

Titel

Interpretability Begins Before Modeling: Background-Aware Curation for Spatial Metabolomics

Engelsk titel

Interpretability Begins Before Modeling: Background-Aware Curation for Spatial Metabolomics

Forfatterens navn

Gorm Hjorth Lütken

Valg af intern bedømmer

Palle Villesen

Indeholder besvarelsen fortroligt materiale?

Nej

Må besvarelsen offentliggøres?

Ja

Sideantal / anslag

26.78 / 64,274

Uddannelsesnavn

Cand.Scient Bioinformatics

Afleverings deadline

12/06/2026 kl 13:00

Abstract

Spatial metabolomics enables spatially resolved molecular profiling of complex biological and forensic samples, but its interpretability is often limited by background contamination, instrumental artifacts, and extreme feature sparsity.

This thesis introduces the DESI/NMF workflow, a structured analysis pipeline designed to transform raw spatial MSI data into curated feature matrices and interpretable latent representations. The approach integrates background-aware preprocessing with unsupervised decomposition, comparing spatially guided background selection, cross-sample artifact identification, and targeted feature filtering prior to NMF. Spatial diagnostics, including Moran's I, are used to evaluate the spatial coherence of latent components.

Results demonstrate that background-aware filtering substantially improves the structure of MSI data, leading to increased spatial coherence of components, reduced dominance of diffuse background signals, and improved separation of overlapping fingerprint patterns, as well as a halving the runtime. Supervised, spatially informed filtering consistently produces components that are less noisy and more easily interpretable than fully unsupervised approaches which in turn improve similarly on unfiltered data.

These findings show that interpretability in spatial metabolomics, defined here as the ability to recover spatially coherent and chemically meaningful patterns, is strongly influenced by preprocessing decisions made prior to model application. The proposed framework highlights the importance of integrating data curation with downstream analysis and provides a modular approach that is transferable to other spatial omics modalities, including MALDI-MSI and spatial transcriptomics.

Contents

Abstract.....	2
1. Introduction	5
2. Methods	9
2.1 Overview of the Analytical Workflow	9
2.2 Data Representation and Spatial Reconstruction	10
2.3 Data Conversion and Ingestion.....	11
2.4 Interactive Background Region Selection	11
2.5 Cross-Sample Shared-Compound Identification	11
2.6 Filtering and Construction of Curated Matrices.....	12
Step 1: Low-signal channel exclusion	12
Step 2: Background filtering (alternating strategies).....	12
2.7 Non-Negative Matrix Factorization (NMF).....	13
2.8 Autoencoders as Alternative Decomposition Methods	13
2.9 Spatial Diagnostics and Component Ranking	15
2.10 Output.....	15
2.11 Code availability	15
3. Results	16
3.1 Filtering Effects on Channel Space	16
3.2 Computational Impact of Filtering.....	20
3.3 Comparison of Filtering Strategies	21
3.4 NMF Results on Raw and Filtered Data	22
3.5 Spatial Coherence and Moran's I Analysis	24
3.6 Impact on Fingerprint Separation Quality.....	26
3.7 Summary of Results	27
4. Discussion and Conclusion	28
4.1 Interpretation of Results and Methodological Implications	28
4.2 Relationship with Existing Work	29
4.3 Practical Implications for Forensic Analysis	29
4.4 Generalization to Other Spatial Omics Data	30
4.5 Limitations and Failure Modes.....	30

4.6 Future Directions31
4.7 Conclusion..... 31
References 33