central BST cluster, highlighting its actual molecular nature as the main cortex layer for all efferent and afferent network connections from/to the neuronal cortex layers. The Extended Data section provides figures and tables with information on the mapped m/z values and a UMAP diagram highlighting the value distribution of all the categories.

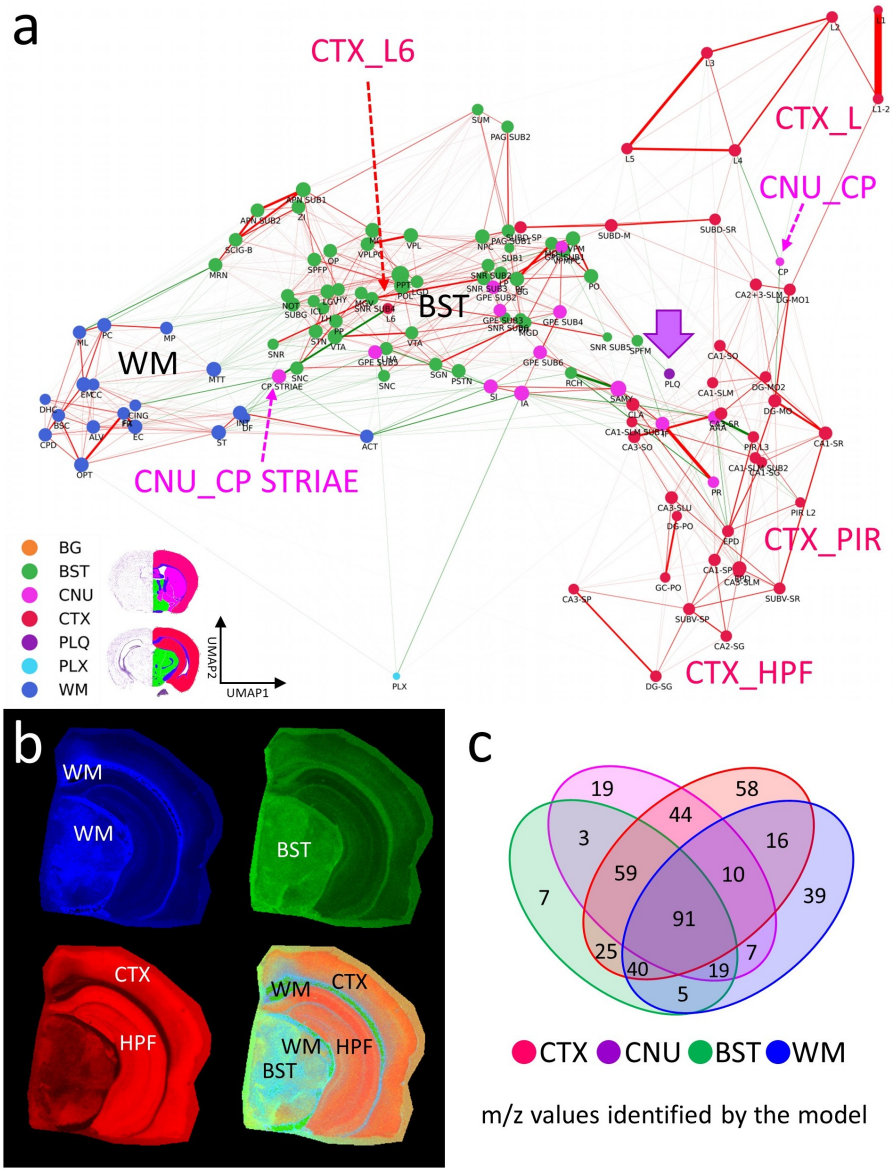
a
CTX_L6
CTX_L
CNU_CP
BST
WM
CNU_CP STRIAE
CTX_PIR
CTX_HPF
BG
BST
CNU
CTX
PLQ
PLX
WM
UMAP1
UMAP2
b
WM
WM
BST
CTX
HPF
WM
CTX
WM
HPF
BST
c
19
58
44
3
16
59
10
7
39
91
25
7
40
19
5
CTX
CNU
BST
WM
m/z values identified by the model

Fig. 6 Brain regions have unique lipidomics profiles. a, The CBLA elucidates the clustered distribution of various major categories (color-coded) and clustered subcategories within the brain. Notably, the category corresponding to layer 6 of the isocortex (CTX_L6) is located amidst the brain stem (BST) structures, attributable to its mixed composition of myelinated tracts and neuronal processes. Furthermore, the caudate putamen striae (CNU_CP STRIAE) within the cerebral nuclei (CNU) category signify myelinated tracts that are positioned between the neurons of the caudate putamen (CNU_CP) in the tissue. These structures are visually represented in the CBLA as situated between the BST and white matter (WM). Also, the Aβ plaques category (violet arrow) is found among the cortex (CTX) subcategories (CTX_L, CTX_PIR, and CTX_HPF). **b**, The RGB images visualize three major brain regions by summing their mapped m/z values and removing shared values: blue, WM connections, encompassing the corpus callosum and brain stem WM; green, the BST nuclei and connections; red, CTX, specifically the isocortex layers (CTX_L) and the hippocampal formation (CTX_HPF). **c**, Venn diagram highlighting the m/z values identified for the gross anatomical regions determined with the ML models. Legend: BG, background; BST, brain stem; CNU, cerebral nuclei; CTX, cerebral cortex; PLQ, Aβ plaques; PLX, choroid plexus; WM, white matter.

Pathological structures reveal their location of origin and their effects on specific brain networks

Aβ plaques are small, dense, spherical deposits of Aβ protein aggregates. These toxic accumulations destroy nearby neurons and their connections in the brains of patients and mouse models with Alzheimer's disease. Using the CBLA, we explored the lipid composition and molecular relationship of these circumscribed pathological structures. During the annotation process, we generated >50 polygonal annotations grouped into the PLQ category, which comprises Aβ plaques. The PLQ category shares numerous m/z values originating from different regions, emphasizing that the lipid signatures of Aβ plaques arise from the originating brain regions (Fig. 7). Moreover, we detect m/z values that originate from bypassing neuronal connections between neurons and their dendritic and axonal networks that were destroyed (Fig. 8), demonstrating how we can perform unsupervised molecular phenotyping of pathological structures through the CBLA. Zooming in on individual plaques within the PLQ category, we can pinpoint their origin based on the molecular pattern and location in the CBLA (Fig. 9). Using the VLVs, we can identify and verify low-abundant m/z values aggregated in Aβ plaques, revealing the disrupted neuronal networks. Fig. 10 shows an example of such a newly discovered, low-abundant m/z value, which is also detected in the hippocampus formation (CTX_HPF), a projecting nucleus - supramamillary nucleus (SUM) -, and the choroid plexus (PLX). Interestingly, the distribution of the m/z values follows a molecular weight-dependent pattern (Fig. 11).

