

Multimomics Tissue Segmentation via Spatially-Informed Nested Biclustering Methods

Francesco Denti^{*1}, Cecilia Balocchi², Vanna Denti³, and Giulia Capitoli^{4,5}

¹Department of Statistical Sciences, University of Padova

²School of Mathematics, University of Edinburgh

³Proteomics and Metabolomics Unit, Department of Medicine and Surgery, University of Milano-Bicocca

⁴Bicocca Bioinformatics Biostatistics and Bioimaging B4 Center, Department of Medicine and Surgery, University of Milano-Bicocca

⁵Biostatistics and Clinical Epidemiology, Fondazione IRCCS San Gerardo dei Tintori

Matrix-Assisted Laser Desorption/Ionization Mass Spectrometry Imaging (MSI) is a powerful technique for spatially resolved molecular profiling and cancer biomarker discovery. Recent advances, including a novel multimomics workflow, enable multiple rounds of MSI on the same tissue section, extracting diverse molecular classes, such as lipids, N-glycans, and peptides, while preserving spatial resolution. This innovation is particularly valuable for studies with limited tissue sample availability, such as rare diseases or small tumors. However, the resulting data are high-dimensional and spatially structured, with the various molecular types sharing the same pixel grid. To address these challenges, we propose Poseidon, a Bayesian nonparametric nested biclustering model that simultaneously segments the common tissue pixels and clusters molecular signals within each molecular class, leveraging the shared spatial structure. A separately exchangeable framework is first considered, and then extended to handle spatial data via Markov random fields. For scalability, we implement an efficient mean-field variational inference algorithm tailored for multi-dataset analysis. After validating the efficacy of our method on simulated scenarios, we demonstrate its applicability in a real-world case study, where multimomics measurements were performed on kidney tissue affected by clear cell renal cell carcinoma. The nested, hierarchical structure of Poseidon, combined with its principled inferential framework, allows the extraction of meaningful biological insights, such as clear tissue segmentation and biomarker detection.

Keywords: *Clear cell renal cell carcinoma; Dirichlet process; Potts model; Shared atoms nested models; Variational inference.*

^{*}Corresponding author: francesco.denti@unipd.it

1 Introduction

Over the last decade, *Matrix-Assisted Laser Desorption/Ionization* (MALDI) Mass Spectrometry Imaging (MSI) has emerged as a key technology for *in situ* molecular profiling and cancer biomarker discovery. The technique measures *mass spectra*, i.e., the molecular abundances of a given molecular class (e.g., lipids) across locations of a biological sample [Rohner et al., 2005]. In MALDI-MSI, a tissue section is first coated with a chemical compound called *matrix*, which isolates a specific class of molecules and prepares them for detection. The slide is then divided into a grid of virtual pixels, typically 10-50 μm in size, each irradiated by a laser beam. The instrument separates the detected molecules - the *analytes* - according to their mass-to-charge ratio (m/z), producing a mass spectrum that records the abundance (intensity) of each m/z signal.

Recent advances in MSI have improved both detection precision and experimental reproducibility, expanding spatial proteomics and enabling finer characterization of tumor microenvironments [Krestensen et al., 2023, Zhang et al., 2024]. A particularly important development was introduced by Denti et al. [2022], who proposed a *multiomics workflow* enabling multiple MSI acquisitions on the *same* tissue section. By sequentially applying and removing different *matrices*, this approach extracts analyte intensities from multiple molecular classes (e.g., lipids, N-glycans, and peptides) while minimizing tissue degradation. As a result, each molecular layer yields a distinct MSI dataset measured over a *common spatial grid*. This strategy is particularly valuable when clinical samples are limited, such as in rare diseases or small tumors, as it maximizes molecular information obtained from a single tissue section.

However, these experimental advances introduce new analytical challenges. Multiomics MSI data are inherently high-dimensional and spatially structured, and the novel *multiomics workflow* further increases complexity by generating nested data structures. The pipeline outputs multiple MSI matrices (one for each molecular class), which share the same spatial grid of pixels (columns), but differ in the analytes they measure (rows). A core objective in this context is tissue segmentation, which involves identifying biologically meaningful subregions of a sample based on their spectral characteristics, to uncover subtle molecular variations that may be missed by morphological inspection alone.

Current approaches typically rely on linear and nonlinear dimensionality reduction techniques to extract simplified representations of MSI data [see Zhvansky et al., 2021]. Building on these representations, clustering methods, most notably k-means and bisecting k-means, remain standard in the field, while probabilistic frameworks such as Gaussian mixture models have also shown promising results [Guo et al., 2019, Prasad et al., 2022]. See also Tuck et al. [2022] for a comprehensive review. Recently, more sophisticated models have been introduced to address the intricate characteristics of omics data, incorporating spatial structure and domain-specific features. These include Bayesian spatial clustering approaches [e.g., Zhao et al., 2021, Sottosanti and Risso, 2023] or graph-based community detection algorithms [e.g., Kim et al., 2022, Hu et al., 2024]. However, most of these methods are designed for single-matrix inputs and therefore fail to capture the shared spatial structure, nested organization, and molecular heterogeneity that arise across different analyte types.

In this paper we propose Poseidon, a novel Bayesian nonparametric (BNP) method for the analysis of multiomics MSI data based on *nested biclustering*, which simultaneously segments spatial pixels and clusters molecular signals within each tissue subregion. Our approach integrates the multiple MSI datasets while explicitly accounting for the shared spatial grid and underlying dependencies across pixels. In particular, Poseidon builds on the concept of separate exchangeability [see Rebaudo et al., 2026, and the references therein], providing a principled foundation for modeling complex nested structures in high-dimensional omics data. This framework naturally induces a two-layer partitioning structure, similar to that discussed in Lee et al. [2013]. Our proposed method contributes in three key ways. First, by jointly clustering pixels and m/z values, the model identifies signal groups with elevated or suppressed

activation, effectively distinguishing biologically meaningful peaks from background noise and enabling the extraction of latent molecular feature representations that support the spatial segmentation of the tissue. Second, it accommodates the nested structure of multiomics MSI data by integrating multiple molecular classes into a unified framework, allowing comparisons across analyte types measured on the same tissue. Finally, to ensure scalability and efficient inference in high-dimensional settings, we implement a tailored variational approximation to the posterior distribution, drawing on recent advances in mean-field variational inference for hierarchical, nested Bayesian models [Blei et al., 2017, D’Angelo and Denti, 2026].

1.1 Clear Cell Renal Cell Carcinoma MALDI-MSI Data

Our application focuses on MALDI-MSI measurements obtained from a human tissue sample diagnosed with clear cell renal cell carcinoma (ccRCC), shown in panel A of Figure 1 [Denti et al., 2026]. The tissue is annotated as follows: the green region indicates the *healthy cortex*, the blue outlines the *capsule*, and the red highlights the *tumoral nodule*. A more detailed segmentation is provided in Figure S.2 of the Supplementary Materials. Renal cell carcinoma is among the ten most frequently diagnosed cancers worldwide, with the clear-cell subtype accounting for roughly 75% of cases and for most related mortality [Hsieh et al., 2017]. Prognosis is often poor due to high recurrence rates and resistance to conventional chemotherapy and radiotherapy. Investigating the phenotypic characteristics of cells within these tumor environments may reveal novel molecular pathways and potential therapeutic strategies [Hsieh et al., 2018].

Using the multiomics pipeline of Denti et al. [2022], mass spectra were acquired for three molecular classes on the same renal tissue section: *lipids*, *N-glycans*, and *tryptic peptides*. These classes were selected for their complementary biological roles. Lipids are major components of cell membranes, provide energy for cell growth, and act as signaling molecules that may contribute to drug resistance [Bian et al., 2021, Fu et al., 2021, Shafiee et al., 2020]. N-glycans are among the most common protein modifications and play a key role in directing proteins to specific cellular compartments [Stanley et al., 2022]. Proteins themselves regulate diverse processes, including cell growth, tissue architecture, and cellular replication [Gobena et al., 2024]. In this protocol, proteins are enzymatically cleaved by trypsin to generate tryptic peptides, enabling more detailed proteomic analysis. Together, these three molecular layers provide a complementary view of tumor biology and can enhance biomarker discovery when analyzed within an integrated spatial framework, as demonstrated in previous spatial multiomic studies [Balluff et al., 2019].

In the preparatory phase, the raw spectra extracted with MALDI underwent routine pre-processing steps to account for physiological, instrumental, and analytical variability. Specifically, the spectra followed the biological five-step protocol described in Capitoli et al. [2025], summarized in Section S.1 of the Supplementary Materials. These pre-processed signals were then stored in three separate matrices, corresponding to lipids, N-glycans, and peptides. In each matrix, columns represent spatial pixels (shared across matrices). At the same time, rows correspond to analyte-specific m/z signals (which differ by molecular class). Each matrix contains approximately 100 rows and over 4,000 columns. Figure 1 provides a visual overview of the data. Panels B, C, and D show the normalized median abundance of each molecular class across the tissue section. These images highlight molecular heterogeneity that is not always apparent from histology alone. Panels F, G, and H illustrate spectral profiles for a single pixel (highlighted in panel E), showing the abundance levels across m/z indices for each molecular class. Note that normalization was performed within each molecular class to enhance visualization, so intensities are not directly comparable across the three panels.

By applying our proposed method to these data, we aim to identify molecular signatures and spatial subregions within the renal tissue that distinguish tumor from non-tumor areas, offering insights into the heterogeneity of ccRCC. Moreover, the biclustering approach aligns with

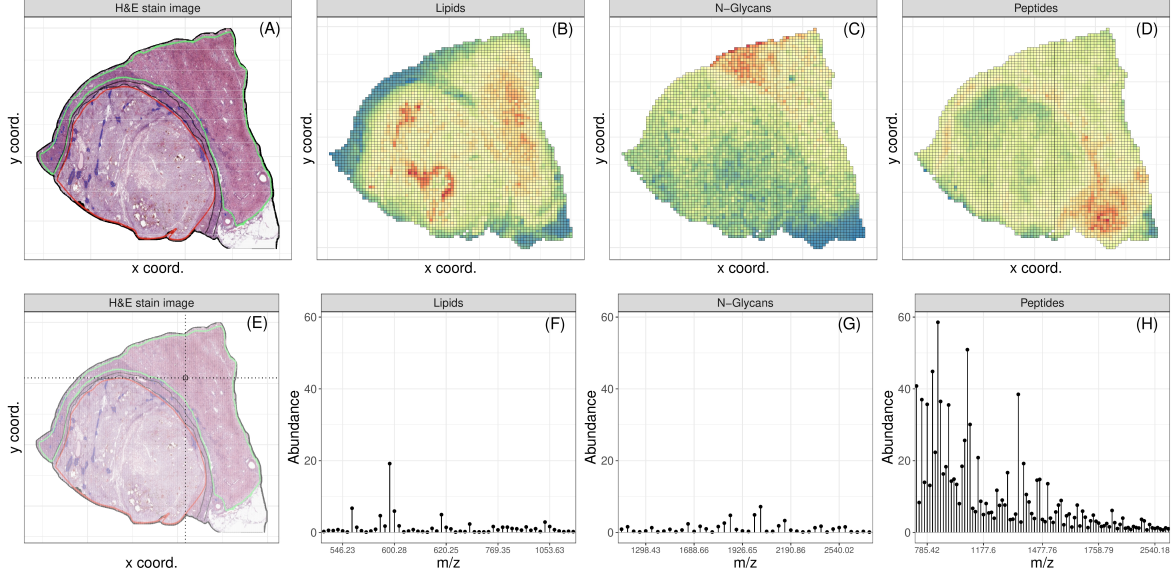


Figure 1: The top row displays the Hematoxylin-and-Eosin stained image of the ccRCC tissue used in the analysis (A) and heatmaps displaying the (normalized) median abundance across the tissue sample for lipids, N-glycans, and peptides (B-D). The bottom row reports, for a specific highlighted pixel (E), abundances of the signals for the different molecular classes (F-H).

our broader goal of reducing data complexity and delivering interpretable, low-dimensional summaries to support pathologists in downstream analyses.

The rest of the paper is organized as follows. In Section 2, we introduce our biclustering model, first outlining the core method for a single molecular layer and then extending it to jointly handle multiple analytes. Section 3 presents the variational inference algorithm used for posterior estimation. Section 4 summarizes the simulation studies reported in Section S.4 of the Supplementary Materials. In Section 5, we apply the model to the ccRCC dataset and analyze the resulting spatial and molecular patterns. Section 6 concludes.

2 A finite-infinite separately exchangeable model for MALDI-MSI data

In line with the data structure outlined in Section 1.1, we consider a collection of T datasets $\mathbf{Y} = \{\mathbf{Y}^{(t)}\}_{t=1}^T$, one for each molecular level. Then, $\mathbf{Y}^{(t)}$ denotes the MALDI-MSI abundance (or transformed abundance) matrix of the t -th molecular class, which is characterized by $N^{(t)}$ rows, representing the specific analytes, and J columns, representing the pixels. We remark that a crucial aspect of the data is that the number of columns is the same across all the datasets. Our goal is to estimate a nested random partition to perform image segmentation, i.e., inducing a *shared partition* of the J pixels driven by the distributional characteristics of the molecular abundance within each location across datasets and, at the same time, inducing an idiosyncratic, *dataset-specific partition* of the signals, different in each dataset. Figure 2 summarizes the objective of our model. In what follows, we begin by discussing our model for the single-dataset case ($T = 1$) and then extend it to the multiple-dataset scenario.

2.1 The single-dataset case: Pose

Let us focus on a single dataset, say $\mathbf{Y}$, with N rows and J columns. From a modeling perspective, we see our data as organized into distinct but related groups defined by the

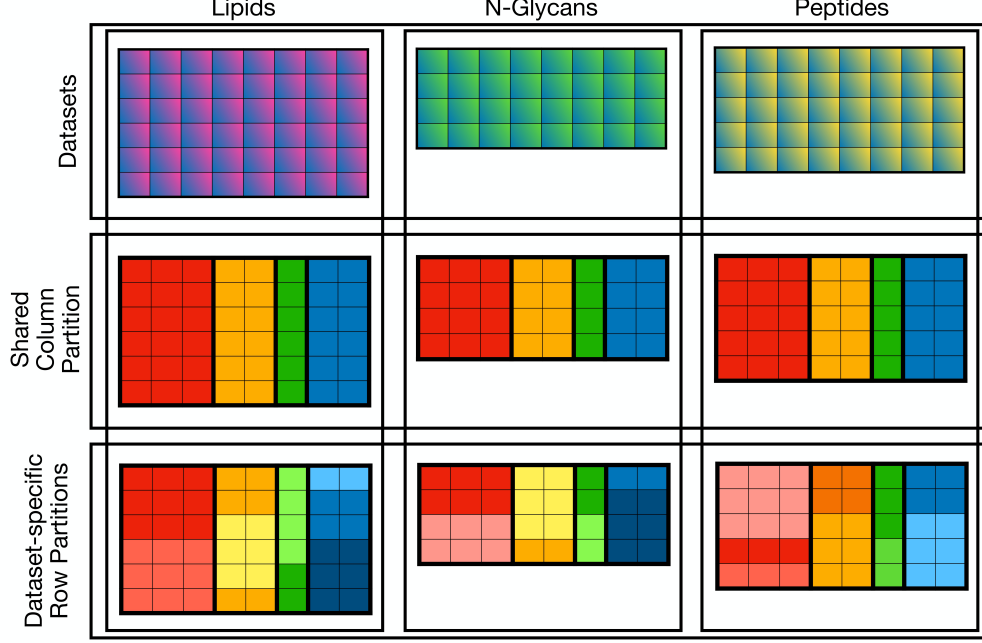


Figure 2: A diagram summarizing the partition structure we aim to recover. The partition over the columns is shared across datasets. The partition over the rows is dataset-specific.

pixels, or columns. This setting is typical of nested designs, an area of research that has thrived in recent years within the Bayesian literature [see, for example, [Rodríguez et al., 2008](#), [Camerlenghi et al., 2019](#), [Denti et al., 2023](#), [Balocchi et al., 2021](#)]. Nested models aim to simultaneously cluster groups and observations; for matrix data, this corresponds to clustering both the columns and the entries. Most of these nested models were proposed under the hypothesis of *partial exchangeability*, where data are assumed to be exchangeable within the different subpopulations. Consequently, the joint probability distribution of observations remains unchanged under any permutation of data points within the same subpopulation. Under this assumption, the probability of clustering matrix entries does not depend on their row identity, and these models completely disregard the information conveyed by the specificity of each row, as [Rebaudo et al. \[2026\]](#) have recently underlined. Indeed, this feature is problematic for us, as each row of our matrix has the same meaning (i.e., it represents the same m/z value) across all the columns. Using a partially exchangeable model in our framework would mean clustering the pixels that have similar abundance distributions by simply considering the proportion of m/z that have high vs low activation levels, regardless of *which* m/z achieve these activation levels. This might lead to an undesirable situation where, for example, two pixels are clustered together because they both have 5% of the signals that are highly activated, even though the high activation is achieved by different sets of m/z in the two pixels, leading to an important loss of information.

To overcome this issue, similarly to [Rebaudo et al. \[2026\]](#), we start by considering a nested model in a *separately exchangeable* setting, so that the identity of each row is preserved across the groups. We will introduce a set of column-specific clusters and, for each of them, a set of row-specific clusters, which are named, respectively, column clusters (CCs) and row clusters (RCs). Specifically, for each column $j = 1, \dots, J$, let C_j denote the CC membership label, with $C_j \in \{1, \dots, K\}$. Given the vector of CC allocations $\mathbf{C} = (C_1, \dots, C_J)$, separate exchangeability implies that in each CC all the entries in a row are restricted to belong to the same RC. In other words, for each row $i = 1, \dots, N$ and *every possible CC* k , we consider a RC membership label $R_{i,k} \in \{1, \dots, L\}$ that is common to all entries (i, j) whose columns

are in the same CC characterized by $C_j = k$. Such RC allocation is considered for all possible CC realizations, with $\mathbf{R}_k = (R_{1,k}, \dots, R_{N,k})$ for all k . The RC assignment variables $R_{i,k}$ are categorical random variables with CC-specific parameters, such that $\mathbb{P}[R_{i,k} = l | \boldsymbol{\omega}_k] = \omega_{l,k}$, where $\boldsymbol{\omega}_k = (\omega_{1,k}, \dots, \omega_{L,k})$ is a vector of probabilities, for all k . In the basic model specification, which we will extend later, we also consider the CC assignment to be independent categorical random variables with probabilities $\boldsymbol{\pi}$, i.e., $\mathbb{P}[C_j = k | \boldsymbol{\pi}] = \pi_k$. Let $f(y | \theta)$ denote a suitable parametric density for the data, and let $\{\theta_l^*\}_{l=1}^L$ be a collection of RC-specific parameters shared across all CCs. The conditional density of entry $y_{i,j}$, given the column-cluster assignment $C_j = k$ and the row-cluster assignment $R_{i,k} = l$, is assumed to be $f(y_{i,j} | \theta_{i,j})$, with $\theta_{i,j} = \theta_{R_{i,C_j}}^*$. Hence, whenever $C_j = k$ and $R_{i,k} = l$, the corresponding parameter satisfies $\theta_{i,j} = \theta_l^*$. Finally, we assume that $\theta_l^* \stackrel{i.i.d.}{\sim} H$, for $l = 1, \dots, L$, where H is a diffuse base measure. Thus, given all the vectors of categorical probabilities $\boldsymbol{\pi}$ and $\{\boldsymbol{\omega}_k\}_{k=1}^K$, we can express our model in terms of the cluster assignment variables $R_{i,k}$ and C_j as

$$\begin{aligned} \theta_l^* &\stackrel{i.i.d.}{\sim} H, \\ C_j | \boldsymbol{\pi} &\sim \text{Cat}_K(\boldsymbol{\pi}), \quad R_{i,k} | \boldsymbol{\omega}_k \sim \text{Cat}_L(\boldsymbol{\omega}_k), \\ y_{i,j} | C_j, R_{i,C_j}, \{\theta_l^*\}_l &\stackrel{i.i.d.}{\sim} f(y_{i,j} | \theta_{R_{i,C_j}}^*). \end{aligned} \tag{1}$$

Note that the matrix entries $y_{i,j}$ associated to the same RC l , i.e., the observations belonging to the set $\{y_{i,j} : R_{i,C_j} = l\}$, are assumed to be exchangeable; in other words, they are modeled as conditionally i.i.d. from $f(\cdot | \theta_l^*)$.

Remark 2.1 *Drawing a parallel with the partially exchangeable nested model, we could describe this model by introducing column-specific random measures G_j , assumed to be distributed according to a discrete random measure $Q = \sum_{k \geq 1} \pi_k \delta_{G_k^*}$. The discreteness of Q induces an implicit clustering on the columns, and we could define $C_j = k$ when $G_j = G_k^*$. For every possible CC k , we could consider row-specific parameters $\theta_{i,k}$, drawn i.i.d. from G_k^* . When the G_k^* 's are also discrete, with $G_k^* = \sum_{l \geq 1} \omega_{l,k} \delta_{\theta_l^*}$, a partition is also induced on the $\theta_{i,k}$ parameters, leading to the definition of the RC assignments $R_{i,k}$.*

Essentially, we cluster together different m/z if they have similar activation levels (a trait captured by the kernel parameters $\{\theta_l^*\}_{l=1}^L$), and we group together different pixels if they have similar activation profiles, as represented by the vector of cluster assignments of the m/z . Note that according to the specification in (1), the distributional law of the data matrix $\mathbf{Y}$ is invariant to permutations of entire rows and columns. Moreover, given the nested nature of the model, the clustering of signals is affected by the CCs, as the cluster assignment of a given signal remains the same for all pixels belonging to the same CC.

Which prior distributions are selected for $\{\boldsymbol{\omega}_k\}_{k=1}^K$ (the *observational weights*) and $\boldsymbol{\pi}$ (the *distributional weights*) represents a crucial modeling choice. In Rebaudo et al. [2026], the authors consider a fully nonparametric approach, allowing both L and K to be infinite, and adopt stick-breaking (SB) Dirichlet process (DP) representations for both weights, inducing a *common atoms* structure. Here, we consider a finite-infinite version of the model, resulting in a *shared atoms* model, following D'Angelo and Denti [2026], and we restrict L to be finite, while K is allowed to be infinite. Specifically, we adopt sparse, symmetric Dirichlet distributions with parameters $\mathbf{b}_0 = b_0 \mathbf{1}_L$ for the row mixture weights, in order to empty useless mixture components. Thus, for all k , $\boldsymbol{\omega}_k \sim \text{Dirichlet}_L(\mathbf{b}_0)$. In the following, we will show how this simple change has a serious impact on the prior probability of coclustering rows. Similarly to Rebaudo et al. [2026], we could assume that $\boldsymbol{\pi}$ is drawn from a SB process with concentration parameter α , denoted with $\text{GEM}(\alpha)$, which is equivalent to setting $\pi_1 = v_1$ and $\pi_k = v_k \prod_{g < k} (1 - v_g)$ where $v_j \stackrel{i.i.d.}{\sim} \text{Beta}(1, \alpha)$ for all $j \geq 1$.

