

Resolving tissue complexity by multimodal spatial omics modeling with MISO

Received: 31 January 2024

Accepted: 21 November 2024

Published online: 15 January 2025

 Check for updates

Kyle Coleman  , Amelia Schroeder², Melanie Loth², Daiwei Zhang², Jeong Hwan Park  ³, Ji-Youn Sung⁴, Niklas Blank^{5,6}, Alexis J. Cowan⁵, Xuyu Qian⁷, Jianfeng Chen⁸, Jiahui Jiang^{9,10}, Hanying Yan  ², Laith Z. Samarah¹¹, Jean R. Clemenceau¹², Inyeop Jang¹², Minji Kim¹², Isabel Barnfather  ¹², Joshua D. Rabinowitz  ¹¹, Yanxiang Deng  ^{13,14}, Edward B. Lee  ¹³, Alexander Lazar  ¹⁵, Jianjun Gao⁸, Emma E. Furth¹³, Tae Hyun Hwang¹², Linghua Wang  ^{9,16}, Christoph A. Thaiss  ⁵, Jian Hu  ¹⁷ & Mingyao Li ^{2,13}

Spatial molecular profiling has provided biomedical researchers valuable opportunities to better understand the relationship between cellular localization and tissue function. Effectively modeling multimodal spatial omics data is crucial for understanding tissue complexity and underlying biology. Furthermore, improvements in spatial resolution have led to the advent of technologies that can generate spatial molecular data with subcellular resolution, requiring the development of computationally efficient methods that can handle the resulting large-scale datasets. MISO (Multi-modal Spatial Omics) is a versatile algorithm for feature extraction and clustering, capable of integrating multiple modalities from diverse spatial omics experiments with high spatial resolution. Its effectiveness is demonstrated across various datasets, encompassing gene expression, protein expression, epigenetics, metabolomics and tissue histology modalities. MISO outperforms existing methods in identifying biologically relevant spatial domains, representing a substantial advancement in multimodal spatial omics analysis. Moreover, MISO's computational efficiency ensures its scalability to handle large-scale datasets generated by subcellular resolution spatial omics technologies.

Multimodal spatial omics represents a paradigm shift in understanding complex biological systems by integrating diverse omics modalities within their original tissue contexts¹. This approach provides a comprehensive view of cellular and tissue functions, which is crucial for unraveling the molecular mechanisms of diseases. Recently developed multimodal spatial omics technologies include spatial transcriptomics^{2–4} (transcriptomics and histology), spatial epigenome–transcriptome co-profiling⁵ (chromatin accessibility, histone modification and transcriptomics), spatial gene and protein expression^{6–8} (transcriptomics, proteomics and histology) and spatial transcriptomics and metabolomics⁹ (transcriptomics, metabolomics and histology).

These technologies provide at least one spatially resolved omics modality and are often accompanied by a high-resolution hematoxylin and eosin (H&E)-stained histology image, yielding two or more modalities from the same tissue slice. The integration of multiple modalities is pivotal in obtaining a more holistic and accurate understanding of tissue architecture and function at cellular and molecular levels.

A critical step in multimodal spatial omics data analysis is to amalgamate features from different modalities to create cohesive low-dimensional data representations that can be harnessed for downstream analyses such as spatial domain detection by clustering. Current methods, however, either ignore spatial information^{10,11} or are limited

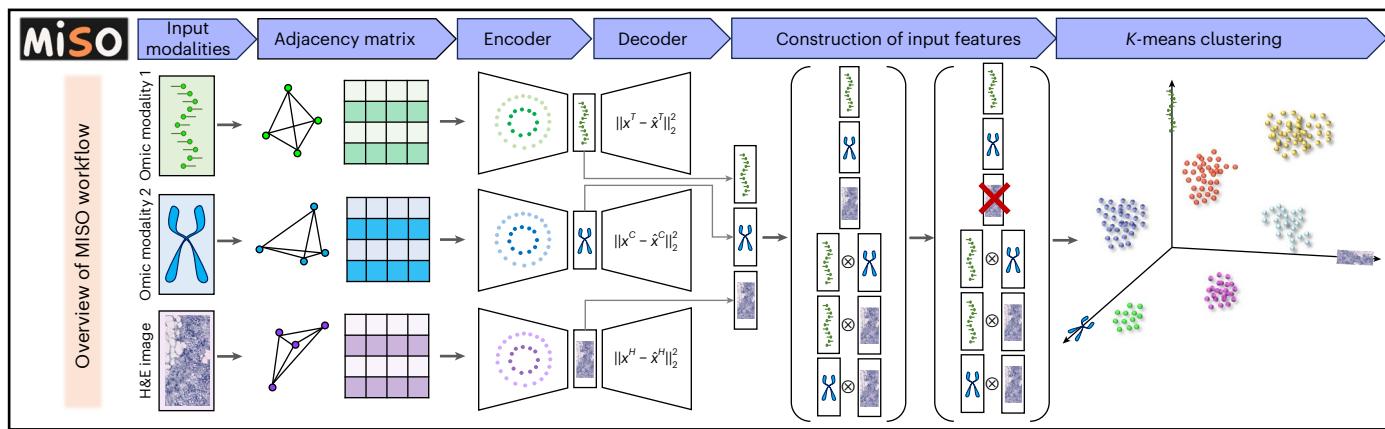


Fig. 1 | MISO workflow for analysis of spatial multi-omics dataset with paired histology image. MISO starts by constructing an adjacency matrix for each modality. This adjacency matrix and the spot-level features for that modality are used as input for a multilayer perceptron that is trained to minimize spectral clustering and reconstruction loss functions. The modality-specific embeddings

are extracted from the multilayer perceptrons, and interactions between each pair of modalities are computed. The features of low-quality modalities are removed, and the final group of embeddings is used as input for the *k*-means clustering algorithm.

to integrating a subset of modalities in spatial omics experiments, typically handling just two modalities, requiring extensive hyperparameter tuning, and being sensitive to modalities of lower quality. For instance, MUSE¹², a spatial clustering algorithm, is designed for combining transcriptomics and imaging data in spatial transcriptomics, while SpatialGlue¹³, a spatial multi-omics clustering algorithm, caters to three distinct experiment types: SPOTS⁸, Stereo-CITE-seq⁷ and spatial assay for transposase-accessible chromatin and RNA using sequencing (spatial ATAC–RNA-seq)⁵. SpatialGlue, although adept at integrating two omics modalities, does not support tissue histology imaging data as an input. As multimodal spatial omics technologies advance and continue to produce data with more than two omics modalities along with tissue histology imaging data, there is a growing need for methods that can seamlessly integrate information from diverse spatial omics experiments.

Here we present MISO (Multi-modal Spatial Omics), a feature extraction and spatial clustering algorithm designed for comprehensive modality integration, including both omics and tissue histology imaging data, in any multimodal spatial omics experiment. MISO's efficacy is demonstrated across various datasets, such as spot-level and single-cell-level spatial transcriptomics (10x Visium, Xenium and Visium HD), spatial ATAC–RNA-seq, spatial CUT&Tag–RNA-seq, spatial gene and protein expression, and spatial transcriptomics and metabolomics. MISO excels in accurately integrating different modalities and segregating spots into distinct, biologically relevant spatial domains. A notable advantage of MISO is its minimal requirement for hyperparameter tuning, coupled with its ability to mitigate artifacts associated with low-quality modalities, a common challenge in real-world studies. This makes MISO not only a powerful tool for multimodal data analysis but also a robust solution for researchers exploring the complexities of spatial omics.

Results

Overview of MISO

The MISO workflow (Fig. 1) begins with the extraction of low-dimensional embeddings for each modality. This is achieved using modality-specific multilayer perceptrons, which are trained to minimize spectral clustering and reconstruction loss functions. MISO then calculates the outer product between pairs of modality-specific embeddings to construct features that capture the interactions between modalities. These modality-specific and interaction feature vectors are concatenated to produce comprehensive embeddings that encapsulate all modalities.

A crucial step in MISO is the filtration of embeddings from low-quality modalities, which are essential in enhancing the robustness of the analysis. The refined subset of embeddings is then used as input for a *k*-means clustering algorithm to separate the spots into distinct clusters, each reflective of the unique characteristics defined by the input modalities. MISO is a computationally fast algorithm, processing medium-scale (1,000–10,000 spots/cells) datasets in under 1 min on a single graphics processing unit. MISO also scales effectively to large datasets (>10,000 spots/cells) and allows for evaluation of datasets containing hundreds of thousands of capture locations (Supplementary Fig. 1).

Application to bladder cancer 10x Visium V2 data

To demonstrate MISO's robust integration capabilities, we applied it to a wide range of multimodal spatial omics datasets. Our initial evaluation focused on a bladder cancer sample generated using 10x Visium (V2)¹⁴, which includes both spatial transcriptomics and tissue histology image data (Fig. 2a). Our primary objective in analyzing this dataset was to assess MISO's ability to identify fine-grained, disease-relevant tissue structures. Specifically, we focused on the segmentation of two tertiary lymphoid structures (TLSs), denoted as TLS1 and TLS2 (Fig. 2a). TLSs are aggregates of immune cells that develop in nonlymphoid tissues in response to chronic inflammation¹⁵. The presence of TLSs in tumors has been associated with improved prognosis for multiple types of cancer, underscoring their potential importance in developing novel therapeutic interventions¹⁶. TLSs contain high endothelial venules (HEVs), specialized blood vessels that recruit lymphocytes from the bloodstream to form the TLS¹⁷. An annotation of the HEVs within TLS1 and TLS2, determined using nuclei segmentation, is shown in Fig. 2a.

Although HEVs are crucial in the immune response to tumor, they are also very small, typically measuring only 40 μm in diameter¹⁸. Spots sequenced using 10x Visium have a diameter of 55 μm , with 100 μm between the centers of adjacent spots, resulting in gaps between spots (Fig. 2b). Consequently, even if a Visium spot encompasses an HEV, it likely contains a mixture of transcripts from immune cells within the TLS, potentially obscuring the gene expression signatures of the endothelial cells lining the HEV boundary. To enhance the detection of HEVs, we increased the spatial resolution of the Visium gene expression data using iStar¹⁹, producing gene expression data at $4 \times 4 \mu\text{m}^2$ super-pixels with no gaps between super-pixels (Fig. 2c). Within the two TLS regions of interest, resolution enhancement increased the number of gene capture units from 17 spots to 8,613 super-pixels for TLS1 and from 12 spots and 9,496 super-pixels for TLS2.

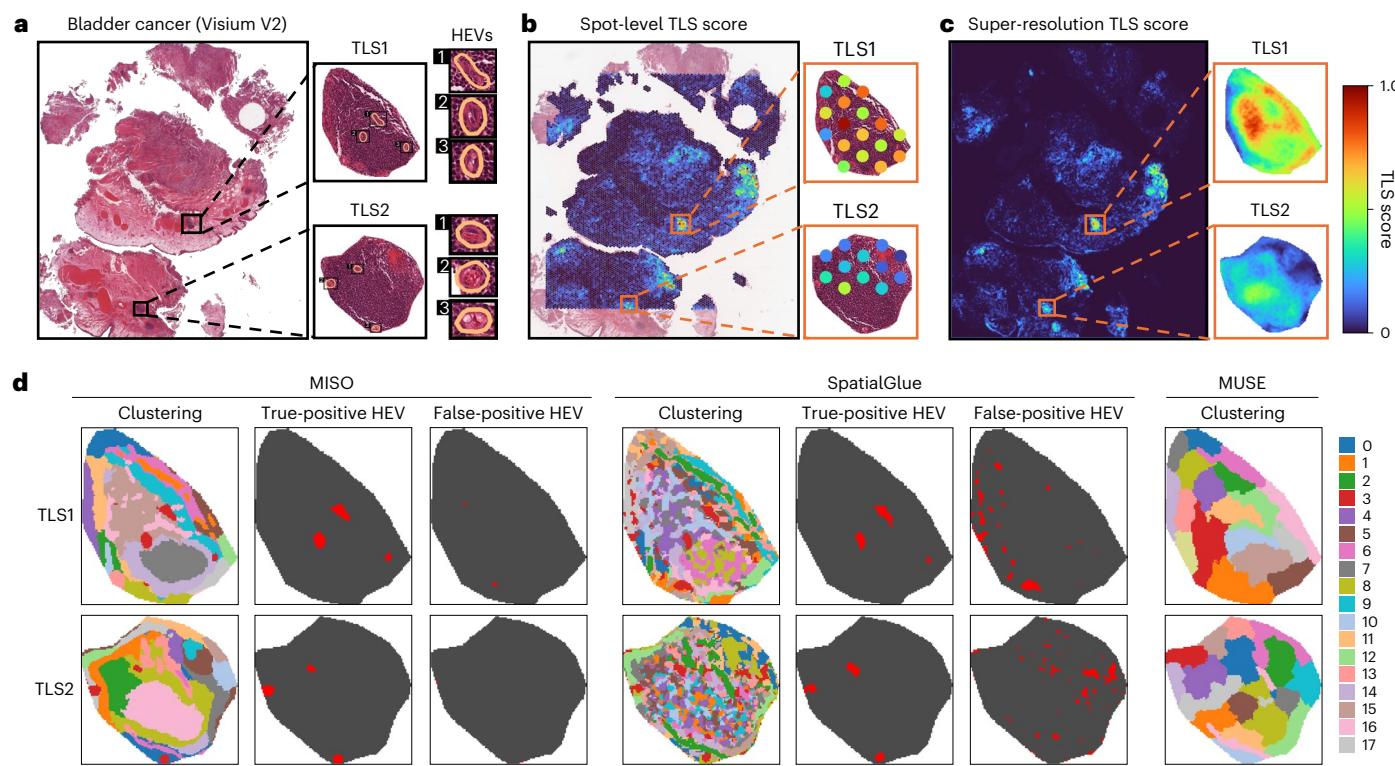


Fig. 2 | Analysis of a 10x Visium bladder cancer spatial transcriptomics dataset. **a**, H&E-stained histology image of the analyzed tissue section with TLS and HEV annotation. **b**, TLS score across all Visium spots. **c**, TLS score across all super-pixels following gene expression resolution enhancement using iStar.

d, Shown from left to right are the clustering results for TLS1 and TLS2 provided by MISO, MUSE and SpatialGlue, with masks showing the true-positive and false-positive HEV super-pixels for MISO and SpatialGlue.

Subsequently, we evaluated MISO on the super-pixels contained within TLS1 and TLS2 (Fig. 2d). Cluster 3 identified by MISO corresponded to the annotated HEVs with high sensitivity and specificity for both TLS regions. MISO grouped only ten non-HEV super-pixels with the HEV cluster for TLS1 and only six for TLS2, demonstrating its precision and effectiveness in identifying these fine-grained critical structures. As a comparison, we also applied SpatialGlue and MUSE to this dataset. Since neither SpatialGlue nor MUSE contains a tissue histology image feature extraction component, we used the tissue histology image features extracted by MISO as input for all methods. Figure 2d shows that SpatialGlue identified the HEV regions with high sensitivity but very low specificity, falsely grouping 289 non-HEV super-pixels with the HEV cluster for TLS1 and 381 for TLS2. MUSE failed to produce a cluster resembling the HEV localization, preventing the generation of true-positive and false-positive HEV masks from its results. The superior performance of MISO over SpatialGlue and MUSE underscores its capability to effectively integrate high-resolution spatial gene expression and histology imaging features to localize fine-grained tissue structures with high accuracy.

Next, we aimed to quantitatively compare the performance of different methods. Traditionally, pathologists' manual annotations serve as the ground truth when assessing clustering accuracy in spatial omics data. However, in the context of multimodal spatial omics data, histology represents just one modality in the analysis. Annotations derived from a single modality may not fully capture the complex tissue structures defined by all modalities. To address this limitation in quantitative evaluation, we calculated the intraclass correlation coefficient (ICC) for each method–modality cluster combination. The ICC measures the consistency of observations within clusters. A higher ICC value indicates increased homogeneity within clusters for a given modality, suggesting an improvement in clustering quality. The box plots in Extended Data Fig. 1 illustrate the distribution of ICC

values across clusters and features for each method with respect to each modality. MISO demonstrates superior performance in clustering based on both RNA and tissue histology data, showing its ability to extract low-dimensional embeddings that capture the complexities of both input modalities.

Application to gastric cancer 10x Xenium data

Encouraged by the above results, we next evaluated a molecular imaging-based spatial transcriptomics dataset from a gastric cancer sample generated using 10x Xenium (Fig. 3a). Xenium generates spatial gene expression data with single-cell resolution for an entire tissue section with the number of profiled capture locations far surpassing that of 10x Visium and most other sequencing-based spatial transcriptomics platforms. Unlike sequencing-based technologies, molecular imaging-based spatial transcriptomics platforms cannot measure gene expression for the full transcriptome and are instead restricted to a preselected gene panel typically consisting of a few hundred genes. For this gastric cancer sample, a total of 377 genes were profiled across 696,314 cells. Because of the low number of genes measured, effectively integrating histology image information becomes even more crucial for accurately segmenting the tissue section.

