Developing an Integrative Spatial Multi-Omics Pipeline to Map Molecular Alterations in Murine Models

A Dissertation Presented for the

Doctor of Philosophy

Degree

The University of Tennessee, Knoxville

Amin Jarrahi

May 2026

Copyright © 2026 by Amin Jarrahi

All rights reserved.

DEDICATION

I dedicate this thesis to my family for their support in every aspect of my life. Throughout this journey, they continuously encouraged me, and their belief in me made it possible for me to achieve this milestone.

این پایان‌نامه را به خانواده‌ام تقدیم می‌کنم که در تمام جنبه‌های زندگی‌ام همواره پشتیبان و حامی من بوده‌اند. در طول این مسیر، آن‌ها پیوسته مرا تشویق کردند و ایمان و باورشان به من، امکان دستیابی به این دستاورد ارزشمند را فراهم ساخت.

ACKNOWLEDGEMENTS

I am deeply grateful to my advisor, Dr. Colleen Crouch, for her continuous support throughout my doctoral studies. Her dedication to creating a positive and stress-free lab environment made this journey much smoother for me. Her commitment to thorough and careful research has greatly influenced this work and helped me develop as a researcher. It has been an honor to conduct my research under her mentorship. Also, I would like to thank my committee members, Dr. Hairong Qi, Dr. Bryan Good, and Dr. Sara Hanrahan, for their valuable feedback and guidance during this research.

I am thankful to Dr. Rebecca Prosser, Dr. Kalynn Schulz, Jaydeep Kolape, Dr. Weisheng Tang, Dr. Madeline Colley, Dr. Martin Dufresne, and Christopher Carter for their support, collaboration, training, and helpful discussions during different stages of this project.

I would like to express my gratitude for the financial support that made this research possible. Significant funding for this work was provided by the UT–Oak Ridge Innovation Institute (UT-ORII) Science Alliance: Support for Affiliated Research Teams (StART) Program. I also received support from the UTK ORIED Human Health and Wellness Gateway FY2026 and AI Tennessee Initiative Seed Funds. In addition, I am grateful to the Department of Biomedical Engineering at the University of Tennessee for supporting me as a teaching assistant, an experience that greatly enriched my doctoral journey.

I want to thank all my labmates – especially Allison Jones – for their help with research and for creating a supportive, collaborative, and enjoyable research environment.

I am also thankful to my friends for their encouragement and for making this journey more enjoyable during both the challenging and rewarding moments.

Finally, I would like to express my deepest gratitude to my family for their unconditional love. Even across the miles, their support and constant encouragement have carried me through the challenges of my academic journey, and their belief in me has made reaching this milestone possible.

ABSTRACT

Understanding spatial heterogeneity in biological tissues is essential for understanding the molecular processes that contribute to complex diseases. Spatial omics technologies such as mass spectrometry imaging (MSI) and spatial transcriptomics make it possible to map lipids, metabolites, and RNA transcripts directly within tissue sections while preserving spatial information. However, the high dimensionality of these datasets and the differences in spatial resolution and sampling geometry between technologies create significant analytical and integration challenges.

This dissertation presents the development of computational frameworks for analyzing and integrating spatial omics data, with a focus on lipidomics and transcriptomics in the context of aging and Alzheimer's disease (AD). AD was selected as a biological model due to its strong associations with lipid dysregulation, gene expression changes, and spatially localized pathology in the brain.

First, a Python-based analytical method was developed to perform analyte preprocessing and quantitatively assess analyte delocalization in matrix-assisted laser desorption/ionization (MALDI) MSI data. A composite delocalization score was introduced that integrates normalized off-tissue area and mean distance from the tissue boundary, providing a robust measure of spatial spreading beyond tissue regions. This metric enables systematic evaluation of analyte reliability and supports quality control in MSI experiments.

Second, a computational framework was created to enable large-scale exploration of MSI datasets. The pipeline integrates preprocessing, dimensionality reduction, clustering, and statistical analysis to identify molecular features exhibiting biologically meaningful spatial patterns. Application to MALDI-MSI lipidomics data from mouse brain tissues representing young, aged, and AD models revealed region-specific lipid variations and candidate analytes associated with aging and disease.

Third, a Python-based workflow for spatial transcriptomic analysis using the 10x Genomics Visium platform was developed. The pipeline successfully performed preprocessing, clustering, and differential expression analysis while preserving spatial context, identifying genes associated with both AD and aging.

Finally, a multi-objective scoring (MOS) framework was developed to integrate spatial transcriptomics and MSI data without requiring exact spatial co-registration. The framework combines multiple spatial descriptors to identify lipid features with transcriptomic patterns. Together, these tools provide a scalable framework for spatial multi-omics analysis and hypothesis generation in complex biological systems.

TABLE OF CONTENTS

Chapter 1: Background and Literature Review 1

1.1 Introduction 2

1.2 Alzheimer's Disease and Aging: Molecular and Cellular Basis 5

1.3 Lipid Metabolism in the Brain and Its Role in AD 5

1.4 Mass Spectrometry Imaging for Lipidomics in Aging and AD 6

1.5 Transcriptomics and Spatial Transcriptomics in Aging and AD 7

1.6 Summary and Research Gap 8

References 10

Chapter 2: Development of an Algorithm for Mass Spectrometry Data and Analyte Delocalization Analysis 17

2.1 Introduction 18

2.2 Methodology 20

2.2.1 Methods for Assessing Delocalization 20

2.2.2 Composite Delocalization Score 41

2.3 Results and Discussion 45

2.4 Conclusion 54

2.4.1 Interpretation of Findings 54

2.4.2 Comparison to Existing Approaches 55

2.4.3 Limitations and Assumptions 56

2.4.4 Factors Contributing to Delocalization in MALDI-MSI 57

2.4.5 Broader Applicability 57
2.5 Code and Data Availability 58
2.6 Acknowledgements 59
References 60
Supplementary Material 68
Chapter 3: Designing an AI-enhanced algorithm for MSI-based lipidomics studies . 70
3.1 Introduction 71
3.2 Methodology 73
3.2.1 Sample Preparation and MALDI-MSI Acquisition 73
3.2.2 Preprocessing of MALDI-MSI Data 75
3.2.3 Dimensionality Reduction and Clustering 76
3.2.4 Quantitative Study 77
3.2.5 Finding Corresponding Lipids 77
3.3 Results and Discussion 78
3.3.1 Unsupervised Dimensionality Reduction and Clustering Reveal Group Differences 78
3.3.2 Differential Analyte Analysis Across Experimental Groups 83
3.4 Conclusion 89
References 90
Chapter 4: Spatial Transcriptomics in Aging and Alzheimer's Disease 94
4.1 Introduction 95

4.2 Methodology 97
4.2.1 Tissue Preparation 97
4.2.2 Visium Spatial Tissue Optimization Workflow 98
4.2.3 Spatial Transcriptomics 99
4.2.4 Data Analysis 99
4.3 Results and Discussion 103
4.4 Conclusion 121
References 122
Chapter 5: A Multi-Objective Scoring (MOS) Framework for Detecting Cross-Modal Spatial Similarity: Conceptual and Direct Formulations 125
5.1 Introduction 126
5.2 Methods 130
5.2.1 Sample Collection 130
5.2.2 MALDI-MSI Lipidomics 131
5.2.3 Spatial Transcriptomics 134
5.2.4 Multi-Objective Scoring Framework 136
5.3 Results and Discussion 156
5.4 Conclusion 170
5.5 Acknowledgements 171
References 173
Chapter 6: Conclusions and Future Directions 185

6.1 Conclusion 186

6.2 Future Directions 187

VITA 190

LIST OF TABLES

Table 2-1 MALDI-MSI Acquisition Parameters 22

Table 2-2. Parameter combinations used for peak detection, including filter type, smoothing parameters, and minimum peak distance (Da)..... 25

Table 3-1 DMACA matrix sublimation conditions 75

Table 3-2 MALDI timsTOF fleX instrumental parameters for positive ion mode..... 75

Table 5-1 DMACA matrix sublimation conditions 133

Table 5-2 MALDI timsTOF fleX instrumental parameters for positive ion mode..... 133

Table 5-3 Positive Control Pairs 151

Table 5-4 Negative Control Pairs (Spatially Distinct Lipid Signals) 151

Table 5-5 Optimized normalized weights of the eight spatial similarity metrics contributing to the final combined scoring function (Φ), derived through data-driven calibration using synthetic training data. 158

Table 5-6 Mean normalized weights (± SD and SE) of spatial similarity metrics used in the pattern-similarity model for identifying isotopes and adducts among MALDI-MSI peaks, evaluated using 10-fold cross-validation. 160

LIST OF FIGURES

Figure 1-1. MALDI-MSI data are collected on a 60 × 60 μm grid, with each pixel formed from 400 laser shots [https://www.biorender.com]. 3

Figure 1-2. Schematic overview of the experimental workflow and analysis pipeline used in this study. 4

Figure 1-3 (A) The Visium Spatial Gene Expression Slide has four Capture Areas, each containing ~5,000 spots with primers carrying a spatial barcode, UMI, and poly(dT) sequence for mRNA capture. (B) A Permeabilization Enzyme releases mRNA from cells, which binds to the primers on the spots [https://www.10xgenomics.com]..... 9

Figure 2-1 Schematic overview of the experimental workflow for MALDI-MSI, highlighting tissue preparation, matrix application, data acquisition, and imaging. 22

Figure 2-2 Flowchart summarizing the preprocessing steps for MALDI-MSI data implemented in Python. 23

Figure 2-3. H&E-stained tissue image overlaid with the spatial distribution heatmap of m/z 800.55. Top row (left to right): (a) Initial overlay with transparency (α = 0.5), (b) Overlay after rotation (α = 0.5), (c) Final aligned overlay (α = 0.5). Bottom row (left to right): (d) Fully opaque overlay (α = 1), (e) Overlay with annotated tissue margins (red dots and lines; α = 0), (f) Distribution of m/z 800.55 after background masking, showing signal localized exclusively to tissue regions. 27

Figure 2-4. Spatial distribution maps of six selected analyte features used to evaluate delocalization rankings based on the Euclidean distance between the global and

tissue-specific centers of mass ($dCOM$). The top row displays m/z 365.32 (rank 7), 180.07 (rank 13), and 348.07 (rank 15); the bottom row shows m/z 283.09 (rank 25), 703.57 (rank 35), and 240.10 (rank 118). 34

Figure 2-5. Spatial map showing background pixels (blue) and their nearest corresponding tissue border points (red) for m/z 651.53. Arrows (gray) illustrate the direction and distance from each sampled background pixel to its closest tissue border. BG: background pixels 35

Figure 2-6. Spatial distribution maps of six analyte features used to evaluate delocalization rankings based on the maximum Euclidean distance of nonzero background pixels to the nearest tissue border. The top row shows m/z 180.07 (rank 1), 185.07 (rank 2), and 247.24 (rank 25); the bottom row shows m/z 369.35 (rank 27), 729.57 (rank 115), and 651.53 (rank 139). 37

Figure 2-7. Spatial distribution maps of six analyte features to evaluate delocalization rankings based on the mean distance of nonzero-intensity background pixels to the nearest tissue pixel. The top row shows m/z 353.33 (rank 23), 341.31 (rank 26), and 753.59 (rank 58); the bottom row displays m/z 427.93 (rank 63), 710.49 (rank 324), and 769.56 (rank 327). 38

Figure 2-8. Spatial distribution maps of six analyte features to evaluate delocalization rankings using the area of nonzero-intensity background pixels as a delocalization metric. The top row includes m/z 799.55 (rank 259), 771.55 (rank 260), and 870.47

(rank 318); the bottom row includes m/z 714.45 (rank 319), 869.46 (rank 326), and 823.63 (rank 328). 40

Figure 2-9. Spatial distribution maps of four analyte features used to evaluate the effect of the weighting parameter α on the final delocalization score rankings. The top row shows m/z 255.21 and m/z 651.53, while the bottom row includes m/z 761.59 and m/z 406.95 43

Figure 2-10. Delocalization scores of analyte groups ranked by their mean score across samples. The top panel shows ranks 1–71 (most delocalized), and the bottom panel shows ranks 72–142. Bars represent mean scores with standard error; red dots indicate individual sample values. 46

Figure 2-11. Pairwise correlations between normalized delocalization metrics across analyte features. Scatter plots include linear regression lines (red), with histograms on the diagonal showing each metric's distribution. 48

Figure 2-12. Principal component analysis (PCA) of delocalization score profiles across samples. Samples are projected onto the first two principal components based on their delocalization score patterns across analyte groups. Each point represents a sample, colored by experimental group. 49

Figure 2-13 Pearson correlation heatmap of delocalization score profiles across samples. Pairwise correlation coefficients (r) are computed between all samples based on their delocalization scores across analyte groups. Warmer colors indicate stronger

positive correlations. The high within-group similarity suggests consistent spatial distribution patterns among replicates. 51
Figure 2-14 Heatmap of delocalization scores across analyte groups and samples. Each row represents an analyte group and each column corresponds to a sample. Color intensity reflects the delocalization score, with brighter values indicating stronger spatial delocalization. This visualization highlights sample-specific and group-specific variation in lipid spatial patterns. 52
Figure 2-15 Spatial distribution of m/z 844.52 in young control samples (from left to right: YC_1, YC_2, YC_5) with corresponding H&E-stained sections. The first row shows ion intensity maps of m/z 844.52, and the second row shows the corresponding histological images. 68
Figure 2-16 Optical images of all stained tissue sections. Sample groups include Young Control (YC), Young AD (YAD), Aged Control (AC), and Aged AD (AAD). 69
Figure 3-1 Schematic overview of the experimental workflow for MALDI-MSI, highlighting tissue preparation, matrix application, data acquisition, and imaging. 75
Figure 3-2 Flowchart summarizing the preprocessing steps for MALDI-MSI data implemented in Python. 76
Figure 3-3 TIC-normalized spatial distribution maps of analyte 921.69 across all samples after filtering background, probable contamination, and noise pixels, with intensity scaled to the 99.9th percentile (Vmax). 79

Figure 3-4 TIC-normalized spatial distribution maps of analyte 921.69 across all samples without filtering for background, probable contamination, or noise pixels..... 80
Figure 3-5 PCA variance ratio plot of the first 50 components, with variance values shown on a logarithmic scale..... 81
Figure 3-6 Dimensionality reduction visualization using UMAP (left) and t-SNE (right), with data points colored according to group identity..... 81
Figure 3-7 Leiden clustering of the dataset at a resolution of 1.5, resulting in 38 distinct clusters that separate pixels according to their analyte intensity profiles..... 82
Figure 3-8 Top contributing analyte features for each cluster (clusters 0-19) were identified based on their corresponding scores, highlighting the analyte signals driving regional distinctions across the tissue and samples..... 84
Figure 3-9 Top contributing analyte features for each cluster (clusters 20-38) were identified based on their corresponding scores, highlighting the analyte signals driving regional distinctions across the tissue and samples..... 85
Figure 3-10 Heatmap of common analyte features (758.56–798.54) showing $\log_2$ fold changes across group comparisons. Statistical significance is indicated as $p < 0.05$ (*), $p < 0.01$ (**), and $p < 0.001$ (***) 86
Figure 3-11 Volcano plots for each group comparison (comparisons 1-12), showing $\log_2$ fold changes versus – $\log_{10}$ p-values. Significantly upregulated and downregulated analytes are highlighted, with the top five most significant features labeled 87

Figure 3-12 Volcano plots for each group comparison (comparisons 13-20), showing $\log_2$ fold changes versus – $\log_{10}$ p-values. Significantly upregulated and downregulated analytes are highlighted, with the top five most significant features labeled 88
Figure 4-1 Flowchart of Visium tissue optimization, spatial gene expression analysis, and data preprocessing pipeline. TO: tissue optimization 102
Figure 4-2 Violin plots displaying the number of genes detected per cell (*n_genes_by_counts*; left), total counts per cell (middle), and percentage of mitochondrial counts (*pct_counts_mt*; right) across samples following preprocessing and filtering. 104
Figure 4-3 Objective selection of principal components based on explained variance .. 105
Figure 4-4 PCA-Loading-Based Refinement of Highly Variable Genes 105
Figure 4-5 UMAP visualization of transcriptional structure (left) clustered by group and (right) Leiden clustering. AAD: aged AD, AC: aged control, YAD: young AD, YC: young control 106
Figure 4-6 Upregulated genes in the young disease comparison (YAD vs. YC). (a) Dotplot illustrating the top upregulated genes in this comparison; (b) volcano plot highlighting differentially expressed genes meeting the significance criteria (adjusted p-value < 0.05 and $\log_2$FC > 0.5); and (c) statistics heatmap summarizing the $\log_2$ fold change and adjusted p-values for the top upregulated genes in the young disease comparison. Significance levels are indicated as follows: $p_{adj} < 0.05$ (*), < 0.01 (**), < 0.001 (***), and ns (not significant; $p_{adj} \geq 0.05$). 108

Figure 4-7 Downregulated genes in the young disease comparison (YAD vs. YC). (a) Dotplot illustrating the top downregulated genes in this comparison; (b) volcano plot highlighting differentially expressed genes meeting the significance criteria (adjusted p-value < 0.05 and $log_2FC < -0.5$); and (c) statistics heatmap summarizing the log_2 fold change and adjusted p-values for the top downregulated genes in the young disease comparison. Significance levels are indicated as follows: $p_{adj} < 0.05$ (*), < 0.01 (**), < 0.001 (***), and ns (not significant; $p_{adj} \geq 0.05$). 109

Figure 4-8 Upregulated genes in the aged disease comparison (AAD vs. AC). (a) Dotplot illustrating the top upregulated genes in this comparison; (b) volcano plot highlighting differentially expressed genes meeting the significance criteria (adjusted p-value < 0.05 and $log_2FC > 0.5$); and (c) statistics heatmap summarizing the log_2 fold change and adjusted p-values for the top upregulated genes in the aged disease comparison. Significance levels are indicated as follows: $p_{adj} < 0.05$ (*), < 0.01 (**), < 0.001 (***), and ns (not significant; $p_{adj} \geq 0.05$). 111

Figure 4-9 Downregulated genes in the aged disease comparison (AAD vs. AC). (a) Dotplot illustrating the top downregulated genes in this comparison; (b) volcano plot highlighting differentially expressed genes meeting the significance criteria (adjusted p-value < 0.05 and $log_2FC > 0.5$); and (c) statistics heatmap summarizing the log_2 fold change and adjusted p-values for the top downregulated genes in the aged disease comparison. Significance levels are indicated as follows: $p_{adj} < 0.05$ (*), < 0.01 (**), < 0.001 (***), and ns (not significant; $p_{adj} \geq 0.05$). 112

Figure 4-10 Upregulated genes in the ageing effect comparison (AC vs. YC). (a) Dotplot illustrating the top 20 upregulated genes in this comparison; (b) volcano plot highlighting differentially expressed genes meeting the significance criteria (adjusted p-value < 0.05 and $log_2FC > 0.5$); and (c) statistics heatmap summarizing the log_2 fold change and adjusted p-values for the top 20 upregulated genes in the ageing effect comparison. Significance levels are indicated as follows: $p_{adj} < 0.05$ (*), < 0.01 (**), < 0.001 (***), and ns (not significant; $p_{adj} \geq 0.05$). 114

Figure 4-11 Downregulated genes in the ageing effect comparison (AC vs. YC). (a) Dotplot illustrating the top 20 downregulated genes in this comparison; (b) volcano plot highlighting differentially expressed genes meeting the significance criteria (adjusted p-value < 0.05 and $log_2FC < -0.5$); and (c) statistics heatmap summarizing the log_2 fold change and adjusted p-values for the top 20 downregulated genes in the ageing effect comparison. Significance levels are indicated as follows: $p_{adj} < 0.05$ (*), < 0.01 (**), < 0.001 (***), and ns (not significant; $p_{adj} \geq 0.05$). 115

Figure 4-12 Combined overview of upregulated genes across comparisons. (a) Grouped dotplot showing up to the top 10 upregulated genes for each contrast (aging effect in control groups, AD effect in young groups, and AD effect in aged groups), organized by comparison. (b) Heatmap summarizing log_2 fold change values of the selected genes across all contrasts, with significance indicated by stars. (c) Table presenting detailed differential expression statistics, including p-values, adjusted p-values, and log_2 fold change for each gene across comparisons. Significance levels

are indicated as follows: $p_{adj} < 0.05$ (*), < 0.01 (**), < 0.001 (***), and ns (not significant; $p_{adj} \geq 0.05$). 117
Figure 4-13 Combined overview of downregulated genes across comparisons. (a) Grouped dotplot showing up to the top 10 downregulated genes for each contrast (aging effect in control groups, AD effect in young groups, and AD effect in aged groups), organized by comparison. (b) Heatmap summarizing log_2 fold change values of the selected genes across all contrasts, with significance indicated by stars. (c) Table presenting detailed differential expression statistics, including p-values, adjusted p-values, and log_2 fold change for each gene across comparisons. Significance levels are indicated as follows: $p_{adj} < 0.05$ (*), < 0.01 (**), < 0.001 (***), and ns (not significant; $p_{adj} \geq 0.05$). 118
Figure 4-14 Spatial autocorrelation analysis of AD candidate genes using Moran's I. Red bars indicate spatially significant genes (I > 0.1 and padj < 0.05) and gray bars indicating non-significant genes. The dashed line represents the Moran's I threshold (I = 0.1). 119
Figure 4-15 Heatmap of the spatial distribution of Thy1 across all samples, illustrating region-specific expression patterns and differences among aging, AD, and control conditions.(Spots are bigger than their normal size for better visualization) 120
Figure 5-1 Schematic illustration of the glass microscope slide, Visium Spatial Gene Expression slide, and Visium Spatial Tissue Optimization slide (from left to right). 131

Figure 5-2 Schematic overview of the experimental workflow for MALDI-MSI, highlighting tissue preparation, matrix application, data acquisition, and imaging. 133
Figure 5-3 Flowchart summarizing the computational steps for MALDI-MSI data implemented in Python. 134
Figure 5-4 Flowchart of Visium tissue optimization, spatial gene expression analysis, and data preprocessing pipeline. TO: tissue optimization 137
Figure 5-5 schematic demonstration of sibling resemblance..... 139
Figure 5-6 Schematic diagram of the MOS framework..... 141
Figure 5-7 Gene–m/z consensus heatmap. Columns represent genes and rows show the top 10 ranked m/z candidates. Color indicates confidence (fraction of animals where the feature was detected in the top 5%). Each cell reports the m/z value, consensus score (S), and detection frequency (X/Y). Asterisks (*) denote samples where isotopic or adduct peaks were detected without the M_{C12} peak. 162
Figure 5-8 Spatial heatmap of *Mbp* expression across all samples; spots enlarged for easier visualization..... 164
Figure 5-9 TIC-normalized spatial distribution heatmaps of analyte 842.6652 across all half-brain samples with filtering of background, probable contamination, or noise pixels 165
Figure 5-10 Spatial distribution heatmap of *Mbp* across a representative sample, shown with both true spatial spot size and enlarged spots for visualization, alongside the

distributions of the 10 matched m/z candidate features identified using the MOS framework. Numbers indicate the ranking of the candidate features. 166

Figure 5-11 Spatial distribution heatmap of *Thy1* expression across all samples; spots enlarged for easier visualization 167

Figure 5-12 TIC-normalized spatial distribution heatmaps of analyte 820.52 across all half-brain samples with filtering of background, probable contamination, or noise pixels 168

Figure 5-13 Spatial distribution heatmap of *Hspa1a* expression across a representative sample (left), with spots enlarged for clearer visualization, and the spatial distribution heatmap of its proposed candidate analyte feature m/z 769.59 (right), identified using the MOS framework. 170

CHAPTER 1:

BACKGROUND AND LITERATURE REVIEW

1.1 Introduction

The spatial heterogeneity, variation in molecular and cellular characteristics within tissue, is crucial for understanding human disease. Mapping this spatial heterogeneity provides insight into disease initiation, progression, and potential therapeutic targets, enabling analysis of cellular neighborhoods and molecular interactions. Techniques such as mass spectrometry imaging (MSI) and spatial RNA-sequencing are essential for generating these molecular maps. Lipids, which can be measured with MSI, are vital to many biological functions, including cell structuring, energy storage, and cell signaling, and there is strong evidence that indicates neurological diseases are correlated with changes in lipid metabolism. Ribonucleic acid (RNA) sequencing, including spatial transcriptomics, can measure gene expression and often provides a more complete representation than DNA sequencing because RNA transcription plays a critical role in cell function and aging.

While lipidomics and transcriptomics each provide valuable insights into pathology, studying them independently overlooks the molecular interactions that drive disease progression. For example, lipid metabolism genes such as *Apoe* directly influence lipid homeostasis, while lipid signaling can modulate gene expression via transcription factors [1].

Spatial multi-omics integration enables:

- Correlation of lipid species with gene expression patterns in situ.

- Identification of pathology-associated molecular neighborhoods (e.g., amyloid plaques for Alzheimer's disease).
- Construction of multi-layered disease models for biomarker discovery.

However, integrating MALDI-MSI and spatial transcriptomic data poses significant technical challenges due to differences in spatial coordinates and resolution. Visium barcoded capture spots are arranged in a hexagonal grid with ~100 µm center-to-center spacing, whereas MALDI-MSI data is collected in a Cartesian grid. An example grid is shown below with 60 × 60 µm pixels, each consisting of 400 laser shots with a 3 µm diameter (Figure 1-1). These differences in pixel size, shape, and spatial alignment complicate fusion of the datasets. To our knowledge, no prior studies have fully addressed this challenge. In this thesis, I developed a computational pipeline for MSI–spatial transcriptomic integration, creating the first comprehensive spatial multi-omics maps of aging and AD, and enabling a deeper understanding of the molecular architecture underlying disease progression (Figure 1-2).

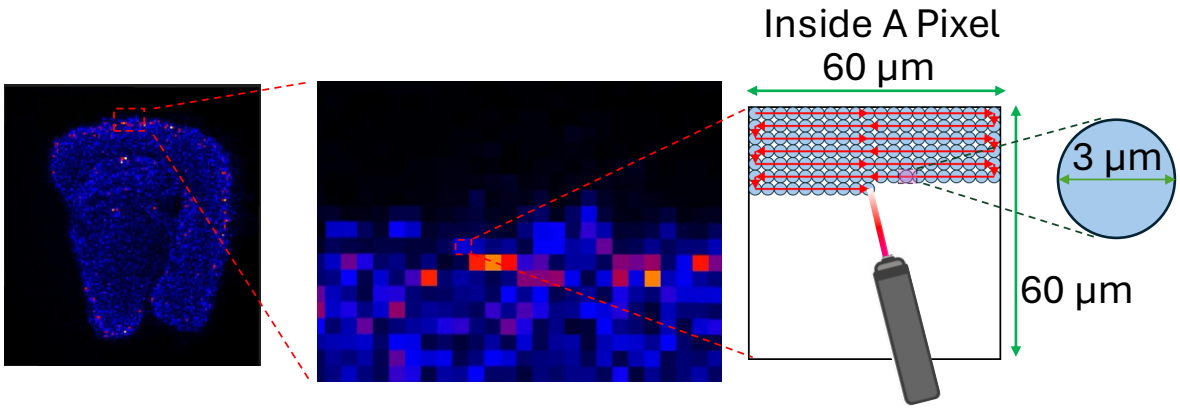


Figure 1-1. MALDI-MSI data are collected on a 60 × 60 µm grid, with each pixel formed from 400 laser shots [https://www.biorender.com].

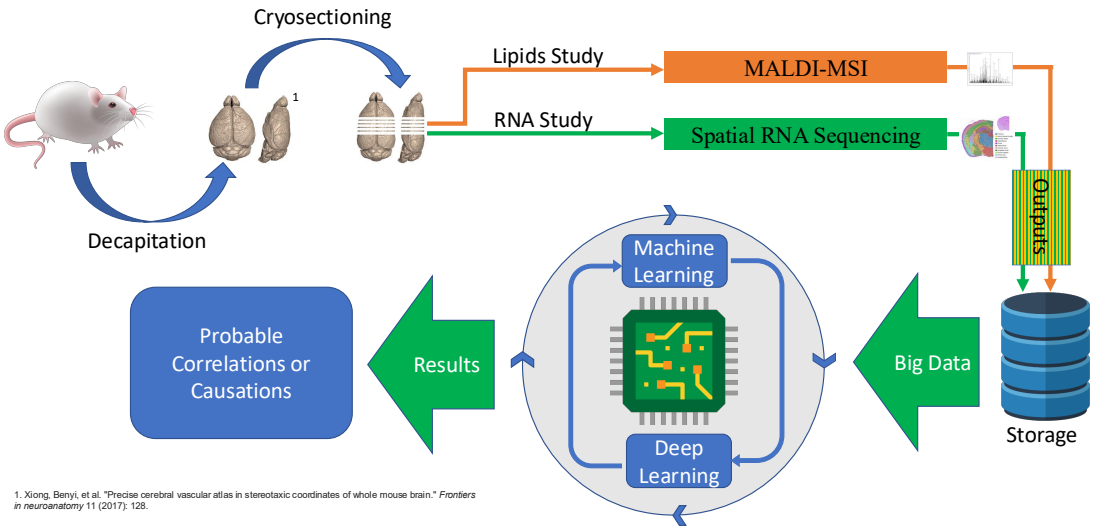


Figure 1-2. Schematic overview of the experimental workflow and analysis pipeline used in this study.

To develop this workflow, a neurological disease, Alzheimer's disease (AD), was chosen because of the extensive literature on lipids and transcriptomics. AD is the sixth leading cause of death in the United States and the fifth leading cause of death among Americans older than 65 [2]. After age 65, the incidence of AD doubles approximately every five years, with nearly half of individuals aged 85 or older affected [3]. AD is the most common cause of dementia, beginning with mild memory loss and progressing to the loss of the ability to carry out daily activities, as thought, memory, and language are controlled by brain regions damaged by the disease [4, 5]. AD has no known cure [6-8] and remains a devastating condition for patients and caregivers.

Molecular changes underlying AD initiation and progression remain incompletely understood. I developed spatial omics techniques to identify alterations in the organization, pattern, and abundance of biomolecules in the brain during AD and aging.

Understanding these molecular mechanisms is critical for discovering potential diagnostic and therapeutic strategies.

1.2 Alzheimer's Disease and Aging: Molecular and Cellular Basis

AD is characterized by alterations in molecular and cellular processes involving DNA, RNA, proteins, metabolites, and lipids. Hallmarks include the accumulation of amyloid-β (Aβ) and tau proteins in the brain [9-11]. Despite decades of research, no FDA-approved medication effectively targets these processes. Astrocyte populations associated with AD have been observed in both aged mice without disease and in aging human brains, suggesting links to genetic and age-related factors [12].

Metabolic changes are also implicated in AD, including alterations in small molecules such as N-acetylaspartate, creatine, glutathione, and myo-inositol [13-17]. The mechanisms linking these metabolic changes to AD pathology remain unclear. Proteins including Aβ, tau, apolipoprotein E, and alpha-synuclein are critical to disease progression [13, 18-23]. Transcriptomic alterations, including changes in gene expression for synaptic plasticity, neuroinflammation, and protein homeostasis, as well as dysregulation of long non-coding and circular RNAs, further contribute to AD pathology [24-28].

1.3 Lipid Metabolism in the Brain and Its Role in AD

The brain is highly enriched in lipids, which comprise at least 50% of its dry weight. Neuronal membranes are primarily lipid-based, with phospholipids as the major component and significant contributions from cholesterol and glycolipids. Long-chain

polyunsaturated fatty acids (LC-PUFAs), including docosahexaenoic acid (DHA) and arachidonic acid (AA), are important constituents of brain lipids [29].

Aging disrupts lipid homeostasis and contributes to the development of AD [30–32]. Lipid dysregulation has been linked to processes such as mitochondrial dysfunction, oxidative stress, inflammation, and synaptic loss [30]. Spatial lipid mapping using MSI enables visualization of lipid distributions within brain tissue and supports integrative analyses with transcriptomic data.

1.4 Mass Spectrometry Imaging for Lipidomics in Aging and AD

Mass spectrometry (MS) techniques, including MSI and liquid chromatography-mass spectrometry (LC-MS), involve ionizing a sample, separating ions by their mass-to-charge ratio (m/z), and detecting them to generate spectra representing each ion's relative abundance. Lipids can be identified in tissues and body fluids using MS-based methods such as liquid chromatography-mass spectrometry (LC-MS) or gas chromatography-mass spectrometry [33, 34].

MALDI-MSI is a form of MSI where the ion source is rastered across the sample surface to generate spatial heat maps of molecule abundance. It is particularly suited for studying peptides, lipids, and metabolites in biological tissues [35-40]. A matrix is applied to aid the extraction and ionization of molecules, which can be applied manually, via sublimation, or with an automated sprayer; in this study, the HTX M3+ automated sprayer was used for the first aim, and HTX SubliMATE was used for the rest of the aims.

MALDI-MSI produces a hyperspectral image where each pixel contains a full mass spectrum, generating a data matrix of *spatial coordinates × analyte (m/z) features.*

Key challenges in MSI lipidomics:

- High dimensionality (thousands of analyte features per sample).
- Ion suppression effects and baseline drift.
- Need for accurate normalization (e.g., total ion current (TIC) normalization) and peak detection.
- Lack of standardized computational pipelines for lipid annotation and spatial analysis.

1.5 Transcriptomics and Spatial Transcriptomics in Aging and AD

Traditional RNA sequencing has revealed profound transcriptomic alterations in AD, including immune activation, synaptic dysfunction, and metabolic reprogramming [41]. However, bulk or single-cell approaches typically lack spatial context, limiting the ability to link gene expression changes to histopathology. Spatial transcriptomics bridges this gap by mapping gene expression profiles back to their tissue coordinates. 10x Genomics Visium is a widely adopted Spatial transcriptomics platform that uses barcoded capture spots (~55 μm diameter) on glass slides to retain spatial information. Each spot captures transcripts from adjacent cells, generating spatially resolved gene expression maps (Figure 1-3A). In AD research, spatial transcriptomics has revealed region-specific transcriptional programs, such as microglial activation near amyloid plaques and

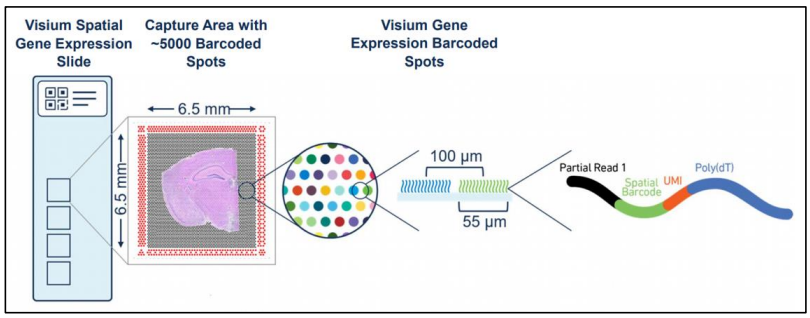
astrocytic responses in white matter [42], highlighting the importance of spatial context in understanding neurodegeneration.

Spatial RNA sequencing allows for high-resolution mapping of gene expression while preserving tissue architecture. Optimizing tissue permeabilization is a critical step; Visium Spatial Tissue Optimization slides are used to determine the optimal permeabilization time, testing 3, 6, 12, 18, 24, and 30 minutes (Figure 1-3B). For subsequent analyses, brain slices are mounted on Visium Gene Expression (GEX) slides containing arrays of barcoded oligonucleotides to capture the spatial location of RNA transcripts. RNA is extracted and reverse-transcribed into cDNA, and sequencing performed on the Illumina NovaSeq platform at the University of Tennessee Genomic Core. Sequencing data are analyzed for differentially expressed genes and their spatial distributions using Space Ranger (10X Genomics), enabling detailed mapping of transcriptomic changes associated with aging and AD.

1.6 Summary and Research Gap

Current biochemistry research, including in AD research is limited by the lack of spatially integrated multi-omics approaches. MSI provides high-resolution lipid maps, while spatial transcriptomics maps gene expression domains, yet no established pipeline exists for their integration in neurodegenerative research. This thesis proposes a systematic framework (including tools) combining MALDI-MSI lipidomics and Visium-based spatial transcriptomics, enabling spatial correlation analysis, multi-omics factor modeling, and deep learning integration for biomarker discovery.

A



B

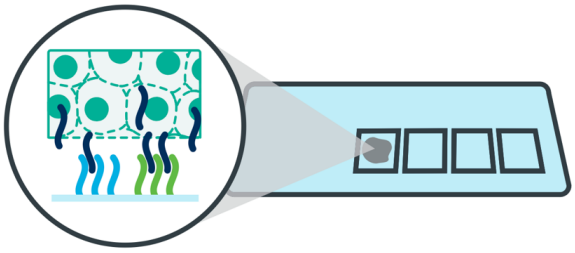


Figure 1-3 (A) The Visium Spatial Gene Expression Slide has four Capture Areas, each containing ~5,000 spots with primers carrying a spatial barcode, UMI, and poly(dT) sequence for mRNA capture. (B) A Permeabilization Enzyme releases mRNA from cells, which binds to the primers on the spots [https://www.10xgenomics.com].

REFERENCES

1. Lefterov I, Wolfe CM, Fitz NF, Nam KN, Letronne F, Biedrzycki RJ, Kofler J, Han X, Wang J, Schug J, Koldamova R. APOE2 orchestrated differences in transcriptomic and lipidomic profiles of postmortem AD brain. Alzheimer's research & therapy. 2019 Dec 30;11(1):113.
2. 2023 Alzheimer's disease facts and figures. Alzheimers Dement [Internet]. 2023 [cited 2023 Apr 21];19(4):1598–695. Available from: https://pubmed.ncbi.nlm.nih.gov/36918389/
3. Liu RM. Aging, cellular senescence, and Alzheimer's disease. International Journal of Molecular Sciences. 2022 Feb 11;23(4):1989.
4. Finch CE, Kulminski AM. The Alzheimer's disease exposome. Alzheimer's & Dementia. 2019 Sep 1;15(9):1123-32.
5. Yousaf K, Mehmood Z, Awan IA, Saba T, Alharbey R, Qadah T, Alrige MA. A comprehensive study of mobile-health based assistive technology for the healthcare of dementia and Alzheimer's disease (AD). Health Care Management Science. 2020 Jun;23:287-309.
6. Srivastava S, Ahmad R, Khare SK. Alzheimer's disease and its treatment by different approaches: A review. European Journal of Medicinal Chemistry. 2021 Apr 15;216:113320.

7. Abeysinghe AA, Deshapriya RD, Udawatte C. Alzheimer's disease; a review of the pathophysiological basis and therapeutic interventions. Life sciences. 2020 Sep 1;256:117996.
8. Jha NK, Jha SK, Kar R, Nand P, Swati K, Goswami VK. Nuclear factor-kappa β as a therapeutic target for Alzheimer's disease. Journal of neurochemistry. 2019 Jul;150(2):113-37.
9. Maggio JE, Stimson ER, Ghilardi JR, Allen CJ, Dahl CE, Whitcomb DC, Vigna SR, Vinters HV, Labenski ME, Mantyh PW. Reversible in vitro growth of Alzheimer disease beta-amyloid plaques by deposition of labeled amyloid peptide. Proceedings of the National Academy of Sciences. 1992 Jun 15;89(12):5462-6.
10. Goedert M. Tau protein and the neurofibrillary pathology of Alzheimer's disease. Trends in neurosciences. 1993 Nov 1;16(11):460-5.
11. Martin L, Latypova X, Wilson CM, Magnaudeix A, Perrin ML, Yardin C, Terro F. Tau protein kinases: involvement in Alzheimer's disease. Ageing research reviews. 2013 Jan 1;12(1):289-309.
12. Habib N, McCabe C, Medina S, Varshavsky M, Kitsberg D, Dvir-Szternfeld R, Green G, Dionne D, Nguyen L, Marshall JL, Chen F. Disease-associated astrocytes in Alzheimer's disease and aging. Nature neuroscience. 2020 Jun;23(6):701-6.

13. Roalf DR, Sydnor VJ, Woods M, Wolk DA, Scott JC, Reddy R, Moberg PJ. A quantitative meta-analysis of brain glutamate metabolites in aging. Neurobiology of aging. 2020 Nov 1;95:240-9.
14. Nativio R, Lan Y, Donahue G, Sidoli S, Berson A, Srinivasan AR, Shcherbakova O, Amlie-Wolf A, Nie J, Cui X, He C. An integrated multi-omics approach identifies epigenetic alterations associated with Alzheimer's disease. Nature genetics. 2020 Oct;52(10):1024-35.
15. Chen Q, Abrigo J, Liu W, Han EY, Yeung DK, Shi L, Au LW, Deng M, Chen S, Leung EY, Ho CL. Lower posterior cingulate N-acetylaspartate to creatine level in early detection of biologically defined Alzheimer's disease. Brain Sciences. 2022 May 31;12(6):722.
16. Chen L, van Zijl PC, Wei Z, Lu H, Duan W, Wong PC, Li T, Xu J. Early detection of Alzheimer's disease using creatine chemical exchange saturation transfer magnetic resonance imaging. NeuroImage. 2021 Aug 1;236:118071.
17. Mandal PK, Shukla D, Tripathi M, Ersland L. Cognitive improvement with glutathione supplement in Alzheimer's disease: a way forward. Journal of Alzheimer's Disease. 2019 Jan 1;68(2):531-5.
18. Jeremic D, Jiménez-Díaz L, Navarro-López JD. Past, present and future of therapeutic strategies against amyloid-β peptides in Alzheimer's disease: A systematic review. Ageing research reviews. 2021 Dec 1;72:101496.

19. Hilbich C, Kisters-Woike B, Reed J, Masters CL, Beyreuther K. Aggregation and secondary structure of synthetic amyloid βA4 peptides of Alzheimer's disease. Journal of molecular biology. 1991 Mar 5;218(1):149-63.
20. Koutsodendris N, Nelson MR, Rao A, Huang Y. Apolipoprotein e and alzheimer's disease: Findings, hypotheses, and potential mechanisms. Annual Review of Pathology: Mechanisms of Disease. 2022 Jan 24;17:73-99.
21. Yamazaki Y, Zhao N, Caulfield TR, Liu CC, Bu G. Apolipoprotein E and Alzheimer disease: pathobiology and targeting strategies. Nature Reviews Neurology. 2019 Sep;15(9):501-18.
22. Hashimoto M, Masliah E. Alpha-synuclein in Lewy body disease and Alzheimer's disease. Brain pathology. 1999 Oct;9(4):707-20.
23. Twohig D, Nielsen HM. α-synuclein in the pathophysiology of Alzheimer's disease. Molecular neurodegeneration. 2019 Dec;14(1):1-9.
24. Reza-Zaldivar EE, Hernández-Sápiens MA, Minjarez B, Gómez-Pinedo U, Sánchez-González VJ, Márquez-Aguirre AL, Canales-Aguirre AA. Dendritic spine and synaptic plasticity in Alzheimer's disease: a focus on MicroRNA. Frontiers in Cell and Developmental Biology. 2020 May 5;8:255.
25. Castro-Hernández R, Berulava T, Metelova M, Epple R, Peña Centeno T, Richter J, Kaurani L, Pradhan R, Sakib MS, Burkhardt S, Ninov M. Conserved reduction of m6A RNA modifications during aging and neurodegeneration is linked to

changes in synaptic transcripts. Proceedings of the National Academy of Sciences. 2023 Feb 28;120(9):e2204933120.

26. Liu Y, Cheng X, Li H, Hui S, Zhang Z, Xiao Y, Peng W. Non-Coding RNAs as Novel Regulators of Neuroinflammation in Alzheimer's Disease. Frontiers in Immunology. 2022;13.
27. Zhang Y, Zhao Y, Ao X, Yu W, Zhang L, Wang Y, Chang W. The role of non-coding RNAs in Alzheimer's disease: from regulated mechanism to therapeutic targets and diagnostic biomarkers. Frontiers in aging neuroscience. 2021 Jul 2;13:654978.
28. Adusumalli S, Ngian ZK, Lin WQ, Benoukraf T, Ong CT. Increased intron retention is a post-transcriptional signature associated with progressive aging and Alzheimer's disease. Aging cell. 2019 Jun;18(3):e12928.
29. Sastry PS. Lipids of nervous tissue: composition and metabolism. Progress in lipid research. 1985 Jan 1;24(2):69-176.
30. Elobeid A, Libard S, Leino M, Popova SN, Alafuzoff I. Altered proteins in the aging brain. Journal of Neuropathology & Experimental Neurology. 2016 Apr 1;75(4):316-25.
31. Abad-Rodriguez J, Ledesma MD, Craessaerts K, Perga S, Medina M, Delacourte A, Dingwall C, De Strooper B, Dotti CG. Neuronal membrane cholesterol loss enhances amyloid peptide generation. The Journal of cell biology. 2004 Dec 6;167(5):953-60.

32. Frank C, Rufini S, Tancredi V, Forcina R, Grossi D, D'arcangelo G. Cholesterol depletion inhibits synaptic transmission and synaptic plasticity in rat hippocampus. Experimental neurology. 2008 Aug 1;212(2):407-14.
33. Ellis SR, Paine MR, Eijkel GB, Pauling JK, Husen P, Jervelund MW, Hermansson M, Ejsing CS, Heeren RM. Automated, parallel mass spectrometry imaging and structural identification of lipids. Nature methods. 2018 Jul;15(7):515-8.
34. Nygren H, Seppänen-Laakso T, Castillo S, Hyötyläinen T, Orešič M. Liquid chromatography-mass spectrometry (LC-MS)-based lipidomics for studies of body fluids and tissues. Metabolic Profiling: Methods and Protocols. 2011:247-57.
35. Cerruti CD, Benabdellah F, Laprévote O, Touboul D, Brunelle A. MALDI imaging and structural analysis of rat brain lipid negative ions with 9-aminoacridine matrix. Analytical chemistry. 2012 Mar 6;84(5):2164-71.
36. Miyake Y, Kusaka S, Murata I, Toyoda M. Matrix-Assisted Laser Desorption/Ionization (MALDI) Mass Spectrometry Imaging of L-4-Phenylalanineboronic Acid (BPA) in a Brain Tumor Model Rat for Boron Neutron Capture Therapy (BNCT). Mass Spectrometry. 2022 Dec 20;11(1):A0105-.
37. Ntshangase S, Mdanda S, Naicker T, Kruger HG, Baijnath S, Govender T. Spatial distribution of elvitegravir and tenofovir in rat brain tissue: Application of matrix-

assisted laser desorption/ionization mass spectrometry imaging and liquid chromatography/tandem mass spectrometry. Rapid Communications in Mass Spectrometry. 2019 Nov 15;33(21):1643-51.

38. Stoeckli M, Staab D, Staufenbiel M, Wiederhold KH, Signor L. Molecular imaging of amyloid β peptides in mouse brain sections using mass spectrometry. Analytical biochemistry. 2002 Dec 1;311(1):33-9. A
39. Steven RT, Bunch J. Repeat MALDI MS imaging of a single tissue section using multiple matrices and tissue washes. Analytical and bioanalytical chemistry. 2013 May;405:4719-28.
40. Heijs B, Tolner EA, Bovée JV, van den Maagdenberg AM, McDonnell LA. Brain region-specific dynamics of on-tissue protein digestion using MALDI mass spectrometry imaging. Journal of proteome research. 2015 Dec 4;14(12):5348-54.
41. Mathys H, Davila-Velderrain J, Peng Z, Gao F, Mohammadi S, Young JZ, Menon M, He L, Abdurrob F, Jiang X, Martorell AJ. Single-cell transcriptomic analysis of Alzheimer's disease. Nature. 2019 Jun 20;570(7761):332-7.
42. Chen WT, Lu A, Craessaerts K, Pavie B, Frigerio CS, Corthout N, Qian X, Laláková J, Kühnemund M, Voytyuk I, Wolfs L. Spatial transcriptomics and in situ sequencing to study Alzheimer's disease. Cell. 2020 Aug 20;182(4):976-91.

CHAPTER 2:

DEVELOPMENT OF AN ALGORITHM FOR MASS SPECTROMETRY DATA AND ANALYTE DELOCALIZATION ANALYSIS

A version of this chapter was originally published as: Amin Jarrahi, Allison Jones, Weisheng Tang, Hairong Qi, and Anna Colleen Crouch, "Mathematical Framework for Quantifying Delocalization in MALDI-MSI via a Composite Scoring Approach," *ACS Measurement Science Au,* 2026, 6 (1), 134–149. https://doi.org/10.1021/acsmeasuresciau.5c00148 (published January 15, 2026). This chapter is reproduced directly from the published article without substantive modification. The author, Amin Jarrahi, led the development of the computational framework, including the design and implementation of the delocalization metrics, and was responsible for the computational methods, sample preparation, experimental work, data collection, data analysis, interpretation of results, and code development. Co-authors contributed as follows: Allison Jones assisted with sample preparation; Weisheng Tang and Hairong Qi provided guidance and feedback on code development; and Anna Colleen Crouch supervised the project and contributed to experimental design and manuscript preparation.

2.1 Introduction

Mass spectrometry imaging (MSI) is an advanced molecular imaging technique that ionizes molecules directly from biological specimens and captures the spatial distribution of them across the sample [1-4]. Unlike conventional imaging methods such as light and fluorescence microscopy [5] with immunohistochemistry and histology [6, 7], MSI operates without the need for labeling analytes. MSI enables untargeted detection of a wide array of molecular ions within a sample. Various ion sources have been used with

MSI including desorption electrospray ionization (DESI) [8-10], secondary ion mass spectrometry (SIMS) [11-13], and matrix-assisted laser desorption/ionization (MALDI) [1, 14, 15].

