



Multiscale biochemical mapping of the brain through deep-learning-enhanced high-throughput mass spectrometry

Received: 5 June 2023

Accepted: 8 January 2024

Published online: 16 February 2024

Check for updates

Yuxuan Richard Xie^{1,2}, Daniel C. Castro^{1,2,3}, Stanislav S. Rubakhin^{2,4}, Timothy J. Trinklein^{2,4}, Jonathan V. Sweedler^{1,2,3,4,5,6} & Fan Lam^{1,2,5,6,7}

Spatial omics technologies can reveal the molecular intricacy of the brain. While mass spectrometry imaging (MSI) provides spatial localization of compounds, comprehensive biochemical profiling at a brain-wide scale in three dimensions by MSI with single-cell resolution has not been achieved. We demonstrate complementary brain-wide and single-cell biochemical mapping using MEISTER, an integrative experimental and computational mass spectrometry (MS) framework. Our framework integrates a deep-learning-based reconstruction that accelerates high-mass-resolving MS by 15-fold, multimodal registration creating three-dimensional (3D) molecular distributions and a data integration method fitting cell-specific mass spectra to 3D datasets. We imaged detailed lipid profiles in tissues with millions of pixels and in large single-cell populations acquired from the rat brain. We identified region-specific lipid contents and cell-specific localizations of lipids depending on both cell subpopulations and anatomical origins of the cells. Our workflow establishes a blueprint for future development of multiscale technologies for biochemical characterization of the brain.

Genomic and transcriptomic tools have transformed neuroscience by allowing us to visualize, untangle and understand the spatiotemporal expression patterns of thousands of genes in the brain, as well as how they are related to various functions and diseases^{1–3}. Beyond gene expression profiles, the biochemical compositions and dynamics of metabolites, lipids, peptides and proteins have essential roles in many neurobiological processes^{4,5}, and they have been implicated in neurodevelopment⁶, learning, memory⁷, aging^{8,9} and a myriad of neurological or neurodegenerative diseases¹⁰. Approaches to characterize these molecular compositions offer invaluable insight complementary to transcriptomics. However, comprehensive biochemical profiling of

both tissue and single cells at a whole-organ level remains challenging. Recent technical advances in single-cell measurements using isolated populations of individual cells and mass spectrometry (MS) have great potential to solve these bottlenecks, prompting single-cell metabolomics to be listed as one of the technologies to watch in 2023¹¹. MS is recognized as a key method of choice for metabolomic and proteomic measurements due to its unique capability of untargeted, sensitive and specific detection of numerous biomolecules in both tissues^{12,13} and single cells^{14–17}. Spatial organizations of biomolecules in the brain have been mapped at cellular and subcellular resolution using advanced MS imaging (MSI) methods^{18–21}. Nevertheless, profiling of small metabolites

¹Department of Bioengineering, University of Illinois Urbana-Champaign, Urbana, IL, USA. ²Beckman Institute for Advanced Science and Technology, University of Illinois Urbana-Champaign, Urbana, IL, USA. ³Department of Molecular and Integrative Physiology, University of Illinois Urbana-Champaign, Urbana, IL, USA. ⁴Department of Chemistry, University of Illinois Urbana-Champaign, Urbana, IL, USA. ⁵Carle-Illinois College of Medicine, University of Illinois Urbana-Champaign, Urbana, IL, USA. ⁶Carl R. Woese Institute for Genomic Biology, University of Illinois Urbana-Champaign, Urbana, IL, USA. ⁷Department of Electrical and Computer Engineering, University of Illinois Urbana-Champaign, Urbana, IL, USA. ✉e-mail: jsweedle@illinois.edu; fanlam1@illinois.edu

and lipids in large brain regions in three dimensions at single-cell resolution with simultaneous brain-wide coverage and chemical detail (important for untargeted and unbiased molecular characterization) has not yet been achieved. We provide several innovations to existing workflows that enable multiscale biochemical profiling at a scale not previously attempted. First, as existing high-resolution MSI is throughput-limited, we integrate deep-learning approaches to enhance high-mass-resolving Fourier-transform MS (FTMS) acquisition by tenfold, enabling imaging of many tissue sections with brain-wide coverage and reconstruction of three-dimensional (3D) molecular distributions or atlases. Second, high-throughput single-cell MS (SCMS) allows populations of individual cells to be characterized²²; however, isolated cells lack spatial context of tissue. We integrate both workflows (high-throughput tissue MSI and SCMS) to map the chemical profiles of single cells onto tissue sections, allowing multiscale characterization of spatial–biochemical organization of the brain.

More specifically, we introduce MEISTER, a framework of MS for integrative single-cell and tissue analysis with deep-learning-based reconstruction that integrates high-throughput MS platforms with several technical innovations: (1) a deep-learning-based signal reconstruction approach capable of producing high-resolution mass spectra with greatly enhanced throughput for both tissue MSI and SCMS; (2) a multimodal image registration technique that produces coherent 3D reconstruction of MSI data from many tissue sections and affords quantitative analysis of regional chemical profiles; and (3) a computational approach that exploits dictionary learning concepts to create and map cell-specific chemical profiles to tissue imaging data for multiscale integration. We validated MEISTER using computational simulations, as well as experimental tissue MSI and SCMS data. With MEISTER, we achieved 3D mapping of the rat brain with an unprecedented combination of large volume coverage, high spatial resolution (50- μm lateral and 16- μm sections) over millions of pixels and high chemical content (>1,000 lipid features). We also profiled 13,566 single cells that were isolated from five rat brain regions and built cell-type-specific chemical dictionaries, which were then mapped to the tissue images, to obtain spatially resolved cell type distributions across the brain. To further demonstrate the capabilities of our framework, we studied how lipids associate with the brain's anatomical structures. We extracted lipid profiles from 11 brain regions by registering serial MSI sections to a rat magnetic resonance imaging (MRI) brain atlas using a data-driven nonlinear image registration method that generated volumetric reconstruction of thousands of lipid features over a large brain volume while identifying region-specific lipid contents. With the single cell-to-tissue data integration approach, we identified heterogeneous lipid distributions and differential lipid features at both tissue and single-cell levels, discovering relationships of single-cell biochemical profiles to region-specific spatial distributions of lipids. We demonstrated the potential of MEISTER as a general multiscale tissue biochemical characterization approach by also applying it to another tissue type, rat pancreas, and to molecules beyond lipids, for example, peptides.

Results

A deep-learning-enabled, high-throughput multiscale MSI framework

MEISTER integrates high-throughput MS experiments, a deep-learning-based signal reconstruction method and data-driven high-dimensional MSI analysis to enable brain-wide, multiscale profiling of brain biochemistry. To resolve detailed chemical contents, we collected both high-resolution tissue MSI and SCMS data, leveraging a high-throughput experimental platform using matrix-assisted laser desorption/ionization (MALDI) Fourier-transform ion cyclotron resonance (FT-ICR) MS. Achieving brain-wide coverage and cell-specific profiling requires probing a large number of tissue sections and cells (Fig. 1a), which is time-prohibitive on high-mass-resolution platforms such as FT-ICR (Methods, 'Signal modeling'). To this end, we developed

a deep-learning model to reconstruct high-resolution MS data from low-mass resolving measurements (Fig. 1b). In short, we model the underlying high-dimensional transient signals \mathbf{S} as points on a low-dimensional nonlinear manifold embedded in the high-dimensional space. These low-dimensional embeddings \mathbf{Z} can be effectively learned by training a deep autoencoder (DAE) network, using experimental and/or simulated full transients with the desired mass resolution. The presence of low-dimensional representations implies that high-mass-resolution spectra can be reconstructed from substantially shorter transients than are conventionally acquired. We realized this by training a 'regressor' network jointly with the DAE to estimate the low-dimensional embeddings \mathbf{Z} from only short transients $\hat{\mathbf{S}}$, which were subsequently decoded (by the same decoder from the DAE) into high-mass-resolution data (Extended Data Fig. 1a and Methods). For 3D MSI, the networks were trained on only a small number of tissue sections and applied to reconstruct data for the remaining sections consisting of millions of pixels (Extended Data Fig. 1b and Methods). For SCMS, a small subset of cells were used for training and applied to large cell populations (Extended Data Fig. 1c), allowing much higher data collection throughput. Particularly, an MSI dataset containing more than 1.5 million pixels required only 20 h of acquisition time, which would have taken about 300 h using the conventional acquisition approach. This allowed us to image 16- μm -thick serial sections from rat brains that covered a range of -10 mm (along z) with a raster width of 50 μm used in MSI. In conjunction with the MSI, 13,566 single cells isolated from five brain regions (the neocortex, hippocampus, thalamus, striatum and corpus callosum) were probed using an image-guided MALDI SCMS²³ approach (Fig. 1a). Detailed experimental parameters for MSI and SCMS can be found in Methods.

To enable biochemical characterization of the brain and knowledge discovery from such unprecedented data, we developed and integrated several data-driven methods for analyzing the high-dimensional, multiscale 3D MSI and SCMS data (Fig. 1c). First, the MSI data were mapped to MRI and the Waxholm Space atlas through a customized nonlinear image registration procedure (Fig. 1c, left), which enabled a coherent volumetric MSI reconstruction from the sections imaged. Through registration, we extracted the mass spectra of 11 major brain structures across the 3D volume and identified spatially differential biochemical profiles. Next, we classified the brain structures on the basis of their lipids for tissue typing, which identified enriched lipid species in each region or tissue type. To connect the tissue MSI and SCMS across different scales, we built cell-type-specific 'chemical dictionaries' and introduced a joint union-of-subspaces (UoS) fitting technique that resolved cell-specific contributions to the spatio-chemical contents at the brain-wide tissue level (Fig. 1c, right).

Validation of the deep-learning-based MSI reconstruction

Using a carefully designed, biochemically relevant simulated MSI dataset that contained rich chemical details and brain-mimicking spatial variations, we trained and validated the proposed deep-learning-based method for reconstructing high-mass-resolution mass spectra and ion images from noisy short transients (Extended Data Fig. 2a,b and Methods). Our method showed near-ground-truth-level fidelity with a gain of more than 10 dB in signal-to-noise ratios (SNRs) over the noisy reference using only 5% data. We compared the performance of our model for spectral and spatial feature recovery in the simulated dataset to that of the standard FT reconstruction and a previously described linear subspace approach^{24,25} (Extended Data Fig. 2c,d). Compared to the standard FT reconstruction from full transients, our method also yielded a higher SNR, owing to the denoising effects of the learned low-dimensional representation.

To evaluate our method on experimental data, we trained the model using high-resolution MSI data acquired from rat brain tissue sections using an FT-ICR mass spectrometer. We then validated the model using reference full-transient data acquired on different days from

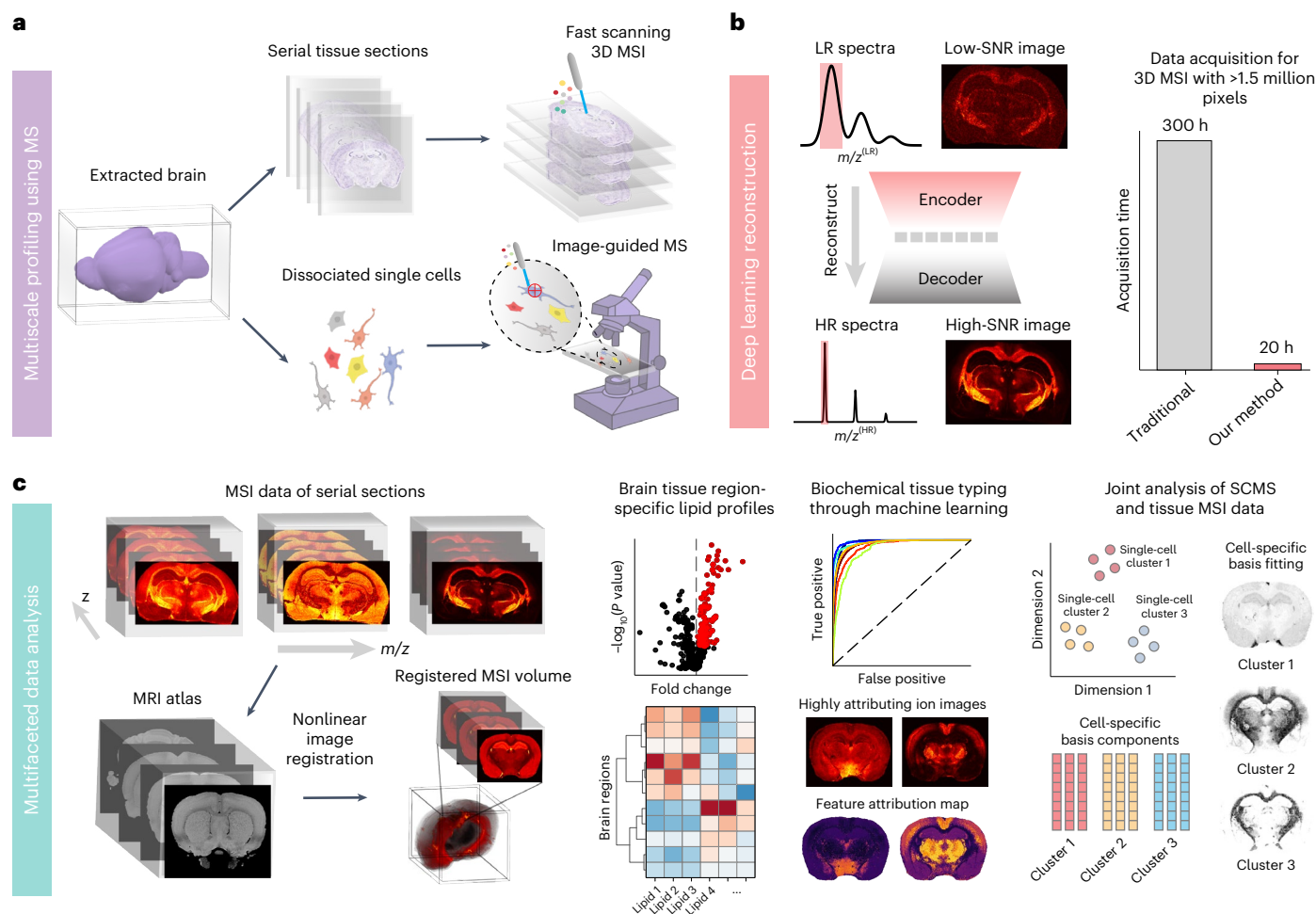


Fig. 1 | The MEISTER framework for multiscale biochemical profiling using high-mass-resolution MS enhanced by computational methods.

a, Obtained from surgically extracted brain, serial tissue sections are imaged for 3D MSI using a fast acquisition strategy, and single-cell populations prepared by tissue dissociation are probed with high-throughput image-guided MS. **b**, A deep-learning model reconstructs high-mass-resolving and high-SNR MS data from the low-mass-resolution measurements acquired with fast acquisitions by exploiting a low-dimensional manifold structure for high-dimensional MS

data, producing large datasets with millions of pixels, which was previously time-prohibitive with the conventional acquisition. LR, low resolution; HR, high resolution. **c**, Our multifaceted data analysis pipeline uses various data-driven methods for multimodal image registration to align MSI with 3D anatomical MRI for volumetric reconstruction, identifying differential lipid distributions, tissue typing using MS data and integrating MSI and SCMS data for joint analysis and resolving cell-type-specific contributions at the tissue level across the brain.

tissue sections not seen during training. For the noise-contaminated reference (transient duration of 0.731 s, 1 million temporal points), images from peaks that were indicated in the single-pixel mass spectra showed ions with distinct spatial distributions, whereas the signals were unresolved in the reduced data with short transients (first 64,000 temporal points) due to poor mass resolution (Supplementary Fig. 1a,b). The deep-learning-based reconstruction from the reduced data successfully resolved nearby mass features, providing enhanced signal strength over the subspace-based reconstruction also from the same reduced data (Fig. 2a,b and Supplementary Fig. 1b). Our method yielded quantitatively better spectral and spatial fidelity with respect to the reference than reduced data and subspace reconstruction. This was further supported by subsequent principal component analysis (PCA) and spatial segmentation through *k*-means clustering on the reconstructed spectral features (Extended Data Fig. 2e–g and Supplementary Fig. 2), with our method producing less noisy spatial parcellation. Our evaluation suggests that the model can learn robust nonlinear low-dimensional features from complex and noisy imaging data, while accurately predicting those features from short transients, even for the highly heterogeneous brain tissue.

Furthermore, we examined the model performance for reconstructing SCMS data. Specifically, we trained a model using high-resolution SCMS data from approximately 4,000 cells, and we tested it on 1,000 independent cells (Methods). We found high correlation scores (Pearson $r > 0.95$) between full-resolution reference and reconstructed single-cell spectra. Consistent molecular profiles across individual cells (Extended Data Fig. 3a) resulted in nearly identical outcomes by uniform manifold approximation and projection (UMAP) and *k*-means clustering (Extended Data Fig. 3b). Even with larger chemical heterogeneity, our model was able to effectively recover the variations within and across cell populations (Extended Data Fig. 3c).