However, the assumption of prior independence in the CC membership labels $\mathbf{C}$, which would be implied under separate exchangeability, could be naive in the context of spatial data. In fact, the columns in our dataset represent pixels of an image, which are likely to be spatially correlated. Therefore, we need to elicit a distribution that allows clustering of the columns, while conveying their spatial information. Specifically, we assume that the clustering assignment for the generic pixel j is influenced by the cluster memberships of its adjacent pixels. This structure can be recovered using discrete-state Markov Random Fields [MRF, Besag, 1986], also known as Potts models when the number of states considered is larger than two. Here, we adopt the BNP-MRF prior introduced by Lü et al. [2020], which extends the classic BNP models for clustering using a MRF. In detail, this distribution assumes that $p(C_j \mid \mathbf{C}_{-j}, \boldsymbol{\pi}, \beta) \propto \pi_{C_j} \cdot \exp(\beta \sum_{q \in \partial_j} \mathbb{1}_{\{C_q = C_j\}})$, where $\boldsymbol{\pi} \sim \text{GEM}(\alpha)$, and ∂_j denotes the set of neighbors of unit j , i.e., the set of pixels adjacent to pixel j . In this work, we employ a second-order neighborhood system (queen adjacency) for each pixel, as it yields a more flexible and realistic characterization of local spatial interactions compared to first-order (rook) adjacency [Besag, 1986]. The quantity β , known as the inverse temperature, governs the level of influence that the neighboring cluster memberships have on a pixel, and can be treated as a fixed hyperparameter or modeled as a random quantity. This conditional distribution can be equivalently stated as a joint distribution for the whole vector of cluster memberships $\mathbf{C}$, as

$$p(\mathbf{C} \mid \boldsymbol{\pi}, \beta) = \frac{1}{\mathcal{K}_{\mathbf{C}}(\boldsymbol{\pi}, \beta)} \left(\prod_{j=1}^J \pi_{C_j} \right) \cdot \exp \left[\beta \sum_{j=1}^J \sum_{q \in \partial_j} \mathbb{1}_{\{C_q = C_j\}} \right], \quad (2)$$

where $\mathcal{K}_{\mathbf{C}}(\boldsymbol{\pi}, \beta)$ is an intractable normalizing constant. Therefore, we redefine the prior distribution for the CC membership $\mathbf{C}$ as (2), and denote it with $\mathbf{C} \mid \boldsymbol{\pi}, \beta \sim \text{BNP-MRF}(\boldsymbol{\pi}, \beta)$.

We refer to the model in (1) with the CC prior redefined as in (2) as the *Potts Over Separate Exchangeability* (Pose) model. Note that under Pose, the model is no longer separately exchangeable, due to the MRF term of the prior for $\mathbf{C}$, and the simple discrete measure representation using Q is no longer possible.

2.1.1 Prior coclustering properties under Pose

A key distinction arises between common atoms and shared atoms models. As discussed in D’Angelo and Denti [2026], combining an infinite set of atoms with a common ordering across random measures and SB priors induces possibly undesired dependencies among the observational weights in partially exchangeable common atoms models. This typically results in high prior coclustering probabilities, which may adversely affect posterior inference. This issue persists even under separate exchangeability (see also Supplementary Section S.4.1). For the Pose model, it is therefore instructive to examine how the prior coclustering probability between two entries, $\mathbb{P}[\theta_{i,j} = \theta_{i',j'}]$, changes when $\boldsymbol{\omega}_k \sim \text{Dirichlet}_L(\mathbf{b}_0)$ (shared atoms), rather than the common atoms induced by a DP. Proposition 1 of Lü et al. [2020] provides the predictive distribution of the BNP-MRF process for general Gibbs-type priors. Specialized to the DP and the case of two observations, we can derive the probability of ties between two groups, equal to $\mathbb{P}[G_j = G_{j'}] = e^\beta / (\alpha + e^\beta)$. We can then prove the following proposition. The proof is deferred to Section S.2.1 of the Supplementary Materials.

Proposition 2.1 *The prior coclustering probability between two entries according to the shared atoms Pose model with $\boldsymbol{\omega}_k \sim \text{Dirichlet}_L(\mathbf{b}_0)$ for all k and $\boldsymbol{\pi} \sim \text{GEM}(\alpha)$ is*

$$\mathbb{P}[\theta_{i,j} = \theta_{i',j'}] = \frac{1}{\alpha + e^\beta} \left[e^\beta \left(\frac{1 + b_0}{1 + Lb_0} \right)^{1 - \mathbb{1}_{i=i'}} + \frac{\alpha}{L} \right].$$

In the same setting, if instead we assume $\omega_k \sim GEM(\nu)$ for all k , we have

$$\mathbb{P}[\theta_{i,j} = \theta_{i',j'}] = \frac{1}{\alpha + e^\beta} \left[e^\beta \left(\frac{1}{1 + \nu} \right)^{1 - \mathbf{1}_{i=i'}} + \frac{\alpha}{2\nu + 1} \right].$$

A graphical depiction of these probabilities is shown in Supplementary Figures S.3-S.5. Consistent with previous findings [D’Angelo and Denti, 2026], the prior coclustering probability is higher under the common atoms model, highlighting its stronger a priori grouping and risk of spurious coclustering. Distinguishing between same-row ($i = i'$) and different-row ($i \neq i'$) entries is important, as when two groups (columns j and j') are clustered, their rows automatically share the same partition. The difference between models is most apparent for different rows, where clustering behavior depends on the observational weight hyperparameter (b_0 for shared atoms, ν for common atoms). For same-row pairs, coclustering under shared atoms is independent of b_0 , but still depends on ν for common atoms. As the spatial regularization parameter β increases, model differences diminish and clustering probabilities converge.

2.2 The multiple-dataset case: Poseidon

Given the definition of Pose for a single molecule dataset, we can simply extend the biclustering framework to multiple datasets. Consider again the collection of T datasets $\mathbf{Y} = \{\mathbf{Y}^{(t)}\}_{t=1}^T$, each characterized by $N^{(t)}$ rows and J columns. We now consider T collections of row partitions $\mathcal{R}^{(t)} = \{\mathbf{R}_k^{(t)}\}_{k=1}^K$ (one for each dataset $t = 1, \dots, T$), with a dataset-specific row partition $\mathbf{R}_k^{(t)}$ for each CC k . We also denote with $\boldsymbol{\theta}^{*(t)}$ the collection of cluster-specific parameters for each dataset $t = 1, \dots, T$, and let $\boldsymbol{\theta}^* = (\boldsymbol{\theta}^{*(1)}, \dots, \boldsymbol{\theta}^{*(T)})$. Finally, we assume the existence of one partition of the columns $\mathbf{C} = (C_1, \dots, C_J)$, shared by all the datasets, resulting in a common image segmentation. Given these partitions, we assume the measurements are conditionally independent across datasets, setting $p(\mathbf{Y} | \{\mathcal{R}^{(t)}\}_{t=1}^T, \mathbf{C}, \boldsymbol{\theta}^*) = \prod_{t=1}^T p(\mathbf{Y}^{(t)} | \mathcal{R}^{(t)}, \mathbf{C}, \boldsymbol{\theta}^{*(t)})$. Then, we can express the model in the following form:

$$\begin{aligned} \theta_l^{*(t)} &\stackrel{i.i.d.}{\sim} H^{(t)}, \quad t = 1, \dots, T \\ \mathbf{C} | \boldsymbol{\pi}, \beta &\sim \text{BNP-MRF}(\boldsymbol{\pi}, \beta), \quad R_{i,k}^{(t)} | \omega_k^{(t)} \sim \text{Cat}_L(\omega_k^{(t)}), \quad t = 1, \dots, T \\ y_{i,j}^{(t)} | C_j, \{R_{i,k}^{(t)}\}_k, \{\theta_l^{*(t)}\}_l &\stackrel{ind.}{\sim} f(y_{i,j}^{(t)} | \theta_{R_{i,C_j}^{(t)}}^{*(t)}), \quad t = 1, \dots, T, \end{aligned} \quad (3)$$

with $\boldsymbol{\pi} \sim GEM(\alpha)$, $\omega_k^{(t)} \sim \text{Dirichlet}_L(\mathbf{b}_0)$, $\forall t, k$, and each $H^{(t)}$ is a diffuse base measure. One can also specify a prior for the concentration parameter of the distributional DP, e.g., $\alpha \sim \text{Gamma}(a_\alpha, b_\alpha)$, and a prior for the inverse temperature, e.g., $\beta \sim \mathcal{U}(0, B)$.

We call this model *Potts Over Separate Exchangeability for the Integration of Different Omics Nested data* (Poseidon). Thanks to the hierarchical structure used in defining Poseidon, the model has the flexibility to model the molecule-specific datasets $\mathbf{Y}^{(t)}$, using dataset-specific parameters $\boldsymbol{\theta}^{*(t)}$; nevertheless, it allows for sharing information across these datasets to infer the shared partition of the columns, leveraging the signal provided by the different molecules (see also Supplementary Section S.4.4).

For our application, we will specialize model (3) to fit a mixture of Gaussian distributions. As the first step, we need to transform the positive measurements of the original abundance matrices, here denoted with $\mathbf{Z}^{(t)}$, into real values as

$$y_{i,j}^{(t)} = \Phi^{-1} \left(\frac{z_{i,j}^{(t)} - \underline{z}^{(t)}}{\bar{Z}^{(t)} - \underline{z}^{(t)}} \right), \quad \text{for } i = 1, \dots, N; \quad j = 1, \dots, J, \quad t = 1, \dots, T, \quad (4)$$

where $\Phi^{-1}(\cdot)$ represents the inverse c.d.f. of a Normal distribution, $\underline{z}^{(t)} = \min_{i,j} (z_{i,j}^{(t)}) - \epsilon$, and $\bar{Z}^{(t)} = \max_{i,j} (z_{i,j}^{(t)}) + \epsilon$, where a small $\epsilon > 0$ ensures the arguments remain within $(0, 1)$ to prevent divergence. To complete our model specification, exploiting conjugacy, we finally specify

Normal-Inverse Gamma priors for the base measures $H^{(t)}$, $t = 1, \dots, T$. Thus, for all l , we assume $\theta_l^{*(t)} = (\mu_l^{*(t)}, \sigma_l^{2*(t)})$ where $\mu_l^{*(t)} \mid \sigma_l^{2*(t)} \sim \mathcal{N}(m_0^{(t)}, \sigma_l^{2*(t)}/k_0^{(t)})$ and $\sigma_l^{2*(t)} \sim \text{IG}(c_0^{(t)}, d_0^{(t)})$. Recommended values for the hyperparameters are $(m_0^{(t)}, k_0^{(t)}, c_0^{(t)}, d_0^{(t)}) = (0, 0.01, 3, 2)$ for all t . We also set $\mathbf{b}_0 = 10^{-4} \cdot \mathbf{1}_L$ to ensure sparse, overfitted finite mixtures across the rows [Malsiner-Walli et al., 2016], effectively estimating the number of non-empty clusters. Finally, we set $a_\alpha = b_\alpha = 1$ and $B = 2$. A detailed discussion of the selected hyperparameters is provided in Supplementary Section S.3.3.

3 Posterior inference via variational Bayes

We perform posterior inference using a tailored variational Bayes (VB) algorithm to ensure an efficient estimation of Poseidon on our high-dimensional MALDI-MSI data. In short, let $\Theta = \left(\{\mathcal{R}^{(t)}\}_t, \mathbf{C}, \theta^*, \pi, \{\omega_k^{(t)}\}_{t,k}, \alpha, \beta \right)$ represent the set of all the parameters in our model. After a family of approximating densities $\mathcal{Q}$ is specified over Θ , we seek to find the member $q^*(\Theta)$ of $\mathcal{Q}$ that minimizes the Kullback-Leibler divergence ($\mathcal{D}_{KL}$) with the true posterior distribution. Note that minimizing this divergence is equivalent to maximizing the Evidence Lower Bound (ELBO), defined as $ELBO(q) = \mathbb{E}[\log p(\Theta, \mathcal{Y})] - \mathbb{E}[\log q(\Theta)]$. The variational family $\mathcal{Q}$ is typically characterized by variational parameters λ . The optimization problem is thus reduced to finding the optimal parameters $\lambda^* = \arg \min_{\lambda} \mathcal{D}_{KL}(q_{\lambda}(\Theta) \parallel p(\Theta \mid \mathcal{Y}))$. We opt for a mean-field variational approximation, which chooses $\mathcal{Q}$ assuming the parameters are mutually independent. To find λ^* , we resort to a Coordinate Ascent Variational Inference (CAVI) procedure, detailed in Section S.3 of the Supplementary Materials. The single-dataset case (Pose model) is easily recovered when $T = 1$. Almost all the CAVI steps can be performed in closed form, as the corresponding full conditionals are known distributions. The only steps that require approximations are the updates for the parameters of the distributions of v_k 's and the inverse temperature β . For the former approximation, we mimic Lü et al. [2020] and assume $\beta = 0$. For the latter, we proceed as delineated in Supplementary Section S.3.1. To fit the model, we initialize our CAVI algorithm using different starting points, stopping the computations when the relative increment in the ELBO is below a certain threshold, such as 10^{-8} . Once all the runs are performed, we retain the estimate that obtained the highest value of the ELBO. The details for the computation of the ELBO components are in Section S.3.2 of the Supplementary Materials.

4 Simulations

We conducted several simulation studies to evaluate the performance of our proposal. All the studies are deferred to the Supplementary Materials. First, Section S.4.1 demonstrates why shared atoms are preferred over common atoms in this context. In Section S.4.2, we validate Pose's ability to recover biclusters accurately, and in Section S.4.3 we show that Pose achieves better cluster recovery than state-of-the-art competitors across datasets with varying degrees of spatial dependence. In Section S.4.4, we show that nested clustering with Poseidon is driven by the shared signal rather than corrupted by noise. Finally, Section S.4.5 analyzes computational scalability, presenting the average per-iteration runtime (in seconds) as J , T , and N vary.

5 ccRCC image segmentation and molecular signal extraction

We fit Poseidon to the ccRCC dataset described in Section 1.1 after applying the transformation in (4). To estimate the model, we set the upper bound for the number of CCs and RCs to $K = 30$ and $L = 40$, respectively. We run the algorithm using 250 different starting points

on a Linux machine with 32 GB of memory. The average running time is 3.2 minutes, with a standard deviation of 1.2 minutes. We summarize the main results below; additional details are provided in Section S.5 of the Supplementary Materials.

5.1 Tissue segmentation

To evaluate our proposed multiomic approach, we first examine the spatial partitions generated by Poseidon, obtained by integrating information across the three molecular classes (Figure 3, panel E). Our proposal reduces modality-specific noise and yields a unified, coherent, and biologically meaningful view of the tissue’s molecular landscape. By imposing a shared column partition, Poseidon produces an interpretable solution with a concise number of clusters (15 CCs). To provide an indirect measure of goodness-of-fit of the estimation, we compute the root mean squared error (RMSE) by aggregating the discrepancies between the observed data values and the corresponding estimated means within each bicluster. For Poseidon, the resulting average RMSE is 0.029. The estimated densities for each CC are provided in the Supplementary Materials. Importantly, Poseidon reveals multiomic alterations in both the tumor microenvironment and adjacent histologically normal tissue, features that may be difficult to capture through standard histopathological evaluation due to limited morphological differences (Figure 1, panel A).

Figure 3 also reports the results of different Pose models estimated on the individual datasets: lipids (panel A), N-glycans (panel B), and peptides (panel C), and allows us to highlight the underlying biological architectures across which Poseidon enables integration. Indeed, single-molecule segmentation reveals that N-glycans and peptides are closely linked, as N-glycans bind to specific peptide sequences and act as labels that direct proteins to their cellular destinations. In the healthy region of the tissue (right side of the section), the clustering of N-glycans and peptides shows greater concordance than the lipid modality, identifying a non-concentric large-scale organization of the healthy cortex. Within the tumor nodule, however, their clustering patterns diverge, likely reflecting disrupted glycosylation processes in cancer [Wallace et al., 2024]. Both of these layers successfully identify the mixed-grade tumor (G2-G3) region. In contrast, individual lipid-based segmentation reveals concentric layers within the healthy region, a pattern also detected by Poseidon (panel E). This suggests the presence of metabolic or inflammatory gradients that operate largely independently of the protein–glycosylation architecture. Lipids also highlight distinct regions within the tumor, including areas associated with inflammatory infiltration and hemorrhagic features (panel F), as well as concentric layers surrounding the tumor nodule that extend into the healthy region, which are less evident in the N-glycan and peptide segmentations. Notably, in the upper portion of the tumor nodule, Poseidon consistently delineates a localized region (Tumor G2–G3 in panel F), which is more prominently captured in the N-glycan and peptide modalities than in the lipid layer.

Finally, we compare our approach against a model estimated on a stacked dataset combining all three molecular classes. While stacking offers a straightforward alternative for data integration, it is generally considered a less principled approach (see Supplementary Section S.4.4 for further discussion). This yields a slightly better fit (RMSE: 0.022) but results in a larger number of CCs (19 CCs), which hinders interpretability. The stacked analysis treats all signals as originating from a single dataset. As a result, some tissue areas appear more fragmented, especially where peptide and N-glycan layers suggest distinct spatial organizations. Furthermore, the resulting clusters predominantly reflect the lipid-driven concentric structure in the healthy region. This suggests that the lipid modality drives the stacked segmentation, while additional molecular-specific patterns may be less distinctly represented. Overall, while similar spatial features are observed with stacking, an integrative method such as Poseidon produces a cleaner segmentation and more effectively captures the distinct characteristics of each molecular modality described above.

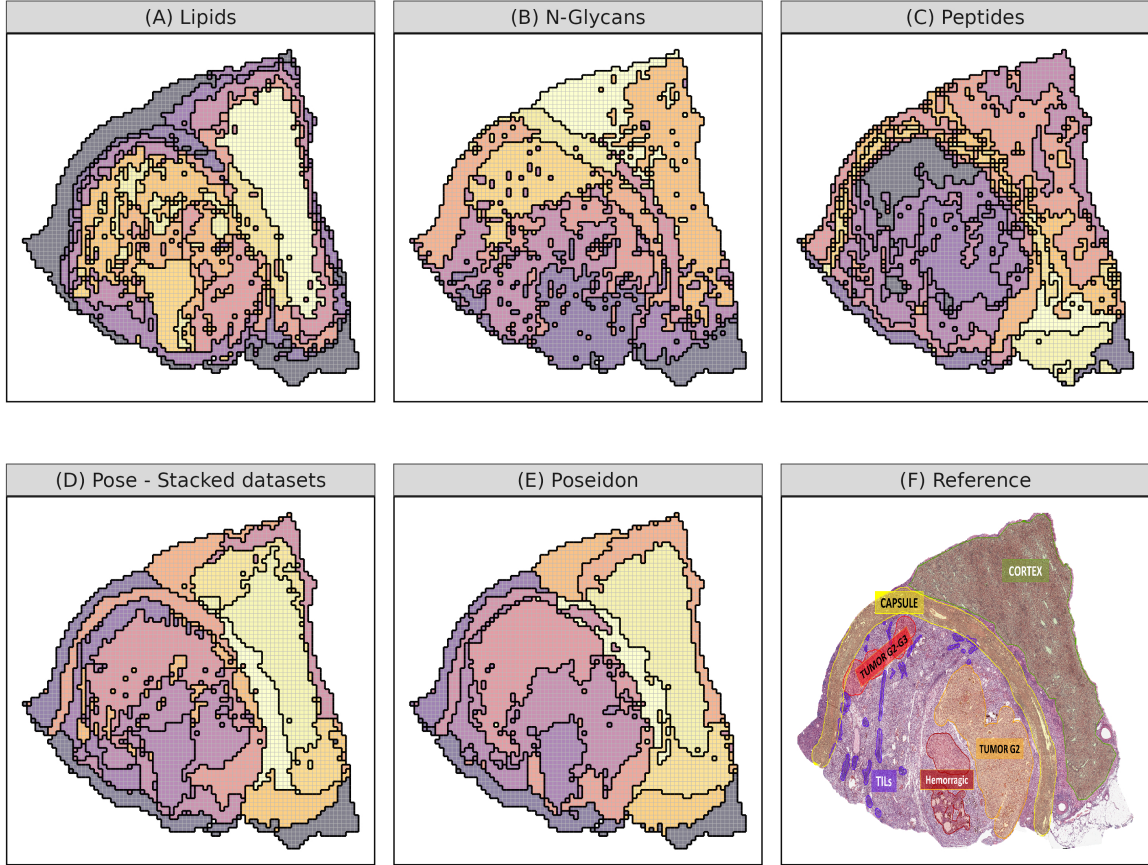


Figure 3: Tissue segmentation obtained by fitting single-dataset Pose models to the separate molecules (panels A–C) and the stacked datasets (panel D). These results are compared with Poseidon’s multi-dataset approach (panel E) and the original ccRCC tissue annotated by a pathologist (panel F).

5.2 Results on m/z in specific CCs

To illustrate the type of inference enabled by the Poseidon formulation, we focus on three CCs associated with biologically meaningful tissue regions (left panels of Figure 4). Two correspond to distinct areas within the tumor nodule (CC-10 and CC-6, top and bottom panels), while the central panel (CC-12) represents a healthy region of the kidney cortex. CC-10 corresponds to a mixed-grade tumor region, annotated by the pathologist as intermediate between *grade 2* and *grade 3*, whereas CC-6 contains hemorrhagic tissue with tumor cells classified as *grade 2*. Distinguishing between these tumor areas is challenging even for experienced pathologists, given their subtle morphological differences. Poseidon’s multiomic integration, in contrast, enables a spatially resolved characterization of molecular behavior across these regions.

The right panels of Figure 4 display the posterior median of each row assigned to the CCs. Dots, colored according to their m/z partition, highlight how specific molecular signals contribute to each molecular class, allowing the identification of the signals that most strongly characterize each region. N-glycans show broadly similar patterns across the tumor clusters (CC-10 and CC-6), consistent with well-documented cancer-associated glycan alterations, while the benign region (CC-12) displays a distinct profile. This difference is particularly evident in the mid- m/z range (approximately 1688–1926), where the tumor clusters exhibit a characteristic low–high–low triplet pattern across adjacent signals, whereas the benign one shows a steadily increasing trend. The similarity between tumor clusters reflects the tendency

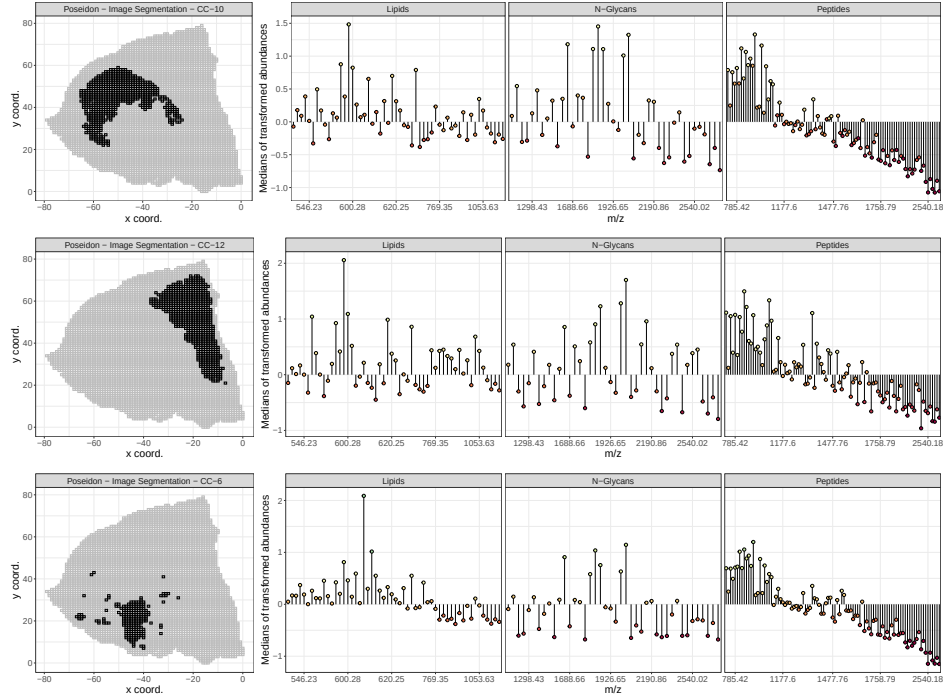


Figure 4: Spatial and signal characterization of three representative regions of interest identified within the tissue by Poseidon. The left panels show the pixels corresponding to each estimated CC, highlighting the localization patterns across the sample. Right panels display the corresponding posterior median of the analytes’ rescaled abundances in each CC, segmented by molecular class. Therefore, each panel illustrates the relative median intensities of m/z features contributing to the cluster-specific molecular profile, and these are colored according to the RC assignments.

of cancer tissues to share glycosylation alterations, even when their histological appearances differ. In contrast, lipidomic signals reveal bigger differences between tumor regions. Marked variation is observed in the higher m/z range, likely reflecting distinct metabolic states in healthy and tumor tissue. CC-6 shows particularly elevated abundance of small lipids in the 606–619 m/z range, consistent with the hemorrhagic nature of this region. Interestingly, these lower m/z lipid signals exhibit more similar behavior between CC-10 and the healthy CC-12, possibly reflecting inflammatory processes present in both tumor-adjacent tissue and portions of the normal cortex.