MALDI MSI can be used for both fresh frozen and formalin-fixed paraffin-embedded (FFPE) tissues, making it compatible with clinical and archival samples. MALDI achieves high spatial resolution (typically 5–200 μm), causes minimal tissue damage (soft ionization), and enables the detection of a broad spectrum of biomolecules, including metabolites, lipids, peptides, and proteins up to 50 kDa [16-18]. These features make MALDI particularly well-suited for a wide range of biological and biomedical applications. MALDI requires the application of a matrix which absorbs the energy from a laser and facilitates the desorption and ionization of analyte molecules into stable gas-phase ions [5, 19]. A wide range of matrices has been developed for MALDI applications, generally classified into three main categories: conventional organic matrices, liquid crystalline matrices, and inorganic matrices [19].

With the application of a matrix, analyte delocalization can occur. Delocalization remains one of the primary challenges in MALDI-MSI, significantly impacting both sensitivity and spatial resolution [20-22]. In recent years, considerable research has focused on mitigating this issue by optimizing various factors such as tissue preparation, storage conditions, handling protocols, and matrix application techniques [3, 23-26]. Despite these efforts, there are few techniques to quantitatively assess analyte delocalization. Specifically, it remains unclear which measurable parameters can be used to compare the

extent of delocalization between two analytes. A standardized criteria to determine when one analyte is more delocalized than another analyte is needed.

In this study, we developed a Python-based technique for assessing delocalization across the entire m/z (analyte) range acquired. We used previously imaged data of murine brains to evaluate various variables as potential indicators of analyte delocalization and identified those most effective at distinguishing between different degrees of delocalization. These key variables were then incorporated into a quantitative formula to describe delocalization more precisely. Furthermore, we used the formula to study whether a given analyte value exhibits a consistent delocalization score across samples within a single group as well as across different groups. This technique can be used to assess delocalization for any MSI study if the data is available in the imzML format.

2.2 Methodology

2.2.1 Methods for Assessing Delocalization

2.2.1.1 Tissue samples

In this study, 3xTg-AD mice (MMRRC Strain #034830-JAX) (n=12) and age-matched controls (n=12) from Jackson Laboratory were used [27]. For each mouse strain, mice were either sacrificed at 8 weeks of age (n=6/strain) or 54 weeks of age (n=6/strain). All animals were euthanized via conscious decapitation, and the brains were rapidly frozen in liquid nitrogen within two minutes post-mortem. Then, the samples were wrapped in aluminum foil, placed in Falcon tubes, transported on dry ice, and stored in a −80 °C freezer. All procedures were performed in accordance with the guidelines of the

Institutional Animal Care and Use Committee (IACUC) at the University of Tennessee. Brains were halved and coronally sectioned at a thickness of 10 μm using Leica Cryostat. The location of each brain slice was determined using a standard mouse brain atlas [28], and tissue samples were collected at a location approximately -2.24 ± 0.21 mm from the bregma and placed on glass microscope slides (Fisher Scientific). Each slide contained brain sections from the six animals within the same experimental group. After sectioning, the slides were placed in plastic slide jars and transferred to -80 °C freezer for storage. The jar caps were tightly sealed to prevent moisture from entering. Samples were transferred in dry ice at all stages.

2.2.1.2 Matrix Application/Spray Parameters

Before matrix application, the slides were taken out of the slide holder and placed in a desiccator immediately and allowed to thaw at room temperature for 30 minutes. The distance between the -80 °C freezer and the desiccator was less than 10 m. 2,5-Dihydroxybenzoic acid (DHB) is a widely used matrix in MALDI mass spectrometry, particularly effective for lipid and peptide analysis in biological samples [14, 18, 29]. DHB (40 mg/mL in 70% methanol) was uniformly applied to tissue sections using an HTX M3+ sprayer at a flow rate of 100 mm/min, velocity of 1200 mm/min, track space of 3 mm, and a drying time of 10 seconds. The matrix was sprayed in a crisscross pattern for 10 passes at 75°C and 10 psi nitrogen pressure to ensure even coverage and optimal crystal formation.

2.2.1.3 MALDI MSI settings

After matrix application, the tissue slide was first scanned using an Epson scanner and subsequently analyzed using a Waters Synapt G2-Si mass spectrometer, with the parameters in Table 2-1. Staining and Imaging

After MALDI MSI acquisition, the tissue sections were stained with hematoxylin and eosin and imaged at 2× resolution using a Leica M205 FCA microscope at the Advanced Microscopy and Imaging Core (AMIC) of the University of Tennessee, Knoxville (Figure 2-1).

Table 2-1 MALDI-MSI Acquisition Parameters

Mass Range (m/z)	Trap CE	Transfer CE	Polarity	Scan Time (s)	Laser Repetition Rate (Hz)	Laser Energy (a.u.)	Laser Pulse Energy Average (µJ)	Pixel Size (µm)
50-2000	4	2	Positive	0.3	1000	300	25	60

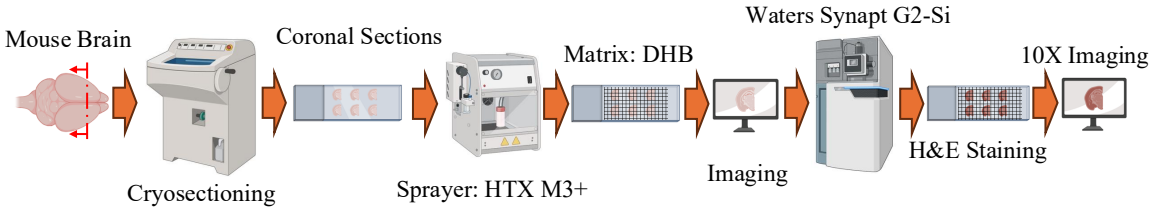


Figure 2-1 Schematic overview of the experimental workflow for MALDI-MSI, highlighting tissue preparation, matrix application, data acquisition, and imaging. [1]

[1] https://www.biorender.com

2.2.1.4 Analysis of Raw Data and Peak Picking

All raw data were converted to the imzML format using High Definition Imaging (HDI) software (Waters Corp). In the young AD group, the data of two samples were lost due to technical difficulties during the data acquisition stage of MSI process. Subsequent preprocessing and data handling were performed using custom-developed Python scripts. Initially, the imzML and ibd files were parsed to extract the complete list of m/z values, their corresponding intensity values, and spatial coordinates for each pixel (Figure 2-2). Following extraction, each m/z spectrum was annotated with relevant metadata, including sample identifier, batch number, age group, and disease status.

2.2.1.5 Global Peak Detection and Targeted Intensity Extraction

To reduce the dimensionality of the MSI dataset while retaining the most informative spectral features, a global peak detection and pixel-wise intensity extraction procedure was implemented.

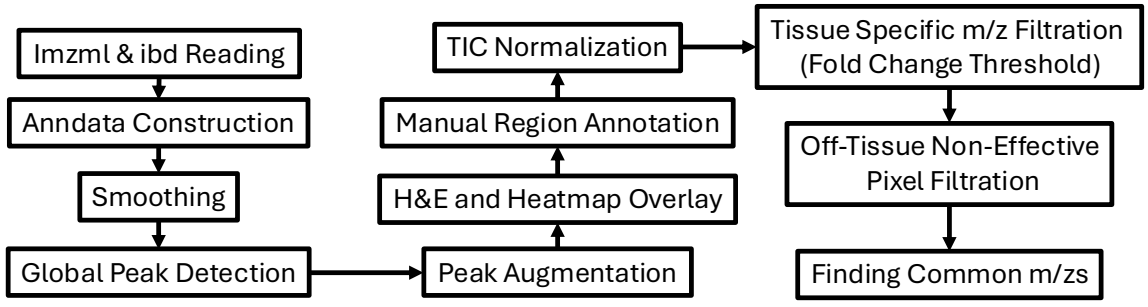


Figure 2-2 Flowchart summarizing the preprocessing steps for MALDI-MSI data implemented in Python.

2.2.1.6 Global Spectrum Construction

First, a global spectrum was constructed by summing intensity values across all spatial pixels for each analyte value.

Let $X \in \mathbb{R}^{n \times m}$ denote the MSI data matrix, where n is the number of spatial pixels and m is the number of m/z (analyte) bins. The global spectrum $S_{\boldsymbol{global}} \in \mathbb{R}^{n}$ was computed by summing the intensity values across all pixels for each analyte:

$S_{\boldsymbol{global}}(j) = \sum_{\boldsymbol{i=1}}^{\boldsymbol{n}} X(i,j), \qquad \text{for } j = 1, 2, \dots, m$

The resulting one-dimensional spectrum represented the aggregate signal intensity profile of the entire sample. This spectrum highlighted the most abundant ions across the tissue section. The summation was performed along the spatial axis, and the resulted single spectrum had the length equal to the number of analyte bins.

2.2.1.7 Peak Detection

To enable reliable peak detection and minimize high-frequency noise, a series of preprocessing steps were applied to the global spectrum. Two types of smoothing filters were tested: the Savitzky–Golay filter and the Gaussian filter, using various parameter combinations as detailed in Table 2-2 [4, 30-32]. Combinations of different window lengths and polynomial orders were evaluated For the Savitzky–Golay filter, and various smoothing sigma values were assessed for the Gaussian filter.

Peak detection was performed using the peak detection function from the SciPy library, which identifies local maxima in a one-dimensional array by comparing each value to its neighboring points. Then, a proximity-based filtering step was applied to reduce

redundancy from closely spaced peaks. Specifically, different minimum distance thresholds between adjacent peaks were tested (Table 2-2). If multiple peaks were detected within this threshold, only the most prominent peak in that group was retained. This method ensured the selection of distinct and non-redundant peaks.

From each parameter combination, the top 1000 most intense peaks were selected. These results – derived from various filter types, smoothing settings, and peak separation distances – were then compared to those obtained from HDI software to validate accuracy and optimize method selection.

Among all tested combinations, the configuration using the Savitzky–Golay filter with a window length of 9 and a polynomial order of 3, along with a minimum peak distance of ±0.024 Da, showed the highest overlap with peaks detected by HDI software.

2.2.1.8 Pixel-wise Intensity Summation Around Target Peaks

To extract peak-specific signal profiles from individual pixels, intensity values around each selected analyte peak were aggregated. For each target peak, a window spanning ±2 index positions was defined, and only m/z values within ±0.012 Da of the target were

Table 2-2. Parameter combinations used for peak detection, including filter type, smoothing parameters, and minimum peak distance (Da)

Filter Type	Smoothing Parameter	Parameter Values	Minimum Peak Distance (Da)
Gaussian Filter	Sigma (σ)	0.2, 0.5, 1.0, 1.5, 2.0, 2.5, 3.0, 3.5, 4.0	0.1, 0.05, 0.024, 0.02, 0.01
Savitzky–Golay Filter	Window Length	3, 5, 7, 9, 11, 13, 15, 17, 19, 21	0.1, 0.05, 0.024, 0.02, 0.01
	Polynomial Order	0, 1, 2, 3	0.1, 0.05, 0.024, 0.02, 0.01

included in the summation to preserve mass accuracy. The choice of ±2 index positions helped minimize the risk of excessive signal augmentation due to overlapping peaks being included in signal calculation for the lower mass range. Alternatively, the ±0.012 Da threshold ensured that overlapping peaks would not be included in the higher mass range. This strategy allowed for the integration of nearby signal contributions, enhancing robustness against minor analyte peak shifts while maintaining specificity.

The resulting dataset comprised a matrix in which each row corresponded to a spatial pixel and each column represented the summed intensity of a selected analyte peak. This reduced representation preserved spatial specificity while significantly lowering computational complexity for downstream analyses.

2.2.1.9 Interactive Registration and Region Annotation of Ion Images

To enable precise anatomical interpretation of spatially resolved MSI data, an interactive framework was developed for overlaying ion heatmaps with corresponding histological images. This workflow allowed for manual alignment between molecular and morphological data (Figure 2-3), followed by user-guided region annotation for spatially targeted analysis.

2.2.1.10 Ion Heatmap Generation

The heat map was constructed for each analyte to visualize the spatial distribution of a specific ion across the sample. The heat map was equipped with a custom continuous colormap and a grey color for zero intensity to enhance visual contrast and perception of chemical gradients across the tissue.

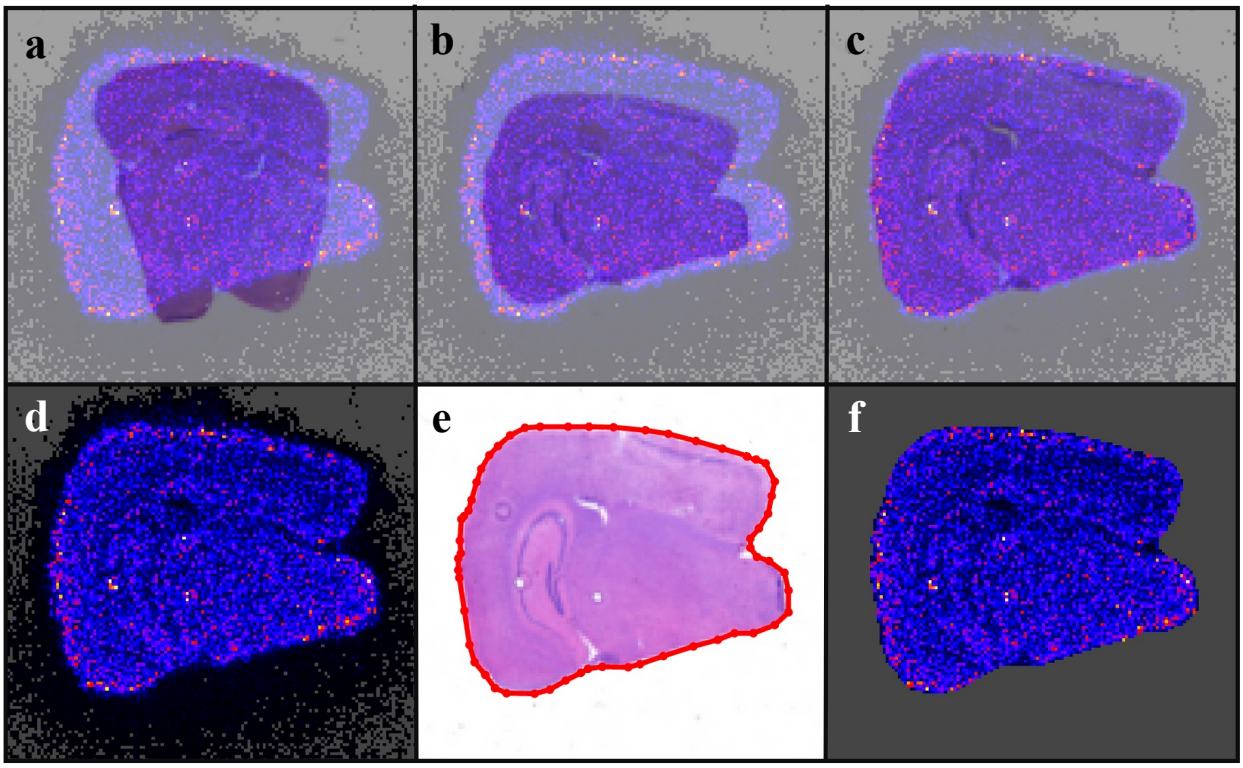


Figure 2-3. H&E-stained tissue image overlaid with the spatial distribution heatmap of m/z 800.55. Top row (left to right): (a) Initial overlay with transparency (α = 0.5), (b) Overlay after rotation (α = 0.5), (c) Final aligned overlay (α = 0.5). Bottom row (left to right): (d) Fully opaque overlay (α = 1), (e) Overlay with annotated tissue margins (red dots and lines; α = 0), (f) Distribution of m/z 800.55 after background masking, showing signal localized exclusively to tissue regions.

2.2.1.11 Manual Image Alignment

The histological image (hematoxylin & eosin stained slides) and the ion image typically differed in orientation, scale, and position. To correct for these discrepancies, an interactive graphical user interface (GUI) was provided, allowing users to manually rotate, resize, shift, and blend the histological image with the ion heatmap. Adjustable parameters including rotation angle, scaling along both axes, horizontal and vertical translation, and the blending ratio between the two images were provided for the

user. These transformations enabled the user to align structural features in the histological image with corresponding molecular features in the MSI data. Real-time updates in the GUI allowed adjusting until optimal visual registration was achieved.

2.2.1.12 Manual Region Annotation

Once the images were aligned, a polygon-based annotation tool was used to manually define regions of interest (ROIs). The user selected points on the image to form a closed polygon surrounding a specific anatomical region. Upon completion, the polygon was converted into a binary mask, indicating which pixels fell within the annotated area.

2.2.1.13 Total Ion Current (TIC) Normalization

To correct for variations in overall ion signal across pixels – caused by technical factors such as matrix heterogeneity, sample thickness, or instrumental fluctuation – total ion current (TIC) normalization was performed on the raw mass spectrometry intensity matrix.

Let the original intensity matrix be denoted as:

$$\mathbf{I} = \begin{bmatrix} I_{11} & I_{12} & \cdots & I_{1m} \\ I_{21} & I_{22} & \cdots & I_{2m} \\ \vdots & \vdots & \ddots & \vdots \\ I_{n1} & I_{n1} & \cdots & I_{nm} \end{bmatrix} \in \mathbb{R}^{n \times m}$$

where:

- n is the number of pixels (observations),
- m is the number of detected analyte bins (variables),
- I_{ij} is the raw intensity of thej^{th} value at the ith pixel.

The total ion current for each pixel is computed as the row-wise sum:

$\mathrm{TIC}_i = \sum_{j=1}^{m} I_{ij}, \qquad \forall i \in \{1, 2, \ldots, n\}$

To prevent division by zero, any zero values of TIC_i were replaced by 1 before normalization.

The normalized intensity matrix $\mathbf{I}' \in \mathbb{R}^{n \times m}$ is then obtained by scaling each row by the inverse of its total ion current:

$I'_{ij} = \frac{I_{ij}}{\mathrm{TIC}_i}, \qquad \forall i \in \{1, 2, \ldots, n\},\ j \in \{1, 2, \ldots, m\}$

This operation ensures that:

$\sum_{j=1}^{m} I'_{ij} = 1$, for all i

Thus, each row of the normalized matrix represents a relative distribution of ion abundances that is invariant to total signal intensity, enabling direct comparisons of spectral patterns across tissue pixels [33, 34].

2.2.1.14 Tissue-Specific Analyte Feature Filtering via Fold Change Thresholding

To improve signal relevance, reduce background noise, and omit most of the matrix related analyte in MSI data, a filtering step was applied to retain only analyte features that exhibited stronger signal intensity within annotated tissue regions compared to surrounding non-tissue areas.

Let the intensity matrix be denoted as:

$$\mathbf{I} \in \mathbb{R}^{n \times m}$$

where:

- n is the number of spatial pixels,
- m is the number of analyte features,

- I_{ij} is the raw intensity of thej^{th} analyte value at the ith pixel.

A binary tissue mask vector $t \in \{1,0\}^n$ was used to distinguish tissue and background pixels:

- $t_i = 1$ indicates a tissue pixel,
- $t_i = 0$ indicates a non-tissue (background) pixel.

Let $\mathcal{T} = \{i|t_i = 1\}$ and $\mathcal{B} = \{i|t_i = 0\}$ denote the sets of indices for tissue and background pixels, respectively.

For each analyte feature $j \in \{1, \dots, m\}$, the mean intensity over tissue and non-tissue regions was computed as [32]:

$$\mu_j^{tissue} = \frac{1}{|\mathcal{T}|}\sum_{i\in\mathcal{T}} I_{ij} \quad , \qquad \mu_j^{non} = \frac{1}{|\mathcal{B}|}\sum_{i\in\mathcal{B}} I_{ij}$$

A fold change threshold $\gamma > 1$was applied to assess tissue enrichment:

$$FoldChange_j = \frac{\mu_j^{tissue}}{\mu_j^{non}}$$

Only those features satisfying:

$$FoldChange_j \geq \gamma$$

were retained. In this study, different thresholds were tested, and a threshold $\gamma = 3$ was selected because it was able to preserve most tissue-specific analytes and omit matrix-related analyte peaks. Because the mean intensity is defined by tissue and non-tissues regions, isobaric molecules of interests that have overlapping mass-to-charge ratios (m/z) with matrix peaks were likely excluded.

This filtering procedure was applied directly to the raw intensity matrix **I** without performing TIC normalization prior to fold change calculation. The resulting matrix contained only those analyte features with substantial tissue-specific abundance, improving signal-to-noise ratio and enhancing the biological interpretability of spatial metabolic signatures.

To increase separability of analytes and focus on meaningful background signal, effective pixels in the off-tissue region (off-tissue pixel enrichment) were defined as those with intensity values exceeding 10% of the maximum intensity for the corresponding analyte feature. Pixels below this threshold were considered background noise and set to zero. The term 'pixel' in the background (off-tissue) region refers specifically to effective pixels.

After initial preprocessing and filtering, several spatial and intensity-based variables hypothesized to influence delocalization were independently evaluated. These variables included:

1. The Euclidean distance between the global center of mass and the tissue-specific center of mass
2. The mean and maximum distances of nonzero-intensity pixels located in the background (off-tissue) region to the nearest tissue boundary
3. The area of nonzero-intensity pixels located in the background (off-tissue)

For each analyte feature, a delocalization score was calculated by integrating these metrics, and the features were subsequently ranked according to their scores. Visual inspection was performed to validate the ranking and spatial distribution patterns.

2.2.1.14.1 Global and Tissue Center of Mass (COM)

To quantify the shift in molecular distribution due to delocalization, the intensity-weighted centers of mass (COM) were calculated for both tissue-only pixels and all pixels (tissue + background).

Let:

- $\mathbf{X} = \{(x_i, y_i)\}_{i=1}^{N} \subset \mathbb{R}^2$: spatial coordinates of all pixels
- $\mathbf{I} = \{I_i\}_{i=1}^{N}$: corresponding intensity values for a given analyte feature
- $\mathcal{T} \subset \{1, \ldots, N\}$: index set of tissue pixels
- $\mathcal{B} = \{1, \ldots, N\} \setminus \mathcal{T}$: index set of background pixels

The intensity-weighted global center of mass (COM) is defined as:

$$\mathbf{c}_{\text{golbal}} = \begin{bmatrix} x_{com} \\ y_{com} \end{bmatrix} = \frac{\sum_{i=1}^{N} I_i \mathbf{x}_i}{\sum_{i=1}^{N} I_i} \quad \text{where } \mathbf{x}_i = \begin{bmatrix} x_i \\ y_i \end{bmatrix}$$

Similarly, the tissue-only center of mass is computed as:

$$\mathbf{c}_{\text{tissue}} = \frac{\sum_{i \in \mathcal{T}} I_i \mathbf{x}_i}{\sum_{i \in \mathcal{T}} I_i}$$

If $\sum I_i = 0$, then $\mathbf{c} = \text{NaN}$.

Then, the Euclidean distance between two centroids was calculated using

$$d_{\text{COM}} = \left\| \mathbf{c}_{\text{golbal}} - \mathbf{c}_{\text{tissue}} \right\|_2 = \sqrt{(x_{\text{com}} - x_{\text{com}}^{\mathcal{T}})^2 + (y_{\text{com}} - y_{\text{com}}^{\mathcal{T}})^2}$$

This distance, d_{COM}, serves as the delocalization score for the analyte feature. All analyte values were ranked based on their corresponding scores to evaluate their localization accuracy.

Evaluation of Euclidean Distance Between Two Centroids

Based on the computed scores, m/z 180.07 (ranked 13th) exhibited more delocalization than m/z 365.32 (ranked 7th), and m/z 703.57 (ranked 35th) showed more delocalization than m/z 240.10 (ranked 118th). Similarly, m/z 348.07 (ranked 15th) was considered more delocalized than m/z 283.09 (ranked 25th) as shown in Figure 2-4. However, visual inspection contradicted these results. These outcomes suggest that d_{COM} was not an effective metric for assessing delocalization because it primarily captured the skewness or displacement of the distribution rather than the true extent of spatial spread.

2.2.1.14.2 Distance of Nonzero Background Pixels to Nearest Tissue Pixel

To investigate whether either the maximum or the average distance from tissue borders could serve as a defining factor for delocalization, the following mathematical framework was implemented.

Let:

- $\mathcal{B}_+ = \{j \in \mathcal{B} \mid I_j > 0\}$: background pixels with nonzero intensity
- $\mathbf{x}_j \in \mathbb{R}^2$: be the spatial coordinate of the background pixel j
- $\mathbf{x}_k \in \mathbb{R}^2$: spatial coordinate of the tissue pixel k

Then, for each $j \in \mathcal{B}_+$, the distance to the nearest tissue pixel was calculated as:

$$d_{\mathrm{j}} = \min_{k \in \mathcal{T}} \left\| \mathbf{x}_j - \mathbf{x}_k \right\|_2 \quad \text{, (Figure 2-5)}$$

From these distances, the definition was:

- Maximum distance of nonzero-intensity background pixels to tissue:

$$d_{\max} = \max_{j \in \mathcal{B}_+} d_j$$

- Mean distance of nonzero-intensity background pixels to tissue:

$$\mu_{\text{bg}} = \frac{1}{|\mathcal{B}_+|} \sum_{j \in \mathcal{B}_+} d_j$$

- Nonzero-intensity background area:

$$A_{\text{bg}} = |\mathcal{B}_+|$$

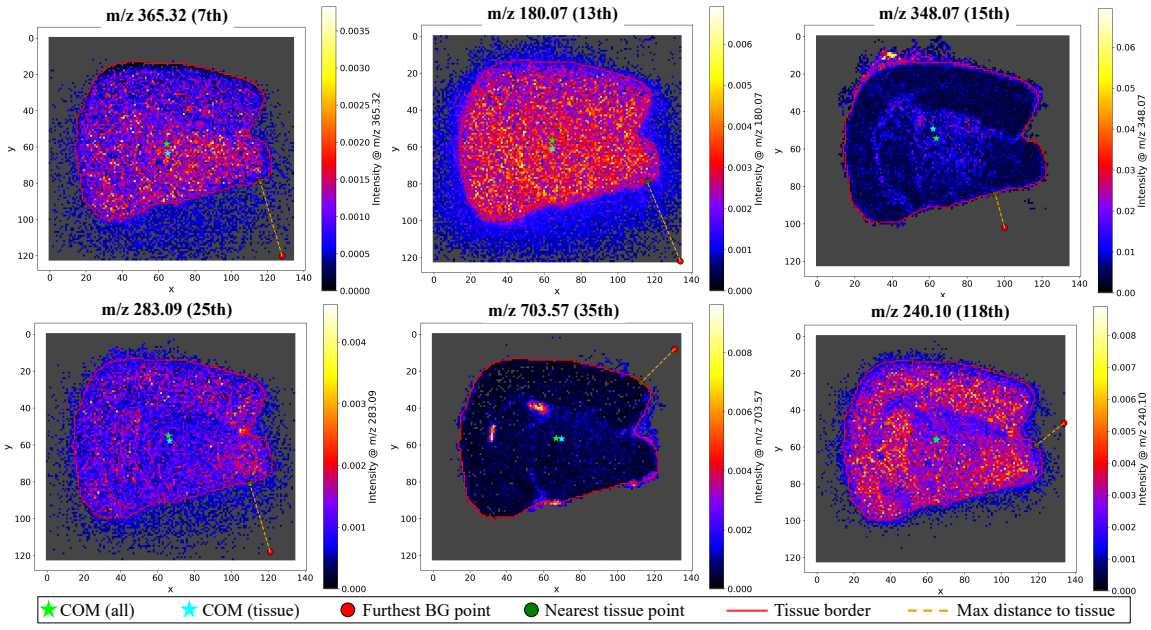


Figure 2-4. Spatial distribution maps of six selected analyte features used to evaluate delocalization rankings based on the Euclidean distance between the global and tissue-specific centers of mass ($\boldsymbol{d_{COM}}$). The top row displays m/z 365.32 (rank 7), 180.07 (rank 13), and 348.07 (rank 15); the bottom row shows m/z 283.09 (rank 25), 703.57 (rank 35), and 240.10 (rank 118).

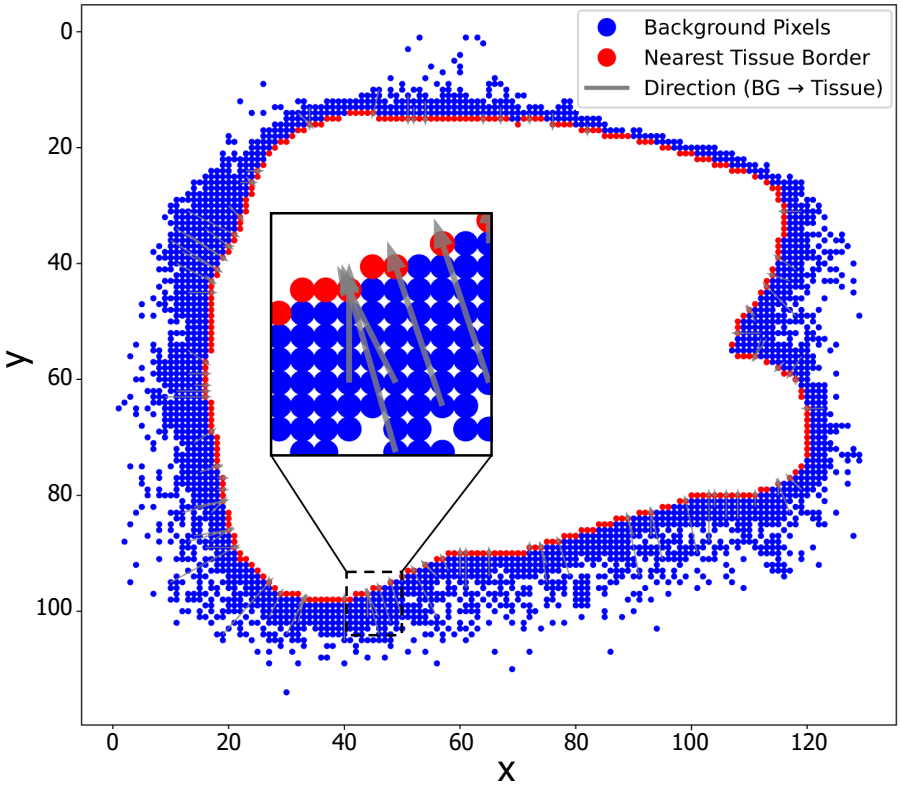


Figure 2-5. Spatial map showing background pixels (blue) and their nearest corresponding tissue border points (red) for m/z 651.53. Arrows (gray) illustrate the direction and distance from each sampled background pixel to its closest tissue border. BG: background pixels

Evaluation of Maximum Distance of Nonzero Background Pixels to Tissue

Based on the results, m/z 180.07 (ranked 1st) and m/z 185.07 (ranked 2nd) were both associated with the same furthest pixel from the tissue border as shown in Figure 2-6. This is likely because the MSI acquisition was spatially limited and did not extend beyond that point – these pixels could not be more distant simply due to the finite acquisition region.

This limitation was further highlighted by the inconsistency between ranking and visual assessment. For instance, m/z 369.35 (ranked 27th) and m/z 247.24 (ranked 25th) showed

opposite delocalization patterns upon inspection, contradicting their rank order. Similarly, m/z 651.53 (ranked 139th) appeared more delocalized than m/z 729.57 (ranked 115th), despite the ranking suggested otherwise. These discrepancies indicate that maximum background distance is not a reliable individual metric for assessing delocalization. This is primarily because:

- The MSI acquisition area is spatially bounded, so some delocalized signals may fall outside the field of view of data acquired
- Using only the furthest pixel neglects the distribution and quantity of delocalized pixels.
- Keeping a large background area to capture distant pixels introduces significant computational burden and potential noise, which compromises interpretability.
- Maximum distance focuses solely on spatial extent**,** ignoring pixel count**,** which is a critical limitation for distinguishing highly vs. weakly delocalized features.

As a result, this metric failed to reliably differentiate between highly delocalized analyte features and more localized ones.

Evaluation of Mean Distance of Nonzero Background Pixels to Tissue

For mean distance of nonzero-intensity background pixels to the nearest tissue pixel, the resulting rankings provided better alignment with visual assessments than both the maximum distance metric and the Euclidean distance between global and tissue centers of mass. However, a substantial number of analyte features were still inaccurately ranked.

For example, m/z 353.33 (ranked 23rd) and m/z 341.31 (ranked 26th) showed conflicting results, where visual inspection clearly indicated that m/z 341.31 exhibited greater delocalization. Similarly, m/z 753.59 (ranked 58th) and m/z 427.93 (ranked 63rd) revealed a discrepancy in delocalization extent, with the latter having significantly more delocalized pixels (Figure 2-7).

These inconsistencies suggest that the number or area of delocalized pixels plays a more crucial role than distance alone in evaluating delocalization. This becomes even more evident in the comparison between m/z 710.49 (ranked 324th) and m/z 769.56 (ranked

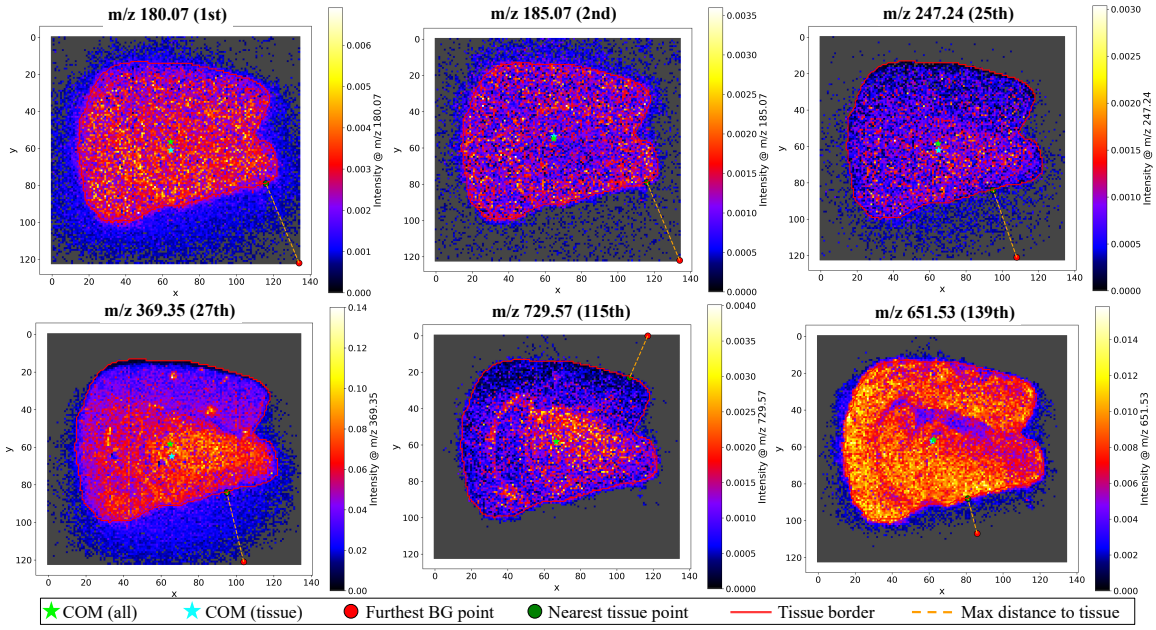


Figure 2-6. Spatial distribution maps of six analyte features used to evaluate delocalization rankings based on the maximum Euclidean distance of nonzero background pixels to the nearest tissue border. The top row shows m/z 180.07 (rank 1), 185.07 (rank 2), and 247.24 (rank 25); the bottom row shows m/z 369.35 (rank 27), 729.57 (rank 115), and 651.53 (rank 139).

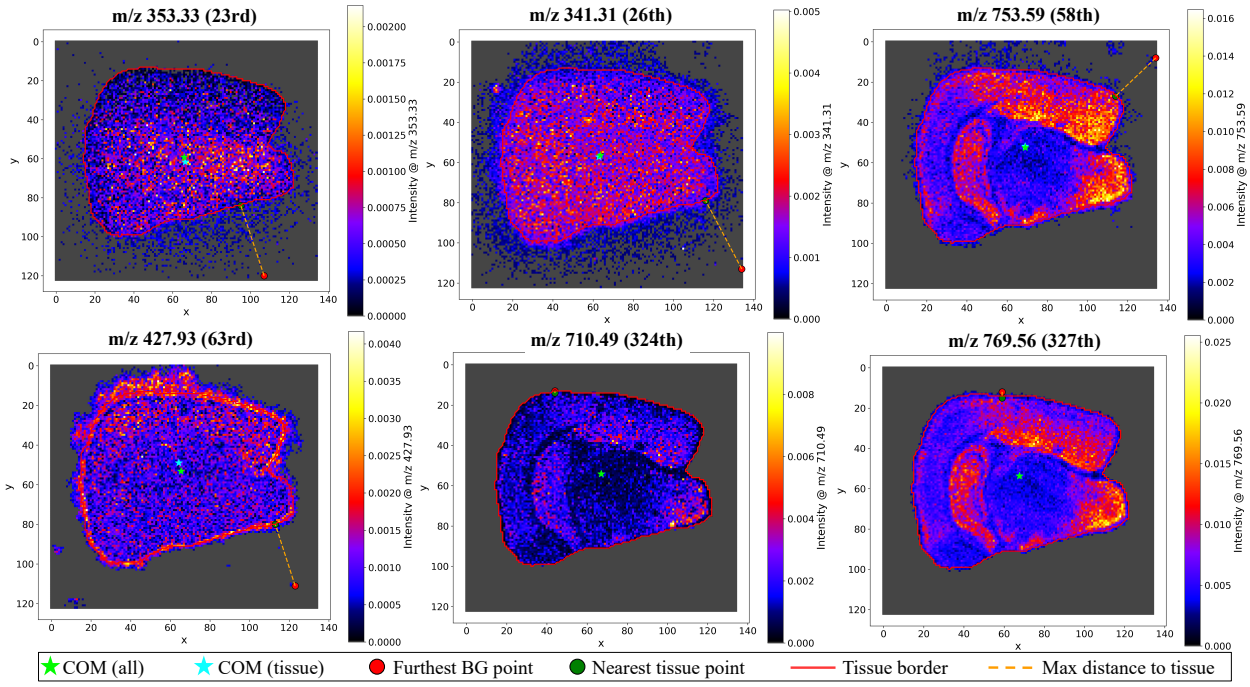


Figure 2-7. Spatial distribution maps of six analyte features to evaluate delocalization rankings based on the mean distance of nonzero-intensity background pixels to the nearest tissue pixel. The top row shows m/z 353.33 (rank 23), 341.31 (rank 26), and 753.59 (rank 58); the bottom row displays m/z 427.93 (rank 63), 710.49 (rank 324), and 769.56 (rank 327).

327th). Although m/z 710.49 had only around 10 delocalized pixels, it was ranked higher than m/z 769.56, which had nearly 100 delocalized pixels (Figure 2-7). The reason for this misranking was that the former had a slightly higher mean background distance. This highlights a limitation of using mean distance alone. It fails to capture the spatial extent or area covered by delocalized pixels**,** which is essential for accurately characterizing delocalization. Therefore, while mean distance improves the ranking results in comparison to previous metrics, it cannot solely quantify the degree of delocalization, and it should be combined with additional spatial descriptors such as delocalized area or pixel count.

2.2.1.14.3 Area of Nonzero-Intensity Background Pixels

To investigate the area of the nonzero-intensity background pixels (a nonzero-intensity background pixel count) could serve as a defining factor for delocalization, the number of nonzero-intensity pixels in the background (A_{bg}) was counted, and the analytes were ranked based on it.

Evaluation of the Area of Nonzero Background Pixels

The results of the analyte delocalization ranking based on the area of nonzero-intensity background pixels showed strong agreement with visual assessments, establishing area as a valuable metric for delocalization evaluation. However, a few minor discrepancies were observed based on closer inspection.

For example, in Figure 2-8, m/z 771.55 (ranked 260th) appeared more delocalized than m/z 799.55 (ranked 259th), despite being ranked lower. In another case, m/z 870.47 was ranked higher (318th) than m/z 714.45 (319th) because the former had a slightly larger background area (27 vs. 25 pixels). Nonetheless, visual assessment suggested otherwise. This discrepancy highlights that delocalization is influenced not only by quantity (i.e., number of background pixels) but also by spatial scattering.

The interaction between quantity and scattering was also evident in the comparison of m/z 869.46 (ranked 326th) and m/z 823.63 (ranked 328th). While their background areas were similar, visual inspection indicated that m/z 823.63 was more delocalized, suggesting that the rankings should be reversed.

It is worth noting that as the quantity of delocalized pixels increases, the scattering often increases as well. This correlation can complicate comparisons, particularly in cases of high delocalization, where visual interpretation gains greater complexity.

Based on the limitations observed in the delocalization rankings using the four previously evaluated parameters, it becomes evident that no single metric is sufficient. Therefore, it is more appropriate to define a composite delocalization score that incorporates both the area and spatial distribution of nonzero background pixels to achieve a more accurate and comprehensive assessment of delocalization.

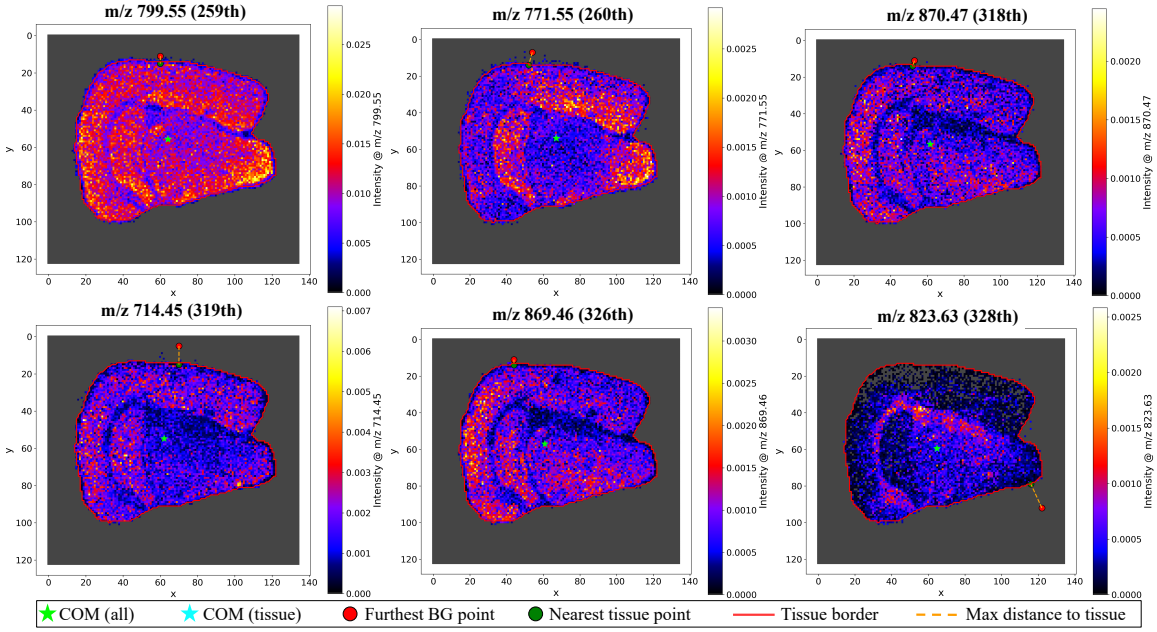


Figure 2-8. Spatial distribution maps of six analyte features to evaluate delocalization rankings using the area of nonzero-intensity background pixels as a delocalization metric. The top row includes m/z 799.55 (rank 259), 771.55 (rank 260), and 870.47 (rank 318); the bottom row includes m/z 714.45 (rank 319), 869.46 (rank 326), and 823.63 (rank 328).

2.2.2 Composite Delocalization Score

To implement this composite delocalization score, a delocalization score was proposed to integrate both the area and spatial distribution characteristics of nonzero-intensity background pixels. The area of these pixels reflected the extent of signal spread, and the average distance between nonzero-intensity background pixels and tissue pixels captured the degree of spatial delocalization. Min-max normalization was applied to both metrics. Then, a weighted linear combination was employed to derive a unified score that more closely aligned with visual assessments and conceptual expectations of delocalization. The detailed mathematical formulation of this score is presented below:

Suppose there are M analyte features, indexed by $i = 1,2, \dots, M$. For each feature i, let:

- d_i: the mean distance from nonzero-intensity background pixels to the nearest tissue pixels
- a_i: the area, i.e., number of nonzero-intensity background pixels

We define the following vectors:

- $d = [d_1, d_2, \dots, d_M]^{\mathrm{T}} \in \mathbb{R}^M$
- $a = [a_1, a_2, \dots, a_M]^{\mathrm{T}} \in \mathbb{R}^M$

2.2.2.1 Feature-wise Min-Max Normalization:

Each vector is normalized using min-max scaling:

$$\tilde{d}_i = \frac{d_i - \min_j d_j}{\max_j d_j - \min_j d_j} \quad \text{for all } i \in \{1, \dots, M\}$$

$$\tilde{a}_i = \frac{a_i - \min_j a_j}{\max_j a_j - \min_j a_j} \quad \text{for all } i \in \{1, \dots, M\}$$

These yield the normalized vectors:

- $\tilde{d} \in [0,1]^M$
- $\tilde{a} \in [0,1]^M$

2.2.2.2 Delocalization Score (Weighted Linear Combination)

Let $\alpha \in [0,1]$ denote the weight assigned to the normalized area metric (e.g., $\alpha = 0.95$). The delocalization score s_i for each analyte feature is then defined as:

$$s_i = \alpha \cdot \tilde{a}_i + (1 - \alpha) \cdot \tilde{d}_i \qquad \text{for all } i \in \{1, \ldots, M\}$$

In vector form, this becomes:

$$\mathbf{s} = \alpha \cdot \tilde{\mathbf{a}} + (1 - \alpha) \cdot \tilde{\mathbf{d}}, \qquad \mathbf{s} \in [0,1]^M$$

A higher $s_i$value indicates greater spatial delocalization of the corresponding analyte peak in the background, reflecting a larger spread area and/or a greater distance from the tissue.

2.2.2.3 Assessing Delocalization Parameters

To determine an appropriate value for α, several factors were considered. The strong performance of the nonzero-intensity background pixel area as an indicator of delocalization suggested that it should contribute more heavily to the final score than the mean background-to-tissue distance. Accordingly, α was varied within the range [0.90, 0.99], and the resulting rankings of analyte features were visually assessed. For instance, when $\alpha = 0.99$, the delocalization score ranked m/z 651.53 30th and m/z 255.21 31st. This ordering conflicted with visual evaluation (Figure 2-9). When α was adjusted to 0.95, this order reversed. Therefore, m/z 255.21 preceded m/z 651.53 which matched

better with visual examination. In the other case, when $\alpha = 0.90$, the delocalization score indicated that m/z 761.59 was more delocalized than m/z 406.95 (Figure 2-9). In contrast, increasing α to 0.95 reversed this ranking, and visual inspection confirmed that m/z 406.95 exhibited greater delocalization. Among the tested values, $\alpha = 0.95$ yielded rankings most consistent with qualitative assessments of spatial delocalization. Notably, in some cases, incrementing α by 0.01 had no effect on the ranking, indicating a degree of robustness in the scoring function.

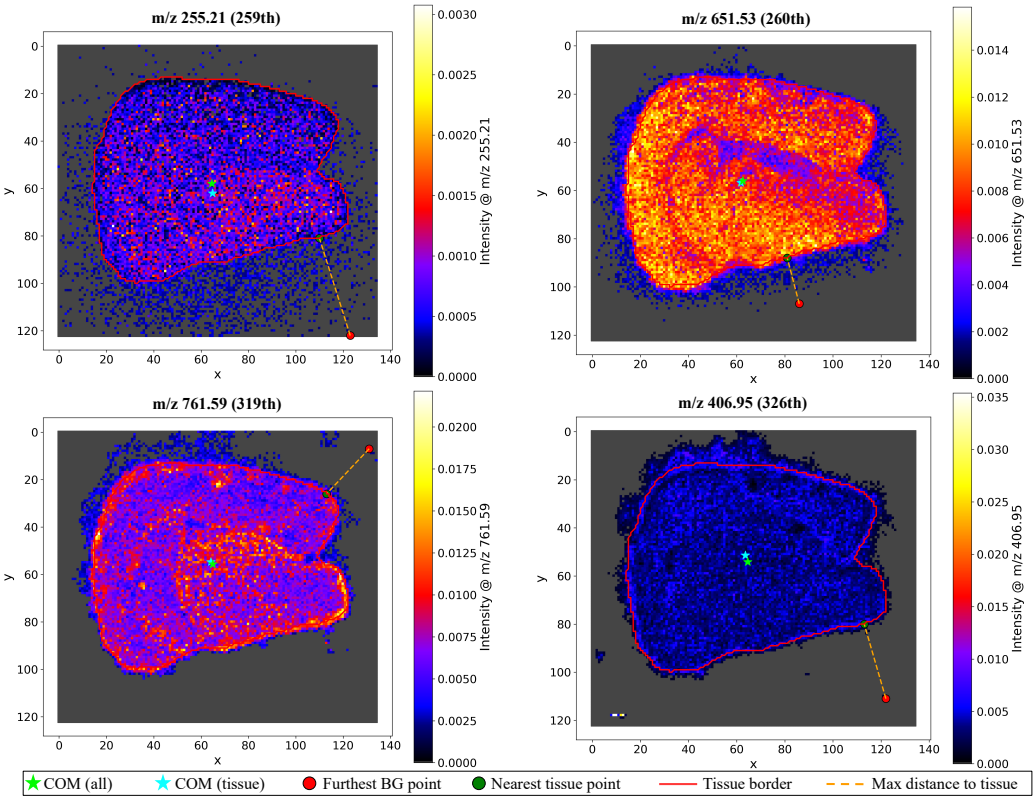


Figure 2-9. Spatial distribution maps of four analyte features used to evaluate the effect of the weighting parameter $\boldsymbol{\alpha}$ on the final delocalization score rankings. The top row shows m/z 255.21 and m/z 651.53, while the bottom row includes m/z 761.59 and m/z 406.95.