To evaluate clustering performance, we obtained a pathologist's annotation of the tissue section (by J.H.P.), distinguishing poorly cohesive carcinoma, invasive discohesive tumor cells, lymphoid aggregates, mucosa, submucosa and muscle (Fig. 3a). As shown in Fig. 3c, MISO accurately localized clusters to all major biological and disease-relevant structures identified in the pathologist's annotation. MISO effectively distinguished regions of poorly cohesive carcinoma (cluster 14) from those of discohesive tumor cells (cluster 5). The poorly cohesive carcinoma cluster was largely contiguous, with little intermixing of other clusters. In the discohesive tumor region, MISO also identified the areas of healthy muscle and lymphoid

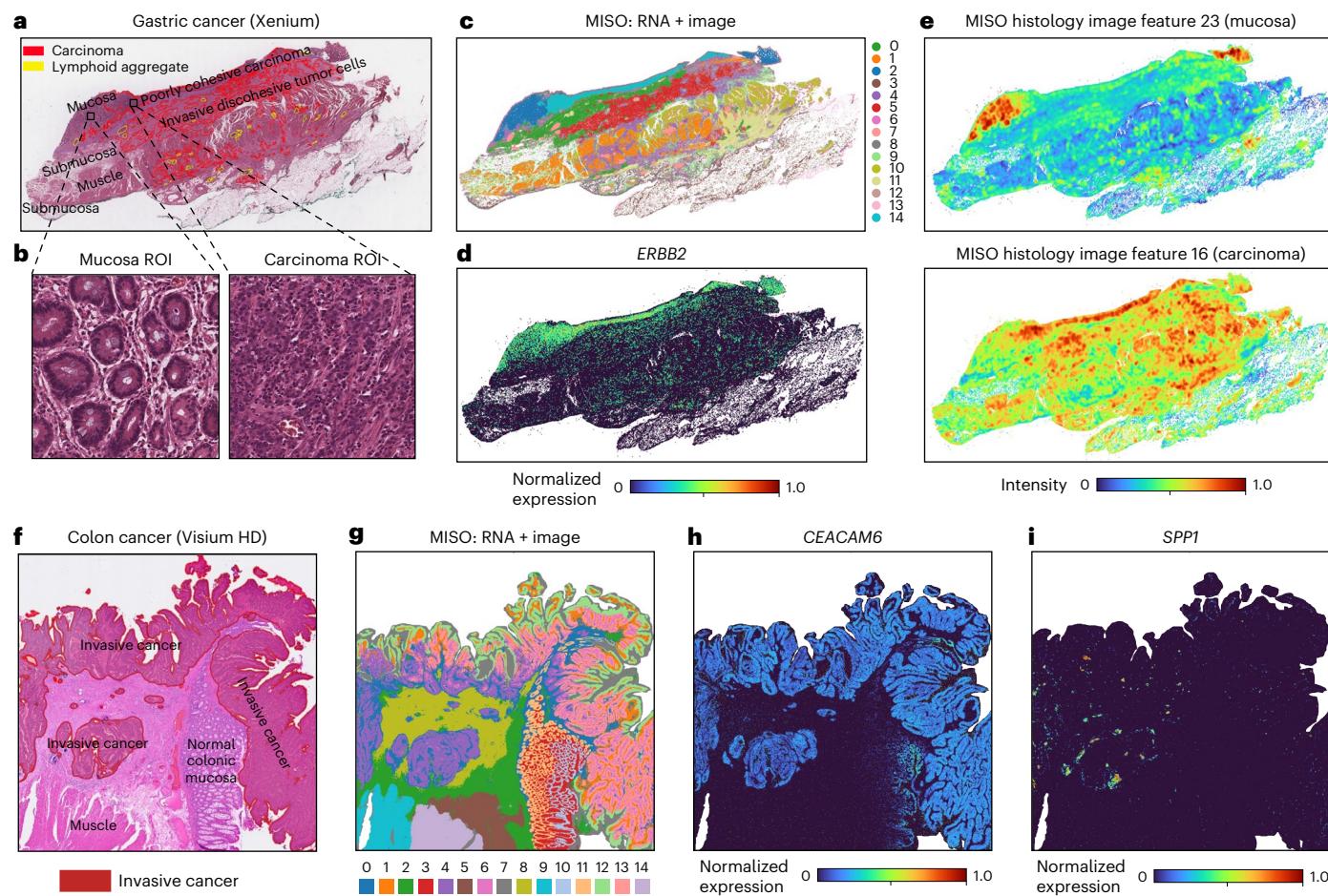


Fig. 3 | Analysis of large-scale spatial transcriptomics datasets. **a–e**, Analysis of a 10x Xenium gastric cancer spatial transcriptomics dataset. **a**, Pathologist manual annotation. **b**, Histology image patches of tissue from mucosa and carcinoma regions. **c**, MISO clustering results. **d**, Spatial gene expression plot of *ERBB2*, a gastric cancer marker gene that shows similar levels of expression in the carcinoma and mucosa regions. **e**, MISO extracted tissue histology image features that enabled the identification of mucosal and carcinoma

clusters. **f–i**, Analysis of a 10x Visium HD CRC spatial transcriptomics dataset. **f**, Pathologist annotation of tissue section. **g**, MISO clustering results. **h**, Spatial gene expression plot of *CEACAM6*, a CRC marker gene that shows similar levels of expression across all annotated invasive cancer regions. **i**, Spatial gene expression plot of *SPP1*, a tumor-specific macrophage marker gene that is colocalized with cluster 4 identified by MISO.

aggregates. Additionally, MISO effectively localized cluster 7 to the annotated lymphoid aggregate regions with high sensitivity. This capability is meaningful because it indicates that MISO can identify small lymphoid aggregates that are not easily detectable through direct visualization. MISO can, therefore, greatly reduce the time required for manual annotation in tissue sections with numerous small and detailed structures, enhancing the efficiency and accuracy of pathological analysis.

MISO was also able to separate the poorly cohesive carcinoma region from the mucosa, a task of particular intricacy due to the similarity in the expression levels of gastric cancer marker genes in these regions. As an example, we included the spatial gene expression plot of *ERBB2*, a gastric cancer marker gene²⁰ that shows a similar level of expression in the annotated carcinoma and mucosa regions (Fig. 3d). However, despite the similar expression patterns of key marker genes, the mucosa and carcinoma regions exhibit very different histological features (Fig. 3b). MISO's extracted tissue histology image features capture these histological differences (Fig. 3e), enabling MISO to clearly segregate the mucosa and poorly cohesive carcinoma regions into distinct clusters. Due to the large number of cells profiled, neither MUSE nor SpatialGlue had the capacity to analyze this tissue section because of memory constraints. MISO's ability to handle large-scale spatial omics datasets is crucial as technological improvements are enabling the

generation of increasingly detailed and comprehensive spatial molecular maps, providing deeper insights into complex biological systems.

Application to CRC 10x Visium HD data

Building on MISO's efficacy in analyzing large-scale spatial transcriptomics data with single-cell resolution, we next applied it to a colorectal cancer (CRC) spatial transcriptomics dataset generated using 10x Visium HD²¹, which provides gene expression data with subcellular resolution and full transcriptome measurements. This dataset includes expression levels for 18,085 genes across 545,913 bins with each bin measuring $8 \times 8 \mu\text{m}^2$. For evaluation, we obtained a pathologist's annotation of the tissue section (by J.-Y.S.), denoting regions of muscle, normal colonic mucosa and invasive carcinoma (Fig. 3f). When integrating gene expression and histology image features, MISO successfully separated bins into clusters representing each of these structures (Fig. 3g). Notably, MISO was able to identify multiple subclasses of invasive carcinoma. These invasive carcinoma clusters showed similar expression of *CEACAM6* (Fig. 3h), a CRC marker gene associated with poor overall survival and disease-free survival^{22,23}. If we were to investigate only the expression of these CRC marker genes, these different cancerous regions identified by MISO might appear equally severe.

However, bins for one of the invasive carcinoma subclusters (cluster 4) were colocalized with *SPP1*⁺ macrophages (Fig. 3i), a type of

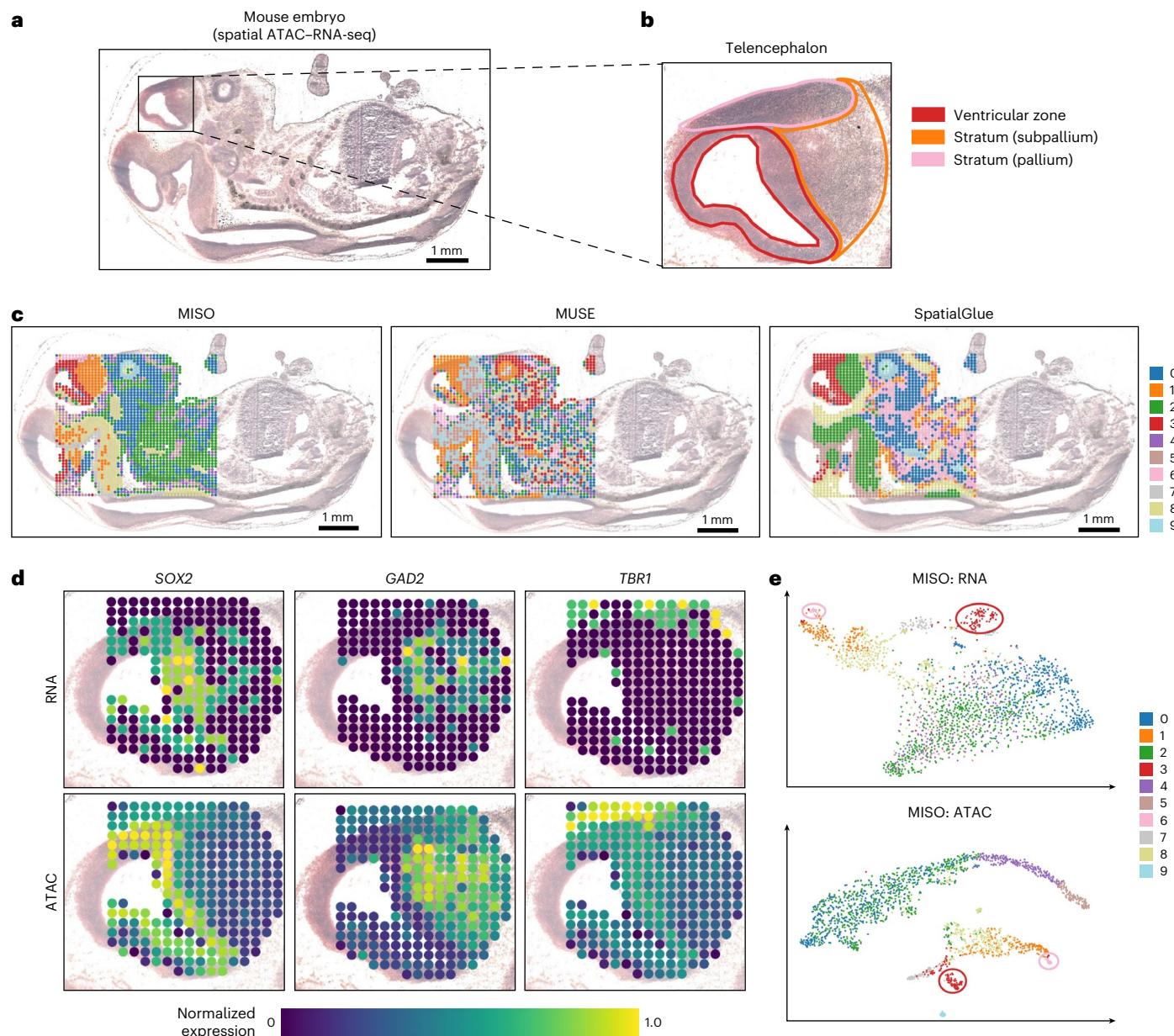


Fig. 4 | Analysis of a spatial ATAC-RNA-seq dataset from a mouse at E13. **a**, H&E-stained histology image of an adjacent tissue section. **b**, Magnification and annotation of telencephalon. **c**, Shown from left to right are the clustering results from MISO, MUSE and SpatialGlue. **d**, Shown from left to right are the

spatial gene expression plots of marker genes for the VZ, subpallium stratum and pallium stratum of the telencephalon. **e**, Spots plotted according to their RNA or ATAC t-SNE coordinates and colored by the MISO clustering results.

tumor-specific macrophage reported by Oliveira et al.²¹. The presence of *SPP1*⁺ macrophages has been associated with CRC progression and worse clinical outcomes^{24,25}. Unlike *CEACAM6* expression, the cancer regions on the periphery of the tissue section assigned to clusters 6 and 12 did not show the same extent of *SPP1* colocalization, indicating potentially less aggressive cancer cells. In fact, *SPP1* was upregulated in cluster 4 compared to clusters 6 and 12, with a log₂ fold change of 1.06 determined by differential expression analysis (*P* value = 1.22×10^{-56}). This means that MISO was able to differentiate a cluster of more aggressive tumor cells based on its proximity to *SPP1*⁺ macrophages. This was achieved because of MISO's approach in extracting hierarchical tissue histology image features that capture both short-range and long-range dependencies across the tissue section, enabling an understanding of the localization of different cancer subclusters relative to *SPP1*⁺ macrophages. The ability of MISO to make such distinctions among cancerous regions of varying

severity is crucial for gaining a deeper understanding of the tumor microenvironment. In contrast, neither MUSE nor SpatialGlue could analyze this tissue section due to memory constraints.

Application to mouse embryo (E13) spatial ATAC-RNA-seq data

To examine the performance of MISO for other omics types, we next analyzed a spatial ATAC-RNA-seq dataset obtained from a mouse at embryonic day 13 (E13; Fig. 4a)⁵. Focusing on the telencephalon region of the forebrain located at the upper left of the tissue (Fig. 4b), MISO uniquely distinguished spots in the pallium's stratum from the ventricular zone (VZ), a capability not observed with MUSE and SpatialGlue (Fig. 4c). MISO's segmentation of the telencephalon subregions was validated by comparing the localization of marker gene expression for the VZ, subpallium stratum and pallium stratum (Fig. 4d). To define the VZ, we visualized expression of the *SOX2* gene, which is highly expressed

in neural stem cells and neural progenitor cells contained within the VZ of the telencephalon, but not in postmitotic neurons that have migrated into the stratum²⁶. Consequently, visualizing the expression of *SOX2* allows differentiation between the VZ and the stratum of the different subregions in the telencephalon. Spots with high expression of *SOX2* closely corresponded to cluster 3 identified by MISO, indicating that MISO effectively distinguished regions of neural stem cells and neural progenitor cells from those containing postmitotic neurons in the telencephalon.

To localize the subpallium stratum, we examined the expression of *GAD2*, a marker gene for GABAergic neurons that produces the GAD65 protein, responsible for converting glutamate to GABA²⁷. In the mouse telencephalon, GABAergic neurons originate exclusively in the stratum of the subpallium before starting migration to the pallium to help in the formation of the cortical plate²⁸. At E13, a majority of GABAergic neurons have not yet migrated, making *GAD2* a suitable marker for the subpallium stratum²⁹. We observed that high expression and chromatin accessibility of *GAD2* correspond to cluster 1 identified by MISO, indicating that this cluster likely encompasses *GAD2*⁺ GABAergic neurons in the subpallium.

To define the pallium stratum, we visualized the expression of *TBR1*, which encodes the T-box brain transcription factor 1, crucial for regulating cerebrocortical development³⁰. In the telencephalon of the developing mouse brain starting at E10, *TBR1* is expressed only in post-mitotic glutamatergic neurons, meaning that neural progenitor cells in the VZ and GABAergic neurons in the subpallium of the telencephalon do not express *TBR1* (ref. 31). In addition, in the telencephalon, *TBR1* is generally not expressed in regions containing cells that express *DLX2*, specifically the subpallium^{31,32}. Therefore, within the telencephalon regions shown in Fig. 4b, *TBR1* serves as a marker gene for cells within the stratum of the pallium. In Fig. 4d, we see high RNA expression of *TBR1* restricted to the upper periphery of the telencephalon, closely corresponding to cluster 6 identified by MISO. There is noticeably high chromatin accessibility for *TBR1* in spots contained within the VZ of the telencephalon, indicating the potential of cells within the VZ to differentiate into *TBR1*⁺ glutamatergic neurons that aid in cortical development. By effectively integrating gene expression and chromatin accessibility information, MISO can distinguish the VZ from the stratum, despite both regions showing high levels of chromatin accessibility for the *TBR1* gene.