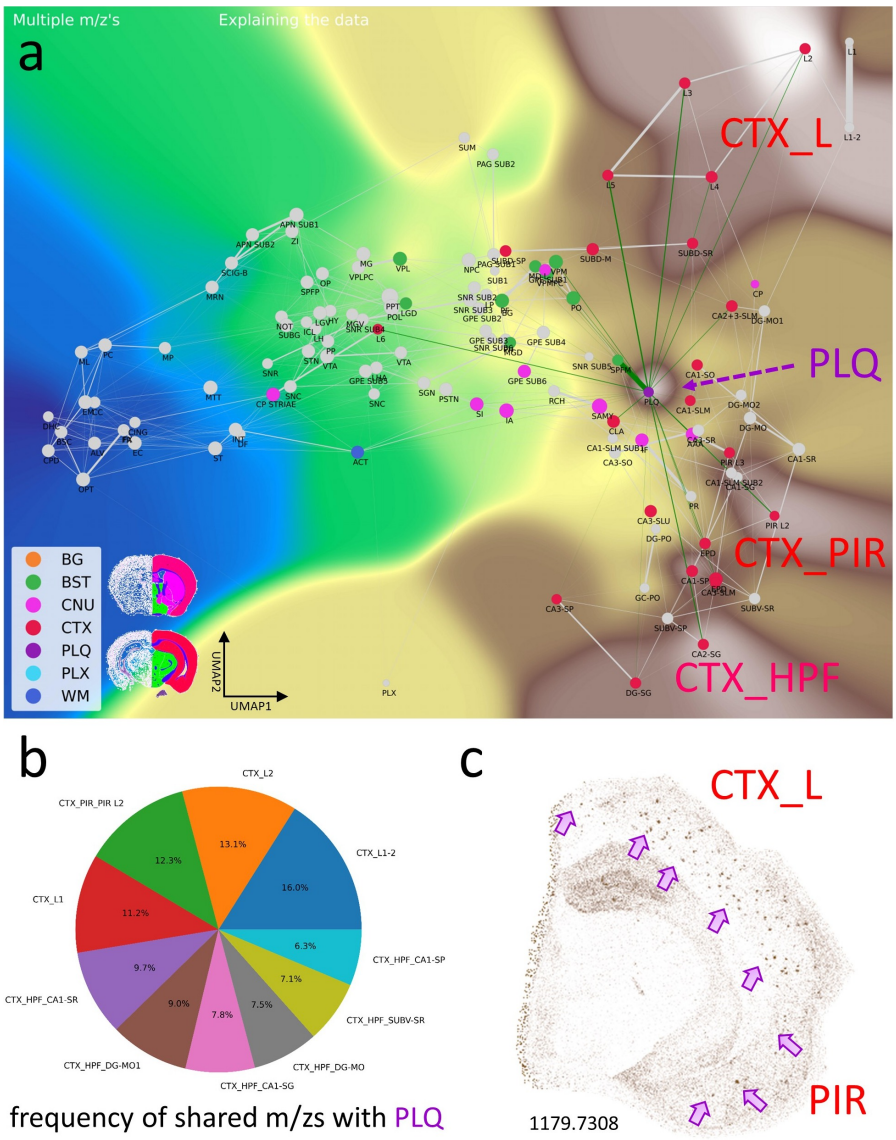

Fig. 7 Aβ plaque-associated m/z signatures reveal relationships with several brain regions. The PLQ category exhibits an interesting pattern of m/z values. The category as such has connections to various brain regions and is therefore located in the center of the cortex brain regions: i) it is composed of plaque-specific m/z values, and ii) shares m/z values from different locations, e.g. CTX_L (cortex layers), CTX_PIR (piriform cortex), and CTX_HPF (hippocampus formation). The PLQ node has edges to different brain regions - these are potential brain connections and indicate the shared m/z values. Of note is the edge towards the cortex layer 6 (CTX_L6). **a** ,*Virtual Landscape Visualization* (VLV) of m/z values significantly associated with the PLQ category. Prominent expression is observed in the isocortex layers 2–5 (CTX_L), piriform cortex (CTX_PIR), and hippocampal formation (CTX_HPF). **b**, Pie chart showing the top anatomical regions based on the frequency of shared m/z's with PLQ, predominantly within the isocortex (L) and hippocampal formation (HPF) of the CTX categories. **c**, a Virtual Pathology Stain (VPS) of m/z 1179.7308 reveals Aβ plaques in different isocortex layers (CTX_ L) and piriform cortex (PIR). Arrows indicate examples of individual plaques. Legend: BG, background; BST, brain stem; CNU, cerebral nuclei; CTX, cerebral cortex; PLQ, Aβ plaques; PLX, choroid plexus; WM, white matter; labeling of pie-chart was done with Allen atlas nomenclature.

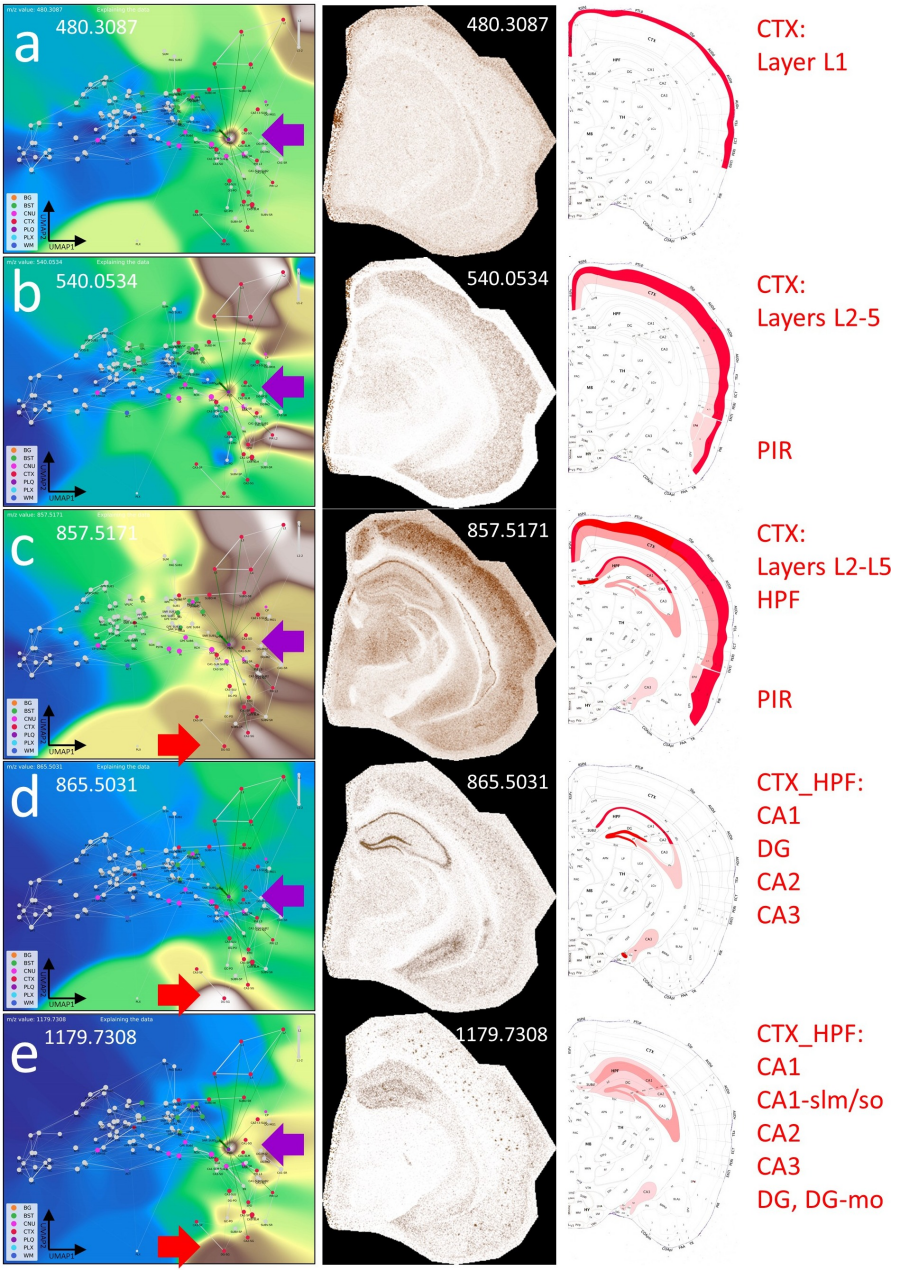

Fig. 8 The CBLA maps Aβ plaque-associated lipid accumulation and brain connectivity. a-e, show the VLV for different m/z values associated with the PLQ category (violet arrow), their Virtual Pathology Stain (VPS) highlighting the corresponding brain regions, and a schematic annotation overlayed on top of the Allen mouse brain atlas [12] explaining the affected regions. In the VLV, only nodes are highlighted that share PLQ m/z values. Aβ plaques disrupt neural connections in the cortex (CTX_L, CTX_PIR) and accumulate lipids from degenerated axons. These lipid accumulations indicate, e.g., the origin of disrupted connections and causes plaques to have shared m/z values with the originating regions, e.g., from the hippocampus (CTX_HPF, red arrow in VLV). Legend: BG, background; BST, brain stem; CNU, cerebral nuclei; CTX, cerebral cortex; PLQ, Aβ plaques; PLX, choroid plexus; WM, white matter.