2.2.2.4 Mean Intensities in Tissue and Background

The mean intensity of each analyte feature was calculated separately for the tissue and background regions to explore the potential relationship between the delocalization score and intensity levels. The calculations were performed using the following formulas:

$$\bar{I}^{\mathcal{T}} = \frac{1}{|\mathcal{T}|}\sum_{i\in\mathcal{T}} I_i \qquad , \qquad \bar{I}^{\mathcal{B}} = \frac{1}{|\mathcal{B}|}\sum_{i\in\mathcal{B}} I_i$$

Where $\bar{I}^{\mathcal{T}}$ and $\bar{I}^{\mathcal{B}}$ denote the mean intensities of a given analyte feature across tissue ($\mathcal{T}$) and background ($\mathcal{B}$) pixels, respectively, and I is the intensity at pixel i. These metrics were used to investigate whether features with higher intensities in the tissue or background regions also tended to exhibit higher delocalization scores. A logarithmic transformation was first applied to the mean intensity values to address potential skewness and to stabilize variance. Then, Min-Max normalization was performed to rescale the transformed values into a comparable range [0, 1]. This procedure facilitated the consistent interpretation and visualization alongside the delocalization score.

2.2.2.5 Candidate Analyte Features

In this study, only analyte features consistently detected across all 22 samples were selected for further analysis to enable a more robust evaluation of delocalization trends and reduce variability. This filtering step ensured that the analysis focused on consistently detected (common) m/z values, and it minimized potential bias from sample-specific or sporadically detected features.

2.3 Results and Discussion

To examine how spatial delocalization varies across all shared analyte values, bar charts were generated to display delocalization scores for each common analyte across the dataset (Figure 2-10). The analyte (m/z) features were first ranked based on their average delocalization scores, then grouped into two subsets (ranks 1–71 and 72–142) for more effective visualization.

Within each chart, the height of the bars reflects the mean delocalization score of each analyte within the respective rank range. Error bars indicate the standard error of the mean across different samples, while red dots overlaid on each bar represent the individual delocalization scores per sample, offering a view into the level of variation between samples.

This method provides an intuitive understanding of how delocalization is distributed across the m/z range, from the most spatially dispersed to the most localized ions. Analyte features that maintain high delocalization in all samples are reflected by consistently elevated red dots. As the average delocalization score declines, the individual sample scores also trend downward. This observation indicates decreased delocalization. For example, the red dots for m/z 864.63 cluster near zero are suggesting minimal delocalization, whereas m/z 652.54 displays higher red dots clustered around values of 0.6 to 0.9 are indicating broader spatial spread. This visualization enhances the clarity and interpretability of the delocalization scoring method, and it helps to identify key analyte features that may deserve further biological or spatial investigation.

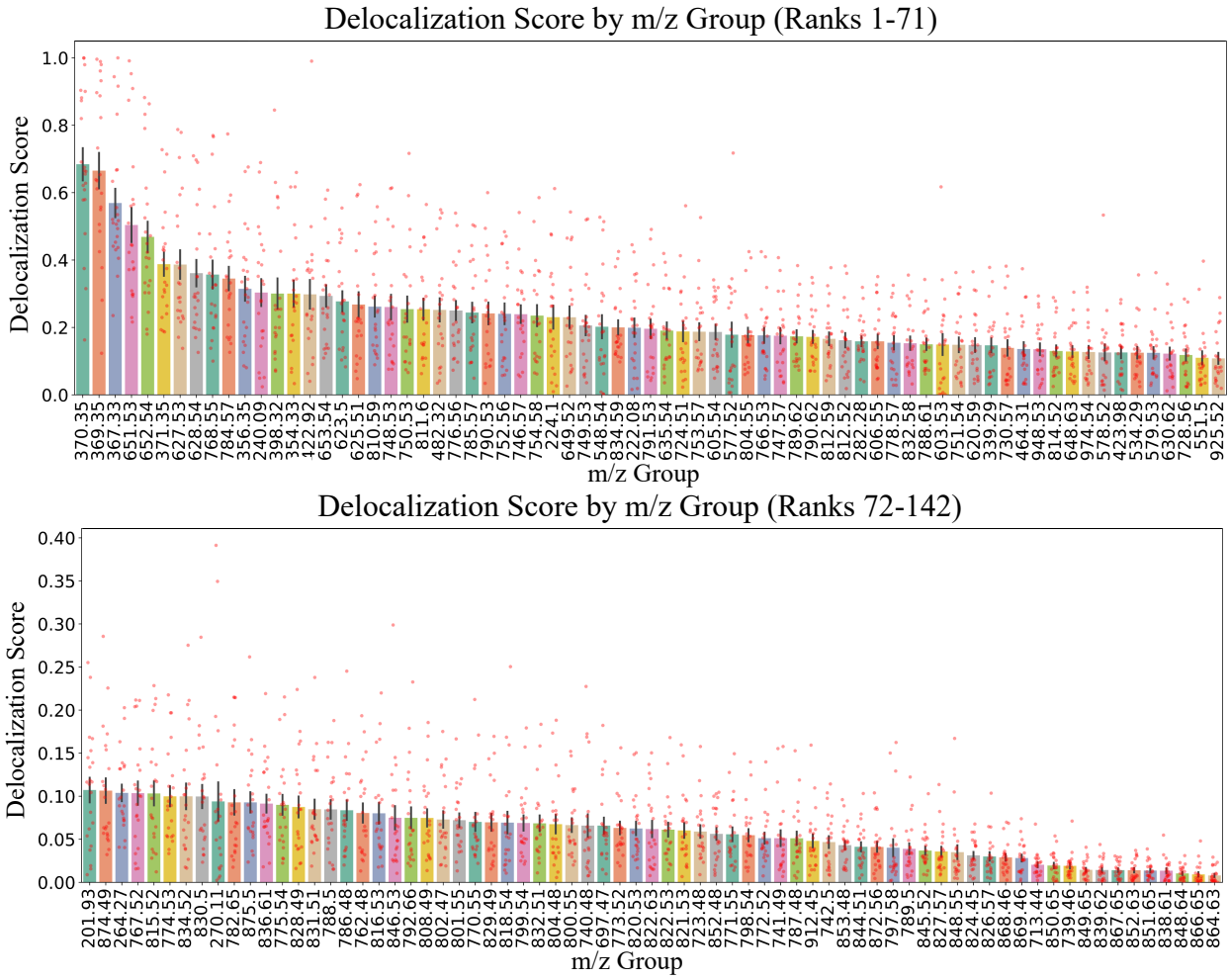


Figure 2-10. Delocalization scores of analyte groups ranked by their mean score across samples. The top panel shows ranks 1–71 (most delocalized), and the bottom panel shows ranks 72–142. Bars represent mean scores with standard error; red dots indicate individual sample values.

A pairwise scatter plot matrix was constructed to investigate the relationships among key spatial metrics and intensity features. This study was performed using five normalized variables: delocalization score, nonzero background area, mean background distance, log-transformed tissue intensity, and log-transformed background intensity (Figure 2-11). Each subplot within the matrix displays a scatter plot for a pair of variables, with a red

regression line overlaid to highlight linear trends. Diagonal elements present histograms that illustrate the distribution of each individual variable.

All axes were normalized to a [0, 1] range to facilitate consistent and direct visual comparison across all metrics. The resulting visualization revealed strong associations between the delocalization score and both mean tissue intensity and mean background intensity. Additionally, nonzero background area exhibited a notable positive correlation with mean background distance, as well as with both intensity-related features. Given that nonzero background area is a major contributor to the delocalization score, these relationships suggest that the influence of intensity and spatial spread is already somewhat embedded in the delocalization score through this variable. This integrative view was able to provide valuable insight into how different spatial and intensity features correlate and support a deeper understanding of the factors contributing to molecular delocalization patterns.

Principal component analysis (PCA) was applied to the delocalization score of candidate analyte values across all 22 samples to examine potential differentiation group-wise patterns in spatial delocalization (Figure 2-12). The first two principal components (PC1 and PC2) were extracted and visualized in a two-dimensional scatter plot, where each point represents a sample and is color-coded by its biological group (e.g., young-control, aged-control, young-AD, aged-AD).

The resulting PCA plot did not reveal any clear separation or clustering by group. All samples appeared intermixed, indicating that delocalization patterns at the analyte level

do not strongly distinguish between biological conditions. As expected, this result suggests that spatial delocalization of individual analyte features may not, on their own,

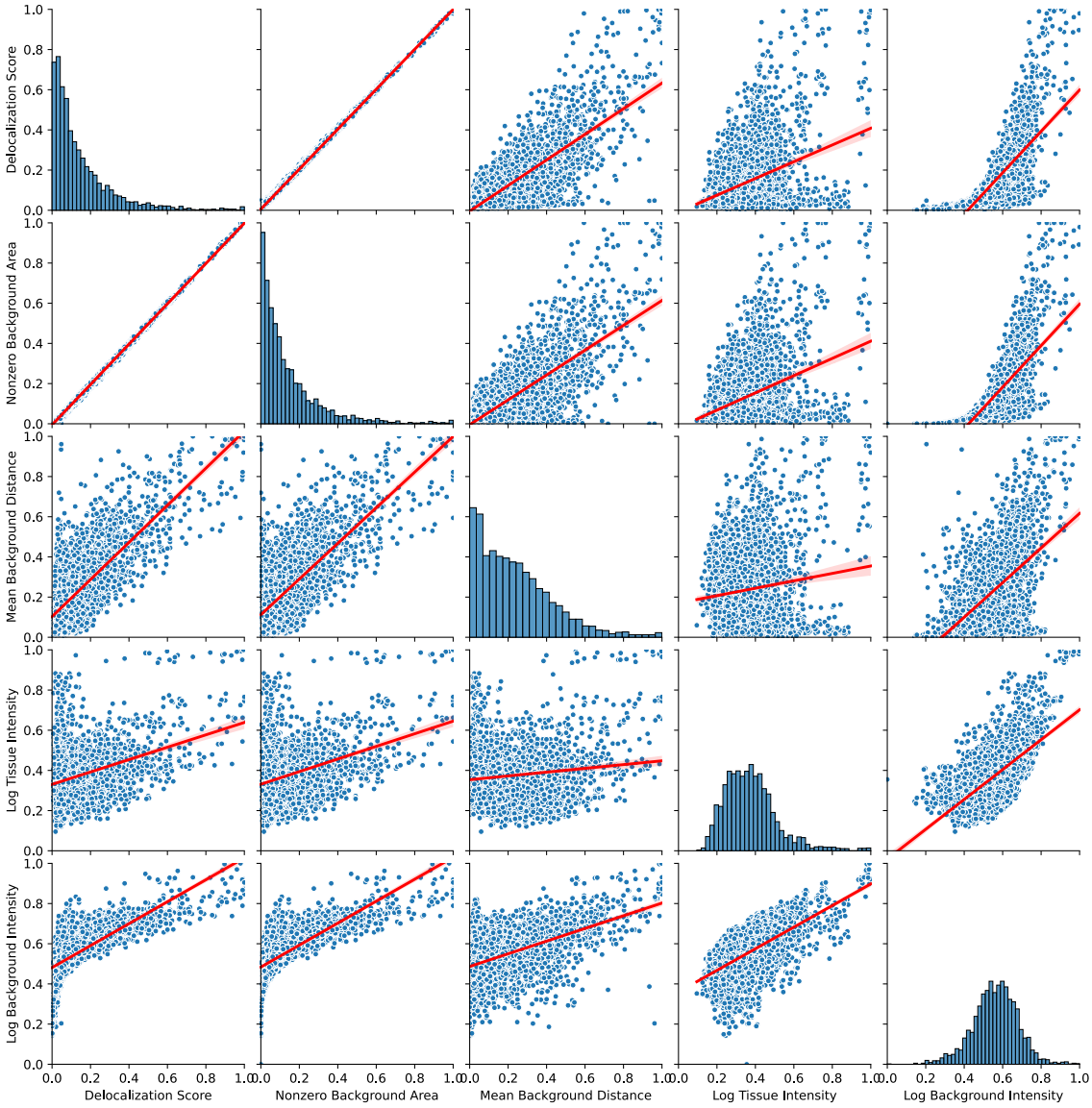


Figure 2-11. Pairwise correlations between normalized delocalization metrics across analyte features. Scatter plots include linear regression lines (red), with histograms on the diagonal showing each metric's distribution.

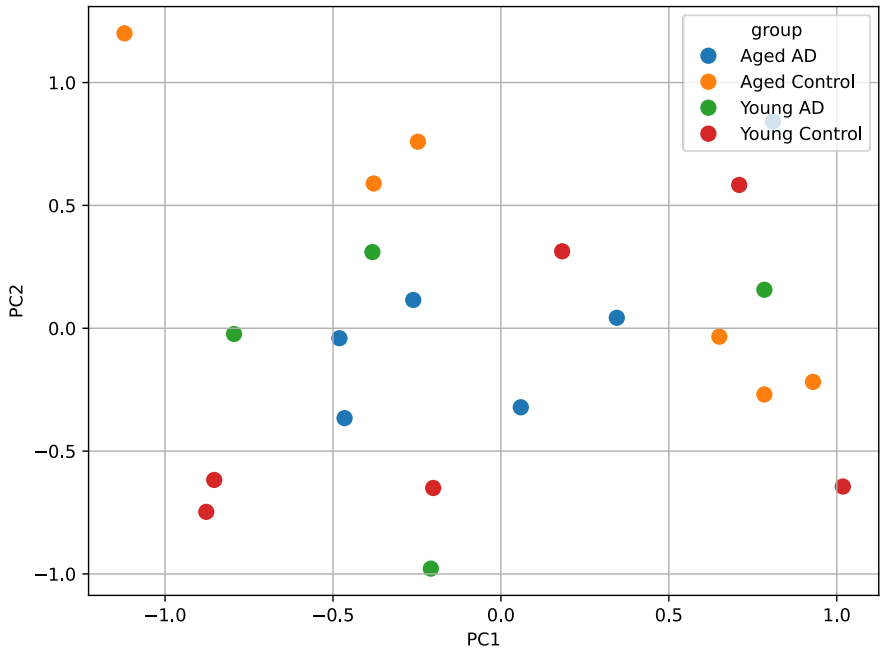


Figure 2-12. Principal component analysis (PCA) of delocalization score profiles across samples. Samples are projected onto the first two principal components based on their delocalization score patterns across analyte groups. Each point represents a sample, colored by experimental group.

reflect major group-level differences, and it underscores that delocalization is largely a property of the analyte and probably the tissue, rather than of the experimental groups.

Pearson correlation coefficients were calculated using the delocalization scores of all common analyte features to evaluate the similarity of spatial delocalization patterns across samples. Subsequently, these pairwise correlations were visualized as a heatmap (Figure 2-13). In the heatmap, each cell represents the correlation between two samples, with values ranging from 0 (no correlation) to 1 (perfect correlation). Pearson correlation can theoretically range from -1 to 1; However, only non-negative correlations were observed in this analysis. All correlation values are directly annotated within the matrix to enhance interpretability.

This visualization offers a comprehensive overview of how consistently delocalization patterns are preserved across the dataset. Most samples demonstrated relatively high correlation coefficients, typically above 0.60 or 0.70, indicating that spatial delocalization behaviors of analyte features are broadly similar across samples. However, a few bluish cells were observed within the Young Control group, suggesting reduced similarity between certain samples in that group and others. Notably, sample YC_1 had undergone a physical cut during preparation, which may have affected its spatial profile. These discrepancies highlight the potential influence of technical variability, such as sample preparation artifacts or batch effects, on spatial delocalization measurements.

A heatmap of delocalization scores was generated to visualize the variation in spatial delocalization across all common analyte features and samples (Figure 2-14). In this heatmap, each row represents a specific analyte group, and each column corresponds to a sample, and the delocalization scores were visualized using a colormap, where darker shades indicate lower scores (more spatially localized features), and brighter shades indicate higher scores (more delocalized features). This representation provides a clear overview of how spatial delocalization patterns differ both across analyte values and among samples.

The heatmap revealed several analyte features with consistently high or low delocalization scores across all samples. These features denoted stable and potentially reproducible delocalization behavior. For example, m/z 866.65 showed a uniform delocalization pattern across all samples, implying that its spatial behavior may be

conserved across other brain tissues as well. However, many common analyte features in samples YC_1 and YC_2 displayed different patterns compared to the rest of the dataset. These differences may reflect somehow underlying biological variation, more probably technical artifacts from sample preparation, or batch effects.

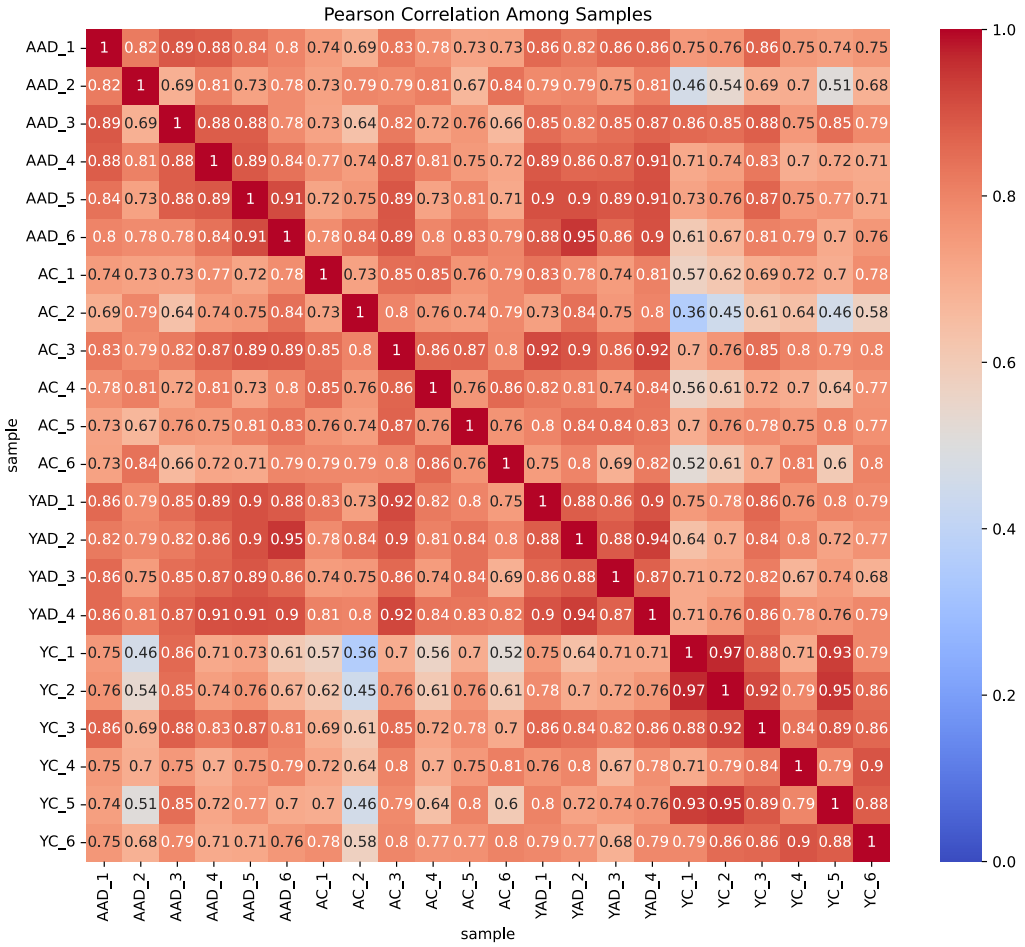
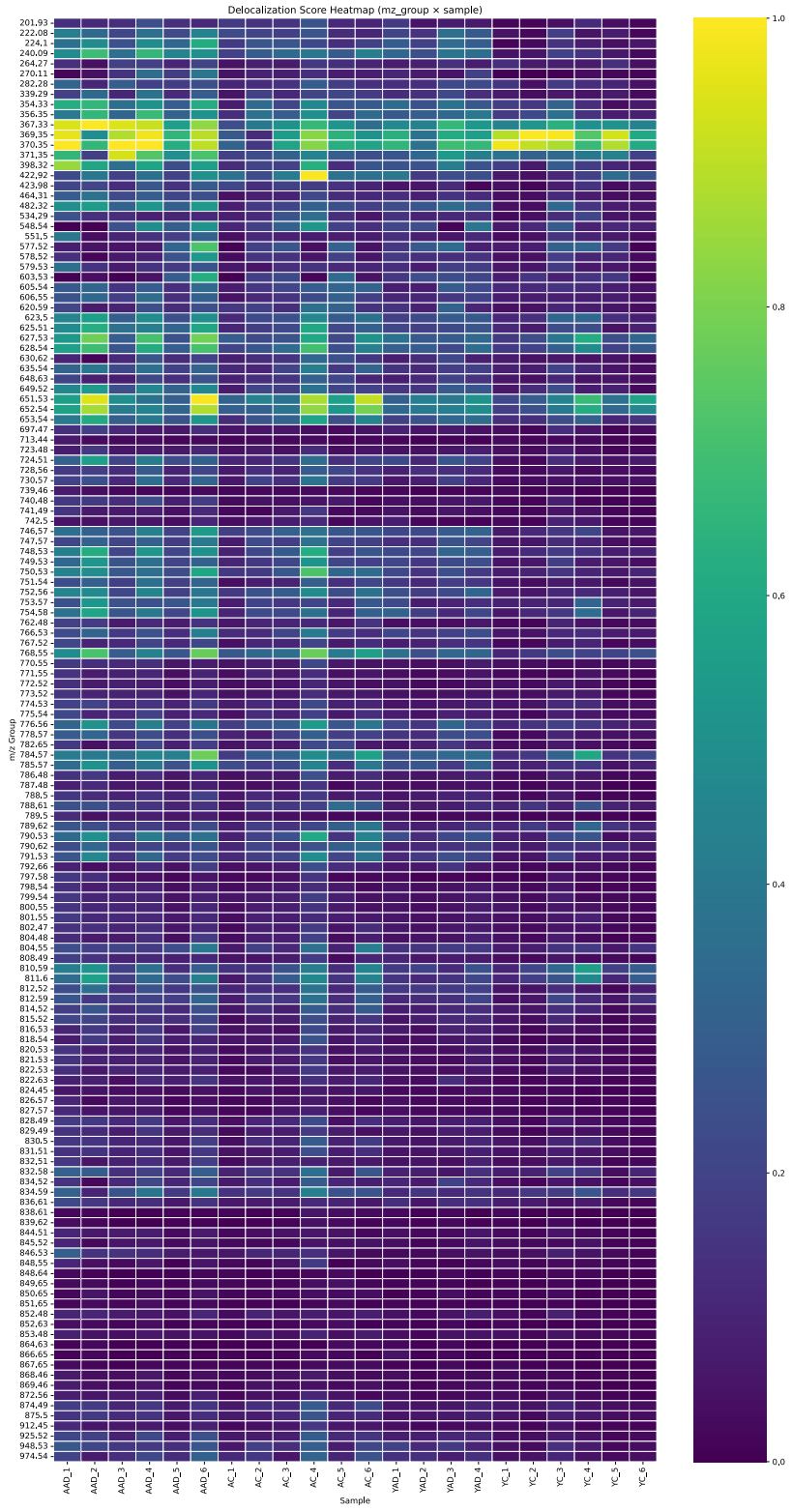


Figure 2-13 Pearson correlation heatmap of delocalization score profiles across samples. Pairwise correlation coefficients (r) are computed between all samples based on their delocalization scores across analyte groups. Warmer colors indicate stronger positive correlations. The high within-group similarity suggests consistent spatial distribution patterns among replicates.

Figure 2-14 Heatmap of delocalization scores across analyte groups and samples. Each row represents an analyte group and each column corresponds to a sample. Color intensity reflects the delocalization score, with brighter values indicating stronger spatial delocalization. This visualization highlights sample-specific and group-specific variation in lipid spatial patterns.


Delocalization Score Heatmap (mz_group × sample)
m/z Group
Sample
AAD_1
AAD_2
AAD_3
AAD_4
AAD_5
AAD_6
AC_1
AC_2
AC_3
AC_4
AC_5
AC_6
YAD_1
YAD_2
YAD_3
YAD_4
YC_1
YC_2
YC_3
YC_4
YC_5
YC_6
1.0
0.8
0.6
0.4
0.2
0.0

2.4 Conclusion

Quantifying delocalization in MALDI-MSI data remains a critical challenge for spatial omics studies and it can impede accurate biological interpretation and reduce spatial resolution. In this work, we proposed and evaluated several spatial and intensity-based metrics to systematically quantify analyte delocalization across tissue sections. Our findings highlight both the limitations of individual metrics and the strengths of a composite delocalization score that integrates two spatial characteristics into a single interpretable value. While only two features were explicitly included in the formulation, the remaining characteristics are implicitly captured through their latent correlation (Figure 2-11).

2.4.1 Interpretation of Findings

Among the individual metrics examined, the Euclidean distance between the global and tissue-specific centers of mass, the maximum and mean distances of nonzero-intensity background pixels from the tissue boundary, and the total area of nonzero-intensity off-tissue pixels each offered partial but incomplete insights into delocalization. The centroid shift metric was conceptually appealing. However, it was limited by its sensitivity to symmetric off-tissue signals and performed poorly when the delocalization was spatially balanced around the tissue. Similarly, background distance metrics captured spatial dispersion but failed to capture the magnitude distribution. The area of nonzero-intensity background pixels, however, emerged as a more consistent and meaningful indicator of delocalization. It directly reflected the physical spread of analyte presence beyond tissue

boundaries, regardless of symmetry. However, even this metric weakly assessed cases where background activity was restricted to a small region but located far from the tissue, or when off-tissue distribution was widespread yet low in intensity.

To overcome these limitations, we introduced a delocalization score that combines the normalized nonzero-intensity off-tissue area and the normalized mean distance of nonzero-intensity off-tissue pixels to the tissue boundary, using a weighted linear combination parameterized by α. This composite score offered a more robust characterization, capturing both the magnitude and spatial pattern of delocalization. By tuning α to emphasize area ($\alpha = 0.95$), we calibrated the score alongside manual visual inspections, indicating that how large the affected area has a greater impact on the perception of delocalization than how widely the distribution is dispersed across the space.

2.4.2 Comparison to Existing Approaches

Although commercial software such as HDI or SCiLS Lab offers visual interfaces for visually assessing delocalization, few open-source or quantitative tools are available to systematically evaluate or prioritize analyte features based on their off-tissue behavior. To address this limitation, we present a reproducible, data-driven metric that can be applied consistently across large feature sets. While existing approaches often employ unsupervised machine learning and clustering primarily for background removal [32, 35], our delocalization score directly captures both the spatial magnitude and location of background distributions in relation to the tissue boundary. These spatial properties are

essential for understanding molecular localization and enhancing downstream analyses, including molecular clustering, biomarker identification, and integration with histological data.

2.4.3 Limitations and Assumptions

In this study, several assumptions and limitations must be acknowledged. First, our framework relies on accurate segmentation of tissue boundaries. Tissue masks were derived using total ion current (TIC), which may miss weakly ionizing regions or fine tissue structures. Inaccuracies in tissue masking can directly affect the classification of background pixels and therefore the computed delocalization metrics.

Second, the current delocalization score does not capture the directional bias of delocalization (e.g., preferential spread toward a specific tissue edge or off-tissue region). While the metric effectively quantifies the magnitude and spatial extent of delocalization, it remains orientation-agnostic. Future implementations could incorporate the weighted centroid of tissue and surrounding regions to detect directional partiality, which may provide deeper insights into the spatial dynamics of analyte redistribution.

Third, the method does not explicitly distinguish between biologically meaningful off-tissue distributions, such as biologically relevant analyte diffusion into adjacent regions [36], and technical artifacts arising from matrix application [36], laser oversampling [37], or tissue preparation such as sectioning. Integrating orthogonal data sources, such as histological staining or ion mobility separation [38], may help resolve these ambiguities in future studies.

Finally, the delocalization score currently uses a fixed weighting parameter (α) which is optimized empirically on a single dataset. This parameter worked well in our setting, but it is essential to examine and validate its general applicability to other tissue types, ionization protocols, or spatial resolutions.

2.4.4 Factors Contributing to Delocalization in MALDI-MSI

Delocalization in MALDI-MSI can arise from multiple sources throughout the experimental workflow, including matrix choice and deposition method (e.g., crystal size and orientation) [39], solvent spreading during matrix application [40], tissue preparation, handling, and storage [24]. In our study, matrix crystallization (DHB applied by spraying) is likely the largest contributor [41], while factors such as storage may contribute less. Different classes of molecules may be affected to varying degrees: small metabolites and lipids are more prone to diffusion than larger peptides or proteins [42]. To minimize delocalization, careful optimization of matrix type and deposition method, reduction of solvent volume, controlled tissue handling and storage, and appropriate laser settings are recommended. Our delocalization score provides a quantitative framework to evaluate these effects and can guide optimization in future experiments.

2.4.5 Broader Applicability

The proposed framework is modular and generalizable. Not only does it support applications across various MALDI-MSI studies, tissue types, and analyte classes (e.g., lipids, metabolites, peptides), but it is also particularly useful in large-scale studies where

manual inspection is not feasible. This ability enables automated filtering or weighting of analyte features based on their spatial reliability.

Aside from delocalization filtering, the delocalization score may also serve as a quality control metric during methods development or matrix optimization. It allows researchers to quantitatively target the spatial integrity of acquired MSI datasets. It can also enhance spatial omics analyses and facilitate the generation of annotated datasets for training and validation by integrating with existing tools, such as unsupervised learning for clustering or pattern discovery, and supervised frameworks (e.g., deep learning) for tasks like tissue classification or biomarker prediction.

2.5 Code and Data Availability

All analyses in this study were performed on a Mac M2 with 24 GB of memory and 1 TB of SSD storage. Peak detection and construction of the sparse (CSR) intensity matrix for each sample took approximately 10 minutes. Overall, the pipeline was not computationally intensive, and memory and CPU usage remained moderate throughout. Subsequent steps, such as calculation of mean distance and composite scoring, were comparatively fast, taking only a few additional minutes per sample.

All source code developed for this study, including scripts for delocalization metric computation, figure generation, and statistical analyses, is openly available at [GitHub repository link, https://github.com/CIL-UTK/Delocalization_Score]. The repository includes detailed documentation, installation instructions, and example Jupyter notebooks

to facilitate reproducibility and adaptation to other MALDI-MSI datasets. Also, the link for the MALDI-MSI dataset is provided is as link in that repository.

2.6 Acknowledgements

The authors would like to thank the Biological and Small Molecule Mass Spectrometry Core (BSMMSC) for time on the MALDI-MSI and the UTK Advanced Microscopy and Imaging Center for use of their microscope. Funding for Amin Jarrahi, animals, and mass spectrometry provided by UT-Oak Ridge Innovation Institute (UT-ORII) Science Alliance: Support for Affiliated Research Teams (StART) Program. The mouse strain used for this research project, B6;129-Tg(APPSwe,tauP301L)1Lfa *Psen1*^{tm1Mpm}/Mmjax, RRID:MMRRC_034830-JAX, was obtained from the Mutant Mouse Resource and Research Center (MMRRC) at The Jackson Laboratory, an NIH-funded strain repository, and was donated to the MMRRC by Frank Laferla, Ph.D., University of California, Irvine. Mark P. Mattson, Ph.D., Johns Hopkins University, School of Medicine

REFERENCES

1. P. Bourceau *et al.*, “Visualization of metabolites and microbes at high spatial resolution using MALDI mass spectrometry imaging and in situ fluorescence labeling,” *Nature Protocols 2023 18:10*, vol. 18, no. 10, pp. 3050–3079, Sep. 2023, doi: 10.1038/s41596-023-00864-1.
2. J. Kim, G. Dwivedi, B. A. Boughton, A. Sharma, and S. Lee, “Advances in cellular and tissue-based imaging techniques for sarcoid granulomas,” *Am J Physiol Cell Physiol*, vol. 326, no. 1, pp. C10–C26, Jan. 2024, doi: 10.1152/AJPCELL.00507.2023/ASSET/IMAGES/LARGE/AJPCELL.00507.2023_F009.JPEG.
3. B. Li, Y. Zhang, J. Ge, K. Liu, and P. Li, “Sample preparation for mass spectrometry imaging of leaf tissues: a case study on analyte delocalization,” *Anal Bioanal Chem*, vol. 410, no. 28, pp. 7449–7456, Nov. 2018, doi: 10.1007/S00216-018-1355-5/METRICS.
4. M. Race, R. T. Steven, A. D. Palmer, I. B. Styles, and J. Bunch, “Memory efficient principal component analysis for the dimensionality reduction of large mass spectrometry imaging data sets,” *Anal Chem*, vol. 85, no. 6, pp. 3071–3078, Mar. 2013, doi: 10.1021/AC302528V/SUPPL_FILE/AC302528V_SI_002.ZIP.
5. Y.-J. Chen, H.-S. Zeng, H. Jin, and H. Wang, “Applications of mass spectrometry imaging in botanical research,” *Advanced Biotechnology 2024 2:1*, vol. 2, no. 1, pp. 1–20, Feb. 2024, doi: 10.1007/S44307-024-00014-Y.

6. M. Holzlechner, E. Eugenin, and B. Prideaux, “Mass spectrometry imaging to detect lipid biomarkers and disease signatures in cancer,” *Cancer Rep (Hoboken)*, vol. 2, no. 6, Dec. 2019, doi: 10.1002/CNR2.1229.
7. G. O’Hurley *et al.*, “Garbage in, garbage out: A critical evaluation of strategies used for validation of immunohistochemical biomarkers,” *Mol Oncol*, vol. 8, no. 4, p. 783, Jun. 2014, doi: 10.1016/J.MOLONC.2014.03.008.
8. R. Yin, K. E. Burnum-Johnson, X. Sun, S. K. Dey, and J. Laskin, “High spatial resolution imaging of biological tissues using nanospray desorption electrospray ionization mass spectrometry,” *Nat Protoc*, vol. 14, no. 12, pp. 3445–3470, Dec. 2019, doi: 10.1038/S41596-019-0237-4;SUBJMETA=11,296,638,639,942;KWRD=IMAGING+STUDIES,MASS+SPECTROMETRY.
9. J. M. Wiseman, D. R. Ifa, Q. Song, and R. G. Cooks, “Tissue Imaging at Atmospheric Pressure Using Desorption Electrospray Ionization (DESI) Mass Spectrometry,” *Angewandte Chemie International Edition*, vol. 45, no. 43, pp. 7188–7192, Nov. 2006, doi: 10.1002/ANIE.200602449.
10. K. Qi, L. Wu, C. Liu, and Y. Pan, “Recent Advances of Ambient Mass Spectrometry Imaging and Its Applications in Lipid and Metabolite Analysis,” *Metabolites 2021, Vol. 11, Page 780*, vol. 11, no. 11, p. 780, Nov. 2021, doi: 10.3390/METABO11110780.

11. D. M. G. Anderson *et al.*, “Lipid Landscape of the Human Retina and Supporting Tissues Revealed by High-Resolution Imaging Mass Spectrometry,” *J Am Soc Mass Spectrom*, vol. 31, no. 12, pp. 2426–2436, Dec. 2020, doi: 10.1021/JASMS.0C00119/SUPPL_FILE/JS0C00119_SI_001.PDF.
12. S. A. Iakab, P. Ràfols, X. Correig-Blanchar, and M. García-Altares, “Perspective on Multimodal Imaging Techniques Coupling Mass Spectrometry and Vibrational Spectroscopy: Picturing the Best of Both Worlds,” *Anal Chem*, vol. 93, no. 16, pp. 6301–6310, Apr. 2021, doi: 10.1021/ACS.ANALCHEM.0C04986/ASSET/IMAGES/LARGE/AC0C04986_0004.JPEG.
13. D. Aoki, K. Saito, Y. Matsushita, and K. Fukushima, “Distribution of Cell Wall Components by TOF-SIMS,” *Secondary Xylem Biology: Origins, Functions, and Applications*, pp. 363–379, Jan. 2016, doi: 10.1016/B978-0-12-802185-9.00017-6.
14. H. Liu *et al.*, “Matrix-assisted laser desorption/ionization mass spectrometry imaging (MALDI MSI) for in situ analysis of endogenous small molecules in biological samples,” *TrAC Trends in Analytical Chemistry*, vol. 157, p. 116809, Dec. 2022, doi: 10.1016/J.TRAC.2022.116809.
15. N. R. Phulara and H. K. Seneviratne, “Visualization of Efavirenz-Induced Lipid Alterations in the Mouse Brain Using MALDI Mass Spectrometry Imaging,” *Curr Protoc*, vol. 5, no. 2, p. e70108, Feb. 2025, doi:

10.1002/CPZ1.70108;REQUESTEDJOURNAL:JOURNAL:26911299;WGROUP:STRING:PUBLICATION.

16. C. Costa *et al.*, “Exploring New Methods to Study and Moderate Proton Beam Damage for Multimodal Imaging on a Single Tissue Section,” *J Am Soc Mass Spectrom*, vol. 33, no. 12, pp. 2263–2272, Dec. 2022, doi: 10.1021/JASMS.2C00226/ASSET/IMAGES/LARGE/JS2C00226_0006.JPEG.
17. L. X. Jiang, M. Yang, S. N. Wali, and J. Laskin, “High-throughput mass spectrometry imaging of biological systems: Current approaches and future directions,” *TrAC Trends in Analytical Chemistry*, vol. 163, p. 117055, Jun. 2023, doi: 10.1016/J.TRAC.2023.117055.
18. X. Ma and F. M. Fernández, “Advances in mass spectrometry imaging for spatial cancer metabolomics,” *Mass Spectrom Rev*, vol. 43, no. 2, pp. 235–268, Mar. 2024, doi: 10.1002/MAS.21804.
19. J. Leopold, P. Prabutzki, K. M. Engel, and J. Schiller, “A Five-Year Update on Matrix Compounds for MALDI-MS Analysis of Lipids.,” *Biomolecules*, vol. 13, no. 3, pp. 546–546, Mar. 2023, doi: 10.3390/BIOM13030546.
20. D. L. Baluya, T. J. Garrett, and R. A. Yost, “Automated MALDI Matrix Deposition Method with Inkjet Printing for Imaging Mass Spectrometry,” *Anal Chem*, vol. 79, no. 17, pp. 6862–6867, Sep. 2007, doi: 10.1021/AC070958D.
21. Iloro, A. Bueno, J. Calvo, H. Urreta, and F. Elortza, “Langartech: A Custom-Made MALDI Matrix Sprayer for MALDI Imaging Mass Spectrometry,” *J Lab*

Autom, vol. 21, no. 2, pp. 260–267, Apr. 2016, doi: 10.1177/2211068215607320/ASSET/E6C777C5-631E-48B9-A742-CF4DBA069F5C/ASSETS/IMAGES/LARGE/10.1177_2211068215607320-FIG7.JPG.

22. Y. Chen *et al.*, "Imaging MALDI mass spectrometry using an oscillating capillary nebulizer matrix coating system and its application to analysis of lipids in brain from a mouse model of Tay-Sachs/Sandhoff disease," *Anal Chem*, vol. 80, no. 8, pp. 2780–2788, Apr. 2008, doi: 10.1021/AC702350G/ASSET/IMAGES/LARGE/AC702350GF00008.JPEG.
23. J. M. Fonville *et al.*, "Robust data processing and normalization strategy for MALDI mass spectrometric imaging," *Anal Chem*, vol. 84, no. 3, pp. 1310–1319, Feb. 2012, doi: 10.1021/AC201767G/SUPPL_FILE/AC201767G_SI_001.PDF.
24. C. Tressler *et al.*, "Factorial Design to Optimize Matrix Spraying Parameters for MALDI Mass Spectrometry Imaging," *J Am Soc Mass Spectrom*, vol. 32, no. 12, pp. 2728–2737, Dec. 2021, doi: 10.1021/JASMS.1C00081/ASSET/IMAGES/LARGE/JS1C00081_0006.JPEG.
25. J. H. Holbrook, G. E. Kemper, and A. B. Hummon, "Quantitative mass spectrometry imaging: therapeutics & biomolecules," *Chemical Communications*, vol. 60, no. 16, pp. 2137–2151, Feb. 2024, doi: 10.1039/D3CC05988J.
26. Dannhorn *et al.*, "Universal Sample Preparation Unlocking Multimodal Molecular Tissue Imaging," *Anal Chem*, vol. 92, no. 16, pp. 11080–11088, Aug. 2020, doi:

10.1021/ACS.ANALCHEM.0C00826/ASSET/IMAGES/LARGE/AC0C00826_0006.JPEG.

27. S. Oddo *et al.*, “Triple-transgenic model of Alzheimer’s Disease with plaques and tangles: Intracellular Aβ and synaptic dysfunction,” *Neuron*, vol. 39, no. 3, pp. 409–421, Jul. 2003, doi: 10.1016/S0896-6273(03)00434-3.
28. K. B. J. Franklin and George Paxinos, *The Mouse Brain in Stereotaxic Coordinates: Compact, The Coronal Plates and Diagrams.*, 3rd ed. Academic Press, 2008.
29. J. Hanrieder, A. Ljungdahl, and M. Andersson, “MALDI Imaging Mass Spectrometry of Neuropeptides in Parkinson’s Disease,” *J Vis Exp*, no. 60, p. 3445, Feb. 2012, doi: 10.3791/3445.
30. C. Yang, Z. He, and W. Yu, “Comparison of public peak detection algorithms for MALDI mass spectrometry data analysis,” *BMC Bioinformatics*, vol. 10, no. 1, pp. 1–13, Jan. 2009, doi: 10.1186/1471-2105-10-4/TABLES/2.
31. Savitzky and M. J. E. Golay, “Smoothing and Differentiation of Data by Simplified Least Squares Procedures,” *Anal Chem*, vol. 36, no. 8, pp. 1627–1639, Jul. 1964, doi: 10.1021/AC60214A047/ASSET/AC60214A047.FP.PNG_V03.
32. González-Fernández, A. Dexter, C. J. Nikula, and J. Bunch, “NECTAR: A New Algorithm for Characterizing and Correcting Noise in QToF-Mass Spectrometry Imaging Data,” *J Am Soc Mass Spectrom*, vol. 34, no. 11, pp. 2443–2453, Nov.

2023, doi: 10.1021/JASMS.3C00116/ASSET/IMAGES/LARGE/JS3C00116_0008.JPEG.

33. S. O. Deininger *et al.*, "Normalization in MALDI-TOF imaging datasets of proteins: Practical considerations," *Anal Bioanal Chem*, vol. 401, no. 1, pp. 167–181, Jul. 2011, doi: 10.1007/S00216-011-4929-Z/FIGURES/9.
34. D. D. Krutkin *et al.*, "To Impute or Not To Impute in Untargeted Metabolomics─That is the Compositional Question," *J Am Soc Mass Spectrom*, vol. 36, no. 4, pp. 742–759, Apr. 2025, doi: 10.1021/JASMS.4C00434/SUPPL_FILE/JS4C00434_SI_001.PDF.
35. L. Blanc, F. Grelard, M. Tuck, V. Dartois, A. Peixoto, and N. Desbenoit, "In tissue spatial single-cell metabolomics by coupling mass spectrometry imaging and immunofluorescences," *bioRxiv*, p. 2024.03.22.586317, Mar. 2024, doi: 10.1101/2024.03.22.586317.
36. K. Ščupáková *et al.*, "Cellular resolution in clinical MALDI mass spectrometry imaging: the latest advancements and current challenges," *Clin Chem Lab Med*, vol. 58, no. 6, p. 914, Jun. 2020, doi: 10.1515/CCLM-2019-0858.
37. J. C. McKinnon *et al.*, "MALDI-2-Enabled Oversampling for the Mass Spectrometry Imaging of Metabolites at Single-Cell Resolution," *J Am Soc Mass Spectrom*, vol. 35, no. 11, pp. 2729–2742, Nov. 2024, doi: 10.1021/JASMS.4C00241/ASSET/IMAGES/LARGE/JS4C00241_0007.JPEG.

38. D. C. Koomen, J. C. May, and J. A. McLean, “Insights and prospects for ion mobility-mass spectrometry in clinical chemistry,” *Expert Rev Proteomics*, vol. 19, no. 1, p. 17, 2022, doi: 10.1080/14789450.2022.2026218.

39. R. J. A. Goodwin, “Sample preparation for mass spectrometry imaging: Small mistakes can lead to big consequences,” *J Proteomics*, vol. 75, no. 16, pp. 4893–4911, Aug. 2012, doi: 10.1016/J.JPROT.2012.04.012.

40. [40] C. Meriaux, J. Franck, M. Wisztorski, M. Salzet, and I. Fournier, “Liquid ionic matrixes for MALDI mass spectrometry imaging of lipids,” *J Proteomics*, vol. 73, no. 6, pp. 1204–1218, Apr. 2010, doi: 10.1016/J.JPROT.2010.02.010.

41. T. Mahamdi *et al.*, “Improving MALDI Mass Spectrometry Imaging Performance: Low-Temperature Thermal Evaporation for Controlled Matrix Deposition and Improved Image Quality,” *J Am Soc Mass Spectrom*, vol. 36, no. 5, pp. 1100–1110, May 2025, doi: 10.1021/JASMS.5C00015/SUPPL_FILE/JS5C00015_SI_001.PDF.

42. R. Buchberger, K. DeLaney, J. Johnson, and L. Li, “Mass Spectrometry Imaging: A Review of Emerging Advancements and Future Insights,” *Anal Chem*, vol. 90, no. 1, p. 240, Jan. 2017, doi: 10.1021/ACS.ANALCHEM.7B04733.

SUPPLEMENTARY MATERIAL

Figure 2-15 illustrates the spatial distribution of m/z 844.52 across the young control samples (YC_1, YC_2, YC_5), which exhibited somewhat different behavior in the Pearson correlation heatmap (Figure 2-13). The observed variation is likely influenced by technical factors rather than biological differences. Notably, sample YC_1 experienced crack during the sectioning and a physical cut during the staining, which may have altered its spatial profile. The corresponding H&E-stained images reveal this cut clearly in YC_1, as well as a smaller cut in YC_2, providing visual evidence of how tissue handling artifacts can affect spatial delocalization measurements.

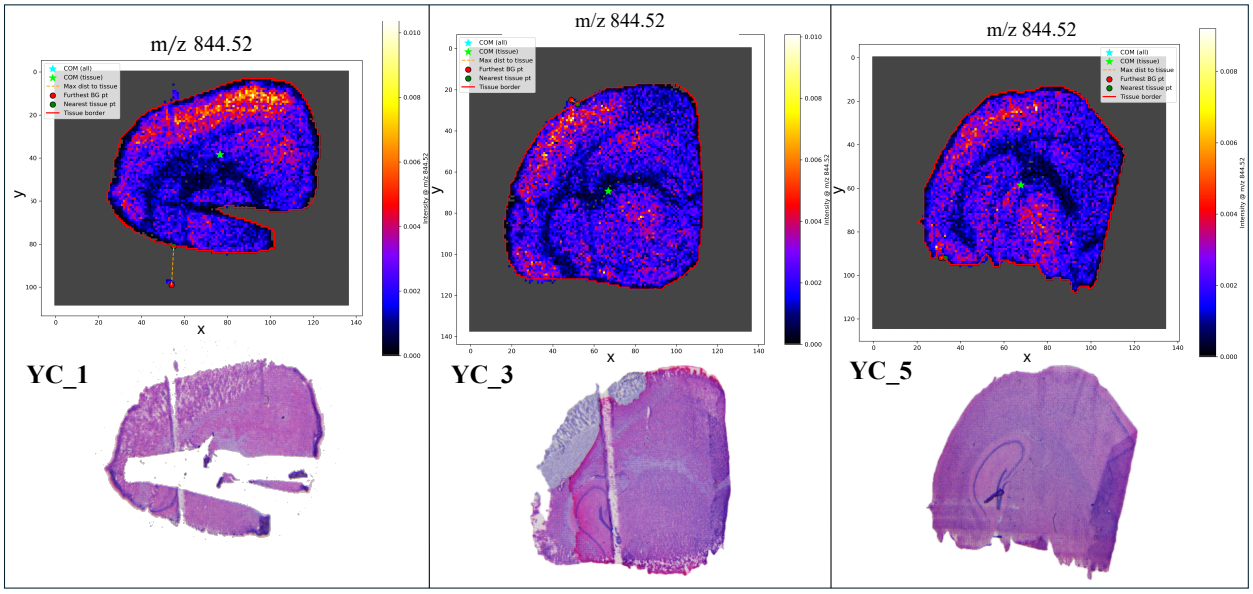


Figure 2-15 Spatial distribution of m/z 844.52 in young control samples (from left to right: YC_1, YC_2, YC_5) with corresponding H&E-stained sections. The first row shows ion intensity maps of m/z 844.52, and the second row shows the corresponding histological images.

Figure 2-16 shows the optical images of all 24 stained tissue sections included in this study. Samples YC_1 and YC_3 exhibited minor physical cuts that occurred during the staining process, while the whole-brain tissue of YAD_5 was partially damaged during brain extraction.