High-resolution 3D MSI with large volume coverage

High-mass-resolving MSI with 3D tissue coverage has been shown^{26,27}. However, the combination of mass and spatial resolution and organ coverage for volumetric imaging with FTMS has been limited, due to the inherent throughput constraint; for example, resolution may be sacrificed (pixel size of $>50 \mu\text{m}$) when imaging multiple sections for 3D imaging or the number of sections may be reduced to maintain a small pixel size. The higher throughput afforded by MEISTER allowed us to

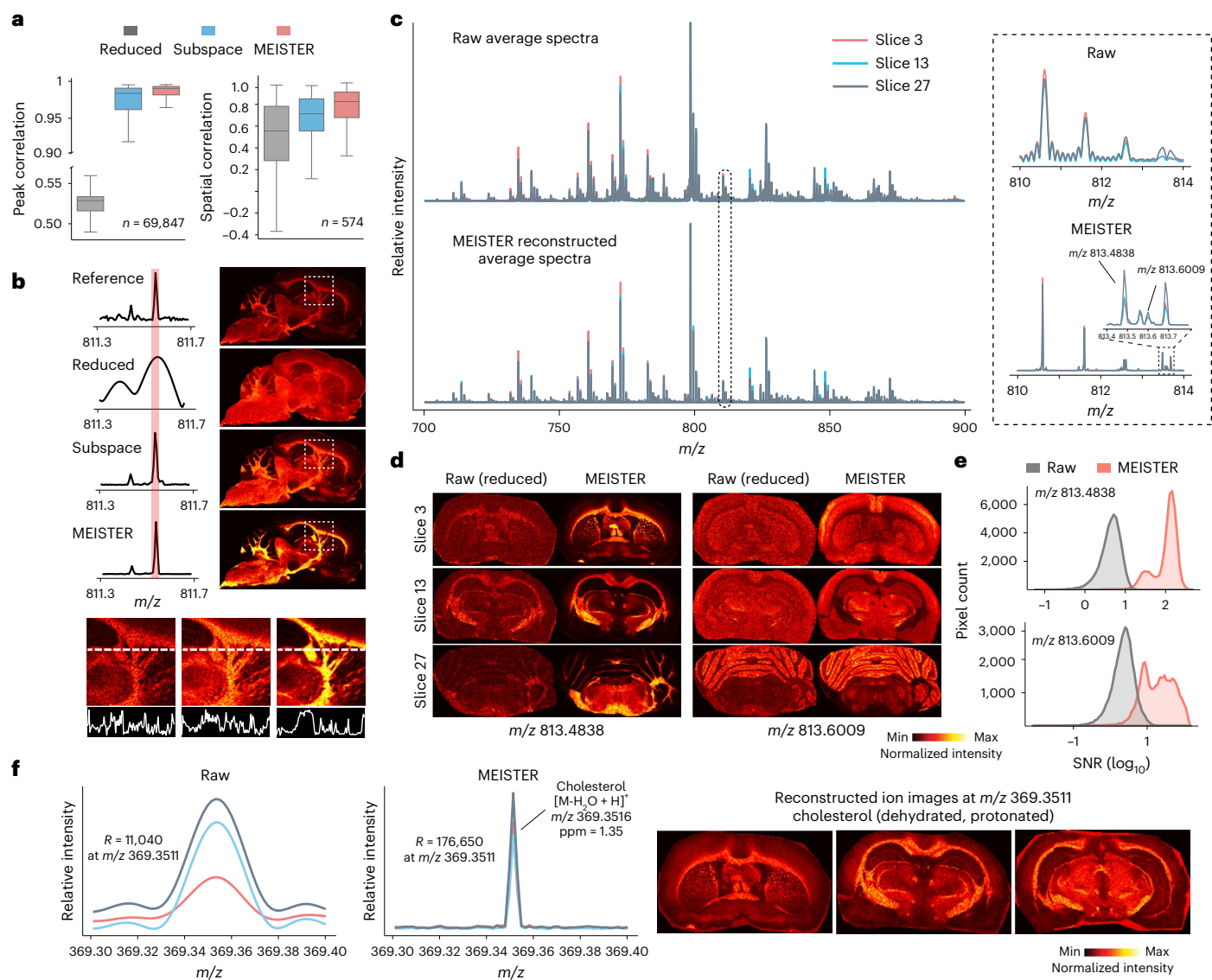


Fig. 2 | High-throughput MSI using MEISTER. **a**, Correlation measures of mass spectra produced by different reconstruction methods; the case ‘reduced’ stands for a standard FT reconstruction from short transients (reduced data); the test data contain $n = 69,847$ pixels and $n = 574$ ion images for peak and spatial correlation, respectively. Data in boxplots are shown as median values (center) with the interquartile range (box), and the whiskers extend to 1.5 times the interquartile range. **b**, Comparisons of mass spectra and corresponding ion images to the reference (full transient; top row) show enhanced signal strength with MEISTER. The bottom row depicts line profiles of signal intensities in reference, subspace and MEISTER images. **c**, Averaged mass spectra from three

different tissue sections for the reduced data (top) and our reconstruction (bottom). The inset displays a small m/z window of the mass spectra, demonstrating mass-resolution (R) enhancement. **d**, Ion images obtained from raw MSI data (left) and our reconstruction (right) at m/z 813.4838 and m/z 813.6009. **e**, Reconstruction improved SNRs as shown in SNR distributions for signals at m/z 813.4838 (top, 100-fold higher) and at m/z 813.6009 (bottom, tenfold higher) compared to raw data. **f**, Raw and reconstructed mass spectra zoomed into the cholesterol range (left columns) and corresponding ion images showing tissue localization of cholesterol (right columns).

spatially profile metabolites and lipids for many serial tissue sections that cover a large volume of the brain. To demonstrate this capability, we imaged 37 coronal and 39 sagittal sections of the rat brain, and we used data with the targeted mass resolution from a few tissue sections for training (Methods). We were able to efficiently reconstruct high-mass-resolution, high-SNR spectra from raw data acquired with short transient duration (that is, <10% collection time per transient) for all remaining serial sections for approximately 2 million total pixels in each 3D dataset (Extended Data Fig. 4a and Supplementary Fig. 3). Reconstructed data exhibited substantially improved quality with a greater than tenfold increment in SNRs over the raw data processed by traditional FT analysis (Fig. 2c–e and Extended Data Fig. 4b,c), while maintaining high mass accuracy on several expected lipid signals in rat

brain and low mass errors on tentatively assigned lipids (Extended Data Fig. 4d). Our method is also applicable to different organ systems and molecules other than lipids. To demonstrate its generalizability, we imaged and trained models on rat pancreas tissue sections (Extended Data Fig. 5a,b and Methods). Faithful detection and reconstruction of rat pancreatic peptides including glucagon, insulin 1 C-peptide and insulin 2 C-peptide from reduced transient data are shown (Extended Data Fig. 5c).

To enable spatially resolved biochemical profiling across the brain, we designed and implemented a multimodal registration strategy to align the misaligned reconstructed MSI sections to a high-resolution rat brain MRI atlas²⁸ to form a volumetric reconstruction. Inspired by a previously proposed approach, we applied parametric UMAP²⁹

to embed the MSI hyperspectral data cube into ‘feature’ images for co-registration with MRI anatomical images (Extended Data Fig. 6a). By learning the embedding process, our method can simultaneously obtain low-dimensional representations of the entire 3D MSI data cube (Extended Data Fig. 6b and Supplementary Fig. 4a,b). These low-dimensional features are effective in delineating tissue morphology for cross-modality image registration (Extended Data Fig. 6b,d and Supplementary Fig. 4c,d). Images of three selected UMAP dimensions of each two-dimensional (2D) tissue section were converted to a single grayscale image to yield anatomical contrast, which was then registered to its corresponding MRI slice by affine and B-spline registration (Extended Data Fig. 6c and Methods). The obtained transformations could then be applied to all ion images, resulting in the final high-quality volumetric reconstruction of the ion distributions (Fig. 3a,b, Extended Data Fig. 6e,f and Supplementary Fig. 5). To the best of our knowledge, this is the first time that a 3D reconstruction of the biochemical distribution of the brain with the combination of coverage, spatial resolution and chemical detail (for example, >1,000 lipid features, as discussed below) has been generated.

Brain region-specific lipid profiling enabled by 3D MSI

Brain lipids serve as both inter- and intracellular signaling molecules and have important functional roles in the formation of morphologically distinct membranes of the diverse neuronal cells^{30,31}. Previous studies on brain lipids showed that lipid enrichments may be distinct across different brain anatomical structures^{8,32,33}. With our data, of a total of 1,156 lipid features annotated by LIPID MAPS³⁴, 728 matched with an error <3 parts per million (ppm) error were extracted from 3D MSI reconstruction for downstream analysis (Fig. 3). We performed both single-pixel analysis of individual tissue sections through UMAP visualization (Fig. 3c and Supplementary Fig. 5) and supervised classification with gradient-boosted trees. Accurate brain region classification of pixels based on lipid profiles could be achieved with an average classifier area under the curve (AUC) of 0.96 ± 0.02 (Supplementary Fig. 6a and Methods). We also investigated the feature attributions to interpret model decisions and selected the lipids that were contributing most to the anatomical classifications (Methods). Multivariate analysis of the 3D data was performed on the mass spectrum intensity profiles of each brain region (averaged per region, per tissue section). UMAP embeddings showed a preservation of the relative spatial organizations of brain regions (Fig. 3d and see Methods for details on anatomical definitions), as well as differentiation among tissues that are gray-matter-dominant, white-matter-dominant or a mixture of both (others). Our analysis uncovered the anatomically differential lipid compositions of the brain, as shown by the ion images and feature attribution maps of the top-scoring features (Fig. 3e and Supplementary Fig. 7). A near-perfect classification accuracy of anatomical structures was achieved (Supplementary Fig. 6e) using averaged region-specific lipids, indicating that MEISTER robustly uncovered anatomically specific biochemical profiles for the entire tissue volume. The mean intensities of the most discriminative lipid features from the classification model were summarized (Supplementary Fig. 6f). Among these features, we identified highly elevated region-specific lipids by comparing one structure against the rest within the tissue volume; for instance, phosphatidylcholine (PC) O-20:4 and hexosylceramide (HexCer) 40:1;O3 were elevated in the hippocampal region and the corpus callosum, respectively. We further examined how sphingolipids, a lipid category critical to brain development and function, are regionally distributed (Fig. 3f). HexCer exhibited higher levels in the thalamus, brainstem and superior colliculus, which contain a large number of nerve fiber projections responsible for sensation. Ceramide (Cer), abundant in the myelin sheath around the nerve fibers, was found elevated in subcortical white matter such as the corpus callosum and corticofugal pathway, with a similar trend observed for sphingomyelin (SM). In subcortical areas, a higher level of lysophosphatidylcholine

(LPC) was also found, perhaps due to its emphasized role in myelination and neuronal membrane synthesis³⁵.

Multiscale single-cell and tissue imaging data integration

Our high-throughput multiscale tissue MSI and SCMS data enabled integrative analysis to combine their powers for investigating cell-specific biochemical composition in tissues. High-throughput 3D MSI allows brain-wide biochemical characterization (for example, lipid mapping as demonstrated above), but the information at each tissue pixel contains convolved biochemical fingerprints of multiple cell types. MSI with subcellular-level resolution has been demonstrated^{18,19} but with limited tissue coverage. Furthermore, brain cells often are not organized into regular grids but interwind in complex ways. On the other hand, SCMS data acquired from individual cells dissociated from brain tissues provide cell-specific mass spectra but with limited spatial information. To integrate these two types of measurements, we reconstructed SCMS data from 13,566 cells sampled from five brain regions (same anatomical definitions as for tissue data; Methods) and annotated SCMS data using the lipid features from the tissue MSI data (considering the presence of both intracellular and extracellular lipids; Methods, ‘Cross-annotation strategy’), resulting in 344 cross-annotated lipids in single cells. Using these lipid species, we obtained 18 single-cell clusters defined by lipid contents, with each cluster containing mixed cell populations from different anatomical regions (Fig. 4a). We characterized the distinct chemical profiles among cells across different clusters, as well as identified region-specific lipid markers, suggesting both intra- and inter-regional diversity of cellular lipids (Extended Data Fig. 7a). Single cells within the corpus callosum and the striatum contained a higher level of sphingolipids, consistent with observations from the tissue MSI data (Extended Data Fig. 7b). Differential single-cell lipid marker analysis was carried out for other brain regions, showing agreement between cellular- and tissue-level lipidomes (Extended Data Fig. 7c,d).

To integrate tissue and single-cell data and resolve cell-type-specific contributions at each image pixel, we developed a new UoSS fitting strategy exploiting cell-specific chemical dictionaries. Specifically, we performed non-negative matrix factorization (NMF) to extract sets of chemical ‘bases’ that represent chemical variations within each cell cluster (Methods). These bases are ‘dictionaries’ that promote sparsity and parts-based representations that delineate the biochemical components in each cell cluster^{36,37}, ideal for stratifying cellular biochemical signatures. For each of the 18 clusters, we extracted 20 non-negative single-cell dictionary items (Fig. 4b and Supplementary Fig. 8b), and we used a UoSS linear regression model to fit all components to tissue MSI data while forcing model weights to be non-negative (Methods). The weights could then be interpreted as the cell-type-specific contributions and yielded deconvolved cellular features at every tissue pixel (Fig. 4c). By analyzing the weights of individual cell type clusters with respect to brain regions, we identified distinct lipid spatial organizations at the single-cell level. For example, two clusters (0 and 2) were more enriched in cortex and hippocampal regions (Fig. 4c), consistent with the observation in Fig. 4a. Although these two clusters had similar cell numbers from each region, one was more localized to the granular layer of the dentate gyrus and pyramidal layer, whereas the other was more general to cornu ammonis (CA) areas. Similarly, several clusters showed strong spatial contributions toward the thalamus, corpus callosum and striatum (Fig. 4c). From single-cell fitting, we found a moderate to high spectral and spatial correlation of the fitted signal intensity to the original tissue signal intensity, indicating the alignment of SCMS and MSI data (Fig. 4d,e). Note that some lipid features (101 of 344) showed negative correlation between the original tissue image and the UoSS model fit using single-cell dictionaries. These might represent extracellular lipid components, false-positive annotations or modeling errors (Discussion).