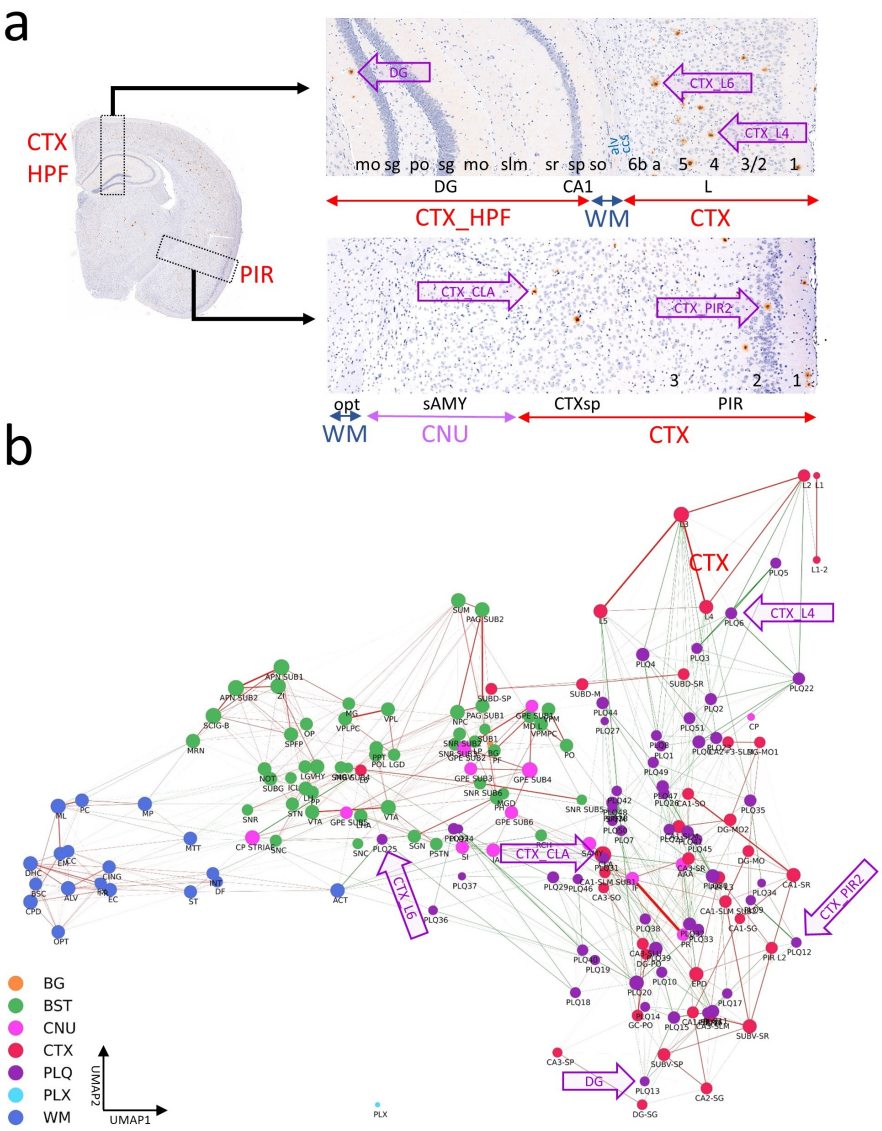

Fig. 9 The CBLA uncovers location-specific lipid signatures of Aβ plaques. a, shows a brain hemisphere immunostained against Aβ and hematoxylin as counterstain, and two magnifications of specific CTX regions marked (with boxes). The magnifications show the anatomical arrangement into specific CTX layers or HPF structures (red - CTX, blue WM, violet - CNU). The annotations, which were used to establish the CBLA, are shown as text (mo→1, and 3→1). The arrows label examples of five plaques in different CTX regions (DG, L4, L6, PIR2, and CLA). **b**, In the CBLA graph, the nodes of the labeled plaques (from **a**) can be found near nodes of their originating anatomical location, confirming their region-specific lipid content. Legend: BG, background; BST, brain stem; CNU, cerebral nuclei; CTX, cerebral cortex; PLQ, Aβ plaques; PLX, choroid plexus; WM, white matter.

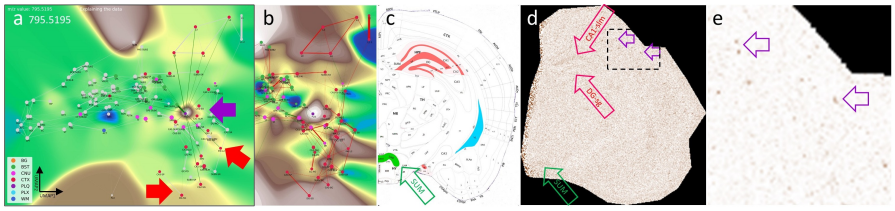

Fig. 10 Visualizing a low-abundant Aβ plaque-related mass and its spatial relationship with other brain regions. a, The VLV of m/z 795.5195 shows a strong signal in the PLQ (violet arrow) and BST_SUM categories, with lower intensity in PLX and CTX_HPF_DG-SG categories. **b**, histogram-normalized VLV to increase low-abundant signal visibility, **c**, Anatomical representation of m/z 795.5195 expression, color-coded by gross anatomical regions (see also Fig. 11). **d**, Linear VPS showing the m/z signal slightly above noise levels; anatomical regions are marked in correspondence with panel c. Examples of Aβ plaques are indicated with violet open arrows **e**, Magnification of d and logarithmic VPS enhancing high-intensity signals and suppressing noise. Aβ plaques are indicated with violet open arrows and become clearly visible. Legend: BG, background; BST, brain stem; CNU, cerebral nuclei; CTX, cerebral cortex; PLQ, Aβ plaques; PLX, choroid plexus; WM, white matter.

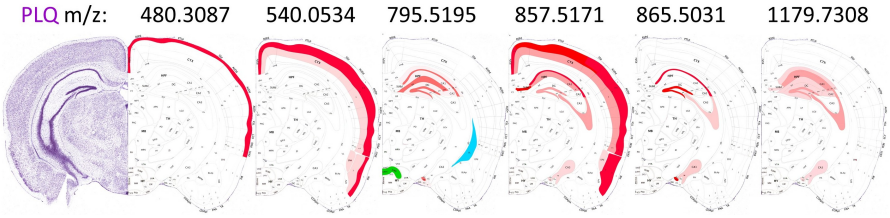

Fig. 11 Schematics of the location of six selected m/z values found in the PLQ category. The scheme highlights the connections that contribute to the lipid profile of Aβ plaques, mostly originating from the isocortex (CTX_L, CTX_PIR categories) and the hippocampus formation (CTX_HPF). The distribution changes with increasing m/z. The newly discovered m/z value 795.5195 is located in the supramamillary nucleus (BST_SUM), which projects to the CTX_HPF, and the choroid plexus (PLX). The Nissl stain and brain layout were modified from the Allen mouse brain atlas [12].

Generating new biological hypotheses with the CBLA

Our objective was to evaluate the CBLA framework in elucidating the distribution of lipids across functional brain networks. While mapping the Aβ plaque m/z values and visualizing them using the CBLA and VLVs (Figures 7-9), we observed a distribution pattern that correlated with the origin of the plaque location and its neuronal connections—such as from the hippocampus for m/z 1179.78 (Fig. 9e). This led us to hypothesize whether distinct lipid distribution patterns are associated specifically with functional-anatomical networks or individual neuronal connections within the brain. Therefore, we named these lipids: *index lipids*. Such patterns may reveal functional networks that could offer novel insights into the intricate lipid connectivity and biological interactions within the brain.