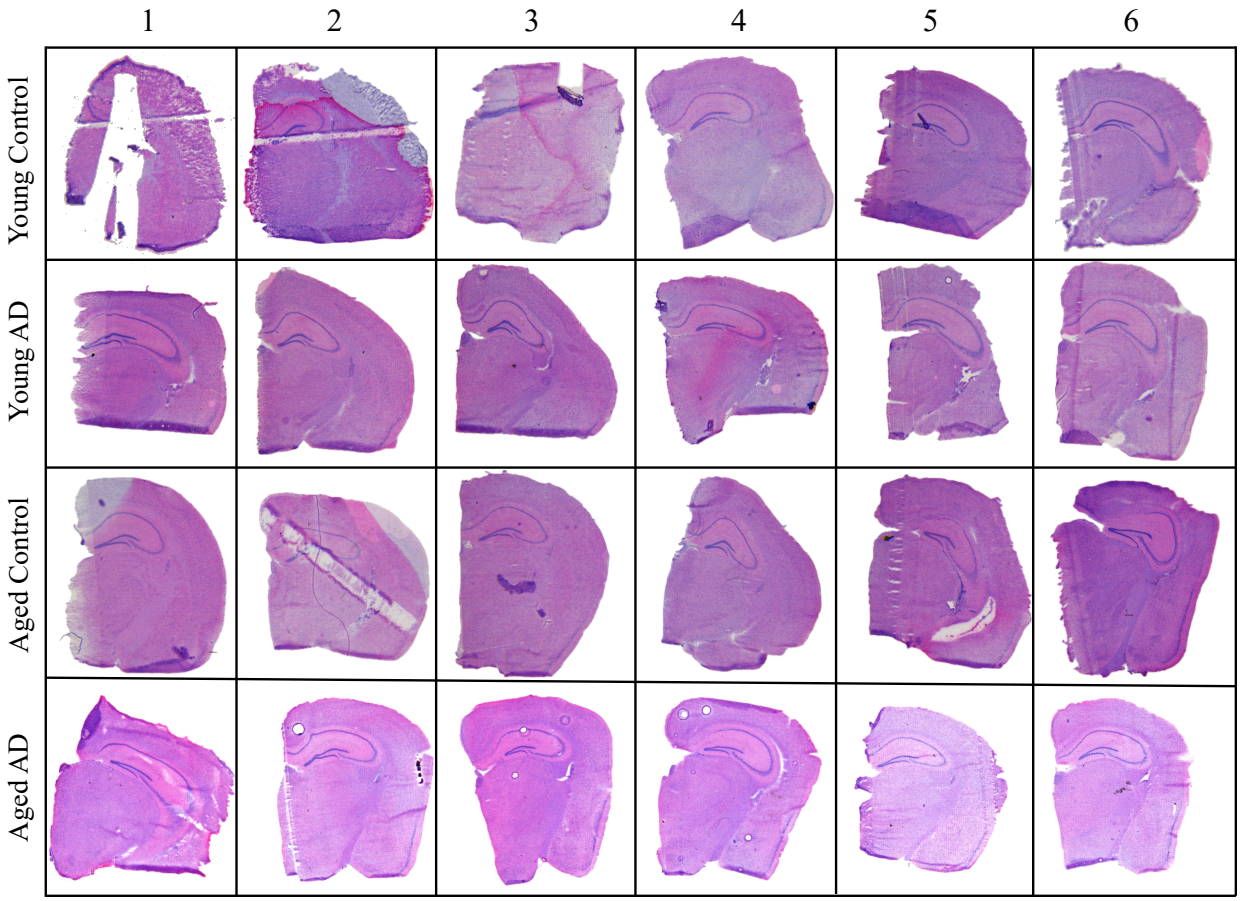


Figure 2-16 Optical images of all stained tissue sections. Sample groups include Young Control (YC), Young AD (YAD), Aged Control (AC), and Aged AD (AAD).

CHAPTER 3:

DESIGNING AN AI-ENHANCED ALGORITHM FOR MSI-BASED LIPIDOMICS STUDIES

3.1 Introduction

Among MSI methods, matrix-assisted laser desorption/ionization mass spectrometry imaging (MALDI-MSI) allows a label-free detection of hundreds to thousands of molecular species, including lipids, metabolites, and small peptides. It is considered a soft ionization technique, so it preserves the spatial architecture of biological tissues [1-6]. By collecting a full mass spectrum at each pixel location, MSI generates high-dimensional datasets that contain both molecular and spatial information. This ability to link chemical composition with anatomical structure makes MSI particularly valuable for investigating complex biological systems in which molecular heterogeneity is closely tied to tissue organization.

Despite these advantages, the large size and complexity of MSI datasets create significant analytical challenges. A single experiment can generate thousands of m/z features across tens of thousands of spatial locations, producing high-dimensional data matrices that are difficult to interpret using conventional statistical approaches. In addition, the spatial nature of the data adds another layer of complexity because biologically meaningful signals are rarely uniform. Instead, they often appear as localized spatial patterns rather than simple global intensity changes. These challenges highlight the need for computational tools that can systematically explore MSI datasets and identify molecular features that exhibit biologically meaningful spatial behavior [5-7].

In recent years, advances in artificial intelligence (AI) and machine learning have provided new opportunities for analyzing complex biological data. Unsupervised learning

approaches – including dimensionality reduction and clustering methods – are particularly useful for discovering patterns in high-dimensional datasets without prior labeling or assumptions about the underlying structure. When applied to MSI data, these methods can reveal spatially organized molecular patterns, group pixels with similar chemical signatures, and highlight analytes that contribute most strongly to spatial variability across the tissue. These approaches provide a scalable framework for exploring large MSI datasets and generating new hypotheses about molecular organization in biological systems [7, 8].

To refine and validate these computational strategies, it is essential to apply them to a biological model characterized by structured molecular change. Aging and neurodegenerative diseases provide such a context. Alzheimer's disease (AD), a progressive neurodegenerative disorder and the sixth leading cause of death in the United States, is characterized by complex molecular and cellular changes that accumulate with age [9,10]. Lipids represent a major component of brain tissue, including phospholipids, glycolipids, cholesterol, and polyunsaturated fatty acids, which play essential roles in neuronal membrane structure and synaptic function [11-13]. Dysregulation of lipid metabolism has been implicated in multiple aspects of AD pathology, including synaptic dysfunction, mitochondrial impairment, oxidative stress, and neuroinflammation [14-16]. Because MSI preserves tissue architecture while mapping molecular distributions, it provides a unique opportunity to investigate how lipid alterations are spatially organized across brain regions. Previous studies using MALDI-MSI have reported region-specific

changes in lipid species such as phosphatidylcholines, sphingomyelins, and ceramides in neurodegenerative conditions [1, 7]. However, extracting meaningful biological insights from these datasets requires analytical frameworks that can efficiently process large numbers of spectra, identify important molecular features, and characterize their spatial relationships.

In this context, this chapter presents the development and application of a computational framework designed to enable large-scale analysis of MSI datasets and facilitate the identification of spatially informative analytes. The framework is intended to support hypothesis generation by enabling systematic exploration of complex MSI data and revealing broader molecular patterns that may not be apparent through conventional analyses. The approach integrates preprocessing, dimensionality reduction, clustering, and classification techniques to systematically explore MSI lipidomics data and highlight molecular species associated with spatial variability. To demonstrate the utility of this framework, it was applied to MALDI-MSI data acquired from mouse brain sections representing young, aged, and AD models. This application provides a biologically relevant test case while illustrating how computational analysis of MSI data can reveal molecular patterns associated with aging and neurodegeneration.

3.2 Methodology

3.2.1 Sample Preparation and MALDI-MSI Acquisition

Murine brains from young, aged, and 3xTg-AD transgenic mice and controls (female, n=4 per group) were harvested, flash-frozen in liquid nitrogen, and cryosectioned at 10

µm thickness with Leica Cryostat. Sections containing the cortex and hippocampus – at a location approximately 2.24 ± 0.21 mm from the bregma – mounted on glass microscope slides (Fisher Scientific) for MALDI-MSI analysis and the next serial section mounted on Visium slides for spatial transcriptomics (Chapter 4) [20-22]. For the MALDI-MSI, the samples were submerged in four consecutive baths containing isotonic ammonium formate solution (150 mM for 45 seconds. Following the washing steps, the slides were gently dried under a continuous stream of nitrogen gas to minimize moisture and prevent analyte delocalization. To further ensure complete dehydration and stabilization of the tissue before subsequent analyses, the dried slides were placed in a vacuum desiccator for 20 minutes [17]. A matrix solution DMACA was applied through the sublimation method (HTX SubliMATE) with optimized parameters (Table 3-1) [17]. The washing, sublimating, scanning, and mass spectrometry imaging were performed at Mass Spectrometry Research Center (MSRC) at Vanderbilt University.

The slides were first fluorescence-scanned using a ZEISS Axio Scan.Z1 slide scanner. They were then analyzed with Bruker's timsTOF fleX mass spectrometer under the parameters listed in Table 3-2. Following imaging, the matrix was removed by three sequential washes in pure methanol, each lasting 1 minute. Finally, the slides were stained and imaged at 2× resolution using a Leica M205 FCA microscope at the Advanced Microscopy and Imaging Core (AMIC) of the University of Tennessee, Knoxville (Figure 3-1).

Table 3-1 DMACA matrix sublimation conditions

Concentration (mg/mL)	Temperature of Sublimation (°C)	Temperature of the Sample (°C)	Time of Sublimation (min)	Crystalization Temperature (°C)	Crystalization Time (s)
3	180	-78	15	100	30

Table 3-2 MALDI timsTOF fleX instrumental parameters for positive ion mode

Transfer	MALDI Plate Offset (V)	50.0
	Deflection 1 Delta (V)	70.0
	Funnel 1 RF (Vpp)	350.0
	isCID Energy (eV)	10.0
	Funnel 1 RF (Vpp)	300.0
	Multipole 1 RF (Vpp)	300.0
Collision Cell	Collision Energy (eV)	10.0
	Collision RF (Vpp)	2500.0
Quadrupole	Ion Energy (eV)	5.0
	Low Mass (analyte)	450.0
Focus Pre Tof	Transfer Time (µs)	90.0
	Pre Pulse Storage (µs)	12.0
Detection	High Sensitivity Detection – Low Sample Amount	False
	Focus Mode	False

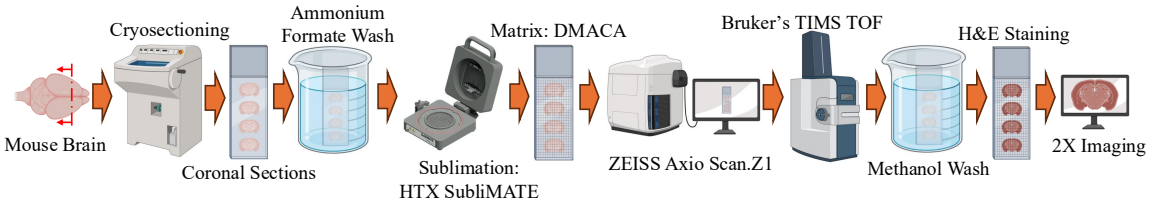


Figure 3-1 Schematic overview of the experimental workflow for MALDI-MSI, highlighting tissue preparation, matrix application, data acquisition, and imaging.

3.2.2 Preprocessing of MALDI-MSI Data

Data preprocessing included different mathematical operations such as smoothing (Savitzky-Golay filter) [18], baseline correction (asymmetric least squares (AsLS)) [19],

peak alignment using 4 lock-masses including 496.339 (LPC 16:0) [20, 21], 734.569 (PC 32:0)[20, 22], 760.585 (PC 34:1)[20, 21, 23], and 806.569 (PC 38:6)[23] as well as drift line, global peak detection (minimum peak distance between two consecutive peak = 0.024Da), peak augmentation (threshold ± 0.012Da or up to ± 2 index), and total ion current (TIC) normalization. During the data preprocessing, Leiden clustering was used in low resolution (res =0.03) to catch the off-tissue pixels and the pixels having contamination. After filtering those pixels, some of the preprocessing steps were repeated for the data (Figure 3-2).

3.2.3 Dimensionality Reduction and Clustering

Dimensionality reductions using PCA, t-SNE, and UMAP were applied to extract dominant analyte patterns. Clustering approaches, including Leiden, K-means, and KNN, were identify subgroups of samples with similar analyte profiles, enabling the discovery of spatially coherent analyte clusters associated with aging and AD pathology [7, 8]. Clustering models were applied to all brain tissue samples to identify regions within each

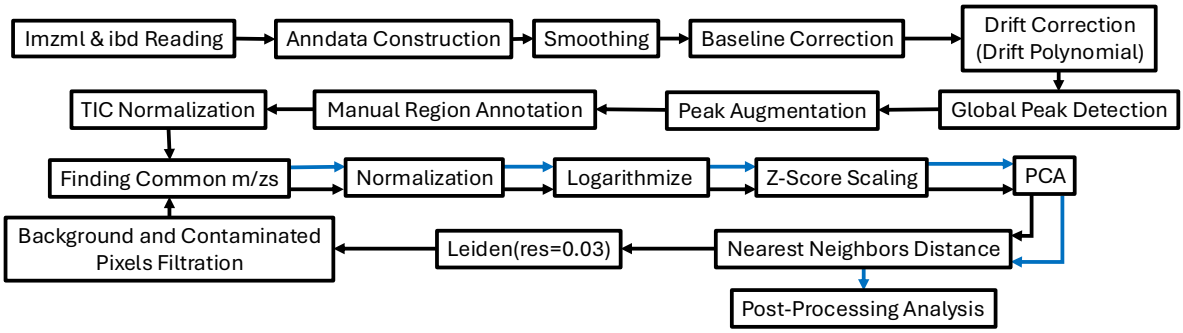


Figure 3-2 Flowchart summarizing the preprocessing steps for MALDI-MSI data implemented in Python.

sample. These regions were classified according to AD, aging, or control groups. The most influential analyte features driving the formation of each region were identified, enabling cross-group comparisons. This approach can facilitate biomarker discovery. [1-7].

3.2.4 Quantitative Study

For each common analyte feature across samples, the maximum and mean intensities were calculated both with and without normalization. Group-level statistics were then obtained by averaging these values across biological replicates. Fold changes and statistical significance (p-values) were computed for relevant group comparisons to identify differentially abundant features.

3.2.5 Finding Corresponding Lipids

After identifying the important analytes, the next step was to determine whether these signals were truly associated with tissue regions rather than matrix-related or background signals. Analytes confirmed to be tissue-related were then subjected to lipid identification by matching their measured masses with entries in reference lipid databases. Finally, the spatial distributions of the identified lipids were mapped onto anatomical brain regions to evaluate reproducibility and biological relevance, ensuring that the observed localization patterns were consistent with known tissue structures.

3.3 Results and Discussion

To showcase the utility of this method, the ion images were generated for a specific analyte that varied with AD. The intensity of analyte 921.69 decreased in aged AD compared to aged controls samples across the hippocampus, cortex, and brainstem regions as evident by the magnitude of the scale (Figure 3-3). To confirm the observed signal is tissue-derived rather than originating from the matrix, Figure 3-4 presents the same analyte without exclusion of background, noise, and contamination pixels, confirming that the observed signal is tissue-derived rather than originating from the matrix. Based on accurate mass matching using the LIPID MAPS database (https://www.lipidmaps.org), the analyte at 921.69 was tentatively assigned to PI O-41:1 ([M+H]+; Δ = 0.0110 Da, $\approx$ 11.9 ppm). However, MS/MS analysis is required for definitive structural confirmation. This result showed our code could generate accurate images without the need for commercial software (e.g. Bruker SCiLS™ Lab).

3.3.1 Unsupervised Dimensionality Reduction and Clustering Reveal Group Differences

Based on the PCA variance ratio plot (log scale) of the MSI dataset, *PC1* captures the majority of the variance, while *PC2* and *PC3* contribute progressively smaller amounts. From *PC4* onward, the explained variance decreases sharply and begins to level off, with *PC* s beyond 6 capturing very little variance (log10 variance ratio below -4.3, corresponding to <0.01% of total variance). Therefore, the elbow point is approximately

at *PC4- PC6*, suggesting that retaining the first six principal components is sufficient to capture the majority of meaningful variation in the data for downstream analyses such as clustering or dimensionality reduction. *PC*s beyond this range likely reflect noise or minor variations and are not expected to contribute substantially to biological interpretation. Although the first 6 *PC*s are enough for interpretation, computing 50 *PC*s ensures the neighbor graph captures subtle differences in pixel intensities for robust Leiden clustering.

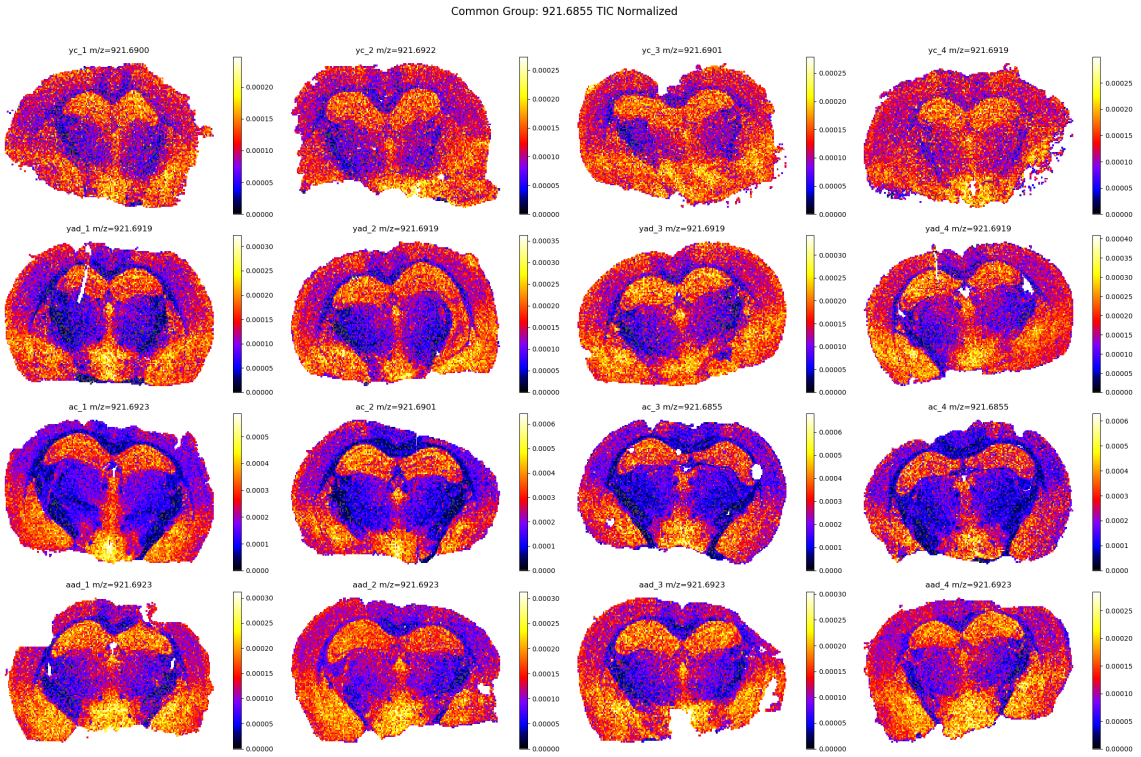


Figure 3-3 TIC-normalized spatial distribution maps of analyte 921.69 across all samples after filtering background, probable contamination, and noise pixels, with intensity scaled to the 99.9th percentile (Vmax).

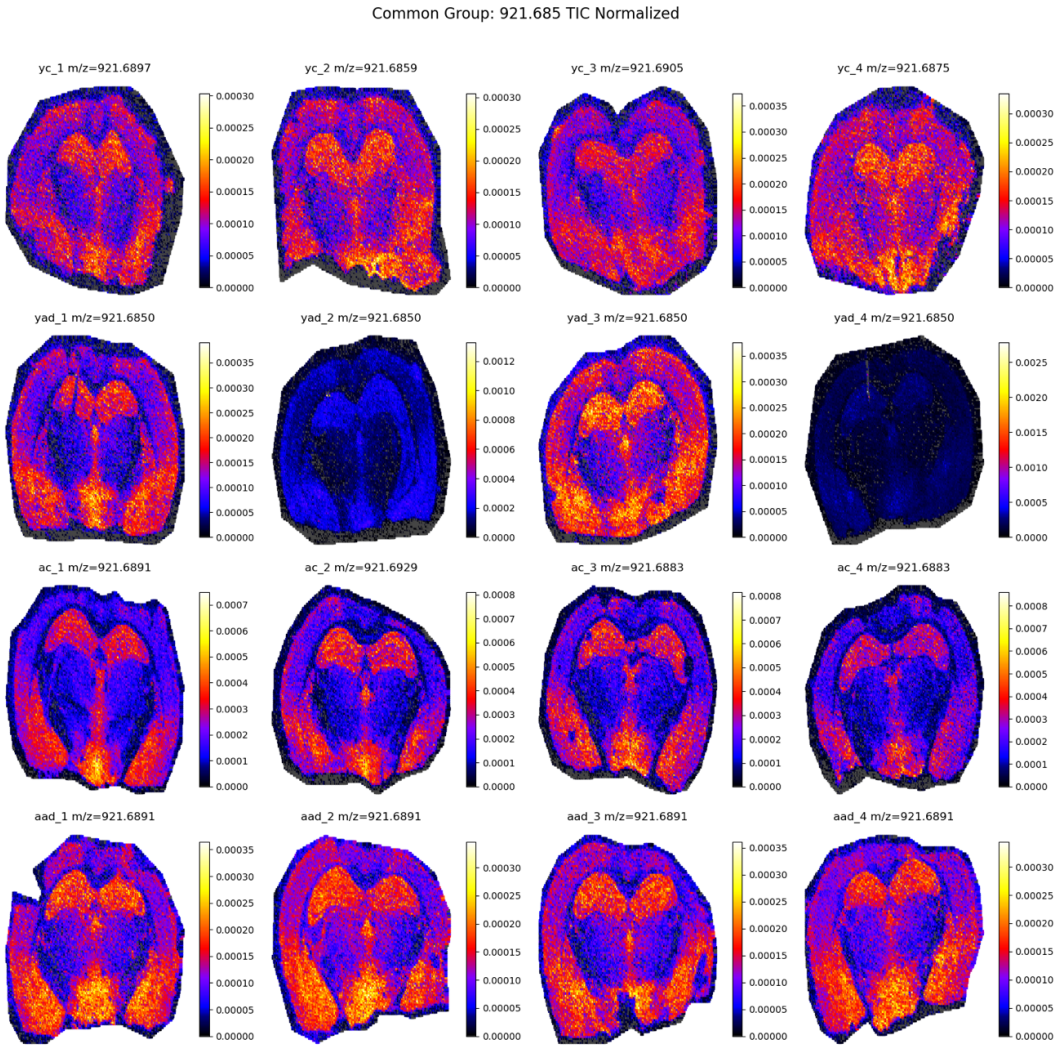


Figure 3-4 TIC-normalized spatial distribution maps of analyte 921.69 across all samples without filtering for background, probable contamination, or noise pixels.

Uniform Manifold Approximation and Projection (UMAP) and t-Distributed Stochastic Neighbor Embedding (t-SNE), K-means, as well as Leiden clustering, were performed based on the selected principal components to reveal spatial patterns and tissue-specific groupings. Figure 3-6 shows the t-SNE and UMAP plots for the common analyte features across all samples. Both algorithms were able to capture differences between young and

aged samples, suggesting that the analyte features contributing to this separation may be potential candidates associated with aging. However, a batch effect cannot be ruled out, as young samples were processed on the first day and aged samples on the following day.

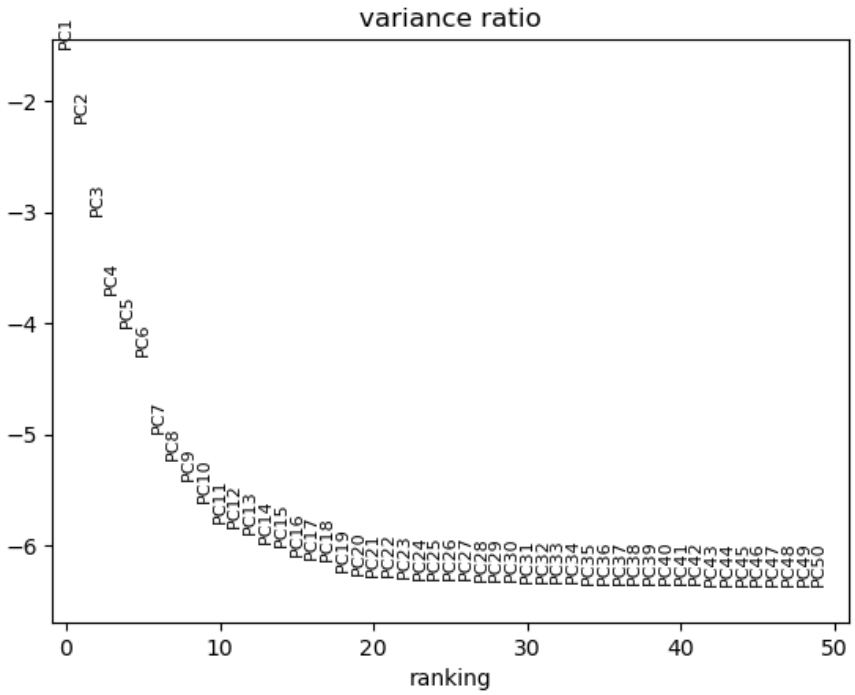


Figure 3-5 PCA variance ratio plot of the first 50 components, with variance values shown on a logarithmic scale.

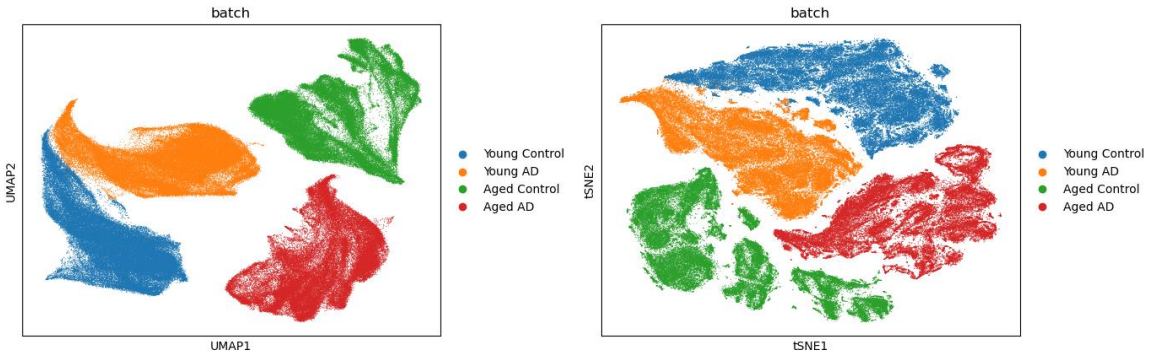


Figure 3-6 Dimensionality reduction visualization using UMAP (left) and t-SNE (right), with data points colored according to group identity.

Machine bias from the MSI instrument is unlikely to account for the separation observed between control and AD samples. The analyte features contributing to this separation may represent potential AD-related candidates. It is also worth noting that the matrix was applied separately for each slide, which could introduce another source of batch effect. Leiden clustering with a resolution of 1.5 was applied to the dataset (Figure 3-7), allowing regional separation of pixels based on the analyte profiles. Following clustering, the analyte features contributing most strongly to each region were identified using their

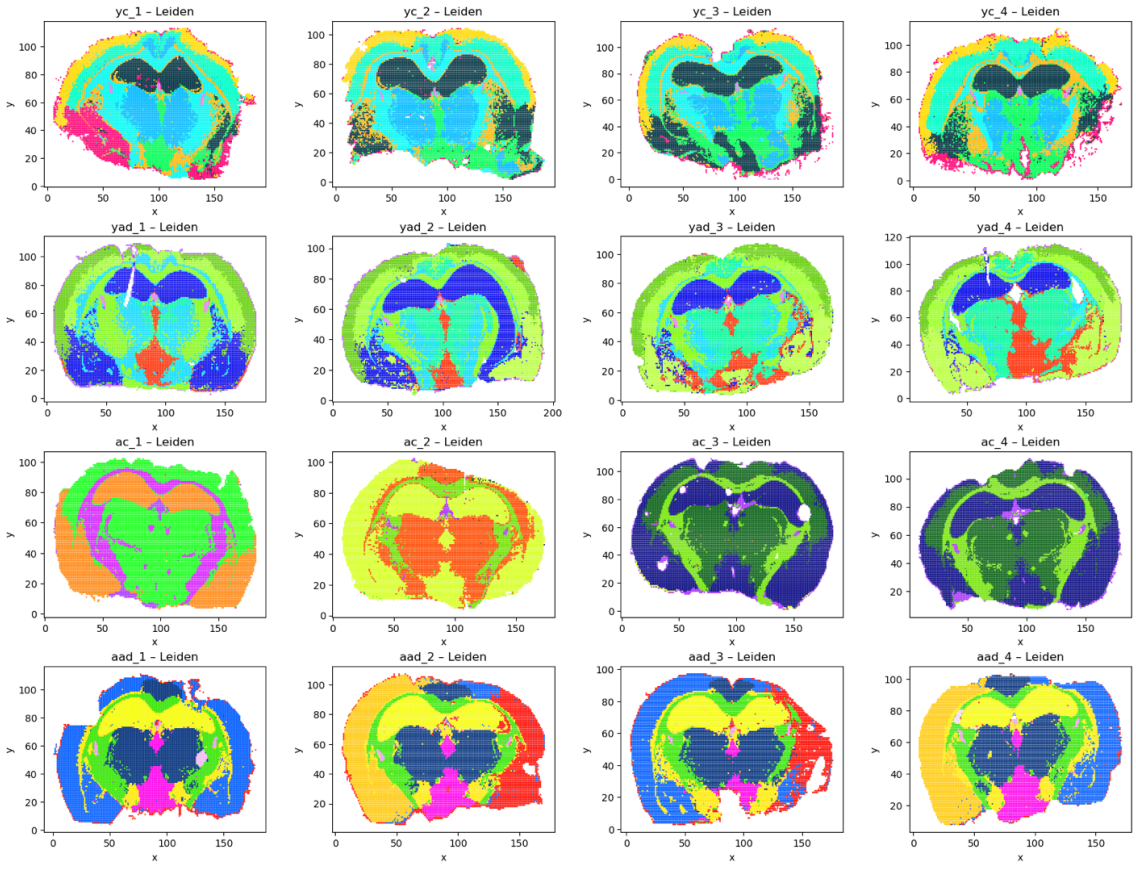


Figure 3-7 Leiden clustering of the dataset at a resolution of 1.5, resulting in 38 distinct clusters that separate pixels according to their analyte intensity profiles.

corresponding scores (Figure 3-8 andFigure 3-9), highlighting the analytes driving regional distinctions across the tissue and samples.

As previously noted, these findings must be interpreted within a biological context. The most prominent analyte features should be evaluated to confirm that they originate from tissue rather than from matrix interference or experimental noise. Once validated, these features can be mapped to potential lipid identities through mass matching with reference databases, generating candidates for subsequent structural validation and functional analysis.

3.3.2 Differential Analyte Analysis Across Experimental Groups

Quantitative and statistical analyses revealed analyte fold changes and their corresponding p-values across different comparisons. Results were sorted by their shared analyte values and visualized using heatmaps and volcano plots. Figure 3-10 illustrates the heatmap of all common analyte features ranging from 758.56 to 798.54, highlighting analyte group comparisons with significant p-values.

Figure 3-11 and Figure 3-12 present the volcano plots for each comparison, showing significantly upregulated and downregulated analytes, with the top five most significant features labeled. Based on the results, the computational framework successfully identified the most significant analytes across the different comparisons. These analytes can subsequently be examined to determine whether they correspond to tissue-specific signals and to evaluate their biological relevance, providing candidates for future mechanistic investigations.

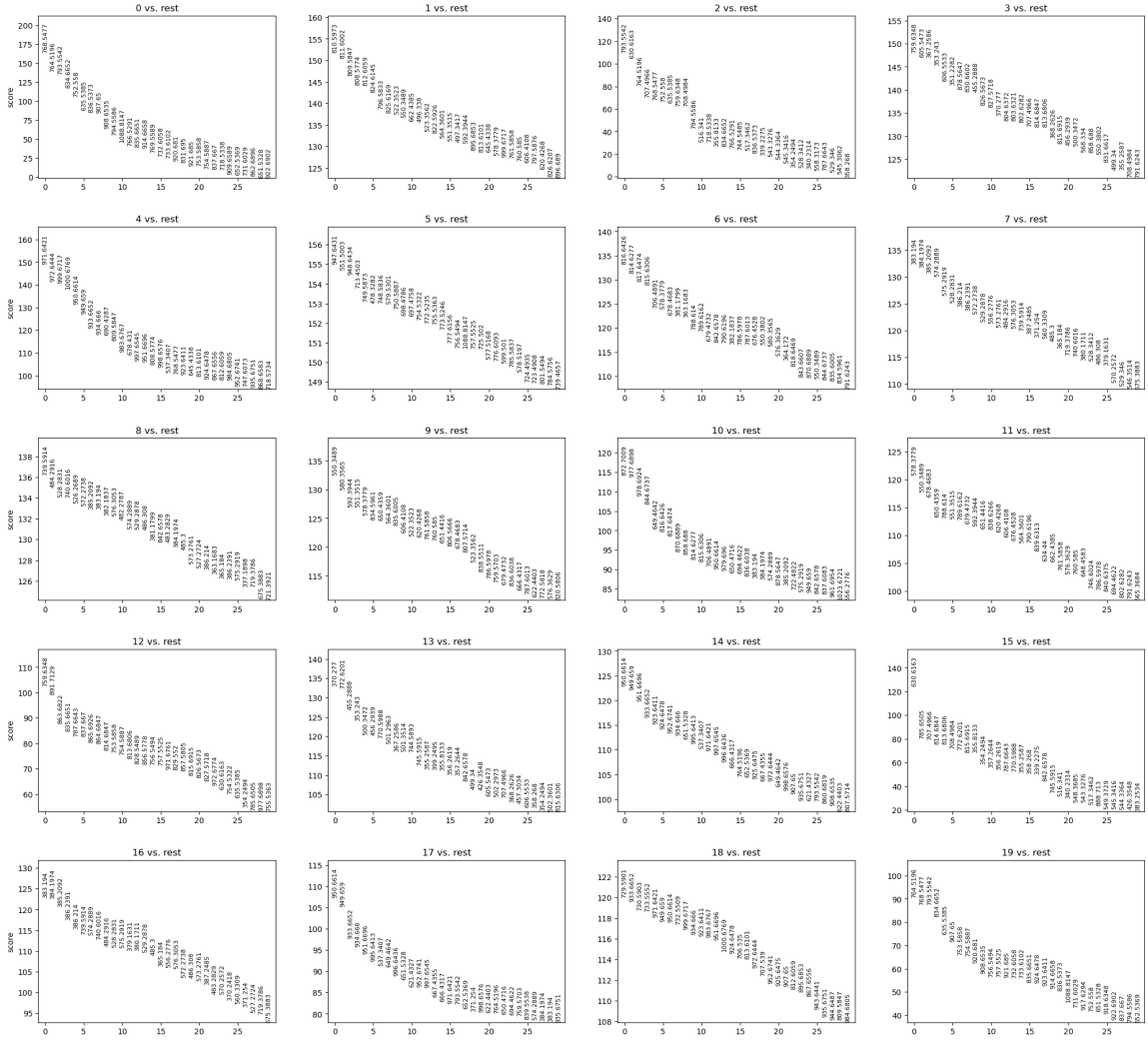


Figure 3-8 Top contributing analyte features for each cluster (clusters 0-19) were identified based on their corresponding scores, highlighting the analyte signals driving regional distinctions across the tissue and samples.

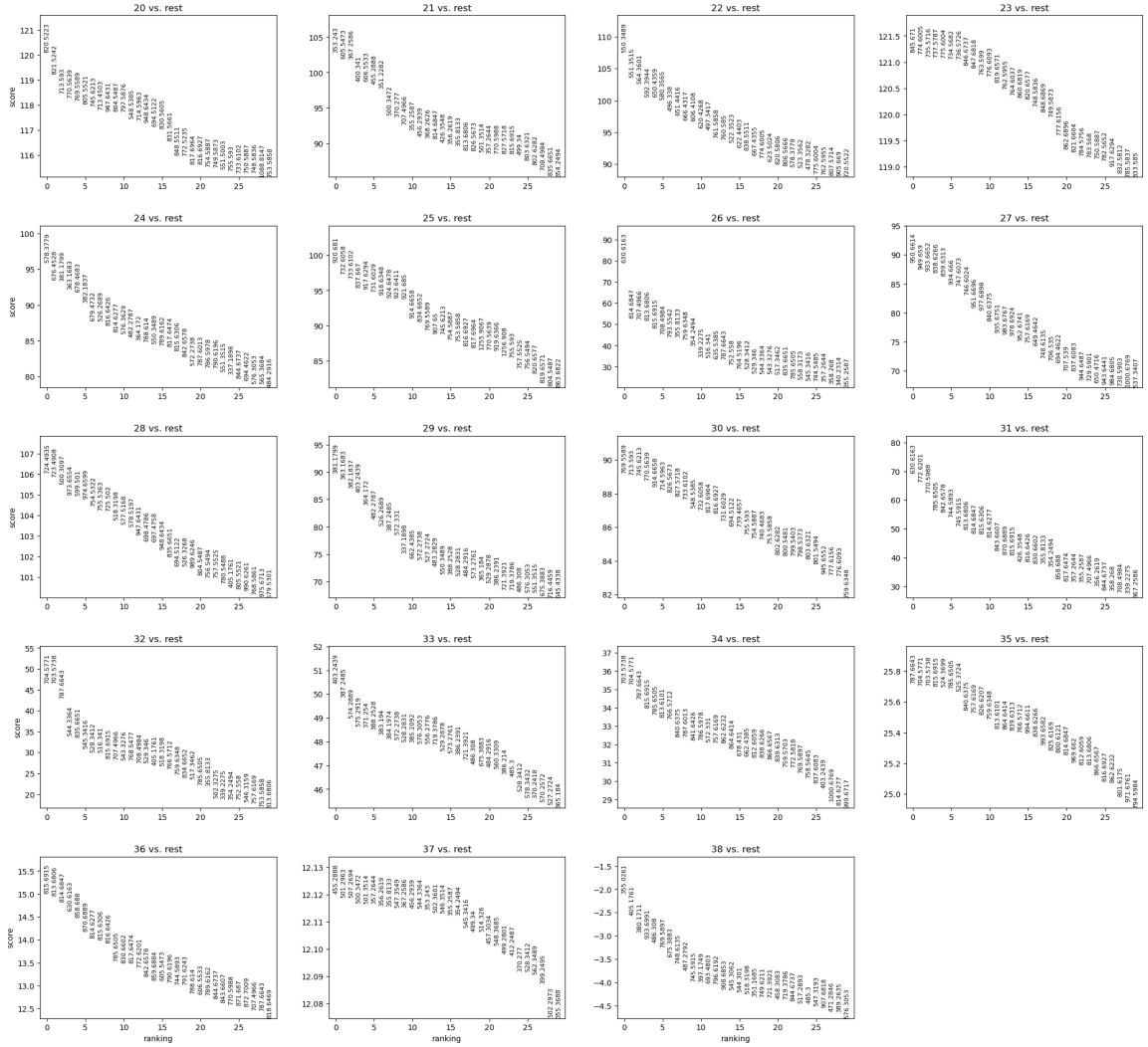


Figure 3-9 Top contributing analyte features for each cluster (clusters 20-38) were identified based on their corresponding scores, highlighting the analyte signals driving regional distinctions across the tissue and samples.

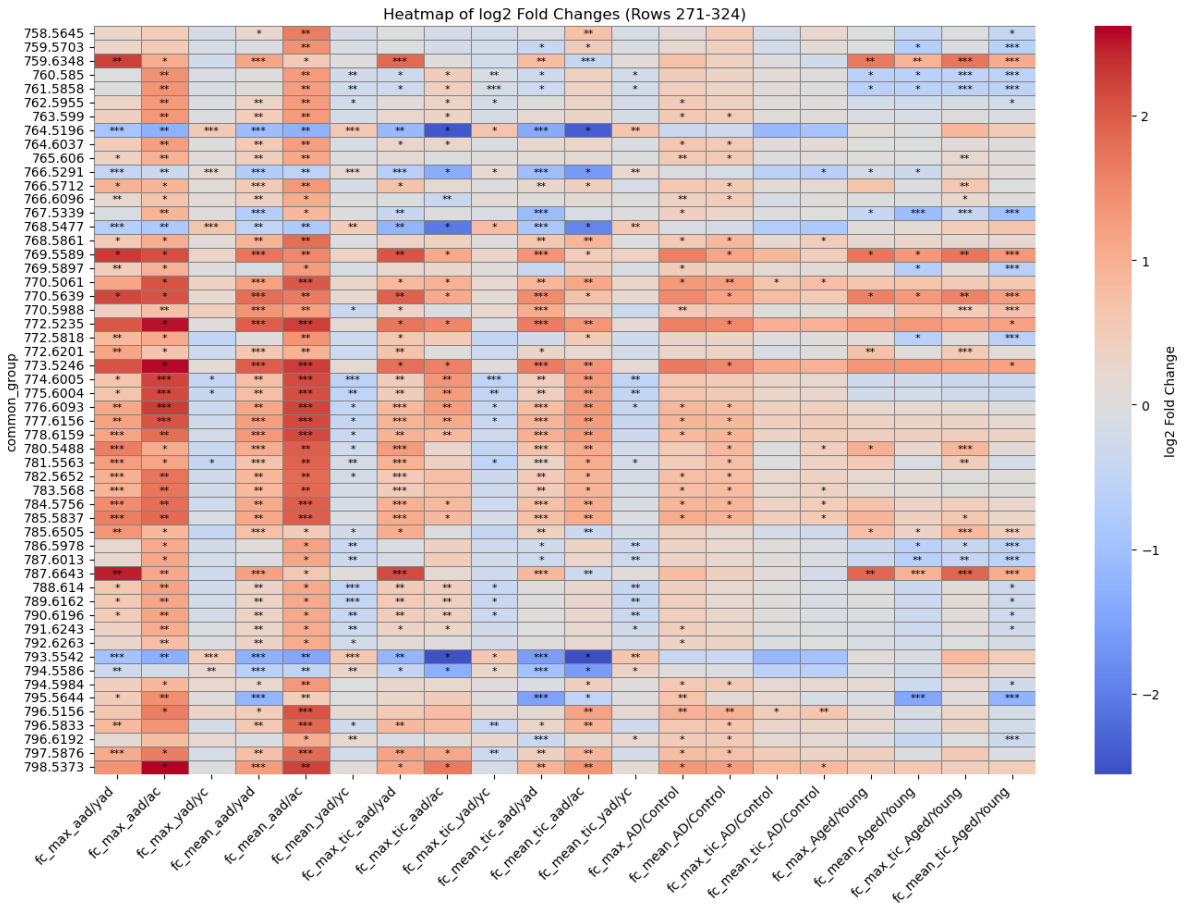


Figure 3-10 Heatmap of common analyte features (758.56–798.54) showing $\log_2$ fold changes across group comparisons. Statistical significance is indicated as $p < 0.05$ (*), $p < 0.01$ (**), and $p < 0.001$ (***)

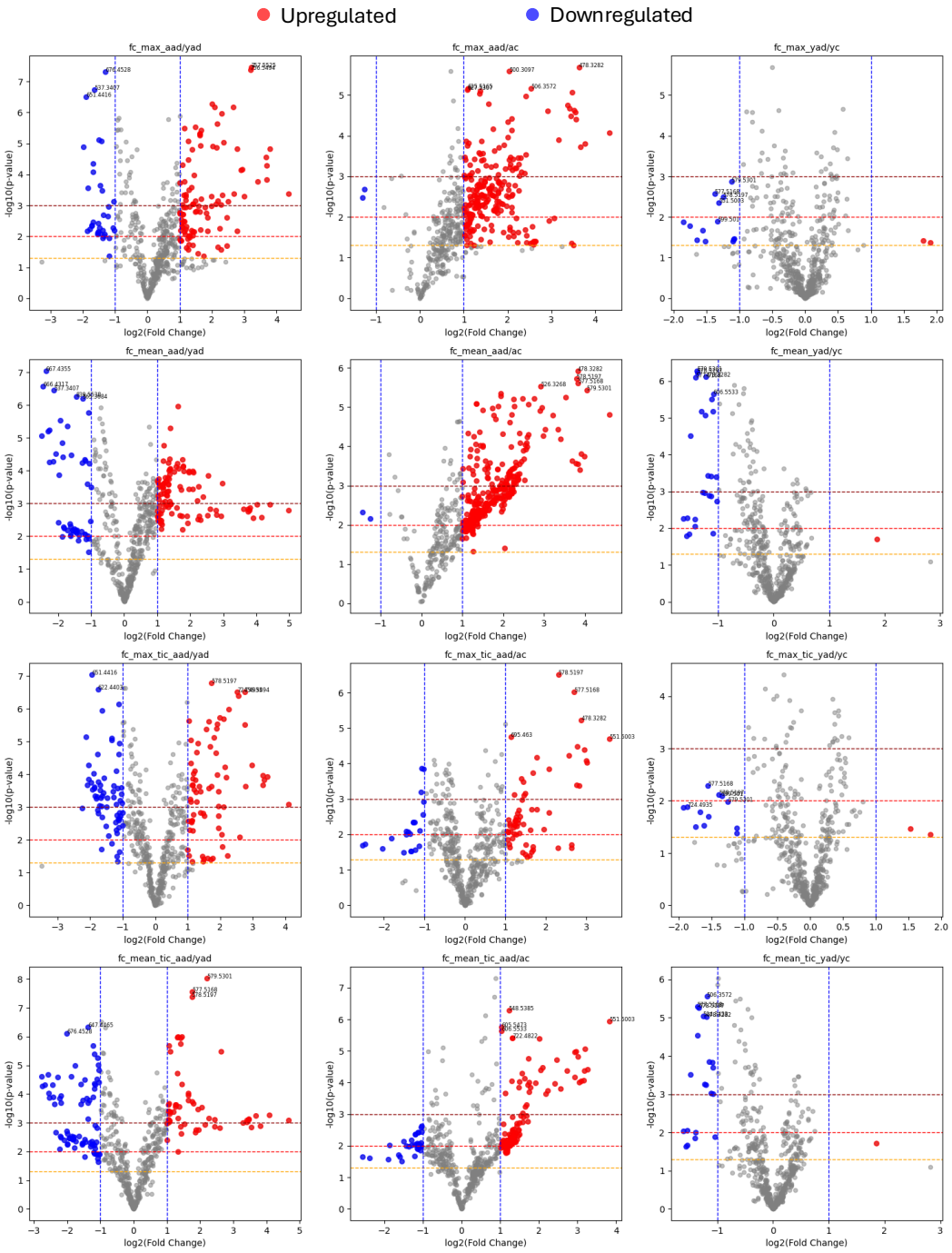


Figure 3-11 Volcano plots for each group comparison (comparisons 1-12), showing $\log_2$ fold changes versus – $\log_{10}$ p-values. Significantly upregulated and downregulated analytes are highlighted, with the top five most significant features labeled

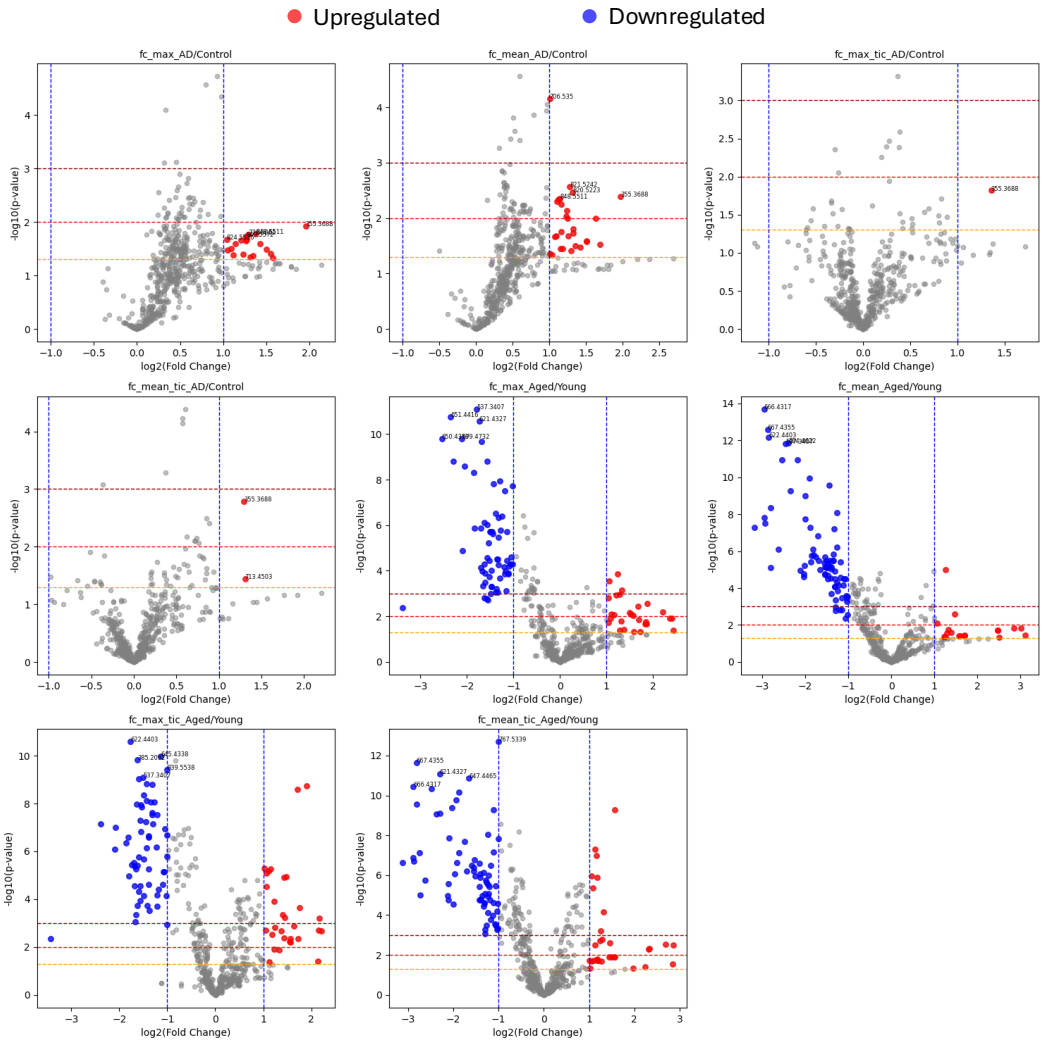


Figure 3-12 Volcano plots for each group comparison (comparisons 13-20), showing $\log_2$ fold changes versus – $\log_{10}$ p-values. Significantly upregulated and downregulated analytes are highlighted, with the top five most significant features labeled

These findings underscore the importance of evaluating analyte significance in the context of tissue specificity and biological meaning. In addition, false discovery rate (FDR) correction can be incorporated into the computational algorithm to control for multiple hypothesis testing. This approach reduces the likelihood of false-positive findings, improves the reliability of statistically significant results, and ensures that observed patterns in fold changes and p-values reflect robust and reproducible signals.