To further confirm MISO's precision, we examined the distribution of RNA and ATAC data modalities in these areas using t-SNE plots, with spots colored based on MISO's cluster assignments (Fig. 4e). Remarkably, cluster 3 (VZ) and cluster 6 (pallium stratum) showed almost complete separation in the t-SNE plots for both modalities, underscoring MISO's ability to discern complex spatial relationships that MUSE and SpatialGlue could not. These observations collectively highlight MISO's ability to integrate complex multimodal spatial omics data, revealing brain structures that are crucial for studying neurodevelopment but otherwise obscured in traditional analysis methods.

Application to mouse hippocampus spatial transcriptomics and metabolomics data

The previous analyses have been limited to two modalities. However, as the field progresses, emerging technologies capable of simultaneously profiling three or more modalities from the same tissue slice are being developed^{6–9}. Anticipating the surge of these complex datasets, developing computational tools capable of handling such multifaceted data is imperative. In addition, given MISO's successful segmentation of the mouse embryo telencephalon, we sought to assess its ability to incorporate more than two modalities to segment finer-grained structural components of the brain. To do so, we focused on the hippocampus of the coronal section of a mouse brain analyzed using 10x Visium and MALDI mass spectrometry imaging (MALDI-MSI)⁹, producing spatial transcriptomics, spatial metabolomics and H&E histology imaging data for the same tissue section (Fig. 5a). As 10x Visium and

MALDI-MSI measure gene and metabolite expression at different spatial resolutions, we enhanced the resolution of both modalities using iStar¹⁹, enabling alignment of the different omics modalities to the same spatial coordinate space. This resolution enhancement also facilitated a more detailed segmentation of the hippocampus.

MISO was able to accurately localize clusters to biologically relevant structural components defined by the Allen Brain Atlas (Fig. 5b). For comparison, we also evaluated this dataset using MUSE and SpatialGlue. Since MUSE and SpatialGlue can only integrate two modalities, we applied MUSE to the transcriptomics and tissue histology image data, as it was not designed to take metabolomics data as input, and SpatialGlue to the transcriptomics and metabolomics data, as it was developed for spatial multi-omics analysis. A main region of interest was the pyramidal layer, which contains different subtypes of pyramidal neurons classified based on areal localization. The pyramidal layer is composed of three main subfields: CA1 stratum pyramidale (CA1sp), CA2 stratum pyramidale (CA2sp) and CA3 stratum pyramidale (CA3sp). While all methods were able to differentiate CA1sp from CA3sp, MISO was the only method that accurately localized a cluster to the much smaller CA2sp subfield.

To validate the precision of the MISO-identified subfields, we used the MERFISH whole mouse brain atlas³³. MERFISH is well-suited for validating these fine-grained subfields because it provides subcellular spatially resolved gene expression information. Based on the localization of CA2 and CA3 glutamatergic neurons and the expression patterns of CA2sp (*Amigo2*) and CA3sp (*Chrm3*) pyramidal neuron marker genes, the CA2sp/CA3sp border inferred by MERFISH corresponds precisely to the border between MISO clusters 11 and 16 (Fig. 5c)³⁴. Furthermore, MISO was able to identify all three subareas of CA3 (refs. 35,36), offering an even more detailed segmentation than can be inferred through analysis of the MERFISH gene panel.

Application to human tonsil 10x Visium spatial gene and protein expression data

As a second demonstration for MISO's ability to integrate three modalities, we analyzed a human tonsil spatial gene and protein expression dataset³⁷, comprising spatial transcriptomics, spatial proteomics (35 proteins) and histology imaging data. Figure 5d shows the germinal centers identified using the high-resolution histology image and expression of PCNA (germinal centers confirmed by E.E.F.), a protein known to have heightened expression in germinal centers³⁸. We evaluated MISO's performance across all available modalities (RNA, protein and tissue histology; Fig. 5e) and compared it with MUSE (RNA and tissue histology; Fig. 5f) and SpatialGlue (RNA and protein; Fig. 5g), each applied to the modalities that these methods were originally designed for. To quantitatively assess clustering accuracy, we focused on each method's ability to localize clusters to the germinal centers. We designated a spot as a ground truth germinal center spot if it resided within an annotated germinal center region. For each method, we defined a cluster as a germinal center cluster if more than half of the spots in that cluster were located within a germinal center. Using this classification criterion, we calculated the F1 score to evaluate the accuracy of germinal center localization for all methods (Fig. 5h). MISO, whether utilizing all three modalities or any two-modality combination, consistently outperformed MUSE and SpatialGlue in terms of F1 score. The maximum F1 score (0.90) was reached by MISO when applied to all three modalities. The lowest F1 score for MISO (0.86) was observed when using only RNA and protein as input, but this performance still surpassed most other methods and was only marginally lower than the highest F1 score achieved by MUSE (0.87) with protein and tissue histology image data as input. SpatialGlue achieved its maximum F1 score (0.83) when RNA and protein were utilized. The mean MISO F1 score across all combinations of two modalities was 0.87, compared to 0.78 for MUSE and 0.68 for SpatialGlue, underscoring the consistently strong performance of MISO across various modality pairs. These observations highlight

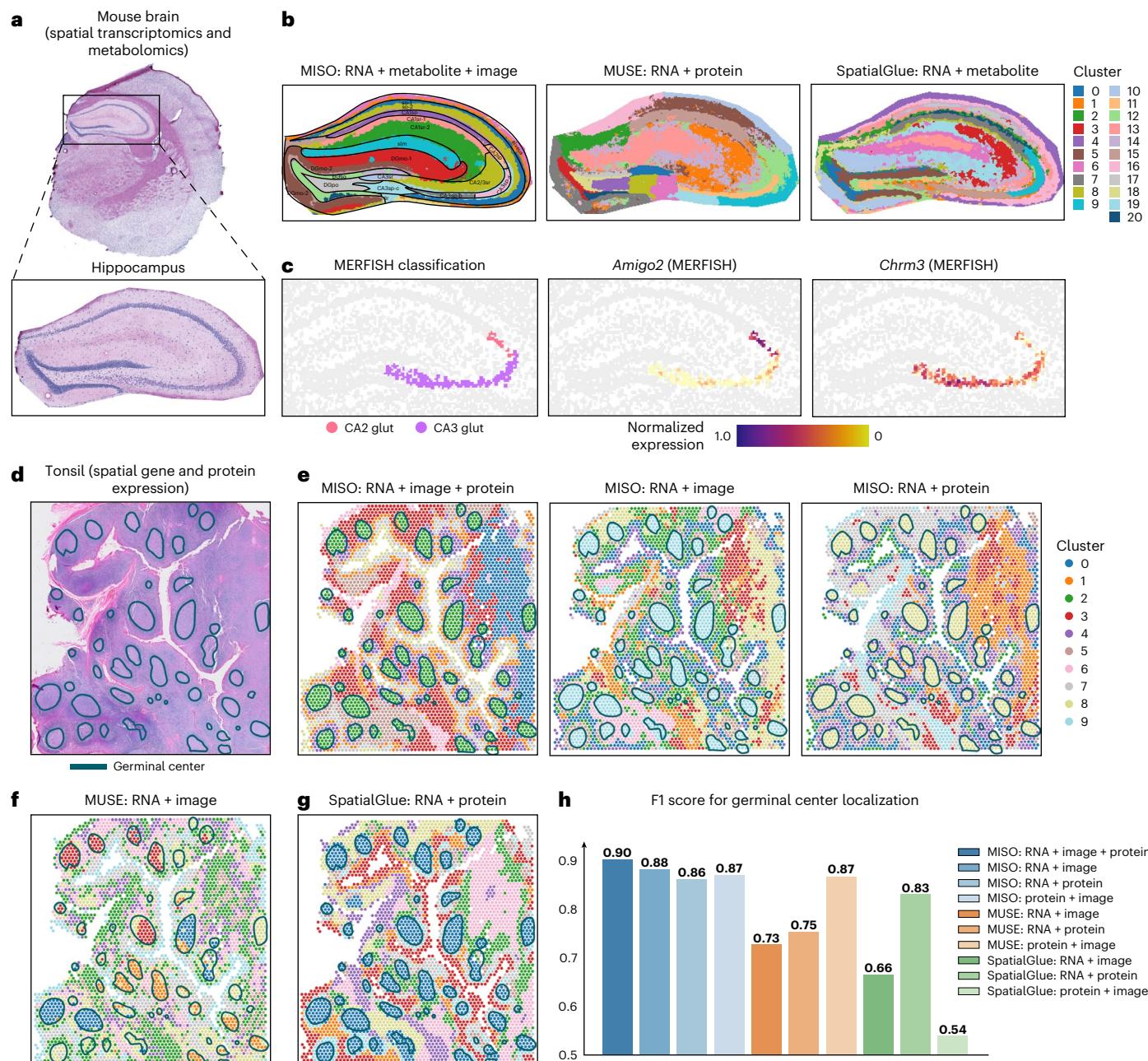


Fig. 5 | Analysis of spatial multi-omics datasets with three modalities.

a–c, Clustering results for a coronal mouse brain spatial transcriptomics and metabolomics dataset. **a**, H&E-stained histology image of an analyzed tissue section, with magnification of hippocampus. **b**, Shown from left to right are the clustering results of the super-pixels in the hippocampus from MISO (RNA, metabolite and tissue histology), MUSE (RNA and tissue histology) and SpatialGlue (RNA and metabolite). **c**, Shown from left to right are the MERFISH distribution of CA2 and CA3 glutamatergic neurons and spatial gene expression plots of *Amigo2* and *Chrm3* in the pyramidal layer of the hippocampus.

MISO's superior capability in accurately identifying complex spatial structures, regardless of the number and type of modalities involved. Such robustness is desired in real studies when multiple modalities are incorporated.

Additional applications

Finally, to showcase MISO's broad applicability to diverse cancer and healthy tissue types, we applied it to spatial omics data with two modalities, including Visium data from mouse anterior brain³⁹ (Extended Data

Fig. 2 and Supplementary Fig. 2), human breast cancer⁴⁰ (Extended Data Fig. 3), zebrafish melanoma⁴¹ (Extended Data Fig. 4), mouse olfactory bulb⁴² (Extended Data Fig. 5), mouse coronal brain⁴³ (Extended Data Fig. 6) and human prostate cancer⁴⁴ (Extended Data Fig. 7) and spatial CUT&Tag–RNA-seq data⁵ (Supplementary Figs. 3 and 4) from mouse coronal brain. We also analyzed additional spatial omics data with three modalities, including spatial gene and protein expression data from human breast cancer⁴⁵ (Extended Data Fig. 8) and mouse colon⁶ (Supplementary Fig. 5), and spatial transcriptomics and metabolomics

data from mouse coronal brain⁹ (Extended Data Figs. 9 and 10). Some of these analyses involve the integration of immunofluorescence and brightfield microscopy image data, histology image data with low resolution or stain artifact, or low RNA quality. In all of these applications, MISO demonstrated superior capability in delineating biologically relevant spatial domains and robustness in handling diverse data types.

Discussion

In summary, we have presented MISO, an algorithm capable of accommodating all modalities in multimodal spatial omics experiments. Distinct from existing methods, MISO's unique strength lies in its flexibility in handling diverse modalities, its ability to overcome the impact of low-quality modalities, and minimal user-dependent parameter tuning. Our comprehensive analyses across 16 datasets—including brain, embryo and colon in mice; melanoma in zebrafish; and bladder cancer, gastric cancer, breast cancer, prostate cancer and tonsil in humans—demonstrated that MISO outperformed MUSE and Spatial-Glue in precisely identifying known biological structures.

Currently, MISO is limited in that its feature extraction method is specific to H&E-stained histology images and cannot be applied to other histology or molecular imaging types, including Trichome, immunofluorescence and DAPI. Additionally, because MISO is an unsupervised algorithm, it requires user interpretation to associate identified spatial domains with biological or disease-relevant structures. MISO also cannot directly quantify the contribution of each modality or automatically detect low-quality modalities, requiring users to identify these through investigation of the spatial omics and imaging data. Despite these limitations, MISO's unique combination of versatility, speed, accuracy and scalability positions it as a valuable tool for modality integration and spatial domain identification in multimodal spatial omics studies. In the future, we plan to extend MISO's capabilities to include automated feature extraction for more histology and molecular imaging types, enhance the model to quantify each modality's contribution in clustering, and broaden its functionality to simultaneously analyze multiple samples. As spatial omics technologies continue to evolve, these improvements will support new data types that are anticipated in the near future.

Online content

Any methods, additional references, Nature Portfolio reporting summaries, source data, extended data, supplementary information, acknowledgements, peer review information; details of author contributions and competing interests; and statements of data and code availability are available at <https://doi.org/10.1038/s41592-024-02574-2>.

References

1. Vandereyken, K., Sifrim, A., Thienpont, B. & Voet, T. Methods and applications for single-cell and spatial multi-omics. *Nat. Rev. Genet.* **24**, 494–515 (2023).
2. Chen, K. H., Boettiger, A. N., Moffitt, J. R., Wang, S. & Zhuang, X. Spatially resolved, highly multiplexed RNA profiling in single cells. *Science* **348**, aaa6090 (2015).
3. Stickels, R. R. et al. Highly sensitive spatial transcriptomics at near-cellular resolution with Slide-seqV2. *Nat. Biotechnol.* **39**, 313–319 (2021).
4. Eng, C.-H. L. et al. Transcriptome-scale super-resolved imaging in tissues by RNA seqFISH+. *Nature* **568**, 235–239 (2019).
5. Zhang, D. et al. Spatial epigenome–transcriptome co-profiling of mammalian tissues. *Nature* **616**, 113–122 (2023).
6. Liu, Y. et al. High-plex protein and whole transcriptome co-mapping at cellular resolution with spatial CITE-seq. *Nat. Biotechnol.* **41**, 1405–1409 (2023).
7. Liao, S. et al. Integrated spatial transcriptomic and proteomic analysis of fresh frozen tissue based on Stereo-seq. Preprint at bioRxiv <https://doi.org/10.1101/2023.04.28.538364> (2023).
8. Ben-Chetrit, N. et al. Integration of whole transcriptome spatial profiling with protein markers. *Nat. Biotechnol.* **41**, 788–793 (2023).
9. Vicari, M. et al. Spatial multimodal analysis of transcriptomes and metabolomes in tissues. *Nat. Biotechnol.* **42**, 1046–1050 (2024).
10. Ghazanfar, S., Guibentif, C. & Marioni, J. C. Stabilized mosaic single-cell data integration using unshared features. *Nat. Biotechnol.* **42**, 284–292 (2024).
11. Hao, Y. et al. Dictionary learning for integrative, multimodal and scalable single-cell analysis. *Nat. Biotechnol.* **42**, 293–304 (2024).
12. Bao, F. et al. Integrative spatial analysis of cell morphologies and transcriptional states with MUSE. *Nat. Biotechnol.* **40**, 1200–1209 (2022).
13. Long, Y. et al. Deciphering spatial domains from spatial multi-omics with SpatialGlue. *Nat. Methods* **21**, 1658–1667 (2024).
14. Jiang, J. et al. METI: deep profiling of tumor ecosystems by integrating cell morphology and spatial transcriptomics. *Nat. Commun.* **15**, 7312 (2024).
15. Schumacher, T. N. & Thommen, D. S. Tertiary lymphoid structures in cancer. *Science* **375**, eabf9419 (2022).
16. Sautès-Fridman, C. et al. Tertiary lymphoid structures in cancers: prognostic value, regulation, and manipulation for therapeutic intervention. *Front. Immunol.* **7**, 407 (2016).
17. Di Caro, G. et al. Occurrence of tertiary lymphoid tissue is associated with T-cell infiltration and predicts better prognosis in early-stage colorectal cancers. *Clin. Cancer Res.* **20**, 2147–2158 (2014).
18. Asrir, A. et al. Tumor-associated high endothelial venules mediate lymphocyte entry into tumors and predict response to PD-1 plus CTLA-4 combination immunotherapy. *Cancer Cell* **40**, 318–334 (2022).
19. Zhang, D. et al. Inferring super-resolution tissue architecture by integrating spatial transcriptomics with histology. *Nat. Biotechnol.* **42**, 1372–1366 (2024).
20. van Cutsem, E. et al. HER2 screening data from ToGA: targeting HER2 in gastric and gastroesophageal junction cancer. *Gastric Cancer* **18**, 476–484 (2015).
21. Oliveira, M. F. et al. Characterization of immune cell populations in the tumor microenvironment of colorectal cancer using high definition spatial profiling. Preprint at bioRxiv <https://doi.org/10.1101/2024.06.04.597233> (2024).
22. Jantschkeff, P. et al. Expression of CEACAM6 in resectable colorectal cancer: a factor of independent prognostic significance. *Jo. Clin. Oncol.* **21**, 3638–3646 (2003).
23. Burgos, M. et al. Prognostic value of the immune target CEACAM6 in cancer: a meta-analysis. *Ther. Adv. Med. Oncol.* **14**, 17588359211072621 (2022).
24. Qi, J. et al. Single-cell and spatial analysis reveal interaction of FAP⁺ fibroblasts and SPP1⁺ macrophages in colorectal cancer. *Nat. Commun.* **13**, 1742 (2022).
25. Ozato, Y. et al. Spatial and single-cell transcriptomics decipher the cellular environment containing HLA-G⁺ cancer cells and SPP1⁺ macrophages in colorectal cancer. *Cell Rep.* **42**, 111929 (2023).
26. Bani-Yaghoub, M. et al. Role of Sox2 in the development of the mouse neocortex. *Dev. Biol.* **295**, 52–66 (2006).
27. Koopman, A. D. et al. The association between GAD65 antibody levels and incident type 2 diabetes mellitus in an adult population: a meta-analysis. *Metabolism* **95**, 1–7 (2019).
28. Petanjek, Z., Kostovic, I. & Esclapez, M. Primate-specific origins and migration of cortical GABAergic neurons. *Front. Neuroanat.* **3**, 26 (2009).
29. Kronman, F. N. et al. Developmental mouse brain common coordinate framework. *Nat. Commun.* **15**, 9072 (2024).