Peptides, instead, tend to be more abundant in the healthy region. In the mid-range of the spectrum (approximately m/z 700–1500), several peptide signals form a group of peaks that are consistently elevated in CC-12 while remaining weaker in the tumor clusters. Among these are peptides corresponding to nuclear proteins such as m/z 944.53 (*histone H2A*) and m/z 1116.56 (putatively identified as *Heterochromatin Protein 1-binding protein 3*), recognized tumor markers in the MSI literature [Denti et al., 2024]. While proteomic alterations in cancer have been extensively studied, the functional roles of N-glycans and lipids remain less explored, particularly with respect to their spatial organization across tissue regions [Vasseur and Guillaumond, 2022, Wallace et al., 2024]. Our findings provide a foundation for future investigations into the spatially resolved roles of these molecular layers in disease processes. Complementary comments on the other CCs are reported in Supplementary Section S.5.2.

6 Discussion

We introduced Poseidon, a BNP method for the joint analysis of multi-molecule MALDI-MSI data, designed to effectively integrate the heterogeneous information provided by different spatially resolved multiomics experiments. By combining a BNP-MRF prior with a separately exchangeable biclustering framework, we could simultaneously segment a tissue and cluster molecular features across omics classes while preserving spatial dependencies. This principled integration improved over both single-dataset and naively combined approaches, as demonstrated in our analysis of lipids, N-glycans, peptides, and stacked datasets. The improvements observed, both in terms of biologically meaningful tissue segmentation and the discovery of plausible molecular biomarkers, underscore the importance of integrated models for spatially structured molecular data.

This work opens up several promising research directions. First, one can consider different likelihood specifications, allowing for a more flexible modeling of the data without resorting to pre-processing transformations. Second, alternative spatially informed priors could be explored as an alternative to the MRF, which is known to pose computational challenges. For example, thresholded Gaussian processes or spatially dependent SB priors [e.g., [Linderman et al., 2015](#)] may offer more flexible and scalable alternatives. Finally, incorporating stochastic variational inference could enable scalability to larger datasets, facilitating information sharing across multiple tissues from the same patient collected over time, or across tissues from different patients with the same disease. This would potentially allow for integrated, one-step differential analyses across space, time, and individuals.

Acknowledgment

FD acknowledges support by the STARS@UNIPD 2025 program, with a contribution from the Fondazione Cassa di Risparmio di Padova e Rovigo.

Supplementary Materials

Web Appendices and Figures referenced in Sections 1-5 are available with this paper at the Biometrics website on Oxford Academic. Furthermore, the code and data retrieval instructions required to reproduce the results are hosted in the GitHub repository at <https://github.com/Fradenti/poseidon-maldi>.

Data Availability

The data [[Denti et al., 2026](#)] mentioned in Sections 1 and 5 of this paper are publicly available on the *Bicocca Open Archive Research Data* and can be accessed at <https://doi.org/10.17632/gxz9kjj2r9.1>.

Supplementary Materials for “Multiomics Tissue Segmentation via Spatially-Informed Nested Biclustering Methods”

Francesco Denti^{*1}, Cecilia Balocchi², Vanna Denti³, and Giulia Capitoli^{4,5}

¹Department of Statistical Sciences, University of Padova

²School of Mathematics, University of Edinburgh

³Proteomics and Metabolomics Unit, Department of Medicine and Surgery, University of Milano-Bicocca

⁴Bicocca Bioinformatics Biostatistics and Bioimaging B4 Center, Department of Medicine and Surgery, University of Milano-Bicocca

⁵Biostatistics and Clinical Epidemiology, Fondazione IRCCS San Gerardo dei Tintori

Contents

1	Introduction	2
1.1	Clear Cell Renal Cell Carcinoma MALDI-MSI Data	3
2	A finite-infinite separately exchangeable model for MALDI-MSI data	4
2.1	The single-dataset case: Pose	4
2.1.1	Prior coclustering properties under Pose	7
2.2	The multiple-dataset case: Poseidon	8
3	Posterior inference via variational Bayes	9
4	Simulations	9
5	ccRCC image segmentation and molecular signal extraction	9
5.1	Tissue segmentation	10
5.2	Results on m/z in specific CCs	11
6	Discussion	13
S.1	Additional details about the MALDI-MSI data used in the analysis	16
S.1.1	The ccRCC tissue section and the biological preprocessing	16
S.2	Additional details on the modeling framework	19
S.2.1	Proof of Proposition 2.1	19
S.3	The CAVI algorithm for Poseidon	23
S.3.1	Approximations in the algorithm	23

^{*}Corresponding author: francesco.denti@unipd.it

S.3.2 ELBO components	25
S.3.3 Recommended hyperparameter specifications	26
S.4 Simulation studies	28
S.4.1 Comparing common and shared atoms	28
S.4.2 Biclustering recovery	29
S.4.3 Varying degree of spatial dependence structure and comparison with alternative methods	31
S.4.4 Informative datasets guide the shared clustering estimation	33
S.4.5 Computational costs	35
S.5 Additional results of Poseidon on ccRCC tissue	37
S.5.1 Additional insights about important analytes of interest	37
S.5.2 Results for all the estimated column clusters	37

S.1 Additional details about the MALDI-MSI data used in the analysis

In this section, we provide additional details on how MALDI-MSI (Matrix-Assisted Laser Desorption/Ionization Mass Spectrometry Imaging) data are generated and how the matrices analyzed by Poseidon are derived from biological tissue. Recall that MALDI-MSI is a technology that measures the molecular composition of a tissue in a spatially resolved manner over a tissue section, which is divided into pixels. A laser scans the tissue section pixel by pixel, recording a mass spectrum for each one. This spectrum contains intensity values associated with thousands of mass-to-charge (m/z) ratios, each corresponding to a molecular signal.

A distinctive aspect of the experimental protocol in this study is that the same histological tissue section is subjected to several consecutive MALDI-MSI acquisitions, each focusing on a different molecular family, such as lipids, N-glycans, or peptides. Between acquisitions, the section undergoes chemical treatment, measurement, and re-preparation to enable the extraction of additional molecular information without changing the spatial grid. As a result, multiple datasets are derived from the same spatial pixel grid. This feature is essential for Poseidon, as it allows us to assume that all datasets share the same spatial structure (the columns of the matrices), while the molecular signals (rows) vary across datasets.

Figure S.1 schematically represents this multiomics acquisition workflow. After histological preparation, the tissue undergoes sequential MALDI-MSI measurements. Each acquisition produces a data matrix where:

- columns correspond to spatial pixels of the tissue,
- rows correspond to detected m/z peaks (molecular signals),
- entries represent the abundance of a specific molecular signal at a specific pixel.

Since all acquisitions are made on the same physical section using the same raster grid, the columns of all matrices are perfectly aligned in space. Consequently, column j in each dataset corresponds to the same tissue location (pixel). This spatial co-registration is essential, as it enables a meaningful statistical search for a common column partition across datasets.

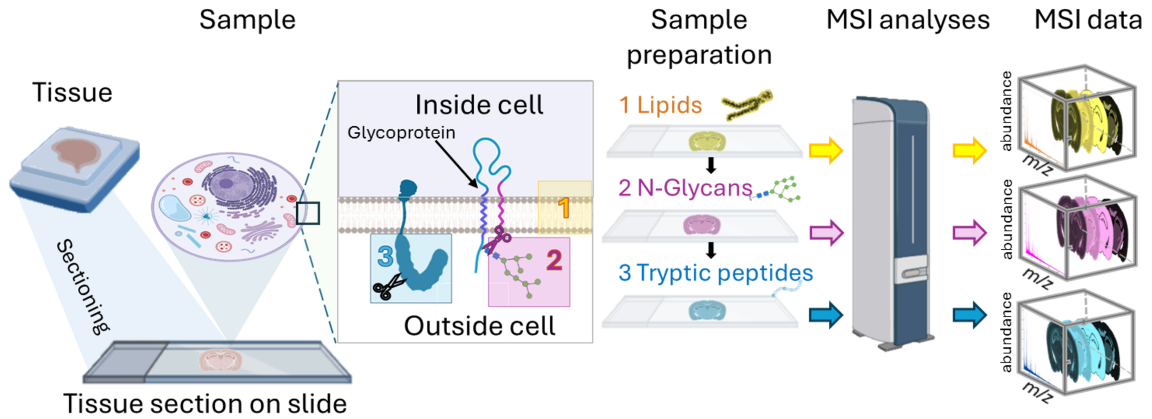


Fig. S.1: A depiction of the *multiomics workflow* developed in [Denti et al. \[2022\]](#). The *same* tissue undergoes multiple consecutive MSI analyses with MALDI, yielding different MSI datasets – one for each molecular class.

S.1.1 The ccRCC tissue section and the biological preprocessing

The sample considered in this study is from clear cell renal cell carcinoma (ccRCC), the most prevalent type of kidney cancer in adults. This tumor develops in the renal cortex and fea-

tures altered cellular metabolism and a dense vascular network. Figure S.2 provides detailed tissue context, aiding the accurate interpretation of the statistical results. The spatial heterogeneity within ccRCC, including regions with varying tumor grades, hemorrhagic areas, and immune cell infiltration, reflects the underlying biological complexity that shapes tumor behavior, prognosis, and response to treatment. Recognizing this heterogeneity is essential for accurately interpreting multiomics data within a biological framework. Indeed, in Figure S.2 regions of interest identified by an expert pathologist are color-coded as follows: green for healthy cortex, yellow for the capsule surrounding the tumor nodule, red for a tumor region with mixed grades (*grade 2* and *grade 3*), orange for a *grade 2* tumor region, brown for a hemorrhagic area, and purple for tumor-infiltrating lymphocytes (TILs). Although these annotations are not used as input in our model, they serve as an external reference to interpret the clustering results obtained by Pose and Poseidon.

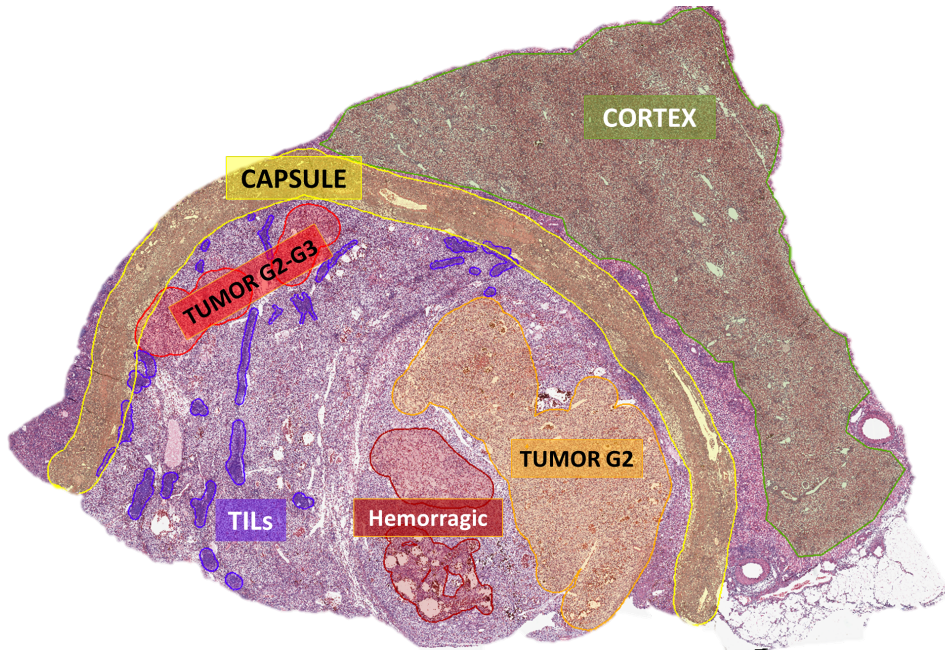


Fig. S.2: Annotated representation of the H&E-stained ccRCC tissue used for the analysis, with regions of interest defined by an expert pathologist: green indicates healthy cortex; yellow, the capsule surrounding the tumor nodule; red, a tumor region with mixed grades (*grade 2* and *grade 3*); orange, a *grade 2* tumor region; brown, a hemorrhagic area; and purple, tumor-infiltrating lymphocytes (TILs).

Raw data from this tissue were extracted and processed for statistical analysis via a standardized preprocessing pipeline [Capitoli et al., 2025]. This step is essential for minimizing physiological, instrumental, and analytical variability often observed in MALDI-MSI data. Overcoming these challenges ensures data quality and reliability for biological analysis and interpretation. Moreover, it enables the generation of comparable intensity profiles across pixels. In detail, the following steps transform raw spectra into the matrices used by Pose and Poseidon:

1. **Baseline subtraction:** Each spectrum contains a slowly varying background signal due to electronic noise and chemical impurities. This background artificially inflates intensity values, especially in low-intensity regions of the spectrum. Baseline subtraction estimates this component and removes it, leaving intensities to reflect only true molecular signals.
2. **Smoothing:** Spectra exhibit high-frequency fluctuations and small spurious peaks

caused by noise. Smoothing reduces these fluctuations while preserving the shape, height, and position of real peaks. This step improves the stability and reliability of peak identification.

3. **Normalization:** The total signal intensity varies across pixels because ionization efficiency is not uniform over the tissue. Without normalization, two pixels with identical molecular composition could appear very different. Each spectrum is therefore rescaled to a common reference level so that intensities can be meaningfully compared across spatial locations.
4. **Alignment:** Due to slight instrumental drift, the same molecule may appear at slightly different m/z values across spectra. Alignment ensures that peaks corresponding to the same molecular species are assigned the same m/z value in all pixels. This guarantees that each row of the data matrix refers to the same molecular signal.
5. **Peak detection:** Finally, peaks are kept only if they have a high enough signal-to-noise ratio. This eliminates non-informative parts of the spectrum and establishes the set of meaningful rows in the matrix.

Following these preprocessing steps, each MALDI-MSI acquisition becomes a data matrix where columns correspond to the spatial pixels shown in Figure S.2, rows represent the set of aligned and filtered molecular peaks, and entries contain the normalized abundance values. These matrices serve as the input for the biclustering model.

S.2 Additional details on the modeling framework

S.2.1 Proof of Proposition 2.1

We aim to derive the expressions for the observational coclustering probability between two data points under our nested prior, i.e., $\mathbb{P}[\theta_{i,j} = \theta_{i',j'}]$. Recall that $\omega_k \sim \text{Dirichlet}_L(b_0, \dots, b_0)$ and consider a BNP-MRF($\pi; \beta$) prior over the group partition, with $\pi \sim \text{GEM}(\alpha)$. First, let $\varrho_{j,j'} = \mathbb{P}[G_j = G_{j'}]$. Then, we have

$$\begin{aligned}\mathbb{P}[\theta_{i,j} = \theta_{i',j'}] &= \mathbb{E}[\mathbb{P}[\theta_{i,j} = \theta_{i',j'} \mid G_j, G_{j'}]] \\ &= \mathbb{E}[\mathbb{P}[\theta_{i,j} = \theta_{i',j'} \mid G_j = G_{j'}] \varrho_{j,j'}] + \mathbb{E}[\mathbb{P}[\theta_{i,j} = \theta_{i',j'} \mid G_j \neq G_{j'}] (1 - \varrho_{j,j'})]\end{aligned}$$

In the case of two samples, we can obtain $\varrho_{j,j'} = \mathbb{P}[G_j = G_{j'}] = e^\beta / (\alpha + e^\beta)$ from the predictive distribution presented in Lü et al. [2020]. The second term in the sum is identical to the partially exchangeable case, see D'Angelo and Denti [2026], Denti et al. [2023]. However, under separate exchangeability, if columns j and j' are clustered together, then the rows must share the same partition. Therefore,

$$\begin{aligned}\mathbb{P}[\theta_{i,j} = \theta_{i',j'} \mid G_j = G_{j'}] &= \mathbb{1}_{\{i=i'\}} \mathbb{P}[\theta_{i,j} = \theta_{i',j'} \mid G_j = G_{j'}] + \mathbb{1}_{\{i \neq i'\}} \mathbb{P}[\theta_{i,j} = \theta_{i',j'} \mid G_j = G_{j'}] \\ &= \mathbb{1}_{\{i=i'\}} \cdot 1 + \mathbb{1}_{\{i \neq i'\}} \sum_{l=1}^L \mathbb{P}[\theta_{i,j} = \theta_l^*, \theta_{i',j'} = \theta_l^* \mid G_j = G_{j'}] \\ &= \mathbb{1}_{\{i=i'\}} + \mathbb{1}_{\{i \neq i'\}} \sum_{l=1}^L \omega_{l,j}^2.\end{aligned}$$

Then, the marginal probability is

$$\begin{aligned}\mathbb{P}[\theta_{i,j} = \theta_{i',j'}] &= \frac{e^\beta}{e^\beta + \alpha} \left(\mathbb{1}_{\{i=i'\}} + \mathbb{1}_{\{i \neq i'\}} \mathbb{E} \left[\sum_{l=1}^L \omega_{l,j}^2 \right] \right) + \frac{\alpha}{e^\beta + \alpha} \mathbb{E} \left[\sum_{l=1}^L \omega_{l,j} \omega_{l,j'} \right] \\ &\stackrel{(\star)}{=} \frac{1}{(e^\beta + \alpha)} \left[e^\beta \left(\frac{1 + b_0}{1 + Lb_0} \right)^{1 - \mathbb{1}_{\{i=i'\}}} + \frac{\alpha}{L} \right],\end{aligned}$$

where in step $(\star)$ we used

$$\mathbb{E} \left[\sum_{l=1}^L \omega_{l,j}^2 \right] = \sum_{l=1}^L \mathbb{E}[\omega_{l,j}^2] = \sum_{l=1}^L \frac{(1 + b_0)}{L(1 + Lb_0)} = \frac{1 + b_0}{1 + Lb_0},$$

and

$$\mathbb{E} \left[\sum_{l=1}^L \omega_{l,j} \omega_{l,j'} \right] = \sum_{l=1}^L \mathbb{E}[\omega_{l,j} \omega_{l,j'}] \stackrel{\text{ind.}}{=} \sum_{l=1}^L \mathbb{E}[\omega_{l,j}]^2 = \frac{1}{L}.$$

In Figures S.3–S.5, we examine the behavior of the derived prior coclustering probabilities for two entries in the data matrix under various parameter specifications. Recall that in the common atoms model, $\pi \sim \text{GEM}(\alpha)$ and $\omega_k \sim \text{GEM}(\nu)$ for all k , whereas in the shared atoms model, $\omega_k \sim \text{Dirichlet}_L(b_0)$ for all k . Here, α is a *distributional parameter*, while ν and b_0 are *observational parameters*. We compute these probabilities for three inverse temperature levels, i.e., $\beta \in \{0, 1, 2\}$. The values of α and ν are varied over the interval $(0, 3)$, and the Dirichlet concentration parameters are constrained to the unit interval to enforce sparsity. We present comparisons for both the same-row and different-row scenarios.

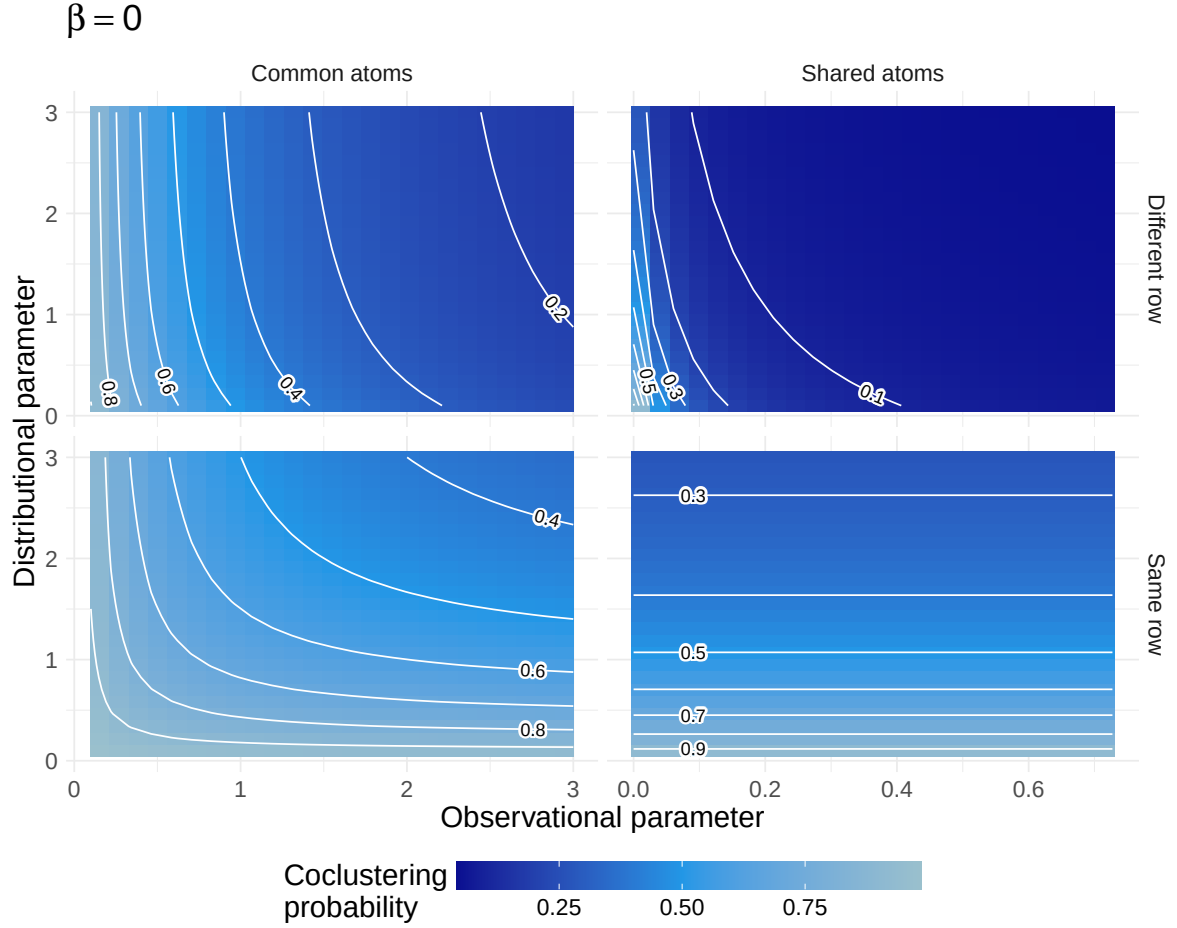


Fig. S.3: Heatmaps depicting the a priori coclustering probabilities for pairs of matrix entries in a two-sample scenario, plotted against the observational and distributional parameters governing the nested mixture weights. The left column illustrates the case with common atoms, while the right column shows the case with shared atoms. The top panels display probabilities for entries on different rows. In contrast, the bottom panels show the corresponding probabilities for entries on the same row. The inverse temperature parameter is fixed at $\beta = 0$.