To further elucidate the spatial organization of cell subpopulations within a certain anatomical region, we jointly examined

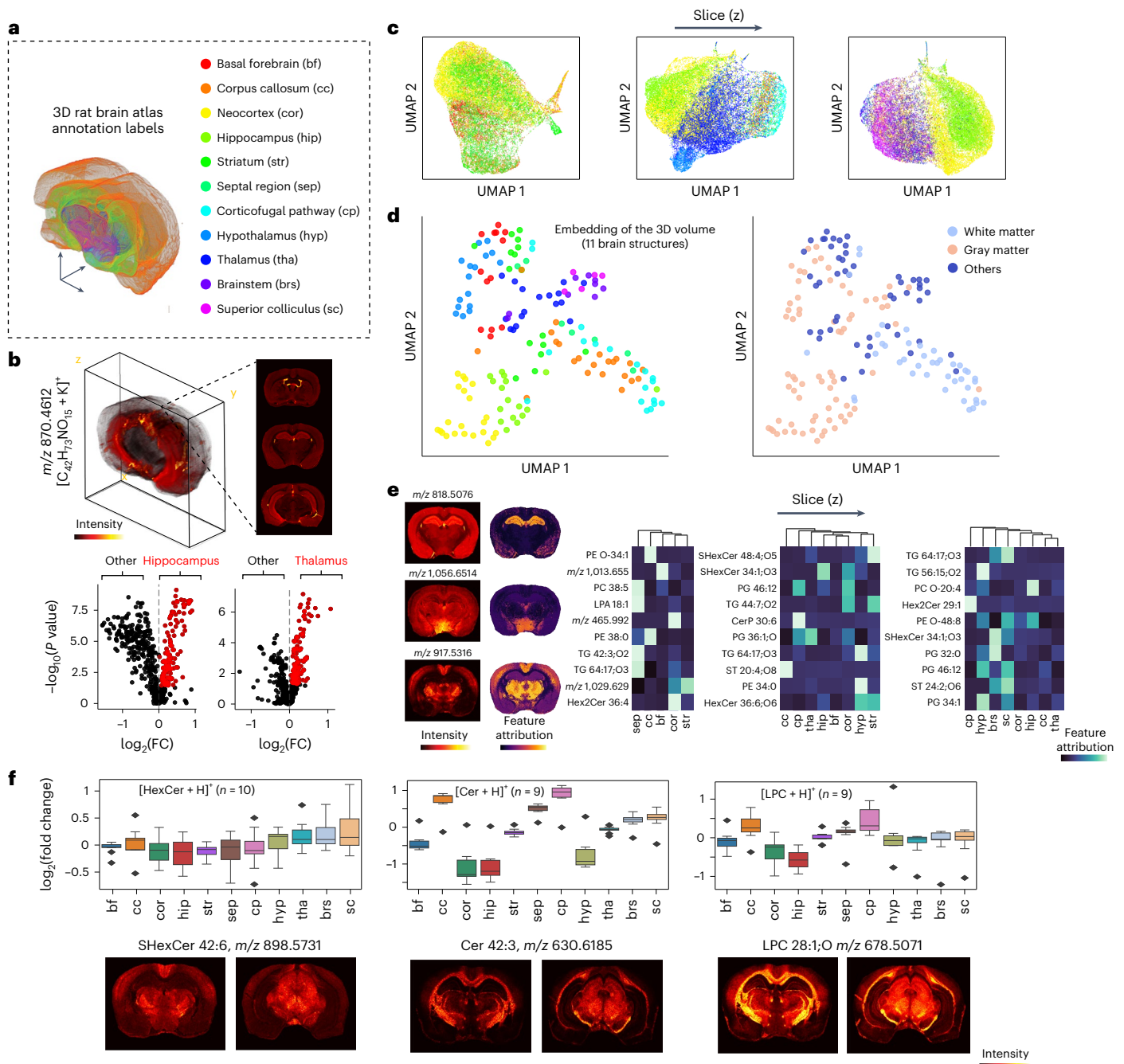


Fig. 3 | Differential lipid profiles across 11 brain structures revealed by high-resolution 3D MSI. a, Atlas annotations colored for 11 brain structures with abbreviations. **b**, Top, the 3D volumetric reconstruction of a representative lipid ion enabled by our reconstruction and analysis methods. Bottom, volcano plots showing differential lipid distributions in hippocampus (left) and thalamus (right). Red dots indicate significantly different lipid features for the brain regions. *P*-values were determined by Wilcoxon rank-sum test (two-sided) and adjusted by Benjamini–Hochberg procedure. FC, fold change. **c**, Low-dimensional UMAP embeddings of the pixel-wise lipid profiles across different brain serial sections, revealing region-specific lipid distributions across a tissue volume. **d**, UMAP analysis of the average lipid profiles from different structures

across the entire 3D volume (each dot represents a region in one tissue section). **e**, Interpreting the machine learning model trained to classify brain regions using lipid profiles reveals attribution of lipids most discriminative for specific brain structures. Examples of top predictive lipid features show distinct spatial distributions and feature attribution maps across the brain (left). **f**, Top, regional distribution of myelination-related lipids (HexCer, Cer and LPC) annotated by protonated ion and quantified by \log_2 fold change. Bottom, representative ion images corresponding to the lipids described in the plots on top. *n* is the number of lipid features for each lipid class. Data in boxplots are shown as median values (center) with the interquartile range (box), and the whiskers extend to 1.5 times the interquartile range.

hippocampus-only SCMS and tissue MSI data (Fig. 5). A total of 2,692 cells (103 annotated lipid features) were analyzed with MSI data through the joint fitting of eight identified cell clusters (with dictionaries estimated using NMF). Single-cell lipids displayed heterogeneous distributions within the hippocampus, with unique lipid markers

(Fig. 5a and Supplementary Fig. 9). The fitted contributions of single-cell dictionaries suggest different spatial organizations of hippocampal cell subpopulations (Fig. 5b,c). Large model weights were found in the dentate gyrus and CA3 for cluster 0, the granular layer of CA1 for cluster 1 and the molecular layer for cluster 5,

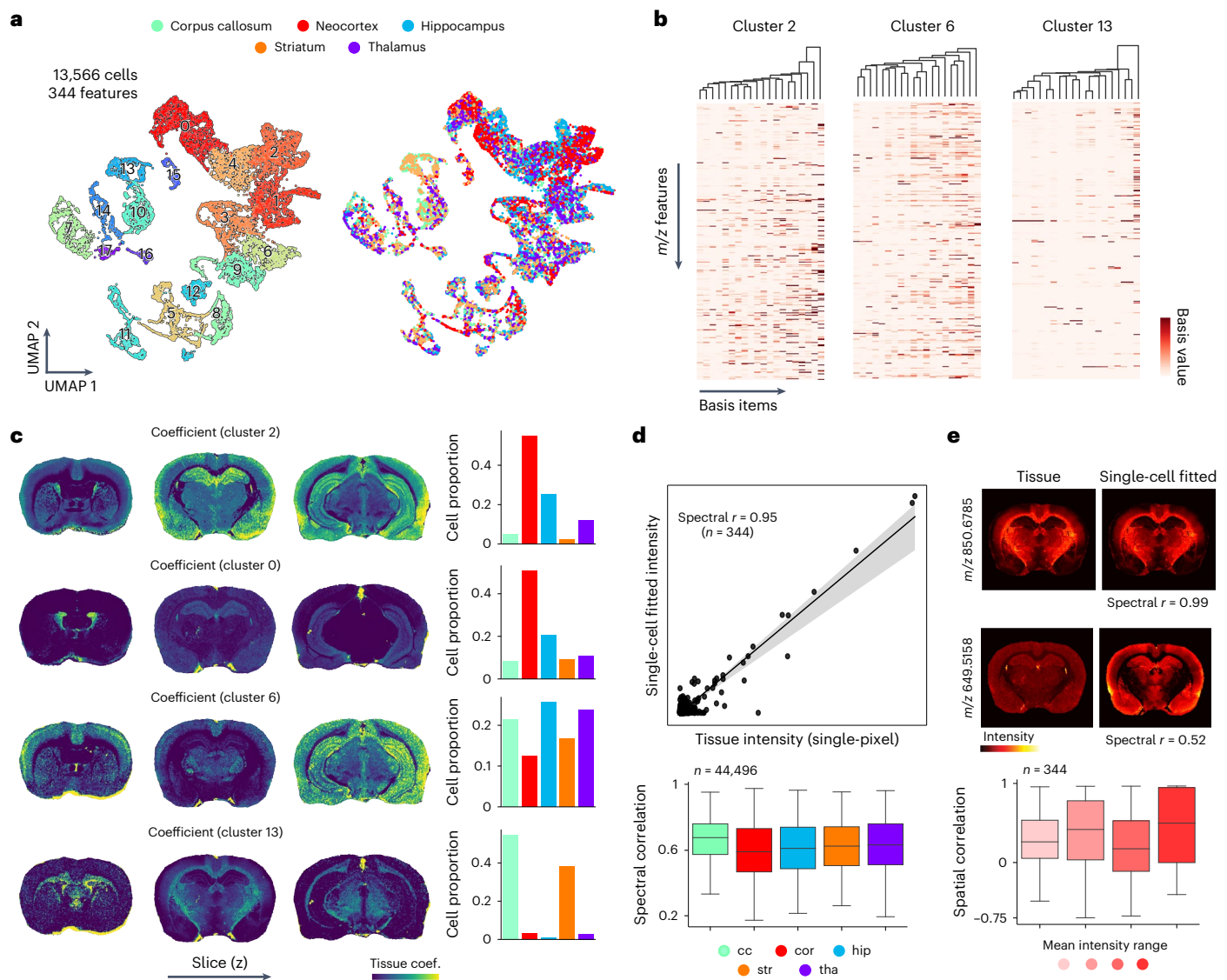


Fig. 4 | Joint visualization and analysis of tissue MSI and SCMS data. **a**, A total of 13,566 cells with 344 cross-annotated lipid features were subjected to UMAP and clustering analysis. As a result, 18 cell clusters were identified (left) with each cluster containing cells dissociated from different brain regions (right; annotated on the top). **b**, Cell-specific chemical dictionaries (lipids) can be extracted, for example, from clusters 2, 6 and 13 shown here (20 basis elements were obtained in each dictionary). **c–e**, Results of our proposed UoSS fitting mapping cell-type-specific dictionaries (obtained from SCMS data) to the tissue MSI data. **c**, Estimated spatially dependent contributions of different cell clusters across the brain (UoSS coefficients for the cell-type-specific dictionaries). Distinct cell compositions can be resolved for different regions. Each row shows results for mapping contributions of one cell cluster to individual pixels in different tissue sections (left three columns), and the percentage compositions of regional cell populations (where they are from; right column) for each cluster.

coef., coefficient. **d**, Signal intensities from tissue pixels ($n = 44,496$ pixels) obtained with tissue MSI are well correlated with the model fitted values (mean correlation coefficients > 0.6 for all regions, bottom box plot). The shaded band (top scatter plot) is the 95% confidence interval for the linear fit. Data in boxplots are shown as median values (center) with the interquartile range (box), and the whiskers extend to 1.5 times the interquartile range. **e**, The top four images illustrate lipids with excellent and moderate consistency between the actual tissue images and the single-cell-dictionary fit. The consistency was evaluated for lipid features ($n = 344$) by spatial correlation across different mean signal intensity ranges. Lipids at high mean intensity have slightly better fitting results than ones at lower intensity (lower boxplots). In total, 101 (of the total 344) lipid features have negative spatial correlation, a result of less accurate fit. Brain region abbreviations: cc, corpus callosum; cor, cortex; hip, hippocampus; str, striatum; tha, thalamus.

approximating the morphological structure of the hippocampus (Fig. 5b). We further analyzed the extracted dictionary items, which showed strong correlation with the lipid fold change (Fig. 5d), serving as indicators of the cluster-specific lipid signatures. Features were then selected on the basis of the magnitude of the averaged basis values, which are a good proxy for lipid specificity to cell clusters. For example, consider lysophosphatidylethanolamine (LPE) O-(16:0) and phosphatidylglycerol (PG)(48:8) (indicated in dictionary items for cluster 1; Fig. 5e). The corresponding tissue distributions for

these two lipids showed agreement with the fitted model weights from cluster 1 (Fig. 5b,f).

To demonstrate the applicability of our joint analysis approach to other tissue types, SCMS data of 13,739 rat pancreatic cells consisting of islets, vasculature and acinar cells (Methods) were acquired and reconstructed. A total of 428 features were annotated using pancreas tissue MSI data. Using the cross-annotated features, we obtained ten single-cell clusters, each containing relatively uniform cell populations (Extended Data Fig. 8a,b). Fitting the extracted single-cell dictionaries

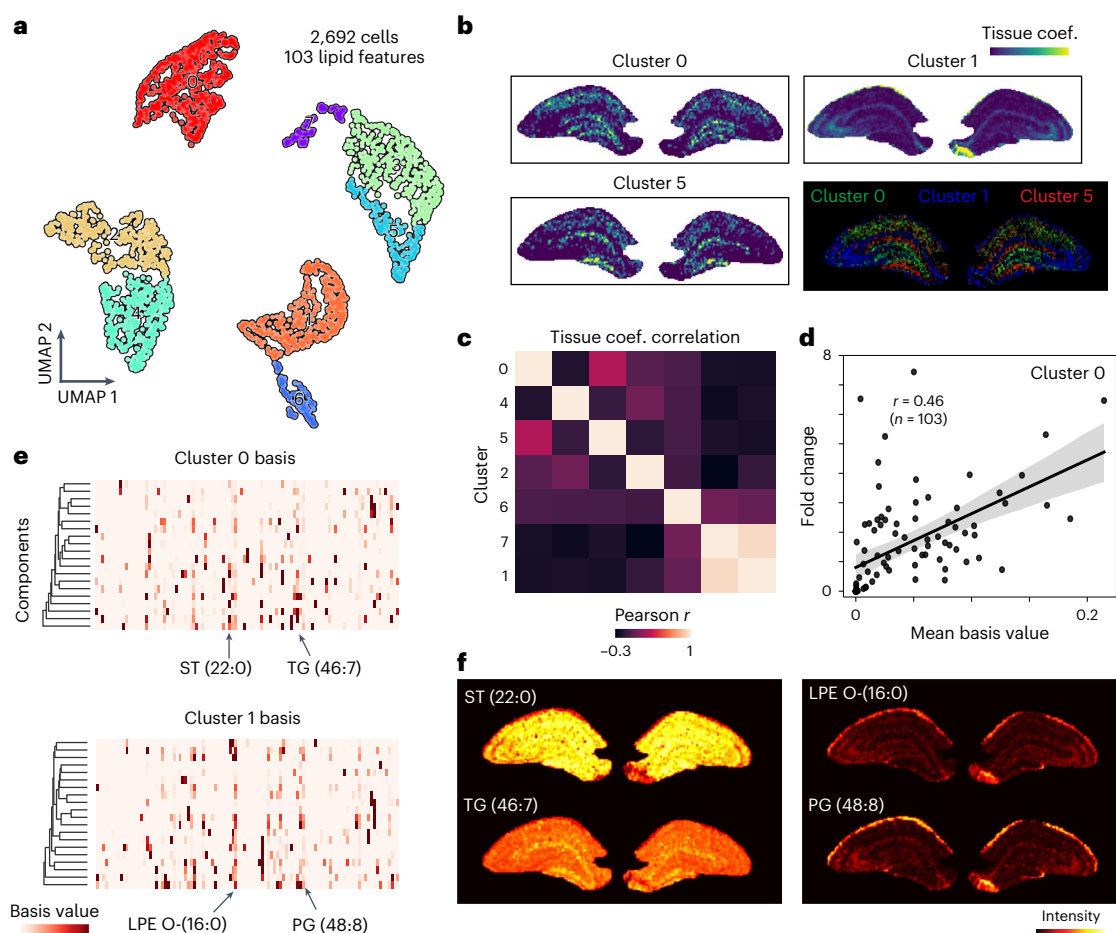


Fig. 5 | Integrative analysis of hippocampal tissue MSI and single-cell data.

a, A total of 2,692 hippocampal cells were probed and data were subjected to UMAP and clustering analysis, producing eight chemically unique single-cell clusters in terms of lipid profiles. **b**, Maps of resolved cell type contributions from fitting the cell-specific chemical dictionaries to tissue data. **c**, Correlation matrix for weights obtained from eight clusters. **d**, Basis values averaged on 20 dictionary items versus fold change for lipid features in cluster 0. The shaded

band is the 95% confidence interval for the linear fit between basis values and fold change. **e**, Chemical dictionaries extracted for cluster 0 and cluster 1, with arrows indicating dominant cluster-specific lipid features. **f**, Ion images showing the distributions of lipids that are more cluster-specific within the hippocampal region. Sterol (ST) (22:0) and triglyceride (TG) (46:7) were selected on the basis of cluster 0 basis values, whereas LPE O-(16:0) and PG (48:8) were determined on the basis of cluster 1 basis values.

to tissue data (Extended Data Fig. 8c), we were able to map spatially dependent cell-specific contributions and resolve tissue organizations of cell populations (Extended Data Fig. 8d). For example, we observed islet populations with distinct spatial localization within the islet region (clusters 0 and 2), likely corresponding to subpopulations of islet cells.

Discussion

We demonstrated integrative 3D tissue and single-cell biochemical mapping of the brain at a large scale using MEISTER. This is enabled by a synergy of unique experimental capabilities and innovations in computational aspects including deep-learning-based reconstruction, image registration and spatially resolved cell-specific dictionary learning and fitting. While the power of deep learning has been illustrated in various imaging modalities including MSI^{38–40}, our method exploits unique signal characteristics in FTMS data and a special network design. Instead of training a deep neural network to generate high-resolution mass spectra from a low-resolution counterpart or interpolating missing pixel values directly⁴¹, we jointly learn low-dimensional embeddings of the high-dimensional data and train a regression network to predict these embeddings from reduced transients, providing a strong generalization ability for both tissue and SCMS data. We validated that training can be achieved using different sections or animals, and it works well for new datasets. Through the unique capability of MEISTER, we

resolved thousands of brain lipid features over millions of pixels across a 3D volume and large cell population, while substantially reducing the data collection time. Our method should be readily adaptable to different types of molecules in addition to lipids and peptides (for example, small metabolites and proteins) and other organ systems.