30. Hsueh, Y.-P., Wang, T.-F., Yang, F.-C. & Sheng, M. Nuclear translocation and transcription regulation by the membrane-associated guanylate kinase CASK/LIN-2. *Nature* **404**, 298–302 (2000).
31. Bulfone, A. et al. T-brain-1: a homolog of Brachyury whose expression defines molecularly distinct domains within the cerebral cortex. *Neuron* **15**, 63–78 (1995).
32. Puelles, L. et al. Pallial and subpallial derivatives in the embryonic chick and mouse telencephalon, traced by the expression of the genes Dlx-2, Emx-1, Nkx-2.1, Pax-6 and Tbr-1. *J. Comp. Neurol.* **424**, 409–438 (2000).
33. Zhang, M. et al. Molecularly defined and spatially resolved cell atlas of the whole mouse brain. *Nature* **624**, 343–354 (2023).
34. Hamilton, D., White, C., Rees, C., Wheeler, D. & Ascoli, G. Molecular fingerprinting of principal neurons in the rodent hippocampus: a neuroinformatics approach. *J. Pharm. Biomed. Anal.* **144**, 269–278 (2017).
35. Lorente de Nò, R. Studies on the structure of the cerebral cortex. II. Continuation of the study of the ammonic system. *Journal für Psychologie und Neurologie* **46**, 113–177 (1934).
36. Li, X. G., Somogyi, P., Ylinen, A. & Buzsáki, G. The hippocampal CA3 network: an in vivo intracellular labeling study. *J. Comp. Neurol.* **339**, 181–208 (1994).
37. 10xGenomics. <https://www.10xgenomics.com/resources/datasets/gene-protein-expression-library-of-human-tonsil-cytassit-ffpe-2-standard> (2023).
38. Lipponen, P. K. & Eskelinen, M. J. Cell proliferation of transitional cell bladder tumours determined by PCNA/cyclin immunostaining and its prognostic value. *Br. J. Cancer* **66**, 171–176 (1992).
39. 10xGenomics. <https://www.10xgenomics.com/resources/datasets/mouse-brain-serial-section-1-sagittal-anterior-1-standard-1-1-0> (2020).
40. 10xGenomics. <https://www.10xgenomics.com/resources/datasets/human-breast-cancer-visium-fresh-frozen-whole-transcriptome-1-standard> (2022).
41. Hunter, M. V., Moncada, R., Weiss, J. M., Yanai, I. & White, R. M. Spatially resolved transcriptomics reveals the architecture of the tumor-microenvironment interface. *Nat. Commun.* **12**, 6278 (2021).
42. 10xGenomics. <https://www.10xgenomics.com/resources/datasets/adult-mouse-olfactory-bulb-1-standard-1> (2022).
43. 10xGenomics. <https://www.10xgenomics.com/resources/datasets/mouse-brain-coronal-section-2-ffpe-2-standard> (2022).
44. Erickson, A. et al. Spatially resolved clonal copy number alterations in benign and malignant tissue. *Nature* **608**, 360–367 (2022).
45. 10xGenomics. <https://www.10xgenomics.com/resources/datasets/gene-and-protein-expression-library-of-human-breast-cancer-cytassit-ffpe-2-standard> (2023).

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

Springer Nature or its licensor (e.g. a society or other partner) holds exclusive rights to this article under a publishing agreement with the author(s) or other rightsholder(s); author self-archiving of the accepted manuscript version of this article is solely governed by the terms of such publishing agreement and applicable law.

© The Author(s), under exclusive licence to Springer Nature America, Inc. 2025

¹Department of Computational Biomedicine, Cedars-Sinai Medical Center, Los Angeles, CA, USA. ²Statistical Center for Single-Cell and Spatial Genomics, Department of Biostatistics, Epidemiology and Informatics, Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA, USA.

³Department of Pathology, Seoul Metropolitan Government-Seoul National University Boramae Medical Center, Seoul National University College of Medicine, Seoul, Korea. ⁴Department of Pathology, College of Medicine, Kyung Hee University Hospital, Kyung Hee University, Seoul, Republic of Korea.

⁵Department of Microbiology, Institute for Immunology, and Institute for Diabetes, Obesity and Metabolism, Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA, USA. ⁶Faculty of Biology, University of Freiburg, Freiburg, Germany. ⁷Division of Genetics and Genomics, Department of Pediatrics, Boston Children's Hospital, Harvard Medical School, Boston, MA, USA. ⁸Department of Genitourinary Medical Oncology, The University of Texas MD Anderson Cancer Center, Houston, TX, USA. ⁹Department of Genomic Medicine, The University of Texas MD Anderson Cancer Center, Houston, TX, USA. ¹⁰School of Public Health, The University of Texas Health Science Center at Houston, Houston, TX, USA. ¹¹Department of Chemistry, Lewis Sigler Institute for Integrative Genomics, and Ludwig Institute for Cancer Research, Princeton University, Princeton, NJ, USA. ¹²Department of Artificial Intelligence and Informatics, Mayo Clinic, Jacksonville, FL, USA. ¹³Department of Pathology and Laboratory Medicine, Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA, USA. ¹⁴Epigenetics Institute, Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA, USA.

¹⁵Department of Pathology, The University of Texas MD Anderson Cancer Center, Houston, TX, USA. ¹⁶The University of Texas MD Anderson Cancer Center UTHealth Graduate School of Biomedical Sciences, Houston, TX, USA. ¹⁷Department of Human Genetics, School of Medicine, Emory University, Atlanta, GA, USA.  e-mail: kyle.coleman@cshs.org; jian.hu@emory.edu; mingyao@pennmedicine.upenn.edu

Methods

Input data

MISO is designed to accommodate a wide range of omics and tissue histology image data from any multimodal spatial omics experiment as input. When utilizing RNA, protein or chromatin accessibility data as input modalities, MISO is capable of automatically handling raw data and performing all necessary preprocessing steps using Scanpy. For RNA, protein and chromatin accessibility data, MISO takes count matrices as input and performs log-normalization. In the case of other types of omics data, such as metabolomics and histone modifications, MISO assumes that the data have been preprocessed. If an H&E-stained histology image is available, MISO can directly accept the raw, high-resolution image and conduct image feature extraction. For other image types, such as microscopy and immunofluorescence imaging, MISO requires preprocessed image features for each spot in the corresponding omics dataset. MISO performs z-score normalization for each input modality.

Extraction of H&E-stained histology image features

Many multimodal spatial omics technologies have the capability to generate high-resolution H&E-stained histology images of analyzed tissue sections. However, the pixel-level resolution of the histology image surpasses that of the corresponding omics data, which typically have spots that are a few hundred pixels in diameter. Consequently, it is imperative to extract histology features for each spot in the dataset to facilitate the integration of tissue histology and omics features. In our approach, we aimed to capture both local and global structures of the histology image in spot-level feature vectors. To accomplish this, we used a hierarchical Vision Transformer (ViT) model pretrained on a substantial dataset of over 10,000 whole-slide images⁴⁶. The model, made available by the Hierarchical Image Pyramid Transformer (HIPT) software, divides the whole-slide image into sub-images of size 256×256 pixels, utilizing each sub-image as input for a local ViT model to extract low-level image features. For a given sub-image, this model assigns 384-dimensional patch embeddings to all 16×16 -pixel patches via linear projection, which are used to generate a 192-dimensional classification (CLS) token for the entire 256×256 -pixel image. The CLS tokens for all 256×256 -pixel images within the same $4,096 \times 4,096$ -pixel region are used as input for a global ViT model to extract high-level image features. The global ViT model aggregates information across these CLS tokens, producing embeddings for each 256×256 -pixel image that capture the long-range dependencies across the tissue section. To construct spot-level histology features, we first assign a 576-dimensional embedding to each 16×16 -pixel patch. This embedding comprises the 384-dimensional embedding from the local ViT model and the 192-dimensional embedding from the global ViT model of the 256×256 -pixel image containing that 16×16 -pixel patch. This process ensures that the histology embeddings encompass both the local and global structures of the image. Subsequently, all 16×16 -pixel patches completely contained within a given omics spot are considered, and their embeddings are averaged to obtain a 576-dimensional embedding for each spot in the dataset. A detailed illustration of the tissue histology image feature extraction procedure is shown in Supplementary Fig. 6.

Extraction of deep modality-specific features

The feature extraction step in MISO relies on a spectral clustering framework. This approach leverages the distances between spots in the raw feature space to generate lower-dimensional spot-level embeddings, meticulously capturing the graph structure inherent in the data. The choice of a spectral clustering methodology for this task is driven by the advantages it offers over alternative clustering methods that impose rigid assumptions about the input data structure. Spectral clustering stands out due to its ability to operate effectively without imposing stringent requirements on the underlying data structure. This adaptability allows it to navigate through complex input structures

and adeptly identify non-convex clusters, a critical capability in domains like gene expression and tissue histology data where clusters often exhibit complex shapes and patterns⁴⁷. The utilization of spectral clustering in MISO's feature extraction not only enhances the algorithm's robustness but also ensures its applicability to real-world datasets characterized by diverse and nuanced structures.

To apply a spectral clustering-based feature extraction method, it is first necessary to transform each input modality into a graph representation. To achieve this, for each modality, we construct an adjacency matrix that represents the similarity between each pair of spots with respect to that modality. For a given modality m , each element of the adjacency matrix is constructed as

$$A^m = \left[\exp\left(\frac{-\|x_i^m - x_j^m\|_2^2}{2\sigma^2}\right) \right]_{i,j},$$

where x_i^m and x_j^m are the modality m input features for spots i and j , respectively, and σ^2 is a scaling parameter that was set to 30 for all modalities in all analyses (based on empirical evidence).

Motivated by SpectralNet⁴⁸, MISO utilizes neural networks for the extraction of spectral clustering features. We start by extracting the top 128 principal components (PCs) for each modality, giving us an initial lower-dimensional representation of the data. For each specific modality, we then use the corresponding adjacency matrix and spot-level PC feature vectors in the training of a deep neural network. This network aims to minimize both reconstruction and spectral clustering loss functions, ensuring a comprehensive understanding and capture of complex patterns within the data. The PC feature vectors for each modality serve as input for an autoencoder with a bottleneck layer of 32 nodes. The matrix of vectors output from the bottleneck layer, Y^m , will be used to cluster the spots after the network has been trained. The spectral clustering loss function for the modality m autoencoder is given by

$$L_{SC}^m = \sum_{i,j} A_{i,j}^m \|y_i^m - y_j^m\|_2^2,$$

where y_i^m is the output of the bottleneck layer corresponding to modality m for spot i . Defining the spectral clustering loss function in this way ensures that spots that are similar with respect to the adjacency matrix will have similar embeddings for that modality. We require that Y^m is orthonormal so that all spots do not have the same embedding. Each autoencoder is also trained to minimize a reconstruction loss, that is,

$$L_R^m = \sum_i \|z_i^m - \hat{z}_i^m\|_2^2,$$

where z_i^m is the modality m PC feature vector for spot i and \hat{z}_i^m is the corresponding output of the decoder layer. The total loss for autoencoder m is a weighted sum of the spectral clustering and reconstruction losses:

$$L^m = L_{SC}^m + \alpha L_R^m,$$

where α , set to 1 by default, is a user-defined hyperparameter that determines the amount of weight given to the reconstruction loss relative to the spectral clustering loss. A higher value for α means that the extracted features will be more accurate representations of the original input data, but the embeddings for similar spots may be more distinct. All results provided in this paper were obtained by training the model with alpha equal to 1. The autoencoders are trained separately for each modality for 1,000 epochs using the Adam optimizer⁴⁹ with a learning rate of 0.001.

Cross-modality feature fusion

The autoencoders described in the previous step associate each spot with a low-dimensional modality-specific vector representation. In many existing methods for multimodal data integration, embeddings for each modality are simply concatenated before utilization in downstream analyses. However, this does not account for the interactions between the different modalities. In the context of multimodal spatial omics, interactions between different modalities can have a great impact on phenotypic properties that are not evident when considering each modality separately. For instance, when considering RNA-seq and ATAC-seq data, interaction features are necessary to uncover the organization of gene regulatory networks by modeling the relationships between expression levels of regulated genes and regions of open chromatin⁵⁰. For a given pair of modalities m_1 and m_2 , we calculate the interaction feature matrix for spot i as

$$y_i^{m_1 \otimes m_2} = y_i^{m_1} \otimes y_i^{m_2} = \begin{pmatrix} y_{i,1}^{m_1} y_{i,1}^{m_2} & \cdots & y_{i,1}^{m_1} y_{i,32}^{m_2} \\ \vdots & \ddots & \vdots \\ y_{i,32}^{m_1} y_{i,1}^{m_2} & \cdots & y_{i,32}^{m_1} y_{i,32}^{m_2} \end{pmatrix}.$$

For each pair of modalities, the interaction feature matrices are flattened for all spots and combined to form a matrix with spot-level interaction features. Subsequently, PC analysis is applied to each matrix of interaction features, extracting the top 32 PCs to ensure uniform dimensionality with modality-specific feature vectors. This step enhances compatibility for downstream analyses. While our general recommendation is to include all pairs of interactions in clustering, MISO has a hyperparameter that allows users to specify which modality interactions they wish to incorporate, offering tailored control over the clustering process based on specific research needs. The selected interaction feature vectors are concatenated with the modality-specific feature vectors and used as input for the k -means clustering algorithm.

Scaling to large datasets

Since the MISO algorithm involves the calculation and storage of an $n \times n$ adjacency matrix for a dataset with n spots, it is not well-suited for datasets containing tens or hundreds of thousands of spots, such as those generated by molecular imaging-based spatial omics technologies like MERFISH. For these situations, we have introduced a modification to the MISO algorithm that facilitates the analysis of large-scale multimodal spatial omics datasets. To effectively handle large datasets, we construct the adjacency matrix for modality m as

$$A_{i,j}^m = \begin{cases} \exp\left(\frac{-\|x_i^m - x_j^m\|_2^2}{2\sigma^2}\right), & j \in \text{KNN}(i) \\ 0, & \text{otherwise.} \end{cases}$$

Specifically, if spot j is one of the k -nearest neighbors of spot i , the j^{th} entry in the spot i adjacency vector is set to the same value as the previously defined similarity measure. However, if spot j is not among the k -nearest neighbors of spot i , the entry is set to 0. For modality m , the k -nearest neighbors of a given spot are determined by the distances between the preprocessed modality m input feature vectors. The default value for k is 100. This modification allows us to store the adjacency matrix as a sparse matrix, greatly reducing the storage requirements for large datasets.

Guidance on feature selection

MISO necessitates users to designate modalities characterized by low quality to determine which modality-specific feature vectors to include in clustering. In the case of histology imaging data, quality assessment involves a visual inspection of the image. Indicators of low quality include artifacts such as ice crystal residue, uneven staining and variations in blurriness across the tissue. These elements serve as

evidence that the image may be of suboptimal quality. On the other hand, for omics data, the evaluation of quality involves the calculation of total unique molecular identifier counts across spots in the dataset. Additionally, assessing the total detected number of genes relative to other gene expression data from the same tissue type provides valuable insights into the quality of transcriptomics data. For MALDI-MSI spatial metabolomics data, we recommend an evaluation of the spatial localization of a subset of metabolites. In high-quality data, a majority of the metabolite signal is contained within the tissue as opposed to the background. In addition, data quality can be assessed using the percentage of detected high mass tissue-specific peaks (above a mass-to-charge (m/z) value of 2,000)⁵¹. An indication of high data quality is if 30% of the tissue-specific peaks are above an m/z of 2,000. Comparison of this percentage with other samples of the same tissue type should also provide a better understanding of the quality of the data. For spatial protein sequencing data, the quality can be inferred through visualization of the spatial expression patterns of the proteins. If there is no spatial variation in the expression levels for the proteins in the panel, then the data are likely of low quality. In addition, to include the protein-specific terms, the protein panel should be reasonably large (at least 35 proteins). When including multiplex RNA or protein imaging data, the most important influence on data quality is cell segmentation accuracy. This can be evaluated by visualizing the cell boundaries relative to the locations of cell nuclei using an available tissue stain (for example, H&E or DAPI). For ATAC-seq, quality can be assessed through signal-to-background ratio using transcription start site enrichment scores.