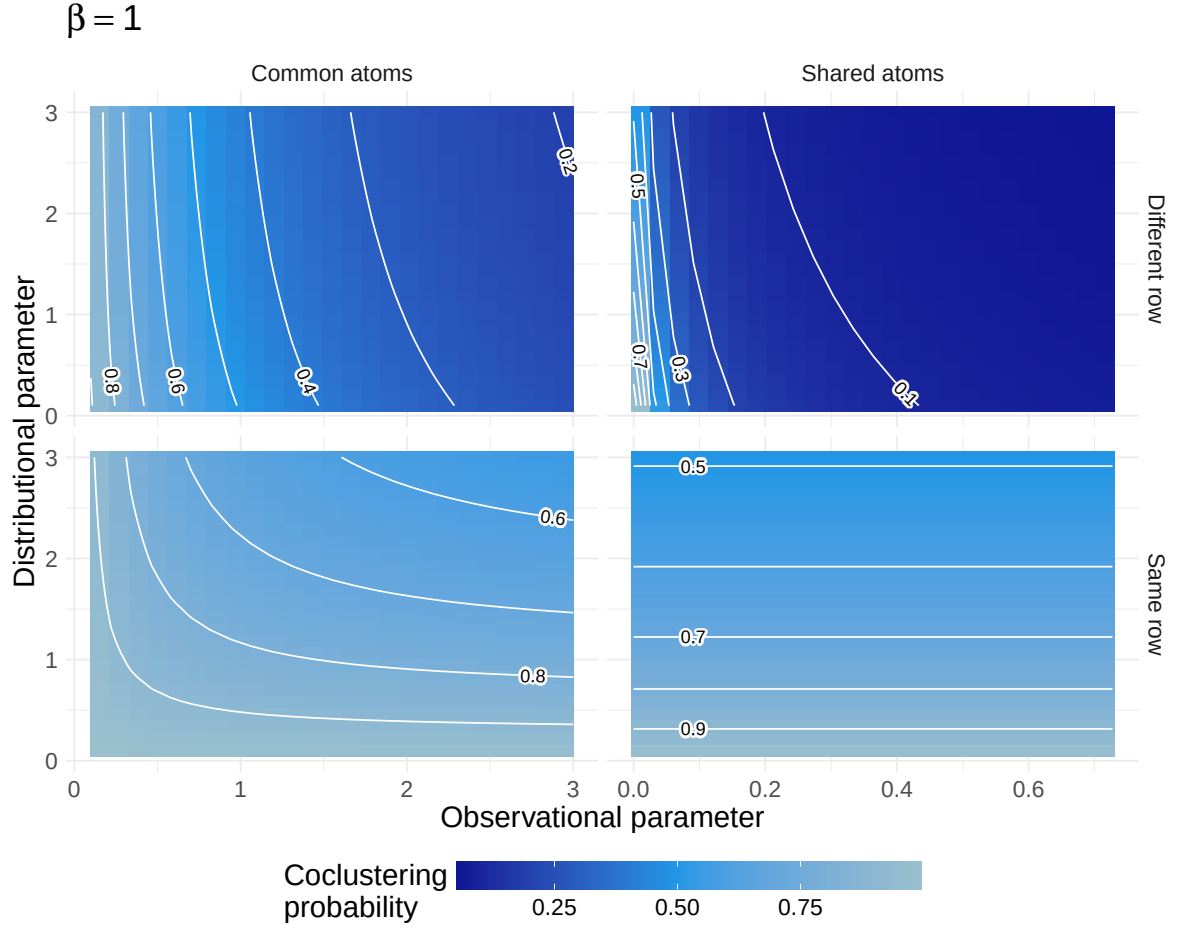


Fig. S.4: Heatmaps depicting the a priori coclustering probabilities for pairs of matrix entries in a two-sample scenario, plotted against the observational and distributional parameters governing the nested mixture weights. The left column illustrates the case with common atoms, while the right column shows the case with shared atoms. The top panels display probabilities for entries on different rows. In contrast, the bottom panels show the corresponding probabilities for entries on the same row. The inverse temperature parameter is fixed at $\beta = 1$.

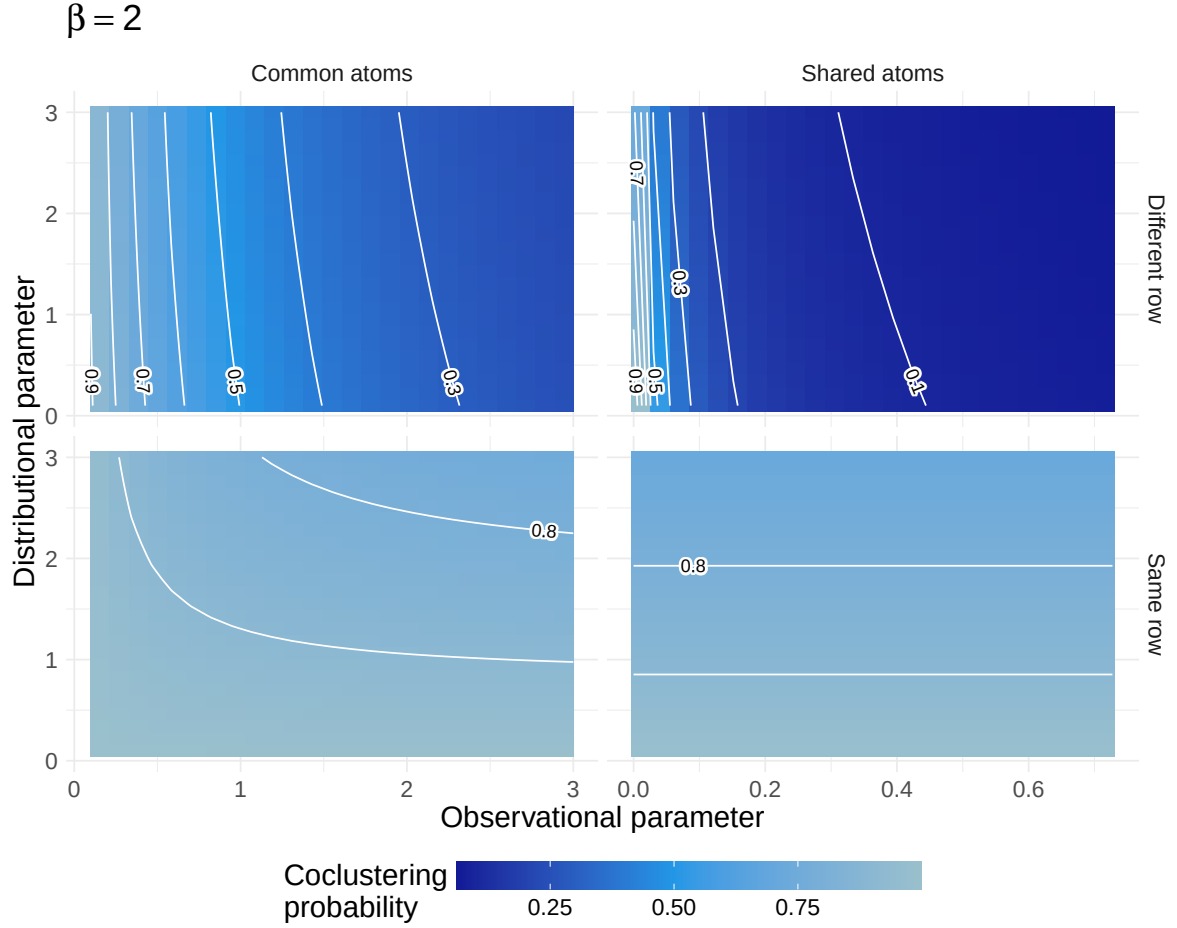


Fig. S.5: Heatmaps depicting the a priori coclustering probabilities for pairs of matrix entries in a two-sample scenario, plotted against the observational and distributional parameters governing the nested mixture weights. The left column illustrates the case with common atoms, while the right column shows the case with shared atoms. The top panels display probabilities for entries on different rows. In contrast, the bottom panels show the corresponding probabilities for entries on the same row. The inverse temperature parameter is fixed at $\beta = 2$.

S.3 The CAVI algorithm for Poseidon

In this section, we detail the Coordinate Ascent Variational Inference (CAVI) algorithm for the Poseidon model and discuss the approximations necessary to make it feasible. The complete algorithm is presented in Algorithm 1.

Note that, similarly to the case of the nested model presented in D’Angelo and Denti [2026], resorting to a variational Bayes approach not only provides estimates in a timely and efficient manner but also addresses the problem of label switching and, more importantly, offers a direct method for estimating the partition of rows and columns. Indeed, once the algorithm has converged, we are left with two sets of posterior variational clustering probabilities that we denote with $\hat{\rho}_{j,k} = q^*(C_j = k)$ and $\hat{\xi}_{i,k,l} = q^*(R_{i,k} = l)$. The former indicates how likely the j -th column is to be assigned to the k -th cluster. Similarly, the latter represents the probability that the i -th row belongs to the l -th cluster within the k -th CC. Thus, we estimate the pixel segmentation as $\hat{C}_j = \arg \max_{k=1,\dots,K} \hat{\rho}_{j,k} \forall j$. Analogously, the RC solutions nested in the k -th column cluster are given by $\hat{R}_{i,k} = \arg \max_{l=1,\dots,L} \hat{\xi}_{i,k,l} \forall i, k$. It is important to note that, with this procedure, we obtain K partitions of the rows, one for each *potential* CC. However, only the ones corresponding to non-empty CCs are interesting.

S.3.1 Approximations in the algorithm

Due to the presence of an intractable normalizing constant, inference on the terms involving the inverse temperature parameter of a Potts model β is extremely challenging. Many approaches have been proposed in the literature, ranging from the use of approximations such as the *pseudo-likelihood* [PL, Besag, 1974], to more refined samplers [Murray et al., 2006], to ABC techniques [Moore et al., 2020].

In the variational Bayes framework, Lü et al. [2020], Durand et al. [2022] treat β as a fixed nuisance parameter w.r.t. maximizing the lower bound, thus resorting to a VB-Expectation Maximization (VBEM) algorithm. This approach translates into computing

$$\hat{\beta} = \arg \max_{\beta} \mathbb{E}_{q_{\mathbf{C}}, q_{\mathbf{v}}} [\log(p(\mathbf{C} \mid \boldsymbol{\pi}; \beta))]. \quad (\text{S.1})$$

Conversely, McGrory et al. [2009] treated β as a random variable and proposed different approximations to the normalizing constant to derive the variational distribution $q^*(\beta)$. In our algorithm, we adopt approximations inspired by McGrory et al. [2009], which we delineate in the following.

A key quantity for both the approximate updates of β and the ELBO components for both $\mathbf{C}$ and β is the expectation of $\log p(\mathbf{C} \mid \boldsymbol{\pi}, \beta)$. Given the intractability of the normalizing constant involved in the expression, computing this expectation analytically is unfeasible. We then resort to two approximations. First, the Potts prior is approximated using a mean-field pseudo-likelihood (PL) approach. Then, every variable in $\mathbf{v}$ is substituted with its variational expectation (A2), switching log and expectation. Therefore, we first write

$$p(\mathbf{C} \mid \boldsymbol{\pi}, \beta) \stackrel{(PL)}{\approx} \prod_{j=1}^J \tilde{p}(C_j \mid \mathbf{C}_{-j}, \boldsymbol{\pi}, \beta) = \prod_{j=1}^J \frac{\exp \left[\log \pi_{C_j} + \beta \sum_{i \in \partial_j} \mathbb{1}_{\{C_i = C_j\}} \right]}{\sum_{k \geq 1} \exp \left[\log \pi_k + \beta \sum_{i \in \partial_j} \mathbb{1}_{\{C_i = k\}} \right]}, \quad (\text{S.2})$$

Algorithm 1: CAVI updates for the Poseidon model

Input: $\text{it} \leftarrow 1$. Randomly initialize $\boldsymbol{\lambda}^{(0)}$. Define the error bound ε and randomly set $\Delta(0, 1) > \varepsilon$.

while $\Delta(\text{it} - 1, \text{it}) > \varepsilon$ **do**

 Set $\text{it} = \text{it} + 1$;

 Update the variational parameters according to the following CAVI steps:

1. For $j = 1, \dots, J$, $q^*(C_j)$ is a K -dimensional multinomial, with $q^*(C_j = k) = \rho_{j,k}$ for $k = 1, \dots, K$, and

$$\log \rho_{j,k} \propto g(\bar{a}_k, \bar{b}_k) + \sum_{r=1}^{k-1} g(\bar{b}_r, \bar{a}_r) + \sum_{l=1}^L \sum_{t=1}^T \sum_{i=1}^{N^{(t)}} \xi_{i,k,l}^{(t)} \ell_{i,j,l}^{(t)} + \beta \sum_{q \in \partial_j} \rho_{q,k},$$

where $g(x, y) = \psi(x) - \psi(x + y)$, with ψ being the digamma function, and

$$\ell_{i,j,l}^{(t)} = -\frac{1}{2} \left[\log(d_l^{(t)}) - \psi(c_l^{(t)}) + \frac{1}{k_l^{(t)}} + \frac{c_l^{(t)}}{d_l^{(t)}} (y_{i,j}^{(t)} - m_l^{(t)})^2 \right].$$

2. For $k = 1, \dots, K$, $t = 1, \dots, T$ and $i = 1, \dots, N^{(t)}$, $q^*(R_{i,k}^{(t)})$ is a L -dimensional multinomial, with $q^*(R_{i,k}^{(t)} = l) = \xi_{i,k,l}^{(t)}$ for $l = 1, \dots, L$, and

$$\log \xi_{i,k,l}^{(t)} \propto \sum_{j=1}^J \rho_{j,k} \ell_{i,j,l}^{(t)} + h_l(\mathbf{p}_k^{(t)}),$$

where $\ell_{i,j,l}^{(t)}$ is defined as in Step (1), and $h_l(\mathbf{p}_k^{(t)}) = \psi(p_{l,k}^{(t)}) - \psi(\sum_{l=1}^L p_{l,k}^{(t)})$.

3. For $k = 1, \dots, K$ and $t = 1, \dots, T$, $q^*(\boldsymbol{\omega}_k^{(t)})$ is Dirichlet $_L(\mathbf{p}_k^{(t)})$ with $p_{l,k}^{(t)} = b_0 + \sum_{i=1}^{N^{(t)}} \xi_{i,k,l}^{(t)}$.

4. For $k = 1, \dots, K$, $q^*(v_k)$ is approximated as in Lü et al. [2020] with a Beta($\bar{a}_k, \bar{b}_k$) distribution with

$$\bar{a}_k = 1 + \sum_{j=1}^J \rho_{j,k}, \quad \bar{b}_k = s_1/s_2 + \sum_{j=1}^J \sum_{q=k+1}^{K-1} \rho_{j,q}.$$

5. For $l = 1, \dots, L$ and $t = 1, \dots, T$, $q^*(\theta_l^{*(t)})$ is a NIG($m_l^{(t)}, k_l^{(t)}, c_l^{(t)}, d_l^{(t)}$) distribution with parameters $k_l^{(t)} = n_l^{(t)} + k_0^{(t)}$, $c_l^{(t)} = c_0^{(t)} + n_l^{(t)}/2$,

$$m_l^{(t)} = \frac{k_0^{(t)} m_0^{(t)} + n_l^{(t)} \bar{y}_l^{(t)}}{n_l^{(t)} + k_0^{(t)}}, \quad d_l^{(t)} = d_0^{(t)} + \frac{n_l^{(t)} k_0^{(t)}}{2k_l^{(t)}} \left(\bar{y}_l^{(t)} - m_0^{(t)} \right)^2 + \frac{1}{2} \sum_{j=1}^J \sum_{i=1}^{N^{(t)}} Q_{i,j,l}^{(t)} \left(y_{i,j}^{(t)} - \bar{y}_l^{(t)} \right)^2,$$

where $Q_{i,j,l}^{(t)} = \sum_{k=1}^K \xi_{i,k,l}^{(t)} \rho_{j,k}$, $n_l^{(t)} = \sum_{i=1}^{N^{(t)}} \sum_{j=1}^J Q_{i,j,l}^{(t)}$, and $\bar{y}_l^{(t)} = (\sum_{i=1}^{N^{(t)}} \sum_{j=1}^J Q_{i,j,l}^{(t)} y_{i,j}^{(t)}) / n_l^{(t)}$.

6. $q^*(\alpha)$ is a Gamma(s_1, s_2) distribution with parameters

$$s_1 = a_\alpha + K - 1, \quad s_2 = b_\alpha - \sum_{k=1}^{K-1} g(\bar{b}_k, \bar{a}_k).$$

7. If β is assumed random, compute the expected value of the inverse temperature, $\bar{\beta}$, following the approximating procedure reported in Section S.3.1 and set $\beta = \bar{\beta}$ in Step 1.

Compute $\Delta(\text{it} - 1, \text{it}) = |\left[\text{ELBO}(\boldsymbol{\lambda}^{(\text{it})}) - \text{ELBO}(\boldsymbol{\lambda}^{(\text{it}-1)}) \right] / \text{ELBO}(\boldsymbol{\lambda}^{(\text{it}-1)})|$.

Let $\boldsymbol{\lambda} = \boldsymbol{\lambda}^{(\text{it})}$;

return $\boldsymbol{\lambda}$

Then, taking the expectation w.r.t. all the variables but β , we have

$$\begin{aligned}
\mathbb{E} [\log p(\mathbf{C} \mid \boldsymbol{\pi}, \beta)] &\stackrel{(PL)}{\approx} \sum_{j=1}^J \mathbb{E} \left[\sum_{k \geq 1} \mathbb{1}_{\{C_j=k\}} \log \pi_k + \beta \sum_{k \geq 1} \sum_{i \in \partial_j} \mathbb{1}_{\{C_i=k\}} \mathbb{1}_{\{C_j=k\}} \right] \\
&\quad - \mathbb{E} \left[\sum_{j=1}^J \log \sum_{k \geq 1} \exp \left[\log \pi_k + \beta \sum_{i \in \partial_j} \mathbb{1}_{\{C_i=k\}} \right] \right] \\
&= \sum_{j=1}^J \sum_{k \geq 1} \rho_{j,k} \mathbb{E} [\log \pi_k] + \beta \sum_{j=1}^J \sum_{k \geq 1} \sum_{i \in \partial_j} \rho_{i,k} \rho_{j,k} \\
&\quad - \sum_{j=1}^J \mathbb{E} \left[\log \sum_{k \geq 1} \exp \left[\log \pi_k + \beta \sum_{i \in \partial_j} \mathbb{1}_{\{C_i=k\}} \right] \right].
\end{aligned}$$

We can further approximate this term by writing

$$\stackrel{(A2)}{\approx} \sum_{j=1}^J \sum_{k \geq 1} \rho_{j,k} \left[\mathbb{E} [\log \pi_k] + \beta \sum_{i \in \partial_j} \rho_{i,k} \right] - \sum_{j=1}^J \log \sum_{k \geq 1} \exp \left[\mathbb{E} [\log \pi_k] + \beta \sum_{i \in \partial_j} \rho_{i,k} \right],$$

where we set $\rho_{j,k} = q^*(C_j = k)$. The two approximations led to a quantity that is easy to evaluate. Thus, given a prior $p(\beta)$ for the inverse temperature parameter, we can approximate

$$q^*(\beta) \stackrel{(PL+A2)}{\propto} p(\beta) \prod_{j=1}^J \frac{\exp \left[\sum_{k \geq 1} \rho_{j,k} \left[\mathbb{E} [\log \pi_k] + \beta \sum_{i \in \partial_j} \rho_{i,k} \right] \right]}{\sum_{k \geq 1} \exp \left[\mathbb{E} [\log \pi_k] + \beta \sum_{i \in \partial_j} \rho_{i,k} \right]}. \quad (\text{S.3})$$

Suppose we assume $p(\beta)$ to be a uniform distribution over a grid of sensible values. In that case, it is then straightforward to numerically evaluate the variational expected value of this distribution.

S.3.2 ELBO components

- $\mathbb{E} [\log p(\mathbf{Y} \mid \mathbf{R}, \mathbf{C}, \boldsymbol{\theta})] = \sum_{t=1}^T \sum_{j=1}^J \sum_{i=1}^{N^{(t)}} \sum_{k=1}^K \sum_{l=1}^L \rho_{j,k} \xi_{i,k,l}^{(t)} \ell_{i,j,l}^{(t)}$ where $\ell_{i,j,l}^{(t)} = \mathbb{E} \left[\log(\phi(y_{i,j}^{(t)} \mid \theta_l^{(t)})) \right] = -\frac{1}{2} \left[\log(d_l^{(t)}) - \psi(c_l^{(t)}) + \frac{1}{k_l^{(t)}} + \frac{c_l^{(t)}}{d_l^{(t)}} (y_{i,j}^{(t)} - m_l^{(t)})^2 \right]$;
- $\mathbb{E} [\log p(\mathbf{R} \mid \boldsymbol{\omega})] = \sum_{t=1}^T \sum_{i=1}^{N^{(t)}} \sum_{k=1}^K \sum_{l=1}^L \xi_{i,k,l}^{(t)} h_l(\mathbf{p}_k^{(t)})$;
- $\mathbb{E} [\log q(\mathbf{R})] = \sum_{t=1}^T \sum_{i=1}^{N^{(t)}} \sum_{k=1}^K \sum_{l=1}^L \xi_{i,k,l}^{(t)} \log(\xi_{i,k,l}^{(t)})$;
- $\mathbb{E} [\log p(\mathbf{C} \mid \boldsymbol{\pi}, \beta)]$ is not available in closed form. One can approximate it, following the rationale presented in Section S.3.1 as

$$\begin{aligned}
\mathbb{E} [\log p(\mathbf{C} \mid \boldsymbol{\pi}, \beta)] &\approx \sum_{j=1}^J \sum_{k \geq 1} \rho_{j,k} \left[\mathbb{E} [\log \pi_k] + \bar{\beta} \sum_{i \in \partial_j} \rho_{i,k} \right] \\
&\quad - \sum_{j=1}^J \log \sum_{k \geq 1} \exp \left[\mathbb{E} [\log \pi_k] + \bar{\beta} \sum_{i \in \partial_j} \rho_{i,k} \right],
\end{aligned}$$

with $\bar{\beta}$ representing the expected value of the variational distribution for β , computed numerically.

- $\mathbb{E} [\log q(\mathbf{C})] = \sum_{j=1}^J \sum_{k=1}^K \rho_{j,k} \log(\rho_{j,k})$
- $\mathbb{E} [\log p(\mathbf{v} \mid \alpha)] = (K-1)(\psi(s_1) - \log(s_2)) + (s_1/s_2 - 1) \sum_{k=1}^{K-1} g(\bar{b}_k, \bar{a}_k);$
- $\mathbb{E} [\log q(\mathbf{v})] = \sum_{k=1}^K \{\log(\mathcal{C}_{\mathbf{v}}(\bar{a}_k, \bar{b}_k)) + (\bar{a}_k - 1)g(\bar{a}_k, \bar{b}_k) + (\bar{b}_k - 1)g(\bar{b}_k, \bar{a}_k)\}$, where $\mathcal{C}_{\mathbf{v}}(\cdot)$ is the normalizing constant of a Beta distribution;
- $\mathbb{E} [\log p(\boldsymbol{\omega})] = T \sum_{k=1}^K \log(\mathcal{K}_{\boldsymbol{\omega}_k}(\mathbf{b}_0)) + \sum_{t=1}^T \sum_{k=1}^K \sum_{l=1}^L (b_0 - 1)h_l^{(t)}(\mathbf{p}_k^{(t)})$, where $\mathcal{K}_{\boldsymbol{\omega}}(\cdot)$ is the normalizing constant of a Dirichlet_L distribution;
- $\mathbb{E} [\log q(\boldsymbol{\omega})] = \sum_{t=1}^T \left[\sum_{k=1}^K \log(\mathcal{K}_{\boldsymbol{\omega}}(\mathbf{p}_k^{(t)})) + \sum_{k=1}^K \sum_{l=1}^L (p_{l,k}^{(t)} - 1)h_l(\mathbf{p}_k^{(t)}) \right];$
- $\mathbb{E} [\log p(\boldsymbol{\mu}^*, \boldsymbol{\sigma}^{2*})] = L \sum_{t=1}^T \log(\mathcal{K}_{(\mu^*, \sigma^{2*})}(m_0^{(t)}, k_0^{(t)}, c_0^{(t)}, d_0^{(t)})) - \sum_{t=1}^T (c_0^{(t)} + 1 + 1/2) \sum_{l=1}^L (\log(d_l^{(t)}) - \psi(c_l^{(t)})) - \sum_{t=1}^T d_0^{(t)} \sum_{l=1}^L c_l^{(t)} / d_l^{(t)} - \sum_{t=1}^T \frac{k_0^{(t)}}{2} \sum_{l=1}^L \left(\frac{1}{k_l^{(t)}} + \frac{c_l^{(t)}}{d_l^{(t)}} (m_l^{(t)} - m_0^{(t)})^2 \right)$, where $\mathcal{K}_{(\mu, \sigma^2)}(\cdot)$ is the normalizing constant of a Normal-Inverse Gamma distribution;
- $\mathbb{E} [\log q(\boldsymbol{\mu}^*, \boldsymbol{\sigma}^{2*})] = \sum_{t=1}^T \sum_{l=1}^L \log(\mathcal{K}_{(\mu^{*(t)}, \sigma^{2*(t)})}(m_l^{(t)}, k_l^{(t)}, c_l^{(t)}, d_l^{(t)})) - \sum_{t=1}^T \sum_{l=1}^L (c_l^{(t)} + 1 + 1/2)(\log(d_l^{(t)}) - \psi(c_l^{(t)})) - \sum_{t=1}^T \sum_{l=1}^L c_l^{(t)} - (L \cdot T)/2;$
- $\mathbb{E} [\log p(\alpha)] = \log(\mathcal{K}_{\alpha}(a_{\alpha}, b_{\alpha})) + (a_{\alpha} - 1)(\psi(s_1) - \log(s_2)) - b_{\alpha}s_1/s_2$, where $\mathcal{K}_{\alpha}(\cdot)$ is the normalizing constant of a Gamma distribution.
- $\mathbb{E} [\log q(\alpha)] = \log(\mathcal{K}_{\alpha}(s_1, s_2)) + (s_1 - 1)(\psi(s_1) - \log(s_2)) - s_1.$
- If β is random and uniformly distributed over a grid, the difference in the ELBO $\mathbb{E} [\log p(\beta) - \log q(\beta)]$ can be approximated with $\mathbb{E} [\log p(\mathbf{C})]$, which then simplifies with the ELBO update given by the variational distribution of the CC membership labels $\mathbf{C}$.