3.4 Conclusion

This study aimed to develop a computational framework for bulk analysis of MALDI-MSI data using machine learning to enable hypothesis generation and testing in AD and aging samples. The framework supported preprocessing, dimensionality reduction, clustering, and statistical analysis of spatial lipid signals across brain regions. Significant analytes were identified, revealing tissue-specific patterns – for example, analyte 921.69 showed decreased intensity in AD samples and was tentatively assigned to PI O-41:1, pending MS/MS confirmation. Clustering and dimensionality reduction highlighted region-specific lipid variations and candidate features associated with aging and disease, while false discovery rate (FDR) correction can be applied to ensure more robust statistical findings. Overall, this framework demonstrates a versatile approach for integrating spatial lipidomics with computational modeling to generate biologically meaningful hypotheses and guide mechanistic studies and biomarker development across a wide range of diseases and conditions.

REFERENCES

1. Cerruti CD, Benabdellah F, Laprévote O, Touboul D, Brunelle A. MALDI imaging and structural analysis of rat brain lipid negative ions with 9-aminoacridine matrix. Analytical chemistry. 2012 Mar 6;84(5):2164-71.
2. Miyake Y, Kusaka S, Murata I, Toyoda M. Matrix-Assisted Laser Desorption/Ionization (MALDI) Mass Spectrometry Imaging of L-4-Phenylalanineboronic Acid (BPA) in a Brain Tumor Model Rat for Boron Neutron Capture Therapy (BNCT). Mass Spectrometry. 2022 Dec 20;11(1):A0105-.
3. Ntshangase S, Mdanda S, Naicker T, Kruger HG, Baijnath S, Govender T. Spatial distribution of elvitegravir and tenofovir in rat brain tissue: Application of matrix-assisted laser desorption/ionization mass spectrometry imaging and liquid chromatography/tandem mass spectrometry. Rapid Communications in Mass Spectrometry. 2019 Nov 15;33(21):1643-51.
4. Stoeckli M, Staab D, Staufenbiel M, Wiederhold KH, Signor L. Molecular imaging of amyloid β peptides in mouse brain sections using mass spectrometry. Analytical biochemistry. 2002 Dec 1;311(1):33-9. A
5. Steven RT, Bunch J. Repeat MALDI MS imaging of a single tissue section using multiple matrices and tissue washes. Analytical and bioanalytical chemistry. 2013 May;405:4719-28.

6. Heijs B, Tolner EA, Bovée JV, van den Maagdenberg AM, McDonnell LA. Brain region-specific dynamics of on-tissue protein digestion using MALDI mass spectrometry imaging. Journal of proteome research. 2015 Dec 4;14(12):5348-54.
7. Strnad Š, Pražienková V, Holubová M, Sýkora D, Cvačka J, Maletínská L, Železná B, Kuneš J, Vrkoslav V. Mass spectrometry imaging of free-floating brain sections detects pathological lipid distribution in a mouse model of Alzheimer's-like pathology. Analyst. 2020;145(13):4595-605.
8. Nativio R, Lan Y, Donahue G, Sidoli S, Berson A, Srinivasan AR, Shcherbakova O, Amlie-Wolf A, Nie J, Cui X, He C. An integrated multi-omics approach identifies epigenetic alterations associated with Alzheimer's disease. Nature genetics. 2020 Oct;52(10):1024-35.
9. Liu RM. Aging, cellular senescence, and Alzheimer's disease. International Journal of Molecular Sciences. 2022 Feb 11;23(4):1989.
10. 2023 Alzheimer's disease facts and figures. Alzheimers Dement [Internet]. 2023 [cited 2023 Apr 21];19(4):1598–695. Available from: https://pubmed.ncbi.nlm.nih.gov/36918389/
11. Sastry PS. Lipids of nervous tissue: composition and metabolism. Progress in lipid research. 1985 Jan 1;24(2):69-176.
12. Abad-Rodriguez J, Ledesma MD, Craessaerts K, Perga S, Medina M, Delacourte A, Dingwall C, De Strooper B, Dotti CG. Neuronal membrane cholesterol loss

enhances amyloid peptide generation. The Journal of cell biology. 2004 Dec 6;167(5):953-60.

13. Frank C, Rufini S, Tancredi V, Forcina R, Grossi D, D'arcangelo G. Cholesterol depletion inhibits synaptic transmission and synaptic plasticity in rat hippocampus. Experimental neurology. 2008 Aug 1;212(2):407-14.
14. Elobeid A, Libard S, Leino M, Popova SN, Alafuzoff I. Altered proteins in the aging brain. Journal of Neuropathology & Experimental Neurology. 2016 Apr 1;75(4):316-25.
15. Abad-Rodriguez J, Ledesma MD, Craessaerts K, Perga S, Medina M, Delacourte A, Dingwall C, De Strooper B, Dotti CG. Neuronal membrane cholesterol loss enhances amyloid peptide generation. The Journal of cell biology. 2004 Dec 6;167(5):953-60.
16. Frank C, Rufini S, Tancredi V, Forcina R, Grossi D, D'arcangelo G. Cholesterol depletion inhibits synaptic transmission and synaptic plasticity in rat hippocampus. Experimental neurology. 2008 Aug 1;212(2):407-14.
17. Dufresne M, Migas LG, Djambazova K, Colley ME, Van de Plas R, Spraggins JM. Aminated cinnamic acid analogs as dual polarity matrices for high spatial resolution MALDI imaging mass spectrometry. Analytica Chimica Acta. 2025 Jul 15:344423.
18. Savitzky A, Golay MJ. Smoothing and differentiation of data by simplified least squares procedures. Analytical chemistry. 1964 Jul 1;36(8):1627-39.

19. Eilers PH, Boelens HF. Baseline correction with asymmetric least squares smoothing. Leiden University Medical Centre Report. 2005 Oct 21;1(1):5.
20. Castro DC, Smith KW, Norsworthy MD, Rubakhin SS, Weisbrod CR, Hendrickson CL, Sweedler JV. Single-cell and subcellular analysis using ultrahigh resolution 21 T MALDI FTICR mass spectrometry. Analytical chemistry. 2023 Apr 18;95(17):6980-8.
21. Blank M, Enzlein T, Hopf C. LPS-induced lipid alterations in microglia revealed by MALDI mass spectrometry-based cell fingerprinting in neuroinflammation studies. Scientific Reports. 2022 Feb 21;12(1):2908.
22. Hall Z, Ament Z, Wilson CH, Burkhart DL, Ashmore T, Koulman A, Littlewood T, Evan GI, Griffin JL. Myc expression drives aberrant lipid metabolism in lung cancer. Cancer research. 2016 Aug 15;76(16):4608-18.
23. Otsuka Y, Okada M, Hashidate-Yoshida T, Nagata K, Yamada M, Goto M, Sun M, Shindou H, Toyoda M. Improved ion detection sensitivity in mass spectrometry imaging using tapping-mode scanning probe electrospray ionization to visualize localized lipids in mouse testes. Analytical and Bioanalytical Chemistry. 2025 Jan;417(2):275-86.

CHAPTER 4:

SPATIAL TRANSCRIPTOMICS IN AGING AND ALZHEIMER'S DISEASE

4.1 Introduction

Recent advances in spatial transcriptomics have made it possible to measure gene expression while preserving the native architecture of tissue. Unlike bulk RNA sequencing, which averages expression signals across heterogeneous cell populations, spatial transcriptomics maintains spatial resolution and enables the detection of localized gene expression changes that may be critical in disease contexts [1]. This approach provides a powerful framework for studying complex disorders such as Alzheimer's disease (AD), where pathology is region-specific and cell-type dependent. When combined with computational techniques, including dimensionality reduction and clustering algorithms, spatial transcriptomics allows for systematic exploration of high-dimensional data and identification of underlying molecular structures [2]. Applying these methods to aging and AD models offers an opportunity to uncover spatially restricted gene expression patterns that contribute to disease progression.

AD is a progressive neurodegenerative disorder marked by cognitive decline, memory impairment, and gradual loss of the ability to perform daily activities. It is the sixth leading cause of death in the United States and the fifth among individuals over 65 years of age [3]. Despite extensive research efforts, the molecular mechanisms underlying AD remain incompletely understood, and no curative treatments are currently available [4-6]. Aging is the strongest known risk factor for AD, and age-related molecular changes in the brain further contribute to the complexity of disease development and progression [7, 8]. Understanding how aging interacts with AD-related processes at the molecular level is

therefore essential for identifying early disease mechanisms and potential therapeutic targets.

The primary objective of this aim is to propose a Python-based pipeline to perform a comprehensive spatial transcriptomic analysis on biological tissues using 10x Genomics Visium slides. The case study tissues are aging and the AD mouse brain tissues. Specifically, this work seeks to:

1. Test the program pipeline to determine whether it identifies spatial domains within brain tissue that are altered in aging and AD.
2. Detect differentially expressed genes (DEGs) between AD and age-matched control mice, as well as those associated with aging in control mice.
3. Integrate spatial transcriptomic data with histological features to generate detailed spatial gene expression maps.

These goals are designed to provide a multidimensional understanding of how gene expression changes are organized across brain regions in both aging and AD.

AD pathology is not uniformly distributed throughout the brain; instead, it tends to affect specific regions, including the cortex and hippocampus, which are particularly vulnerable to neurodegeneration [9-12]. Spatial transcriptomics preserves tissue structure and enables the examination of gene expression within its biological context. This makes it possible to determine how specific genes are differentially regulated in distinct anatomical or pathological regions – information that is essential for understanding disease mechanisms.

To analyze these complex datasets, unsupervised learning approaches such as Leiden, Louvain, and K-means clustering are particularly valuable. These algorithms can reveal hidden spatial domains and molecular patterns that may not be apparent through traditional analyses. By uncovering previously unrecognized structures associated with AD or aging, these computational methods help characterize regional vulnerability and molecular heterogeneity in the brain [1, 2]. Integrating spatial transcriptomics with these analytical strategies provides a systematic framework for identifying localized molecular alterations that may serve as early biomarkers or potential therapeutic targets.

4.2 Methodology

4.2.1 Tissue Preparation

Brains from young (10–12 weeks) and aged (52–54 weeks), female 3xTg-AD and control mice ($n = 4$ per group, Chapter 3) were harvested and rapidly flash-frozen in liquid nitrogen to preserve RNA integrity. The brains were cryosectioned at 10 μm thickness using a Leica Cryostat with one slice used for Chapter 3 mass spectrometry and the next serial slice in this study. Sections encompassing the cortex and hippocampus – approximately 2.24 ± 0.21 mm from the bregma – were mounted onto 10x Genomics Visium Spatial Gene Expression slides to capture spatially resolved gene expression patterns. To determine optimal tissue permeabilization conditions, seven serial sections from a single animal were mounted on 10x Genomics Tissue Optimization slides. This tissue optimization step enabled precise calibration of permeabilization time for efficient mRNA release while preserving tissue morphology, ensuring that downstream spatial

transcriptomic profiling yielded high-quality and reproducible data. After tissue mounting, the slides were placed in -80 C freezer.

4.2.2 Visium Spatial Tissue Optimization Workflow

In this process, seven serial sections from a representative tissue were mounted onto 10x Genomics Tissue Optimization (TO) slides, which contained arrays of spatially barcoded capture spots. The tissue sections were then treated with a permeabilization enzyme, typically a mild protease, for varying durations. This enzyme partially digested the extracellular matrix and cell membranes, allowing mRNA molecules to diffuse out of the tissue and bind to the spatially barcoded capture oligonucleotides on the slide. After permeabilization, reverse transcription was performed to convert the captured mRNA into spatially indexed cDNA. The cDNA was subsequently fluorescently labeled, enabling visualization of RNA capture efficiency across different permeabilization conditions. By analyzing the fluorescence intensity while monitoring tissue integrity, the optimal permeabilization time was determined to maximize RNA yield without compromising tissue morphology. This optimized condition was then applied to subsequent Visium experiments, ensuring high-quality spatial transcriptomic profiling, accurate gene expression quantification, and correct mapping of spatial gene expression patterns within the tissue.

4.2.3 Spatial Transcriptomics

10x Genomics Visium Spatial Gene Expression slides contain an array of spatially barcoded capture spots that enabled the assignment of transcriptomic information to precise tissue locations. The tissue sections were briefly warmed to ensure proper adhesion to the slide surface and were subsequently fixed to preserve tissue morphology and RNA integrity. The sections were then stained using hematoxylin and eosin to visualize tissue architecture and imaged under bright-field microscopy, generating high-resolution images to guide downstream analyses, at Advanced Microscopy and Imaging Core (AMIC) of the University of Tennessee, Knoxville. Following imaging, the slides were processed for spatial transcriptomic library preparation according to the 10x Genomics Visium protocol. The resulting cDNA libraries were sent to the Genomics Core at the University of Tennessee, Knoxville, for sequencing on an Illumina platform (NovaSeq). Sequencing outputs were delivered as FASTQ files, containing raw reads indexed by spatial barcodes and ready for computational processing and spatial gene expression analysis.

4.2.4 Data Analysis

4.2.4.1 Preprocessing

Brightfield H&E-stained tissue images were first examined in Loupe Browser for fiducial frame alignment and determination of the region of interest (ROI). Sequenced data were obtained as FASTQ files from the Genomic Core and were processed using 10x Genomics Space Ranger to align reads, assign them to spatial barcodes, and generate

spot-level count matrices. The data were then imported into Python for downstream analysis. In this study, YC, AC, YAD, and AAD denote Young Control, Aged Control, Young AD, and Aged AD, respectively. The design program in Python enabled investigation of aging effects (AC vs YC) and disease effects within age strata (YAD vs YC; AAD vs AC), thereby preventing confounding between aging and disease. All samples were loaded as individual AnnData objects and concatenated using an inner gene intersection to ensure a consistent feature space across samples. Sample-level metadata (Group, Condition, Age, and Sample ID) were appended to each observation (spot).

4.2.4.1.1 Quality Control

Quality control was conducted at both gene and spot levels to ensure reliable downstream analysis. Genes with zero counts across all spots were removed. Spots expressing fewer than 200 genes and genes detected in fewer than 10 spots were excluded to eliminate low-quality or weakly expressed features. Spots with more than 20% mitochondrial transcript content were filtered out, as high mitochondrial expression often reflects stressed or damaged tissue. Mitochondrial genes were identified by the “mt-” prefix, and ribosomal genes by Rpl/Rps naming conventions. Standard metrics – including total counts, number of detected genes, and percentage of mitochondrial reads – were computed prior to filtering.

4.2.4.1.2 Normalization and Data Representation

To account for sequencing depth differences, counts were normalized to 10,000 counts per spot and log-transformed (log1p) to stabilize variance. Normalized data were stored

separately from raw counts to preserve count integrity for statistical testing while enabling exploratory analysis and visualization.

4.2.4.1.3 Feature Selection

Highly variable genes (HVGs) were identified using the Seurat v3 method with sample-aware variance modeling to account for batch effects. The top 3,000 HVGs were selected based on raw count variability.

4.2.4.1.4 PCA-Loading-Based HVG Refinement

Principal component analysis (PCA) was performed on scaled HVGs. The optimal number of components was determined using an elbow detection (Kneedle) approach. Genes contributing strongly to the selected components were identified using the L2 norm of their loadings, and those exceeding mean + 1 standard deviation were retained. This two-step strategy reduces noise and focuses analysis on genes driving major transcriptional variation (Figure 4-1).

4.2.4.2 Downstream Analysis

4.2.4.2.1 Dimensionality Reduction and Clustering

Using the selected principal components, a k-nearest neighbor graph was constructed. UMAP was applied for non-linear embedding, and Leiden clustering (resolution = 0.5) identified transcriptionally coherent spatial clusters. UMAP visualization enabled assessment of group separation and cluster structure.

4.2.4.2.2 Pseudobulk Construction

To avoid pseudoreplication, spatial spots were aggregated at the biological replicate

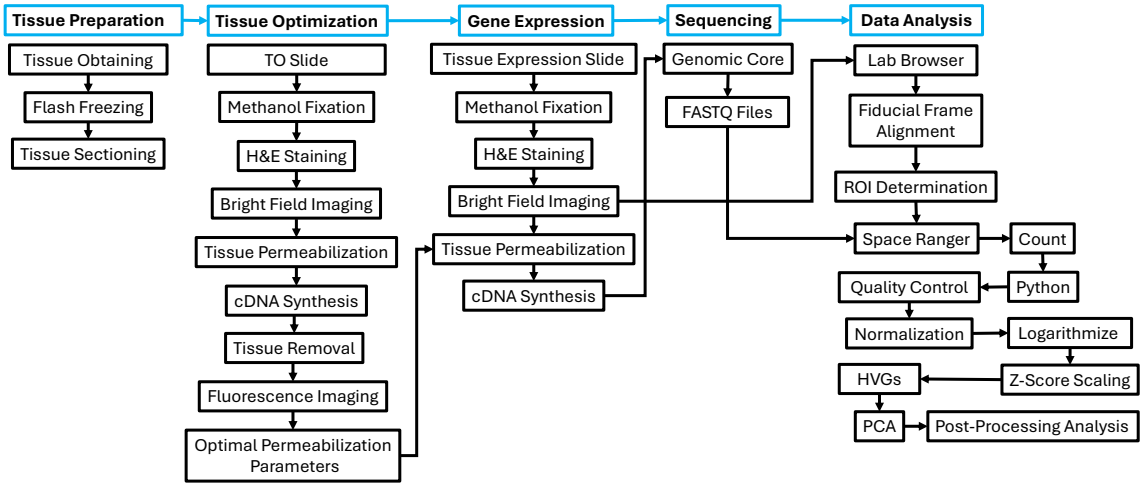


Figure 4-1 Flowchart of Visium tissue optimization, spatial gene expression analysis, and data preprocessing pipeline. TO: tissue optimization

level. Raw counts were summed per sample, preserving metadata and recording the number of contributing spots. This yielded 16 pseudobulk samples, maintaining true biological replication. CPM-normalized log-transformed data were additionally generated for visualization and complementary testing.

4.2.4.2.3 Differential Expression Analysis

(a) Biologically Meaningful Contrasts

Differential expression analysis focused on interpretable comparisons:

- Young AD vs Young Control (disease effect in young mice)
- Aged Control vs Young Control (aging effect)
- Aged AD vs Aged Control (disease effect in aged mice)

This structure prevents confounding between aging and disease effects.

(b) Statistical Framework

Welch's t-test was applied to CPM-log-transformed pseudobulk data, as it does not assume equal variance between groups. P-values were adjusted using the Benjamini–Hochberg false discovery rate (FDR) method.

4.2.4.2.4 Significance Criteria and Gene Selection

Genes were considered significantly differentially expressed if FDR < 0.05 and $|\log_2$ fold change$| > 0.5$. For each contrast, the top 20 upregulated and top 20 downregulated genes were reported. If no genes met FDR thresholds, genes with the largest fold changes were described.

4.2.4.2.5 Spatial Variability Analysis

To assess spatial organization of AD-associated genes, AD samples were subset and a spatial neighbor graph (k = 6) was constructed. Moran's I was computed with 999 permutations. Genes with Moran's $I > 0.1$ and permutation p-value < 0.05 were considered spatially variable. This analysis identifies genes that are both differentially expressed and spatially patterned within diseased tissue.

4.3 Results and Discussion

After concatenation, the dataset comprised 43,020 spatial spots and 32,285 genes. Quality control filtering removed 224 spots expressing fewer than 200 genes and excluded 4,346 genes detected in fewer than 10 spots. Once cells whose mitochondrial gene expression accounted for 20% or more of their total counts were filtered out, leaving only higher-quality cells in the AnnData object were left, resulting in a refined dataset of 25,351 spots and 20,562 genes (Figure 4-2). Counts were then normalized per spot to account for

sequencing depth differences. From this dataset, 3,000 highly variable genes (HVGs) were identified for downstream analysis. Principal component analysis (PCA) was performed using 50 components, and objective elbow detection indicated that the first 8 principal components captured the most meaningful transcriptional variation (Figure 4-3). To further refine the feature space, highly variable genes were filtered based on their contributions to the top eight principal components using an L2 norm of PCA loadings. Genes with loading scores exceeding the mean plus one standard deviation (0.073) were retained as PCA-informative (Figure 4-4). Based on this criterion, 595 genes were selected from the original 3,000 HVGs, restricting the feature set to those driving the dominant transcriptional structure and improving signal specificity.

A k-nearest neighbor graph was constructed using 15 neighbors and the top eight principal components, followed by UMAP embedding for visualization of the global transcriptional landscape. When colored by experimental group, the UMAP projection revealed the overall relationships among samples, indicating partial segregation driven by

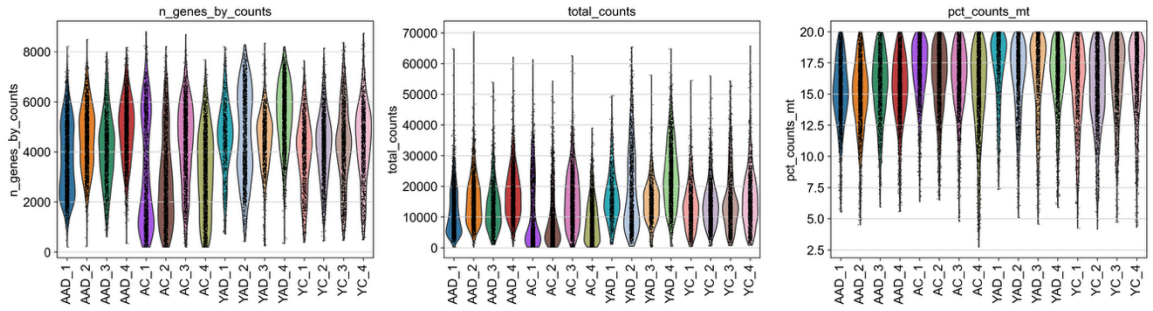


Figure 4-2 Violin plots displaying the number of genes detected per cell (*n_genes_by_counts*; left), total counts per cell (middle), and percentage of mitochondrial counts (*pct_counts_mt*; right) across samples following preprocessing and filtering.

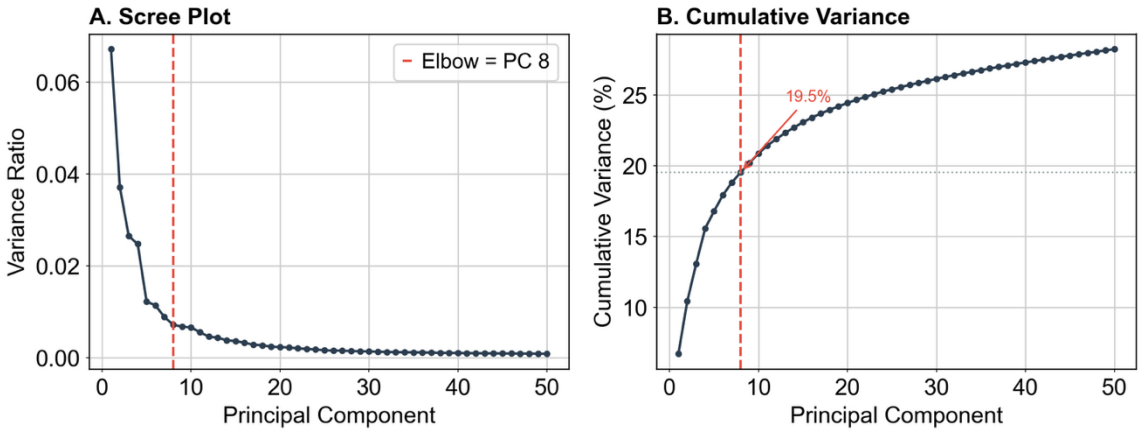


Figure 4-3 Objective selection of principal components based on explained variance

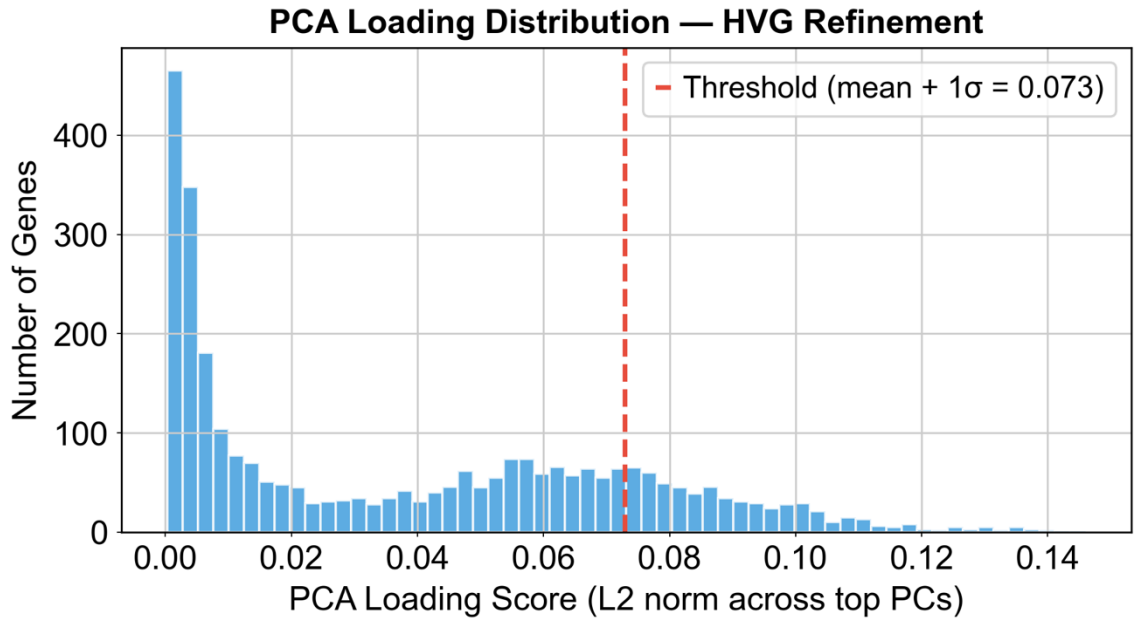


Figure 4-4 PCA-Loading-Based Refinement of Highly Variable Genes

biological condition. Leiden clustering (resolution = 0.5) was not able to completely identify transcriptionally coherent subpopulations (Figure 4-5), reflecting non-discrete molecular states within the dataset. These results suggest overlapping transcriptional programs and gradual transitions rather than sharply separated cellular states.

After implementing the pseudobulk strategy, raw counts were aggregated across all spots within each biological sample to generate sample-level expression profiles. Differential expression was then performed on log-CPM–normalized data using Welch's t-test, with Benjamini–Hochberg correction applied to control the false discovery rate. Group-wise comparisons were conducted using biologically relevant contrasts, with YC serving as the baseline, YAD versus YC to assess the young disease effect, AC versus YC to evaluate the aging effect, and AAD versus AC to examine the disease effect in aged animals. Then, the significant upregulated genes (padj < 0.05, log2FC > 0.5) and downregulated ones (padj < 0.05, log2FC < -0.5) for each comparison were determined.

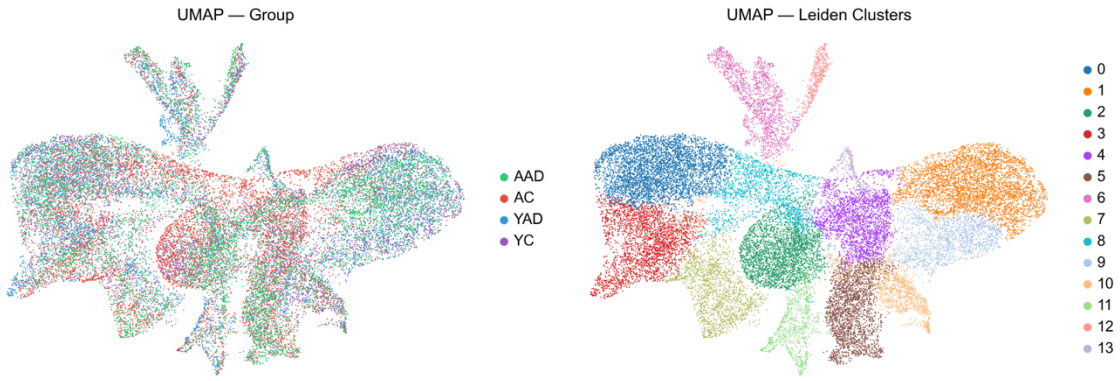


Figure 4-5 UMAP visualization of transcriptional structure (left) clustered by group and (right) Leiden clustering. AAD: aged AD, AC: aged control, YAD: young AD, YC: young control

Figure 4-6 illustrates the top 20 upregulated genes identified in the young-group disease comparison (YC vs. YAD), visualized using a dot plot, volcano plot, and heatmap. Although Hbb-bs showed higher expression in the young AD samples compared to young controls and was detected in nearly 100% of cells (Figure 4-6a), its adjusted *p*-value exceeded 0.05, indicating that the difference was not statistically significant. Overall, none of the upregulated genes in this comparison reached statistical significance after multiple testing correction. A similar pattern was observed for the downregulated genes (Figure 4-7). While several genes met the fold-change threshold, their adjusted *p*-values remained above the significance cutoff. One possible explanation for these results is technical variability across samples. In particular, partial loss of cortical regions during cryosectioning or suboptimal placement on the spatially barcoded capture area may have reduced tissue coverage in some sections, thereby limiting statistical power in this comparison.

The comparison between aged AD samples and age-matched controls revealed that *Thy1, Cldn11,* and *Bcas1* were significantly upregulated in the aged AD group (Figure 4-8), whereas *Ccl27a* was significantly downregulated (Figure 4-9). These statistically significant differences indicate AD-associated molecular alterations in the aged brain. In addition to these genes, several others exhibited upregulated or downregulated expression patterns in AD samples relative to controls; however, these changes did not reach statistical significance. As discussed previously, the lack of statistical significance may be attributed to factors related to sample preparation and experimental variability, as well

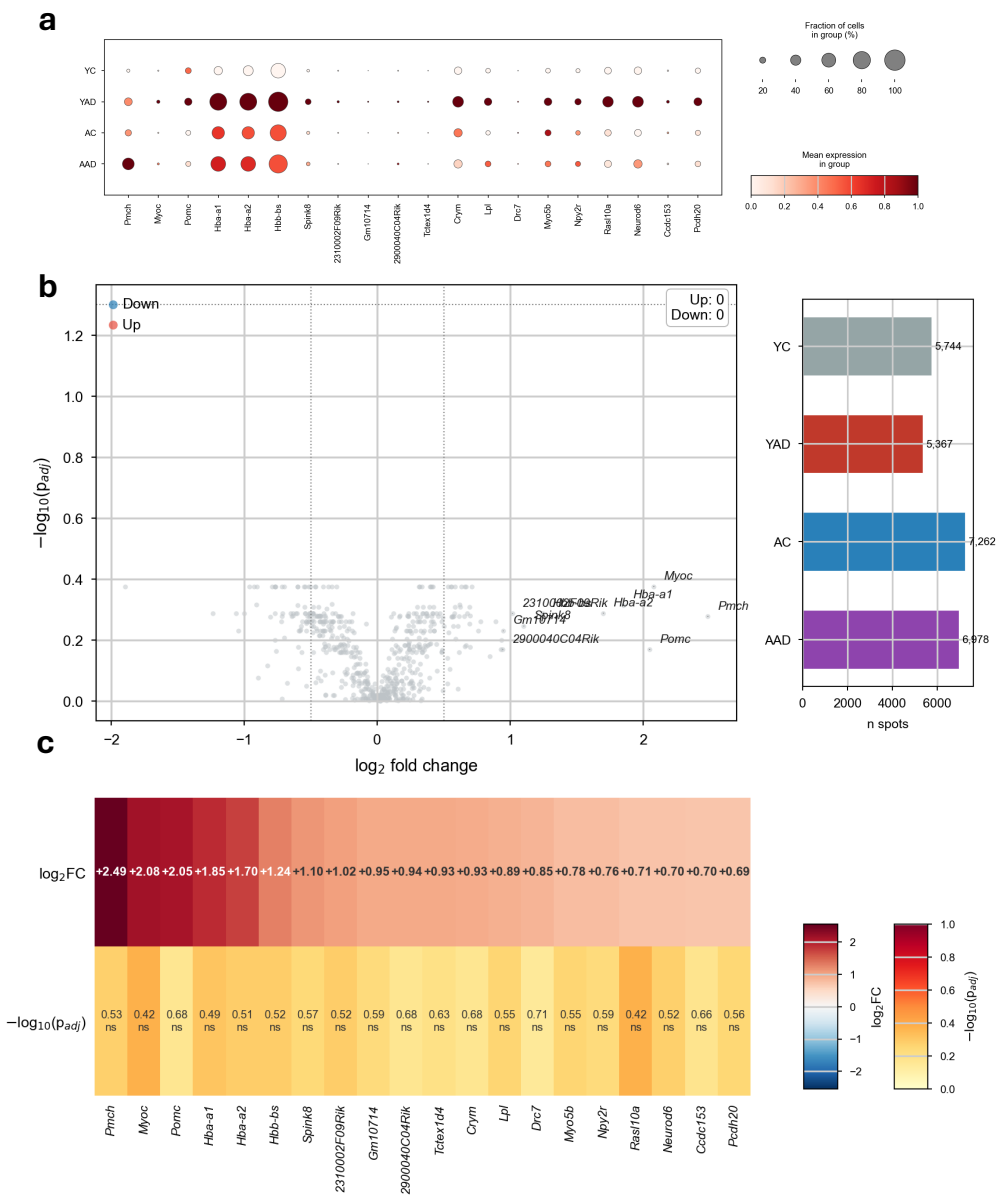


Figure 4-6 Upregulated genes in the young disease comparison (YAD vs. YC). (a) Dotplot illustrating the top upregulated genes in this comparison; (b) volcano plot highlighting differentially expressed genes meeting the significance criteria (adjusted p-value < 0.05 and $log_2FC > 0.5$); and (c) statistics heatmap summarizing the log_2 fold change and adjusted p-values for the top upregulated genes in the young disease comparison. Significance levels are indicated as follows: $p_{adj} < 0.05$ (*), < 0.01 (**), < 0.001 (***), and ns (not significant; $p_{adj} \geq 0.05$).

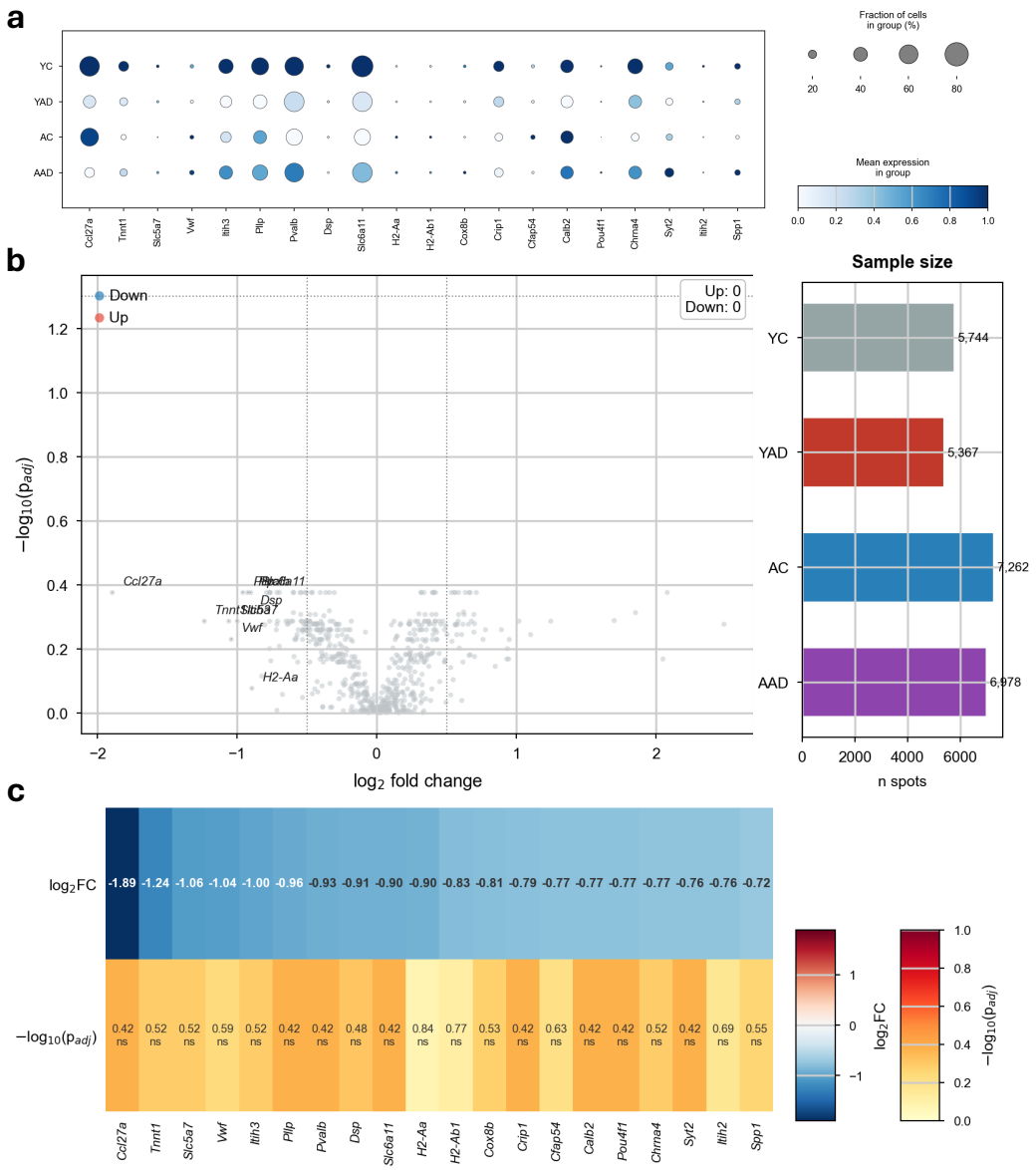


Figure 4-7 Downregulated genes in the young disease comparison (YAD vs. YC). (a) Dotplot illustrating the top downregulated genes in this comparison; (b) volcano plot highlighting differentially expressed genes meeting the significance criteria (adjusted p-value < 0.05 and $log_2FC < -0.5$); and (c) statistics heatmap summarizing the log_2 fold change and adjusted p-values for the top downregulated genes in the young disease comparison. Significance levels are indicated as follows: $p_{adj} < 0.05$ (*), < 0.01 (**), < 0.001 (***), and ns (not significant; $p_{adj} \geq 0.05$).

as preprocessing steps such as cell and gene filtering, normalization, and quality control thresholds.

Differential expression analysis comparing AC with young controls YC revealed pronounced aging-related transcriptional changes in the mouse brain. 25 genes were significantly upregulated (padj < 0.05, log2FC > 0.5), including *Ftl1, Hba-a1*, and *Hbb-bs* (Figure 4-10). *Hba-a1* and *Hbb-bs* encode hemoglobin subunits. Although hemoglobin is classically associated with red blood cells, its expression has also been reported in neurons and may be involved in maintaining oxygen homeostasis and responding to oxidative stress in the aging brain [13]. The increased expression of *Ftl1*, which encodes the ferritin light chain responsible for intracellular iron storage, is consistent with previous evidence showing that iron metabolism becomes altered during aging and may contribute to increased oxidative vulnerability. Aging has been shown to modify the balance and spatial distribution of ferritin subunit transcripts in the mouse hippocampus, indicating changes in iron handling in the aged brain [14]. In addition, more recent studies have reported elevated neuronal FTL1 levels in aged mice, which were associated with cognitive decline [15]. Together, these findings support the relevance of ferritin-related pathways in age-associated brain alterations.

In contrast, a substantially larger number of genes (n = 73) were significantly downregulated with aging (padj < 0.05, log2FC < −0.5) (Figure 4-11). Notably, *Chrna3* and *Chrna4*, which encode nicotinic acetylcholine receptor subunits that play key roles in cholinergic signaling [16, 17], were reduced in aged mice in our analysis. Previous

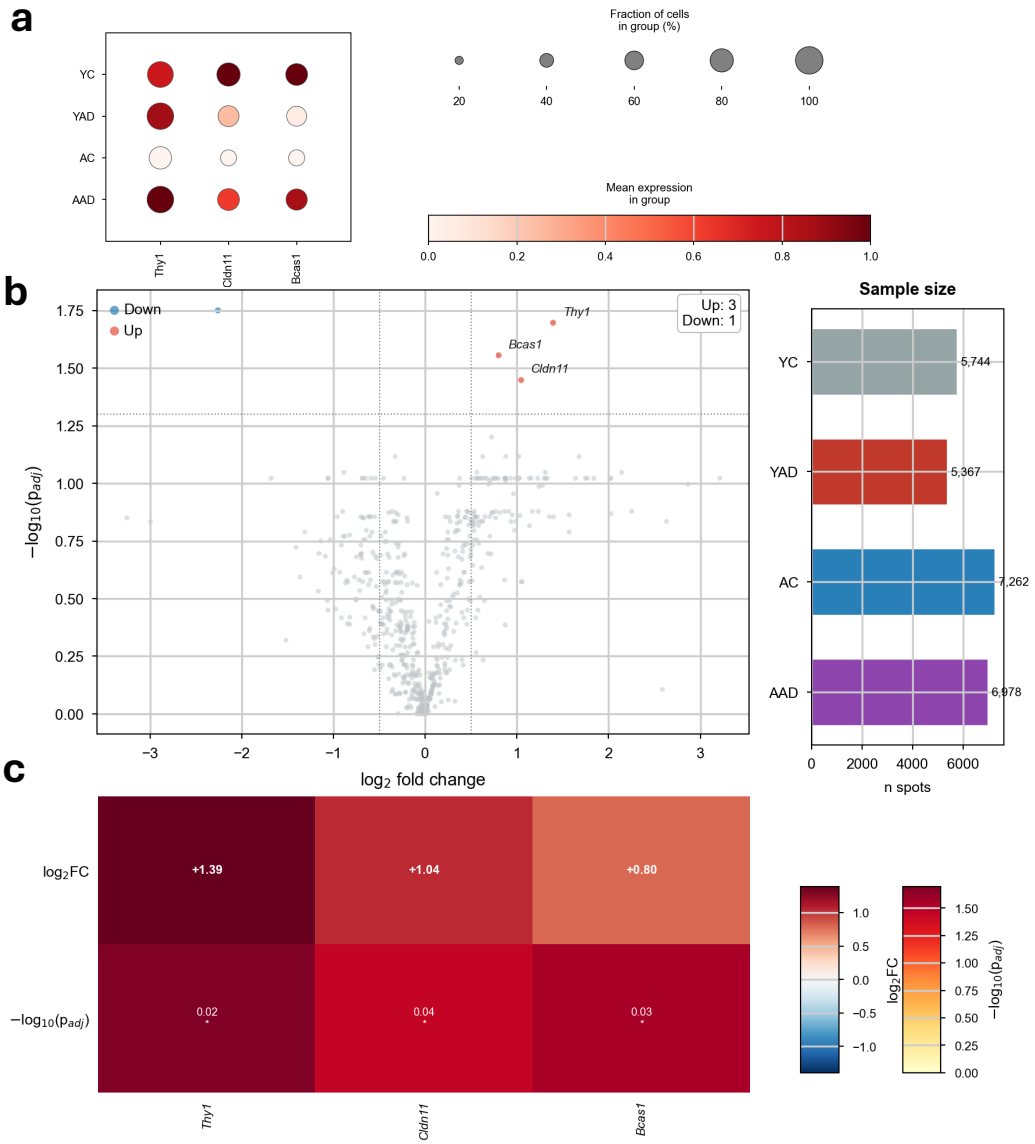


Figure 4-8 Upregulated genes in the aged disease comparison (AAD vs. AC). (a) Dotplot illustrating the top upregulated genes in this comparison; (b) volcano plot highlighting differentially expressed genes meeting the significance criteria (adjusted p-value < 0.05 and $log_2FC > 0.5$); and (c) statistics heatmap summarizing the log_2 fold change and adjusted p-values for the top upregulated genes in the aged disease comparison. Significance levels are indicated as follows: $p_{adj} < 0.05$ (*), < 0.01 (**), < 0.001 (***), and ns (not significant; $p_{adj} \geq 0.05$).

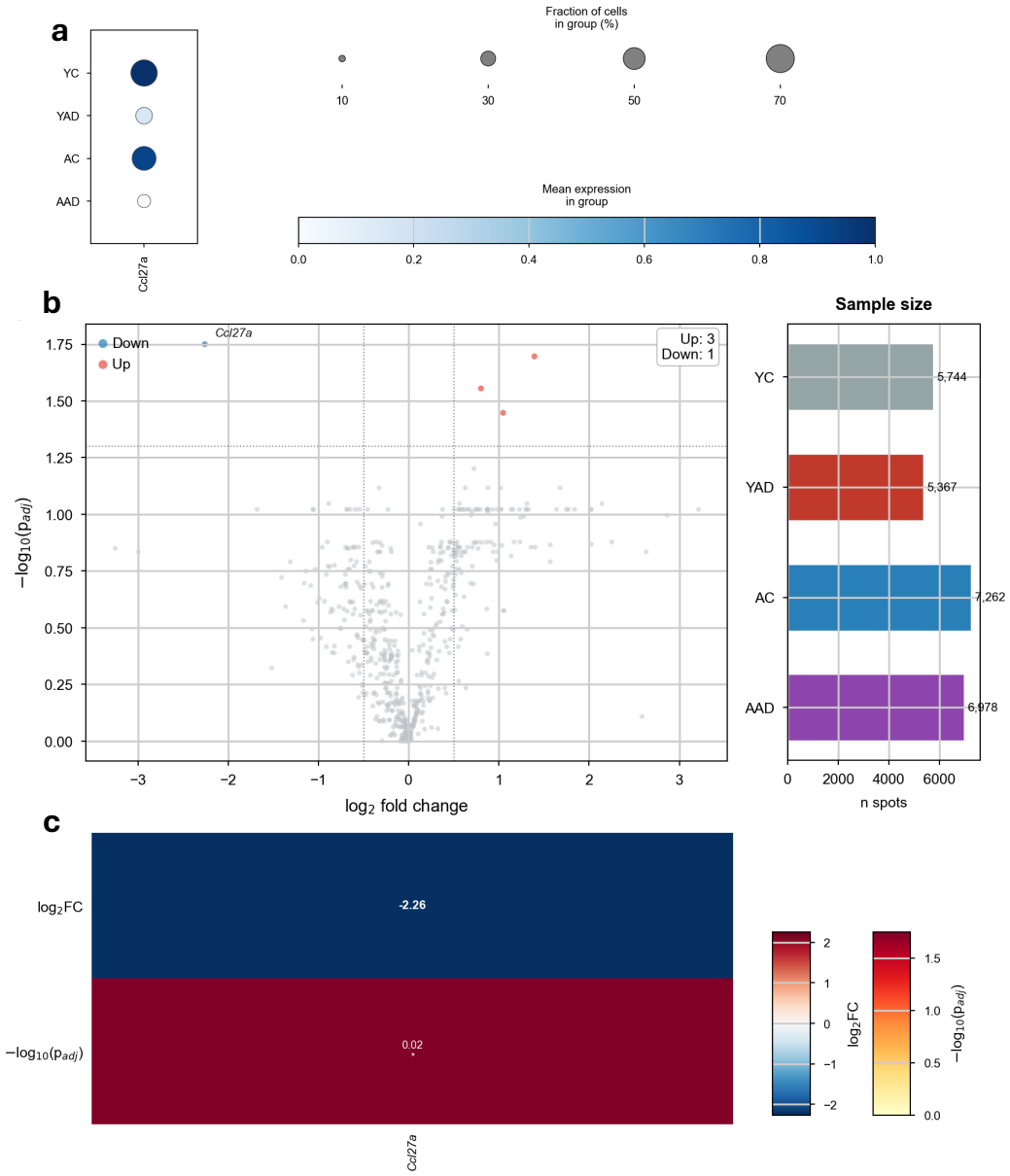


Figure 4-9 Downregulated genes in the aged disease comparison (AAD vs. AC). (a) Dotplot illustrating the top downregulated genes in this comparison; (b) volcano plot highlighting differentially expressed genes meeting the significance criteria (adjusted p-value < 0.05 and $log_2FC > 0.5$); and (c) statistics heatmap summarizing the log_2 fold change and adjusted p-values for the top downregulated genes in the aged disease comparison. Significance levels are indicated as follows: $p_{adj} < 0.05$ (*), < 0.01 (**), < 0.001 (***), and ns (not significant; $p_{adj} \geq 0.05$).

studies have reported that adult mice (approximately 47 weeks old) with early cognitive impairment exhibit significantly lower hippocampal *Chrna3* expression compared with younger animals, suggesting that reduced α3 nicotinic receptor levels may accompany aging-related cognitive changes. In this context, the downregulation of *Chrna3* observed in our study aligns with earlier evidence describing age-associated declines in cholinergic neurotransmission [16], further supporting the involvement of cholinergic pathway alterations in the aging brain. Additionally, decreased expression of transcription factors such as *Lef1*, key components of the Wnt signaling pathway, is consistent with studies showing that Wnt signaling activity diminishes with age and contributes to reduced neuronal plasticity and maintenance [18].