To address low-quality modalities, MISO simplifies the decision-making process by requiring users to make a binary determination on whether to include the modality-specific features for each modality in clustering based on its quality. By requiring a binary decision on modality inclusion based on quality, MISO offers versatility across multimodal spatial omics data with minimal preprocessing requirements. This user-centric design not only enhances the ease of application but also ensures broader adaptability, making MISO an accessible and robust choice for researchers working with diverse types of data in the multimodal spatial omics domain.

ICC

The ICC is a commonly used metric to evaluate the degree of similarities for observations within the same cluster. For a given cluster i and modality k , the ICC is defined as

$$\text{ICC}_i^{(k)} = \frac{\sigma_i^2}{\sigma_i^2 + \sigma_m^2},$$

where σ_i^2 is the within-cluster variance for cluster i , and σ_m^2 is the variance across cluster means of features from modality k . A higher ICC indicates more coherent clusters with respect to that modality and, therefore, suggests superior clustering results. For each modality, we computed the ICC for all clusters using the top 50 PCs of the preprocessed data for that modality. For all comparisons of ICC values across different methods, P values were obtained using two-sample t -tests.

Gastric cancer 10x Xenium data generation

Formalin-fixed, paraffin-embedded sections of signet-ring cell carcinoma were cut with 5-μm thickness into corresponding Xenium slides. Deparaffinization was performed by incubating slides at 60 °C for 2 h, followed by 10 min in xylene jar 1 and 3 min in xylene jar 2, 100% ethanol jar 1, 100% ethanol jar 2, 96% ethanol jar 1, 96% ethanol jar 2 and 70% ethanol. After a 20-s water rinse, RNA decrosslinking was performed at 80 °C for 30 min. The 10x Genomics Human Multi-Tissue and Cancer gene panel VI was used to hybridize probes for 377 genes for 18 h at 50 °C. Following 10x Genomics protocol no. CG000582, revision E, unhybridized probes were washed away at 37 °C for 30 min, followed by

probe ligation for 2 h at 37 °C, and rolling circle amplification at 30 °C for 2 h. After washing, background autofluorescence was quenched, and nuclei were fluorescently stained. Prepared tissue sections were then imaged using the Xenium Analyzer instrument following 10x Genomics protocol CG000584, revision C. The instrument automatically performed fluorescent probe hybridization, imaging and probe removal as a series of 15 cycles using firmware version 1.7.6.0. The Xenium Analyzer then automatically processed and analyzed the images by performing image co-registration, probe quality filtering, nuclei and cell segmentation using software analysis version Xenium 1.7.1.0.

Mouse coronal brain spatial transcriptomics and metabolomics data generation

Visium slide preparation for MALDI imaging mass spectrometry: Visium (10x Genomics) slides were adapted for imaging mass spectrometry by trimming off 1 mm of the slide width using a diamond-tip scriber, followed by sanding the trimmed edge with sandpaper while preventing glass debris from covering the fiducial frames and the tissue capture areas on the slide during sanding. Further clearing of any debris was done by blowing nitrogen gas onto the slide surface. Resizing the Visium slide was necessary for fitting it into a slide adaptor (Bruker Daltonics) later for MALDI imaging mass spectrometry.

Brain tissue preparation for MALDI imaging mass spectrometry: Fresh brain tissue samples were obtained from 9-month-old mice without perfusion. Tissues were immediately snap-frozen in liquid nitrogen without embedding. Coronal brain sections were cut frozen at 12-μm thickness using a cryostat (Leica, CM3050S) at -20 °C, thaw mounted on adapted Visium 10x slides and dehydrated in a vacuum chamber for 15 min. N-(1-naphthyl) ethylenediamine dihydrochloride (NEDC) was used as the MALDI matrix at a concentration of 10 mg ml⁻¹ in a 70:30 (vol:vol) methanol:water mixture and sprayed on the tissue sections using a HTX-TM sprayer (HTX Technologies) under the following spraying conditions: nozzle temperature, 70 °C; matrix solution flow rate, 0.1 ml min⁻¹; nitrogen gas pressure, 10 psi; nozzle velocity, 1,200 mm min⁻¹; nozzle height, 40 mm; spray spacing, 2 mm; ten passes and 10 s of drying time between each pass.

MALDI imaging mass spectrometry: All MALDI imaging mass spectrometry experiments were done using a MALDI-2 timsTOF flex (Bruker Daltonics) equipped with microGRID technology. Mass calibration was performed using a Tuning Mix solution (Agilent Technologies, G1969-85000) before starting the imaging run. The laser Smart Beam was set to 'Custom' mode while engaging the 'Beam scan' to deliver a single burst of 200 laser shots at a frequency of 10 kHz in a pixel area of 20 × 20 μm². The raster step size was set to 20 μm in x and y directions. Acquisitions were recorded in negative ion mode at an *m/z* range of 100–1,000. Energy difference between the collision cell and the quadrupole was set to 8 eV with a collision RF of 2,000 Vpp. The quadrupole ion energy was set to 8 eV and the minimum *m/z* was set to 150, with the pre-TOF transfer time set to 50 μs. Lock masses of *m/z* 124.0068 (taurine), *m/z* 255.2330 (palmitate), *m/z* 346.0558 (AMP) and *m/z* 426.0222 (ADP) were used for online calibration. All MALDI imaging experiments were done without engaging trapped ion mobility separation.

Spatial transcriptomics with 10x Visium: The Visium Spatial Gene Expression Reagent Kit (10x Genomics, PN-1000189) was used to capture spatially resolved transcriptomic data from the tissue sections according to the manufacturer's protocol (CG000239). H&E images were acquired with a Leica Aperio Slidescanner. The optimal tissue permeabilization time was determined with the Visium Spatial Optimization Kit (PN-1000191) according to the manufacturer's instructions to be 25 min (CG000238). Libraries for sequencing were prepared with 10x genomics library construction kits (PN-1000196).

Sequencing and data analysis: Prepared libraries were sequenced on an Illumina platform (NextSeq 2000, Illumina) using the P2 kit (100 cycles) to generate high-throughput transcriptomic data. Raw sequencing data were processed using the Space Ranger pipeline

(10x Genomics) to align reads, perform spot calling and generate a spatially resolved gene expression matrix.

Alignment of 10x Visium and MALDI-MSI data: The 10x Visium and MALDI-MSI measure omics data at different spatial resolutions. Specifically, 10x Visium measures transcriptomics data at spots that are 55 μm in diameter, whereas MALDI-MSI measures metabolomics data at spots that are 20 μm in diameter. To obtain data with the same spatial resolution, we ran iStar¹⁹, a method that enhances the resolution of spatial omics data using histology image information. iStar predicts molecular expression at super-pixels of 16 × 16 pixels in size across the entire tissue section. The iStar output for the 10x Visium and MALDI-MSI data are thus aligned in the same coordinate space, facilitating their integration.

Pseudocode

Number of inputs: X_1, \dots, X_n (matrices of features for each of the n modalities)

```

for  $i$  in range( $n$ ):
     $X_i^{PC}$  = PCA( $X_i$ )[:, range(128)]
     $A_i$  = calculate_affinity( $X_i$ )
    while iteration < 1,000:
         $Y_i$  = Encoder( $X_i^{PC}$ )
         $X_i^{PC}$  = Decoder( $Y_i$ )
         $L_R$  = MSE( $X_i^{PC}, X_i^{PC}$ )
         $L_{SC}$  = Spectral_loss( $A_i, Y_i$ )
         $L = L_R + \alpha L_{SC}$ 
        Update weights to minimize  $L$ 
    interactions = [ $Y_i \otimes Y_j$  for  $i, j \in \text{range}(n)$  s.t.  $i \neq j$ ]
     $Y = [Y_i \text{ for } i \in \text{range}(n)]$ 
    interactions = [PCA( $Z$ )[:, range(32)] for  $Z$  in interactions]
    emb = concatenate( $Y$ , interactions)
    clusters = KMeans(emb)

```

Reporting summary

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

References

46. Chen, R. J. et al. Scaling vision transformers to gigapixel images via hierarchical self-supervised learning. In Proc. IEEE/CVF Conference on Computer Vision and Pattern Recognition 16144–16155 (2022).
47. Ng, A., Jordan, M. & Weiss, Y. On spectral clustering: analysis and an algorithm. *Adv. Neural Inf. Process. Syst.* **14**, 849–856 (2001).
48. Shaham, U. et al. Spectralnet: Spectral clustering using deep neural networks. Preprint at <https://doi.org/10.48550/arXiv.1801.01587> (2018).
49. Kingma, D. P. & Ba, J. Adam: a method for stochastic optimization. In Proc. International Conference on Learning Representations (ICLR, 2015)
50. Lowe, E. K., Cuomo, C., Voronov, D. & Arnone, M. I. Using ATAC-seq and RNA-seq to increase resolution in GRN connectivity. *Methods Cell. Biol.* **151**, 115–126 (2019).
51. Høiem, T. S. et al. An optimized MALDI MSI protocol for spatial detection of tryptic peptides in fresh frozen prostate tissue. *Proteomics* **22**, e2100223 (2022).

Acknowledgements

M. Li was partly supported by National Institutes of Health (NIH) grants R01HG013185, R01LM014592, R01EY030192, U19NS135582, R01HL171595 and U01CA294518. L.W. was partly supported by NIH grants R01CA266280, U01CA264583, U01CA294518 and U24CA274274; the start-up research fund provided by the University of Texas MD Anderson Cancer Center; The Break Through Cancer Foundation; and the Andrew Sabin Family Foundation. L.W. and J.J.

were also supported by NIH grant R01 CA254988. T.H.H. was partly supported by NIH grants R01CA276690 and U01CA294518, DOD grant CA190578, the Eric and Wendy Schmidt Foundation's AI Innovation Award through the Mayo Clinic Foundation, and the Torrey Coast Foundation. X.Q. was partly supported by NIH grant K99NS135123. J.G. was partly supported by the Doris Duke Clinical Scientist Development Award (2018097), MD Anderson Faculty Scholar Award, the David H. Koch Center for Applied Research of Genitourinary Cancers, Wendy and Leslie Irvin Barnhart Fund, Joan and Herb Kelleher Charitable Foundation, KCA Advanced Discovery Award, the Williams TNT Fund, the V Foundation Translational Award, the DOD KCRP Translational Research Partnership Award, NIH/NCI R01 CA254988-01A1, NIH/National Cancer Institute (NCI) R01 CA269489-01A1 and NIH/NCI R01 CA282282-01; as well as in part by the Cancer Center Support Grant to MDACC (P30 CA016672) from the NCI, by MD Anderson's Prometheus informatics system and by the Department of Genitourinary Medical Oncology's Eckstein and Alexander Laboratories. J.D.R. was supported by Ludwig Cancer Research, the Penn Diabetes Research Center grant (P30-DK19525) and the Chan Zuckerberg Initiative DAF (2023-331955), an advised fund of Silicon Valley Community Foundation.

Author contributions

This study was conceived of and led by M. Li and J.H. K.C. designed the model and algorithm with input from M. Li and J.H., implemented the MISO software and led data analyses. A.S., M. Loth and H.Y. performed data analyses. D.Z. proposed the histology image feature extraction approach. T.H.H., J.H.P., J.-Y.S., J.R.C., I.J., M.K. and I.B. generated and processed the Xenium gastric cancer data. J.H.P. annotated the Xenium gastric cancer data. J.-Y.S. annotated the Visium HD colon cancer data. L.W., J.G., J.C., A.L. and J.J. generated and processed the Visium bladder cancer data. A.L. annotated the Visium bladder cancer data. C.A.T., J.D.R., N.B., A.J.C. and L.Z.S. generated and processed the mouse brain spatial transcriptomics and metabolomics data.

X.Q. annotated the hippocampus region and interpreted the results of the mouse brain spatial transcriptomics and metabolomics data. Y.D. provided input for the spatial CUT&Tag-RNA-seq data analysis. E.B.L. provided input for mouse brain data analysis. E.E.F. confirmed tissue annotation and provided input for interpretation of the Visium HD human colon cancer data. K.C. and M. Li wrote the paper with feedback from the other co-authors.

Competing interests

M. Li receives research funding from Biogen unrelated to the current manuscript. M. Li and D.Z. are cofounders of OmicPath AI. T.H. is a cofounder of Kure.ai therapeutics and has received consulting fees from IQVIA; these affiliations and financial compensations are unrelated to the current paper. The other authors declare no competing interests.

Additional information

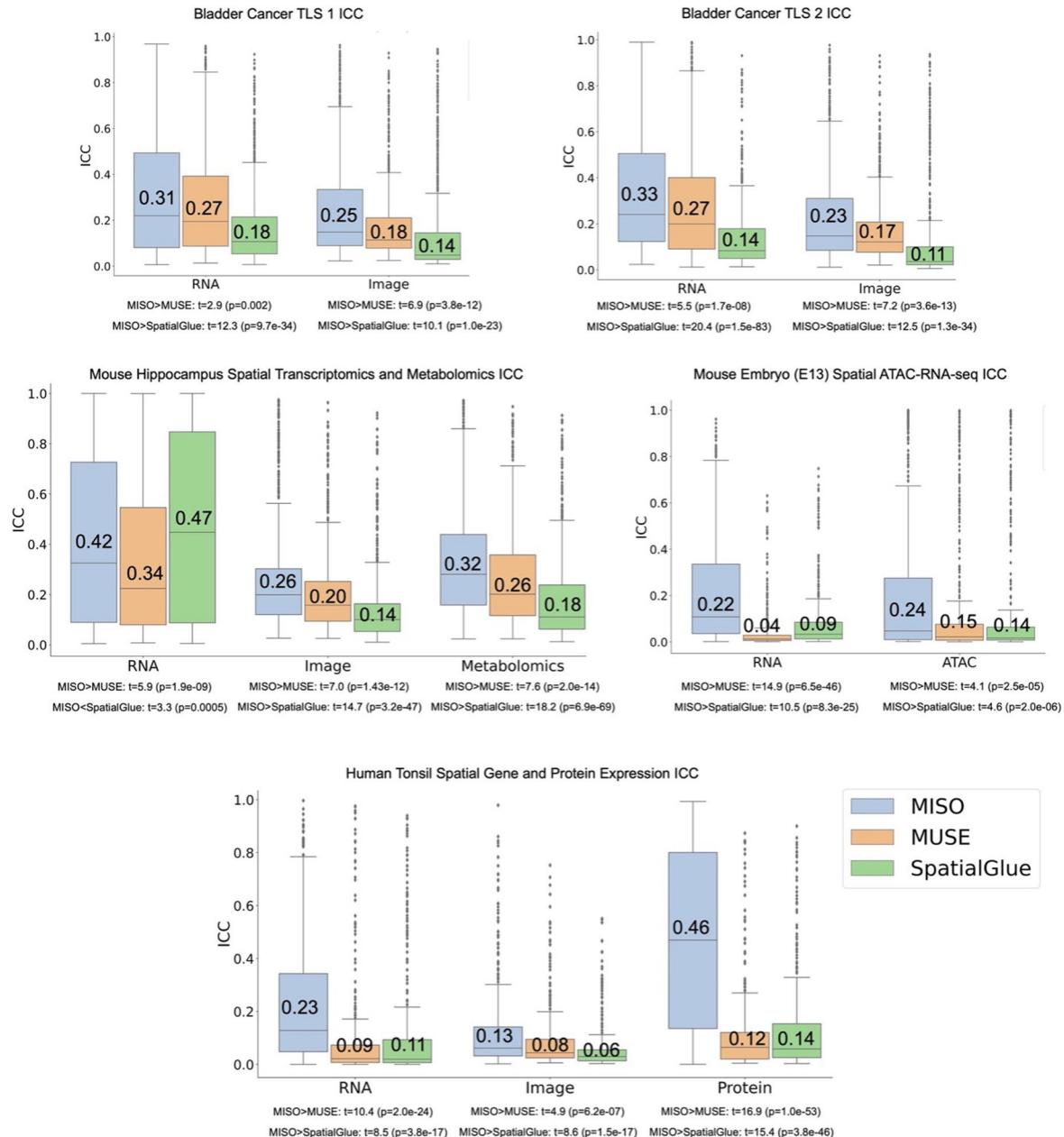
Extended data is available for this paper at
<https://doi.org/10.1038/s41592-024-02574-2>.

Supplementary information The online version contains supplementary material available at <https://doi.org/10.1038/s41592-024-02574-2>.

Correspondence and requests for materials should be addressed to Kyle Coleman, Jian Hu or Mingyao Li.