MEISTER relies on several alignment steps for knowledge extraction from integrating high-dimensional MSI and SCMS data. First, while alignment between MSI and other imaging modalities has been performed^{42–47}, we chose to register MSI data to brain MRI that was acquired from the intact brain (without deformation) and offered a coherent 3D volumetric ‘atlas’ for registration. We realize that the majority of prior approaches emphasized on extracting feature images from individual 2D sections, which can lead to incoherent image registration across different serial sections in large 3D MSI datasets. Our pixel-wise parametric UMAP strategy generated feature images with a similar contrast to MRI images (Extended Data Fig. 6) and provided structurally informative features across the entire 3D dataset for easier registration. The 3D registration capability can also enable richer analysis leveraging both in vivo MRI and ex vivo MSI. Second, we align SCMS and MSI data through cross-annotation to facilitate integrative analysis. Recent progress has made false discovery rate (FDR)-controlled metabolite annotation for MSI possible^{48,49}. However, it is difficult to leverage such methods to annotate image-guided SCMS

data due to lack of spatial information, which is a crucial statistical consideration in the annotation algorithm. Rather, we leverage the biochemical information that mutually exists in tissue and single cells, which are searched against the ‘tissue feature database’. This approach boosts confidence in selecting biologically viable features, as retaining mutual features can minimize the inclusion of the experiment- or sample-specific artifacts. Another advantage for cross-annotation is to consider only molecules that are present in the intracellular spaces for more accurate single-cell-to-tissue mapping. Meanwhile, we think it is possible to extend our approach to look at changes in extracellular spaces (for example, in different diseases) by interrogating the fitting residual, as well as lipid species filtered out by cross-annotation. For the UoSS fitting, modeling errors may occur due to falsely cross-annotated features from the tissue data caused by mass shifts and an inappropriate ppm window or due to nonlinearities between MSI and SCMS data. These may have contributed to the poor lipid fits with negative correlation shown in Fig. 4e. A close examination of those poorly fitted lipid components may lead to important insights into potential directions for improvement in future research.

Computational methods for integrating single-cell sequencing and spatial transcriptomics (ST) data have been explored⁵⁰, including deconvolution of cell type fractions^{51–53}, joint clustering for mapping single-cell transcriptomics to ST data⁵⁴ and estimating the number of cells per ST spot⁵⁵. Our work is the first attempt for similar cross-scale integration for large metabolomics and lipidomics data. Our UoSS regression model is also distinct in that it does not assume only a ‘reference signature’ but a more general mathematical representation of the ensemble chemical profiles for each cell type, capturing the intrinsic variations within each cell cluster. We used all the cross-annotated lipid species making the approach ‘unbiased’. In addition to identifying brain region-specific lipid variability, clustering of single cells suggests that a continuity of lipid-defined cell type diversity exists across brain regions (Fig. 4a). Similar observations have been made by transcriptomics of single cells from various brain regions, supporting the idea that many cell types are shared between brain regions^{56,57}. We demonstrate a proof of concept for linking single-cell and spatial organizations of lipids, paving the way to build biochemical cartography of tissue and primary cells.

Previous studies have used liquid chromatography–tandem MS (LC–MS²) measurements to study differential lipid contents within different brain regions and cell types^{8,32}. Comparing the LC–MS²-based shotgun lipidomics and our MS imaging-based lipid profiling, we found agreements in region-specific distributions of major lipid classes, including HexCer, Cer and SM that are more enriched in regions such as the corpus callosum with higher myelin content³² (Fig. 3d,e). Our multiscale imaging-based approach offers not only the capability of resolving hundreds of lipid molecules, but also a new tool for understanding spatial–biochemical tissue architecture with cellular specificity, transforming how we study brain chemistry just as how spatial transcriptomics transforms the determination of gene expression. We envision future endeavors on creating multiscale biochemical atlases, with increasingly powerful profiling technology for metabolites, lipids, peptides and proteins, as well as integrative analysis with other omics data.

Online content

Any methods, additional references, Nature Portfolio reporting summaries, source data, extended data, supplementary information, acknowledgements, peer review information; details of author contributions and competing interests; and statements of data and code availability are available at <https://doi.org/10.1038/s41592-024-02171-3>.

References

- Hawrylycz, M. J. et al. An anatomically comprehensive atlas of the adult human brain transcriptome. *Nature* **489**, 391–399 (2012).
- Darmanis, S. et al. A survey of human brain transcriptome diversity at the single cell level. *Proc. Natl Acad. Sci. USA* **112**, 7285–7290 (2015).
- Rao, A., Barkley, D., França, G. S. & Yanai, I. Exploring tissue architecture using spatial transcriptomics. *Nature* **596**, 211–220 (2021).
- Tsui-Pierchala, B. A., Encinas, M., Milbrandt, J. & Johnson, E. M. Lipid rafts in neuronal signaling and function. *Trends Neurosci.* **25**, 412–417 (2002).
- Allaman, I., Bélanger, M. & Magistretti, P. J. Astrocyte–neuron metabolic relationships: for better and for worse. *Trends Neurosci.* **34**, 76–87 (2011).
- Iwata, R. et al. Mitochondria metabolism sets the species-specific tempo of neuronal development. *Science* **379**, eabn4705 (2023).
- Lynch, G. & Baudry, M. The biochemistry of memory: a new and specific hypothesis. *Science* **224**, 1057–1063 (1984).
- Ding, J. et al. A metabolome atlas of the aging mouse brain. *Nat. Commun.* **12**, 6021 (2021).
- Mutlu, A. S., Duffy, J. & Wang, M. C. Lipid metabolism and lipid signals in aging and longevity. *Dev. Cell* **56**, 1394–1407 (2021).
- Hou, Y. et al. Ageing as a risk factor for neurodegenerative disease. *Nat. Rev. Neurol.* **15**, 565–581 (2019).
- Eisenstein, M. Seven technologies to watch in 2023. *Nature* **613**, 794–797 (2023).
- Piehowski, P. D. et al. Automated mass spectrometry imaging of over 2000 proteins from tissue sections at 100- μ m spatial resolution. *Nat. Commun.* **11**, 8 (2020).
- Guo, G. et al. Automated annotation and visualisation of high-resolution spatial proteomic mass spectrometry imaging data using HIT-MAP. *Nat. Commun.* **12**, 3241 (2021).
- Mund, A. et al. Deep Visual Proteomics defines single-cell identity and heterogeneity. *Nat. Biotechnol.* **40**, 1231–1240 (2022).
- Brunner, A.-D. et al. Ultra-high sensitivity mass spectrometry quantifies single-cell proteome changes upon perturbation. *Mol. Syst. Biol.* **18**, e10798 (2022).
- Budnik, B., Levy, E., Harmange, G. & Slavov, N. SCoPE-MS: mass spectrometry of single mammalian cells quantifies proteome heterogeneity during cell differentiation. *Genome Biol.* **19**, 161 (2018).
- Specht, H. et al. Single-cell proteomic and transcriptomic analysis of macrophage heterogeneity using SCoPE2. *Genome Biol.* **22**, 50 (2021).
- Passarelli, M. K. et al. The 3D OrbiSIMS—label-free metabolic imaging with subcellular lateral resolution and high mass-resolving power. *Nat. Methods* **14**, 1175–1183 (2017).
- Kompauer, M., Heiles, S. & Spengler, B. Atmospheric pressure MALDI mass spectrometry imaging of tissues and cells at 1.4- μ m lateral resolution. *Nat. Methods* **14**, 90–96 (2017).
- Rappez, L. et al. SpaceM reveals metabolic states of single cells. *Nat. Methods* **18**, 799–805 (2021).
- Capolupo, L. et al. Sphingolipids control dermal fibroblast heterogeneity. *Science* **376**, eabh1623 (2022).
- Rubakhin, S. S., Romanova, E. V., Nemes, P. & Sweedler, J. V. Profiling metabolites and peptides in single cells. *Nat. Methods* **8**, S20–S29 (2011).
- Comi, T. J., Neumann, E. K., Do, T. D. & Sweedler, J. V. microMS: a Python platform for image-guided mass spectrometry profiling. *J. Am. Soc. Mass Spectrom.* **28**, 1919–1928 (2017).
- Xie, Y. R., Castro, D. C., Lam, F. & Sweedler, J. V. Accelerating Fourier transform-ion cyclotron resonance mass spectrometry imaging using a subspace approach. *J. Am. Soc. Mass Spectrom.* **31**, 2338–2347 (2020).
- Xie, Y. R., Castro, D. C., Rubakhin, S. S., Sweedler, J. V. & Lam, F. Enhancing the throughput of FT mass spectrometry imaging using joint compressed sensing and subspace modeling. *Anal. Chem.* **94**, 5335–5343 (2022).

26. Mallah, K. et al. Lipid changes associated with traumatic brain injury revealed by 3D MALDI-MSI. *Anal. Chem.* **90**, 10568–10576 (2018).
27. Randall, E. C. et al. Localized metabolomic gradients in patient-derived xenograft models of glioblastoma. *Cancer Res.* **80**, 1258–1267 (2020).
28. Papp, E. A., Leergaard, T. B., Calabrese, E., Johnson, G. A. & Bjaalie, J. G. Waxholm Space atlas of the Sprague Dawley rat brain. *NeuroImage* **97**, 374–386 (2014).
29. Sainburg, T., McInnes, L. & Gentner, T. Q. Parametric UMAP embeddings for representation and semisupervised learning. *Neural Comput.* **33**, 2881–2907 (2021).
30. Bazinet, R. P. & Layé, S. Polyunsaturated fatty acids and their metabolites in brain function and disease. *Nat. Rev. Neurosci.* **15**, 771–785 (2014).
31. Piomelli, D. & Sasso, O. Peripheral gating of pain signals by endogenous lipid mediators. *Nat. Neurosci.* **17**, 164–174 (2014).
32. Fitzner, D. et al. Cell-type- and brain-region-resolved mouse brain lipidome. *Cell Rep.* **32**, 108132 (2020).
33. Li, T. et al. Ion mobility-based sterolomics reveals spatially and temporally distinctive sterol lipids in the mouse brain. *Nat. Commun.* **12**, 4343 (2021).
34. Fahy, E., Sud, M., Cotter, D. & Subramaniam, S. LIPID MAPS online tools for lipid research. *Nucleic Acids Res.* **35**, W606–W612 (2007).
35. Tan, S. T., Ramesh, T., Toh, X. R. & Nguyen, L. N. Emerging roles of lysophospholipids in health and disease. *Prog. Lipid Res.* **80**, 101068 (2020).
36. Devarajan, K. Nonnegative matrix factorization: an analytical and interpretive tool in computational biology. *PLoS Comput. Biol.* **4**, e1000029 (2008).
37. Townes, F. W. & Engelhardt, B. E. Nonnegative spatial factorization applied to spatial genomics. *Nat. Methods* **20**, 229–238 (2023).
38. Li, X. et al. Reinforcing neuron extraction and spike inference in calcium imaging using deep self-supervised denoising. *Nat. Methods* **18**, 1395–1400 (2021).
39. Lecoq, J. et al. Removing independent noise in systems neuroscience data using DeepInterpolation. *Nat. Methods* **18**, 1401–1408 (2021).
40. Abdelmoula, W. M. et al. Peak learning of mass spectrometry imaging data using artificial neural networks. *Nat. Commun.* **12**, 5544 (2021).
41. Hu, H. et al. High-throughput mass spectrometry imaging with dynamic sparse sampling. *ACS Meas. Sci. Au* **2**, 466–474 (2022).
42. Sinha, T. K. et al. Integrating spatially resolved three-dimensional MALDI IMS with in vivo magnetic resonance imaging. *Nat. Methods* **5**, 57–59 (2008).
43. Van de Plas, R., Yang, J., Spraggins, J. & Caprioli, R. M. Image fusion of mass spectrometry and microscopy: a multimodality paradigm for molecular tissue mapping. *Nat. Methods* **12**, 366–372 (2015).
44. Vollnhals, F. et al. Correlative microscopy combining secondary ion mass spectrometry and electron microscopy: comparison of intensity–hue–saturation and Laplacian pyramid methods for image fusion. *Anal. Chem.* **89**, 10702–10710 (2017).
45. Patterson, N. H., Tuck, M., Van de Plas, R. & Caprioli, R. M. Advanced registration and analysis of MALDI imaging mass spectrometry measurements through autofluorescence microscopy. *Anal. Chem.* **90**, 12395–12403 (2018).
46. Neumann, E. K. et al. Multimodal chemical analysis of the brain by high mass resolution mass spectrometry and infrared spectroscopic imaging. *Anal. Chem.* **90**, 11572–11580 (2018).
47. Porta Siegel, T. et al. Mass spectrometry imaging and integration with other imaging modalities for greater molecular understanding of biological tissues. *Mol. Imaging Biol.* **20**, 888–901 (2018).
48. Palmer, A. et al. FDR-controlled metabolite annotation for high-resolution imaging mass spectrometry. *Nat. Methods* **14**, 57–60 (2017).
49. Alexandrov, T. et al. METASPACE: a community-populated knowledge base of spatial metabolomes in health and disease. Preprint at *bioRxiv* <https://doi.org/10.1101/539478> (2019).
50. Longo, S. K., Guo, M. G., Ji, A. L. & Khavari, P. A. Integrating single-cell and spatial transcriptomics to elucidate intercellular tissue dynamics. *Nat. Rev. Genet.* **22**, 627–644 (2021).
51. Li, B. et al. Benchmarking spatial and single-cell transcriptomics integration methods for transcript distribution prediction and cell type deconvolution. *Nat. Methods* **19**, 662–670 (2022).
52. Nitzan, M., Karaiskos, N., Friedman, N. & Rajewsky, N. Gene expression cartography. *Nature* **576**, 132–137 (2019).
53. Biancalani, T. et al. Deep learning and alignment of spatially resolved single-cell transcriptomes with Tangram. *Nat. Methods* **18**, 1352–1362 (2021).
54. Korsunsky, I. et al. Fast, sensitive and accurate integration of single-cell data with Harmony. *Nat. Methods* **16**, 1289–1296 (2019).
55. Vahid, M. R. et al. High-resolution alignment of single-cell and spatial transcriptomes with CytoSPACE. *Nat. Biotechnol.* **41**, 1543–1548 (2023).
56. Yao, Z. et al. A taxonomy of transcriptomic cell types across the isocortex and hippocampal formation. *Cell* **184**, 3222–3241 (2021).
57. Yao, Z. et al. A high-resolution transcriptomic and spatial atlas of cell types in the whole mouse brain. *Nature* **624**, 317–332 (2023).

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2024

Methods

Experimental details

Animals. The male Sprague-Dawley outbred rats (*Rattus norvegicus*) used in this study were sourced from Inotivco (www.inotivco.com). These rats were provided with ad libitum access to food and housed on a 12-h light cycle. All animal euthanasia procedures were carried out in strict accordance with the guidelines set forth by the Illinois Institutional Animal Care and Use Committee, as well as the federal and ARRIVE guidelines to ensure the humane care and treatment of animals.

Tissue dissociation and preparation of single cells. A total of three 2.0- to 2.5-month-old male rats were used for brain tissue isolation. Each isolated tissue region was individually treated with a papain dissociation system (Worthington Biochemical) and incubated for 120 min at 34 °C with oxygenation. The treated tissue regions were then mechanically dissociated in ice-cold modified Gey's balanced salt solution (mGBSS) containing (in mM) 1.5 CaCl₂, 5 KCl, 0.2 KH₂PO₄, 11 MgCl₂, 0.3 MgSO₄, 138 NaCl, 28 NaHCO₃, 0.8 Na₂HPO₄ and 25 HEPES, pH 7.2, and supplemented with 0.08% paraformaldehyde to stabilize cells against damage during dissociation and other methodological steps. A solution of 80% glycerol and 20% mGBSS was added to a final glycerol concentration of 40% (v/v). The cells were stained with Hoechst 33342 (0.1 μg ml⁻¹ in mGBSS), and a 30-μl aliquot of cell suspension was plated onto an indium tin oxide (ITO)-coated unpolished float glass slide, $R_s = 70\text{--}100\ \Omega$ (Delta Technologies). After -16 h, glycerol was aspirated off the dissociated cells before rinsing with 150 mM ammonium acetate. Each slide held three biological replicates of each brain region, placed in discrete but random locations on the ITO-coated glass slide to mitigate batch and spatial-dependent artifacts.

Islet isolation was performed as previously described^{58,59} with some modifications. Briefly, the pancreas was surgically removed and treated with lyberase. Islets were manually collected by mechanical dissociation of tissue using a micropipette under visual control using an inverted microscope. Islets, acinar tissue and vasculature regions were incubated for 20 min at 37 °C in Trypsin LE solution before mechanical dissociation into single cells and deposition onto ITO-coated glass slides.

Tissue sectioning. Coronal and sagittal brain slices were obtained from the rats in this study. The entire rat brain was quickly removed and flash-frozen after decapitation, before being sectioned. Sagittal brain slices were prepared at a temperature of -20 °C and sliced into 16-μm-thick tissue sections using a cryostat-microtome (3050 S, Leica Biosystems). The tissue slices were then thaw-mounted onto ITO-coated glass slides for MALDI matrix application. The pancreata from three male rats were surgically removed and immediately frozen. Six 16-μm-thick adjacent sections were cut from the frozen pancreata using a cryostat-microtome and similarly thaw-mounted onto ITO-coated glass slides for matrix application.