Anatomical regions and functional networks show specific lipid patterns

Nuclei and connections within the extrapyramidal motoric network exhibit characteristic index lipids. Utilizing the CBLA, we have identified lipid distribution patterns that correlate with specific functional networks. Fig. 12 illustrates two examples of m/z values that reveal a shared network pattern consisting of nuclei and cortical regions involved in the modulation of motor function. These lipid patterns could be used to investigate why neurodegenerative diseases have predilection sites at typical disease locations, how these diseases start, and the effects that promote the pathogenesis process.

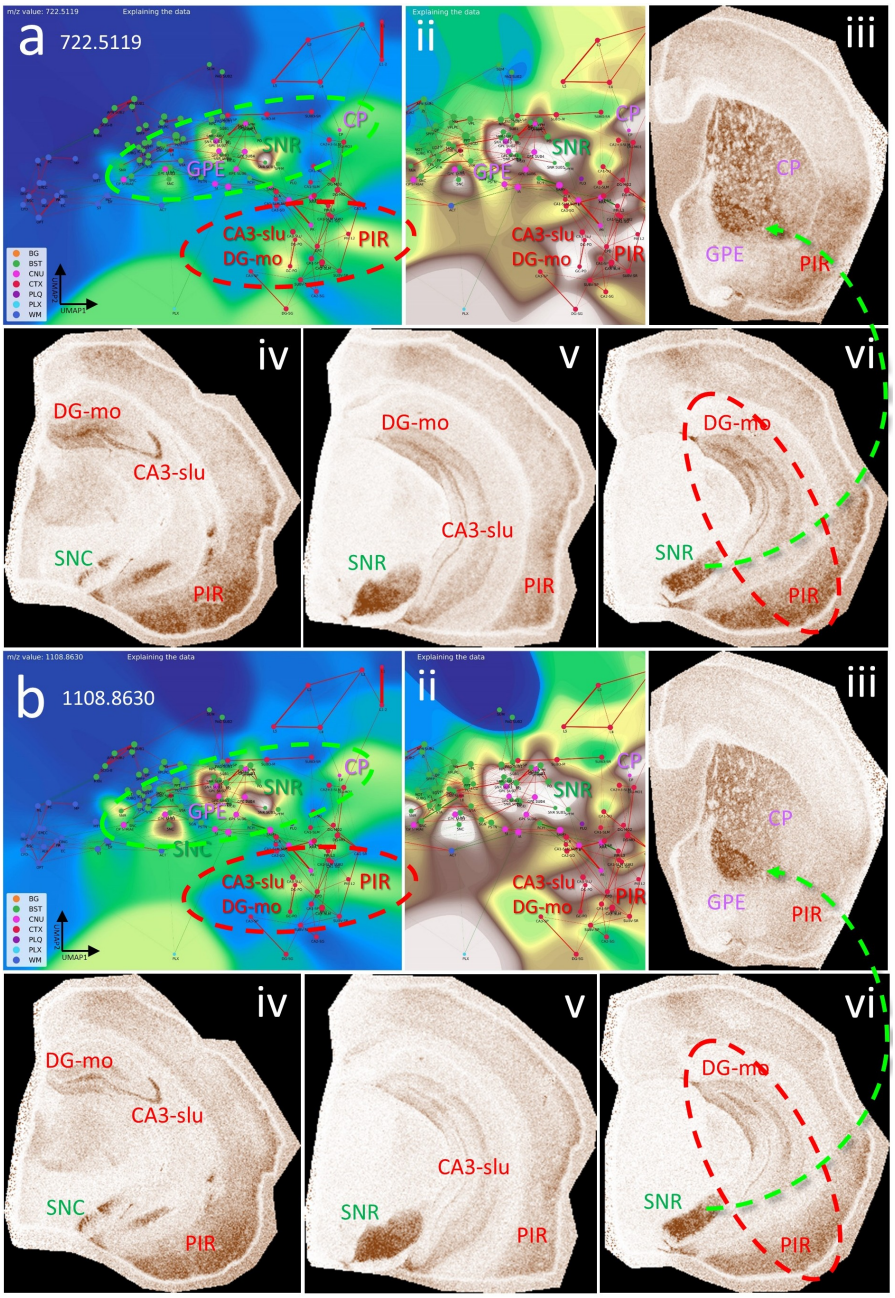

Fig. 12 Nodes and connections of the extrapyramidal network share specific index lipids. **a**, shows the CBLA of the m/z 722.5119 highlighting its distribution in BST nuclei (green ellipse, SNR, SNC, and GPE), cerebral nuclei (CP), and cortex structures (red ellipse, CA3, PIR). ii) shows a histogram-normalized landscape to increase visibility of the nodes. ii-vi show VPS of the expression of the m/z and its distribution pattern in SNC, SNR, DG-mo, CA3-slu, GPE, CP, and PIR. SNR/SNC are functionally connected to GPE/CP (iii and vi, green arrow), similar to PIR and HPF (DG, CA3) are connected (vi - red ellipse). **b**, m/z 1108.8630 shows a nearly identical expression and distribution pattern as **a**, except for a much higher expression in the GPE (iii) and SNC (iv). Legend: BG, background; BST, brain stem; CA3, cornu ammonis part 3; CNU, cerebral nuclei; CP, caudate putamen; CTX, cerebral cortex; DG, dentate gyrus; GPE, Globus pallidus, pars externa; HPF, hippocampus formation; PIR, piriform cortex; PLQ, Aβ plaques; PLX, choroid plexus; SNC, substantial nigra, pars compact; SNR, substantial nigra, pars reticulata; WM, white matter.

The hippocampal formation reveals organized lipidomic structures. The HPF consists of multiple functional-anatomical layers (see also Fig. 9a) which are connected to several cortical brain areas. These connections are affected during disease development and persistence, e.g., temporal lobe epilepsy. Assessment of specific metabolic and lipidomics markers that are linked to these connections can be analyzed and used to understand the generation of grand mal seizures originating locally in the HPF and their long-term effect on the organization of the hippocampal-cortical networks (Fig. 13).

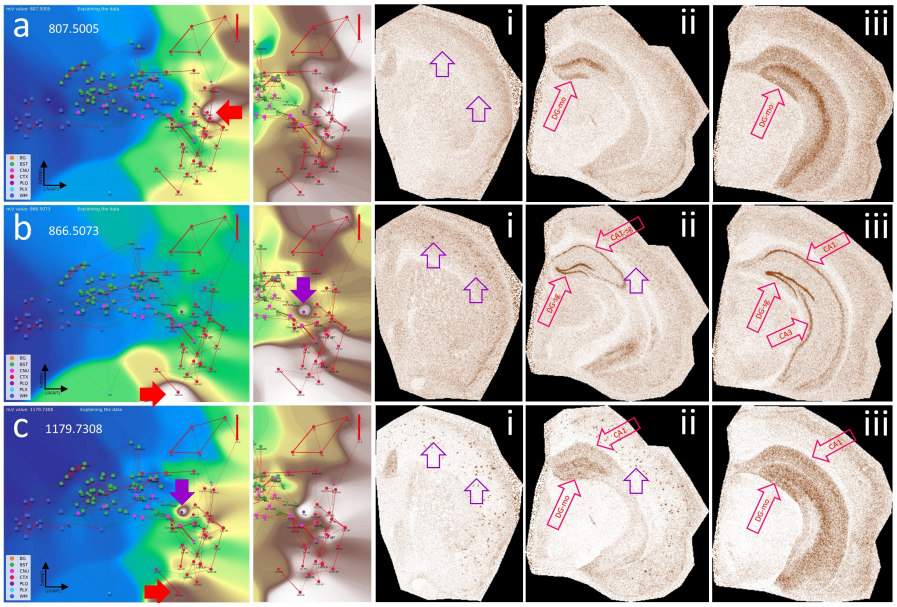

Fig. 13 The hippocampus formation (HPF) has specific lipid signatures for its substructures. a, shows the CBLA and a histogram-normalized part with the VLV of the m/z 807.5005, which strongly emphasizes its specific location in the molecular layer of the Dentate gyrus (red arrows, DG-mo). The violet arrows show tiny speckles of Aβ plaques. **b**, characteristics for the m/z 866.5073 showing the neuron band of the DG and CA1-3 as well as Aβ plaques in the cortex (i), revealing the cortical connections of the DG and CA. **c**, characteristics for the m/z 1179.7308 showing extensive expression in connecting structures orthe neurons from **b**, and cortical Aβ plaques (i), again emphasizing the destruction of hippocampo-cortical connections due to Aβ plaques deposition.