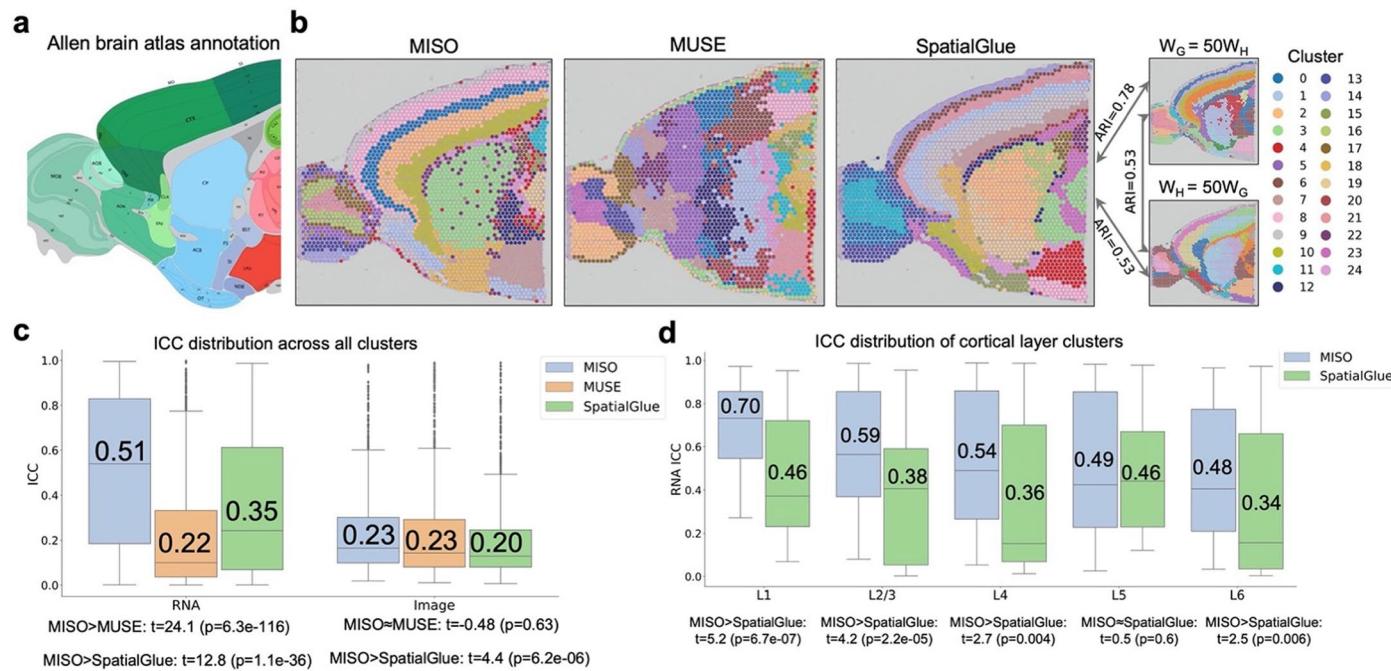
Peer review information *Nature Methods* thanks the anonymous reviewers for their contribution to the peer review of this work. Peer reviewer reports are available. Primary Handling Editor: Rita Strack, in collaboration with the *Nature Methods* team.

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



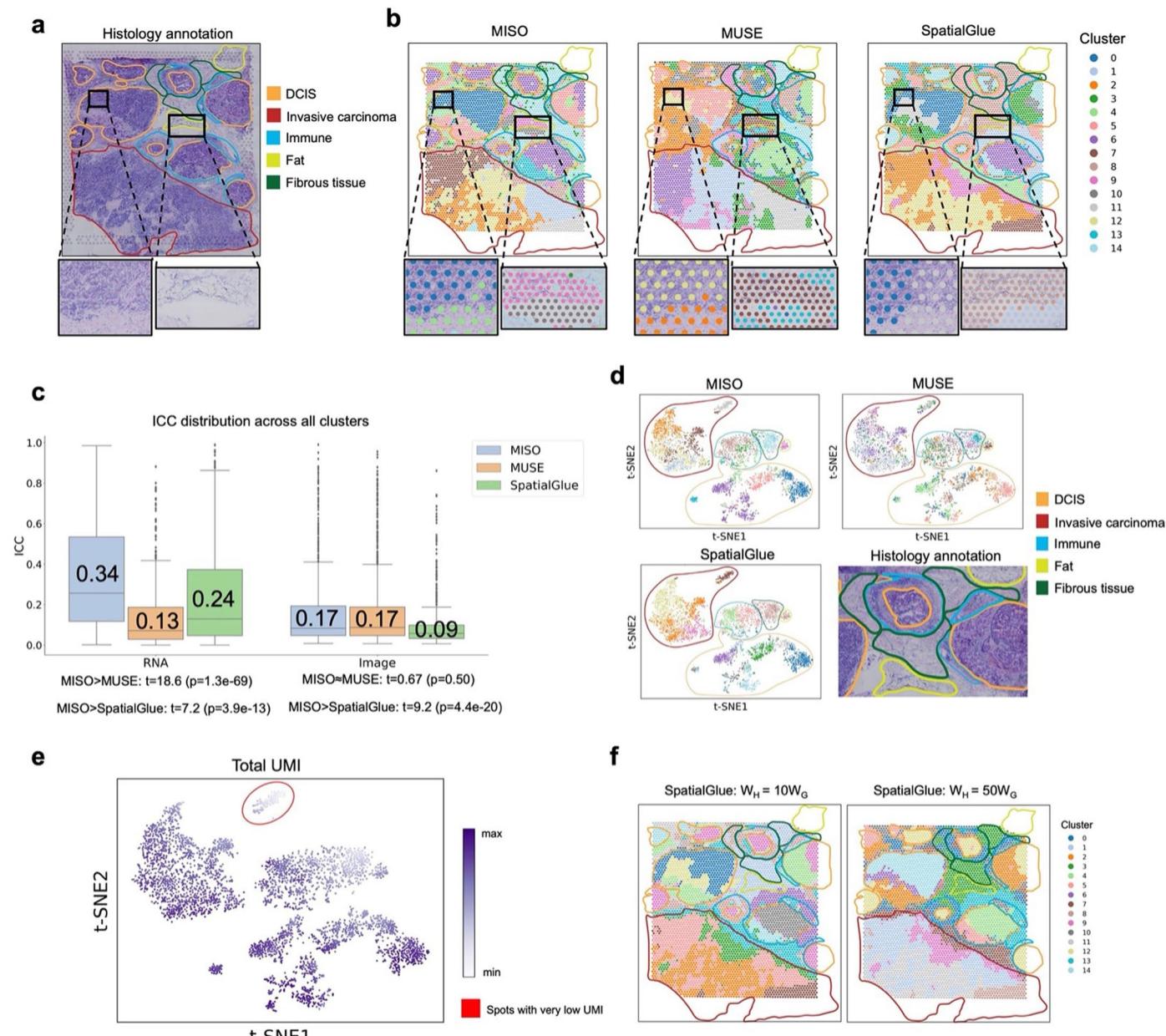
Extended Data Fig. 1 | Intraclass correlation coefficient (ICC) results for all datasets provided in Figs. 2–5 that were evaluated using MISO, MUSE, and SpatialGlue. The mean ICC for each method and each modality is printed on the corresponding box plot. Test statistics and p-values were obtained using one-sided t-tests ($n = 1250$ ICC values for each group). For a vast majority of the modalities across all datasets, the MISO clustering results produced a higher ICC compared to the other methods. The only instance in which the MISO

ICC was lower was for the RNA modality in the mouse hippocampus spatial transcriptomics and metabolomics dataset, where the ICC for SpatialGlue surpassed that of MISO. The likely cause of this is that, because the RNA data was of low quality, MISO did not use the RNA-specific terms in clustering, and only accounted for this modality through the RNAxImage and RNAxMetabolite interaction terms. Box plots: center line, median; box limits, upper and lower quartiles; whiskers, 1.5x interquartile range; points, outliers.



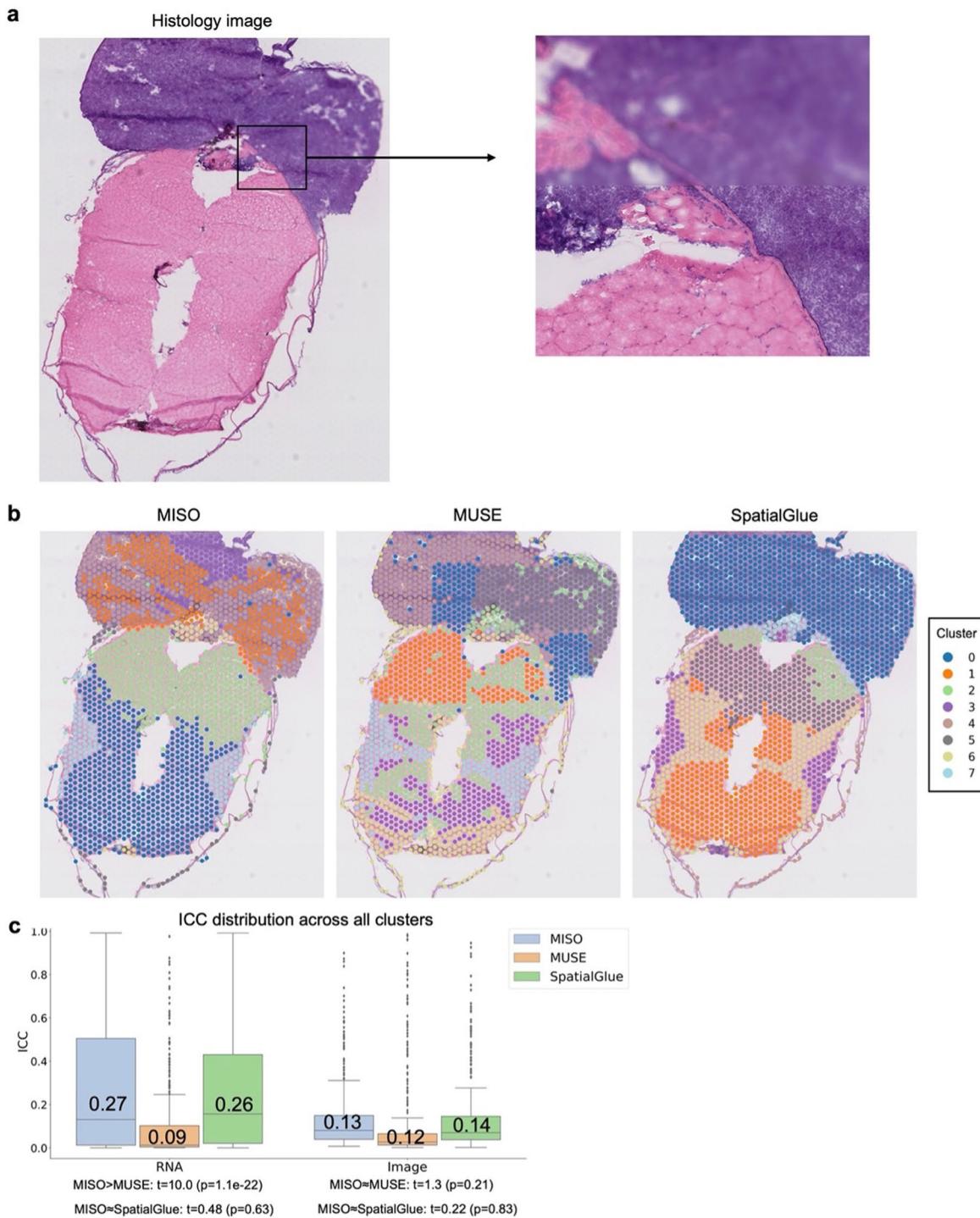
Extended Data Fig. 2 | Clustering results for a mouse anterior brain spatial transcriptomics dataset. **a**, Allen Brain Atlas annotation of mouse anterior brain. **b**, Shown from left to right are clustering results from MISO, MUSE, and SpatialGlue, respectively. SpatialGlue is sensitive to weight specified for each modality. W_G is the weight for gene expression and W_H is the weight for histology. Adjusted Rand Index (ARI) is calculated between SpatialGlue clustering with different weights. **c**, RNA and image ICC distributions across all clusters and features for each method in the mouse anterior brain data ($n = 1250$ ICC values for each group). The mean ICC for each method and each modality is printed on

the corresponding box plot. Test statistics and p-values were obtained using one-sided ($<$, $>$) or two-sided (\approx) t-tests. **d**, MISO and SpatialGlue RNA ICC distributions for all clusters corresponding to the cortical layers in the mouse anterior brain data ($n = 50$ ICC values for each group except SpatialGlue L2/3 and SpatialGlue L6, which contain $n = 100$ ICC values). Test statistics and p-values were obtained using one-sided ($<$, $>$) or two-sided (\approx) t-tests. **e**, Illustration of image artifact. Box plots: center line, median; box limits, upper and lower quartiles; whiskers, 1.5x interquartile range; points, outliers.



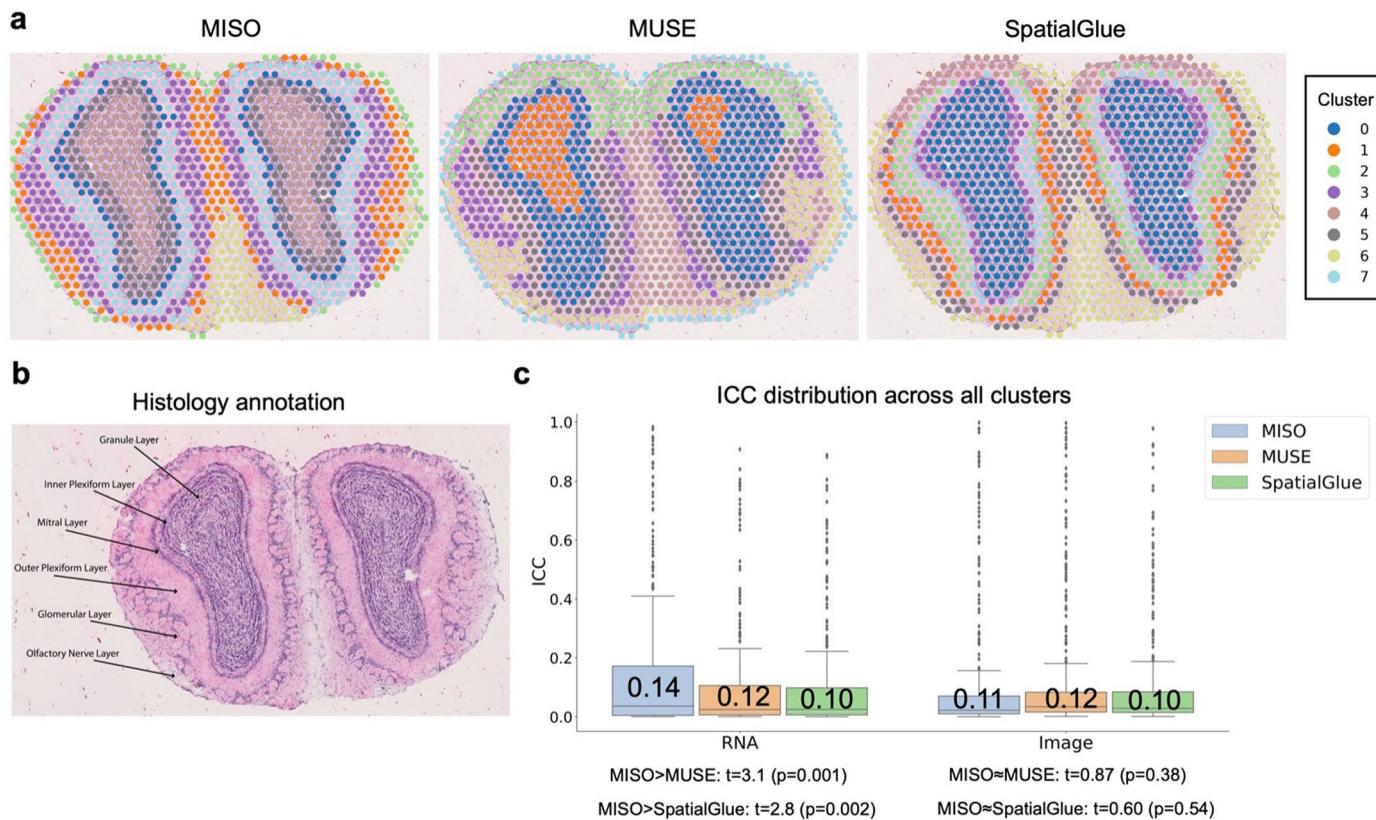
Extended Data Fig. 3 | Clustering results for a human breast cancer spatial transcriptomics dataset. **a**, Pathologist manual annotation of tissue section. DCIS: Ductal carcinoma in situ. **b**, Shown from left to right are clustering results from MISO, MUSE, and SpatialGlue, respectively. Two patches were selected to highlight that MISO's results agree better with histological patterns. **c**, RNA and image ICC distributions across all clusters and features for each method in the breast cancer data ($n = 750$ ICC values for each group). The mean ICC for each method and each modality is printed on the corresponding box plot. Test statistics and p-values were obtained using one-sided ($<$, $>$) or two-sided (\approx) t-tests. **d**, Spots plotted according to their RNA t-SNE coordinates and colored by the clustering results for each method. The MISO clustering results demonstrate

coherence with respect to gene expression patterns and the annotated histological regions. **e**, RNA t-SNE plot for the breast cancer Visium dataset with spots colored according to total UMI count. MISO was able to localize a sub-cluster in the annotated invasive carcinoma region with much lower total UMI counts compared to other sub-clusters in this region. **f**, SpatialGlue clustering results when increasing the weight given to histology in the loss function. SpatialGlue was not able to detect the fat region of the tissue section when making the weight given to histology 10 or 50 times greater than that given to gene expression. Box plots: center line, median; box limits, upper and lower quartiles; whiskers, 1.5x interquartile range; points, outliers.