S.3.3 Recommended hyperparameter specifications

The model includes several hyperparameters that must be specified prior to estimation.

Base measure hyperparameters Recall that we assume Normal-Inverse Gamma priors for the base measures $H^{(t)}$, $t = 1, \dots, T$. For each l , the atom-specific parameters are $\theta_l^{*(t)} = (\mu_l^{*(t)}, \sigma_l^{2*(t)})$, where

$$\mu_l^{*(t)} \mid \sigma_l^{2*(t)} \sim \mathcal{N}\left(m_0^{(t)}, \frac{\sigma_l^{2*(t)}}{k_0^{(t)}}\right), \quad \sigma_l^{2*(t)} \sim \text{IG}(c_0^{(t)}, d_0^{(t)}).$$

Our recommended hyperparameter values are $(m_0^{(t)}, k_0^{(t)}, c_0^{(t)}, d_0^{(t)}) = (0, 0.01, 3, 2)$, for all t , which imply, for every l and t ,

$$\mathbb{E} [\mu_l^{*(t)}] = 0, \quad \mathbb{V} [\mu_l^{*(t)}] = 100, \quad \mathbb{E} [\sigma_l^{2*(t)}] = 1, \quad \mathbb{V} [\sigma_l^{2*(t)}] = 1.$$

We regularize the variance component to ensure that all modes in the data can be detected, avoiding overly diffuse kernels that may cover multiple modes simultaneously. Conversely, the location parameter is allowed to vary freely through a weakly informative prior. This strategy has led to satisfactory performance in nested modeling with both finite and infinite mixtures [Rodríguez et al., 2008, D’Angelo and Denti, 2026], and we adopted it in this work. Alternatively, one can follow the rationale presented in Fraley and Raftery [2007] to obtain alternative methods to regularize the variance.

Mixture weights hyperparameters Setting $\mathbf{b}_0 = 0.0001 \cdot \mathbf{1}_L$ helps promote sparse finite mixtures across the rows [Malsiner-Walli et al., 2016], a popular approach that enables the estimation of the effective number of clusters $L^* < L$ by emptying some of the L fitted components, typically chosen to be a large conservative value. For further discussion of this important distinction, see, for example, Nobile [2004], Miller and Harrison [2018], Frühwirth-Schnatter et al. [2021]. The truncation level K for the BNP-MRF prior is selected following a similar rationale. Finally, setting $a_\alpha = b_\alpha = 1$ yields the standard $\text{Gamma}(1, 1)$ prior for the Dirichlet process concentration parameter [see, for example, Blei and Jordan, 2006]. The upper bound $B = 2$ is chosen to cautiously avoid the phase transition, i.e., an abrupt change from a disordered to an ordered state (i.e., a single cluster), in the Potts model [Moore et al., 2020].

S.4 Simulation studies

In this section, we describe a series of simulation studies designed to evaluate the properties of our model. In Section S.4.1, we explain why shared atoms are preferred over common atoms. Following this, Section S.4.2 demonstrates Pose’s ability to recover biclusters. Section S.4.3 presents the results obtained by Pose on synthetic datasets presenting different levels of spatial dependence. Additionally, we benchmark these results with two state-of-the-art methods, namely UTAG [Kim et al., 2022] and BayesSpace [Zhao et al., 2021]. In Section S.4.4, we show that, when using Poseidon, clustering is driven by the shared signal rather than being corrupted by noise. Section S.4.5 reports on how the computational cost per iteration changes as J , T , and N vary.

S.4.1 Comparing common and shared atoms

A pertinent question arises regarding the practical implications of employing *common atoms* versus *shared atoms*. To provide more context, when dealing with nested BNP mixtures, we need to distinguish between distributional and observational mixtures: the former drives the model-based clustering of groups (distributional clusters), while the latter governs the grouping of observations. Each cluster of groups (in our case, a cluster of columns (CC)) is represented by an observational mixture.

Following the nomenclature introduced in D’Angelo and Denti [2026], common atoms belong to nested, fully nonparametric models where stick-breaking priors are utilized to model the observational mixtures. Conversely, shared atoms are related to “finite-infinite” models, where the random observational weights follow a finite Dirichlet distribution while the distributional weights remain nonparametric. While adopting a finite mixture over an infinite one may appear to be a modeling simplification, this choice strategically bypasses the problematic interaction between the stochastic decreasing order of stick-breaking weights and the fixed sequence of atoms across mixtures, discussed in D’Angelo and Denti [2026], Denti and D’Angelo [2025]. Indeed, in common atoms models, this interaction implies that, a priori, the initial atoms in a sequence are assigned disproportionately high weights. This feature is reflected by a high prior correlation between the random measures, and a high observational coclustering probability among observations. As discussed in Section 2.1, this phenomenon is not confined to partially exchangeable nested mixtures but extends to the separately exchangeable setting.

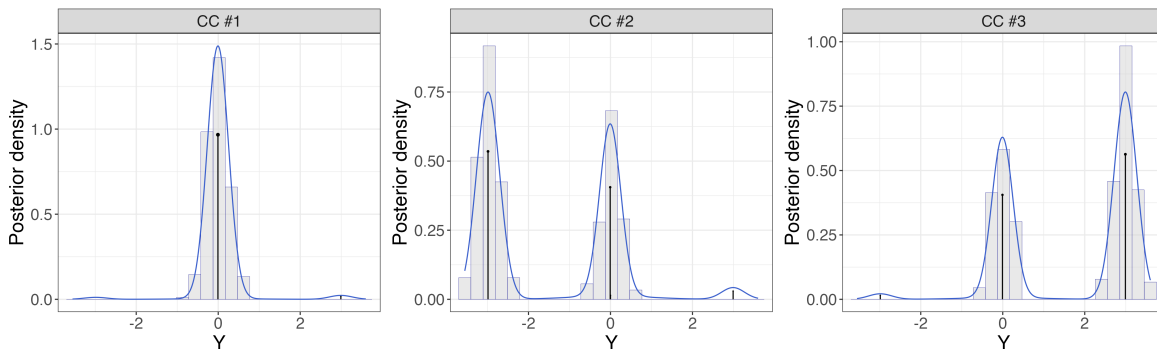
A posteriori, this characteristic can lead to unwanted “extra-sharing” of information across the atoms characterizing the distributional clusters, resulting in the estimation of “ghost modes”, i.e., spurious density peaks that do not exist in the data-generating process. Although density estimation is not the primary focus of this work, and recognizing that such spurious effects typically diminish with sufficiently large datasets, we now demonstrate how posterior density estimates can be biased when the sample size is moderate even in a very simple setting.

Indeed, we consider a synthetic dataset with 30 rows and 30 columns, structured into three column clusters (CCs) of ten columns each. For each CC, rows are independently sampled from the following simple distributions:

$$\begin{aligned} P_1 &: \mathcal{N}(0, 0.25^2), \\ P_2 &: \tfrac{1}{2}\mathcal{N}(0, 0.25^2) + \tfrac{1}{2}\mathcal{N}(-3, 0.25^2), \\ P_3 &: \tfrac{1}{2}\mathcal{N}(0, 0.25^2) + \tfrac{1}{2}\mathcal{N}(3, 0.25^2). \end{aligned}$$

The separation between the mixture components is deliberately large to clearly highlight differences in model behavior. We fit Pose models with either common or shared atoms, setting hyperparameters as previously explained, $L = 20$, $K = 15$, $\varepsilon = 10^{-8}$, and 100 random initializations. After variational estimation of the hyperparameters, we compute posterior expectations in closed form for the Dirichlet and stick-breaking weights - using the posterior

CC posterior density estimate with common atoms



CC posterior density estimate with shared atoms

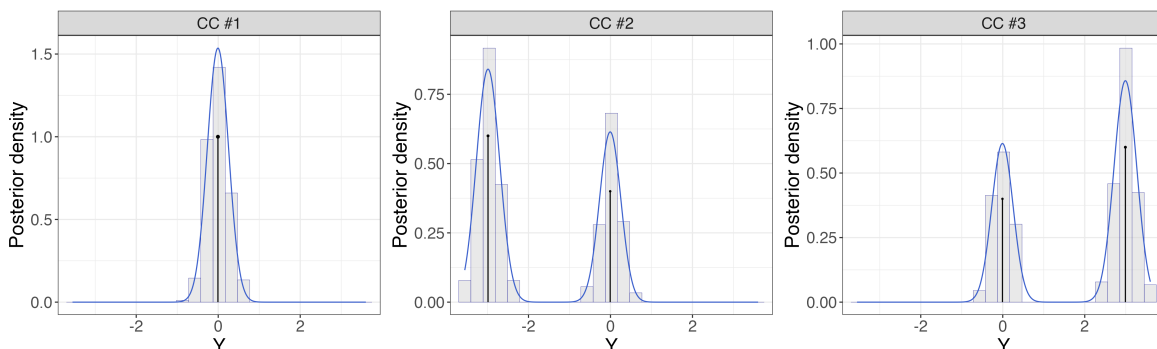


Fig. S.6: Variational posterior density estimates for the three CCs are shown. The top row corresponds to the Pose model with common atoms, whereas the bottom row displays results obtained with shared atoms. When common atoms are used, mixture components detected in one cluster tend to propagate to others, resulting in spurious modes. This behavior disappears under the shared-atom specification, which yields cleaner and more interpretable density estimates.

means of the associated Beta random variables - as well as for the Normal-Inverse Gamma atoms. This yields variational posterior density estimates of the CC-specific Normal mixture models underlying the data. Given the simplicity of the data-generating process, the CCs are perfectly recovered under both specifications¹. For each model, we report the density estimates corresponding to the run achieving the highest ELBO among the 100 initializations in Figure S.6. The top and bottom rows of the figure display results obtained with common and shared atoms, respectively. When common atoms are employed, strong dependence is induced across mixture densities, leading to the appearance of spurious modes in CCs where they are not supported by the data - simply because the corresponding atom is active in another mixture. This issue is effectively solved when shared atoms are used, yielding cleaner and more interpretable density estimates.

S.4.2 Biclustering recovery

One of the key features that distinguishes Pose from most biclustering methods is its ability to recover different row-clustering structures within different CCs. In contrast, most block-clustering approaches [see, e.g., Tan and Witten, 2014, Govaert and Nadif, 2013] assume a single row-clustering structure shared across all CCs. This flexibility is particularly important for MALDI-MSI data, where we expect peaks to occur at different m/z (rows) in different CCs, making traditional block-clustering methods poorly suited to the task.

¹With fewer observations (e.g., ten rows), only the finite-infinite specification successfully recovers the partition; these results are omitted for brevity but are available in the accompanying code.

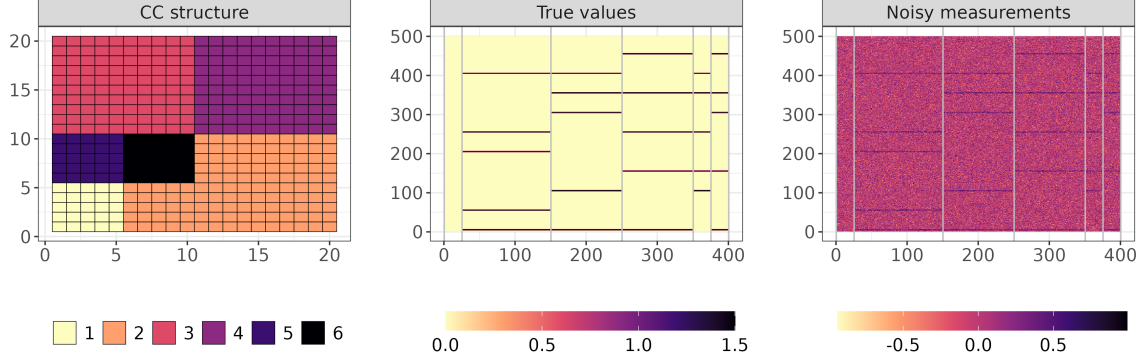


Fig. S.7: *Left*: the column partition structure used for the simulation study. *Center*: an example of the biclustering structure (with row cluster allocations within each CC), for all rows and all columns (ordered and labeled by their CCs). The colors represent the mean activation for each element. *Right*: an example of the data generated given the biclustering allocation (for all rows and columns). Colors correspond to the intensity activation level of each m/z in the corresponding pixel, and columns have been sorted by their CCs.

To assess Pose’s performance in recovering biclustering structure, we use synthetic datasets designed to mimic the characteristics of MALDI data. In these datasets, certain m/z values appear as peaks with high activation levels (mean = 1.5), their immediate neighbors show medium activation (mean = 0.75), and all other rows are treated as noise (mean = 0). Peak locations vary across CCs, and we simulate a moderately high noise level (standard deviations = 1).

Synthetic data generating process We generate 50 synthetic datasets sharing the same column partition. Specifically, we consider a grid of 20×20 pixels, with a fixed partition structure (shared by all synthetic datasets) that divides the pixels into six clusters, each containing 25, 25, 25, 100, 100, and 125 contiguous pixels, respectively. A visual representation of the partition structure is provided in the left panel of Figure S.7.

We consider 500 rows, each corresponding to a hypothetical m/z level. To generate the m/z intensity level and create a biclustering structure, for each CC, we assign a number of rows to a group of “high-activation” m/z values and other rows to a group of “medium-activation” m/z values. The remaining rows are classified as “noise”. These three categories correspond to the row clusters that are generated for each CC. The rows assigned to the “high-activation” m/z , which correspond to the activation peaks, are selected randomly, so that there are five groups of peaks, each corresponding to 2 adjacent rows. Note that different CCs display different sets of “high-activation” m/z , and by default, the first CC is designed to have no “high-activation” m/z . Given the group of rows assigned to the “high-activation” m/z , the “medium-activation” m/z are selected by considering the rows immediately above and below the “high-activation” m/z . Given each row’s allocation to these groups, we generate the activation levels from a Gaussian distribution with means of 1.5, 0.75, and 0 for the high-activation, medium-activation, and noise groups, respectively, and a standard deviation of 1. A visual example of the row clustering allocation is provided in the center panel of Figure S.7, where the matrix entries have been reordered by the CC assignment. The colors correspond to the mean intensity associated with each m/z (row) in the corresponding pixel (column). The right panel of Figure S.7 instead visualizes the entire data matrix for one instance of the data generation process. Colors represent the activation level of each m/z in the corresponding pixel.

Model specification and competing alternatives We compare the biclustering performance of: (a) Pose (fixing at $\beta = 1$), (b) `sparseBC` [Tan and Witten, 2014] with the true number of CCs K (`trueK`), (c) `sparseBC` with K selected via its internal procedure (`chooseKR`), and (d) `Double k-means`, which first applies k-means to columns and then, within each CC, applies k-means to the corresponding rows. More in detail, given the spatial structure of the chosen column clustering partition (see Figure S.7), we fit the Pose model with a strong spatial prior fixing $\beta = 1$. We set the upper bound on the number of clusters as $L = 10$ and $K = 10$. The hyperparameters are set across all l as follows: $(m_0, k_0, c_0, d_0) = (0, 0.01, 3, 2)$, $\mathbf{b}_0 = 0.0001 \cdot \mathbf{1}_L$, and $a_\alpha = b_\alpha = 1$. We run the CAVI algorithm from 50 different starting points and retain only the one that reaches the highest ELBO value.

The other biclustering models we consider do not have a spatial component, but can still be fit to spatial data. The `sparseBC` algorithm [Tan and Witten, 2014] was run both fixing the number of CCs K to the true value 6 (`trueK`) and using their function `sparseBC.chooseKR` that performs cross-validation to select the number of column and row clusters (`chooseKR`). In both cases, we set the regularization parameter λ to 10. The `Double k-means` algorithm was designed to mimic the main ideas behind the Pose model. First, the k-means algorithm is run on the columns of the matrix, and the number of clusters is chosen using the gap method (using the `cluster::clusGap` function in R [Tibshirani et al., 2001, Maechler et al., 2025]). Given the selected CCs, the average intensity is computed for each m/z across all columns belonging to the same CC. Another k-means algorithm is run on these average intensity values, determining the row cluster allocation for each CC. The number of clusters is again selected using the gap method.

Results To compute the Adjusted Rand Index (ARI) between the true and estimated biclustering structures, the biclustering labels are vectorized, transforming the cluster allocation matrix into vectors for both the true and estimated allocations. Then, the ARI is applied between these vectors. To evaluate the biclustering RMSE, when the mean values associated with each cluster are not estimated by the algorithm, we compute the empirical average of the activation levels in each estimated bicluster and compare this with the generated mean activation levels (ground truth) to calculate the RMSE.

Figure S.8 shows recovery of the true CC and biclustering structure, measured by the Adjusted Rand Index (ARI), as well as the root mean squared error (RMSE) in recovering the biclustering matrix. Overall, Pose delivers the best performance, with median ARI values close to 1 and the lowest RMSE. `sparseBC` performs substantially worse, with a median ARI of around 0.5 for CC recovery and larger RMSE values, reflecting the limitations of its block-clustering assumption. Surprisingly, `Double k-means`, despite its greater flexibility compared to `sparseBC`, yields the poorest results.

S.4.3 Varying degree of spatial dependence structure and comparison with alternative methods

Here, we evaluate the robustness of the Pose model to datasets with varying degrees of spatial dependence. Thanks to the possibility of estimating, albeit approximately, the inverse temperature β , Pose’s clustering and biclustering recovery performance should generally be robust to different degrees of spatial dependence in the latent partition.

Synthetic data generating process The latent spatial partition of pixels into clusters is randomly generated using the Ewens-Pitman Attraction (EPA) distribution [Dahl et al., 2017], chosen for its interpretability, and the simplicity with which samples can be drawn. In fact, this distribution maintains some important properties of the popular Ewens-Pitman distribution over partitions [Pitman and Yor, 1997], such as the asymptotic growth of the number of clusters, while imposing a spatial dependence on the clusters. Because of its definition relying

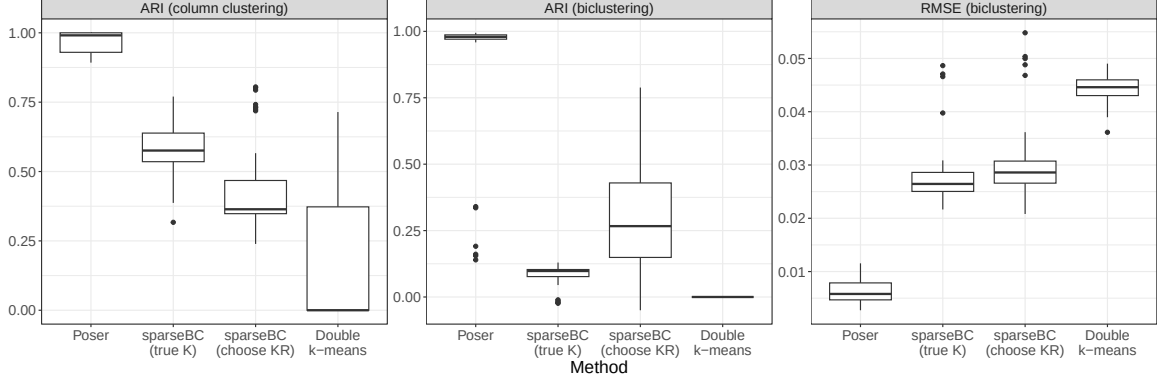


Fig. S.8: Recovery of column partition (left panel) and biclustering structure (central and right panel), measured by ARI (left and center) and RMSE (right) in synthetic data.

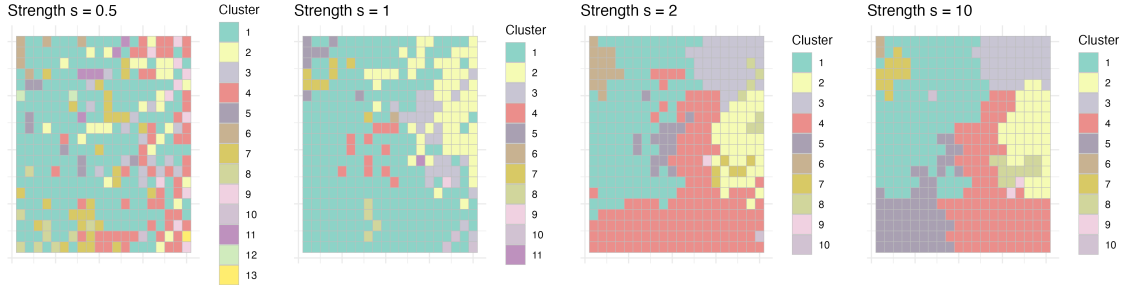


Fig. S.9: Instances of latent partitions drawn from the EPA distribution for different values of s .

on a sequential allocation of pixels to their clusters, sampling from it is much easier compared to other distributions; moreover, the parameter regulating the spatial dependence is easily interpretable. Specifically, the EPA distribution relies on a similarity $\lambda(i, j)$ between any two pixels i and j , and we use one of the default recommendations $\lambda(i, j) = \exp(-s \cdot d(i, j))$. To incorporate a queen adjacency structure, we choose the distance $d(i, j) = \|(x_i, y_i) - (x_j, y_j)\|_\infty$, where (x_i, y_i) denotes the coordinate for pixel i and $\|\cdot\|_\infty$ is the supremum norm. We vary the strength parameter $s \in \{0.5, 1, 2, 10\}$. See Figure S.9 for an example of a latent partition drawn from the EPA for different values of s . Given the latent partition determining the column-clusters, the m/z intensity level is generated as in Section S.4.2.

Model specification and competing alternatives We evaluate how the column-clusters and biclustering recovery of Pose changes for different values of spatial dependence, as regulated by s . At the same time, we also compare Pose’s segmentation recovery performance with other methods that are commonly used in spatial transcriptomics scenarios. While not specifically designed for MALDI-MSI data, these methods are used to infer latent spatial partitions of the data, and can be adapted to our setting. In particular, we focus on comparing with BayesSpace [Zhao et al., 2021] and UTAG [Kim et al., 2022]. When applying these methods to our synthetic MALDI data, we treat the activation levels at different m/z values in a certain pixel as the expression of different genes at a given spot.