Matrix application. The MALDI matrix 2,5-dihydroxybenzoic acid (DHB) was prepared to a concentration of 30 mg ml⁻¹ in 70% methanol for brain samples. For pancreatic samples, DHB was prepared at a concentration of 10 mg ml⁻¹ in 50% ethanol. The matrix was applied using an HTX-M5Sprayer (HTX Technologies), with a spray spacing of 2.5 mm at a temperature of 75 °C using a flow rate of 100 μl min⁻¹. The distance of the sprayer nozzle was 50 mm from the sample, and a spray pressure of 10 psi with a spray nozzle motion velocity of 1,200 mm min⁻¹ was used. Four passes were used to apply the MALDI matrix.

Image-guided SCMS analysis. The brightfield and fluorescence microscopy images were obtained using an Axio Imager M2 (Zeiss) equipped with an AxioCam ICc 5 camera and a ×63 camera adapter. For transmitted light, a visible light (Vis) light-emitting diode (LED) lamp was used, while, for fluorescence, an X-cite Series 120 Q mercury

lamp (Lumen Dynamics) was used. The imaging was performed using DAPI (excitation 335–383 nm; emission 420–470 nm) dichroic filter cubes. The images were acquired in mosaic mode with a ×10 objective and 10% tile overlap. The resulting tiles were stitched together before being exported in TIFF file format using ZEN 2.0 Pro edition (Zeiss) software. The single-cell coordinates, geometry files and an Excel file required for the target automation function on fmsControl (v.2.1.0, Bruker) were obtained using microMS, as described previously²³. To ensure data quality, cells were filtered from lists of analyzed structures on the basis of their distance from each other, with cells closer than 200 μm being removed, and on the basis of their size, with any free nuclei resulting from cell lysis being removed. High-throughput single-cell analysis was performed using a Solarix 7 T FT-ICR mass spectrometer (Bruker), with a mass window of 50–1,000 *m/z* (rat brain) or 150–6,000 *m/z* (rat pancreas). MALDI mass spectra were acquired in positive mode using a Smartbeam-II UV laser in 'ultra' mode, which produces a 100-μm-diameter laser-spot size. Each MALDI acquisition comprised one ICR accumulation, consisting of 150 or 500 laser shots, for brain and pancreatic samples, respectively, at a frequency of 1,000 Hz.

Signal modeling. A transient can be modeled as a temporal signal that contains many frequencies corresponding to different ions, following the generic signal model proposed by Marshall^{60,61}:

$$s(t) = \sum_i A_i \cdot \exp\left(-\frac{t}{\tau_i}\right) \cos(\omega_i t) + \varepsilon(t).$$

For ion *i*, constant A_i represents the initial signal amplitude, τ_i is the decay rate of the excited ICR signal due to ion collisions, ω_i is the ion cyclotron frequency and ε is the independent noise. The theoretical mass resolution is calculated as:

$$\frac{m}{\Delta m} = \frac{1.274 \times 10^7 z B_0 T_{\text{aqn}}}{m}$$

where B_0 is the magnetic field strength and T_{aqn} is the transient acquisition time. Given a fixed B_0 , the theoretical mass resolution is, thus, proportional to T_{aqn} ; a certain T_{aqn} is required for a target high mass resolution. We used the described signal model for the MSI data simulation.

Simulation of MSI data. We followed a previously described procedure to simulate the MSI data²⁵. Briefly, transients were generated through the generic signal model discussed above with a list of 30 chemical formulas. Frequencies were reverse-calculated for all possible ions including H⁺, Na⁺ and K⁺, adducts were assigned to each formula and theoretical isotopic distributions were calculated using the Python version of the BRAIN algorithm⁶². Allen Brain Atlas (ABA) mouse brain annotation was used as the spatial reference⁶³ to generate eight pseudo-tissue regions with different combinations of chemical formulas. All transients were simulated for 262,144 temporal data points in a total of 26,497 pixels. Independent Gaussian noise was added to each simulated transient.

Model design for MEISTER

Reconstruction model. The signal reconstruction model consists of three parts: (1) an encoder network encoding input high-resolution transient signals into lower-dimensional latent features; (2) a regression network transforming corresponding low-resolution signals to their latent features; and (3) a decoder network decoding the estimated low-dimensional latents back to high-resolution signals. We use a DAE architecture to learn low-dimensional features directly from raw high-resolution transient signals for both tissue and single-cell measurements. Each transient signal $s(r, t)$ is sampled with a specific temporal sampling rate, with $t = \{t_1, t_2, \dots, t_{N_r}\}$, where N_r is the number of

discrete time points and the duration T for a defined mass resolution, and $n = 1, 2, \dots, N_r$, where N_r corresponds to the number of pixels in MSI or number of cells in SCMS data. Denoting $\mathbf{S} = \{s(r_n, t_i)\}_{n,i=1}^{N_r, N_r}$ as the ensemble of training data, our objective is to train the network to encode \mathbf{S} into a set of low-dimensional features and produce reconstructions $\hat{\mathbf{S}} = \{\hat{s}(r_n, t_i)\}_{n,i=1}^{N_r, N_r}$ that are as close to \mathbf{S} as possible. Specifically, our network can be described mathematically as:

$$\mathbf{Z} = \mathbf{E}(\mathbf{S}; \mathbf{W}_E, \mathbf{B}_E)$$

$$\hat{\mathbf{S}} = \mathbf{D}(\mathbf{Z}; \mathbf{W}_D, \mathbf{B}_D)$$

where \mathbf{Z} represents a 32-dimensional latent vector encoding \mathbf{S} , $\mathbf{E}(\cdot)$ and $\mathbf{D}(\cdot)$ denote the encoder and decoder functions, respectively, each containing three fully connected layers with 512, 256 and 64 neurons (symmetric design). Denoting the whole network as $\phi(\cdot; \Theta)$ (combining encoder and decoder) with $\Theta = [\mathbf{W}_E, \mathbf{B}_E, \mathbf{W}_D, \mathbf{B}_D]$ containing all the network parameters, the mean squared error (MSE) loss was minimized during training:

$$\Theta = \arg \min_{\Theta} \|\mathbf{S} - \phi(\mathbf{S}; \Theta)\|_2^2 + \lambda \|\Theta\|_2^2$$

We then trained a regressor network $\mathbf{R}(\cdot)$ to map the low-resolution measurements $\mathbf{S}' = \{s'(r_n, t_j)\}_{n,j=1}^{N_r, N_r}$ with N_r transients ($N_r \gg N$) and first N_r temporal points corresponding to a shorter acquisition duration T' ($T' \ll T$), to the latent features \mathbf{Z}' :

$$\mathbf{Z}' = \mathbf{R}(\mathbf{S}'; \mathbf{W}_R, \mathbf{B}_R)$$

which can be decoded into full-resolution transients. Denoting all the regressor network parameters as $\Theta_R = [\mathbf{W}_R, \mathbf{B}_R]$, the model was trained by minimizing the MSE between \mathbf{Z} (encoded from \mathbf{S}) and the regressor output \mathbf{Z}' :

$$\Theta_R = \arg \min_{\Theta_R} \|\mathbf{Z} - \mathbf{R}(\mathbf{S}'; \Theta_R)\|_2^2 + \|\Theta_R\|_2^2$$

Low-resolution measurements \mathbf{S}' can then be transformed into high-resolution data by:

$$\hat{\mathbf{S}} = \mathbf{D}(\mathbf{R}(\mathbf{S}'; \Theta_R); \mathbf{W}_D, \mathbf{B}_D)$$

Evaluation on simulated and experimental MSI data. First, we validated MEISTER's signal reconstruction performance on simulated MSI data. We trained MEISTER on 2,000 randomly sampled pixels containing noisy transients (simulation data described above), and we performed reconstruction on all pixels containing reduced noisy transients (taking the first 10,000 temporal points). We compared the reconstructed data against the ground-truth high-resolution data without Gaussian noise. All sets of transients (clean, noisy, noisy reduced and reconstructed) were transformed into mass spectra and converted to peak intensity lists, which consist of peak centroids identified from the ground-truth spectra. Peak and spatial correlation scores against the ground truth were computed as the Pearson correlation coefficients between each peak list and ion image pairs. Encoded features of the latent space were extracted from the bottleneck layer and subjected to UMAP for visualization.

For the experimental data, we trained the networks using a set of high-resolution data (1 million temporal points) acquired on rat sagittal and coronal brain sections and the corresponding low-resolution data (64,000 temporal points). Reconstruction was performed on the 160- μm , 1-mm and 2-mm sections away from the training sections to validate the model's generalizability across tissue volume. High-resolution data were acquired from these tissue sections (serving as the reference) and reduced to 64,000 temporal points as the input to the MEISTER

reconstruction model. Reduced (zero-padded), reconstructed and reference transients were then transformed into mass spectra and converted to peak intensity lists. For each tissue section, the peak centroids were determined on the average mass spectra obtained from the high-resolution reference data. Peak and spatial correlation scores were calculated the same way as for the simulation, but against the high-resolution reference. SNRs were defined as the ratios between the signal intensity and the standard deviation of the noise, which was obtained over a spectral region without apparent signals. k -means clustering was performed for different reconstructions with $k = 6$. The components and scores for the first five principal components (PCs) were compared between reconstructed and reference data.

MEISTER for 3D MSI. For 3D MSI of rat coronal sections, training data (transients collected for 1 million temporal points) were collected on three tissue sections with a total of 124,370 pixels. For rat sagittal sections, data from two tissue sections in a total of 105,954 pixels were used for the model training. Training sections were roughly 2 mm apart to ensure the coverage of diverse tissue types. The autoencoders were trained for 20 epochs, and the regressors were trained for 50 epochs. A batch size of 128 and the Adam optimizer were applied to train both networks. We then acquired low-resolution data with 64,000 temporal points (mass resolution 10,000 at m/z 400) for all remaining tissue sections (37 coronal and 39 sagittal). During reconstruction, these low-resolution signals served as input for the regressor network, which predicted 32-dimensional latent vectors for each signal. The predicted latent vectors were decoded to transient signals with 1 million temporal points (mass resolution 160,000 at m/z 400) by the previously trained decoder. Finally, high-resolution mass spectra were generated from the decoded transients.

MEISTER for image-guided SCMS. We trained MEISTER using 3,840 random cells (data with 1 million temporal points) from five brain regions (neocortex, hippocampus, thalamus, striatum and corpus callosum) using microMS. The autoencoder and regressor were trained for 20 and 50 epochs, respectively, with a batch size of 64. The model was validated on a validation set containing 1,000 cells. The spectral correlation scores were calculated between the reference peak intensity and the reconstructed peak intensity. To show that MEISTER reconstruction provides consistent downstream analysis, we compared UMAP and Leiden clustering results between the reference (validation set) and reconstructed single-cell data, and we visualized the single-cell distributions of m/z features over UMAP (Extended Data Fig. 3). A total of 13,566 cells (64,000 temporal points) were reconstructed (1 million temporal points) to obtain high-resolution single-cell mass spectra.

Analysis of 3D MSI data

Data preparation. To prepare for data analysis, we first determined the peak centroids on the average mass spectra for each tissue section, and we extracted the intensities of the peak centroids for all pixels. The peak lists were then processed by m/z binning in 3-ppm increments to align peaks affected by potential mass shift. After extracting peak lists from a 3D dataset, we retained m/z bins common across all tissue sections for further analysis. Each pixel was normalized by total ion count (TIC). The processed data were finally converted into imzML file format.

Data-driven image registration. To enable 3D reconstruction and analysis of MSI data with respect to brain anatomy, we registered MSI serial sections to T_2^* -weighted anatomical MRI from the Waxholm Space atlas of the Sprague-Dawley rat brain. To ensure precise and accurate registration across serial sections, we adapted parametric UMAP to extract both structurally informative and consistent feature images from high-dimensional MSI data. Previous work has demonstrated using low-dimensional feature images (embeddings) obtained using nonparametric dimensionality reduction methods (both t -distributed

stochastic neighbor embedding (*t*-SNE) and UMAP) for image registration tasks. However, feature images from different tissue sections can provide disparate morphological contrasts, because data are essentially embedded into different embedding space for each tissue section. Embedding the entire 3D MSI dataset can overcome such an issue, but it is computationally intractable for *t*-SNE and UMAP optimization over millions of input pixels with thousands of dimensions. The major advantage of the parametric version of UMAP is to use a neural network to learn a relationship between data and embedding. Thus, a small subset of pixels can be sampled from 3D MSI data for training the network, which can rapidly embed a large number of pixels into a single embedding space. We used an autoencoder in conjunction with UMAP, of which the encoder is trained to minimize UMAP loss and the decoder is trained to minimize reconstruction loss. The autoencoder input and output size is set to be the number of *m/z* features, followed by fully connected layers with 256, 128 and 64 neurons. The UMAP loss function between two data points *i* and *j* is the cross-entropy, defined as:

$$C_{\text{UMAP}} = \sum_{i \neq j} v_{ij} \log \left(\frac{v_{ij}}{w_{ij}} \right) + (1 - v_{ij}) \log \left(\frac{1 - v_{ij}}{1 - w_{ij}} \right)$$

Ten percent of total pixels were randomly sampled as training data, and the network was trained for 20 epochs. For each tissue section, embedding vectors from three UMAP dimensions were encoded into red, green and blue (RGB) channels to form feature images, which were then converted to grayscale ready for registration.

The anatomical images from the MRI atlas were selected on the basis of the tissue sectioning distance with manual inspection. We applied a two-step multimodal image registration to align grayscale MSI feature images as the moving images with the reference anatomical images. First, rigid affine registration was performed to roughly align the two with nine hand-selected initial transformation points. After rough alignment, a nonrigid cubic B-spline registration was performed with mutual information as the similarity measure with 200 maximum optimization steps. The registration quality was evaluated by the Dice index (DI), which assesses the image mask overlap between the *i*th brain region labels from the atlas and the human-annotated masks from the *j*th registered tissue section (Supplementary Fig. 10):

$$DI_{ij} = \frac{2|Label_{ij} \cap Annotation_{ij}|}{|Label_{ij}| + |Annotation_{ij}|}$$

The transformation matrices were applied to each MSI section to visualize registered ion images. The region-specific mass spectral profiles were extracted from pixels on the basis of the atlas brain region labels.

Brain region-specific analysis. To analyze regional lipid distributions at the single-pixel level, we extracted all MSI pixels belonging to 11 brain structures on the basis of MRI atlas labels, including the neocortex, hippocampus, thalamus and hypothalamus as gray matter-dominant regions, the corpus callosum and corticofugal pathway as white matter-dominant regions and the superior colliculus, basal forebrain, brainstem, striatum and septal regions as regions (others) that contain collections of clusters of cell bodies, as well as processes for signal communication. For individual tissue sections, the extracted pixels were first subjected to UMAP for visualization and used to train gradient boosting tree (GBT) models to predict the brain regions, a multiclass classification task. Training and testing set sizes were set to 0.8 and 0.2. GBT models were further interpreted through SHAP (SHapley Additive exPlanations) values. In SHAP, each pixel provides the lipid feature attributions toward predicting certain brain regions, which can be used to generate feature attribution maps for ion images. The most contributing lipid features were selected by ranking mean absolute SHAP values. Regional average lipid profiles were obtained from every

tissue section, which was repeated for the aforementioned analysis. Differential analysis of lipid features was performed for each brain region to obtain the \log_2 fold change and *P*-values tested by Wilcoxon rank-sum test and adjusted by Benjamini–Hochberg. For putative lipid annotation, we searched the *m/z* values against LIPID MAPS³⁴ experimental and virtual databases with a ± 0.005 *m/z* threshold for chemical formula and lipid species assignments. From the combined list, the matches were sorted according to their ppm errors from the accurate masses. In cases when experimentally or structurally validated lipids (biologically relevant lipids present in LIPID MAPS) were matched, they were given priority for assignment.

Joint analysis of MSI and SCMS data

Cross-annotation strategy. We applied a straightforward strategy to annotate lipids in SCMS data using features observed in MSI data. Similar to putative annotation, features in MSI served as a database to search the SCMS peak lists for matching lipids within a 3-ppm *m/z* window. Features present in less than 5% of cells were discarded, and cells with less than 5% total number of features were filtered out. Using an alternative method, we first annotated the tissue MSI data using METASPACE with the CoreMetabolome database and obtained the monoisotopic *m/z* features from the annotation results with a 50% FDR. These monoisotopic ions were then used to search the SCMS data.