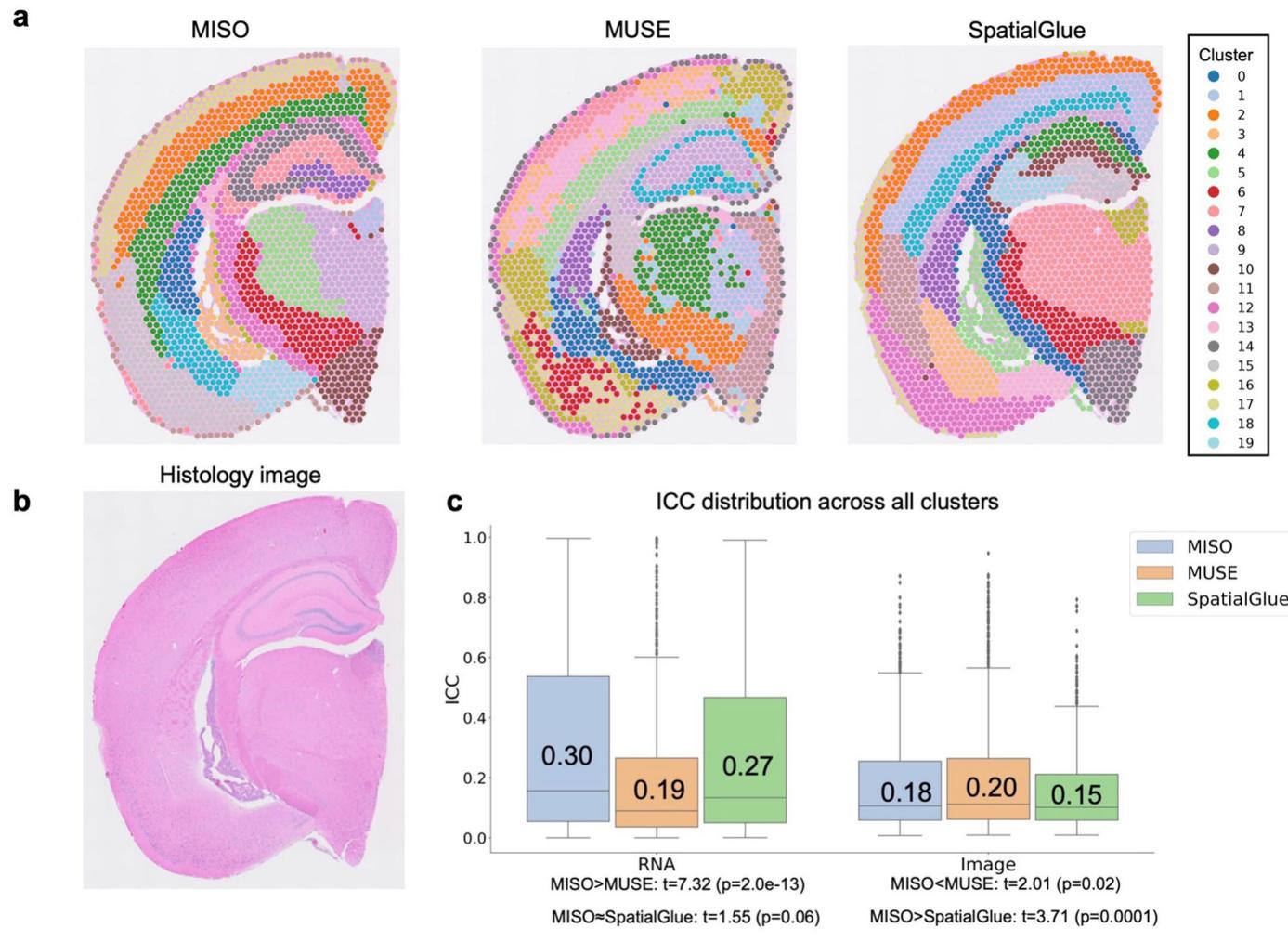
Extended Data Fig. 4 | Clustering results for a zebrafish melanoma spatial transcriptomics dataset. **a**, Blurriness artifact in the H&E-stained histology image. **b**, Clustering results from MISO, MUSE, and SpatialGlue. MISO did not include the image-specific features in clustering because of the low quality of the image, but the image features were still accounted for in the RNAimage interaction terms. Clusters in the MUSE results are driven by the blurriness

artifact. **c**, RNA and image ICC distributions across all clusters and features for each method ($n = 400$ ICC values for each group). The mean ICC for each method and each modality is printed on the corresponding box plot. Test statistics and p-values were obtained using one-sided ($<$, $>$) or two-sided (\approx) t-tests. Box plots: center line, median; box limits, upper and lower quartiles; whiskers, 1.5x interquartile range; points, outliers.



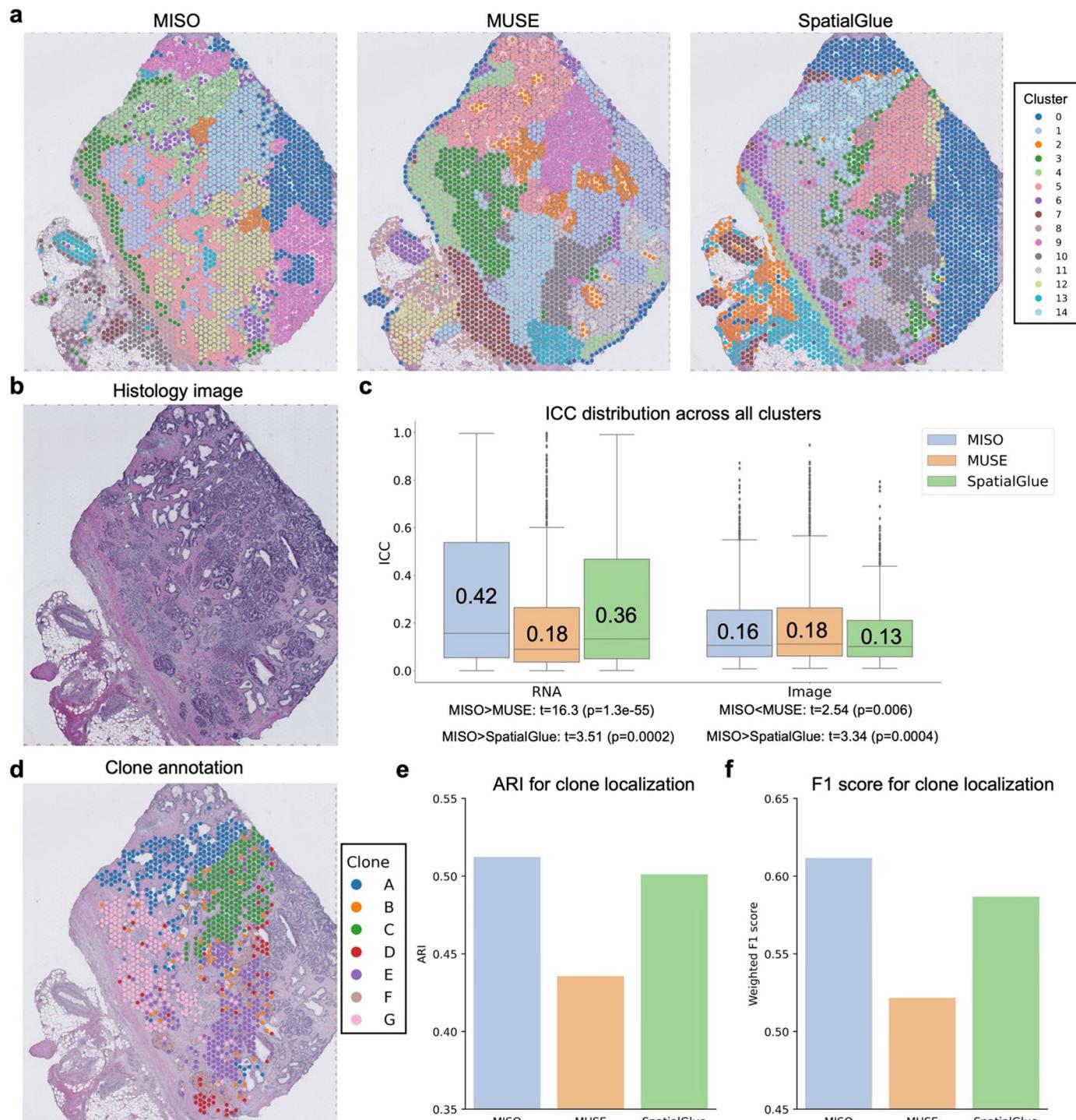
Extended Data Fig. 5 | Clustering results for a mouse olfactory bulb spatial transcriptomics dataset. **a**, Clustering results from MISO, MUSE, and SpatialGlue. **b**, H&E-stained histology image of analyzed tissue section with layer annotation. The MISO results align well with the annotation, assigning clusters to each of the annotated layers. MUSE was not able to accurately localize clusters to the annotated layers, and instead separated spots in the upper region of the

tissue section from those in the lower region. **c**, RNA and image ICC distributions across all clusters and features for each method ($n = 400$ ICC values for each group). The mean ICC for each method and each modality is printed on the corresponding box plot. Test statistics and p-values were obtained using one-sided ($<$, $>$) or two-sided ($=$) t-tests. Box plots: center line, median; box limits, upper and lower quartiles; whiskers, 1.5x interquartile range; points, outliers.



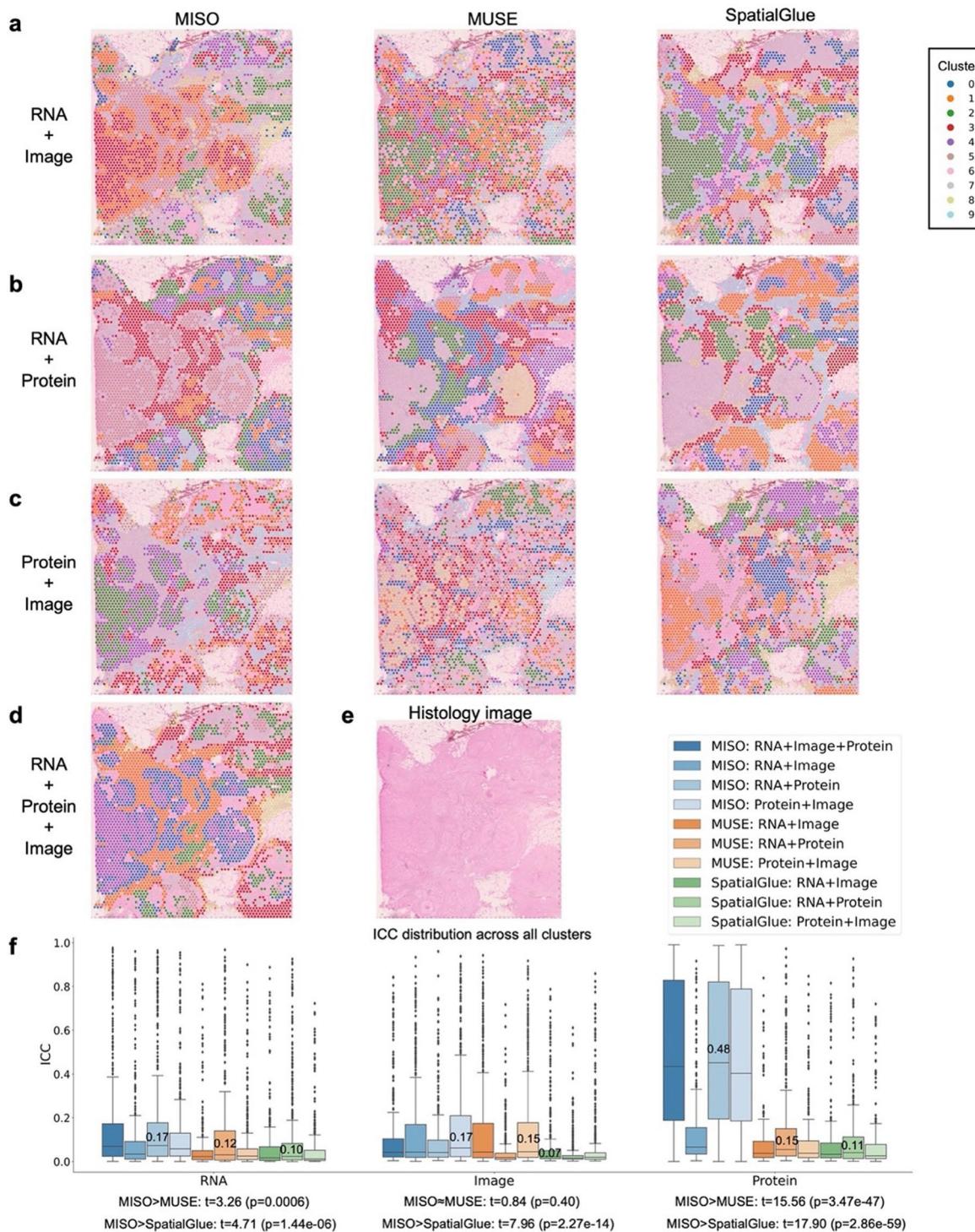
Extended Data Fig. 6 | Clustering results for a mouse coronal brain spatial transcriptomics dataset. **a**, Clustering results from MISO, MUSE, and SpatialGlue. **b**, H&E-stained histology image of analyzed tissue section. **c**, RNA and image ICC distributions across all clusters and features for each method ($n=1000$ ICC values for each group). The mean ICC for each method and each

modality is printed on the corresponding box plot. Test statistics and p-values were obtained using one-sided ($<$, $>$) or two-sided (\approx) t-tests. Box plots: center line, median; box limits, upper and lower quartiles; whiskers, 1.5x interquartile range; points, outliers.