Note that UTAG relies on some parameter specifications: (1) a maximum distance to determine which pixels fall in the spatial neighborhood of other pixels, and for this we choose `max_dist` = 1.5, which is equivalent to using a Queen adjacency structure; (2) a resolution parameter r for the Leiden clustering algorithm, which regulates the number of clusters detected, and we run the algorithm multiple times with different values of $r \in \{0.2, 0.3, 0.4, 0.5\}$ and only keep the result achieving the best performance across these values. BayesSpace was

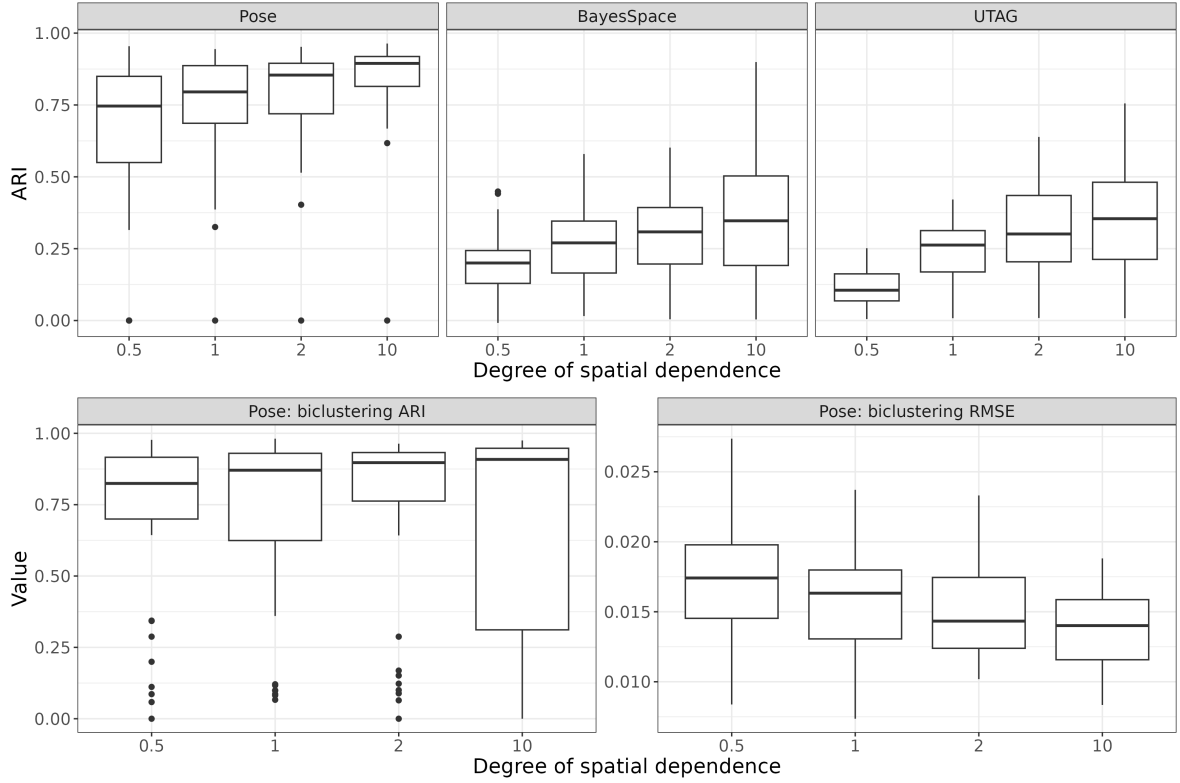


Fig. S.10: Results on 40 simulated datasets with varying spatial dependence. Top row: ARI for the column partition (image segmentation) obtained by the competing methods, stratified by the spatial dependence level. Bottom row: performance of Pose in terms of biclustering ARI and RMSE.

run with default parameters, using the true number of clusters.

Results Each method has been evaluated on 40 simulated datasets. Figure S.10 reports the results. The top row compares the ARI values obtained by the different methods for the column partition (which corresponds to the image segmentation), stratified by the level of spatial dependence. As expected, since the data are generated from a process that differs from the assumed models, none of the methods achieves perfect performance. However, Pose, which allows the spatial strength to adapt through the estimation of the inverse temperature parameter, provides the best results across all scenarios.

The performance of Pose can be further examined in the second row of Figure S.10, which displays the distribution of the biclustering ARI and RMSE values. The row partition is satisfactorily recovered in all cases. When $s = 10$, the ARI is slightly lower (indicating less regularization with respect to the true partition), and, in turn, it is accompanied by a smaller RMSE.

S.4.4 Informative datasets guide the shared clustering estimation

A desired feature of Poseidon is that the shared column partition should be driven primarily by datasets with informative observations rather than being negatively affected by those potentially dominated by noise. To investigate this property, we compare the performance of Pose and Poseidon models, without spatial interaction, on specifically designed datasets.

Synthetic data generation We consider five datasets, denoted as $\mathbf{Y}^{(t)}$, each comprising $J = 20$ columns and $N^{(t)} = 50$ rows, with $t = 0, 1, 2, 3, 4$. A common biclustering structure

is assumed across these datasets. As ground truth, we assume the presence of two CCs, each consisting of 10 columns. Specifically, the first CC contains 25 rows, with cell entries sampled from a Normal distribution with mean $\mu^{(t)}$ and variance 1. The remaining 25 rows in this cluster are generated from a standard Gaussian distribution. The second CC contains a single RC, and all cell entries are sampled from a standard Gaussian distribution. Given this setting, we first generate a reference dataset, $\mathbf{Y}^{(0)}$, with $\mu^{(0)} = 4$. This case represents a scenario where a clear separation exists among the row clusters and, consequently, among the CCs. For the remaining datasets, $\mathbf{Y}^{(t)}$ with $t = 1, 2, 3, 4$, we set $\mu^{(t)} \in \{1, 0.75, 0.50, 0.25\}$. In these datasets, the magnitude of $\mu^{(t)}$ controls the signal-to-noise ratio: as the index t grows, the two RC distributions increasingly overlap and noise dominates, complicating the recovery of the true partitions. A visual example of these datasets is reported in Figure S.11.

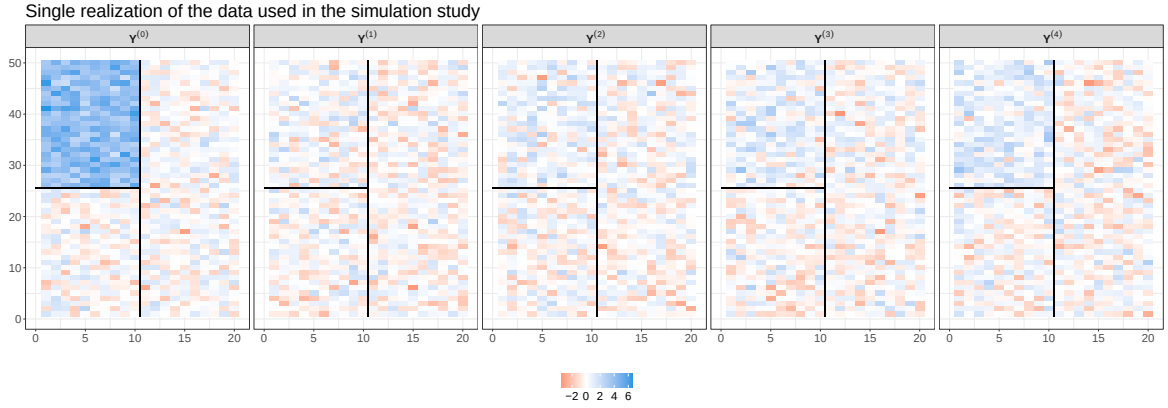


Fig. S.11: Example of realization of a collection of matrix datasets used in the second simulation study.

Hyperparameters and results The hyperparameters for the models are set, for $t = 1, \dots, T$ as $(m_0^{(t)}, k_0^{(t)}, c_0^{(t)}, d_0^{(t)}) = (0, 0.01, 3, 2)$, $\mathbf{b}_0 = 0.0001 \cdot \mathbf{1}_L$, and $a_\alpha = b_\alpha = 1$. Additionally, we set $L = 30$ and $K = 15$, and $\beta = 0$. We consider the best CAVI run out of 50 random initializations. Each experiment is simulated 100 times to assess Monte Carlo variability in the results.

The results of the Pose model across all datasets are summarized in the top two rows of Table S.1, where we report the averages (and standard deviations) over 100 replicates of the ARIs computed between the ground truth and the estimated column and row partitions, respectively. While both partitions are largely recovered for $\mathbf{Y}^{(0)}$ and $\mathbf{Y}^{(1)}$, performance deteriorates sharply for the remaining datasets as the signal-to-noise ratio decreases. We then estimate a multi-dataset model with $T = 2$, considering four “collections” of datasets formed by combining the reference dataset with each noisy dataset: $\{\mathbf{Y}^{(0)}, \mathbf{Y}^{(t)}\}$ for $t = 1, \dots, 4$. The results, summarized in the central rows of Table S.1, report the ARI indices for the CCs and the two distinct RCs estimated on the reference and noisy datasets, respectively. First, we note that including $\mathbf{Y}^{(0)}$ drives the shared CC solution, effectively mitigating the noise in the other datasets. Second, the estimation of the RCs for the noisy datasets shows significant improvement over single-dataset models, without affecting results on the reference dataset. Thus, the multi-dataset approach has an advantage in recovering the underlying structure.

Integrating data via stacking The observed improvement raises an important question. Since we are working with multiple datasets sharing the same number of columns, one may ask whether similar results can be obtained with a simpler strategy: applying the single-dataset Pose model to tables constructed by row-stacking the reference dataset with one of

Pose	$\mathbf{Y}^{(1)}$		$\mathbf{Y}^{(2)}$		$\mathbf{Y}^{(3)}$		$\mathbf{Y}^{(4)}$	
CC	0.957	(0.090)	0.509	(0.423)	0.034	(0.139)	0.000	(0.000)
RC	0.848	(0.116)	0.355	(0.335)	0.016	(0.055)	0.000	(0.000)
Poseidon	$\{\mathbf{Y}^{(0)}, \mathbf{Y}^{(1)}\}$		$\{\mathbf{Y}^{(0)}, \mathbf{Y}^{(2)}\}$		$\{\mathbf{Y}^{(0)}, \mathbf{Y}^{(3)}\}$		$\{\mathbf{Y}^{(0)}, \mathbf{Y}^{(4)}\}$	
CC	1.000	(0.000)	1.000	(0.000)	0.994	(0.029)	0.996	(0.028)
RC Ref.	1.000	(0.000)	0.995	(0.047)	0.995	(0.047)	0.995	(0.047)
RC Noisy	0.858	(0.108)	0.716	(0.109)	0.350	(0.196)	0.098	(0.127)
Stacked	$[\mathbf{Y}^{(0)}, \mathbf{Y}^{(1)}]$		$[\mathbf{Y}^{(0)}, \mathbf{Y}^{(2)}]$		$[\mathbf{Y}^{(0)}, \mathbf{Y}^{(3)}]$		$[\mathbf{Y}^{(0)}, \mathbf{Y}^{(4)}]$	
CC	1.000	(0.000)	1.000	(0.000)	0.994	(0.031)	1.000	(0.000)
RC Aggr.	0.855	(0.056)	0.763	(0.075)	0.507	(0.150)	0.475	(0.046)
RC Ref. rows	0.967	(0.052)	0.935	(0.087)	0.801	(0.212)	0.948	(0.148)
RC Noisy rows	0.860	(0.093)	0.692	(0.114)	0.283	(0.207)	0.027	(0.078)

Tab. S.1: Averages and standard deviations of the adjusted Rand indices obtained comparing the ground truth and the estimated column and row partitions on the noisy datasets using Pose, Poseidon, and stacking the reference and noisy datasets (in all models, $\beta = 0$). For comparison, the indices for the RC and CC solutions estimated on the reference dataset $\mathbf{Y}^{(0)}$ using Pose are always equal to 1.

the noisy datasets. We denote these datasets as $[\mathbf{Y}^{(0)}, \mathbf{Y}^{(t)}]$, for $t = 1, \dots, 4$. The clustering performances are summarized in the last four rows of Table S.1. We observe that the CC solutions exhibit robustness, albeit with a slight under-performance compared to those obtained with Poseidon. In contrast, contamination affects the joint RC configurations. As expected, RC estimation improves compared to the one on the individual noisy dataset. However, especially when evaluated after unstacking the two datasets, this comes at the cost of compromising the recovery of the true partition in the reference dataset.

S.4.5 Computational costs

A key question concerns the scalability of the proposed method. Assessing the time required for convergence is not straightforward, as the performance of Poseidon may depend on several factors. The most relevant ones are the characteristics of the dataset collection $\mathbf{Y} = \{\mathbf{Y}^{(t)}\}_{t=1}^T$, namely the number of datasets T , the common number of columns J , and the dataset-specific number of rows $N^{(t)}$. Computational time is also affected by the initialization, the number of mixture components L and K , and the convergence threshold ε . To effectively reduce some of the degrees of freedom of this problem, we decide to use the dataset $\mathbf{Y}^{(0)}$ presented in the simulation study of Supplementary Section S.4.4, and we fix $L = 30$ and $K = 15$. The hyperparameter specification follows the one presented in the corresponding simulation study. Moreover, to eliminate the effect of the stopping rule, we set $\varepsilon = 0$ and terminate each run after 10 iterations. To evaluate the impact of initialization, we consider 50 different starting points.

Figure S.12 reports boxplots of the average computational time per iteration (averaged over the ten iterations) of the variational inference algorithm for varying $T \in \{1, 2, 3, 4\}$, $J \in \{100, 200, 500\}$, and $N^{(t)} = N \in \{50, 100, 200, 500\}$. We observe that increasing T and J has a relatively modest impact on computational cost. Notably, the code used to fit the model is specifically designed to scale efficiently with respect to J . In contrast, the number of rows N has the largest effect on runtime: across nearly all panels, the computational time

exhibits an approximately linear increase on the $\log_{10}$ scale.

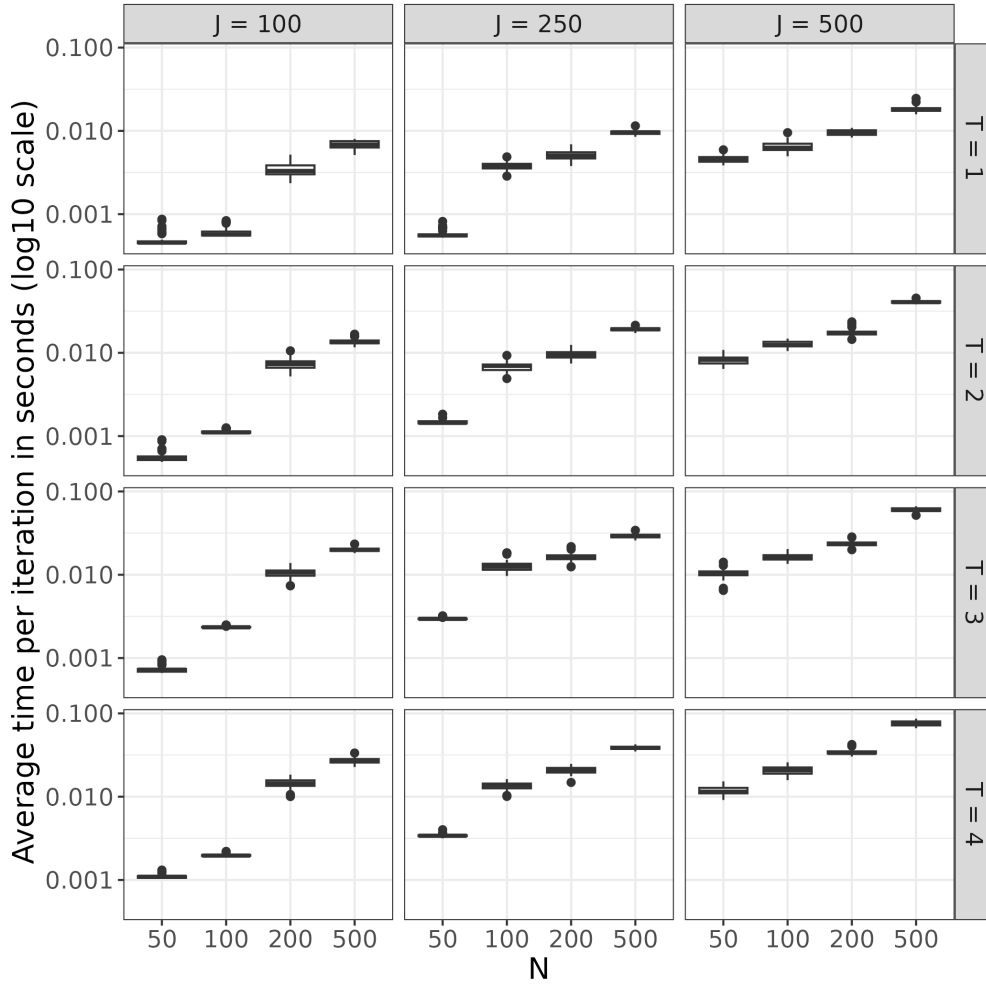


Fig. S.12: Evolution of the computational time (in seconds, $\log_{10}$ scale) required to perform a single step of the Poseidon variational inference algorithm, averaged over ten iterations. Boxplots summarize the distribution of computational times obtained from 50 different starting points. Each row of the figure corresponds to a different number of datasets T in the collection, while each column represents a different number of dataset columns J . Within each panel, multiple values of the number of dataset rows N are considered.

S.5 Additional results of Poseidon on ccRCC tissue

S.5.1 Additional insights about important analytes of interest

In the main paper, we demonstrated how the image segmentation estimated by Poseidon effectively captures distinct multiomics clusters, reflecting the complex histopathological landscape of the kidney tumor resection previously reported in [Denti et al. \[2022\]](#). This segmentation mirrors known tissue heterogeneity and provides a framework for identifying region-specific molecular signatures. Pinpointing the most influential molecules in each area could offer valuable insights for biomedical research, particularly in uncovering potential diagnostic markers and therapeutic targets. As a proof of concept, in Section 5, we focused on three representative clusters: a benign region (central row of Figure 4, CC-12), a tumor area adjacent to the capsule (top row, CC-10), and a tumor region with hemorrhagic features (bottom row, CC-6). As previously discussed, Poseidon enables us to summarize and quantify the contribution of each analyte across the three molecular classes, studying the row partitions within each CC. These results confirm that lipids, N-glycans, and peptides convey complementary biological information.

In contrast to approaches that analyze each molecular class independently, our method integrates all modalities simultaneously, enabling direct comparison of signal intensities across distinct molecular families. For instance, in Figure S.13, we illustrate the spatial distribution of three representative molecules associated with specific tissue regions. In particular, given the summary provided in Figure 4, we were able to focus on a specific lipid with m/z 606.06, which marks the hemorrhagic area; a specific N-glycan with m/z 1905.67 (putatively identified as Hex9HexNAc2), which highlights the high-grade tumor (*grade 2* and *grade 3*) within the tumor nodule; and a specific peptide with m/z 1116.56 (putatively identified as Heterochromatin Protein 1-Binding Protein 3), which is localized in the cortex surrounding the tumor nodule.

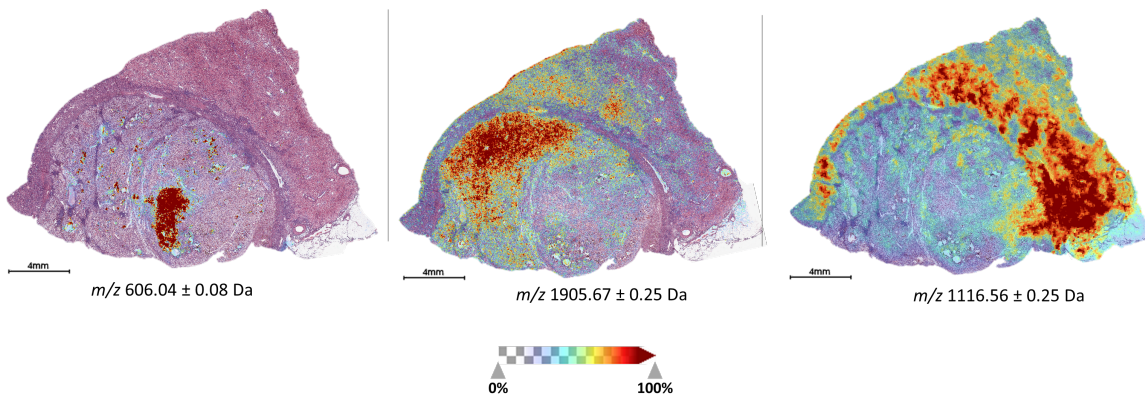


Fig. S.13: Spatial distribution of three selected m/z features across kidney tissue in the three molecular classes. From left to right: lipid (m/z 606.06), N-glycan (m/z 1905.67), and peptide (m/z 1116.56).

S.5.2 Results for all the estimated column clusters

We complement the results presented in the main paper by reporting the complete set of findings for the ccRCC tissue. Figures S.14, S.15, and S.16 illustrate the goodness of fit of the CC-specific density estimates obtained from nested mixture models, with the corresponding portion of the image segmentation displayed in the first column of each panel. The remaining columns show the fitted distributions of the transformed abundances for lipids, N-glycans, and peptides. Overall, the model demonstrates a strong ability to adapt to the data. Note that the

same atoms are shared across dataset-specific distributional clusters. Moreover, inspection of the leftmost panels, which decompose the image segmentation into fifteen components, reveals no evidence of singleton or very small clusters—features that would have complicated the interpretation of the results. For completeness, Table S.2 showcases the number of pixels assigned to each CC.

Additionally, we display the medians in each estimated RC for every molecular signal for all the recovered CCs in Figures S.17, S.18, and S.19. Each dot, representing the median transformed abundance, is colored according to the RC assignment. Finally, Figures S.20, S.21, and S.22 report the estimated posterior means representing each RC in all the recovered CCs. We conclude this section by commenting on these results.

CC #	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
Freq.	130	160	196	213	250	210	169	563	196	613	156	660	172	267	168

Tab. S.2: Frequencies of the estimated CCs in the real data analysis.

With these different plots, we can examine how the RC distributions behave within each molecular modality (lipids, N-glycans, and peptides) across pixels assigned to different CCs, and how these patterns relate to the underlying tissue regions. We observe that some CCs share very similar RC distributions within the same molecular family, indicating tissue areas with comparable internal molecular organization. For instance, CC-12 and CC-13 correspond to benign and healthy tissue, including the healthy region surrounding the tumor capsule. In these areas, the RC structure is relatively stable across modalities, consistent with the expectation that molecular organization is better preserved in healthy tissue.

Within the tumor region, a distinct yet structured behavior emerges. In particular, CC-6 and CC-7 correspond to two biologically meaningful tumor subregions: CC-6 is associated with a hemorrhagic area, whereas CC-7 corresponds to the tumor capsule region, which also contains blood cells. In these clusters, we observe that lipids and N-glycans show similar RC distributions, suggesting that these two modalities may capture related molecular changes in these tumor environments. In contrast, CC-8, CC-9, and CC-10 correspond to different portions of the tumor nodule characterized by increasing heterogeneity. In this case, lipids maintain relatively similar RC structures across these CCs, while peptides display a marked shift in CC-9, which is associated with regions showing inflammatory features. This behavior is biologically plausible, as peptide signals are expected to be particularly sensitive to inflammation-related processes.

More broadly, these patterns suggest that RC structures are more stable across CCs in healthy regions than in diseased or inflamed areas. This also helps explain a feature observed in the single-modality segmentations: clusters obtained separately within each molecular family tend to align more consistently in healthy tissue, whereas in tumor and inflammatory regions, the segmentations become less overlapping, reflecting modality-specific responses to the underlying biological changes. In this context, the joint multiomics modeling provided by Poseidon enables the integration of these complementary signals, yielding a more coherent clustering of tissue regions driven by the RC structures induced by the different modalities.