Brain stem nuclei have a unique composition of lipids

The brain stem (BST) consists of diverse nuclei and intricate connections. We therefore hypothesize that the BST lipid composition would show similarities to other regions of the brain. As depicted in Figure 14, CBLA and VLV identified masses such as m/z 786.5286 and 881.5165 that are heterogeneously enriched within the BST, with certain nuclei, such as SUM and APN, demonstrating a pronounced expression pattern. Some BST connections toward the cortex become visible with this m/z, especially to layers 2 and 3. Generally, knowledge of specific lipid patterns of BST nuclei could be used to understand complex neurodegenerative diseases. Some pathological processes of such

diseases start in specific brain ganglia/regions but are linked to the same pathological protein. α-synuclein is an example, which is found in several neurodegenerative diseases with motor impairment: Parkinson's disease, Multiple System Atrophy (MSA), and Lewy-body dementia (LBD).

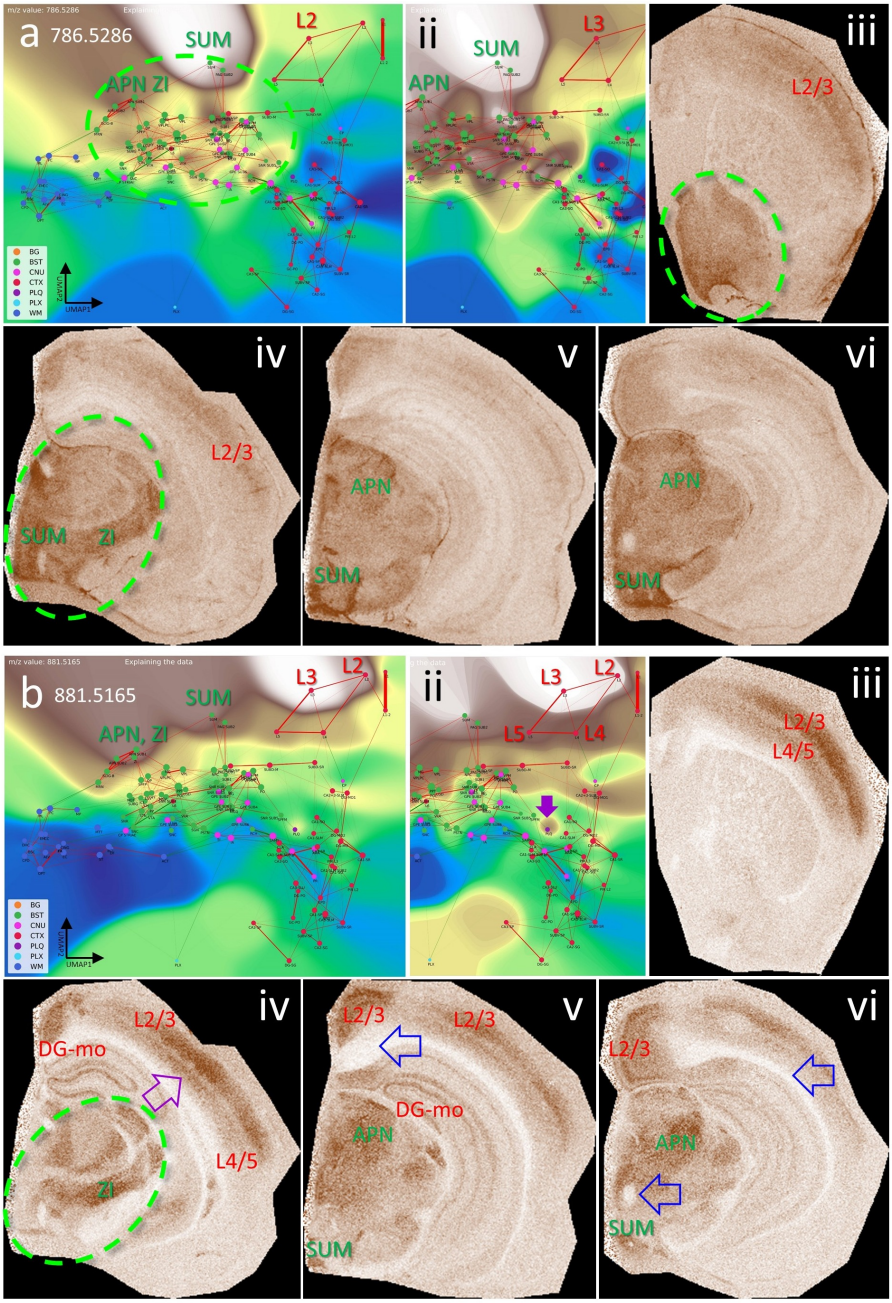

Fig. 14 The brain stem (BST) has specific lipid signatures for its substructures and connections. a,b, show the CBLA and a histogram-normalized selection (ii) with the VLV overlay of the m/z 789.5286 and 881.5165, which strongly emphasizes their specific location in BST nuclei (green ellipse). **ii-vi**, show VPS at 4 locations in the mouse brain stem (green ellipse), highlighting specific nuclei.**aiv**, APN, SUM, and ZI are examples of regions that exhibit higher expression than the surrounding nuclei and tracts. **iii-vi**, Cortex layers 2/3 and 4/5 (CLX_L) show increased expression and are examples of input cortical layers for tracts from the brain stem nuclei. **b iv-vi**, m/z 881.5165 also shows specific expression in HPF regions, specifically in DG-mo, CA1-sr/slm. **b v-vi**, the open blue arrows point to WM tracts that are fully negative, and in **b ii+iv**, the violet arrows point to tiny Aβ plaques residues of the destroyed cortical projections. Legend: BG, background; BST, brain stem; CNU, cerebral nuclei; CTX, cerebral cortex; PLQ, Aβ plaques; PLX, choroid plexus; WM, white matter.

White matter tracts show specific lipid content

The white matter (WM) of the brain represents most of its lipid content as an effective isolation of the neuronal connections to increase connection speed. The WM establishes connections between different brain regions, e.g., between the two hemispheres, called the corpus callosum (cc), but also long tracts within the hemispheres from frontal to occipital lobes. Using CBLA and VLV, we identified lipid signatures that were specific to the WM (see 7, as exemplified by m/z 756.5894 shown in Fig. 15. Thus, the combination of CBLA and VLV enables the identification of comprehensive lipid profiles, which are shared or are unique to specific anatomical regions of the brain, and can serve as resources to generate novel biological hypotheses. Moreover, using other MSI modalities, such as metabolomics and proteomics, these tools can be expanded to integrate complex region-specific information patterns.