Integrative analysis using UoSS fitting. The $m \times n$ single-cell lipid feature matrix \mathbf{X} was first processed by TIC normalization, with *m* being the number of cells in each identified cluster and *n* being the number of lipid features. Leiden clustering was performed on the first 40 PCs with the parameters `n_neighbors = 30`, `min_dist = 0.5` and `resolution = 0.25`, using cosine as the distance metric. The single-cell matrix $\mathbf{X}^{(l)}$ of cell cluster *l* can be decomposed into:

$$\mathbf{X}^{(l)} = \mathbf{W}^{(l)} \mathbf{D}^{(l)}$$

where $\mathbf{W}^{(l)}$ represents the $m \times k$ weight matrix (with *k* being the number of dictionary items capturing the chemical variations within each cluster) and $\mathbf{D}^{(l)}$ is the $k \times n$ non-negative dictionary matrix that contains sparse representations of lipid signatures in the *l*th single-cell cluster. We chose *k* = 20 for the NMF algorithm. The union-of-dictionary items concatenated across *L* clusters $\mathbf{U} = [\mathbf{D}^{(0)}; \mathbf{D}^{(1)}; \dots; \mathbf{D}^{(L)}]$ was fitted to the $p \times n$ tissue imaging data matrix \mathbf{Y} , where *p* is the number of tissue pixels, by a constrained linear least-squares fitting, with model weights constrained to be non-negative:

$$\mathbf{Y} = \mathbf{C}\mathbf{U}$$

$$\mathbf{C} = \underset{\mathbf{C}}{\text{argmin}} \|\mathbf{Y} - \mathbf{C}\mathbf{U}\|_F^2, \mathbf{C} \geq 0$$

The $p \times k \times L$ tissue weight matrix \mathbf{C} has the row vectors $\mathbf{C}_{p_i} = [c_{p_i,0}, c_{p_i,1}, \dots, c_{p_i,L}]$ and $c_{p_i,l}$ contains the weights for pixel *p_i* and *k* dictionary items from the *l*th single-cell cluster. The vector norm $\|c_{p_i,l}\|_2$ was taken corresponding to the summarized contributions from all dictionary items of the *l*th cluster. The vector norms were mapped back to original pixel locations to visualize the spatial contributions of single-cell lipid signatures at the tissue level.

Reporting summary

Further information on research design is available in the Nature Portfolio Reporting Summary linked to this article.

Data availability

The processed 3D MSI, SCMS and other relevant imaging data that support the findings of this study are publicly available and free to download from Illinois Data Bank⁶⁴ at https://doi.org/10.13012/B2IDB-9740536_V1. Due to large file sizes, raw data including simulated

and experimental high-mass-resolution transients can be available upon reasonable request to the corresponding authors to arrange data sharing.

Code availability

The code used in this study is free for noncommercial use and available on GitHub (<https://github.com/richardxie1119/MEISTER>).

References

- Jansson, E. T., Comi, T. J., Rubakhin, S. S. & Sweedler, J. V. Single cell peptide heterogeneity of rat islets of Langerhans. *ACS Chem. Biol.* **11**, 2588–2595 (2016).
- Lee, D.-K., Rubakhin, S. S. & Sweedler, J. V. Chemical decrosslinking-based peptide characterization of formaldehyde-fixed rat pancreas using fluorescence-guided single-cell mass spectrometry. *Anal. Chem.* **95**, 6732–6739 (2023).
- Marshall, A. G., Comisarow, M. B. & Parisod, G. Relaxation and spectral line shape in Fourier transform ion resonance spectroscopy. *J. Chem. Phys.* **71**, 4434–4444 (1979).
- Marshall, A. G., Hendrickson, C. L. & Jackson, G. S. Fourier transform ion cyclotron resonance mass spectrometry: a primer. *Mass Spectrom. Rev.* **17**, 1–35 (1998).
- Dittwald, P., Claesen, J., Burzykowski, T., Valkenburg, D. & Gambin, A. BRAIN: a universal tool for high-throughput calculations of the isotopic distribution for mass spectrometry. *Anal. Chem.* **85**, 1991–1994 (2013).
- 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).
- Xie, Y. R. et al. Integrative multiscale biochemical mapping of the brain via deep-learning-enhanced high-throughput mass spectrometry. Illinois Data Bank. https://doi.org/10.13012/B2IDB-9740536_V1 (2023).

Acknowledgements

This project was supported by the National Institute on Drug Abuse under award no. P30 DA018310 (S.S.R., J.V.S.), the National Institute on

Aging under award no. 1R01AG078797 (J.V.S., F.L.), the National Human Genome Research Institute under award no. RM1HG010023 (J.V.S.), and the National Institute of General Medical Sciences under award no. 1R35GM142969 (F.L.). The content is solely the responsibility of the authors and does not necessarily represent the official views of the awarding agencies.

Author contributions

Y.R.X., F.L. and J.V.S. conceptualized the project. Y.R.X. and F.L. developed the deep-learning and computational methods. S.S.R. and D.C.C. designed and performed the sample preparation for tissues and single cells. Y.R.X., T.J.T. and D.C.C. acquired MS data. Y.R.X. implemented and evaluated the computational algorithms. Y.R.X., T.J.T., F.L. and J.V.S. performed data analysis. Y.R.X. created the figures. All authors wrote the manuscript. F.L. and J.V.S. acquired funding and provided direction throughout the project.

Competing interests

The authors declare no competing interests.

Additional information

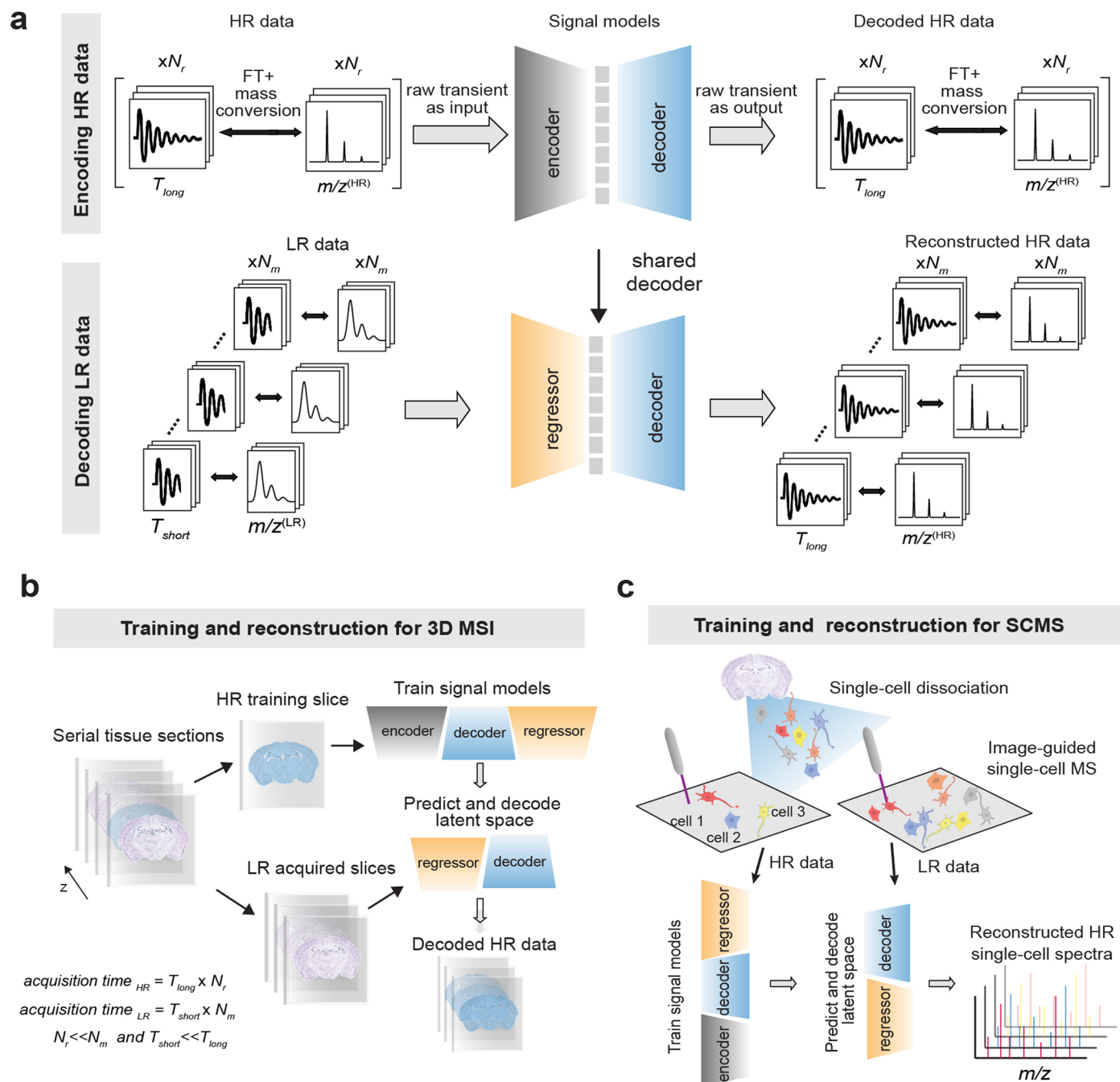
Extended data is available for this paper at <https://doi.org/10.1038/s41592-024-02171-3>.

Supplementary information The online version contains supplementary material available at <https://doi.org/10.1038/s41592-024-02171-3>.

Correspondence and requests for materials should be addressed to Jonathan V. Sweedler or Fan Lam.

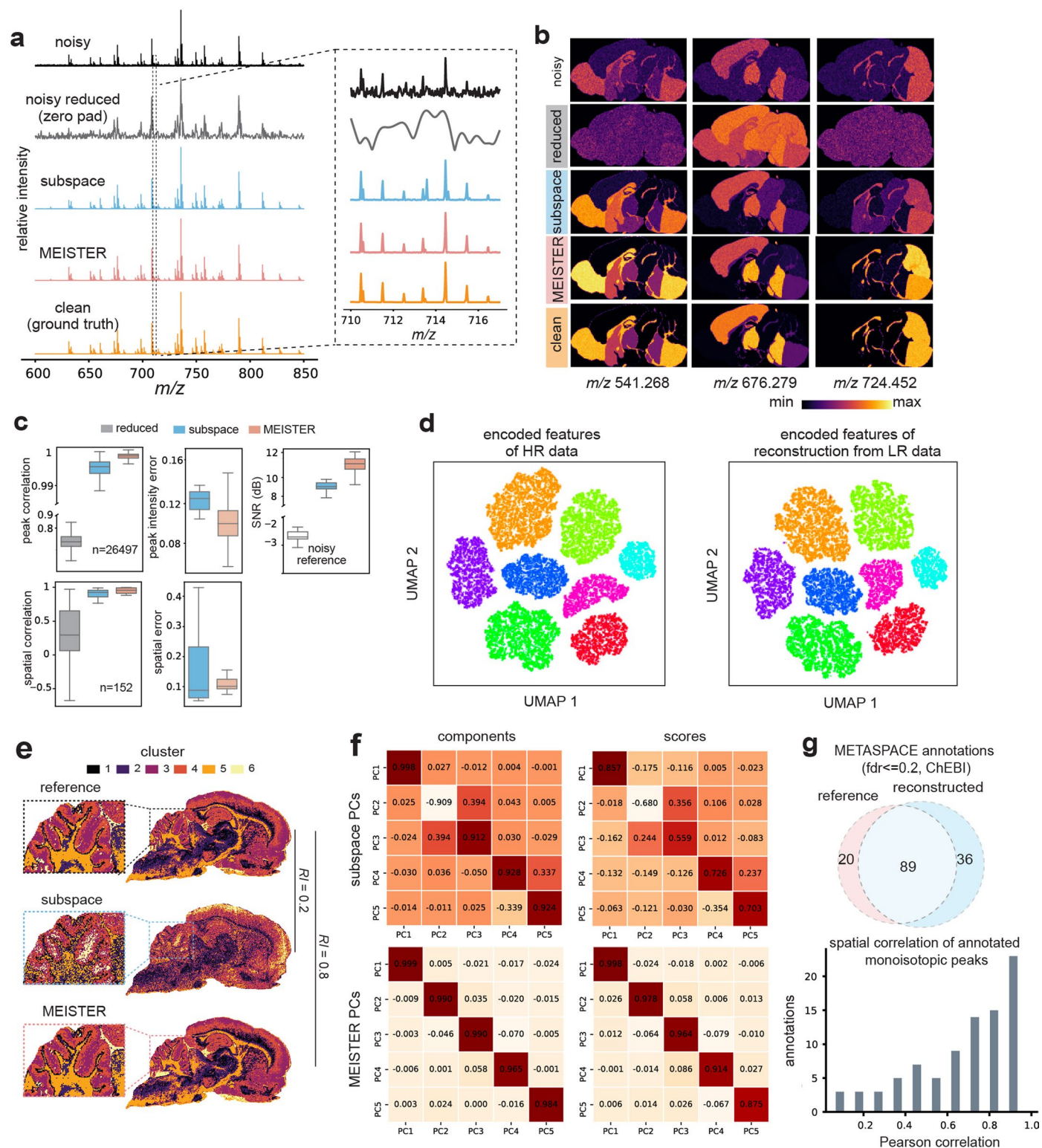
Peer review information *Nature Methods* thanks Christopher Anderton, Mitsutoshi Setou and the other, anonymous, reviewer(s) for their contribution to the peer review of this work. Primary Handling Editor: Rita Strack, in collaboration with the *Nature Methods* team.

Reprints and permissions information is available at www.nature.com/reprints.



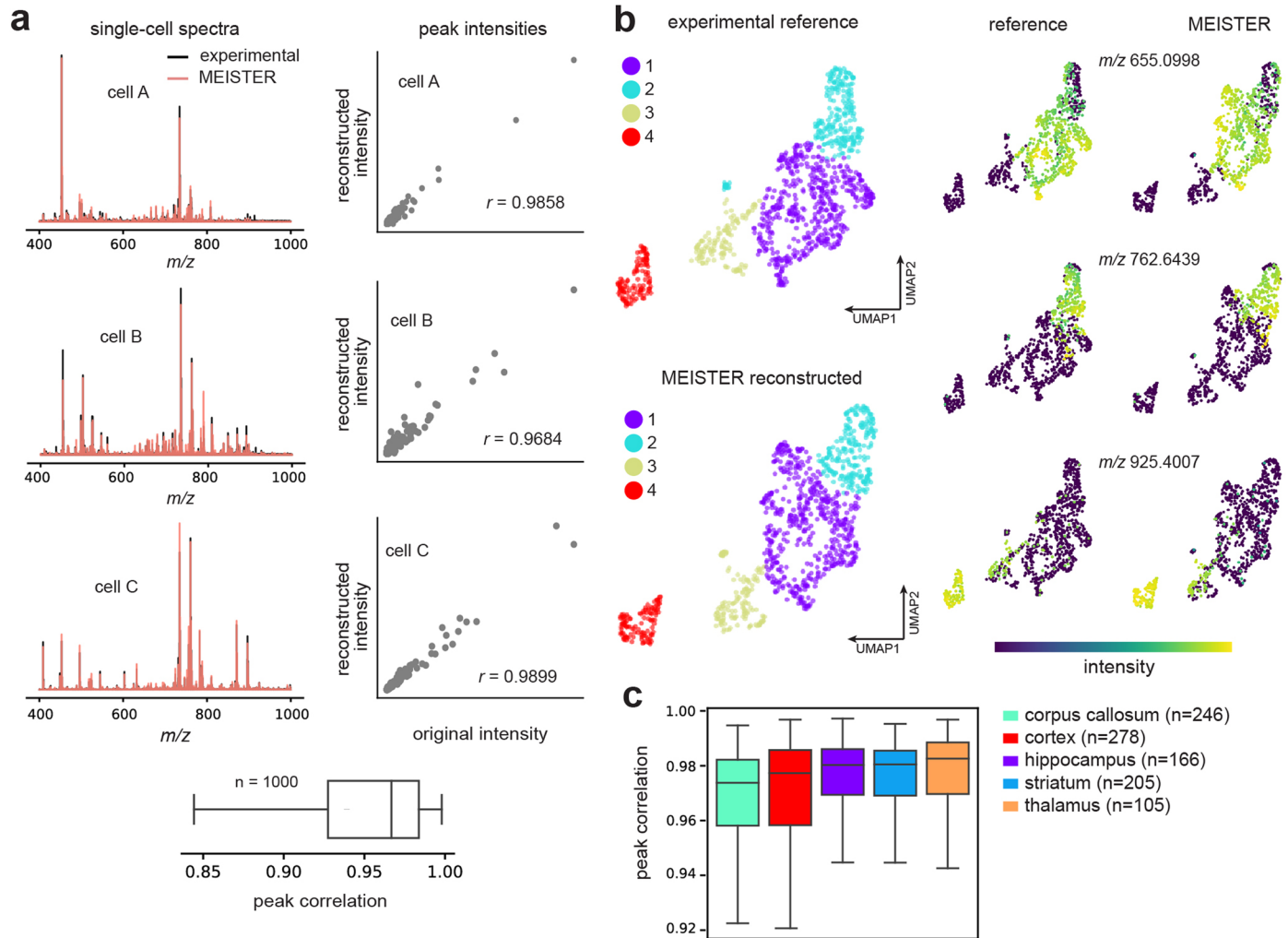
Extended Data Fig. 1 | MEISTER model design. a, MEISTER reconstruction model that contains an autoencoder to learn latent features from high-resolution signals, and a regressor network that maps low-resolution signals to encoded

latent features. MEISTER training workflow for **b**, 3D MSI using high-mass resolution data acquired on a small number of tissue sections, and **c**, SCMS using high-mass resolution data acquired on a subset of individual cells.