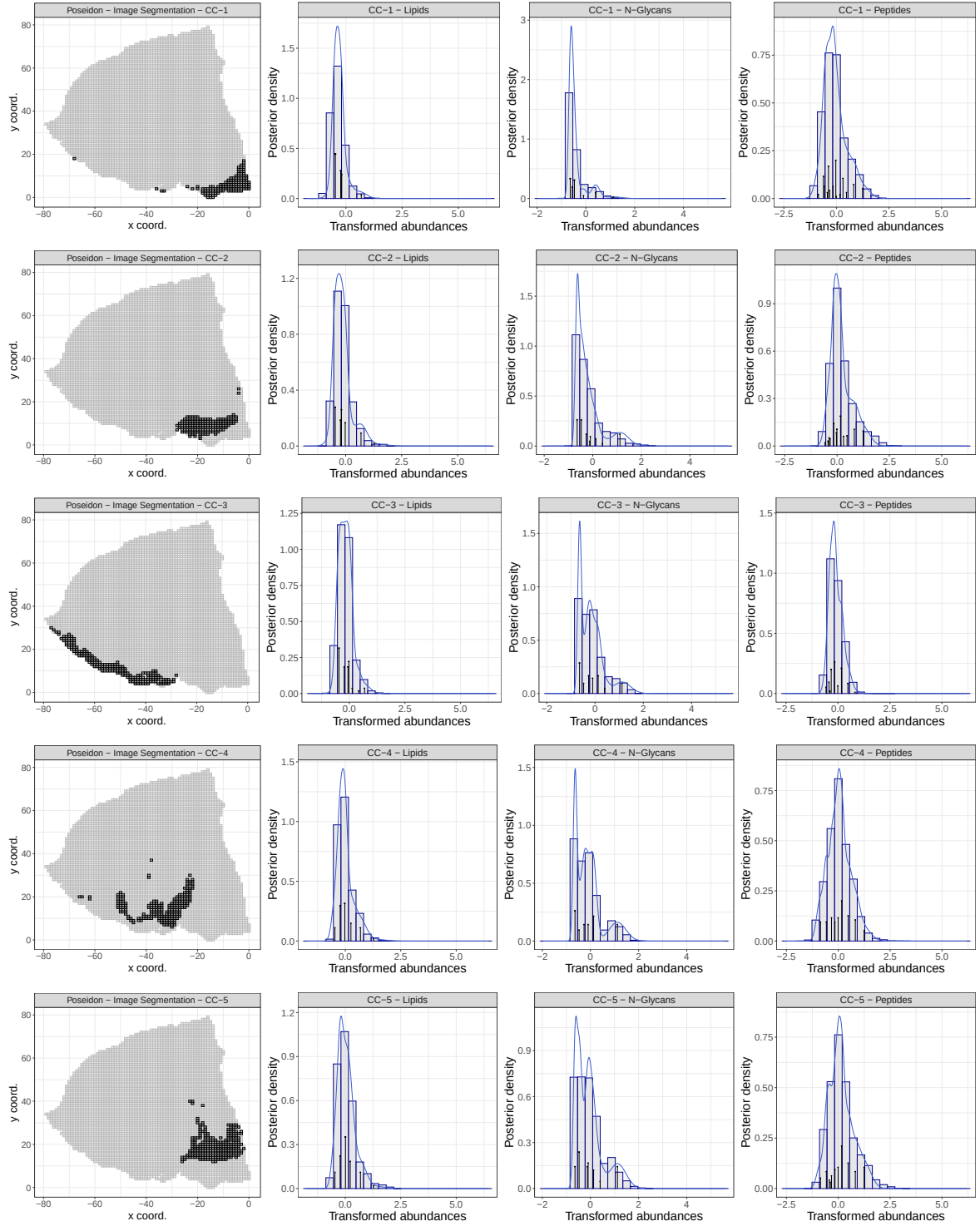


Fig. S.14: Posterior density estimates of the transformed molecular abundances (columns 2–4) for CCs 1–5. The corresponding pixels associated with each cluster are highlighted in the first column.

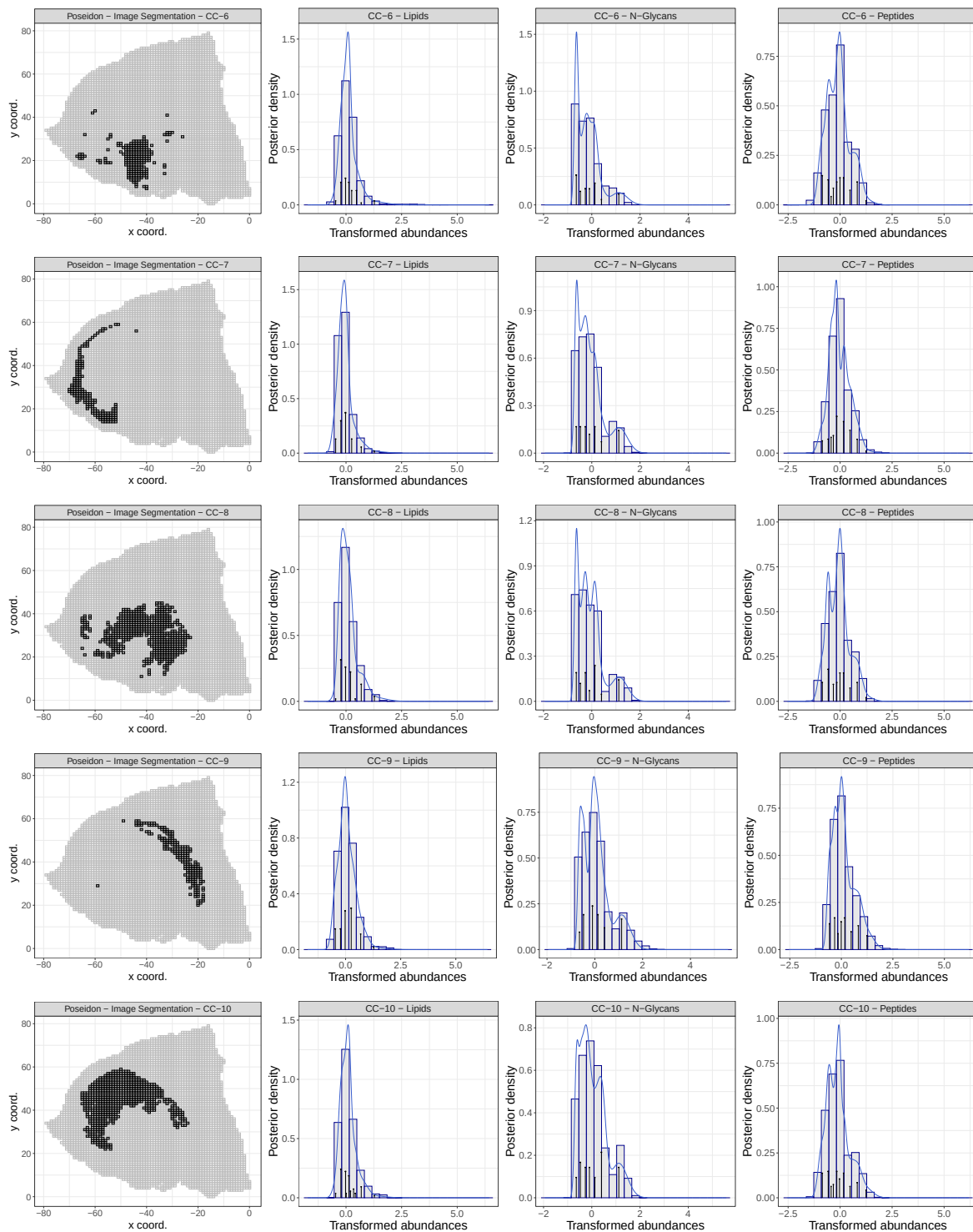


Fig. S.15: Posterior density estimates of the transformed molecular abundances (columns 2–4) for CCs 6–10. The corresponding pixels associated with each cluster are highlighted in the first column.

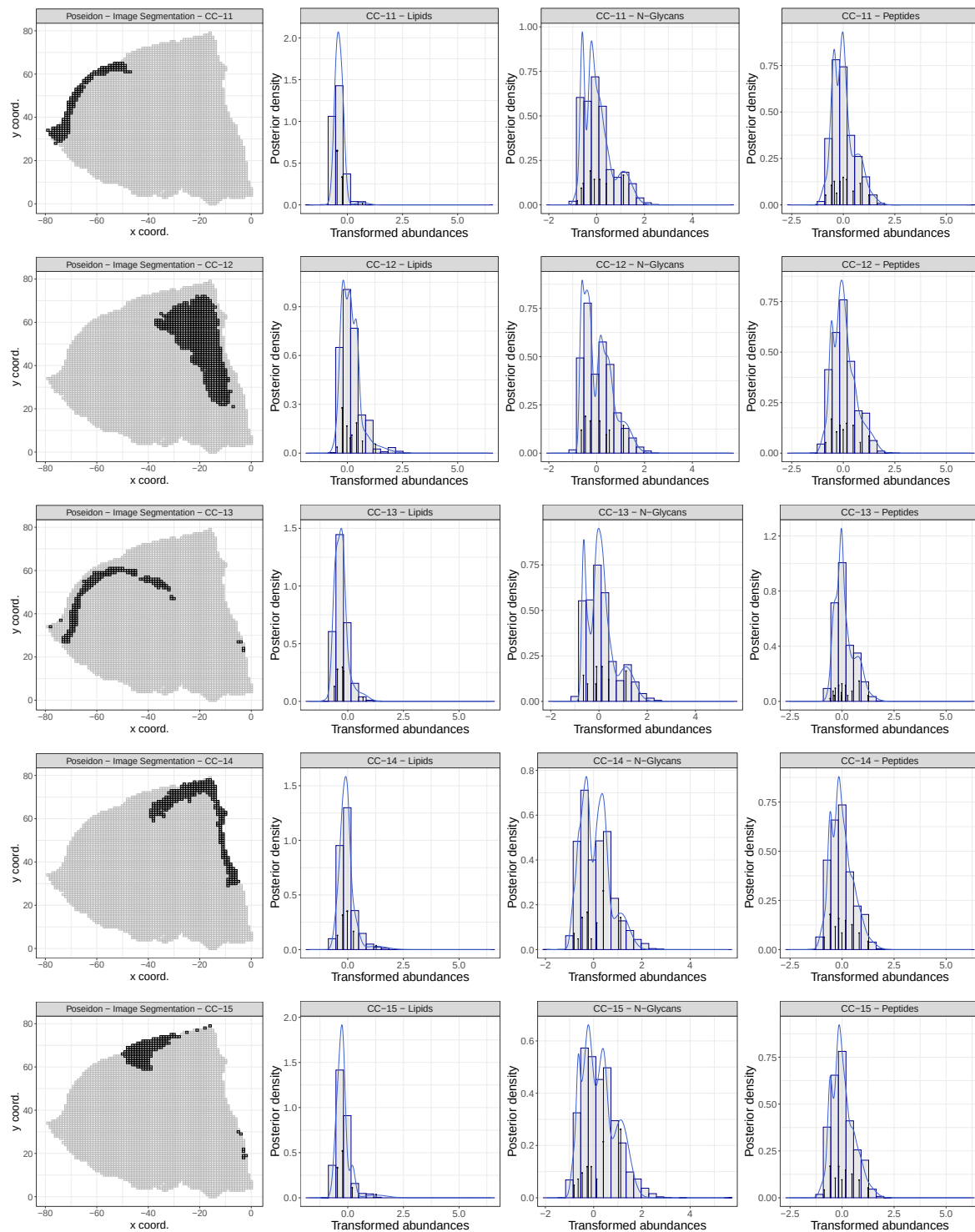


Fig. S.16: Posterior density estimates of the transformed molecular abundances (columns 2–4) for CCs 11–15. The corresponding pixels associated with each cluster are highlighted in the first column.

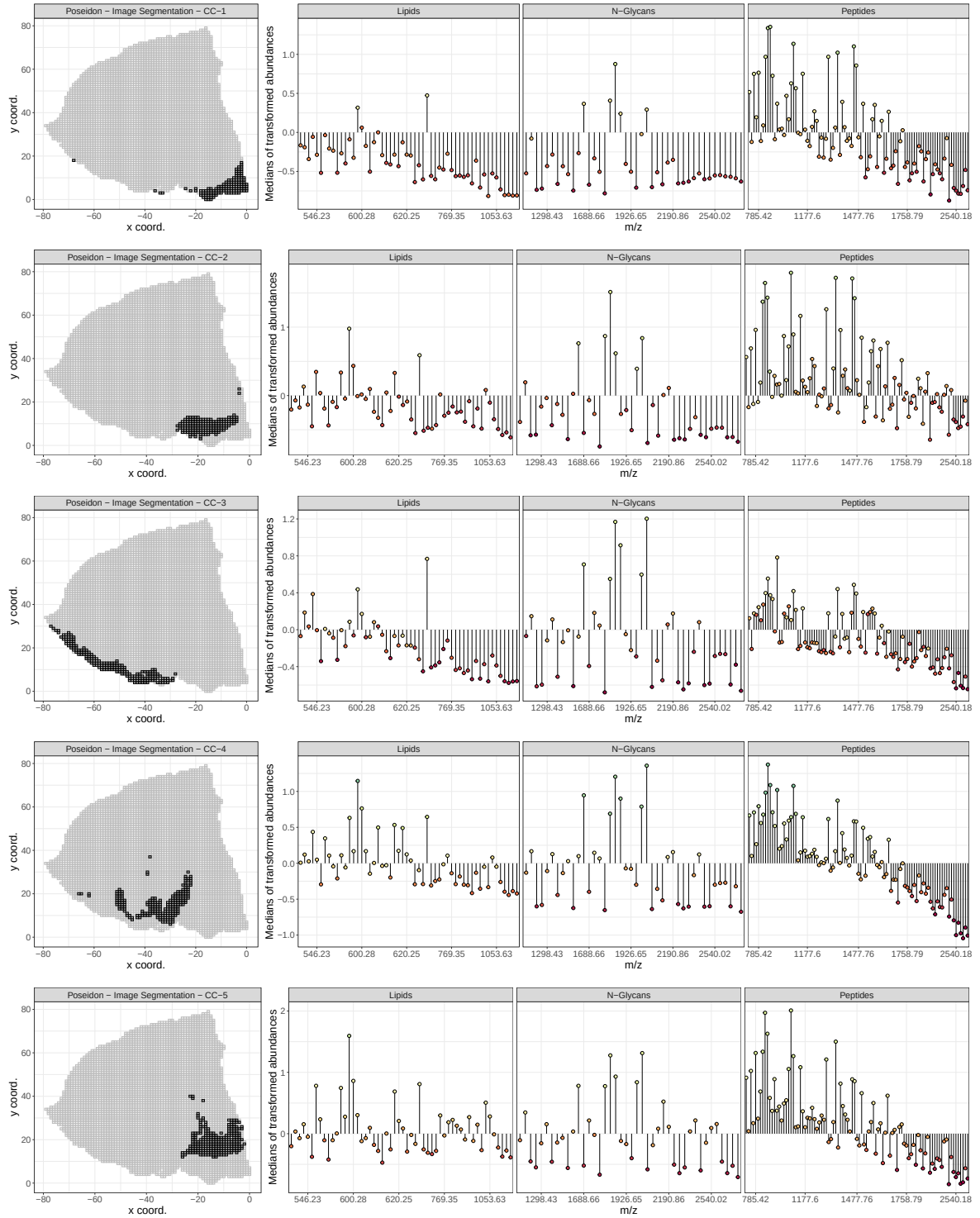


Fig. S.17: Visual representation of the sample medians of transformed abundances across CCs 1–5. Panels in columns 2–4 display these medians colored according to their estimated RC. The corresponding pixels associated with each cluster are highlighted in the first column.

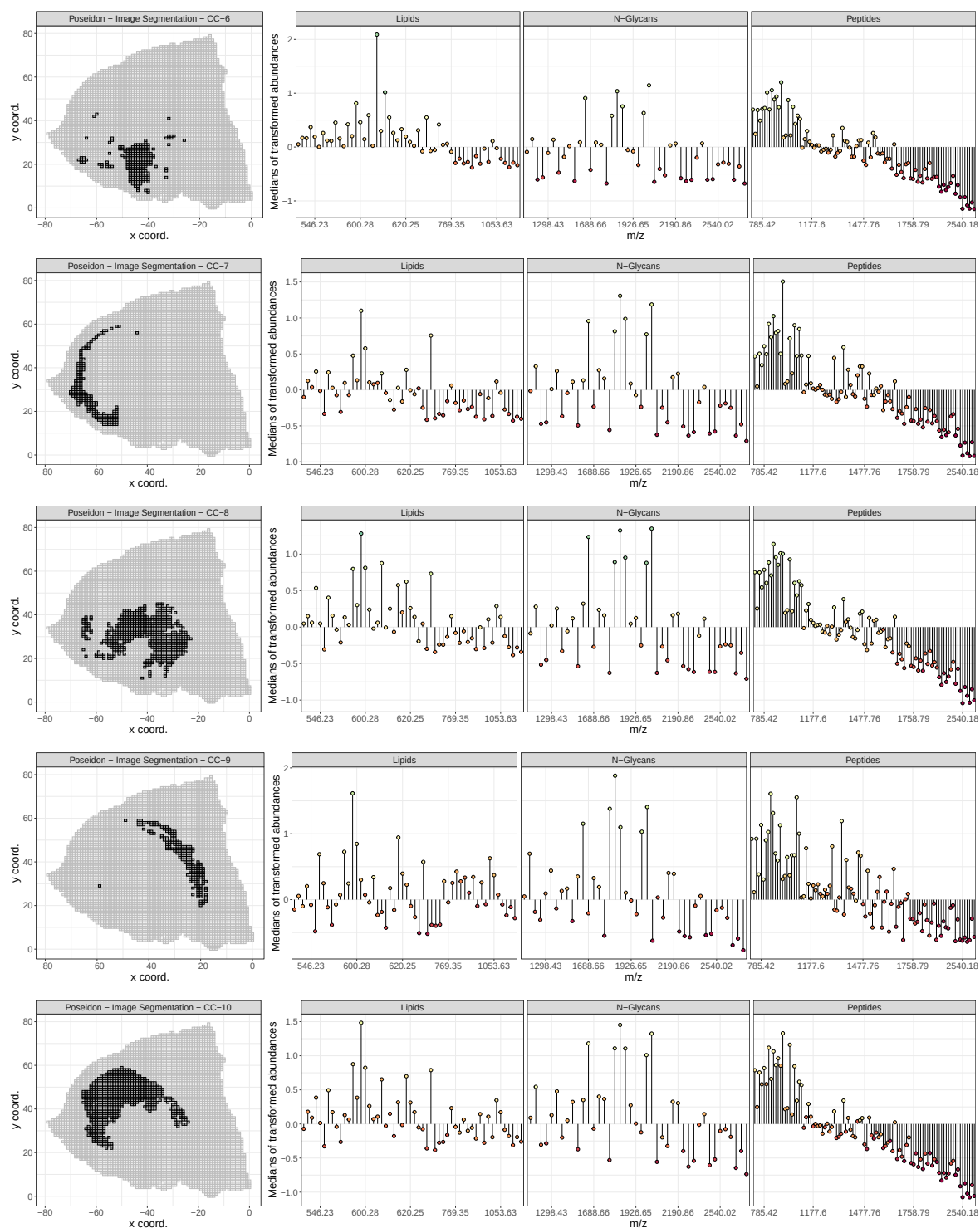


Fig. S.18: Visual representation of the sample medians of transformed abundances across CCs 6–10. Panels in columns 2–4 display these medians colored according to their estimated RC. The corresponding pixels associated with each cluster are highlighted in the first column.

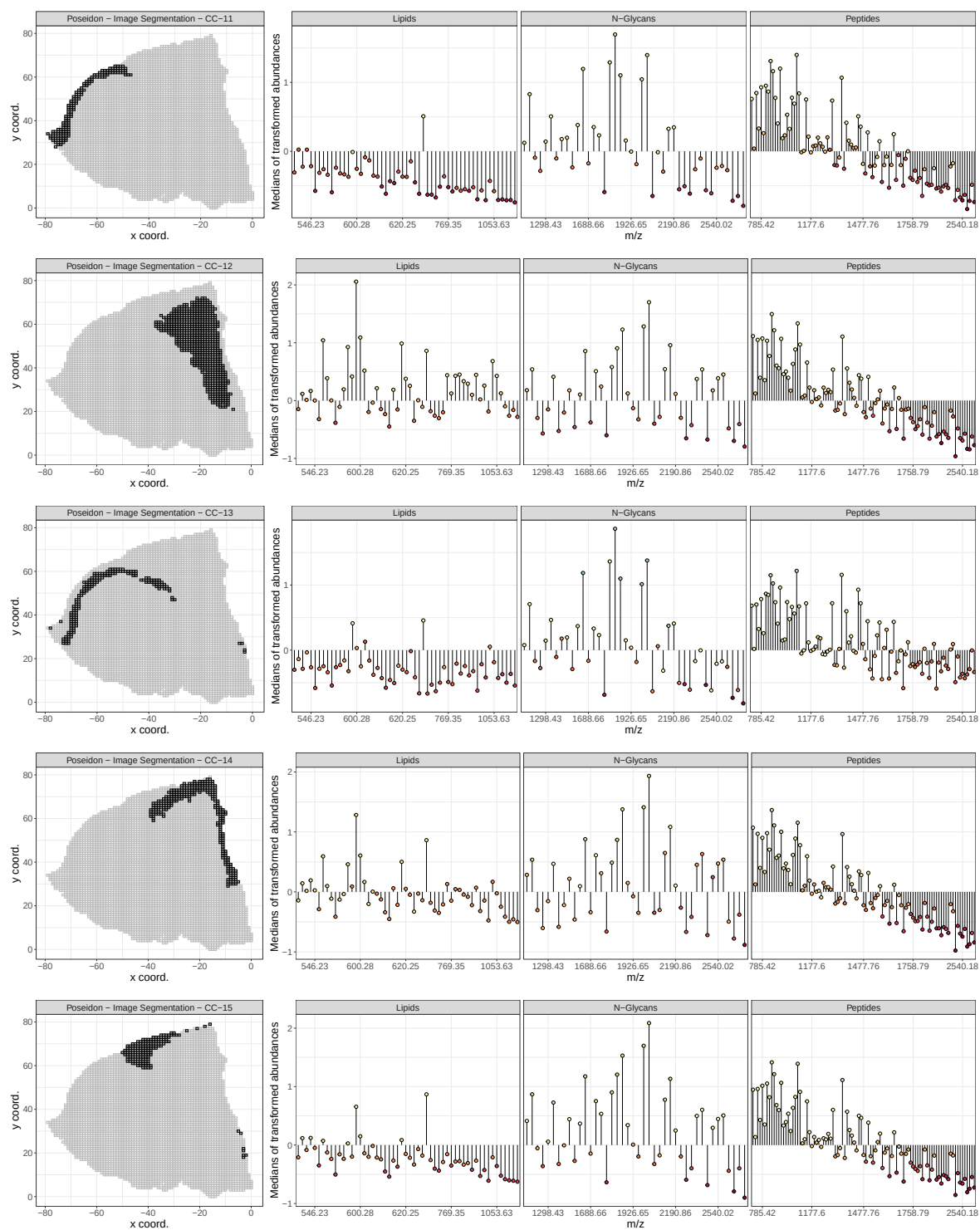


Fig. S.19: Visual representation of the sample medians of transformed abundances across CCs 11–15. Panels in columns 2–4 display these medians colored according to their estimated RC. The corresponding pixels associated with each cluster are highlighted in the first column.