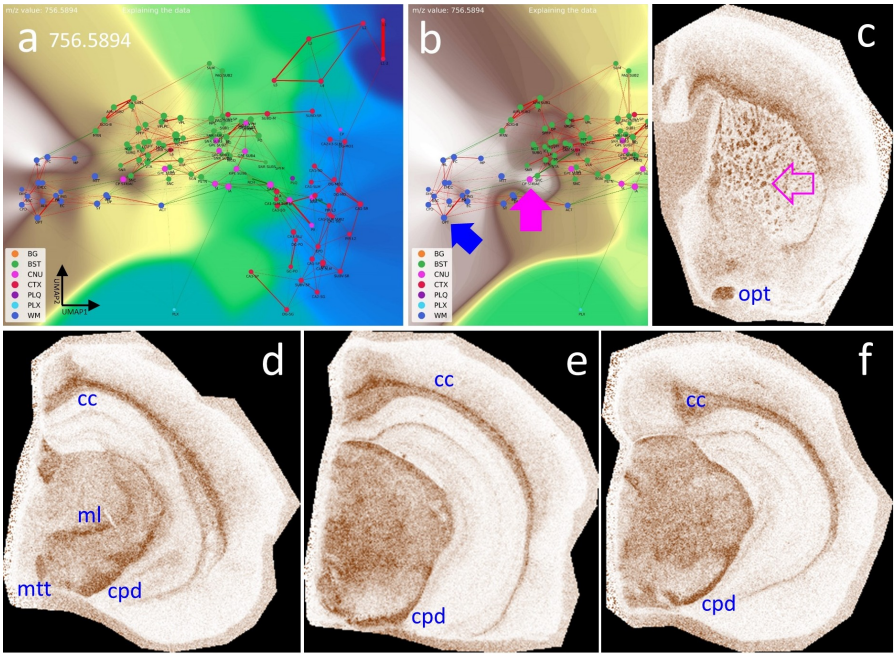

Fig. 15 White matter tracts (WM) have specific lipid signatures. a, shows the CBLA and a histogram-normalized selection (**b**) with the VLV overlay of the m/z 756.5894, which strongly emphasizes their specific location in WM tracts (left cluster, dark blue nodes) and lower in BST tracts (green nodes). The blue arrow points towards the optic tract (opt) as an exposed example highlighted in (**c**), the pink arrow points at the striae in the caudate putamen (CP striae), which are numerous small white matter tracts (**c**) crossing the CP from BST nuclei to the CTX. **c-f**, show four VPS at different locations to visualize the in-situ distribution in the WM tracts in the cc and BST. Legend: BG, background; BST, brain stem; CNU, cerebral nuclei; CTX, cerebral cortex; PLQ, Aβ plaques; PLX, choroid plexus; WM, white matter; cc, corpus callosum; cpd, cerebral peducule; ml, medial lemniscus; mtt, mamillo-thalamic tract; opd, optic tract.

Utilizing the CBLA for visualizing region-specific differences between mouse models.

Finally, we wanted to prove the applicability of the CBLA for general basic research purposes, such as investigations of disease mouse models. In our dataset, we have analyzed tissue from a new mouse model that has a deficiency of the ABCA7 lipid transporter [16]. The ABCA7 transporter is one of the major genetic risk factors for Alzheimer's disease, transporting lipids via the plasma membrane and from the plasma membrane into the intracellular space or cytoplasm [31, 32]. Therefore, investigating its biological function using MSI and appropriate mouse models is of high interest for neurodegenerative disease research. Indeed, using CBLA, we show for the first time examples of m/z values from MSI that are differentially expressed between ABCA7-deficient mice and non-deficient controls (Fig. 16). These m/z values represent lipids that are linked to ABCA7-dependent distribution patterns in the brain and may represent disease-modifying lipids or those directly involved in the pathogenetic processes.

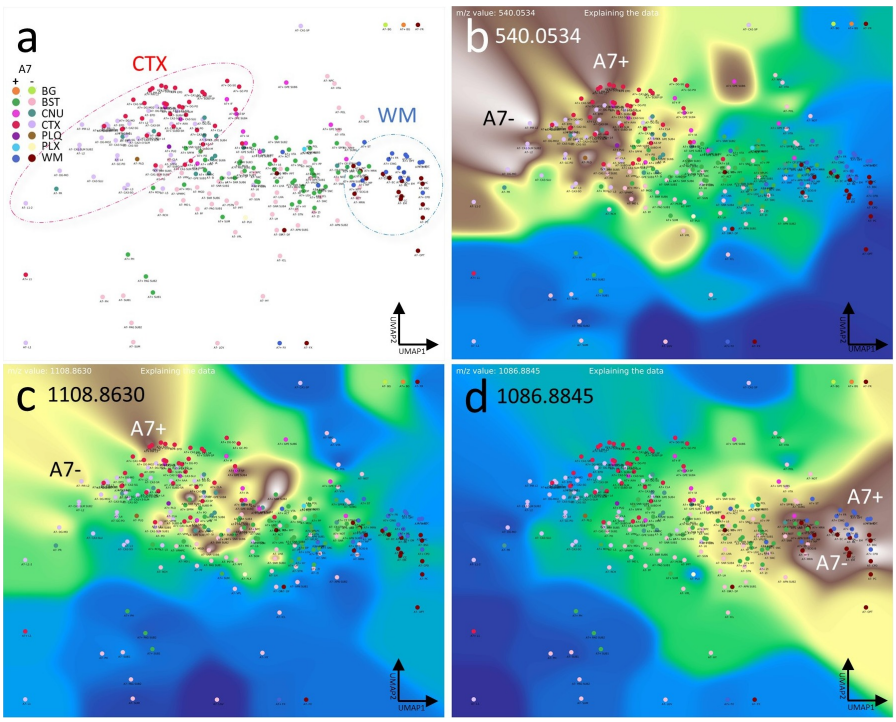

Fig. 16 Lipid transporter ABCA7-deficiency results in altered brain lipid distribution and abundance profile. a, shows the CBLA. Not all annotation categories appear in all mice. Therefore, we designed an ML model to generate annotations for this comparison. We run an ensemble of 100 models and take the majority prediction. We group the mice into the A7- and A7+ mice. For the nodes, we measure the average spectra for each predicted brain region type without TIC normalization to be able to compare raw intensity differences between mice. The circles highlight the CTX (red) and WM (blue) nodes, which are clustered but show a distinct shift in their location due to slight differences in their composition between the A7+ (wild-type) and the A7- (deficient) animals. **b-d**, show examples of three VLVs of differentially expressed m/z values highlighting either A7- or A7+ preference. b and d highlight m/z values that are increased in A7- mice, whereas c shows an m/z that is reduced in A7- mice (increased in A7+ controls). Legend: BG, background; BST, brain stem; CNU, cerebral nuclei; CTX, cerebral cortex; PLQ, Aβ plaques; PLX, choroid plexus; WM, white matter

Discussion

This study introduces the inaugural large-scale annotation and mapping of the brain utilizing solely Mass Spectrometry Imaging (MSI) data, covering a significant portion of categories outlined in the Allen mouse brain atlas [12]. To effectively manage the complexities inherent in high-dimensional MSI data, the multitude of anatomical regions, and their connections to the spatial distribution of selected m/z values, we developed the *Computational Brain Lipid Atlas* (CBLA). The CBLA facilitated both the enhancement of anatomical annotations and the generation of novel biological insights and hypotheses. Distinct from previous studies that concentrated largely on broad anatomical annotations [5, 6, 33], our methodology captures intricate brain substructures without the need for other modalities, e.g. immunostains [7]. By incorporating multiple continuous visualization panels based on raw MSI data, we streamline the annotation process by eliminating the need for external imaging modalities, while enabling the annotation of a substantially larger array of brain regions.

Our findings illustrate that MSI, even when constrained to a small lipidomics dataset with limited spatial resolution as shown here, can uncover very detailed brain anatomy and act as a foundational tool for biological exploration.

Leveraging CBLA, we provide pioneering evidence that lipid distributions exhibit unique regional signatures, permitting differentiation of brain sub-regions based solely on lipid composition. This work presents a comprehensive mapping of m/z values linked to specific brain regions, which is utilized to scrutinize the spatial organization of Aβ plaques. Our analysis indicates that these plaques localize m/z values from their originating region, with spatial distributions mirroring functional connectivity and forming linkages to distant brain areas. These revelations are enabled through the synergistic m/z mapping and CBLA visualization framework. We further demonstrate the application's utility in investigating hypotheses regarding the hippocampal network, extrapyramidal motor pathways, and other neuronal systems.