Figure 4-12 and Figure 4-13 demonstrate the most significant transcriptional changes observed across all comparisons (aging effect in control groups, AD effect in young groups, and AD effect in aged groups) in a clear and unified format. Figure 4-12 focuses on the top upregulated genes, where the grouped dotplot shows both the expression level and the proportion of expressing cells for each contrast, the heatmap summarizes the corresponding $\log_2$ fold change values across all groups with significance indicated by star annotations, and the accompanying table provides the full statistical details, including p-values, adjusted p-values, and $\log_2$ fold change values. Figure 4-13 follows the same structure for the top downregulated genes, allowing straightforward comparison of decreased expression patterns across conditions. The consistent presentation across both figures makes it easier to identify genes that are shared between comparisons as well as

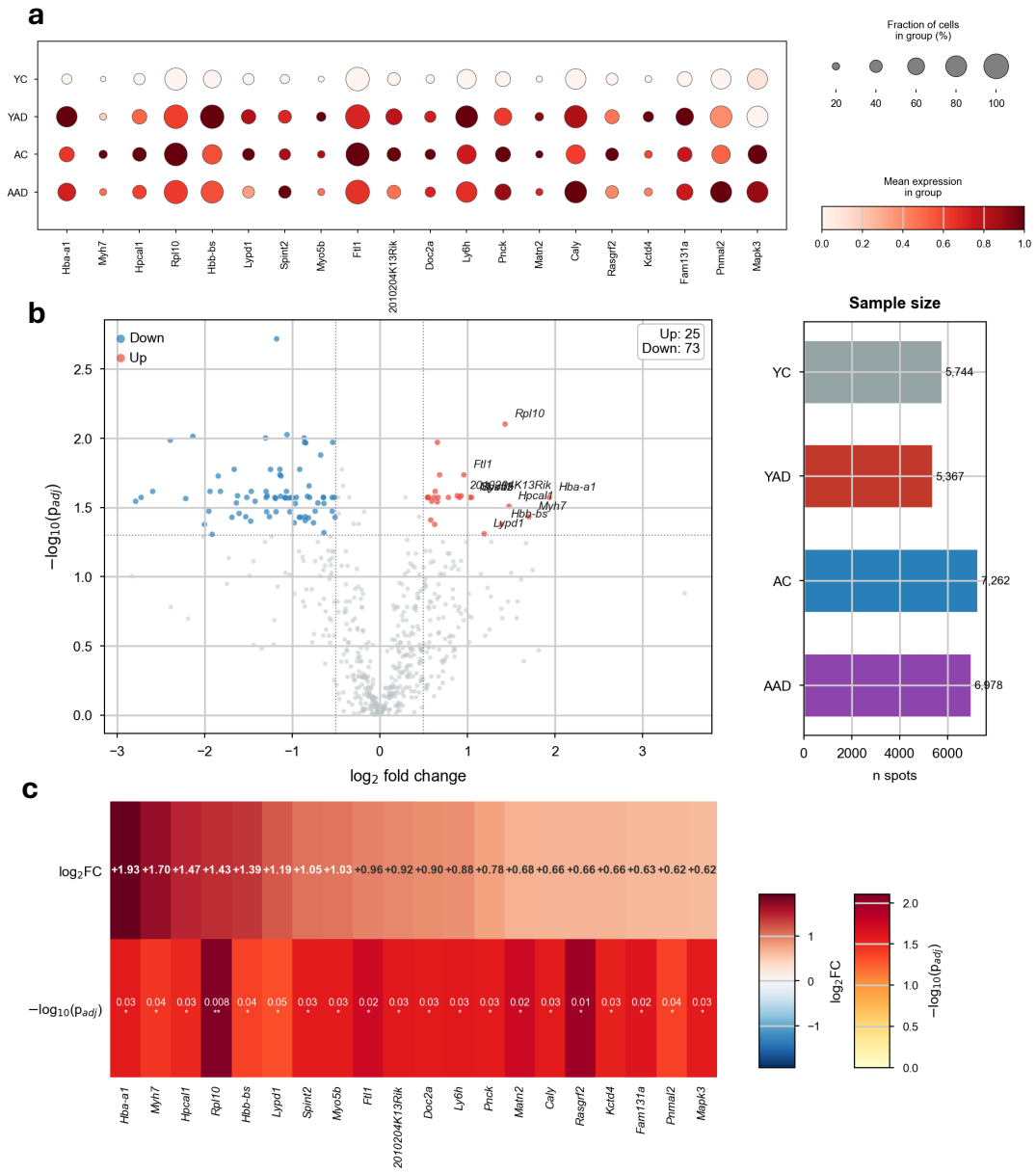


Figure 4-10 Upregulated genes in the ageing effect comparison (AC vs. YC). (a) Dotplot illustrating the top 20 upregulated genes in this comparison; (b) volcano plot highlighting differentially expressed genes meeting the significance criteria (adjusted p-value < 0.05 and $\log_2$FC > 0.5); and (c) statistics heatmap summarizing the $\log_2$ fold change and adjusted p-values for the top 20 upregulated genes in the ageing effect comparison. Significance levels are indicated as follows: p_{adj} < 0.05 (*), < 0.01 (**), < 0.001 (***), and ns (not significant; $p_{adj} \geq 0.05$).

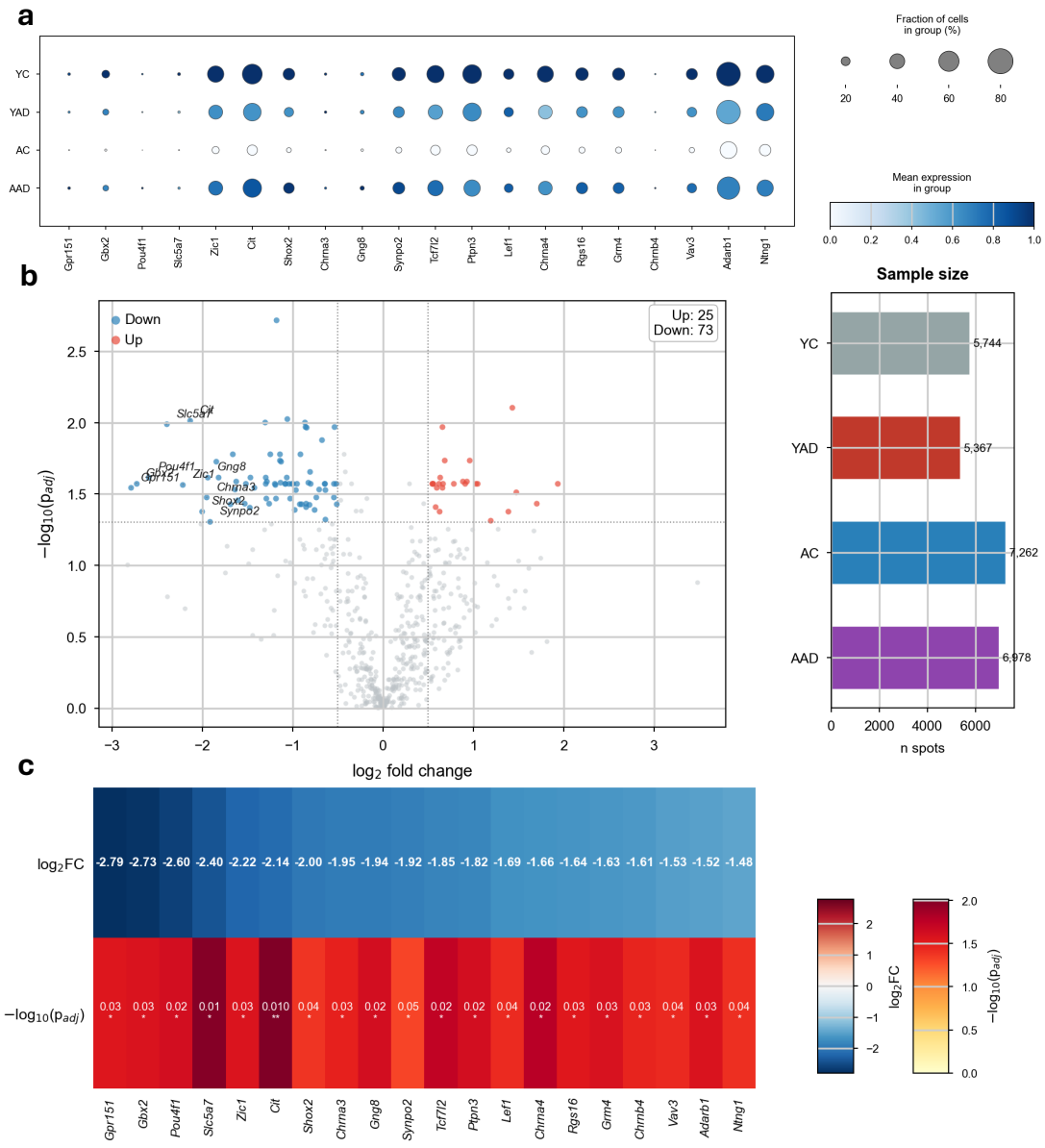
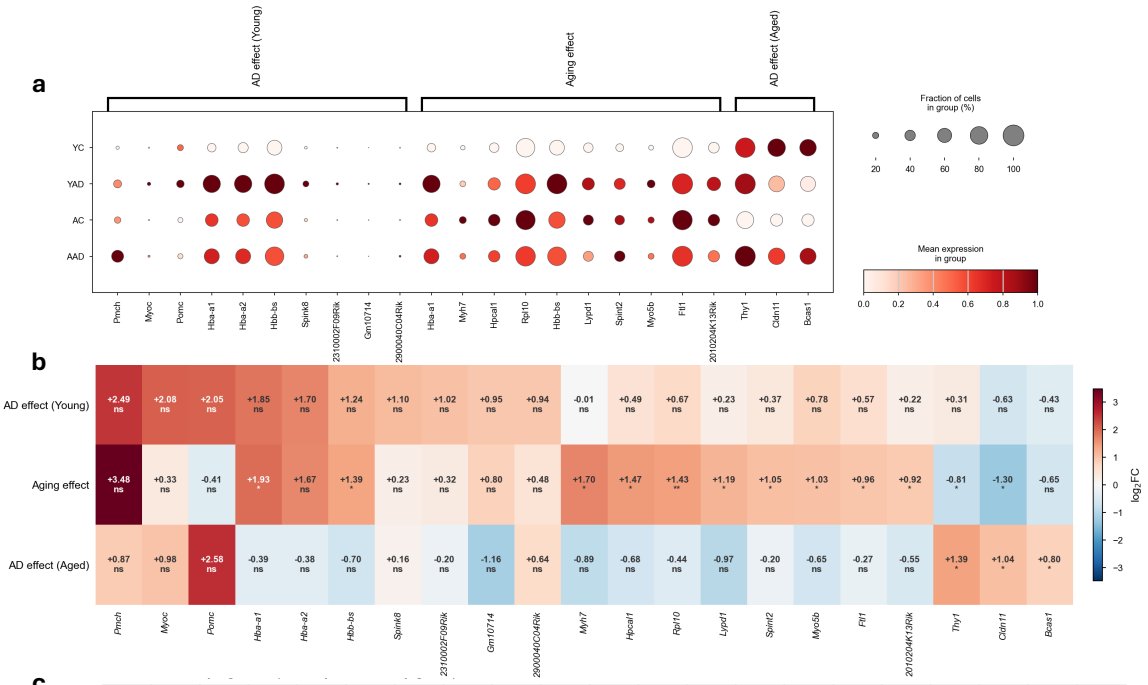


Figure 4-11 Downregulated genes in the ageing effect comparison (AC vs. YC). (a) Dotplot illustrating the top 20 downregulated genes in this comparison; (b) volcano plot highlighting differentially expressed genes meeting the significance criteria (adjusted p-value < 0.05 and $log_2FC < -0.5$); and (c) statistics heatmap summarizing the log_2 fold change and adjusted p-values for the top 20 downregulated genes in the ageing effect comparison. Significance levels are indicated as follows: $p_{adj} < 0.05$ (*), < 0.01 (**), < 0.001 (***), and ns (not significant; $p_{adj} \geq 0.05$).

those that are contrast-specific, while maintaining transparency through clear reporting of significance levels (padj < 0.05 (*), < 0.01 (**), < 0.001 (***), and ns for padj ≥ 0.05). Although these figures do not introduce new comparisons or additional genes, they provide valuable context by placing the results side by side. This makes it possible to see how a gene that is significant in one comparison behaves in the others – whether it remains consistently upregulated or downregulated, shows a weaker effect, or loses statistical significance. Overall, this integrated view helps clarify the broader patterns of aging- and AD-related molecular changes and strengthens the interpretability of the findings.

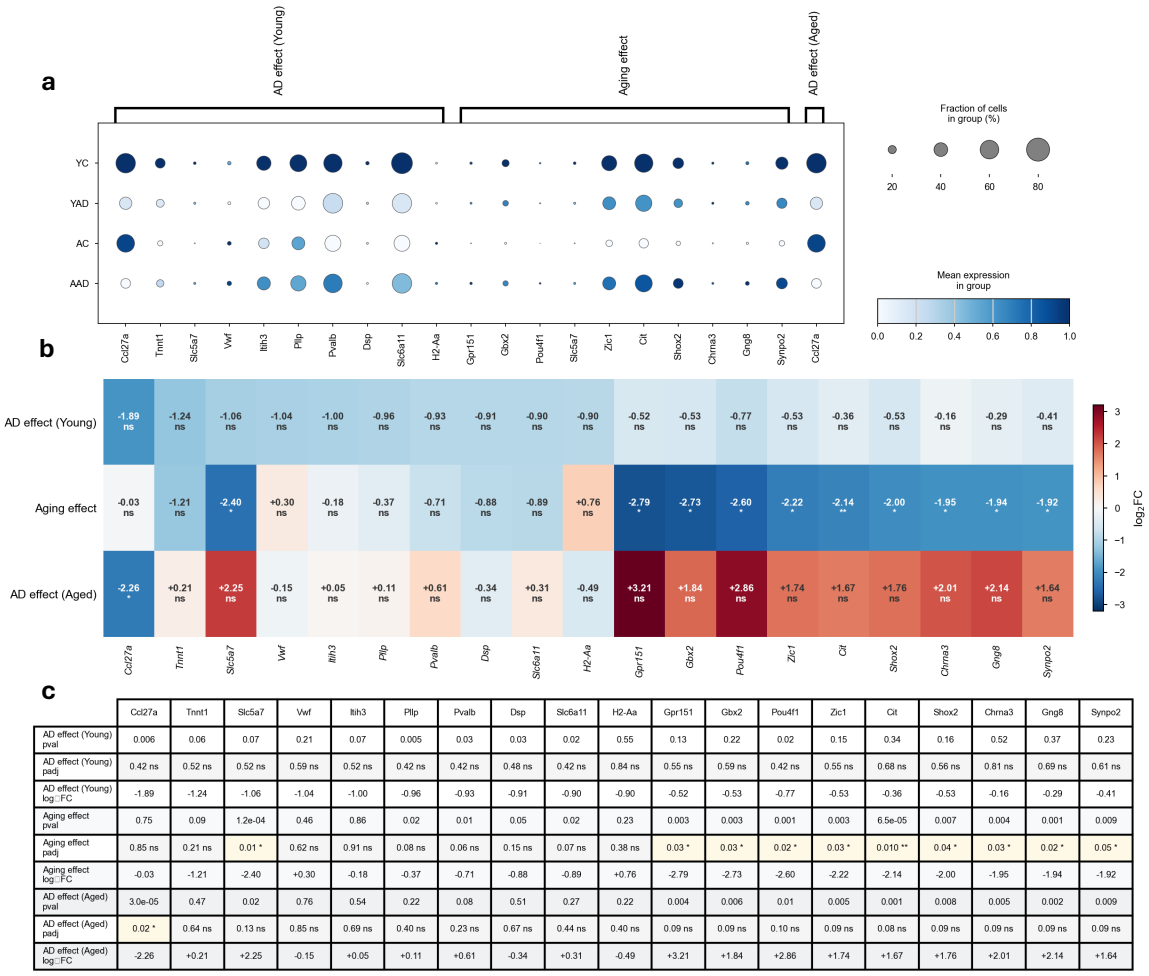
Figure 4-14 presents the spatial autocorrelation analysis computed for four AD candidate genes identified from the YAD and AAD comparisons using 999 permutations. The genes were ranked according to their Moran's I values, and red bars indicate spatially significant genes (I > 0.1 and padj < 0.05). The results indicate that all four AD candidate genes exhibit significant spatial autocorrelation within the tissue. Specifically, each gene showed a Moran's I value greater than 0.1 with an adjusted p-value below 0.05, demonstrating that their expression patterns are not randomly distributed but instead form spatially clustered regions. This suggests that these genes are associated with localized transcriptional patterns in the AD brain, potentially reflecting region-specific cellular activity or pathological processes. Overall, the findings highlight that these AD-related genes display strong and statistically significant spatial organization within the tissue. For instance, the spatial distribution of *Thy1* further supports this finding, as it demonstrates

significant spatial clustering within the tissue, consistent with the observed Moran's I results (Figure 4-15). This indicates that *Thy1* expression is not randomly distributed but instead follows a localized spatial pattern in the AD brain.



	Pmch	Myoc	Pomc	Hba-a1	Hba-a2	Hbb-bs	Spink8	2310002F09Rik	Gm10714	2900040C04Rik	Myh7	Hpcal1	Rpl10	Lypd1	Spint2	Myo5b	Ftl1	2010204K13Rik	Thy1	Cldn11	Bcas1
AD effect (Young) pval	0.10	0.02	0.34	0.04	0.05	0.08	0.17	0.07	0.22	0.33	0.95	0.12	0.29	0.29	0.10	0.15	0.18	0.36	0.15	0.09	0.13
AD effect (Young) padj	0.53 ns	0.42 ns	0.68 ns	0.49 ns	0.51 ns	0.52 ns	0.57 ns	0.52 ns	0.59 ns	0.68 ns	0.98 ns	0.53 ns	0.66 ns	0.66 ns	0.53 ns	0.55 ns	0.57 ns	0.69 ns	0.55 ns	0.52 ns	0.55 ns
AD effect (Young) log2FC	+2.49	+2.08	+2.05	+1.85	+1.70	+1.24	+1.10	+1.02	+0.95	+0.94	-0.01	+0.49	+0.67	+0.23	+0.37	+0.78	+0.57	+0.22	+0.31	-0.63	-0.43
Aging effect pval	0.04	0.47	0.89	0.002	0.01	0.007	0.70	0.28	0.15	0.36	0.005	0.004	2.6e-05	0.009	0.002	0.002	6.1e-04	0.001	8.8e-04	0.003	0.02
Aging effect padj	0.13 ns	0.63 ns	0.93 ns	0.03 *	0.06 ns	0.04 *	0.82 ns	0.44 ns	0.30 ns	0.53 ns	0.04 *	0.03 *	0.008 **	0.05 *	0.03 *	0.03 *	0.02 *	0.03 *	0.02 *	0.03 *	0.09 ns
Aging effect log2FC	+3.48	+0.33	-0.41	+1.93	+1.67	+1.39	+0.23	+0.32	+0.80	+0.48	+1.70	+1.47	+1.43	+1.19	+1.05	+1.03	+0.96	+0.92	-0.81	-1.30	-0.65
AD effect (Aged) pval	0.23	0.03	0.64	0.21	0.18	0.06	0.55	0.51	0.06	0.15	0.002	0.05	0.13	0.02	0.25	0.007	0.11	0.005	6.8e-05	2.4e-04	1.4e-04
AD effect (Aged) padj	0.41 ns	0.15 ns	0.78 ns	0.39 ns	0.36 ns	0.20 ns	0.70 ns	0.68 ns	0.20 ns	0.32 ns	0.09 ns	0.17 ns	0.30 ns	0.14 ns	0.42 ns	0.09 ns	0.27 ns	0.09 ns	0.02 *	0.04 *	0.03 *
AD effect (Aged) log2FC	+0.87	+0.98	+2.58	-0.39	-0.38	-0.70	+0.16	-0.20	-1.16	+0.64	-0.89	-0.68	-0.44	-0.97	-0.20	-0.65	-0.27	-0.55	+1.39	+1.04	+0.80

Figure 4-12 Combined overview of upregulated genes across comparisons. (a) Grouped dotplot showing up to the top 10 upregulated genes for each contrast (aging effect in control groups, AD effect in young groups, and AD effect in aged groups), organized by comparison. (b) Heatmap summarizing $\log_2$ fold change values of the selected genes across all contrasts, with significance indicated by stars. (c) Table presenting detailed differential expression statistics, including p-values, adjusted p-values, and $\log_2$ fold change for each gene across comparisons. Significance levels are indicated as follows: $p_{adj} < 0.05$ (*), < 0.01 (**), < 0.001 (***), and ns (not significant; $p_{adj} \geq 0.05$).



	Ccl27a	Tnnt1	Slc5a7	Vwf	Itih3	Pllp	Pvalb	Dsp	Slc6a11	H2-Aa	Gpr151	Gbx2	Pou4f1	Zic1	Cit	Shox2	Chrna3	Gng8	Synpo2
AD effect (Young) pval	0.006	0.06	0.07	0.21	0.07	0.005	0.03	0.03	0.02	0.55	0.13	0.22	0.02	0.15	0.34	0.16	0.52	0.37	0.23
AD effect (Young) padj	0.42 ns	0.52 ns	0.52 ns	0.59 ns	0.52 ns	0.42 ns	0.42 ns	0.48 ns	0.42 ns	0.84 ns	0.55 ns	0.59 ns	0.42 ns	0.55 ns	0.68 ns	0.56 ns	0.81 ns	0.69 ns	0.61 ns
AD effect (Young) log□FC	-1.89	-1.24	-1.06	-1.04	-1.00	-0.96	-0.93	-0.91	-0.90	-0.90	-0.52	-0.53	-0.77	-0.53	-0.36	-0.53	-0.16	-0.29	-0.41
Aging effect pval	0.75	0.09	1.2e-04	0.46	0.86	0.02	0.01	0.05	0.02	0.23	0.003	0.003	0.001	0.003	6.5e-05	0.007	0.004	0.001	0.009
Aging effect padj	0.85 ns	0.21 ns	0.01 *	0.62 ns	0.91 ns	0.08 ns	0.06 ns	0.15 ns	0.07 ns	0.38 ns	0.03 *	0.03 *	0.02 *	0.03 *	0.010 **	0.04 *	0.03 *	0.02 *	0.05 *
Aging effect log□FC	-0.03	-1.21	-2.40	+0.30	-0.18	-0.37	-0.71	-0.88	-0.89	+0.76	-2.79	-2.73	-2.60	-2.22	-2.14	-2.00	-1.95	-1.94	-1.92
AD effect (Aged) pval	3.0e-05	0.47	0.02	0.76	0.54	0.22	0.08	0.51	0.27	0.22	0.004	0.006	0.01	0.005	0.001	0.008	0.005	0.002	0.009
AD effect (Aged) padj	0.02 *	0.64 ns	0.13 ns	0.85 ns	0.69 ns	0.40 ns	0.23 ns	0.67 ns	0.44 ns	0.40 ns	0.09 ns	0.09 ns	0.10 ns	0.09 ns	0.08 ns	0.09 ns	0.09 ns	0.09 ns	0.09 ns
AD effect (Aged) log□FC	-2.26	+0.21	+2.25	-0.15	+0.05	+0.11	+0.61	-0.34	+0.31	-0.49	+3.21	+1.84	+2.86	+1.74	+1.67	+1.76	+2.01	+2.14	+1.64

Figure 4-13 Combined overview of downregulated genes across comparisons. (a) Grouped dotplot showing up to the top 10 downregulated genes for each contrast (aging effect in control groups, AD effect in young groups, and AD effect in aged groups), organized by comparison. (b) Heatmap summarizing $\log_2$ fold change values of the selected genes across all contrasts, with significance indicated by stars. (c) Table presenting detailed differential expression statistics, including p-values, adjusted p-values, and $\log_2$ fold change for each gene across comparisons. Significance levels are indicated as follows: $p_{adj} < 0.05$ (*), < 0.01 (**), < 0.001 (***), and ns (not significant; $p_{adj} \geq 0.05$).

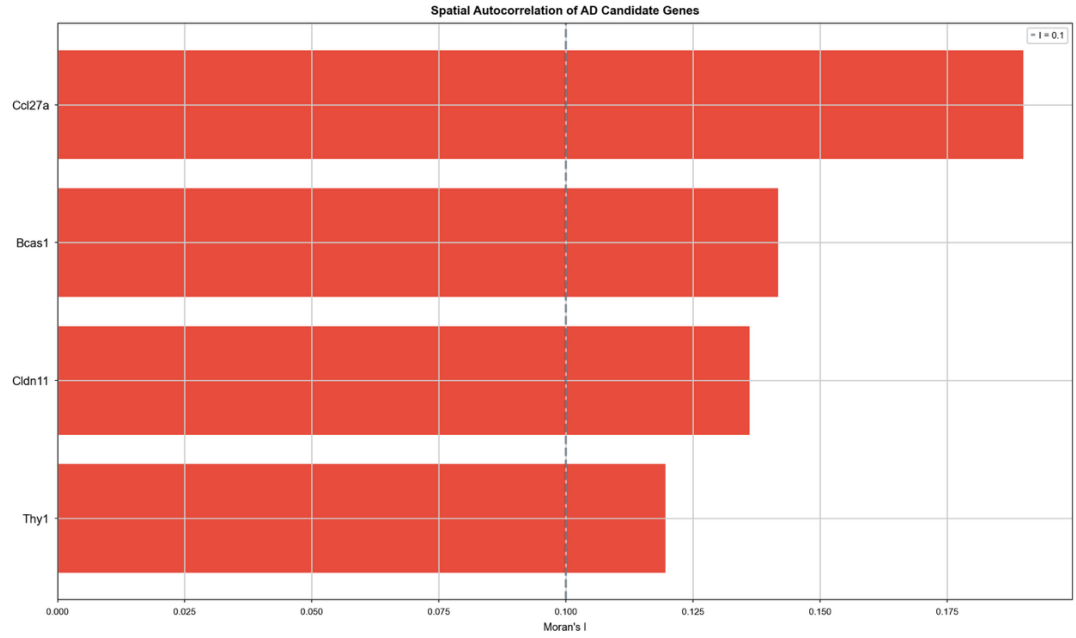


Figure 4-14 Spatial autocorrelation analysis of AD candidate genes using Moran's I. Red bars indicate spatially significant genes ($I > 0.1$ and padj < 0.05) and gray bars indicating non-significant genes. The dashed line represents the Moran's I threshold ($I = 0.1$).

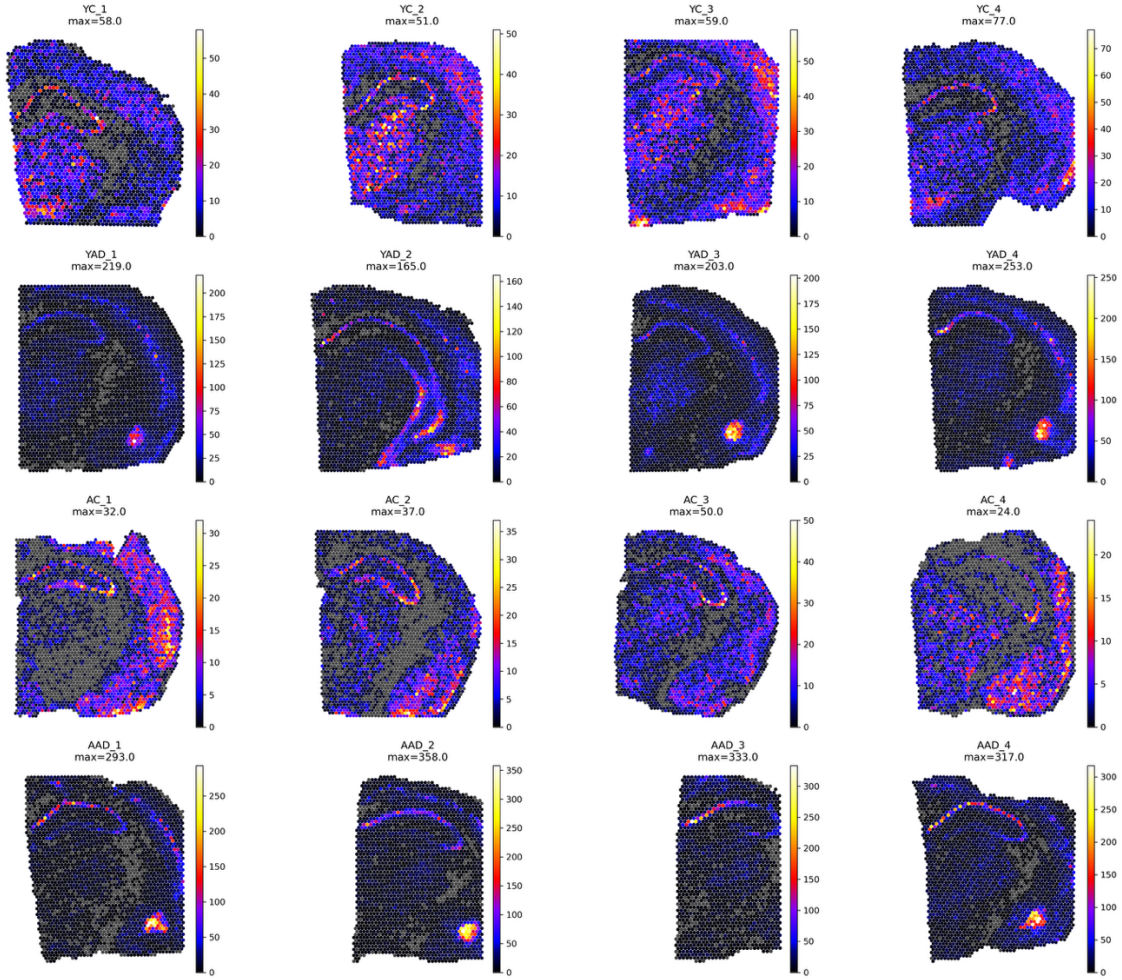


Figure 4-15 Heatmap of the spatial distribution of Thy1 across all samples, illustrating region-specific expression patterns and differences among aging, AD, and control conditions.(Spots are bigger than their normal size for better visualization)

4.4 Conclusion

Spatial transcriptomics, when combined with computational modeling, provides a unique view of the molecular architecture of aging and Alzheimer's disease (AD) brains. By preserving spatial context and mapping gene expression to specific anatomical regions, this approach enables the detection of localized molecular alterations that cannot be captured through conventional bulk RNA sequencing.

In this work, the developed program successfully implemented the complete standard spatial transcriptomics analysis pipeline, from preprocessing and normalization to clustering and differential expression analysis. The pipeline was specifically applied to investigate the effect of AD in both young and aged mouse groups, as well as to assess the independent impact of aging. Through these comparative analyses, the program identified genes that were significantly differentially expressed in specific groups, producing biologically meaningful and interpretable results.

Importantly, the program also demonstrated robustness in handling real experimental variability. Sample-related conditions – such as tissue tearing or partial tissue loss – had noticeable effects on statistical metrics, including adjusted p-values. Nevertheless, the pipeline was still able to pinpoint several significant genes whose roles in AD have been previously validated in the literature. This highlights not only the biological relevance of the findings but also the reliability and practical applicability of the developed computational framework in real-world datasets.

REFERENCES

1. M. Finicelli and R.-M. Liu, “Aging, Cellular Senescence, and Alzheimer’s Disease,” *International Journal of Molecular Sciences 2022, Vol. 23, Page 1989*, vol. 23, no. 4, p. 1989, Feb. 2022, doi: 10.3390/ijms23041989.
2. S. Srivastava, R. Ahmad, and S. K. Khare, “Alzheimer’s disease and its treatment by different approaches: A review,” *Eur. J. Med. Chem.*, vol. 216, no. 3, p. 113320, Apr. 2021, doi: 10.1016/j.ejmech.2021.113320.
3. A. D. T. Abeysinghe, R. D. U. S. Deshapriya, and C. Udawatte, “Alzheimer’s disease; a review of the pathophysiological basis and therapeutic interventions,” *Life Sci.*, vol. 256, p. 117996, Sep. 2020, doi: 10.1016/j.lfs.2020.117996.
4. N. K. Jha, S. K. Jha, R. Kar, P. Nand, K. Swati, and V. K. Goswami, “Nuclear factor-kappa β as a therapeutic target for Alzheimer’s disease,” *J. Neurochem.*, vol. 150, no. 2, pp. 113–137, Jul. 2019, doi: 10.1111/jnc.14687.
5. “2023 Alzheimer’s disease facts and figures,” *Alzheimer’s and Dementia*, vol. 19, no. 4, pp. 1598–1695, Apr. 2023, doi: 10.1002/alz.13016.
6. N. Habib *et al.*, “Disease-associated astrocytes in Alzheimer’s disease and aging,” *Nature Neuroscience 2020 23:6*, vol. 23, no. 6, pp. 701–706, Apr. 2020, doi: 10.1038/s41593-020-0624-8.
7. J. F. Navarro *et al.*, “Spatial Transcriptomics Reveals Genes Associated with Dysregulated Mitochondrial Functions and Stress Signaling in Alzheimer Disease,” *iScience*, vol. 23, no. 10, Oct. 2020, doi: 10.1016/j.isci.2020.101556.

8. P. L. De Jager *et al.*, “Data Descriptor: A multi-omic atlas of the human frontal cortex for aging and Alzheimer’s disease research Background & Summary,” *Sci. Data*, vol. 5, no. 1, pp. 180142-, Aug. 2018, doi: 10.1038/sdata.2018.142.
9. E. Schueller *et al.*, “Dysregulation of histone acetylation pathways in hippocampus and frontal cortex of Alzheimer’s disease patients,” *European Neuropsychopharmacology*, vol. 33, pp. 101–116, Apr. 2020, doi: 10.1016/j.euroneuro.2020.01.015.
10. D. Wang *et al.*, “Elevated plasma levels of exosomal BACE1-AS combined with the volume and thickness of the right entorhinal cortex may serve as a biomarker for the detection of Alzheimer’s disease,” *Mol. Med. Rep.*, vol. 22, no. 1, pp. 227–238, Jul. 2020, doi: 10.3892/mmr.2020.11118.
11. T. Du *et al.*, “Magnetic resonance imaging of the entorhinal cortex and hippocampus in mild cognitive impairment and Alzheimer’s disease,” *J. Neurol. Neurosurg. Psychiatry*, vol. 71, no. 4, pp. 441–447, Oct. 2001, doi: 10.1136/jnnp.71.4.441.
12. C. Raulin, S. V. Doss, Z. A. Trottier, T. C. Ikezu, G. Bu, and C. C. Liu, “ApoE in Alzheimer’s disease: pathophysiology and therapeutic strategies,” *Molecular Neurodegeneration 2022 17:1*, vol. 17, no. 1, pp. 72-, Nov. 2022, doi: 10.1186/s13024-022-00574-4.
13. M. Biagioli *et al.*, “Unexpected expression of α- and β-globin in mesencephalic dopaminergic neurons and glial cells,” *Proceedings of the National Academy of*

Sciences, vol. 106, no. 36, pp. 15454–15459, Sep. 2009, doi: 10.1073/pnas.0813216106.

14. R. Tortuyaux *et al.*, "Physiopathological changes of ferritin mRNA density and distribution in hippocampal astrocytes in the mouse brain," *J. Neurochem.*, vol. 164, no. 6, pp. 847–857, Mar. 2023, doi: 10.1111/jnc.15747.
15. L. Remesal *et al.*, "Targeting iron-associated protein Ftl1 in the brain of old mice improves age-related cognitive impairment," *Nat. Aging*, vol. 5, no. 10, pp. 1957–1969, Oct. 2025, doi: 10.1038/s43587-025-00940-z.
16. H. Ota, T. Ohnuma, A. Kodama, T. Shimizu, K. Sugawara, and F. Yamamoto, "The Cholinergic Receptor Nicotinic α3 Was Reduced in the Hippocampus of Early Cognitively Impaired Adult Male Mice and Upregulated by Nicotine and Cytisine in HT22 Cells," *Cells*, vol. 14, no. 5, p. 340, Mar. 2025, doi: 10.3390/cells14050340.
17. J. Rybnicek *et al.*, "CHRNA5 links chandelier cells to severity of amyloid pathology in aging and Alzheimer's disease," *Translational Psychiatry 2024 14:1*, vol. 14, no. 1, pp. 83-, Feb. 2024, doi: 10.1038/s41398-024-02785-3.
18. J. W. Hofmann, T. McBryan, P. D. Adams, and J. M. Sedivy, "The effects of aging on the expression of Wnt pathway genes in mouse tissues," *AGE 2014 36:3*, vol. 36, no. 3, pp. 1033–1040, Feb. 2014, doi: 10.1007/s11357-014-9618-3.

CHAPTER 5:

A MULTI-OBJECTIVE SCORING (MOS) FRAMEWORK FOR DETECTING CROSS-MODAL SPATIAL SIMILARITY: CONCEPTUAL AND DIRECT FORMULATIONS

5.1 Introduction

Spatial multi-omics has become a powerful approach for exploring complex biological systems by integrating corresponding molecular layers. This can be achieved either by profiling multiple modalities on the same tissue section or by aligning consecutive, closely matched sections to maintain spatial continuity [1], [2], [3], [4], [5]. With platforms such as 10x Genomics Visium Spatial Gene Expression, researchers can characterize gene expression patterns of tissue through spatial transcriptomics[6], [7]. Similarly, mass spectrometry imaging (MSI), including matrix-assisted laser desorption/ionization (MALDI-MSI), enables spatially resolved mapping of other biomolecules, including lipids, directly from tissue sections[8], [9], [10], [11], [12], [13]. Combining these technologies together provides a more extensive and complementary molecular view of tissue structure at high spatial resolution[4], [5].

Computational approaches for multi-omics integration have evolved from non-spatial models to frameworks that incorporate tissue architecture [5]. Early integrative methods, such as Multi-Omics Factor Analysis (MOFA) and its successor MOFA+, established foundational approaches for identifying major patterns across molecular layers by understanding underlying sources of variation [14], [15]. These models effectively separate shared and modality-specific sources of heterogeneity. However, these models were primarily developed for non-spatial single-cell data and do not explicitly incorporate anatomical coordinates. To address this limitation, more recent methods such as COSMOS in 2025 and MAGPIE in 2026 have been introduced. COSMOS uses graph

neural networks to integrate paired spatial datasets. It can capture complementary features while preserving shared spatial coordinates [16]. Similarly, the MAGPIE framework was developed to support the co-registration of spatial transcriptomics and metabolomics data from either the same or consecutive tissue sections. This framework emphasizes the importance of mapping different modalities into a shared coordinate space to enable cross-modal analysis [4].

With significant progress in spatial profiling technologies, there are still major methodological challenges in robust integration of spatial transcriptomics and MALDI-MSI datasets [16]. There are significant differences in these modalities including sampling geometry, spatial resolution, signal scaling, noise structure, and underlying physical measurement principles. Among these differences, the mismatch in grid architecture poses one of the most substantial obstacles and further intensifies resolution disparities. The 10x Genomics Visium platform employs a hexagonal array of discrete capture spots with defined center-to-center spacing for the Visium Gene Expression Slides. These spots do not form a continuous sampling surface, and gaps exist between neighboring capture locations. Therefore, it creates a partially discrete spatial representation of the tissue. The overall coverage of spots is less than 30% of the capturing area; however, the use of the hexagonal grid is more efficient compared to rectangular grid by reducing the sampling size by 13.4% [17], [18]. In contrast, MALDI-MSI produces data on a dense Cartesian grid in which adjacent pixels are tightly connected. Therefore, exact spatial alignment between modalities is rarely achievable in

practice, and relying on a single similarity metric fails to adequately represent the complex, multidimensional structure of biological spatial relationships[4].

The importance of robust integration becomes especially evident when considering complex molecular interactions that drive many diseases. Spatial multi-omics is particularly valuable for revealing how intracellular processes are formed by, and interact with, the surrounding microenvironment. These processes become more important in areas such as tumor immunology, developmental biology, and tissue injury. For example, integrated analyses have linked metabolic changes – such as increased lactate production – to specific transcriptomic signatures in damage-associated macrophages during pulmonary fibrosis [4]. In oncology, mapping gene expression alongside metabolic gradients has revealed how changes in the extracellular matrix and alterations in cellular metabolism work together to drive tumor progression and enable immune evasion [4].

Considering these advancements in this field, major challenges remain unsolved due to fundamental differences in sampling geometry and measurement principles between technologies such as spatial transcriptomics and mass spectrometry. This highlights the need for generalized, geometry-aware frameworks that can quantify spatial similarity without relying on perfect co-registration across modalities.

We present a Multi-Objective Scoring (MOS) Framework designed to quantify spatial similarity between heterogeneous omics modalities without requiring perfect co-registration. This framework is designed to remain robust in the presence of minor spatial misalignments and orientation differences, while adapting to variations in data quality. It

integrates complementary geometric and statistical descriptors, including radial profiles, spatial histograms, quadrant statistics, spatial autocorrelation (Moran's I), and intensity-distribution similarity, to capture both direct spatial correspondence and higher-order structural relationships. The component metrics are combined via a weighted ensemble model, with weights calibrated using synthetic data that systematically simulate spatial misalignment, orientation perturbations, resolution differences, acquisition variability, and realistic mouse half-brain anatomical geometry.

Our framework uses selected spatial patterns within the gene modality and subsequently searches for candidate lipid features that exhibit similar structural signatures. This targeted strategy substantially reduces computational cost compared to exhaustively evaluating cross-modal similarity for all gene–lipid combinations. We applied the framework to Alzheimer's disease (AD) mouse brain tissue as a case study, given its extensive investigation over several decades and its relevance as a complex neurodegenerative disorder [19], [20], [21], [22]. This highlights the utility of the proposed framework in a biologically meaningful context. The main aim of this work presented here is not to generate AD-specific insights, but to design and validate a generalized, geometry-aware computational framework for cross-modal spatial integration. However, this framework can be applied either directly or conceptually to a wide range of spatial omics combinations and provides a scalable foundation for future multi-omic investigations across tissues, organisms, and disease models, which reflects its modality-agnostic nature.

5.2 Methods

5.2.1 Sample Collection

Murine brains from young (10–12 weeks) and aged (52–54 weeks) female 3xTg-AD transgenic mice and age-matched controls (n = 4 per group) [23] were harvested and rapidly flash-frozen in liquid nitrogen to preserve tissue morphology and RNA integrity. The brains were cryosectioned coronally at 10 μm thickness using a Leica cryostat. Sections encompassing the cortex and hippocampus, at a location approximately 2.24 ± 0.21 mm from the bregma, were collected in series. To enable cross-modal spatial analysis, consecutive sections were mounted in an alternating manner: one section was mounted onto glass microscope slides (Fisher Scientific) for MALDI mass spectrometry imaging [24], [25], [26], and the immediately adjacent section was mounted onto 10x Genomics Visium Spatial Gene Expression slides for spatial transcriptomic profiling. This serial mounting strategy ensured that lipidomic and transcriptomic measurements were obtained from closely matched anatomical regions. Additionally, seven serial sections from a single animal were mounted on 10x Genomics Tissue Optimization (TO) slides for performing the Visium Tissue Optimization Protocol. This protocol was implemented to determine the optimal permeabilization time, allowing efficient mRNA release while preserving tissue morphology and ensuring high-quality, reproducible gene expression data (Figure 5-1). After mounting, all slides were placed in plastic slide jars (the lids of which were tightly sealed to avoid moisture exposure), transferred in dry ice, and stored at −80 °C until further processing.

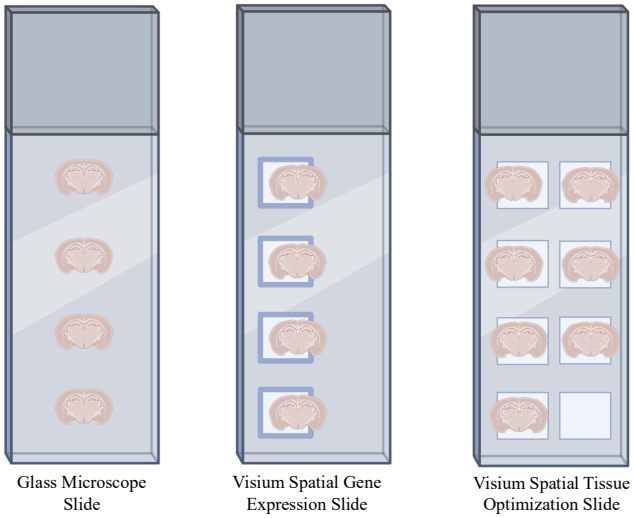


Figure 5-1 Schematic illustration of the glass microscope slide, Visium Spatial Gene Expression slide, and Visium Spatial Tissue Optimization slide (from left to right)[2].

5.2.2 MALDI-MSI Lipidomics

5.2.2.1 Sample Preparation and MALDI-MSI Acquisition

The washing, sublimating, scanning, and mass spectrometry imaging were performed at the Mass Spectrometry Research Center (MSRC) at Vanderbilt University. Samples, mounted on glass slides, were transferred in the plastic slide jars on dry ice to Vanderbilt. The samples on glass slides were submerged in four consecutive baths containing isotonic ammonium formate solution (150 mM for 45 seconds. Following the washing steps, the slides were gently dried under a continuous stream of nitrogen gas to minimize moisture and prevent analyte delocalization. To further ensure complete dehydration and stabilization of the tissue before subsequent analyses, the dried slides were placed in a vacuum desiccator for 20 minutes[27]. A matrix solution DMACA was applied through

[2] https://www.biorender.com

the sublimation method (HTX SubliMATE) with optimized parameters (Table 5-1) [27]. The slides were first fluorescence-scanned using a ZEISS Axio Scan.Z1 slide scanner. They were then analyzed with Bruker's timsTOF fleX mass spectrometer under the parameters listed in Table 5-2. Following imaging, the samples were placed in the slide jar and vacuumed, sealed, and stored at −80 °C freezer at the University of Tennessee., After removing from the freezer, the matrix was removed by three sequential washes in pure methanol, each lasting 1 minute. Finally, the slides were stained and imaged at 2× resolution using a Leica M205 FCA microscope at the Advanced Microscopy and Imaging Core (AMIC) of the University of Tennessee, Knoxville (Figure 5-2).

5.2.2.2 MALDI-MSI Data Processing

For MALDI-MSI data processing, our previously published algorithm was employed and enhanced, and new features were added to the algorithm to enable it to read and analyze datasets from various companies and also to perform better peak detection and peak alignment [28]. For this study, data preprocessing involved smoothing (Savitzky–Golay filter)[29], baseline correction (asymmetric least squares, AsLS) [30], peak alignment using four lock masses [496.339 (LPC 16:0) [31], [32], 734.569 (PC 32:0) [31], [33], 760.585 (PC 34:1) [31], [32], [34], and 806.569 (PC 38:6) [34]] together with drift line correction, global peak detection (minimum peak distance = 0.024 Da), peak augmentation (± 0.012 Da or up to ± 2 indices), and total ion current (TIC) normalization. As the goal was to examine the distribution pattern of the analyte within each sample to

Table 5-1 DMACA matrix sublimation conditions

Concentration (mg/mL)	Temperature of Sublimation (°C)	Temperature of the Sample (°C)	Time of Sublimation (min)	Crystalization Temperature (°C)	Crystalization Time (s)
3	180	-78	15	100	30

Table 5-2 MALDI timsTOF fleX instrumental parameters for positive ion mode

Transfer	MALDI Plate Offset (V)	50.0
	Deflection 1 Delta (V)	70.0
	Funnel 1 RF (Vpp)	350.0
	isCID Energy (eV)	10.0
	Funnel 1 RF (Vpp)	300.0
	Multipole 1 RF (Vpp)	300.0
Collision Cell	Collision Energy (eV)	10.0
	Collision RF (Vpp)	2500.0
Quadrupole	Ion Energy (eV)	5.0
	Low Mass (analyte)	450.0
Focus Pre Tof	Transfer Time (µs)	90.0
	Pre Pulse Storage (µs)	12.0
Detection	High Sensitivity Detection – Low Sample Amount	False
	Focus Mode	False

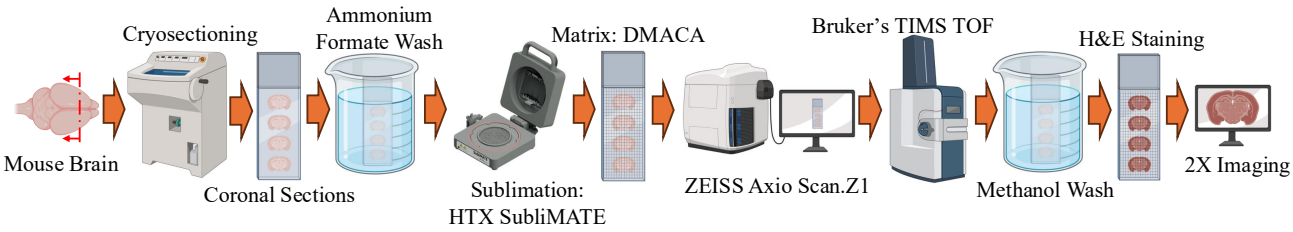


Figure 5-2 Schematic overview of the experimental workflow for MALDI-MSI, highlighting tissue preparation, matrix application, data acquisition, and imaging[3].