Extended Data Fig. 2 | Evaluating model performance on simulated and experimental MSI data. Comparisons of **a**, mass spectra from the simulated MSI data. **b**, ion images extracted from several m/z features, showing enhanced spectral and image quality enabled by MEISTER reconstruction. **c**, Correlation coefficient and error distributions by evaluating mass spectra and ion images against the ground truth. The simulation data contain $n = 26497$ pixels and $n = 152$ ion images for peak and spatial correlation respectively. Data in boxplots are shown as median values (center) with the interquartile range (box), and the whiskers extend to 1.5 times the interquartile range. **d**, UMAP embeddings

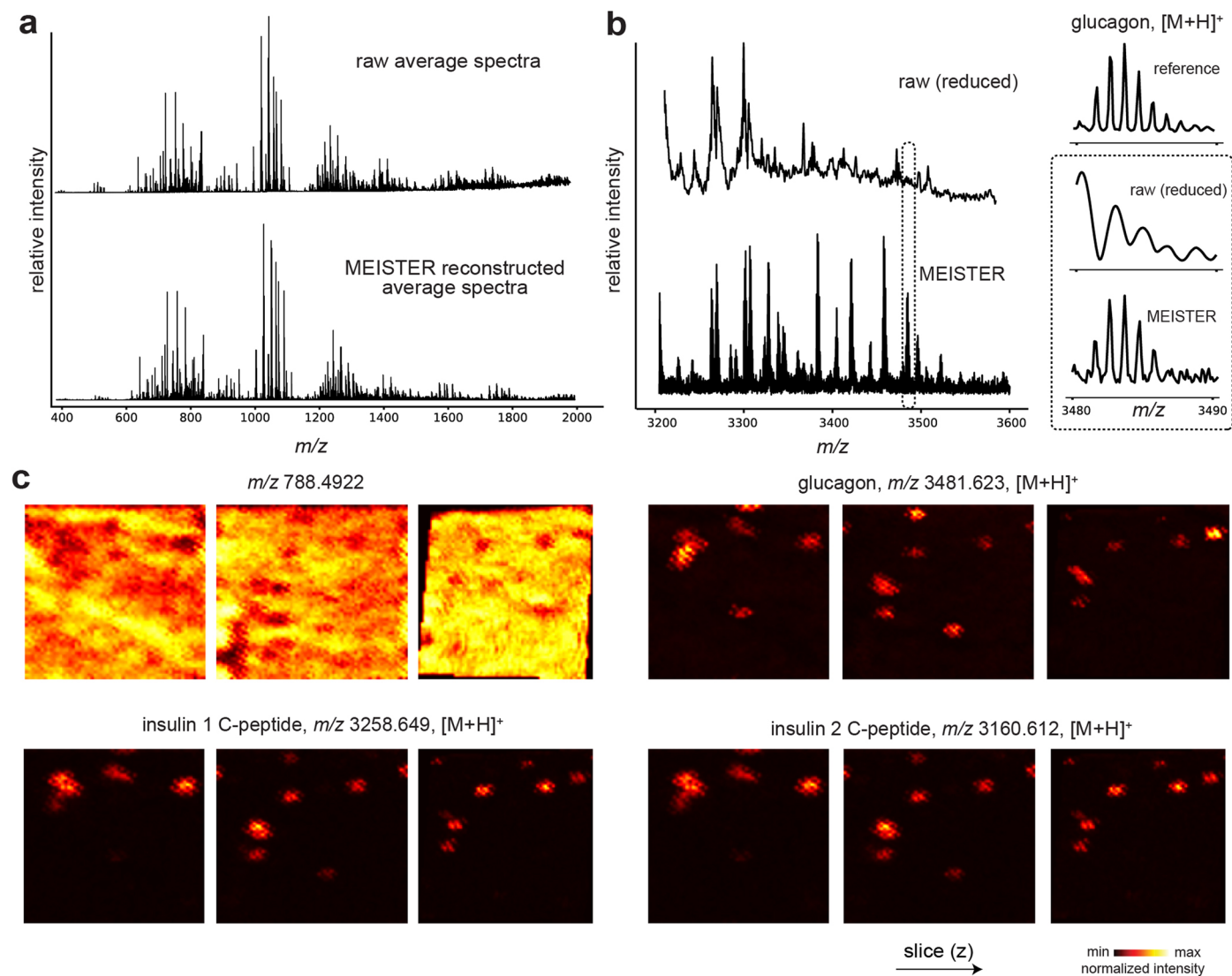
of encoded features of the simulated high-resolution data and the features of reconstruction from low-resolution data. Colors indicate different pseudo-tissue regions. **e**, K-means clustering for the experimental reference, subspace reconstruction, and MEISTER reconstruction. **f**, Pearson correlation coefficients between top-5 PCs extracted from the experimental reference versus from the data reconstructed by subspace (top) and MEISTER (bottom). **g**, Comparison of number of annotated lipids (top) and correlation of ion images (bottom) using METASPACE with FDR set to 20%.



Extended Data Fig. 3 | Reconstructing image-guided SCMS data.

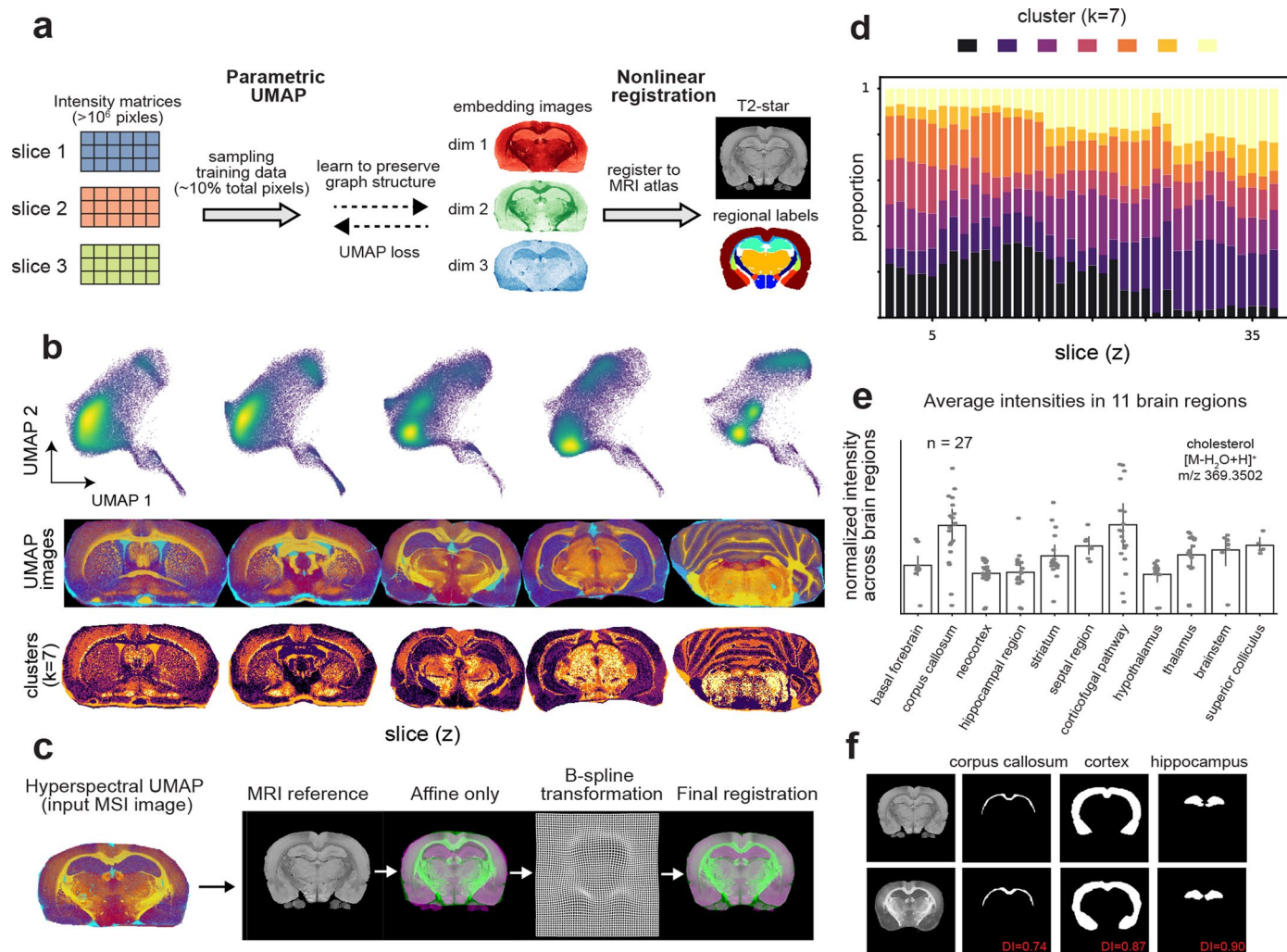
a, Experimental high-resolution single-cell mass spectra versus MEISTER reconstructed mass spectra. High peak correlation scores were obtained on 1000 validation cells (bottom box plot). **b**, Downstream analysis of the reference (full transients) and reconstructed data shows nearly identical clustering patterns

through k-means ($k = 4$; 1-4 denote cluster numbers and each cluster of cells are coded with a different color) and ion distributions at the single-cell level. **c**, Peak correlation scores for cells sampled from five different brain regions. Data in boxplots are shown as median values (center) with the interquartile range (box), and the whiskers extend to 1.5 times the interquartile range.



Extended Data Fig. 5 | Generalizability to imaging peptides in rat pancreas. **a,b**, Averaged mass spectra of rat pancreas tissue sections from the reduced data (top) and our reconstruction (bottom) for m/z range of 400 to 2000 (**a**; lipids) and 3200 to 3600 (**b**; peptides). Inlet displays a zoomed-in m/z window with signals of protonated glucagon for high-resolution reference (top), reduced

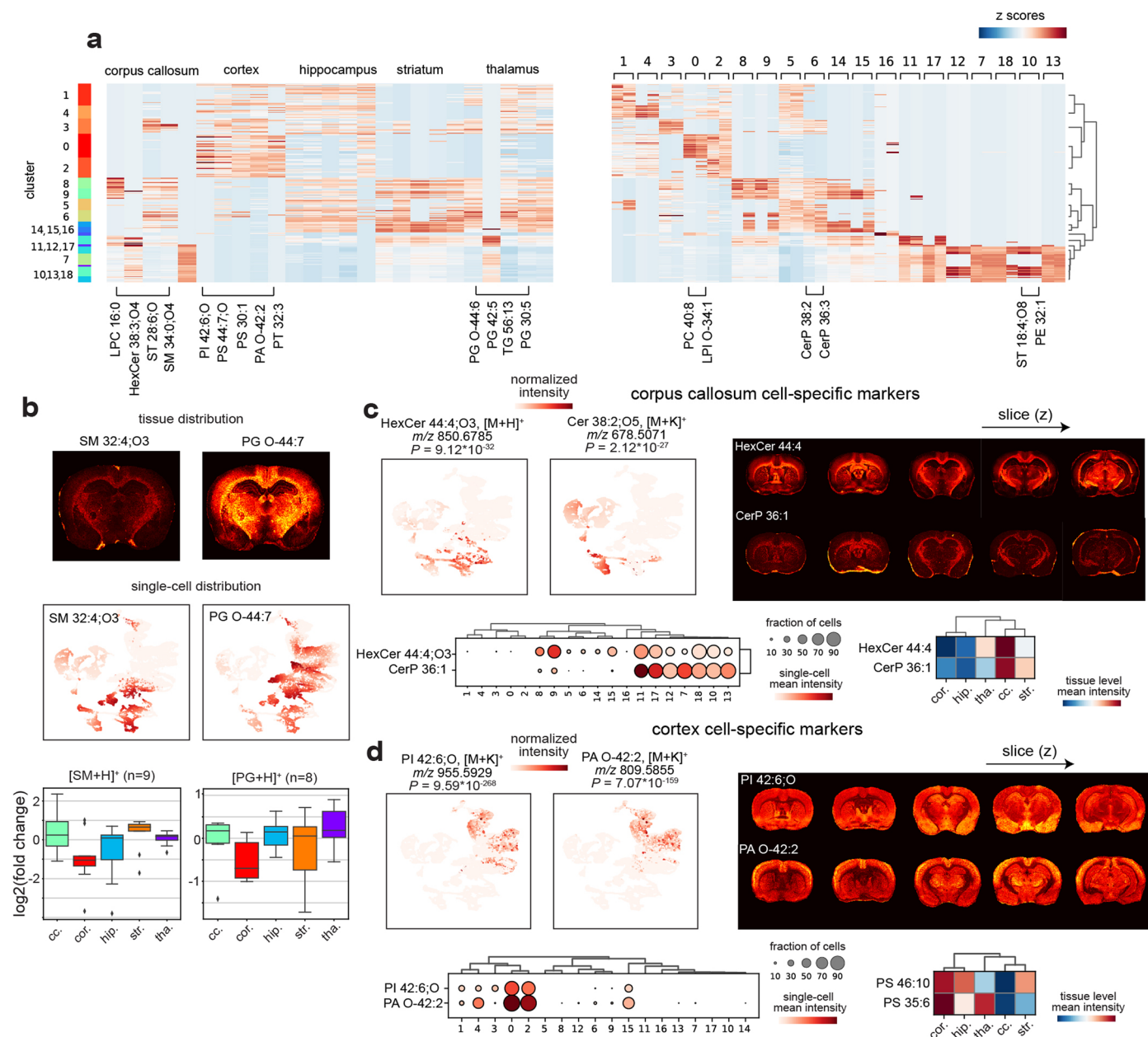
(middle) and reconstructed (bottom) data. High-fidelity reconstruction by the proposed method w.r.t. the reference can be observed. **c**, Ion images of different sections obtained from deep learning reconstructed tissue MSI data for m/z 788.4922, glucagon, insulin 1 C-peptide, and insulin 2 C-peptide.



Extended Data Fig. 6 | Data-driven registration to align 3D MSI and anatomical atlas.

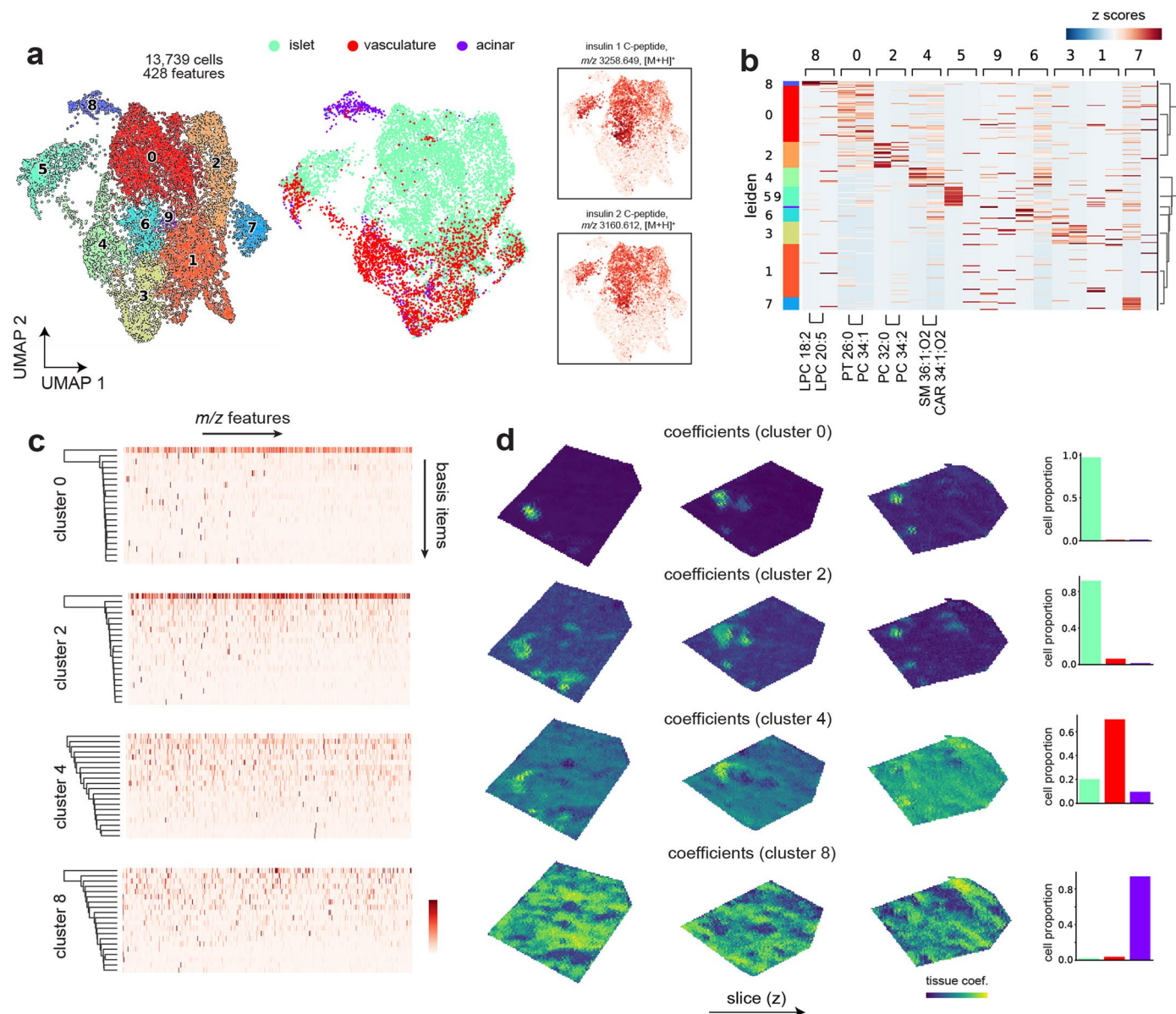
a, The proposed workflow leveraging pixel-wise parametric UMAP to efficiently align MSI data from serial tissue sections to a 3D MRI volume. **b**, Across serial sections in z-axis, parametric UMAP embeddings (top, colored by point density) formed consistent and structurally informative feature images (middle) and k-means clusters (bottom). **c**, Nonlinear image registration aligning the hyperspectral UMAP image to the target MR image via sequential affine and B-spline registration. The combination produced excellent alignment and enabled coherent volumetric reconstruction of MSI sections. **d**, The cluster

proportions for 7 clusters varying with the slice order. **e**, Differential signal intensity distributions for lipids across 11 brain anatomical structures (for $n = 27$ tissue sections) identified using the ROI labels from a rat MRI atlas, that is, cholesterol shown here with distinct regional differences. The bars indicate the mean value intensity values, and the error bars indicate the 95 percent confidence intervals of the intensity distributions. **f**, Anatomical masks from the atlas (top) and manual annotations of MSI post registration (bottom). Dice Indices are shown in red indicating good alignment.