Several limitations are noteworthy. To address m/z values beyond fixed binning resolution, we devised a novel Gaussian mixture model (GMM)-based approach [22]. Although this enhances precision, it also introduces a trade-off: greater sensitivity to subtle differences comes with the risk of fragmenting a single biological signal into multiple, closely related peaks. One promising direction to mitigate this challenge is the incorporation of spatial priors, enabling the resolution of m/z values within the anatomical context of annotated regions. Alternatively, the use of Trapped Ion-Mobility Time-of-Flight (TIMS-TOF) mass spectrometry [34] could further disambiguate signals and reduce redundancy. Another challenge lies in the annotation of highly similar sub-regions. Our approach involved iterative refinement using panels of varied visualizations, each highlighting distinct structural characteristics. Future advancements could include active learning strategies that prioritize uncertain or under-annotated regions, thereby focusing expert attention more effectively. Also of use would be quantitative scoring for the annotations. This information is indirectly embedded in the atlas as visual cues—annotation errors manifest as overlapping nodes and the appearance of spurious strong edges between them. However, the introduction of quantitative region-annotation scores for MSI could further minimize such errors and improve overall accuracy.

MSI data is inherently rich and informative, yet remains underexploited due to the absence of tools for high-resolution and interpretable analysis. The MSI-ATLAS presented herein constitutes the first large-scale, intelligible mapping of lipid distributions across brain regions. It establishes a foundational basis for subsequent studies aiming to delve into the brain's intricate molecular architecture and bridge spatial metabolomics with neuroanatomical function.

Conclusion

MSI data is inherently rich and informative, yet remains underexploited due to the absence of tools for high-resolution and interpretable analysis. The MSI-ATLAS presented herein constitutes the first large-scale, intelligible mapping of lipid distributions across brain regions. It establishes a foundational basis for subsequent studies aiming to delve into the brain's intricate molecular architecture and bridge spatial metabolomics with neuroanatomical function.

List of Abbreviations

Aβ - amyloid-β, BG - background category, BST - brain stem category, CBLA - Computational Brain Lipidomics Atlas, CNU - cerebral nuclei category, CTX - cortex category, GMM - Gaussian mixed model, IHC - immunohistochemistry MSI - Mass spectrometry imaging, MSI-ATLAS - Mass spectrometry imaging atlas computational tool, MSI-VISUAL - Mass spectrometry imaging visualization tool, PLQ - Aβ plaque category, PLX - plexus category, TIC - total ion current TIMSTOF - Trapped-Ion-Mobility Time-of-flight, VLV - Virtual Landscape Visualization, VPS - Virtual Pathology Stain, WM - white matter category

Anatomical abbreviations were used according to the Allen mouse brain atlas nomenclature [12].

Declarations

Ethics approval and consent to participate

Not applicable.

Consent for publication

Not applicable.

Availability of Data and Material

The datasets generated and/or analyzed during the current study are available in the ProteomeXchange PRIDE repository with the dataset number PXD056609.

Availability of code

We release the code as an open-source package on GitHub: https://github.com/jacobgil/msi-atlas

Competing interests

The authors do not declare competing interests.

Funding

J.P. received funding from Norges forskningsråd [NFR, Norway; 327571 (PETABC), 295910 (NAPI, www.napi.uio.no)], HelseSØ (Norway; 2022046), and the EIC Pathfinder Open Challenges program (EU commission; 101185769). The Proteomics Core Facility at the University of Oslo and the Oslo University Hospital is supported by the Core Facilities Program of the South-Eastern Norway Regional Health Authority (HSØ) and NAPI (www.napi.uio.no, NFR, Norway; 295910).

Authors' contributions

J.G. analyzed the MSI data and wrote and revised the manuscript. J.S. performed MSI measurements and revised the final manuscript. J.P. analyzed the MSI data, annotated data, wrote and revised the manuscript, acquired funding, and provided infrastructure.

Acknowledgements

We thank Thomas Brüning for preparing the brain samples and Intellislides for MSI measurements.

References

[1] Gildenblat, J. & Pahnke, J. Dimensionality reduction with strong global structure preservation. *Pattern Analysis and Applications* **38** (2026). URL https://doi.org/10.1007/s10044-025-01585-9.

[2] Gildenblat, J. & Pahnke, J. Truthful visualizations for mass spectrometry imaging enable high spatial resolution interactive m/z mapping and exploration. *BioRxiv* (2025). URL https://doi.org/10.1101/2025.03.18.643852.

[3] Huang, H. X. *et al.* Mass spectrometry imaging highlights dynamic patterns of lipid co-expression with aβ plaques in mouse and human brains. *Journal of Neurochemistry* **168**, 1193–1214 (2024). URL https://onlinelibrary.wiley.com/doi/10.1111/jnc.16042.

[4] Guo, G. *et al.* Automated annotation and visualisation of high-resolution spatial proteomic mass spectrometry imaging data using hit-map. *Nature Communications* **12**, 3241 (2021). URL https://www.nature.com/articles/s41467-021-23461-w.

[5] Sun, C. *et al.* Spatially resolved multi-omics highlights cell-specific metabolic remodeling and interactions in gastric cancer. *Nature Communications* **14**, 2692 (2023). URL https://www.nature.com/articles/s41467-023-38360-5.

[6] Xie, Y. R. *et al.* Multiscale biochemical mapping of the brain through deep-learning-enhanced high-throughput mass spectrometry. *Nat Methods* **21**, 521–530 (2024). URL https://www.ncbi.nlm.nih.gov/pubmed/38366241.

[7] Trinklein, T. J. *et al.* Multimodal mass spectrometry imaging for plaque- and region-specific neurolipidomics in alzheimer's disease mouse models. *Nat Commun* **16**, 10969 (2025). URL https://www.ncbi.nlm.nih.gov/pubmed/41365882.

[8] Muller, E. *et al.* Exploring the abeta plaque microenvironment in alzheimer's disease model mice by multimodal lipid-protein-histology imaging on a benchtop mass spectrometer. *Pharmaceuticals (Basel)* **18** (2025). URL https://www.ncbi.nlm.nih.gov/pubmed/40006065.

[9] Schede, H. H. *et al.* Unified mass imaging maps the lipidome of vertebrate development. *Nat Methods* (2025). URL https://www.nature.com/articles/s41592-025-02771-7.pdf.

[10] Gildenblat, J. & Pahnke, J. Preserving clusters and correlations: a dimensionality reduction method for exceptionally high global structure preservation (2025). URL https://arxiv.org/abs/2503.07609. arXiv:2503.07609.

[11] McInnes, L., Healy, J. & Melville, J. Umap: Uniform manifold approximation and projection for dimension reduction. *arXiv preprint arXiv:1802.03426* (2018).

URL https://arxiv.org/abs/1802.03426.

[12] Sunkin, S. M. *et al.* Allen brain atlas: an integrated spatio-temporal portal for exploring the central nervous system. *Nucleic Acids Res* **41**, D996–D1008 (2013). URL https://www.ncbi.nlm.nih.gov/pubmed/23193282.

[13] Brackhan, M. *et al.* Isotope-labeled amyloid-beta does not transmit to the brain in a prion-like manner after peripheral administration. *EMBO Rep* **23**, e54405 (2022). URL https://www.ncbi.nlm.nih.gov/pubmed/35620875.

[14] Bascunana, P., Brackhan, M. & Pahnke, J. Machine learning-supported analyses improve quantitative histological assessments of amyloid-beta deposits and activated microglia. *J Alzheimers Dis* **79**, 597–605 (2021). URL https://www.ncbi.nlm.nih.gov/pubmed/33337377.

[15] Radde, R. *et al.* Abeta42-driven cerebral amyloidosis in transgenic mice reveals early and robust pathology. *EMBO Rep* **7**, 940–6 (2006). URL https://www.ncbi.nlm.nih.gov/pubmed/16906128.