[3] https://www.biorender.com

compare across groups, only analytes consistently detected across all sample groups were retained, and those not commonly present were excluded from further analysis. Subsequently, normalization, logarithmic transformation, Z-score scaling, principal component analysis (PCA), and k-nearest neighbors (kNN) were performed. Leiden clustering at a low resolution (res = 0.03) was then applied to identify and remove off-tissue and contaminated pixels. The processed data were stored in AnnData (.h5ad) format (Figure 5-3).

5.2.3 Spatial Transcriptomics

5.2.3.1 Visium Spatial Tissue Optimization Workflow

In this process, 10x Genomics Tissue Optimization (TO) slides with sections on were used. 10x Genomics TO slides contained arrays of spatially barcoded capture spots. The tissue sections were then treated with a permeabilization enzyme for varying durations. This enzyme partially digested the extracellular matrix and cell membranes, allowing mRNA molecules to diffuse out of the tissue and bind to the spatially barcoded capture

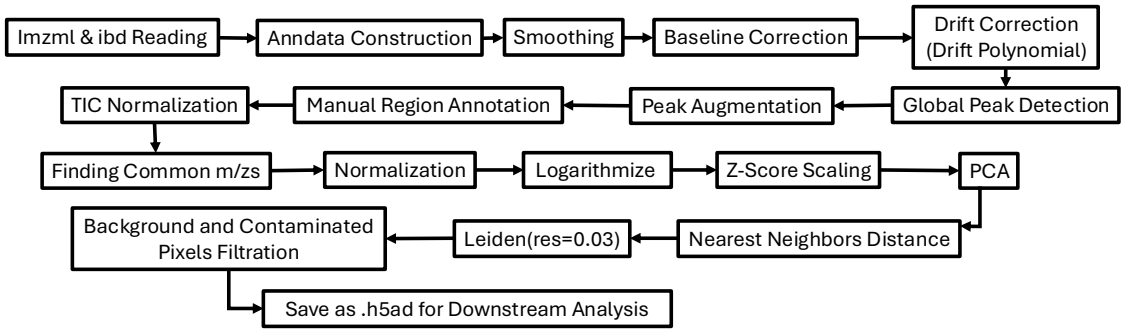


Figure 5-3 Flowchart summarizing the computational steps for MALDI-MSI data implemented in Python.

oligonucleotides on the slide. After permeabilization, reverse transcription was performed to convert the captured mRNA into spatially indexed cDNA. The cDNA was subsequently fluorescently labeled, enabling visualization of RNA capture efficiency across different permeabilization conditions. By analyzing the fluorescence intensity while monitoring tissue integrity, the optimal permeabilization time was determined to maximize RNA yield without compromising tissue morphology. This optimized condition was then applied to subsequent Visium experiments, ensuring high-quality spatial transcriptomic profiling, accurate gene expression quantification, and correct mapping of spatial gene expression patterns within the tissue (Figure 5-4).

5.2.3.2 Spatial Transcriptomics

10x Genomics Visium Spatial Gene Expression slides contained an array of spatially barcoded capture spots that enabled the assignment of transcriptomic information to precise tissue locations. The 10x Genomics Visium Spatial Gene Expression protocol was followed. The tissue sections were briefly warmed to ensure proper adhesion to the slide surface and were subsequently fixed to preserve tissue morphology and RNA integrity. The sections were then stained using hematoxylin and eosin to visualize tissue architecture and imaged under bright-field microscopy, generating high-resolution images to guide downstream analyses, at Advanced Microscopy and Imaging Core (AMIC) of the University of Tennessee, Knoxville. Following imaging, the slides were processed for spatial transcriptomic library preparation according to the 10x Genomics Visium protocol with a permeabilization time of 12 minutes. The resulting cDNA

libraries were sent to the Genomics Core at the University of Tennessee, Knoxville, for sequencing on an Illumina platform (NovaSeq). Sequencing outputs were delivered as FASTQ files, containing raw reads indexed by spatial barcodes and ready for computational processing and spatial gene expression analysis (Figure 5-4).

5.2.3.3 Spatial Transcriptomic Data Analysis

Brightfield H&E-stained tissue images were first examined in Loupe Browser for fiducial frame alignment and determination of the region of interest (ROI). Sequenced data were obtained as FASTQ files from the Genomic Core and were processed using 10x Genomics Space Ranger to align reads, assign them to spatial barcodes, and generate spot-level count matrices. The processed data were then imported into Python for normalization, and the spatial coordinates of each spot (X and Y, in micrometers) were incorporated into the corresponding samples' AnnData objects (Figure 5-4).

No further large-scale analysis was performed on the spatial transcriptomics dataset. Instead, eight genes – *Mbp* [35], [36], *Plp1* [37], [38], *S100b* [39], *Hspa1a* [40], [41], *Mapt* [42], [43], [44], *Thy1* [45], *App* [46], [47], and *Apoe* [46], [48], [49], [50] – were selected for subsequent analysis based on previously reported evidence of their upregulation, downregulation, dysregulation, or involvement in pathological mechanisms associated with aging, AD, or both.

5.2.4 Multi-Objective Scoring Framework

We employed a MOS pipeline that did not require pixel-to-pixel overlay. We focused on identifying differentially expressed genes within the spatial transcriptomic dataset and

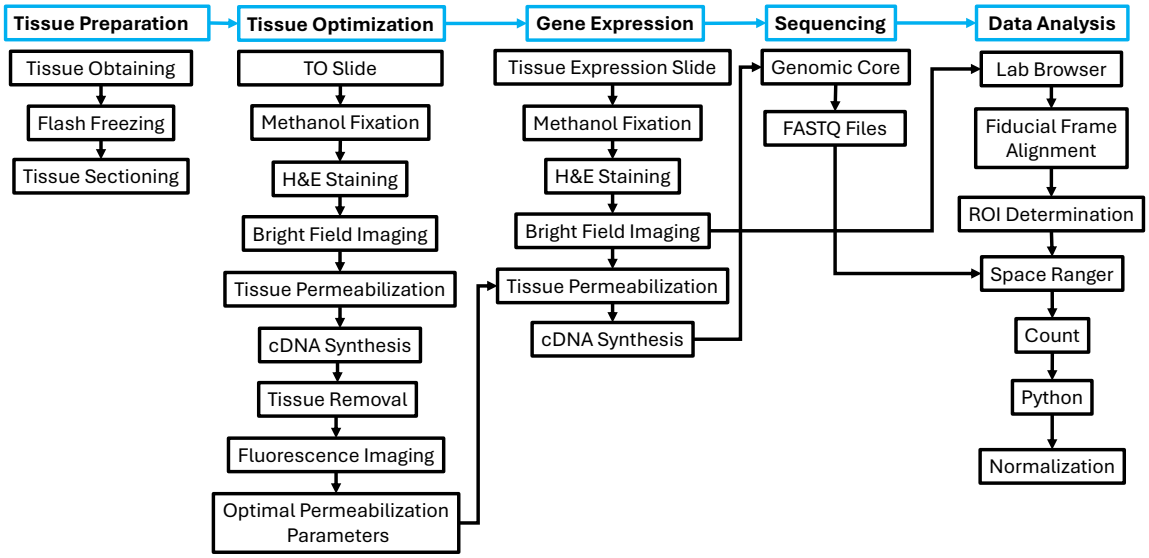


Figure 5-4 Flowchart of Visium tissue optimization, spatial gene expression analysis, and data preprocessing pipeline.[4] TO: tissue optimization

evaluated whether corresponding spatial patterns can be detected in the lipidomic images. Integrating these coordinate systems (Visium: hexagonal spot; MALDI-MSI: Cartesian) requires resampling or interpolation and resolution differences (Visium spots: diameter ~55 µm with a center-to-center spacing of 100 µm; MALDI-MSI pixels: ~60 µm) further complicate alignment. These disparities in geometry and resolution can lead to misalignment, loss of spatial detail, and challenges in downstream multimodal analyses. Our framework can identify similarities even with slight misalignment between modalities without the need for pixel-to-pixel co-registration.

[4] Some parts of the flowchart was adopted from https://www.bmkgene.com/10x-genomics-visium-spatial-transcriptome-product/

Rationale and Conceptual Foundation

A central methodological decision in this work was to avoid defining gene–lipid similarity using a single statistical metric. The primary reason for this is that spatial biological similarity is inherently multidimensional. Gene expression and lipid composition are measured using fundamentally different technologies that operate on different physical principles and exhibit distinct intensity distributions, spatial resolutions, and noise characteristics. As a result, no single metric is sufficient to characterize cross-modal spatial agreement robustly.

Therefore, we designed a MOS framework that integrates complementary spatial descriptors into a unified similarity score. Conceptually, this approach is analogous to how humans evaluate resemblance between related individuals [51], [52]. When we say that two siblings look alike, we are not basing that impression on a single trait – such as hair color or the shape of their eyes. Instead, the sense of resemblance emerges from the combined similarity of many features: facial structure, skin tone, proportions, and even subtle expressions (Figure 5-5). No single attribute is enough on its own. It is the consistent alignment across several independent traits that creates a convincing and reliable perception of similarity.

The same principle applies to spatial molecular similarity. A gene and a lipid may share similar global intensity trends but differ in local clustering structure. Alternatively, they may exhibit similar clustering strength but be spatially shifted. If similarity were defined solely by a single metric – such as one Pearson correlation of intensity value – important

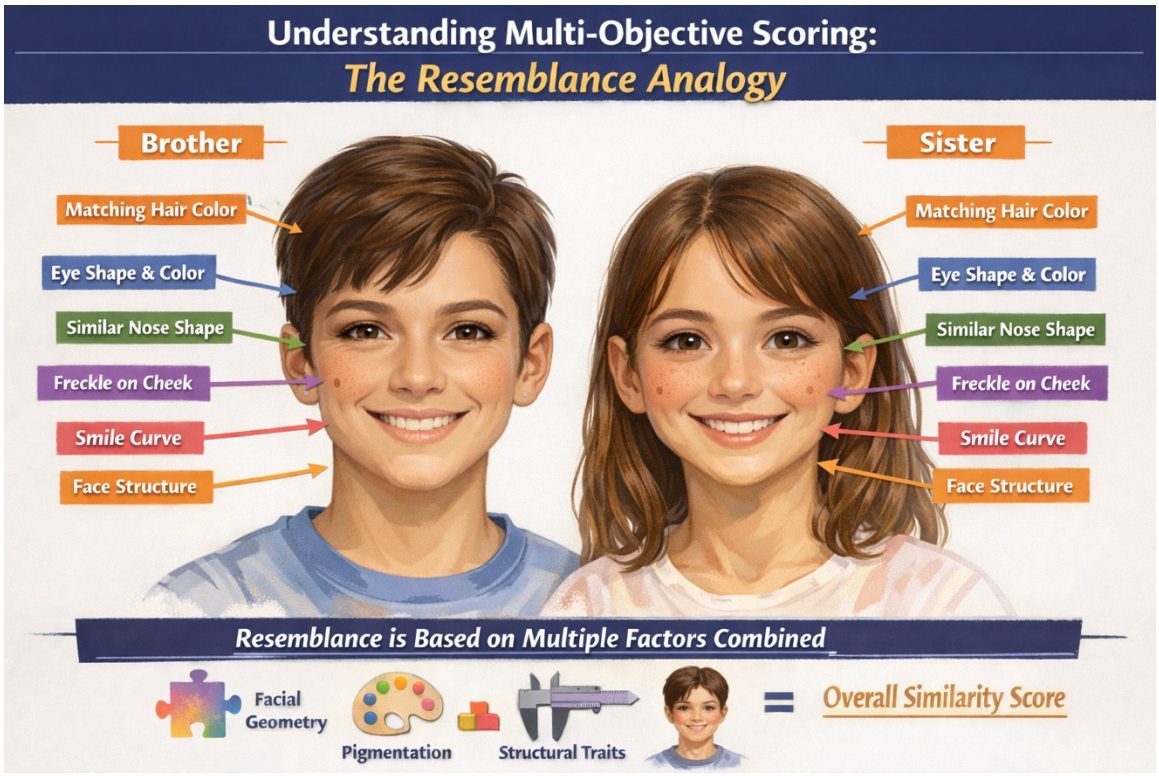


Figure 5-5 schematic demonstration of sibling resemblance[5]

structural differences could be overlooked, or false matches could be incorrectly accepted. For this reason, the similarity evaluation was decomposed into complementary components, each capturing a distinct dimension of spatial organization.

5.2.4.1 Architecture and Mathematical Formula of Multi-Objective Scoring Framework

The MOS Framework was engineered as a robust framework for cross-modal spatial correlation. Unlike deep-learning approaches that require extensive training, this pipeline utilizes a heuristic-analytic architecture.

---

[5] The image is produced by AI (https://chatgpt.com)

The "match" between a target gene (G) and a candidate lipid (M) is quantified by a comprehensive scoring engine (final combined score). This engine integrates direct spatial overlap with geometric descriptor similarity into a single scalar value (Figure 5-6).

5.2.4.1.1 Component Metrics

Let $G(x, y)$ represent the spatial pattern of a gene, and let $M(x, y)$ represent the corresponding m/z. To quantify similarity, we divided the evaluation into two complementary components:

$$\text{Sim} = \{\text{Sim}_{\text{coord}}, \text{Sim}_{\text{desc}}\}$$

where coordinate-based similarity measures direct spatial alignment, and descriptor-based similarity evaluates higher-order structural relationships between the modalities.

5.2.4.1.1.1 Coordinate-Based Similarity (Sim_{coord})

All spatial maps were first geometrically aligned and then interpolated onto a shared 50×50 grid to ensure a common spatial resolution. Let $\tilde{G}$ and $\tilde{M}$ denote the interpolated gene and m/z grids, respectively.

a) Value Correlation

Spatial intensity similarity was quantified using Pearson's correlation coefficient[53]:

$$r(\tilde{G}, \tilde{M}) = \frac{\sum_i (\tilde{G}_i - \bar{G})(\tilde{M}_i - \bar{M})}{\sqrt{\sum_i (\tilde{G}_i - \bar{G})^2 \sum_i (\tilde{M}_i - \bar{M})^2}}$$

$$r_{\text{val}} = r(\tilde{G}, \tilde{M})$$

This metric measures the linear relationship of spatial intensity distributions, which means it measures how strongly the intensities of the two signals vary together across

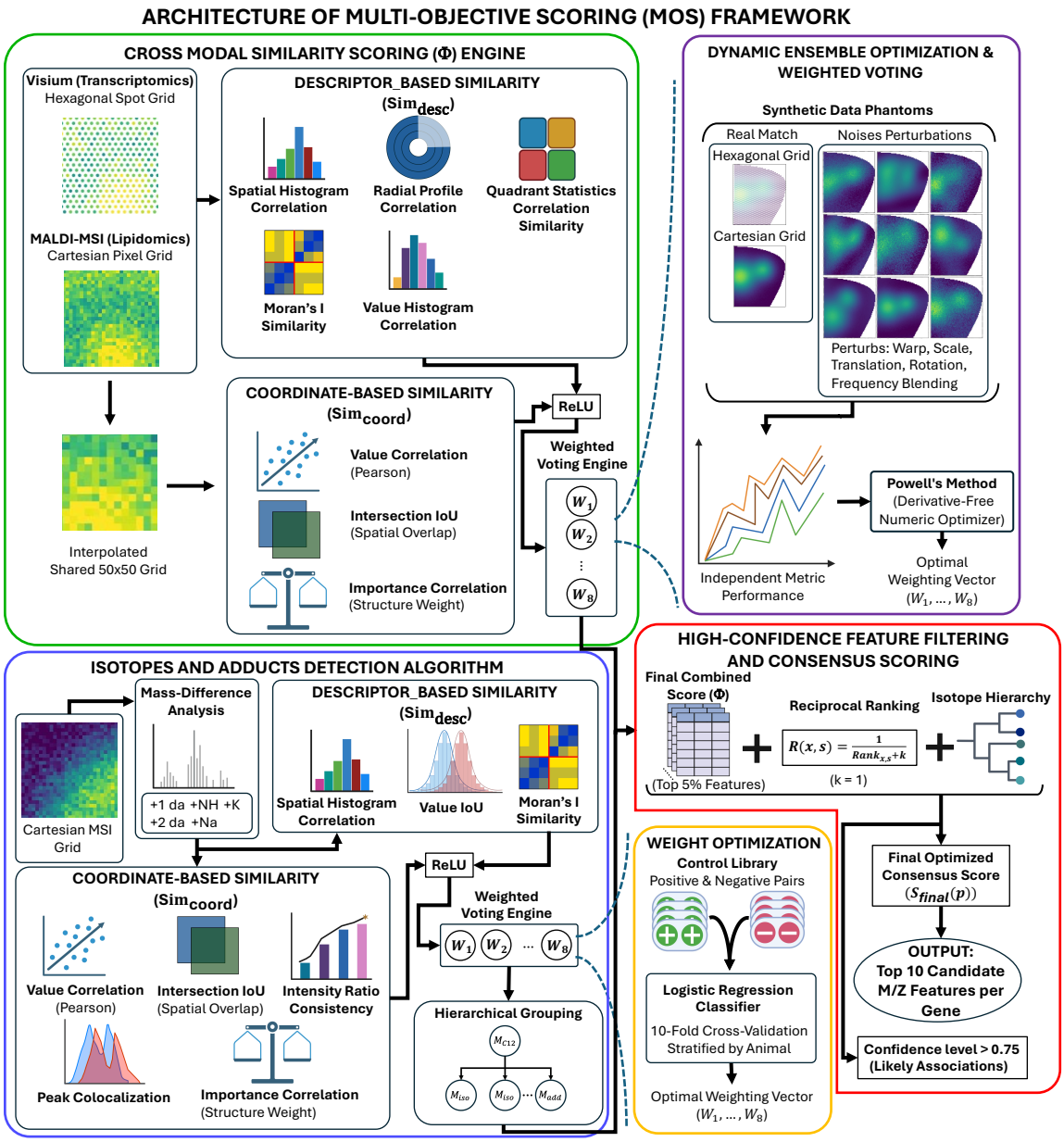


Figure 5-6 Schematic diagram of the MOS framework[6]

[6] https://www.biorender.com

the tissue. Signals that increase and decrease in the same spatial regions produce higher similarity scores, while negative correlations are treated as non-similar relationships.

b) Importance Intersection over Union (IoU)

The importance map was computed for each signal by combining normalized intensity with local spatial variance among neighboring pixels. Two metrics were derived from these maps: Importance Intersection-over-Union and Importance Correlation. Let I_G and I_M represent normalized biological importance maps. Spatial overlap is defined as:

$$\mathrm{IoU}(I_G, I_M) = \frac{\sum_i \min\,(I_{G,i}, I_{M,i})}{\sum_i \max\,(I_{G,i}, I_{M,i})}$$

$$\mathrm{IoU}_{\mathrm{imp}} = \mathrm{IoU}(I_G, I_M)$$

This metric quantifies the degree of spatial overlap of high-importance regions.

c) Importance Correlation

Pearson's correlation coefficient is used to measure spatial similarity between the importance maps:

$$r_{imp} = r(I_G, I_M)$$

which evaluates the linear relationship in biologically weighted spatial structure. Together, these metrics quantify **pixel-level spatial correspondence** after mild alignment and interpolation.

5.2.4.1.1.2 Descriptor-Based Similarity (Sim_{desc})

To capture structural similarity independent of exact pixel-to-pixel alignment, each spatial map is converted into a set of global descriptors. These descriptors summarize its distributional characteristics, geometric properties, and spatial autocorrelation patterns.

a) Spatial Histogram Correlation

Let H_G and H_M represent normalized 2D spatial histograms. Similarity is defined as:

$$r_{spat} = r(H_G, H_M)$$

This metric compares coarse spatial distributions by dividing the tissue into grid regions and evaluating whether both signals show similar large-scale intensity patterns across those regions.

b) Radial Profile Correlation

Let R_G and R_M represent radial intensity profiles computed from the centroid outward. Similarity is defined as:

$$r_{rad} = r(R_G, R_M)$$

This metric measures how intensity changes from the center toward the periphery, enabling comparison of overall gradient patterns between the two maps.

c) Quadrant Statistics Correlation

The spatial domain is divided into directional subregions. Let Q_G and Q_M represent the concatenated quadrant-wise summary statistics from each map. Similarity is then computed as:

$$r_{quad} = r(Q_G, Q_M)$$

This metric assesses the degree of directional symmetry and regional heterogeneity shared between the two spatial patterns.

d) Moran's I Similarity

Spatial autocorrelation is measured using Moran's I [54], [55]:

$$I = \frac{n}{W} \frac{\sum_i \sum_j w_{ij} (x_i - \bar{x})(x_j - \bar{x})}{\sum_i (x_i - \bar{x})^2}$$

Where w_{ij} represents the spatial weights between locations i and j, and $W = \sum_{i,j} w_{ij}$ is the total sum of weights.

In this implementation, the weight matrix W is constructed using a k-nearest neighbor (kNN) graph based on normalized spatial coordinates. The value of k was set to 6 for gene expression data and 8 for m/z data. This approach helps preserve local spatial structure while maintaining robustness to variations in sampling density.

Similarity between gene and m/z spatial autocorrelation is defined as:

$$\text{Sim}_{\text{Moran}} = 1 - \frac{| I_G - I_M |}{2}$$

This formulation guarantees that the similarity score remains bounded within the interval [0,1]. Moran's I spatial autocorrelation measures the tendency of neighboring pixels to exhibit similar intensities. The similarity between two signals is then defined by how closely their spatial autocorrelation values match.

e) Value Histogram Correlation

Let V_G and V_M represent normalized intensity histograms. Structural similarity in signal distribution is quantified via:

$$r_{hist} = r(V_G, V_M)$$

This metric captures how closely the two distributions are similar in terms of their overall shape and dynamic range, independent of spatial location.

5.2.4.1.2 Final Scoring Equation

The final combined score (Φ) is calculated as the sum of the weighted coordinate and descriptor scores. To ensure the model favors positive associations, correlation values are rectified using a $\max(x\ , 0)$ function where appropriate.

The total score Φ is defined as:

$$\Phi = Score_{coord} + Score_{desc}$$

Breaking this down into the specific weights (w_1 through w_8) used in the pipeline:

$$Score_{coord} = w_1.\max(r_{val}, 0) + w_2.IoU_{imp} + w_3.\max\left(r_{imp}, 0\right)$$

$$Score_{desc} = w_4.\max\left(r_{spat}, 0\right) + w_5.\max(r_{rad}, 0) + w_6.\max\left(r_{quad}, 0\right) + w_7.Sim_{moran} + w_8.\max(r_{hist}, 0)$$

The weights were assigned to prioritize direct spatial overlap ($Score_{coord}$) as the primary signal, while using the descriptor ensemble ($Score_{desc}$) to validate the structural integrity of the match. The weights were calculated based on an optimization algorithm using synthetic data.

5.2.4.2 Synthetic data

To evaluate the performance of the final combined score (Φ) framework and optimize its weight assignments (w_1 through w_8), a generative framework was developed to produce synthetic "digital phantoms." These datasets are designed to replicate the biological variability and technical challenges of the spatial transcriptomics (RNA) and MSI. The simulation begins by defining a reproducible tissue morphology modeled after a murine brain hemisphere, constructed using a composite masking function that combines trigonometric and elliptical curves. To capture the physical acquisition differences between modalities, the tissue region was sampled using two distinct coordinate systems: a 100 μm hexagonal lattice for transcriptomic spots and a 60 μm Cartesian grid for mass spectrometry pixels. This variation in sampling density and geometry ensures that the synthetic data realistically captures the non-aligned nature of real-world spatial omics experiments.

Within this defined tissue region, the framework constructed 50 distinct spatial archetypes, ranging from simple linear gradients and periodic wave patterns to more biologically inspired structures such as cortical layers, which serve as the ground-truth (GT) signals. Each GT gene pattern is matched with a mathematically identical GT lipid pattern, establishing a known 1.0 correlation for validation purposes. To rigorously test the robustness of the scoring engine, a non-rigid distortion engine generates 10 "decoy" noise variants for every pattern. These decoys are produced through five independent spatial perturbations: sinusoidal coordinate warp (shape deformation), anisotropic

scaling, center translation, rotational shifts, and frequency blending. The severity of each perturbation is set to 8 on a user-defined 1–20 scale. This controlled setup enables a systematic evaluation of the algorithm's ability to distinguish true cross-modal biological associations from spatially coincident noise, providing the positive and negative control pairs required for the weight optimization process.

5.2.4.3 Independent Metric Performance and Rotational Accuracy

The framework evaluated the robustness of the final combined score (Φ) engine by measuring the performance of each spatial metric separately across all ground-truth (GT) patterns and their associated orientations. In this process, gene expression data was systematically rotated, and the corresponding m/z matches were identified. The start, stop, and step angles are user-defined based on the orientation of the real data. For example, the tissue may be slightly rotated between consecutive slices during mounting causing an angular shift between MSI and Visium slices. Based on our samples, the start, stop, and step angles were set to 354°, 6° (±6°), and 2°, respectively. For every combination of gene expression, sample, and rotation angle, each of the eight scoring metrics – such as Value Correlation, Moran's I Similarity, and Radial Profile Correlation – is treated as an independent classifier. By isolating these metrics, the analysis can determine which specific spatial descriptors are most resilient to the five types of spatial perturbations (warp, scale, translation, rotation, and frequency blending) introduced by the distortion engine.

At each orientation, the algorithm selects the m/z feature with the highest score for a given metric. A match is considered correct only if the top-scoring m/z corresponds to the pre-defined ground-truth lipid paired with the gene expression pattern. This approach makes it possible to generate detailed accuracy profiles for each metric, showing, for example, whether the Radial Profile metric better captures concentric ring structures, while Moran's I performs more effectively for clustered "blob" patterns. The framework recorded these accuracies for every gene expression (phantom), sample, rotation, and scoring metric, and these results were subsequently used to inform and refine the weight assignments.

5.2.4.4 Dynamic Ensemble Optimization and Weighted Voting

To maximize the predictive performance of the final combined score (Φ), a dynamic ensemble optimization strategy was implemented to determine the optimal weighting vector (w_1, …, w_8). Instead of relying on fixed weights, the framework employs a derivative-free numerical optimizer (Powell's method [56], [57]) to maximize a global accuracy objective function within defined rotational windows. This approach allowed the model to adaptively prioritize different spatial descriptors based on their empirical performance within specific angular tolerances.

5.2.4.5 Isotopes and Adducts Detection Algorithm

In many MALDI-MSI studies, isotopic peaks and adduct forms are removed during preprocessing to reduce redundancy in downstream analyses. In this framework, however, these related signals were intentionally retained and used as additional evidence

to strengthen the MOS system. Because isotopes and adducts of the same molecule are expected to share highly similar spatial distributions, they provide an internal reference for evaluating the robustness of spatial similarity metrics.

To identify isotopes and adducts, a dedicated algorithm was developed that integrates two complementary components: (i) mass-difference analysis and (ii) spatial pattern similarity assessment. In the mass-difference step, candidate pairs are screened based on established isotopic spacing rules and commonly reported adduct mass shifts described in the literature. In the spatial component, candidate signals are evaluated to determine whether they exhibit consistent spatial organization across the tissue section. For spatial comparison, the same conceptual framework used in the primary MOS pipeline was adopted, with the simplification that all computations were performed directly within a Cartesian grid. In the gene-lipid matching pipeline, spatial alignment is more complex because the two modalities originate from different coordinate systems, requiring hexagonal-to-Cartesian interpolation and alignment that accounts for probable rotational differences. In contrast, isotope and adduct analysis was performed entirely within a single 60 μm Cartesian MSI grid. Because all signals originated from the same tissue section and coordinate system, no geometric transformation or rotational normalization was required. This allowed spatial similarity metrics to be computed directly across corresponding pixels without cross-modality alignment.

To evaluate spatial similarity between two m/z signals, the analysis was divided into two complementary components: coordinate-based similarity and descriptor-based similarity.

Coordinate-based similarity incorporates some metrics from the final combined score, including Value Correlation, Importance IoU, and Importance Correlation. It also uses two new metrics, including Intensity ratio consistency and Peak colocalization. Intensity ratio consistency assesses whether the relative intensity between the two signals stays consistent across spatial locations, as expected for isotopic peaks. Peak colocalization measures the spatial overlap of the highest-intensity pixels of each signal, indicating whether the strongest signal regions coincide within the tissue.

The descriptor-based similarity incorporates Moran's I, Spatial Histogram Correlation – from the final combined score, and value intersection-over-union. The value intersection-over-union metric evaluates the overall similarity of the raw intensity distributions across spatial locations, capturing similarity in signal magnitude independent of spatial structure. Together, these eight complementary metrics captured multiple aspects of spatial similarity expected for isotopes.

The weighting vector (w_1, …, w_8) for the spatial metrics was refined using a supervised machine learning model (Logistic Regression) based on previously validated isotope and adduct relationships. For each sample, a control library consisting of 20 positive control pairs and 18 negative control pairs (lipid signals with clearly distinct spatial distributions) was constructed. The positive pairs were detected based on mass difference, intensity ratio, and overall pattern [58], [59]. The complete list of positive and negative controls is provided in Table 5-3 and Table 5-4. The same control library was applied consistently across all samples to ensure comparability.

Table 5-3 Positive Control Pairs

#	M/Z 1	M/Z 2	#	M/Z 1	M/Z 2
1	760.5856	761.5864	**11**	971.6514	972.6575
2	788.6156	789.6178	**12**	500.3482	501.3524
3	810.5997	811.6054	**13**	484.2952	485.3011
4	798.5392	799.5461	**14**	627.5340	628.5358
5	813.6858	814.6872	**15**	455.2904	456.2954
6	524.3706	525.3731	**16**	496.3390	497.3426
7	522.3557	523.3569	**17**	666.4319	667.4356
8	734.5686	735.5720	**18**	678.4684	679.4734
9	947.6513	948.6564	**19**	703.5740	704.5774
10	560.3341	561.3381	**20**	706.5388	707.5393

Table 5-4 Negative Control Pairs (Spatially Distinct Lipid Signals)

#	M/Z 1	M/Z 2	#	M/Z 1	M/Z 2
1	760.5856	810.5997	**10**	579.5304	592.3947
2	788.6156	798.5951	**11**	759.6355	760.5856
3	810.5997	792.6320	**12**	772.6211	773.5294
4	792.6320	524.3706	**13**	788.6156	797.5895
5	813.6858	522.3557	**14**	840.6410	842.6652
6	524.3706	947.6513	**15**	848.5591	848.6185
7	740.4725	739.5955	**16**	848.5591	848.6907
8	739.4661	740.6021	**17**	868.6629	870.6936
9	629.5422	630.6165	**18**	971.6514	971.6854

In the Logistic Regression classifier, model training employed 10-fold cross-validation, stratified by animal, to ensure that learned weights generalized across biological variability among the YC, YAD, AC, and AAD groups. This design prevented overfitting to tissue-specific spatial characteristics and ensured robustness to inter-animal heterogeneity.

The primary objective of this calibration procedure was to identify metric weights that maximize discrimination between true isotope/adduct pairs and spatially unrelated signals that may appear close in mass. Model performance was evaluated using the standard error (SE) of metric coefficients across folds. Metrics demonstrating consistently low SE values were prioritized due to their stability across tissue samples and biological conditions.

5.2.4.5.1 Isotopic and Adduct Hierarchy Mapping

After determining the weights, the algorithm was employed to identify and merge chemically related m/z features based on their spatial co-localization. Let's M_{C12} denote the most abundant isotope of an ion (e.g. ^{12}C) and $M_{iso/add}$ its other isotopes or adducts (e.g. ^{13}C).

- **Pattern Matching:** The algorithm searched for five predefined mass shifts accounting for isotopes and adducts: +1 da, +2 da, +NH, +Na, and +K.
- **Frequency Filtering:** A candidate relationship was only validated if it surpassed a 60% score threshold in at least 12 out of the 16 samples (>=75% of the cohort).

- **Hierarchical Grouping:** For each M_{C12} m/z feature, the algorithm identified the closest match within a 0.01 Da tolerance. Validated features were organized into a strict M_{C12}-$M_{iso/add}$ hierarchy, limiting each M_{C12} to a maximum of five $M_{iso/add}$ to maintain chemical plausibility.

5.2.4.6 High-Confidence Feature Filtering

The MOS framework was developed to systematically evaluate and rank each m/z feature within the dataset, assigning a performance score to all 528 unique values for every gene under investigation. A key component of the framework is the estimation of a confidence level, which is defined by the frequency with which an m/z feature is detected across the murine cohort. This strategy was designed to balance sensitivity and specificity. Wider inclusion thresholds may increase the risk of false positives and artificially raised confidence estimates, whereas overly restrictive thresholds may exclude true positives and potentially filter out biologically meaningful lipid associations.

Because the optimal threshold depends on factors, such as the number of m/z features, spatial patterns, sample quality, and overall data quality, only the top 5% of features were retained in this study. From the initial pool of 528 unique m/z features ranked in each murine brain sample using the final combined score (Φ), the highest 26 features were selected as primary candidates for the gene-to-lipid matching process.

5.2.4.7 Consensus Scoring and Chemical Synergy

For each gene, the retained top 5% of m/z features ranked using the final combined score (Φ) were further evaluated to determine a consensus ranking across all samples. To

integrate evidence across samples while accounting for chemical structure, a hierarchical rank-based scoring framework was developed that combines each M_{C12} m/z feature with its validated isotopes or adducts.

5.2.4.7.1 Rank-Based Contribution

Within each sample, detected m/z features were ranked based on their Φ. To transform rank into a quantitative contribution, we applied a modified Reciprocal Rank formulation:

$$R(x,s) = \frac{1}{\text{Rank}_{x,s} + k}$$

where:

$\text{Rank}_{x,s}$ represents the rank of m/z feature *x* within sample *s*.

The parameter *k* is a smoothing constant that controls how strongly rank influences the score *R(x, s)*. In this study, it was set to 1 to give greater emphasis to higher-ranked features. This formulation ensures that top-ranked features contribute more significantly to the final score, while still allowing a gradual decrease in influence for lower-ranked features.

5.2.4.7.2 Hierarchical Integration of Isotopes and Adducts

Each M_{C12} m/z feature p may have validated isotopes or adducts, represented as $\mathcal{A}(p)$, defined from the determined M_{C12}-$M_{iso/add}$ hierarchy file.

For each sample s, the contribution of the M_{C12} feature, represented as $S_{p,s}$, is calculated as follows:

- *Case 1: M_{C12} m/z detected*

If the M_{C12} m/z feature p is detected in the sample s, its contribution is:

$$S_{p,s} = R(p,s) + \alpha \sum_{a \in \mathcal{A}_s(p)} R(a,s)$$

where:

$\mathcal{A}_s(p) \subseteq \mathcal{A}(p)$ are isotopes or adducts of p detected in the sample s,

The parameter α represents the synergy weight. In this study, α was set to 0.2 to reduce the influence of isotopes and adducts. Lower α ensures that the isotopes and adducts combined contribution does not dominate the overall ranking or mask other high-ranking m/z features within the top 10.

Based on this formulation, the M_{C12} feature contributes fully based on its rank, and co-detected isotopes or adducts provide additional, proportionally weighted reinforcement.

- *Case 2: M_{C12} absent, isotopes/adducts detected*

If the M_{C12} m/z feature is not detected in sample *s*, but one or more of its isotopes or adducts are present, these related signals are treated as supporting chemical evidence. In this case, the per-sample contribution is calculated as:

$$S_{p,s} = \alpha \sum_{a \in \mathcal{A}_s(p)} R(a,s)$$

Accordingly, indirect chemical support contributes to the M_{C12} feature's score, but at a reduced weight to avoid overestimating its importance.

- *Case 3: No detection*

If neither the M_{C12} nor any of its isotopes or adducts is detected:

$$S_{p,s} = 0$$

5.2.4.7.3 Final Optimized Consensus Score

The final optimized consensus score for a M_{C12} m/z feature p across all samples $\mathcal{S}$ is defined as:

$$S_{\text{final}}(p) = \sum_{s \in \mathcal{S}} S_{p,s}$$

Expanded form:

$$S_{\text{final}}(p) = \sum_{s \in \mathcal{S}} \left[\mathbf{1}_{p \in s} \cdot R(p,s) + \alpha \sum_{a \in \mathcal{A}_s(p)} R\,(a,s) \right]$$

Where $\mathbf{1}_{p \in s}$ equals 1 if the M_{C12} m/z is detected in the sample s, and 0 otherwise.

5.2.4.7.4 Confidence Score

The confidence level is defined as the proportion of samples in which either the M_{C12} or at least one of its isotopes/adducts is detected:

$$\text{Confidence(p)} = \frac{|\,\{s \in \mathcal{S} \mid p \in s \text{ or } \mathrm{A}_s(p) \neq \emptyset\}\,|}{|\,\mathcal{S}\,|}$$

The MOS framework reports the top 10 candidate m/z features for each gene based on their final score $S_{\text{final}}(p)$ and assigns a confidence level to each candidate. Candidates with a confidence level exceeding 0.75 are considered more likely to be associated with the expressed gene.

5.3 Results and Discussion

After the preprocessing stage of the MALDI-MSI data, 528 peaks were retained for each of the 16 samples for downstream analysis. Subsequently, two independent optimization

algorithms were applied to determine metric weights for two distinct analytical objectives. The first optimization procedure calibrated the weights for the final combined spatial similarity score (Φ) used in gene-to-MSI pattern matching. The second optimization procedure determined the weights for the isotope and adduct detection algorithm, which integrates spatial similarity metrics within a classification framework. Although both approaches rely on spatial similarity analysis, they employ partially different metrics and descriptor implementations, and they were optimized using separate algorithms tailored to their respective objectives. This independent design ensures fair evaluation and robust performance for each task.

Table 5-5 reports the optimized relative contribution of the eight spatial similarity metrics used in the final combined score (Φ), which integrates coordinate-based overlap and structural descriptor similarity within the rotation-based pattern matching framework. The weights were calculated using synthetic data consisting of 50 structured spatial patterns and 500 noise patterns per each of the 16 samples, ensuring robust calibration under controlled ground-truth conditions. The Powell's method optimization algorithm assigned the highest weights to Value_Hist_Corr (0.1850) and Radial_Corr (0.1814), indicating that similarity in intensity distribution and radial spatial organization provides the strongest evidence for true pattern correspondence. Importance_Correlation (0.1356) and Quadrant_Corr (0.1133) also contribute substantially, highlighting the importance of localized structural agreement across spatial regions. In contrast, Value_Correlation (0.0832) and Morans_Sim (0.0914) received lower weights, suggesting that global

intensity similarity or spatial autocorrelation alone is less informative than the integrated set of structured spatial descriptors. Importantly, the optimized weighted combination achieved an average accuracy of 96.14% in the 354°–6° region, demonstrating that integrating complementary spatial features significantly enhances robustness and discriminative performance. These results confirm that the multi-component formulation of Φ – combining spatial overlap with descriptor validation – provides a reliable and empirically supported strategy for identifying biologically meaningful peak relationships in MALDI-MSI data.

Table 5-6 summarizes the calibrated contribution of eight spatial similarity metrics used to determine whether two MALDI-MSI peaks exhibit similar spatial patterns. These

Table 5-5 Optimized normalized weights of the eight spatial similarity metrics contributing to the final combined scoring function (Φ), derived through data-driven calibration using synthetic training data.

Metric	Optimal Weight
Value_Correlation	0.0832
Importance_IoU	0.1102
Importance_Correlation	0.1356
Value_Hist_Corr	0.1850
Spatial_Hist_Corr	0.1000
Radial_Corr	0.1814
Quadrant_Corr	0.1133
Morans_Sim	0.0914

weights were estimated using a 10-fold cross-validated logistic regression model trained on positive and negative control *m/z* pairs extracted from 16 MSI brain samples.

As described in the Methods section, to ensure robustness and minimize overfitting, model performance was evaluated using 10-fold cross-validation grouped by animal sample. In each fold, data from specific animals were used for training, while data from previously unseen animals were reserved for testing. This animal-level separation ensures that the model generalizes across biological replicates rather than memorizing sample-specific patterns. The resulting mean AUC of 0.9675 demonstrates excellent discrimination between positive and negative peak relationships, confirming that the combined spatial similarity metrics reliably capture biologically meaningful ion co-localization.

The Weight (0–1) column represents the average normalized contribution of each metric across all ten folds. Because the logistic regression coefficients were normalized to sum to one, these values reflect the relative importance of each feature in the final calibrated model. Higher weights indicate greater influence in distinguishing true peak associations. Among the metrics, intensity ratio consistency (0.2231) shows the highest contribution, indicating that consistent intensity relationships between ions across spatial pixels are the strongest indicator of true peak associations. Spatial structure metrics also contribute substantially, particularly importance correlation (0.1948), Moran's similarity (0.1713), and Importance IoU (0.1226), highlighting shared biologically important regions, spatial autocorrelation, and the importance of spatial overlap. Moderate contribution is observed

Table 5-6 Mean normalized weights (± SD and SE) of spatial similarity metrics used in the pattern-similarity model for identifying isotopes and adducts among MALDI-MSI peaks, evaluated using 10-fold cross-validation.

Metric Name	Weight (0–1)	SD	Std. Error (SE)
Intensity Ratio Consistency	0.2231	0.0032	0.0010
Value Correlation	0.0414	0.0049	0.0016
Peak Colocalization	0.0850	0.0069	0.0022
Importance IoU	0.1226	0.0038	0.0012
Importance Correlation	0.1948	0.0030	0.0009
Spatial Histogram Correlation	0.1111	0.0056	0.0018
Value IoU	0.0508	0.0057	0.0018
Moran's Similarity	0.1713	0.0023	0.0007
TOTAL (Sum)	**1.0000**		

for spatial histogram correlation (0.1111), while peak colocalization (0.0850) has a smaller influence. In contrast, value IoU (0.0508) and value correlation (0.0414) contribute minimally, suggesting that global intensity correlation alone is not a strong indicator of true MSI peak relationships compared with spatially informed descriptors. After applying the weights in the isotope and adduct detection algorithm, containing mass-difference analysis and spatial pattern similarity assessment metrics, and using it on all data, 136 peaks were identified as having at least one additional isotope or adduct. Subsequently, MOS introduced 10 candidates for each expressed gene by incorporating

528 m/z peaks for each sample, isotopes/adducts hierarchy, and 8 selected expressed genes (Figure 5-7).

Based on the results (Figure 5-7) and considering the confidence score (>=0.75), the most robust associations were observed for the myelin-related genes *Mbp* and *Plp1* [60], [61]. Several m/z features linked to these genes were detected in nearly all animals, including 630.6165, 870.6936, and 842.6652, each appearing in all 16 samples. The high reproducibility and strong consensus scores of these peaks suggest that they represent stable analyte signatures associated with myelin-rich regions of the brain. Both *Mbp* and *Plp1* encode structural components of the myelin sheath produced by oligodendrocytes [60], [61], and their consistent association with the same analyte features indicates that these peaks may correspond to lipid species involved in myelin structure or maintenance. Interestingly, one of the most recurrent features across multiple genes was m/z 842.6652, which appeared among the top-ranked peaks for *Mbp*, *Plp1*, and the astrocyte-associated gene *S100b* [62]. This peak was detected in all animals for *Mbp* and *Plp1* and in the majority of samples for *S100b*. The S100B protein is widely used as a molecular marker of astrocytes and glial activity in the central nervous system, although it can also be expressed in oligodendrocytes during development [62], [63], [64]. The repeated detection of the same m/z feature across these genes indicates a strong spatial similarity (overlap) between the gene-expression patterns and the analyte distributions across samples, suggesting overlapping molecular signatures in regions enriched with glial cells and myelinated structures.

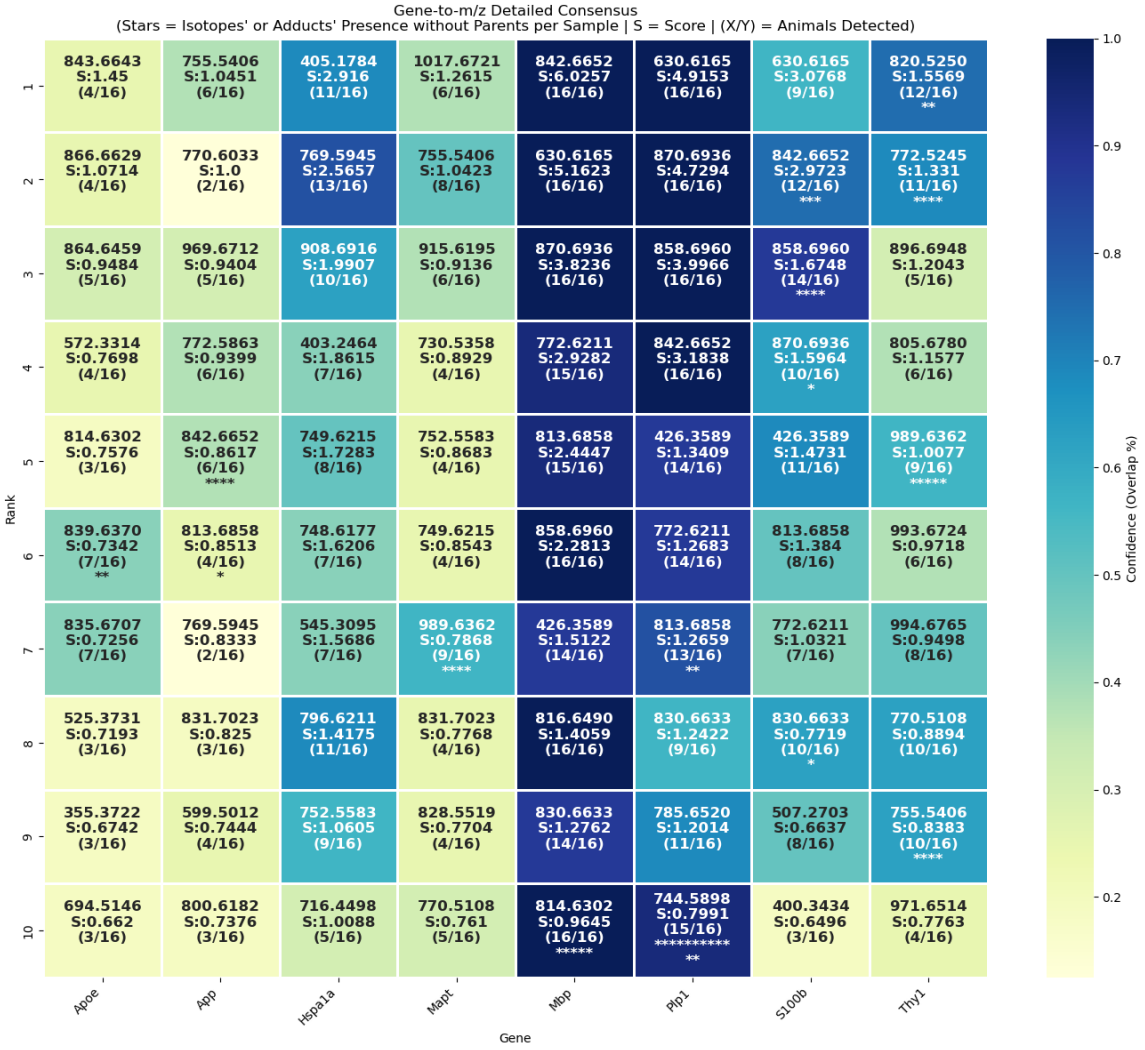


Figure 5-7 Gene–m/z consensus heatmap. Columns represent genes and rows show the top 10 ranked m/z candidates. Color indicates confidence (fraction of animals where the feature was detected in the top 5%). Each cell reports the m/z value, consensus score (S), and detection frequency (X/Y). Asterisks (*) denote samples where isotopic or adduct peaks were detected without the M_{C12} peak.

Figure 5-8 shows the spatial heatmap of *Mbp* expression across all samples. The expression pattern follows the expected anatomical distribution of myelinated white matter tracts in the mouse brain. The corresponding MALDI-MSI distribution of analyte m/z 842.66 (Figure 5-9) displays a highly similar spatial pattern after TIC normalization and removal of background and noise pixels. Both maps show strong signal enrichment along white matter structures and reduced signal in cortical gray matter regions [65]. The close spatial similarity between the transcriptomic signal and the MSI analyte suggests that the MOS framework can successfully identify analyte features associated with myelin-related molecular patterns.

Figure 5-10 further demonstrates this relationship by comparing the spatial distribution of *Mbp* with the top ten candidate m/z features with high confidence level proposed by the MOS framework in a representative sample. All candidate features exhibit spatial distributions that closely resemble the *Mbp* expression map, indicating that the algorithm consistently identifies analytes localized to white matter regions. This result supports the robustness of the MOS system when the gene expression pattern is spatially well-defined and densely sampled. The similarity among multiple candidate analytes may reflect the presence of several lipid species associated with myelin sheaths, which are known to be abundant in white matter and detectable by MALDI-MSI.