Extended Data Fig. 7 | Identification of brain region-specific lipid features at the single-cell level. **a**, Top five single-cell lipid features identified to be brain-region specific (left), and top two lipid features identified to be cluster-specific (right). Lipid features were selected based on p-values and log₂ of fold change obtained by differential analysis. Rows and columns correspond to cells organized by the clusters and lipid features organized by regions respectively. Region and cluster specific lipids can be identified. For instance, PI 42:6;O is significantly elevated (adj. p-value = 9.6*10⁻²⁶⁸) in cells from cortex. Inspecting the PI 42:6;O column in the heatmap, we can observe that most cells in cluster 0, 2, 3, 4 contain this particular lipid. **b**, Tissue and single-cell distributions of lipid markers identified by region-specific lipid analysis, demonstrated by data of two representative lipids, SM (32:4);O3 and PG O-(44:7). Bottom: regional

distributions of SM and PG signal intensities quantified by log₂ of fold change at the single-cell level. n is the number of lipid features for each lipid classes. Data in boxplots are shown as median values (center) with the interquartile range (box), and the whiskers extend to 1.5 times the interquartile range. **c**, **d**, Highly-specific lipid markers (significance indicated by p-values) were identified for different brain regions, showing agreement between single-cell and tissue imaging data for **c**, corpus callosum and **d**, cortex regions. For **c**, **d**, top left: single-cell UMAP, top right: corresponding ion images, bottom left: relations between clusters, mean signal intensity for single cells in cluster, and size of cell fraction per cluster, bottom right: relation between mean of signal intensities, brain locations of collected signals. p-values were tested by Wilcoxon rank-sum (two-sided) and adjusted by Benjamini-Hochberg procedure.



Extended Data Fig. 8 | Integrative analysis of MSI and SCMS data from rat pancreas. **a**, A total of 13,739 cells from rat pancreas with 428 cross-annotated features (tissue and single cells) are subjected to UMAP and Leiden clustering analysis. 10 cell clusters were identified (left), which can also be mapped to three major pancreatic cell types (right). Inlet shows the distributions of insulin 1 and 2 C-peptides within single cells. **b**, Top two features identified to be

cluster-specific across all clusters. **c**, Cell-cluster-specific dictionaries extracted from representative cluster 0, 2, 4, and 8. **d**, Estimated spatial contributions of individual cell clusters across pancreas tissue. Each row shows results of mapping the contributions of one cluster to individual pixels, revealing distinct spatial organizations of islet, vasculature, and acinar cells.

Reporting Summary

Nature Portfolio wishes to improve the reproducibility of the work that we publish. This form provides structure for consistency and transparency in reporting. For further information on Nature Portfolio policies, see our [Editorial Policies](#) and the [Editorial Policy Checklist](#).

Please do not complete any field with "not applicable" or n/a. Refer to the help text for what text to use if an item is not relevant to your study.

For final submission: please carefully check your responses for accuracy; you will not be able to make changes later.

Statistics

For all statistical analyses, confirm that the following items are present in the figure legend, table legend, main text, or Methods section.

n/a Confirmed

- | | | |
|-------------------------------------|-------------------------------------|--|
| <input type="checkbox"/> | <input checked="" type="checkbox"/> | The exact sample size (n) for each experimental group/condition, given as a discrete number and unit of measurement |
| <input type="checkbox"/> | <input checked="" type="checkbox"/> | A statement on whether measurements were taken from distinct samples or whether the same sample was measured repeatedly |
| <input type="checkbox"/> | <input checked="" type="checkbox"/> | The statistical test(s) used AND whether they are one- or two-sided
<i>Only common tests should be described solely by name; describe more complex techniques in the Methods section.</i> |
| <input checked="" type="checkbox"/> | <input type="checkbox"/> | A description of all covariates tested |
| <input type="checkbox"/> | <input checked="" type="checkbox"/> | A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons |
| <input type="checkbox"/> | <input checked="" type="checkbox"/> | A full description of the statistical parameters including central tendency (e.g. means) or other basic estimates (e.g. regression coefficient) AND variation (e.g. standard deviation) or associated estimates of uncertainty (e.g. confidence intervals) |
| <input type="checkbox"/> | <input checked="" type="checkbox"/> | For null hypothesis testing, the test statistic (e.g. F , t , r) with confidence intervals, effect sizes, degrees of freedom and P value noted
<i>Give P values as exact values whenever suitable.</i> |
| <input checked="" type="checkbox"/> | <input type="checkbox"/> | For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings |
| <input checked="" type="checkbox"/> | <input type="checkbox"/> | For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes |
| <input type="checkbox"/> | <input checked="" type="checkbox"/> | Estimates of effect sizes (e.g. Cohen's d , Pearson's r), indicating how they were calculated |

Our web collection on [statistics for biologists](#) contains articles on many of the points above.

Software and code

Policy information about [availability of computer code](#)

Data collection

Microscopy data were collected using ZEN Pro 3.1 (Zeiss) and exported to bigTIFF files for downstream analysis. FTMS data were collected with fMS Control 2.3.0 (Bruker Daltonics).

Data analysis

Detailed instructions on reproducing the data analysis are provided in the supplemental protocol; Data analysis is performed using Python 3.8; code can be downloaded from <https://github.com/richardxie1119/MEISTER>.

For manuscripts utilizing custom algorithms or software that are central to the research but not yet described in published literature, software must be made available to editors and reviewers. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Portfolio [guidelines for submitting code & software](#) for further information.

Data

Policy information about [availability of data](#)

All manuscripts must include a [data availability statement](#). This statement should provide the following information, where applicable:

- Accession codes, unique identifiers, or web links for publicly available datasets
- A description of any restrictions on data availability
- For clinical datasets or third party data, please ensure that the statement adheres to our [policy](#)

LIPID MAPS database was used for putative lipid assignments. The processed 3D MSI, single-cell MS and other relevant imaging data that support the findings of this study are publicly available and free to download from Illinois Data Bank64 at https://doi.org/10.13012/B21DB-9740536_V1. Due to large file sizes, raw data including simulated and experimental high-mass-resolution transients can be available upon reasonable request to the corresponding authors to arrange data sharing.

Research involving human participants, their data, or biological material

Policy information about studies with [human participants or human data](#). See also policy information about [sex, gender \(identity/presentation\), and sexual orientation](#) and [race, ethnicity and racism](#).

Reporting on sex and gender	n/a
Reporting on race, ethnicity, or other socially relevant groupings	n/a
Population characteristics	n/a
Recruitment	n/a
Ethics oversight	n/a

Note that full information on the approval of the study protocol must also be provided in the manuscript.

Field-specific reporting

Please select the one below that is the best fit for your research. If you are not sure, read the appropriate sections before making your selection.

Life sciences Behavioural & social sciences Ecological, evolutionary & environmental sciences

For a reference copy of the document with all sections, see [nature.com/documents/nr-reporting-summary-flat.pdf](https://www.nature.com/documents/nr-reporting-summary-flat.pdf)

Life sciences study design

All studies must disclose on these points even when the disclosure is negative.

Sample size	Three animals were used for both brain tissue isolation and for pancreas tissue isolation.
Data exclusions	Cells closer than 200 μm were excluded from microMS single-cell analysis.
Replication	The same methodology was repeated using both coronal and sagittal sections.
Randomization	Cells from three different animals were randomly mixed and placed on slides for single-cell mass spectrometry.
Blinding	Blinding was not relevant to this study as data were only generated to demonstrate a new computational mass spectrometry workflow.

Behavioural & social sciences study design

All studies must disclose on these points even when the disclosure is negative.

Study description	
Research sample	
Sampling strategy	
Data collection	
Timing	
Data exclusions	
Non-participation	
Randomization	

Ecological, evolutionary & environmental sciences study design

All studies must disclose on these points even when the disclosure is negative.

Study description	<input type="text"/>
Research sample	<input type="text"/>
Sampling strategy	<input type="text"/>
Data collection	<input type="text"/>
Timing and spatial scale	<input type="text"/>
Data exclusions	<input type="text"/>
Reproducibility	<input type="text"/>
Randomization	<input type="text"/>
Blinding	<input type="text"/>

Did the study involve field work? Yes No

Field work, collection and transport

Field conditions	<input type="text"/>
Location	<input type="text"/>
Access & import/export	<input type="text"/>
Disturbance	<input type="text"/>

Reporting for specific materials, systems and methods

We require information from authors about some types of materials, experimental systems and methods used in many studies. Here, indicate whether each material, system or method listed is relevant to your study. If you are not sure if a list item applies to your research, read the appropriate section before selecting a response.

Materials & experimental systems

n/a	Involved in the study
<input checked="" type="checkbox"/>	<input type="checkbox"/> Antibodies
<input checked="" type="checkbox"/>	<input type="checkbox"/> Eukaryotic cell lines
<input checked="" type="checkbox"/>	<input type="checkbox"/> Palaeontology and archaeology
<input type="checkbox"/>	<input checked="" type="checkbox"/> Animals and other organisms
<input checked="" type="checkbox"/>	<input type="checkbox"/> Clinical data
<input checked="" type="checkbox"/>	<input type="checkbox"/> Dual use research of concern
<input checked="" type="checkbox"/>	<input type="checkbox"/> Plants

Methods

n/a	Involved in the study
<input checked="" type="checkbox"/>	<input type="checkbox"/> ChIP-seq
<input checked="" type="checkbox"/>	<input type="checkbox"/> Flow cytometry
<input checked="" type="checkbox"/>	<input type="checkbox"/> MRI-based neuroimaging

Antibodies

Antibodies used	<input type="text" value="n/a"/>
Validation	<input type="text" value="n/a"/>

Eukaryotic cell lines

Policy information about [cell lines and Sex and Gender in Research](#)

Cell line source(s)	<input type="text"/>
Authentication	<input type="text"/>
Mycoplasma contamination	<input type="text"/>
Commonly misidentified lines (See ICLAC register)	<input type="text"/>

Palaeontology and Archaeology

Specimen provenance	<input type="text"/>
Specimen deposition	<input type="text"/>
Dating methods	<input type="text"/>
<input type="checkbox"/> Tick this box to confirm that the raw and calibrated dates are available in the paper or in Supplementary Information.	
Ethics oversight	<input type="text"/>

Note that full information on the approval of the study protocol must also be provided in the manuscript.

Animals and other research organisms

Policy information about [studies involving animals](#); [ARRIVE guidelines](#) recommended for reporting animal research, and [Sex and Gender in Research](#)

Laboratory animals	Male 2.0-2.5 Sprague-Dawley® outbred rats (Rattus norvegicus) used in this study and were sourced from Inotivco.
Wild animals	No wild animals were used in this study.
Reporting on sex	Male
Field-collected samples	No field-collected samples were used in this study.
Ethics oversight	Illinois Institutional Animal Care and Use Committee as well as the federal and ARRIVE guideline

Note that full information on the approval of the study protocol must also be provided in the manuscript.

Clinical data

Policy information about [clinical studies](#)

All manuscripts should comply with the ICMJE [guidelines for publication of clinical research](#) and a completed [CONSORT checklist](#) must be included with all submissions.

Clinical trial registration	<input type="text"/>
Study protocol	<input type="text"/>
Data collection	<input type="text"/>
Outcomes	<input type="text"/>

Dual use research of concern

Policy information about [dual use research of concern](#)

Hazards

Could the accidental, deliberate or reckless misuse of agents or technologies generated in the work, or the application of information presented in the manuscript, pose a threat to:

- | No | Yes |
|--------------------------|---|
| <input type="checkbox"/> | <input type="checkbox"/> Public health |
| <input type="checkbox"/> | <input type="checkbox"/> National security |
| <input type="checkbox"/> | <input type="checkbox"/> Crops and/or livestock |
| <input type="checkbox"/> | <input type="checkbox"/> Ecosystems |
| <input type="checkbox"/> | <input type="checkbox"/> Any other significant area |

Experiments of concern

Does the work involve any of these experiments of concern:

- | No | Yes |
|--------------------------|--|
| <input type="checkbox"/> | <input type="checkbox"/> Demonstrate how to render a vaccine ineffective |
| <input type="checkbox"/> | <input type="checkbox"/> Confer resistance to therapeutically useful antibiotics or antiviral agents |
| <input type="checkbox"/> | <input type="checkbox"/> Enhance the virulence of a pathogen or render a nonpathogen virulent |
| <input type="checkbox"/> | <input type="checkbox"/> Increase transmissibility of a pathogen |
| <input type="checkbox"/> | <input type="checkbox"/> Alter the host range of a pathogen |
| <input type="checkbox"/> | <input type="checkbox"/> Enable evasion of diagnostic/detection modalities |
| <input type="checkbox"/> | <input type="checkbox"/> Enable the weaponization of a biological agent or toxin |
| <input type="checkbox"/> | <input type="checkbox"/> Any other potentially harmful combination of experiments and agents |

Plants

- Seed stocks
- Novel plant genotypes
- Authentication

ChIP-seq

Data deposition

- Confirm that both raw and final processed data have been deposited in a public database such as [GEO](#).
- Confirm that you have deposited or provided access to graph files (e.g. BED files) for the called peaks.

- Data access links
May remain private before publication.
- Files in database submission
- Genome browser session (e.g. [UCSC](#))

Methodology

- Replicates
- Sequencing depth
- Antibodies
- Peak calling parameters
- Data quality
- Software

Flow Cytometry

Plots

Confirm that:

- The axis labels state the marker and fluorochrome used (e.g. CD4-FITC).
- The axis scales are clearly visible. Include numbers along axes only for bottom left plot of group (a 'group' is an analysis of identical markers).
- All plots are contour plots with outliers or pseudocolor plots.
- A numerical value for number of cells or percentage (with statistics) is provided.

Methodology

- Sample preparation
- Instrument
- Software
- Cell population abundance
- Gating strategy

Tick this box to confirm that a figure exemplifying the gating strategy is provided in the Supplementary Information.

Magnetic resonance imaging

Experimental design

- Design type
- Design specifications
- Behavioral performance measures

- Imaging type(s)
- Field strength
- Sequence & imaging parameters
- Area of acquisition
- Diffusion MRI Used Not used

Preprocessing

- Preprocessing software
- Normalization
- Normalization template
- Noise and artifact removal
- Volume censoring

Statistical modeling & inference

- Model type and settings
- Effect(s) tested
- Specify type of analysis: Whole brain ROI-based Both

Statistic type for inference

(See [Eklund et al. 2016](#))

Correction

Models & analysis

n/a | Involved in the study

 Functional and/or effective connectivity Graph analysis Multivariate modeling or predictive analysis

Functional and/or effective connectivity

Graph analysis

Multivariate modeling and predictive analysis