[16] Santos-García, I. *et al.* The abc transporter a7 modulates neuroinflammation via nlrp3 inflammasome in alzheimer's disease mice. *Alzheimer's Research & Therapy* **17**, 30 (2025). URL https://www.ncbi.nlm.nih.gov/pubmed/39871385.

[17] Andersen, M. K. *et al.* Spatial differentiation of metabolism in prostate cancer tissue by maldi-tof msi. *Cancer Metab* **9**, 9 (2021). URL https://www.ncbi.nlm.nih.gov/pubmed/33514438.

[18] Pearson, K. On lines and planes of closest fit to systems of points in space. *Philosophical Magazine* **2**, 559–572 (1901). URL https://www.tandfonline.com/doi/abs/10.1080/14786440109462720.

[19] Lee, D. D. & Seung, H. S. Learning the parts of objects by non-negative matrix factorization. *Nature* **401**, 788–791 (1999). URL https://www.nature.com/articles/44565.

[20] Bachem, O., Lucic, M. & Krause, A. Scalable k-means clustering via lightweight coresets. *arXiv* (2017). URL https://arxiv.org/abs/1702.08248.

[21] Kibbe, R. R. & Muddiman, D. C. Quantitative mass spectrometry imaging (qmsi): A tutorial. *Journal of Mass Spectrometry* **59**, e5009 (2024). URL https://analyticalsciencejournals.onlinelibrary.wiley.com/doi/10.1002/jms.5009.

[22] Dempster, A. P., Laird, N. M. & Rubin, D. B. Maximum likelihood from incomplete data via the em algorithm. *Journal of the Royal Statistical Society: Series B (Methodological)* **39**, 1–38 (1977). URL https://rss.onlinelibrary.wiley.com/doi/10.1111/j.2517-6161.1977.tb01600.x.

[23] Pedregosa, F. *et al.* Scikit-learn: Machine learning in python. *Journal of Machine Learning Research* **12**, 2825–2830 (2011). URL https://www.jmlr.org/papers/volume12/pedregosa11a/pedregosa11a.pdf.

[24] *The VGG Image Annotator (VIA)* (ACM, 2019). URL https://dl.acm.org/doi/10.1145/3343031.3350535.

[25] Bingham, E. *et al.* Pyro: Deep universal probabilistic programming. *The Journal of Machine Learning Research* **20**, 973–978 (2019). URL https://pyro.ai.

[26] Kursa, M. B. & Rudnicki, W. R. Feature selection with the boruta package. *Journal of Statistical Software* **36**, 1–13 (2010). URL https://doi.org/10.18637/jss.v036.i11.

[27] Parisot, A. *et al.* Stabl: Stable attribution of biomarker lists in high-dimensional omics data. *Nature Communications* **14**, 2163 (2023). URL https://www.nature.com/articles/s41467-023-37848-w.

[28] IEEE. *Probabilistic Interpretation of Feedforward Classification Network Outputs, with Relationships to Statistical Pattern Recognition.* URL https://ieeexplore.ieee.org/document/118638.

[29] Gorska, A. M. *et al.* Abca7 deficiency exacerbates glutamate excitotoxicity in alzheimer's disease mice - a new pharmacological target for glu-related neurotoxicity. *Prog Neurobiol* **259**, 102891 (2026). URL https://www.ncbi.nlm.nih.gov/pubmed/41651103.

[30] Hunter, J. D. Matplotlib: A 2d graphics environment. *Computing in Science & Engineering* **9**, 90–95 (2007).

[31] Villa, M., Wu, J., Hansen, S. & Pahnke, J. Emerging role of abc transporters in glia cells in health and diseases of the central nervous system. *Cells* **13** (2024). URL https://www.ncbi.nlm.nih.gov/pubmed/38727275.

[32] Dib, S., Pahnke, J. & Gosselet, F. Role of abca7 in human health and in alzheimer's disease. *Int J Mol Sci* **22** (2021). URL https://www.ncbi.nlm.nih.gov/pubmed/33925691.

[33] Clarke, H. A. *et al.* Spatial mapping of the brain metabolome lipidome and glycome. *Nature Communications* **16** (2025). URL https://www.nature.com/articles/s41467-025-59487-7.

[34] Spraggins, J. M. *et al.* High-performance molecular imaging with maldi trapped ion-mobility time-of-flight (timstof) mass spectrometry. *Anal Chem* **91**, 14552–14560 (2019). URL https://www.ncbi.nlm.nih.gov/pubmed/31593446.

Supplementary Data

Comparison of lipidomic profiles per brain region

m/z mapping data. The csv-files with the mapping of the m/z values to the categories can be found here:

1. the model-based mapping - link to csv file
2. the mPAUC mapping - link to csv file

UMAP visualization of all annotations. The UMAP visualizations of the m/z content of the annotations are given in the Extended Data, Fig. 17, showing that different brain regions have distinct lipidomic profiles, enabling their visualization and annotation.

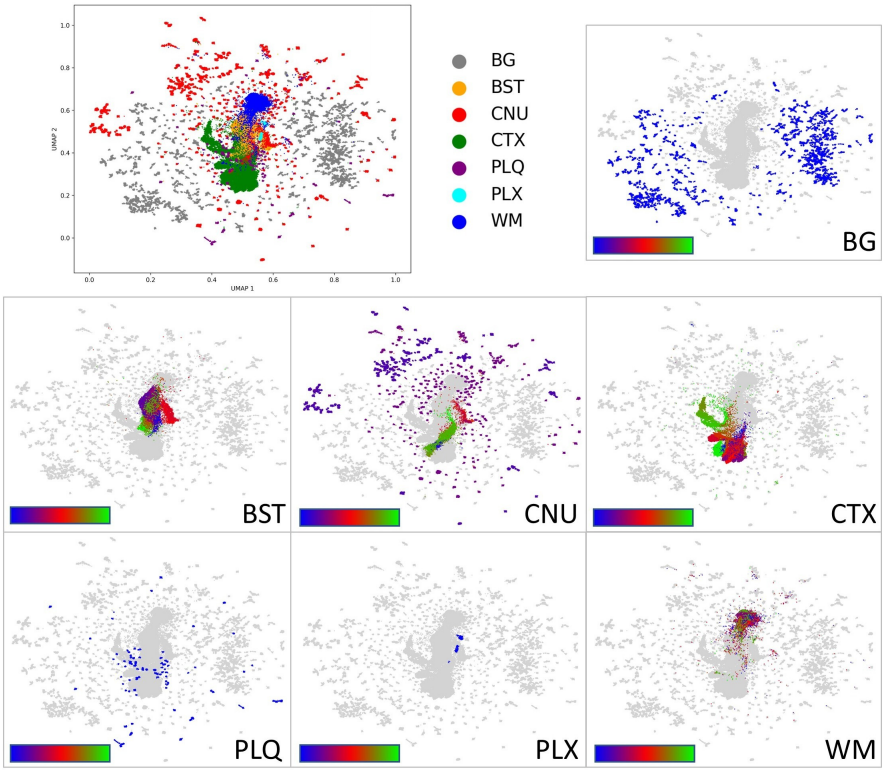

Fig. 17 Analysing lipid profiles of brain regions. The high-dimensional MSI data was reduced to two dimensions using UMAP [11]. The 123 regions were grouped into the main categories according to their anatomical locations [12] or as background (BG, ground truth). Each main category was also colorized with a BRG color scheme, showing the diversity within the categories. This visualization shows that annotated points from the same categories tend to distribute similarly in the low-dimensional visualization, which would be expected if the annotations were correct. On the other hand, our annotations capture extreme diversity in the data, especially in the CNU, WM, CTX, and BST categories. Legend: BG, background; BST, brain stem; CNU, cerebral nuclei; CTX, cerebral cortex; PLQ, Aβ plaques; PLX, choroid plexus; WM, white matter.