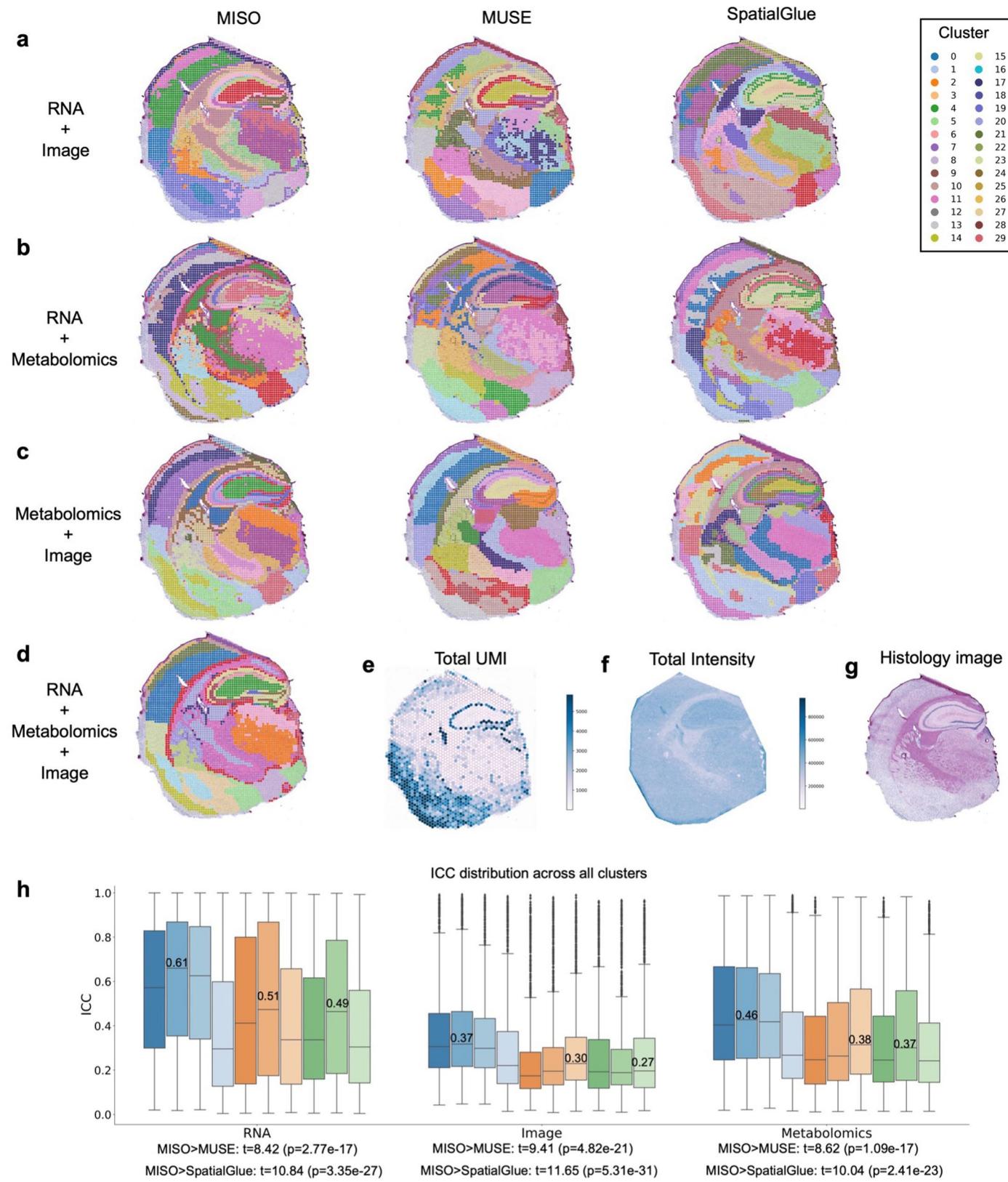
Extended Data Fig. 7 | Clustering results for a human prostate cancer spatial transcriptomics dataset. **a**, Clustering results from MISO, MUSE, and SpatialGlue. **b**, H&E-stained histology image of analyzed tissue section. **c**, RNA and image ICC distributions across all clusters and features for each method ($n = 750$ ICC values for each group). The mean ICC for each method and each modality is printed on the corresponding box plot. Test statistics and p-values were obtained using one-sided ($<$, $>$) or two-sided (\approx) t-tests. **d**, Clone annotation of cancer spots. **e**, ARI between the clone annotation and the clustering results

across all cancer spots for MISO (0.51), MUSE (0.44), and SpatialGlue (0.50). **f**, Weighted F1 score for localization of clusters to the annotated clones for MISO (0.61), MUSE (0.52), and SpatialGlue (0.59). To calculate F1 score for a given method, a cluster was assigned to a clone if more than half of the spots from that cluster overlapped with the clone annotation. F1 score was weighted by the number of spots belonging to each clone in the annotation. Box plots: center line, median; box limits, upper and lower quartiles; whiskers, 1.5x interquartile range; points, outliers.



Extended Data Fig. 8 | Clustering results for a human breast cancer spatial gene and protein expression dataset. Image features used as input for each method were extracted from an immunofluorescence image with 3 channels (DAPI, Vimentin, and PCNA) using a pre-trained InceptionV3 model. Patches corresponding to the omics spots were extracted from the immunofluorescence image, resized to 299×299 pixels, and normalized prior to extracting features for each patch using InceptionV3. **a**, Clustering results from MISO, MUSE, and SpatialGlue when taking RNA and image data as input. **b**, Clustering results from MISO, MUSE, and SpatialGlue when taking RNA and protein data as input. **c**, Clustering results from MISO, MUSE, and SpatialGlue when taking protein and image data as input. **d**, Clustering results from MISO when taking RNA,

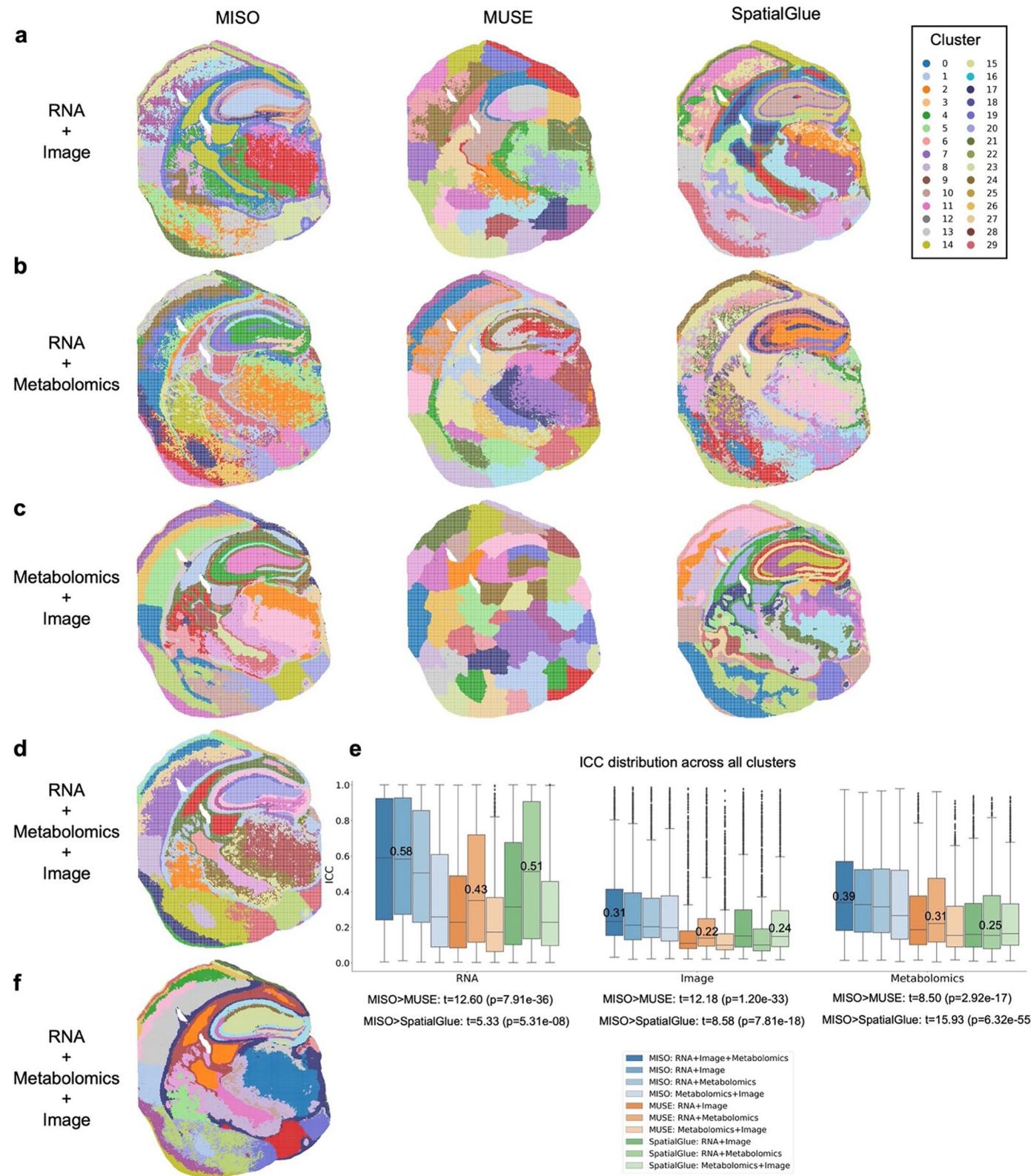
image, and protein data as input. **e**, Eosin-stained histology image of analyzed tissue section. **f**, RNA, image, and protein ICC distributions across all clusters and features for each method when taking each possible combination of modalities as input ($n = 500$ ICC values for each group). For each method, the mean ICC for each modality is printed on the corresponding box plot for its top-performing combination of modalities. Test statistics and p-values were obtained using one-sided ($<$) or two-sided (\approx) t-tests when comparing each method's top-performing results for a given modality. Box plots: center line, median; box limits, upper and lower quartiles; whiskers, 1.5x interquartile range; points, outliers.



Extended Data Fig. 9 | See next page for caption.

Extended Data Fig. 9 | Clustering results for a mouse brain spatial transcriptomics (10x Visium) and metabolomics (MALDI-MSI) dataset, which was generated following the protocol described in Vicari et al. [9]. To make the super-resolution spatial molecular data inferred by iStar more suitable for input to all methods, we merged superpixels obtained from iStar to create 4,687 pseudo-spots of size 128×128 pixels, containing paired gene expression and metabolite information. Because the RNA data is of low quality (e), the RNA-specific features extracted by MISO were not used to produce any of the results provided, but RNA was still accounted for in its interactions with metabolomics and image features. For all applicable results, metabolomics data were normalized by total intensity and log transformed. **a**, Clustering results from MISO, MUSE, and SpatialGlue when taking RNA and histology image data as input. **b**, Clustering results from MISO, MUSE, and SpatialGlue when taking RNA and metabolomics data as input. **c**, Clustering results from MISO,

MUSE, and SpatialGlue when taking metabolomics and histology image data as input. **d**, Clustering results from MISO when taking RNA, histology image, and metabolomics data as input. **e**, Total UMI counts across all spots in the dataset. The UMI counts are low because the tissue section was analyzed using MALDI-MSI prior to Visium. **f**, Total metabolite intensities across all spots in the dataset. **g**, H&E-stained histology image of analyzed tissue section. **h**, RNA, image, and metabolomics ICC distributions across all clusters and features for each method when taking each possible combination of modalities as input ($n = 1500$ ICC values for each group). For each method, the mean ICC for each modality is printed on the corresponding box plot for its top-performing combination of modalities. Test statistics and p-values were obtained using one-sided t-tests when comparing each method's top-performing results for a given modality. Box plots: center line, median; box limits, upper and lower quartiles; whiskers, 1.5x interquartile range; points, outliers.



Extended Data Fig. 10 | See next page for caption.

Extended Data Fig. 10 | Clustering results for large-scale mouse brain spatial transcriptomics (10x Visium) and metabolomics (MALDI-MSI) dataset described in Extended Data Fig. 9. Results were obtained using RNA, metabolomics, and image features as input, but RNA was only accounted for in the interaction terms due to its low quality. For all applicable results, other than the MUSE results obtained when taking metabolomics and image data as input, metabolomics data were normalized by total intensity and log transformed. MUSE was not able to evaluate the combination of metabolomics and image data when the metabolomics data were log normalized, so this step was not utilized to obtain the corresponding results. The dataset used to generate the results in (a-e) contains 74,851 pseudo-spots of size 32×32 pixels. The dataset used to generate the results in (f) contains 299,350 pseudo-spots of size 16×16 pixels. Due to memory requirements, MISO was the only method that could evaluate the dataset with pseudo-spots of size 16×16 pixels. **a**, Clustering results from MISO, MUSE, and SpatialGlue when taking RNA and histology image data as input. **b**, Clustering results from MISO, MUSE, and SpatialGlue when

taking RNA and metabolomics data as input. **c**, Clustering results from MISO, MUSE, and SpatialGlue when taking metabolomics and histology image data as input. **d**, Clustering results from MISO when taking RNA, histology image, and metabolomics data as input. **e**, RNA, image, and metabolomics ICC distributions across all clusters and features for each method when taking each possible combination of modalities as input ($n = 1500$ ICC values for each group). For each method, the mean ICC for each modality is printed on the corresponding box plot for its top-performing combination of modalities. Test statistics and p-values were obtained using one-sided t-tests when comparing each method's top-performing results for a given modality. **f**, Clustering results from MISO when taking RNA, histology image, and metabolomics data from the dataset with pseudo-spots of size 16×16 pixels as input. Test statistics and p-values were obtained using one-sided t-tests when comparing each method's top-performing results for a given modality. Box plots: center line, median; box limits, upper and lower quartiles; whiskers, 1.5x interquartile range; points, outliers.

Corresponding author(s): Mingyao Li, Kyle Coleman, Jian Hu

Last updated by author(s): Oct 8, 2024

Reporting Summary

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

Statistics

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

n/a Confirmed

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

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

Software and code

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

Data collection	No software was used for data collection.
Data analysis	MISO v1.0.0 (https://github.com/kpcoleman/miso), MUSE v1.0.0 (https://github.com/AltschulerWu-Lab/MUSE), and SpatialGlue v1.0.0 (https://github.com/JinmiaoChenLab/SpatialGlue) were used for spatial clustering. Scampy v1.9.0 was used for data preprocessing. Motivation for using neural networks for extraction of spectral clustering features came from SpectralNet v1.0 (https://github.com/KlugerLab/SpectralNet). Pre-trained vision transformer models are from HPT v1.0 (https://github.com/mahmoodlab/HPT).

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

Data

Policy information about [availability of data](#)

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

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

We analyzed the following datasets: (1) 10x Visium human bladder cancer spatial transcriptomics data (GEO GSE246011, sample BLCA-B1); (2) 10x Xenium human

gastric cancer spatial transcriptomics data will be made publicly available after January 1st, 2025. Until then, data requests will be reviewed by Dr. Tae Hyun Hwang and Dr. Jeong Hwan Park, and reasonable requests may be accommodated upon approval. IRB No. 30-2023-1; (3) 10x Visium HD human colorectal cancer data (<https://www.10xgenomics.com/datasets/visium-hd-cytassist-gene-expression-libraries-of-human-crc>); (4) Spatial ATAC-RNA-seq mouse embryonic day 13 (E13) data reported in Zhang et al.5 (<https://cells.ucsc.edu/?ds=brain-spatial-omics>); (5) Spatial transcriptomics and metabolomics mouse coronal brain data (<https://upenn.box.com/s/3o8dq5j4x29ic6zo7iugdo83scnv4qis>). This data will be made publicly available once the paper is accepted for publication. (6) 10x Visium human tonsil gene and protein expression data (<https://www.10xgenomics.com/resources/datasets/gene-protein-expression-library-of-human-tonsil-cytassist-ffpe-2-standard>); (7) 10x Visium mouse anterior brain spatial transcriptomics data (<https://www.10xgenomics.com/resources/datasets/mouse-brain-serial-section-1-sagittal-anterior-1-standard-1-1-0>); (8) 10x Visium human breast cancer spatial transcriptomics data (<https://www.10xgenomics.com/resources/datasets/human-breast-cancer-visium-fresh-frozen-whole-transcriptome-1-standard>); (9) 10x Visium zebrafish melanoma spatial transcriptomics data reported in Hunter et al.45 (GSE159709); (10) 10x Visium mouse olfactory bulb spatial transcriptomics data (<https://www.10xgenomics.com/resources/datasets/adult-mouse-olfactory-bulb-1-standard-1>); (11) 10x Visium mouse coronal brain spatial transcriptomics data (<https://www.10xgenomics.com/resources/datasets/mouse-brain-coronal-section-2-ffpe-2-standard>); (12) 10x Visium human prostate cancer spatial transcriptomics data reported in Erickson et al.48 (<https://data.mendeley.com/datasets/swv96g68dv/1>); (13) Spatial CUT&Tag-RNA-seq (H3K27AC) mouse coronal brain data reported in Zhang et al.5 (<https://cells.ucsc.edu/?ds=brain-spatial-omics>); (14) Spatial CUT&Tag-RNA-seq (H3K27ME3) mouse coronal brain data reported in Zhang et al.5 (<https://cells.ucsc.edu/?ds=brain-spatial-omics>); (15) 10x Visium human breast cancer gene and protein expression data (<https://www.10xgenomics.com/resources/datasets/gene-and-protein-expression-library-of-human-breast-cancer-cytassist-ffpe-2-standard>); (16) Spatial CITE-seq mouse colon data reported in Liu et al.56 (GSE213264). Details of the datasets analyzed in this paper are described in Supplementary Table 1.

Human research participants

Policy information about [studies involving human research participants and Sex and Gender in Research](#).

Reporting on sex and gender

n/a

Population characteristics

n/a

Recruitment

n/a

Ethics oversight

n/a

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

Field-specific reporting

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

Life sciences

Behavioural & social sciences

Ecological, evolutionary & environmental sciences

For a reference copy of the document with all sections, see nature.com/documents/nr-reporting-summary-flat.pdf

Life sciences study design

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

Sample size

For each of the spatial omics datasets we analyzed, there was a predefined number of spots for which the omics data were measured. We utilized all spots when analyzing a given tissue section, and did not use statistical methods to determine sample sizes.

Data exclusions

No data were excluded from the analyses. We did not perform spot/cell filtering, so all capture locations were included in the analysis.

Replication

Each experiment was performed independently without replication. For the analyses, we did not replicate our results, as we instead compared them with known biological properties and evaluated performance through quantitative metrics.

Randomization

This is not relevant to our study as all tissue sections were analyzed separately.

Blinding

Blinding was not relevant to the data collection since each experiment involved the generation of data from only one tissue section/state. For the analyses involving comparison with tissue annotation, the investigators were blinded to the annotation and group allocation when running the methods.

Reporting for specific materials, systems and methods

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

Materials & experimental systems

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

Methods

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