Figure 5-11 and Figure 5-12 illustrate a case in which only partial spatial similarity was observed between transcriptomic and MSI data. The spatial heatmap of *Thy1* expression (Figure 5-11) shows enrichment in cortical and hippocampal regions, particularly in AD

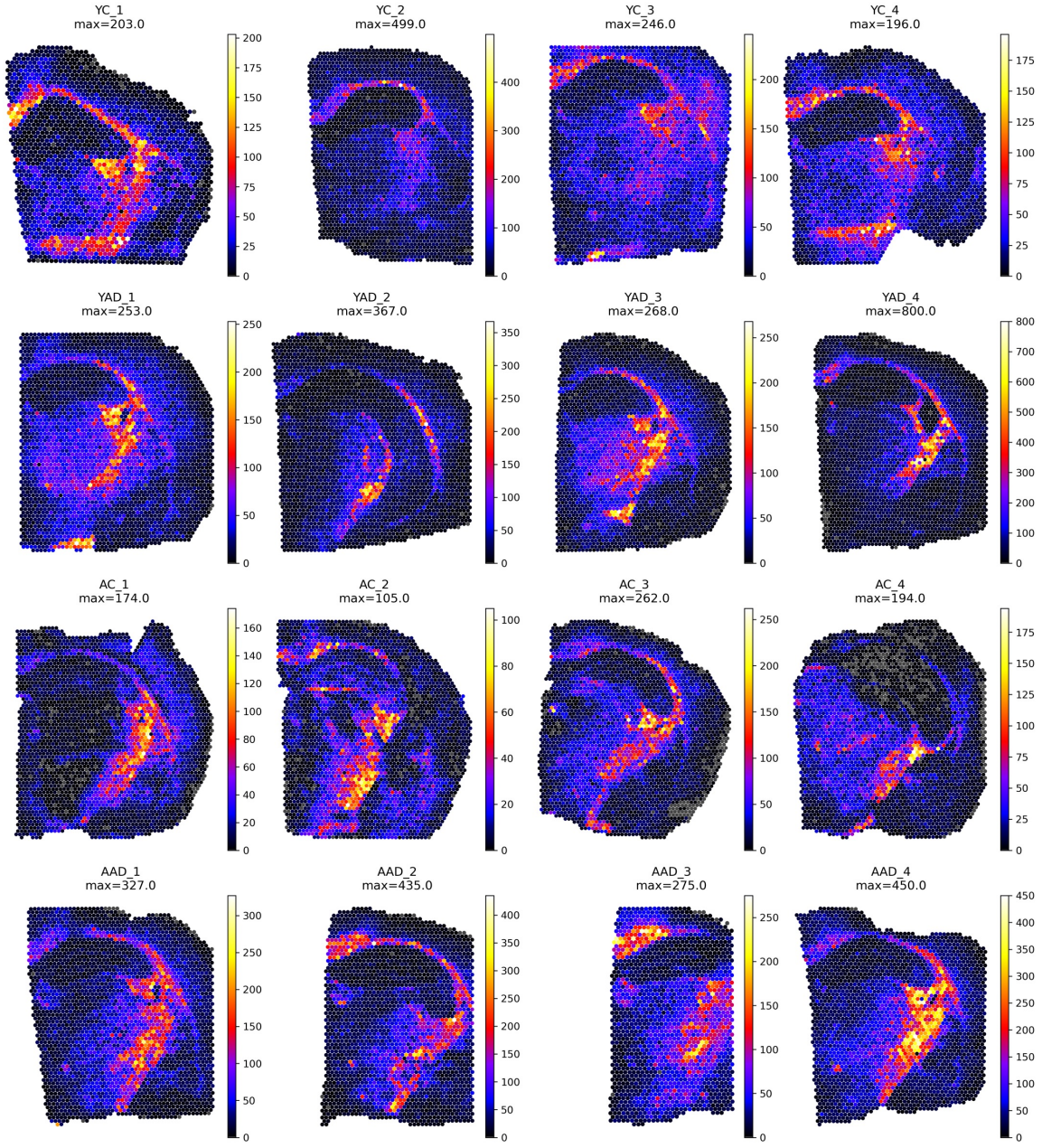


Figure 5-8 Spatial heatmap of *Mbp* expression across all samples; spots enlarged for easier visualization

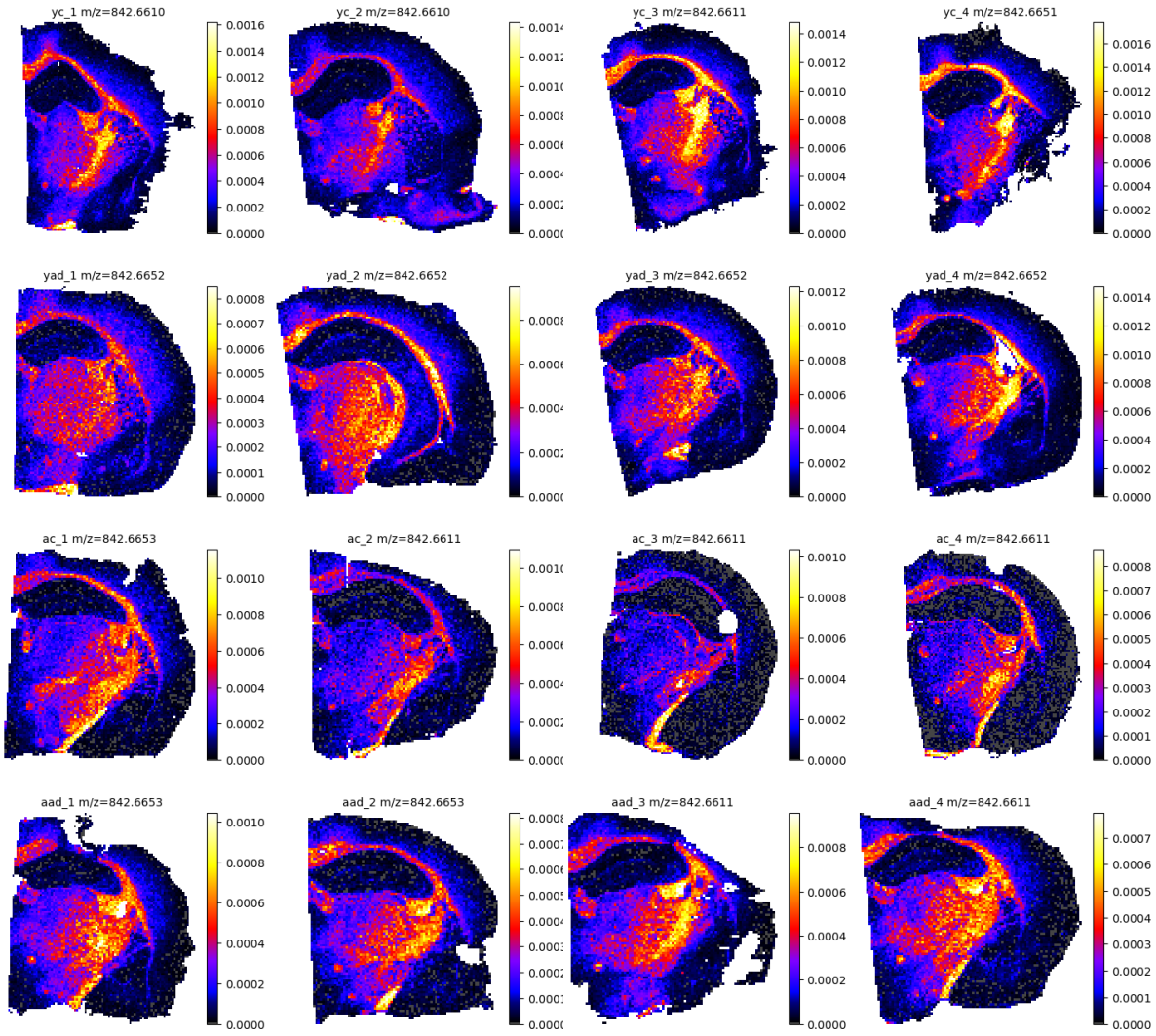


Figure 5-9 TIC-normalized spatial distribution heatmaps of analyte 842.6652 across all half-brain samples with filtering of background, probable contamination, or noise pixels

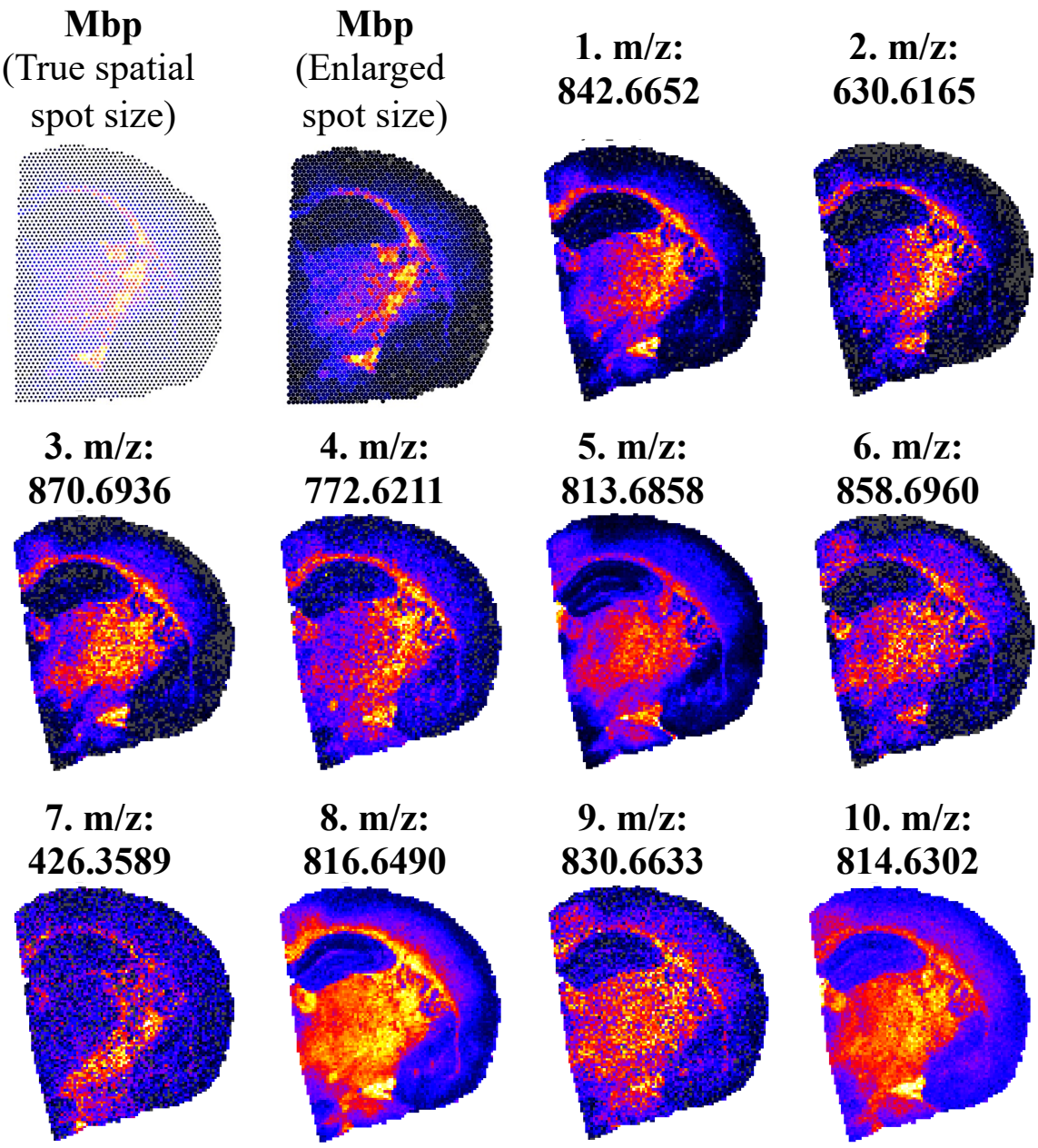


Figure 5-10 Spatial distribution heatmap of *Mbp* across a representative sample, shown with both true spatial spot size and enlarged spots for visualization, alongside the distributions of the 10 matched m/z candidate features identified using the MOS framework. Numbers indicate the ranking of the candidate features.

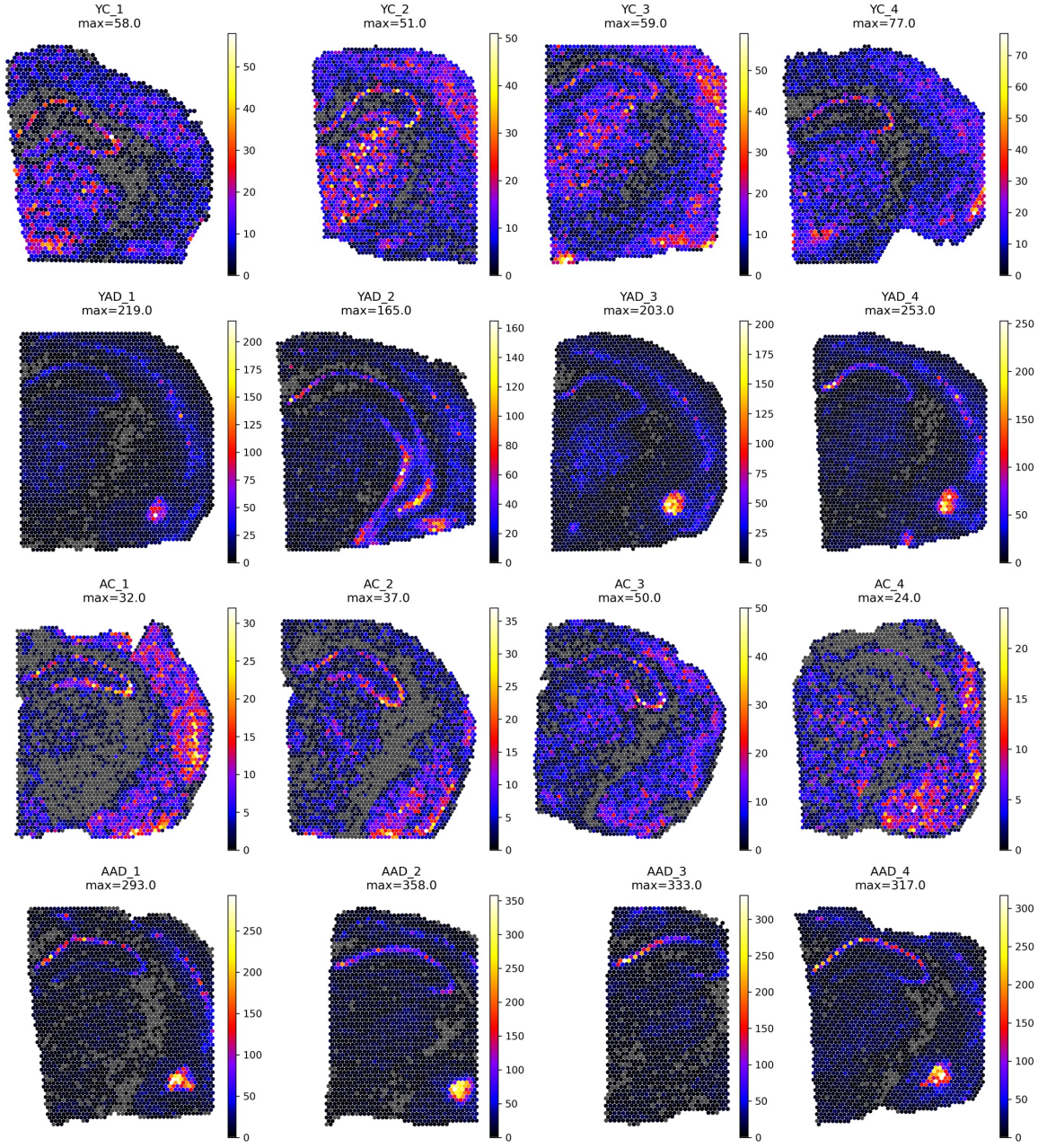


Figure 5-11 Spatial distribution heatmap of *Thy1* expression across all samples; spots enlarged for easier visualization

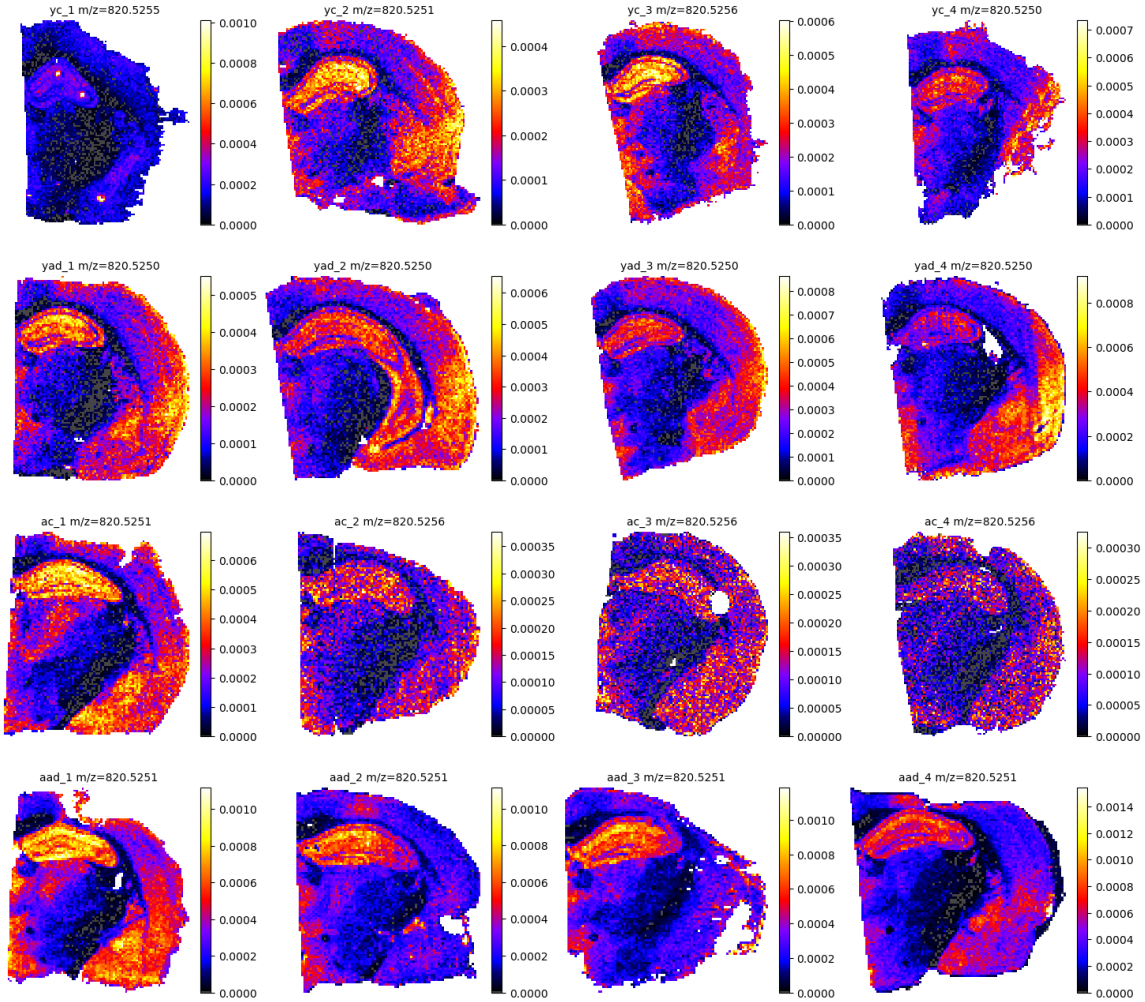


Figure 5-12 TIC-normalized spatial distribution heatmaps of analyte 820.52 across all half-brain samples with filtering of background, probable contamination, or noise pixels

samples, while the corresponding MSI analyte at m/z 820.52 (Figure 5-12) exhibits a broadly similar but imperfect spatial distribution. The discrepancy appears to be largely driven by artifacts at tissue boundaries, as the most prominent signal regions in both spatial maps are located near or directly along the tissue borders.

In both modalities, edge regions are prone to distortions caused by issues during freezing, sectioning, and mounting. Additional inconsistencies in spatial transcriptomics data can

be due to an incomplete capture or misaligned spots at the margins. As a result, some transcriptomic spots may fall partially outside the actual tissue region, while MSI pixels located near tissue edges may experience altered ionization efficiency or matrix-related effects. These factors can introduce spatial inconsistencies that reduce the accuracy of similarity assessment between the gene and analyte maps, as seen in the *Mbp* and *Thy1* cases. In the *Mbp* example, the most informative regions of expression were located closer to the center of the tissue section, where the data are less affected by edge-related distortions. In contrast, the prominent regions of *Thy1* expression and the corresponding MSI signal are positioned closer to the tissue boundaries, making them more susceptible to sample preparation artifacts and spatial misalignment. This difference likely contributes to the lower spatial agreement observed in the *Thy1* case compared with the *Mbp*-associated analytes.

Figure 5-13 presents an example highlighting a limitation of the MOS framework when gene expression data are sparse. The spatial heatmap of *Hspa1a* shows relatively few detected spots across the tissue section, resulting in a highly sparse expression map. Because the MOS framework assumes that available gene expression signals correspond to meaningful spatial information, the limited number of gene spots can artificially increase similarity scores across many MSI features. Consequently, the framework may assign a relatively high score to m/z value, such as the proposed candidate at m/z 769.59, even when a clear spatial correspondence is not evident. This issue suggests that gene quality-control and filtering steps as well as PCA and finding HVGs (highly variable

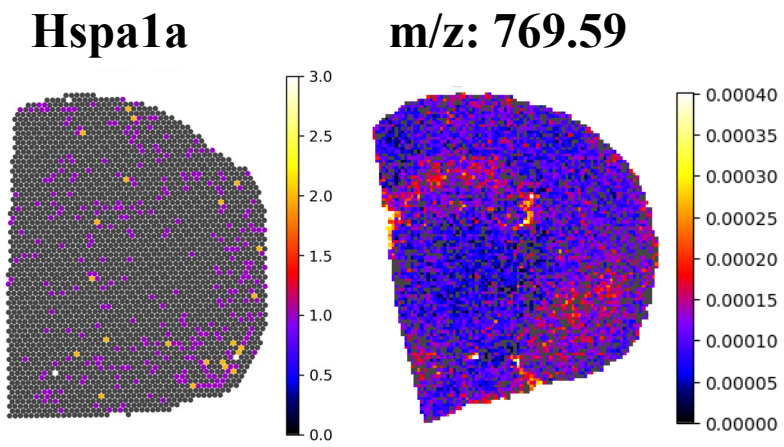


Figure 5-13 Spatial distribution heatmap of *Hspa1a* expression across a representative sample (left), with spots enlarged for clearer visualization, and the spatial distribution heatmap of its proposed candidate analyte feature m/z 769.59 (right), identified using the MOS framework.

genes) should be applied prior to the MOS analysis to remove sparsely expressed genes or low-confidence spots. By implementing such preprocessing steps in the pipeline, genes with insufficient spatial information will likely be excluded and will not appear among the proposed gene–lipid associations.

5.4 Conclusion

Overall, these results demonstrate that the MOS framework can effectively identify MALDI-MSI analyte features whose spatial distributions correspond to transcriptomic patterns. The optimized multi-metric scoring approach showed strong performance in both synthetic validation and experimental data, indicating that combining complementary spatial descriptors improves the reliability of pattern matching. In the biological examples, consistent associations were observed for the myelin-related genes *Mbp* and *Plp1*, whose expression patterns closely matched several analyte features,

particularly m/z 842.6652, suggesting that these peaks may reflect lipid species associated with myelin-rich brain regions.

More broadly, the MOS framework is designed as a hypothesis-generation tool that examines large-scale spatial relationships while preserving structural details. The concept of a consensus score can be extended to other pattern-matching and similarity analyses, and its metric components could be adapted or replaced with alternative statistical, machine-learning, or deep-learning approaches – such as graph neural networks – to further improve precision and robustness in future implementations. In addition, the isotope/adduct detection algorithm and its underlying concept could be further developed by incorporating additional chemical constraints, improved mass-difference modeling, more advanced learning-based classification methods, and real-time integration with online annotation platforms such as METASPACE [66] to enhance the accuracy of peak relationship identification in MALDI-MSI datasets.

5.5 Acknowledgements

The mouse strain used for this research project, B6;129-Tg(APPSwe,tauP301L)1Lfa *Psen1*^{tm1Mpm}/Mmjax, RRID:MMRRC_034830-JAX, was obtained from the Mutant Mouse Resource and Research Center (MMRRC) at The Jackson Laboratory, an NIH-funded strain repository, and was donated to the MMRRC by Frank Laferla, Ph.D., University of California, Irvine, and Mark P. Mattson, Ph.D., Johns Hopkins University, School of Medicine.

Funding for animals was provided by UT-Oak Ridge Innovation Institute (UT-ORII) Science Alliance: Support for Affiliated Research Teams (StART) Program. Amin Jarrahi was supported by UTK ORIED Human Health and Wellness Gateway FY2026 and the ORIED AI Tennessee Initiative Seed Funds.

REFERENCES

1. M. Wess et al., “Spatial integration of multi-omics data from serial sections using the novel Multi-Omics Imaging Integration Toolset,” Gigascience, vol. 14, pp. 1–19, Jan. 2025, doi: 10.1093/gigascience/giaf035.
2. A. Enninful et al., “Integration of imaging-based and sequencing-based spatial omics mapping on the same tissue section via DBiTplus,” Nature Methods 2026, pp. 1–13, Jan. 2026, doi: 10.1038/s41592-025-02948-0.
3. T. F. E. Hendriks, G. B. Eijkel, T. Visvikis, B. Balluff, R. M. A. Heeren, and E. Cuypers, “One section, two worlds: single-cell integration of MALDI-MSI and spatial transcriptomics on the same single tissue section,” Scientific Reports 2025 15:1, vol. 15, no. 1, pp. 42660-, Nov. 2025, doi: 10.1038/s41598-025-26735-1.
4. E. C. Williams et al., “Spatially resolved integrative analysis of transcriptomic and metabolomic changes in tissue injury studies,” Nature Communications 2026 17:1, vol. 17, no. 1, pp. 205-, Jan. 2026, doi: 10.1038/s41467-025-68003-w.
5. X. Liu et al., “Spatial multi-omics: deciphering technological landscape of integration of multi-omics and its applications,” Journal of Hematology & Oncology 2024 17:1, vol. 17, no. 1, pp. 72-, Aug. 2024, doi: 10.1186/s13045-024-01596-9.
6. J. F. Navarro et al., “Spatial Transcriptomics Reveals Genes Associated with Dysregulated Mitochondrial Functions and Stress Signaling in Alzheimer Disease,” iScience, vol. 23, no. 10, Oct. 2020, doi: 10.1016/j.isci.2020.101556.

7. T. C. Gripshover et al., “Visium spatial transcriptomics and proteomics identifies novel hepatic cell populations and transcriptomic signatures of alcohol-associated hepatitis,” Alcohol: Clinical and Experimental Research, vol. 49, no. 1, pp. 106–116, Jan. 2025, doi: 10.1111/acer.15494.
8. B. Heijs, E. A. Tolner, J. V. M. G. Bovée, A. M. J. M. Van Den Maagdenberg, and L. A. McDonnell, “Brain Region-Specific Dynamics of On-Tissue Protein Digestion Using MALDI Mass Spectrometry Imaging,” J. Proteome Res., vol. 14, no. 12, pp. 5348–5354, Dec. 2015, doi: 10.1021/acs.jproteome.5b00849.
9. R. T. Steven and J. Bunch, “Repeat MALDI MS imaging of a single tissue section using multiple matrices and tissue washes,” Analytical and Bioanalytical Chemistry 2013 405:14, vol. 405, no. 14, pp. 4719–4728, Mar. 2013, doi: 10.1007/s00216-013-6899-9.
10. M. Stoeckli, D. Staab, M. Staufenbiel, K. H. Wiederhold, and L. Signor, “Molecular imaging of amyloid β peptides in mouse brain sections using mass spectrometry,” Anal. Biochem., vol. 311, no. 1, pp. 33–39, Dec. 2002, doi: 10.1016/S0003-2697(02)00386-X.
11. S. Ntshangase, S. Mdanda, T. Naicker, H. G. Kruger, S. Baijnath, and T. Govender, “Spatial distribution of elvitegravir and tenofovir in rat brain tissue: Application of matrix-assisted laser desorption/ionization mass spectrometry imaging and liquid chromatography/tandem mass spectrometry,” Rapid

Communications in Mass Spectrometry, vol. 33, no. 21, pp. 1643–1651, Nov. 2019, doi: 10.1002/rcm.8510.

12. Y. Miyake, S. Kusaka, I. Murata, and M. Toyoda, “Matrix-Assisted Laser Desorption/Ionization (MALDI) Mass Spectrometry Imaging of L-4-Phenylalanineboronic Acid (BPA) in a Brain Tumor Model Rat for Boron Neutron Capture Therapy (BNCT),” Mass Spectrometry, vol. 11, no. 1, pp. A0105–A0105, Dec. 2022, doi: 10.5702/massspectrometry.A0105.
13. C. D. Cerruti, F. Benabdellah, O. Laprévote, D. Touboul, and A. Brunelle, “MALDI Imaging and Structural Analysis of Rat Brain Lipid Negative Ions with 9-Aminoacridine Matrix,” Anal. Chem., vol. 84, no. 5, pp. 2164–2171, Mar. 2012, doi: 10.1021/ac2025317.
14. R. Argelaguet et al., “Multi-Omics Factor Analysis—a framework for unsupervised integration of multi-omics data sets,” Molecular Systems Biology 2018 14:6, vol. 14, no. 6, pp. MSB178124-, Jun. 2018, doi: 10.15252/msb.20178124.
15. R. Argelaguet et al., “MOFA+: a statistical framework for comprehensive integration of multi-modal single-cell data,” Genome Biology 2020 21:1, vol. 21, no. 1, pp. 111-, May 2020, doi: 10.1186/s13059-020-02015-1.
16. Y. Zhou, X. Xiao, L. Dong, C. Tang, G. Xiao, and L. Xu, “Cooperative integration of spatially resolved multi-omics data with COSMOS,” Nature

Communications 2024 16:1, vol. 16, no. 1, pp. 27-, Jan. 2025, doi: 10.1038/s41467-024-55204-y.

17. R. M. Mersereau, “The Processing of Hexagonally Sampled Two-Dimensional Signals,” Proceedings of the IEEE, vol. 67, no. 6, pp. 930–949, 1979, doi: 10.1109/PROC.1979.11356.
18. W. E. Snyder, H. Qi, and W. A. Sander, “Coordinate system for hexagonal pixels,” in Proceeding Volume 3661, Medical Imaging, Image Processing, San Diego, CA: Society of Photo-Optical Instrumentation Engineers (SPIE), May 1999, pp. 716–727. doi: 10.1117/12.348629.
19. J. A. Soria Lopez, H. M. González, and G. C. Léger, “Alzheimer’s disease,” Handb. Clin. Neurol., vol. 167, pp. 231–255, Jan. 2019, doi: 10.1016/B978-0-12-804766-8.00013-3.
20. M. W. Bondi, E. C. Edmonds, and D. P. Salmon, “Alzheimer’s Disease: Past, Present, and Future,” Journal of the International Neuropsychological Society, vol. 23, no. 9–10, pp. 818–831, Oct. 2017, doi: 10.1017/S135561771700100X.
21. I. Kaya et al., “Novel Trimodal MALDI Imaging Mass Spectrometry (IMS3) at 10 μm Reveals Spatial Lipid and Peptide Correlates Implicated in Aβ Plaque Pathology in Alzheimer’s Disease,” ACS Chem. Neurosci., vol. 8, no. 12, pp. 2778–2790, Dec. 2017, doi: 10.1021/acschemneuro.7b00314.
22. N. Kakuda et al., “Distinct deposition of amyloid-β species in brains with Alzheimer’s disease pathology visualized with MALDI imaging mass

spectrometry," Acta Neuropathologica Communications 2017 5:1, vol. 5, no. 1, pp. 73-, Oct. 2017, doi: 10.1186/s40478-017-0477-x.

23. S. Oddo et al., "Triple-transgenic model of Alzheimer's Disease with plaques and tangles: Intracellular Aβ and synaptic dysfunction," Neuron, vol. 39, no. 3, pp. 409–421, Jul. 2003, doi: 10.1016/S0896-6273(03)00434-3.
24. A. T. Du et al., "Magnetic resonance imaging of the entorhinal cortex and hippocampus in mild cognitive impairment and Alzheimer's disease," J. Neurol. Neurosurg. Psychiatry, vol. 71, no. 4, pp. 441–447, Oct. 2001, doi: 10.1136/jnnp.71.4.441.
25. D. Wang et al., "Elevated plasma levels of exosomal BACE1 AS combined with the volume and thickness of the right entorhinal cortex may serve as a biomarker for the detection of Alzheimer's disease," Mol. Med. Rep., vol. 22, no. 1, pp. 227–238, Jul. 2020, doi: 10.3892/mmr.2020.11118.
26. E. Schueller et al., "Dysregulation of histone acetylation pathways in hippocampus and frontal cortex of Alzheimer's disease patients," European Neuropsychopharmacology, vol. 33, pp. 101–116, Apr. 2020, doi: 10.1016/j.euroneuro.2020.01.015.
27. M. Dufresne, L. G. Migas, K. V. Djambazova, M. E. Colley, R. Van de Plas, and J. M. Spraggins, "Aminated cinnamic acid analogs as dual polarity matrices for high spatial resolution MALDI imaging mass spectrometry," Anal. Chim. Acta, vol. 1371, no. 23, p. 344423, Oct. 2025, doi: 10.1016/j.aca.2025.344423.

28. A. Jarrahi, A. Jones, W. Tang, H. Qi, and A. C. Crouch, “Mathematical Framework for Quantifying Delocalization in MALDI-MSI via a Composite Scoring Approach,” ACS Measurement Science Au, vol. 6, p. 58, Jan. 2026, doi: 10.1021/acsmeasuresciau.5c00148.

29. A. Savitzky and M. J. E. Golay, “Smoothing and Differentiation of Data by Simplified Least Squares Procedures,” Anal. Chem., vol. 36, no. 8, pp. 1627–1639, Jul. 1964, doi: 10.1021/AC60214A047/ASSET/AC60214A047.FP.PNG_V03.

30. P. H. C. Eilers and H. F. M. Boelens, “Baseline Correction with Asymmetric Least Squares Smoothing,” Leiden University Medical Centre Report, vol. 1, no. 1, p. 5, Oct. 2005.

31. D. C. Castro et al., “Single-Cell and Subcellular Analysis Using Ultrahigh Resolution 21 T MALDI FTICR Mass Spectrometry,” Anal. Chem., vol. 95, no. 17, pp. 6980–6988, May 2023, doi: 10.1021/acs.analchem.3c00393.

32. M. Blank, T. Enzlein, and C. Hopf, “LPS-induced lipid alterations in microglia revealed by MALDI mass spectrometry-based cell fingerprinting in neuroinflammation studies,” Scientific Reports 2022 12:1, vol. 12, no. 1, pp. 2908-, Feb. 2022, doi: 10.1038/s41598-022-06894-1.

33. Z. Hall et al., “MYC expression drives aberrant lipid metabolism in lung cancer,” Cancer Res., vol. 76, no. 16, pp. 4608–4618, Aug. 2016, doi: 10.1158/0008-5472.CAN-15-3403.
34. Y. Otsuka et al., “Improved ion detection sensitivity in mass spectrometry imaging using tapping-mode scanning probe electrospray ionization to visualize localized lipids in mouse testes,” Analytical and Bioanalytical Chemistry 2024 417:2, vol. 417, no. 2, pp. 275–286, Nov. 2024, doi: 10.1007/s00216-024-05641-x.
35. A. Shenfeld and A. Galkin, “Role of the MBP protein in myelin formation and degradation in the brain,” Biological Communications, vol. 67, no. 2, pp. 127–138, Jun. 2022, doi: 10.21638/spbu03.2022.206.
36. J. H. Ahn et al., “Age-dependent differences in myelin basic protein expression in the hippocampus of young, adult and aged gerbils,” Lab. Anim. Res., vol. 33, no. 3, p. 237, Jul. 2017, doi: 10.5625/lar.2017.33.3.237.
37. C. Dimovasili et al., “Aging compromises oligodendrocyte precursor cell maturation and efficient remyelination in the monkey brain,” Geroscience, vol. 45, no. 1, p. 249, Feb. 2022, doi: 10.1007/s11357-022-00621-4.
38. M. Gadek et al., “Aging activates escape of the silent X chromosome in the female mouse hippocampus,” Science Advances , vol. 11, no. 10, p. 8169, Mar. 2025, doi: 10.1126/sciadv.ads8169.

39. J. Wang et al., “S100B gene polymorphisms are associated with the S100B level and Alzheimer’s disease risk by altering the miRNA binding capacity,” Aging, vol. 13, no. 10, pp. 13954–13967, May 2021, doi: 10.18632/aging.203005.

40. [40] N. V. Bobkova et al., “Exogenous Hsp70 delays senescence and improves cognitive function in aging mice,” Proc. Natl. Acad. Sci. U. S. A., vol. 112, no. 52, pp. 16006–16011, Dec. 2015, doi: 10.1073/pnas.1516131112.

41. M. Ximerakis et al., “Heterochronic parabiosis reprograms the mouse brain transcriptome by shifting aging signatures in multiple cell types,” Nature Aging 2023 3:3, vol. 3, no. 3, pp. 327–345, Mar. 2023, doi: 10.1038/s43587-023-00373-6.

42. M. Babić Leko et al., “Further validation of the association between MAPT haplotype-tagging polymorphisms and Alzheimer’s disease: neuropsychological tests, cerebrospinal fluid biomarkers, and APOE genotype,” Front. Mol. Neurosci., vol. 17, p. 1456670, Sep. 2024, doi: 10.3389/fnmol.2024.1456670.

43. C. C. Zhang, A. Xing, M. S. Tan, L. Tan, and J. T. Yu, “The Role of MAPT in Neurodegenerative Diseases: Genetics, Mechanisms and Therapy,” Molecular Neurobiology 2015 53:7, vol. 53, no. 7, pp. 4893–4904, Sep. 2015, doi: 10.1007/s12035-015-9415-8.

44. H. Mori et al., “Blood MAPT expression and methylation status in Alzheimer’s disease,” Psychiatry and Clinical Neurosciences Reports, vol. 1, no. 4, p. e65, Dec. 2022, doi: 10.1002/pcn5.65.

45. K. Schindowski et al., “Alzheimer’s Disease-Like Tau Neuropathology Leads to Memory Deficits and Loss of Functional Synapses in a Novel Mutated Tau Transgenic Mouse without Any Motor Deficits,” Am. J. Pathol., vol. 169, no. 2, p. 599, 2006, doi: 10.2353/ajpath.2006.060002.
46. N. Bin Abid, M. I. Naseer, and M. O. Kim, “Comparative Gene-Expression Analysis of Alzheimer’s Disease Progression with Aging in Transgenic Mouse Model,” Int. J. Mol. Sci., vol. 20, no. 5, p. 1219, Mar. 2019, doi: 10.3390/ijms20051219.
47. C. Bellenguez et al., “New insights into the genetic etiology of Alzheimer’s disease and related dementias,” Nature Genetics 2022 54:4, vol. 54, no. 4, pp. 412–436, Apr. 2022, doi: 10.1038/s41588-022-01024-z.
48. E. Bellou et al., “Age-dependent effect of APOE and polygenic component on Alzheimer’s disease,” Neurobiol. Aging, vol. 93, p. 69, Sep. 2020, doi: 10.1016/j.neurobiolaging.2020.04.024.
49. A. Labuza et al., “Aging, Rather than Genotype, Is the Principal Contributor to Differential Gene Expression Within Targeted Replacement APOE2, APOE3, and APOE4 Mouse Brain,” Brain Sci., vol. 15, no. 10, p. 1117, Oct. 2025, doi: 10.3390/brainsci15101117.
50. A. C. Raulin, S. V. Doss, Z. A. Trottier, T. C. Ikezu, G. Bu, and C. C. Liu, “ApoE in Alzheimer’s disease: pathophysiology and therapeutic strategies,” Molecular

Neurodegeneration 2022 17:1, vol. 17, no. 1, pp. 72-, Nov. 2022, doi: 10.1186/s13024-022-00574-4.

51. O. Catz and M. B. Lewis, “Exploring distinctiveness, attractiveness and sexual dimorphism in actualized face-spaces,” Vis. cogn., vol. 28, no. 9, pp. 453–469, Oct. 2020, doi: 10.1080/13506285.2020.1797967.
52. [52] T. Valentine, “A Unified Account of the Effects of Distinctiveness, Inversion, and Race in Face Recognition,” The Quarterly Journal of Experimental Psychology Section A, vol. 43, no. 2, pp. 161–204, May 1991, doi: 10.1080/14640749108400966.
53. K. Pearson, “VII. Note on regression and inheritance in the case of two parents,” Proceedings of the Royal Society of London, vol. 58, no. 347–352, pp. 240–242, Dec. 1895, doi: 10.1098/rspl.1895.0041.
54. P. A. P. Moran, “Notes on Continuous Stochastic Phenomena,” Biometrika, vol. 37, no. 1/2, p. 17, Jun. 1950, doi: 10.2307/2332142.
55. L. Anselin, “Local Indicators of Spatial Association—LISA,” Geogr. Anal., vol. 27, no. 2, pp. 93–115, Apr. 1995, doi: 10.1111/j.1538-4632.1995.tb00338.x.
56. M. J. D. Powell, “An efficient method for finding the minimum of a function of several variables without calculating derivatives,” Comput. J., vol. 7, no. 2, pp. 155–162, Jan. 1964, doi: 10.1093/comjnl/7.2.155.

57. P. Virtanen et al., “SciPy 1.0: fundamental algorithms for scientific computing in Python,” Nature Methods 2020 17:3, vol. 17, no. 3, pp. 261–272, Feb. 2020, doi: 10.1038/s41592-019-0686-2.

58. M. Höring et al., “Correction of Isobaric Overlap Resulting from Sodiated Ions in Lipidomics,” Anal. Chem., vol. 92, no. 16, pp. 10966–10970, Aug. 2020, doi: 10.1021/acs.analchem.0c02408.

59. J. Dehairs et al., “LipidQMap - An Open-Source Tool for Quantitative Mass Spectrometry Imaging of Lipids,” bioRxiv, p. 2025.10.15.682422, Oct. 2025, doi: 10.1101/2025.10.15.682422.

60. Y. Miyamoto, T. Torii, A. Tanoue, and J. Yamauchi, “Pelizaeus–Merzbacher disease-associated proteolipid protein 1 inhibits oligodendrocyte precursor cell differentiation via extracellular-signal regulated kinase signaling,” Biochem. Biophys. Res. Commun., vol. 424, no. 2, pp. 262–268, Jul. 2012, doi: 10.1016/j.bbrc.2012.06.101.

61. D. L. Fulton, E. Denarier, H. C. Friedman, W. W. Wasserman, and A. C. Peterson, “Towards resolving the transcription factor network controlling myelin gene expression,” Nucleic Acids Res., vol. 39, no. 18, pp. 7974–7991, Oct. 2011, doi: 10.1093/nar/gkr326.

62. F. Michetti et al., “The S100B story: from biomarker to active factor in neural injury,” J. Neurochem., vol. 148, no. 2, pp. 168–187, Jan. 2019, doi: 10.1111/jnc.14574.

63. J. Du et al., “S100B is selectively expressed by gray matter protoplasmic astrocytes and myelinating oligodendrocytes in the developing CNS,” Mol. Brain, vol. 14, no. 1, p. 154, Dec. 2021, doi: 10.1186/s13041-021-00865-9.
64. J. Steiner et al., “Evidence for a wide extra-astrocytic distribution of S100B in human brain,” BMC Neuroscience 2007 8:1, vol. 8, no. 1, pp. 2-, Jan. 2007, doi: 10.1186/1471-2202-8-2.
65. G. J. Siegel, R. Wayne Albers, S. T. Brady, and D. L. Price, Eds., Basic Neurochemistry: Molecular, Cellular and Medical Aspects, 7th ed. Amsterdam, The Netherlands: Elsevier Academic Press, 2006.
66. D. D. Nguyen et al., “Facilitating imaging mass spectrometry of microbial specialized metabolites with METASPACE,” Metabolites, vol. 11, no. 8, p. 477, Aug. 2021, doi: 10.3390/metabo11080477.

CHAPTER 6:

CONCLUSIONS AND FUTURE DIRECTIONS

6.1 Conclusion

This dissertation presents comprehensive computational frameworks for the analysis and integration of spatial multi-omics data. Although the focus is on lipidomics and spatial transcriptomics modalities, the proposed approaches are broadly applicable to other modalities, either conceptually or through direct implementation. The case study in this work is on aging and Alzheimer's disease. Overall, this research contributes to advancing the field of spatial biology by addressing key methodological challenges in both individual modalities and their integration using computer science and engineering.

First in Chapter 2:, a quantitative approach for assessing analyte delocalization in MALDI-MSI data was developed. The introduction of a composite delocalization score, integrating spatial distribution metrics, provides a robust and scalable method for evaluating analyte reliability. This framework enhances quality control in MSI experiments and establishes a standardized basis for comparing spatial integrity across analytes and samples.

In Chapter 3:, a scalable pipeline for lipidomics analysis was developed to enable efficient preprocessing, dimensionality reduction, clustering, and statistical evaluation of high-dimensional MSI datasets. Application of this pipeline to murine brain tissues revealed spatially distinct lipid patterns associated with aging and Alzheimer's disease, underscoring the biological significance of spatial lipid distributions. Furthermore, this method can be used for hypothesis generation in MSI-based studies.

In Chapter 4:, a spatial transcriptomics workflow was established using Visium data, enabling the identification of region-specific gene expression patterns while preserving spatial context. This analysis provided insights into transcriptional alterations linked to both aging and disease states. Also, it highlights the importance of spatially resolved gene expression profiling.

Finally, in Chapter 5:, a novel MOS framework was introduced for integrating MSI and spatial transcriptomics data without requiring exact spatial co-registration. This approach addresses a fundamental limitation in spatial multi-omics integration by leveraging spatial descriptors rather than strict alignment, enabling the identification of cross-modal molecular relationships.

Collectively, these contributions form an end-to-end computational framework for spatial multi-omics analysis. The methodologies developed in this work not only improve data reliability and interpretability but also provide a foundation for discovering spatially resolved biomarkers and generating biologically meaningful hypotheses in complex diseases such as Alzheimer's disease.

6.2 Future Directions

While this work establishes robust frameworks for spatial single-omics and multi-omics analysis, several opportunities remain to further extend and enhance its capabilities.

First, the delocalization analysis framework can be extended to capture the directionality of spatial dispersion, enabling the distinction between asymmetric (sided) and symmetric delocalization patterns. Incorporating directional metrics would facilitate the

identification of potential sources of delocalization – such as artifacts arising from tissue preparation or matrix application – while also enhancing biological interpretability and providing deeper insight into spatial molecular dynamics.

Additionally, the preprocessing pipeline for MSI data can be improved by incorporating isotope and adduct detection algorithms. These approaches would enable identification of overlapping spectral features, allowing either intensity correction to recover underlying peaks or removal of confounding signals. Additionally, integration with real-time online annotation platforms could facilitate automated peak annotation. These improvements together may enhance statistical power, improve adjusted p-values (e.g., FDR correction), and provide more reliable data for downstream hypothesis generation.

The computational pipeline for MSI analysis can further be enhanced by the integration of modern machine learning and deep learning models. Models such as graph neural networks and convolutional neural networks could be more effective in the identification of complex spatial patterns and heterogeneity across samples and experimental groups.

Finally, the MOS framework can be enhanced by employing advanced machine learning models for cross-modal integration. Approaches such as graph neural networks, spatially aware autoencoders, and multimodal transformers may perform better in capturing nonlinear relationships between lipidomic and transcriptomic data while preserving spatial structure.

Future work may also focus on extending the framework to additional omics layers, including proteomics and metabolomics. Integrating multiple molecular modalities would

enable a more comprehensive characterization of biological systems and support the construction of fully integrated spatial molecular networks.

From a biological perspective, applying these methods to larger and more diverse datasets – particularly human tissues and longitudinal studies – will improve the generalizability of findings. These efforts could facilitate the discovery of important biomarkers and enhance the clinical applicability of spatial multi-omics analyses by supporting their use in diagnosis, prognosis, and therapeutic development.

In summary, future developments will aim to increase the scalability, automation, and biological depth of spatial multi-omics analysis. These advancements have the potential to significantly enhance our understanding of complex diseases by enabling more precise and integrative mapping of molecular processes within tissue architecture.

VITA

Amin Jarrahi was born in Iran. He received his Bachelor of Science degree in Mechanical Engineering from Urmia University in 2012. He then pursued graduate studies in Biomedical Engineering at the Iran University of Science and Technology, where he earned his Master of Science degree in 2015. In 2022, he began his doctoral studies in Biomedical Engineering at the University of Tennessee, Knoxville.

Amin's research focuses on developing computational methods and integrating them with AI and machine learning models for the analysis of spatial multi-omics datasets. His work particularly emphasizes combining spatial lipidomics and spatial transcriptomics to uncover shared spatial patterns across biological modalities. As a case study, these approaches were applied to investigate molecular changes associated with aging and Alzheimer's disease.

During his doctoral studies, Amin has served as both a Graduate Research Assistant and a Graduate Teaching Assistant in the Department of Biomedical Engineering at the University of Tennessee. In his research role, he has developed both conceptual and computational tools for spatial and cross-modal pattern analysis to detect overlapping molecular features and has contributed to interdisciplinary projects involving large-scale biomedical datasets. His broader research interests include biomedical data science, AI and machine learning for healthcare applications, and the development of computational tools to advance precision medicine.