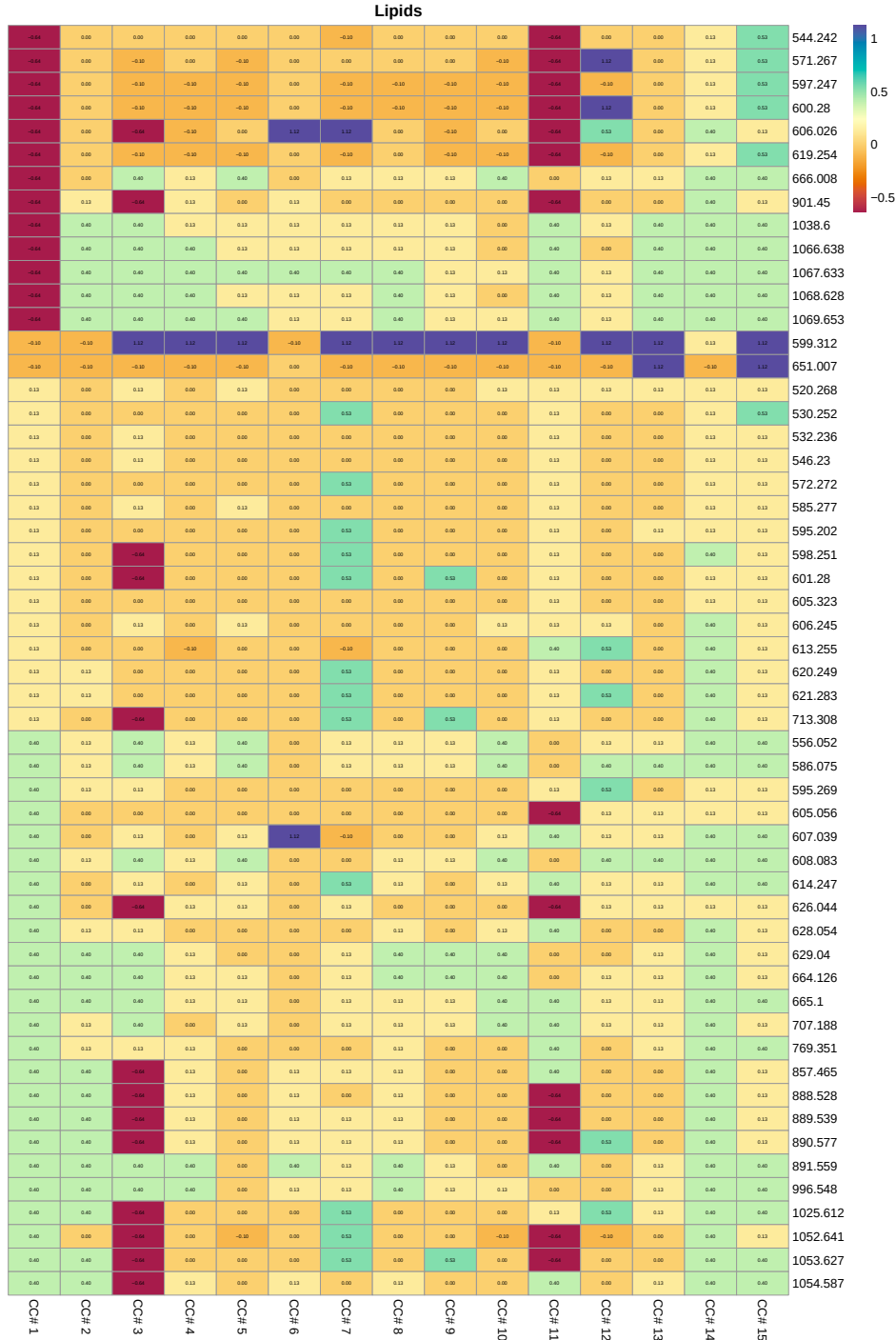


Fig. S.20: Lipids. Heatmap of a matrix with CCs across the columns and m/z values along the rows. Each cell is colored according to the estimated posterior variational mean. The m/z values are ordered by the RC magnitude in the first CC.

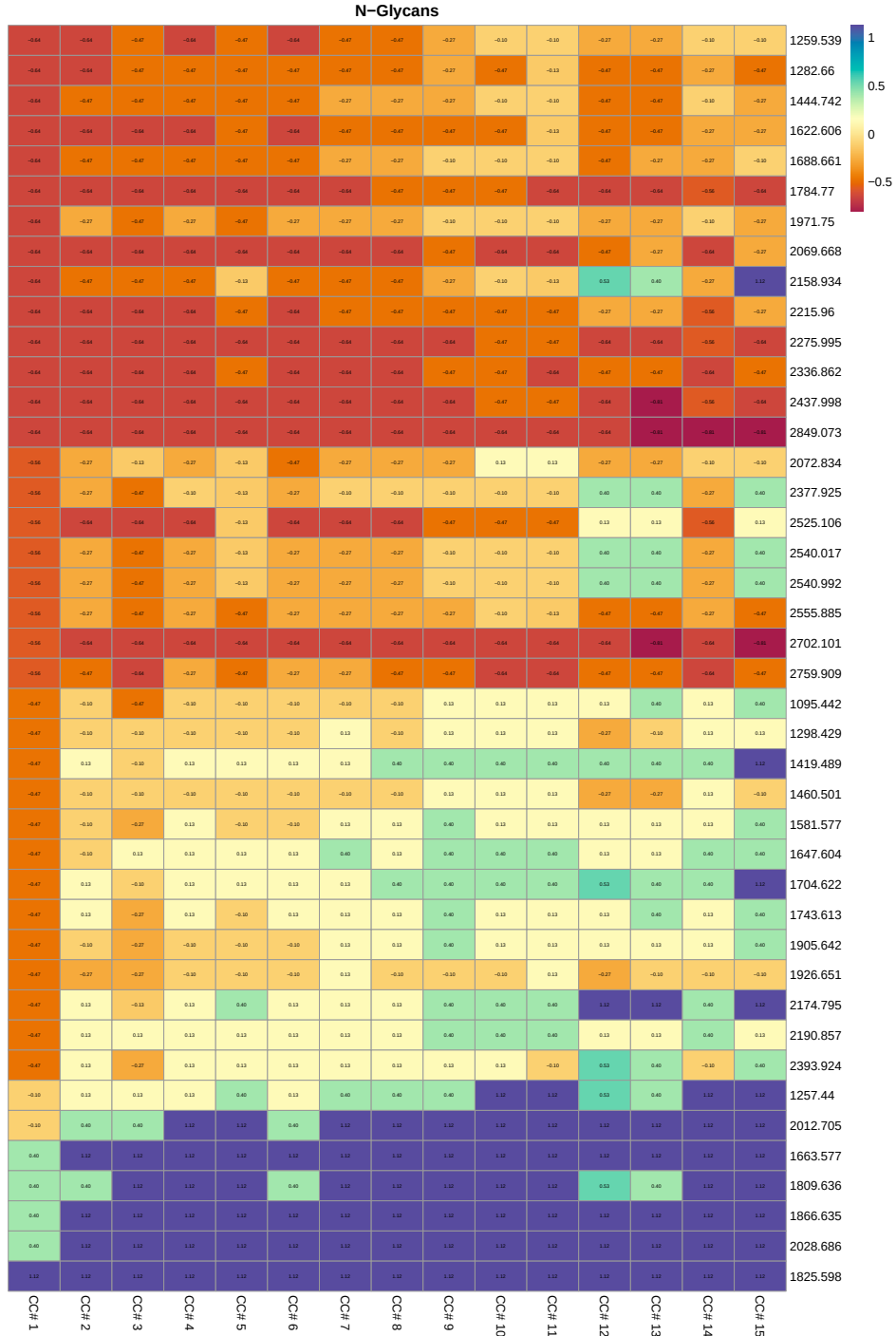


Fig. S.21: N-glycans. Heatmap of a matrix with CCs across the columns and m/z values along the rows. Each cell is colored according to the estimated posterior variational mean. The m/z values are ordered by the RC magnitude in the first CC.

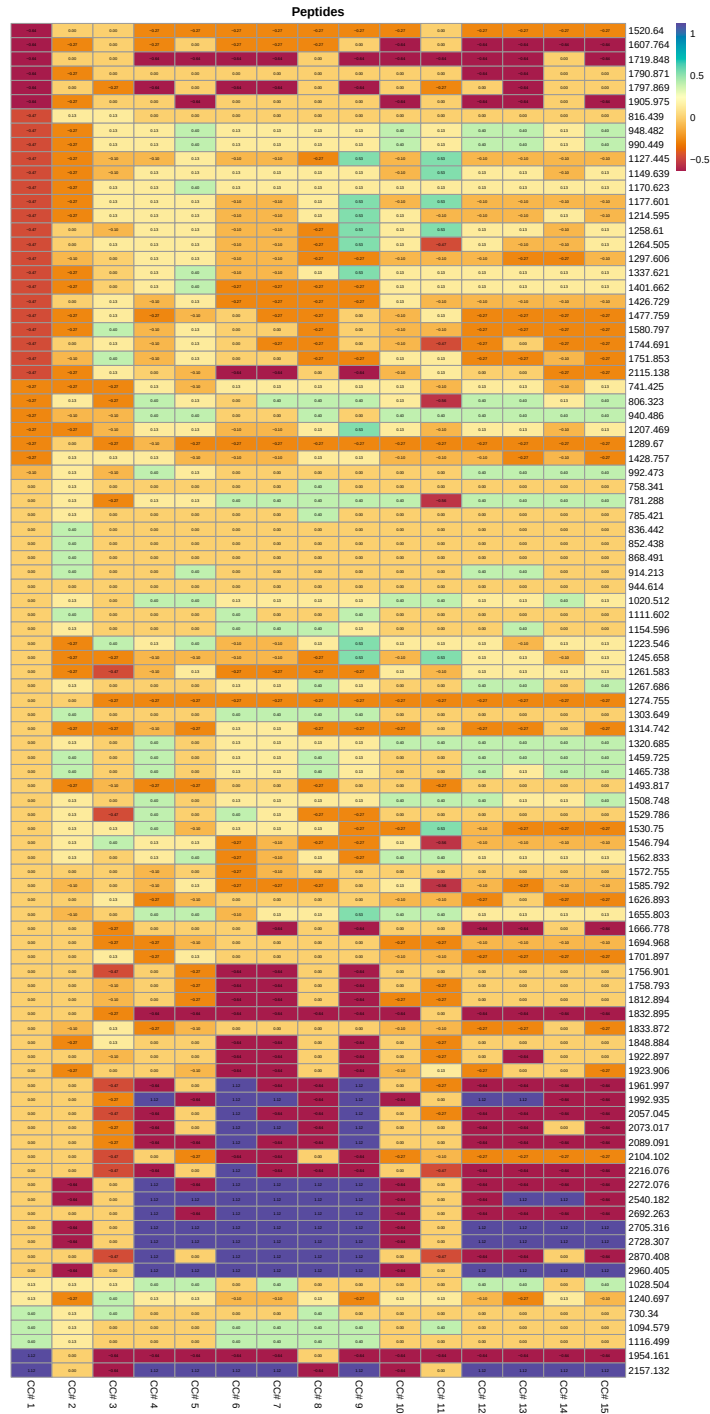


Fig. S.22: Peptides. Heatmap of a matrix with CCs across the columns and m/z values along the rows. Each cell is colored according to the estimated posterior variational mean. The m/z values are ordered by the RC magnitude in the first CC.

References

- Benjamin Balluff, Achim Buck, Marta Martin-Lorenzo, Frédéric Dewez, Rupert Langer, Liam A McDonnell, Axel Walch, and Ron MA Heeren. Integrative clustering in mass spectrometry imaging for enhanced patient stratification. *PROTEOMICS–Clinical Applications*, 13(1):1800137, 2019.
- Cecilia Balocchi, Edward I George, and Shane T Jensen. Clustering areal units at multiple levels of resolution to model crime in Philadelphia. *arXiv preprint arXiv:2112.02059*, 2021.
- Julian Besag. Spatial interaction and the statistical analysis of lattice systems. *Journal of the Royal Statistical Society: Series B (Methodological)*, 36(2):192–236, 1974.
- Julian Besag. On the Statistical Analysis of Dirty Pictures. *Journal of the Royal Statistical Society. Series B (Methodological)*, 48(3):259–302, 1986.
- Xueli Bian, Rui Liu, Ying Meng, Dongming Xing, Daqian Xu, and Zhimin Lu. Lipid metabolism and cancer. *Journal of Experimental Medicine*, 218(1):1–34, 2021.
- David M. Blei and Michael I. Jordan. Variational inference for Dirichlet process mixtures. *Bayesian Analysis*, 1(1), March 2006. ISSN 1936-0975. doi: 10.1214/06-BA104. URL <http://dx.doi.org/10.1214/06-BA104>.
- David M. Blei, Alp Kucukelbir, and Jon D. McAuliffe. Variational Inference: A Review for Statisticians. *Journal of the American Statistical Association*, 112(518):859–877, 2017. doi: 10.1080/01621459.2017.1285773.
- Federico Camerlenghi, David B. Dunson, Antonio Lijoi, Igor Prünster, and Abel Rodríguez. Latent Nested Nonparametric Priors (with Discussion). *Bayesian Analysis*, 14(4):1–54, 2019. doi: 10.1214/19-BA1169. URL <http://arxiv.org/abs/1801.05048>.
- Giulia Capitoli, Kirsten C. J. van Abeelen, Isabella Piga, Vincenzo L’Imperio, Marco S. Nobile, Daniela Besozzi, and Stefania Galimberti. Well Begun Is Half Done: The Impact of Pre-Processing in MALDI Mass Spectrometry Imaging Analysis Applied to a Case Study of Thyroid Nodules. *Stats*, 8(3):1–14, 2025.
- David B Dahl, Ryan Day, and Jerry W Tsai. Random partition distribution indexed by pairwise information. *Journal of the American Statistical Association*, 112(518):721–732, 2017.
- Laura D’Angelo and Francesco Denti. A Finite-Infinite Shared Atoms Nested Model for the Bayesian Analysis of Large Grouped Data Sets. *Bayesian Analysis*, 21(1):105 – 138, 2026. doi: 10.1214/24-BA1458. URL <https://doi.org/10.1214/24-BA1458>.
- Francesco Denti and Laura D’Angelo. The generalized nested common atoms model. *Econometrics and Statistics*, pages 1–28, 2025. doi: 10.1016/j.ecosta.2025.01.001.
- Francesco Denti, Federico Camerlenghi, Michele Guindani, and Antonietta Mira. A Common Atoms Model for the Bayesian Nonparametric Analysis of Nested Data. *Journal of the American Statistical Association*, 118(541):405–416, 2023.
- Francesco Denti, Cecilia Balocchi, Giulia Capitoli, and Vanna Denti. Dataset: Poseidon Data ccRCC Application. *Bicocca Open Archive Research Data*, (v1.0), 2026. doi: 10.17632/gxz9kjj2r9.1. URL <https://doi.org/10.17632/gxz9kjj2r9.1>.

- Vanna Denti, Giulia Capitoli, Isabella Piga, Francesca Clerici, Lisa Pagani, Lucrezia Criscuolo, Greta Bindi, Lucrezia Principi, Clizia Chinello, Giuseppe Paglia, et al. Spatial multiomics of lipids, N-glycans, and tryptic peptides on a single FFPE tissue section. *Journal of Proteome Research*, 21(11):2798–2809, 2022.
- Vanna Denti, Angela Greco, Antonio Maria Alviano, Giulia Capitoli, Nicole Monza, Andrew Smith, Daniela Pilla, Alice Maggioni, Mariia Ivanova, Konstantinos Venetis, et al. Spatially resolved molecular characterization of noninvasive follicular thyroid neoplasms with papillary-like nuclear features (NIFTPs) identifies a distinct proteomic signature associated with RAS-mutant lesions. *International Journal of Molecular Sciences*, 25(23):13115, 2024.
- J. B. Durand, F. Forbes, C. D. Phan, L. Truong, H. D. Nguyen, and F. Dama. Bayesian non-parametric spatial prior for traffic crash risk mapping: A case study of Victoria, Australia. *Australian and New Zealand Journal of Statistics*, 64(2):171–204, 2022.
- Chris Fraley and Adrian E. Raftery. Bayesian regularization for normal mixture estimation and model-based clustering. *Journal of Classification*, 24(2):155–181, September 2007. ISSN 1432-1343. doi: 10.1007/s00357-007-0004-5. URL <http://dx.doi.org/10.1007/s00357-007-0004-5>.
- Sylvia Frühwirth-Schnatter, Gertraud Malsiner-Walli, and Bettina Grün. Generalized Mixtures of Finite Mixtures and Telescoping Sampling. *Bayesian Analysis*, 16(4):1279–1307, 2021.
- Yan Fu, Tiantian Zou, Xiaotian Shen, Peter J Nelson, Jiahui Li, Chao Wu, Jimeng Yang, Yan Zheng, Christiane Bruns, Yue Zhao, et al. Lipid metabolism in cancer progression and therapeutic strategies. *MedComm*, 2(1):27–59, 2021.
- Semira Gobena, Bemrew Admassu, Mebrie Zemene Kinde, and Abebe Tesfaye Gessese. Proteomics and its current application in biomedical area: Concise review. *The Scientific World Journal*, 2024(1):4454744, 2024.
- G rard Govaert and Mohamed Nadif. *Co-clustering: models, algorithms and applications*. John Wiley & Sons, 2013.
- Dan Guo, Kylie Bemis, Catherine Rawlins, Jeffrey Agar, and Olga Vitek. Unsupervised segmentation of mass spectrometric ion images characterizes morphology of tissues. *Bioinformatics*, 35(14):i208–i217, 2019.
- James J Hsieh, Mark P Purdue, Sabina Signoretti, Charles Swanton, Laurence Albiges, Manuela Schmidinger, Daniel Y Heng, James Larkin, and Vincenzo Ficarra. Renal cell carcinoma. *Nature Reviews Disease Primers*, 3(1):1–19, 2017.
- James J Hsieh, Valerie Le, Dengfeng Cao, Emily H Cheng, and Chad J Creighton. Genomic classifications of renal cell carcinoma: A critical step towards the future application of personalized kidney cancer care with pan-omics precision. *The Journal of Pathology*, 244(5):525–537, 2018.
- Yuxuan Hu, Jiazhen Rong, Yafei Xu, Runzhi Xie, Jacqueline Peng, Lin Gao, and Kai Tan. Unsupervised and supervised discovery of tissue cellular neighborhoods from cell phenotypes. *Nature Methods*, 21(2):267–278, January 2024. ISSN 1548-7105. doi: 10.1038/s41592-023-02124-2. URL <http://dx.doi.org/10.1038/s41592-023-02124-2>.
- Junbum Kim, Samir Rustam, Juan Miguel Mosquera, Scott H. Randell, Renat Shaykhiev, Andr  F. Rendeiro, and Olivier Elemento. Unsupervised discovery of tissue architecture in multiplexed imaging. *Nature Methods*, 19(12):1653–1661, October 2022.

ISSN 1548-7105. doi: 10.1038/s41592-022-01657-2. URL <http://dx.doi.org/10.1038/s41592-022-01657-2>.

- Kasper K Krestensen, Ron MA Heeren, and Benjamin Balluff. State-of-the-art mass spectrometry imaging applications in biomedical research. *Analyst*, 148(24):6161–6187, 2023.
- Juhee Lee, Peter Müller, Yitan Zhu, and Yuan Ji. A nonparametric Bayesian model for local clustering with application to proteomics. *Journal of the American Statistical Association*, 108(503):775–788, 2013. ISSN 01621459. doi: 10.1080/01621459.2013.784705.
- Scott W. Linderman, Matthew J. Johnson, and Ryan P. Adams. Dependent multinomial models made easy: Stick-breaking with the pólya-gamma augmentation. In *Advances in Neural Information Processing Systems 28*, pages 3456–3464, 2015.
- Hongliang Lü, Julyan Arbel, and Florence Forbes. Bayesian nonparametric priors for hidden Markov random fields. *Statistics and Computing*, 30(4):1015–1035, 2020.
- Martin Maechler, Peter Rousseeuw, Anja Struyf, Mia Hubert, and Kurt Hornik. *cluster: Cluster Analysis Basics and Extensions*, 2025.
- Gertraud Malsiner-Walli, Sylvia Frühwirth-Schnatter, and Bettina Grün. Model-based Clustering Based on Sparse Finite Gaussian Mixtures. *Statistics and Computing*, 26:303–324, 2016. doi: 10.1007/s11222-014-9500-2.
- Clare A McGrory, D Michael Titterton, Robert Reeves, and Anthony N Pettitt. Variational Bayes for estimating the parameters of a hidden Potts model. *Statistics and Computing*, 19(3):329–340, 2009.
- Jeffrey W Miller and Matthew T Harrison. Mixture models with a prior on the number of components. *Journal of the American Statistical Association*, 113(521):340–356, 2018.
- Matthew Moores, Geoff K. Nicholls, Anthony N. Pettitt, and Kerrie Mengersen. Scalable Bayesian inference for the inverse temperature of a hidden Potts model. *Bayesian Analysis*, 15(1):1–27, 2020.
- Iain Murray, Zoubin Ghahramani, and David J.C. MacKay. MCMC for doubly-intractable distributions. *Proceedings of the 22nd Conference on Uncertainty in Artificial Intelligence, UAI 2006*, pages 359–366, 2006.
- Agostino Nobile. On the Posterior Distribution of the Number of Components in a Finite Mixture. *The Annals of Statistics*, 32(5):2044–2073, 2004.
- Jim Pitman and Marc Yor. The two-parameter Poisson-Dirichlet distribution derived from a stable subordinator. *The Annals of Probability*, pages 855–900, 1997.
- Mridula Prasad, Geert Postma, Pietro Franceschi, Lutgarde MC Buydens, and Jeroen J Jansen. Evaluation and comparison of unsupervised methods for the extraction of spatial patterns from mass spectrometry imaging data (msi). *Scientific Reports*, 12(1):15687, 2022.
- Giovanni Rebaudo, Qiaohui Lin, and Peter Mueller. Separate Exchangeability as Modeling Principle in Bayesian Nonparametrics. *Statistical Science*, (In press), 2026.
- Abel Rodríguez, David B. Dunson, and Alan E. Gelfand. The nested Dirichlet process. *Journal of the American Statistical Association*, 103(483):1131–1154, 2008. ISSN 01621459. doi: 10.1198/016214508000000553.
- T.C. Rohner, D. Staab, and M. Stoeckli. MALDI mass spectrometric imaging of biological tissue sections. *Mechanisms of Ageing and Development*, 126(1):177–185, 2005.

- Mohamad Nasir Shafiee, Catharine A Ortori, David A Barrett, Nigel P Mongan, Jafaru Abu, and William Atiomo. Lipidomic biomarkers in polycystic ovary syndrome and endometrial cancer. *International journal of molecular sciences*, 21(13):4753, 2020.
- Andrea Sottosanti and Davide Risso. Co-clustering of spatially resolved transcriptomic data. *The Annals of Applied Statistics*, 17(2):1444–1468, June 2023. ISSN 1932-6157. doi: 10.1214/22-AOAS1677. URL <http://dx.doi.org/10.1214/22-aos1677>.
- Pamela Stanley, Kelley W. Moremen, Nathan E. Lewis, Naoyuki Taniguchi, and Markus Aebi. *N-Glycans*. Cold Spring Harbor Laboratory Press, Cold Spring Harbor (NY), 4th edition, 2022. ISBN 9781621824220. URL <http://europepmc.org/books/NBK579964>.
- Kean Ming Tan and Daniela M Witten. Sparse biclustering of transposable data. *Journal of Computational and Graphical Statistics*, 23(4):985–1008, 2014.
- Robert Tibshirani, Guenther Walther, and Trevor Hastie. Estimating the number of clusters in a data set via the gap statistic. *Journal of the Royal Statistical Society: Series B*, 63(2):411–423, 2001.
- Michael Tuck, Florent Grélard, Landry Blanc, and Nicolas Desbenoit. MALDI-MSI Towards Multimodal Imaging: Challenges and Perspectives. *Frontiers in Chemistry*, 10:1–11, 2022. ISSN 2296-2646. doi: 10.3389/fchem.2022.904688. URL <http://dx.doi.org/10.3389/fchem.2022.904688>.
- Sophie Vasseur and Fabienne Guillaumond. Lipids in cancer: a global view of the contribution of lipid pathways to metastatic formation and treatment resistance. *Oncogenesis*, 11(1):46, 2022.
- Elizabeth N Wallace, Connor A West, Colin T McDowell, Xiaowei Lu, Evelyn Bruner, Anand S Mehta, Kiyoko F Aoki-Kinoshita, Peggi M Angel, and Richard R Drake. An n-glycome tissue atlas of 15 human normal and cancer tissue types determined by MALDI-imaging mass spectrometry. *Scientific Reports*, 14(1):489, 2024.
- Hua Zhang, Kelly H Lu, Malik Ebbini, Penghsuan Huang, Haiyan Lu, and Lingjun Li. Mass spectrometry imaging for spatially resolved multi-omics molecular mapping. *npj Imaging*, 2(1):20, 2024.
- Edward Zhao, Matthew R. Stone, Xing Ren, Jamie Guenthoer, Kimberly S. Smythe, Thomas Pulliam, Stephen R. Williams, Cedric R. Uytingco, Sarah E. B. Taylor, Paul Nghiem, Jason H. Bielas, and Raphael Gottardo. Spatial transcriptomics at subspot resolution with BayesSpace. *Nature Biotechnology*, 39(11):1375–1384, June 2021. ISSN 1546-1696. doi: 10.1038/s41587-021-00935-2. URL <http://dx.doi.org/10.1038/s41587-021-00935-2>.
- Evgeny Zhvansky, Anatoly Sorokin, Vsevolod Shurkhay, Denis Zavorotnyuk, Denis Bormotov, Stanislav Pekov, Alexander Potapov, Evgeny Nikolaev, and Igor Popov. Comparison of dimensionality reduction methods in mass spectra of astrocytoma and glioblastoma tissues. *Mass Spectrometry*, 10(1):A0094, 2021.