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