Genome Medicine

METHODOLOGY

Open Access

MuCST: restoring and integrating heterogeneous morphology images and spatial transcriptomics data with contrastive learning

Yu Wang[1,2†], Zaiyi Liu[3,4†] and Xiaoke Ma[1,2*†]

Abstract

Spatially resolved transcriptomics (SRT) simultaneously measure spatial location, histology images, and transcriptional profiles of cells or regions in undissociated tissues. Integrative analysis of multi-modal SRT data holds immense potential for understanding biological mechanisms. Here, we present a flexible multi-modal contrastive learning for the integration of SRT data (MuCST), which joins denoising, heterogeneity elimination, and compatible feature learning. MuCST accurately identifies spatial domains and is applicable to diverse datasets platforms. Overall, MuCST provides an alternative for integrative analysis of multi-modal SRT data (https://github.com/xkmaxidian/MuCST).

Keywords Spatial transcriptomics, Spatial domain, Contrastive learning, Multi-modality

Background

Cells are the fundamental units of tissues in multicellular organisms, which are physically clustered together with various states. Recognizing the structure and spatial location of cells is vital for understanding the emergent properties and pathology of tissues [1]. The traditional microscopy technology identifies and characterizes cell groups (also called cell types or sub-populations) through similarities of morphology, including shapes, sizes and physical appearance of cells [2]. However, morphology alone is insufficient to fully characterize structure of cells because of the unstable states of cells [3]. Fortunately, the single-cell RNA sequencing (scRNA-seq) technology enables generation of whole genome-wide expression at cell level, providing complementary information to characterize structure of cells at molecular level [4, 5].

However, the dissociation step of scRNA-seq erases spatial context of cells from their original tissues that is crucial for understanding cellular functions and organizations [6]. Spatial Transcriptomics (ST) [7] simultaneously allows morphological and transcriptional profiling of cells in the same tissue regions, which also retains spatial context of cells [8]. Typically, current ST technologies can be broadly divided into two categories, i.e., imaging- and next-generation sequencing (NGS)-based methods, where the former one uses probes to localize mRNA transcripts, including FISH and MERFISH [9], seqFISH [10], and STARmap [11], which are criticized for their limited capacity to detect RNA transcripts. To overcome this limitation, the latter one utilizes spatial bar-code and next-generation sequencing technologies to retain transcription and spatial information, including Legacy

†Yu Wang, Zaiyi Liu, and Xiaoke Ma contributed equally to this work.

*Correspondence:
Xiaoke Ma
xkma@xidian.edu.cn
[1] School of Computer Science and Technology, Xidian University, No.2 South Taibai Road, Xi'an 710071, Shaanxi, China
[2] Key Laboratory of Smart Human-Computer Interaction and Wearable Technology of Shaanxi Province, Xidian University, No.2 South Taibai Road, Xi'an 710071, Shaanxi, China
[3] Department of Radiology, Guangdong Provincial People's Hospital (Guangdong Academy of Medical Sciences), Southern Medical University, 106 Zhongshan Er Road, Guangzhou 510080, Guangdong, China
[4] Guangdong Provincial Key Laboratory of Artificial Intelligence in Medical Image Analysis and Application, 106 Zhongshan Er Road, Guangzhou 510080, Guangdong, China

Spatial Transcriptomics [12], 10× Visium [13], Slide-seq [14], and Stereo-seq [15]. The accumulated spatially resolved transcriptomics (SRT) data provide an opportunity to investigate functions and cellular structure of tissues by exploiting interesting patterns and features that cannot be discerned from scRNA-seq data [16].

Therefore, integrative analysis of spatially resolved data is a prominent task since it sheds light on revealing mechanisms of tissues. On the basis of principles of algorithms, current approaches are roughly divided into two categories, i.e., transcript- and image-based methods, where the former ones are devoted to integrate transcriptomics and spatial information, and the latter ones fuse morphology, transcript, and spatial information. Specifically, transcript-based approaches concentrate on learning cell features by balancing transcriptomics and spatial coordinates of cells with various strategies. For example, algorithms for scRNA-seq data, such as SCANPY [17], DRjCC [18], and jSRC [19] are directly applied to ST data by ignoring spatial information of cells, resulting in the undesirable performance. To overcome this limitation, many algorithms are developed by incorporating spatial coordinates into feature learning with various manners. For instance, Gitto [20] employs the hidden Markov random field model, whereas BayesSpace [21] adopts the Bayesian statistical method. STAGATE [22], GraphST [23], and Spatial-MGCN [24] utilize graph neural networks to learn features of cells, while SEDR [25], DRSC [26], and SpatialPCA [27] adopt subspace learning. constructs neighbor graphs with transcriptional feature and spatial information, and employs graph neural network to learn features of cells. And, SpiceMix [28] and CellCharter [29] fuse multiple adjacent slices of tissues to jointly model and characterize structure of spatial domains.

Nevertheless, these algorithms ignore morphological information in histological images that usually provide vital supplementary information for transcriptomics. For example, transcriptional variations within distinct spatial domains are often mirrored in morphology [30]. However, integrating morphology and SRT data is highly non-trivial because of the extra-ordinary heterogeneity of multi-modal data. Current algorithms leverage morphological information to complement spatial transcriptomics with different strategies. For example, stLearn [31] calculates morphological distance between cells to smooth and augment expression of cells. SpaGCN [30] and DeepST [32] integrate spatial and morphological information into cell networks, and then learn features with graph convolution network (GCN). stMVC [33] and stMGATF [34] transform spatial domain identification in SRT data into the multi-view clustering, and then adopt the semi-supervised strategy for down-stream analysis. MUSE [35] integrates information of morphology and transcription to learn joint representation with deep learning, while ConGI [36] and conST [37] perform integrative analysis with contrastive learning.

Even though few attempts are devoted to the integration of histological images and spatial transcriptomics, there are still many unsolved and critical problems. First, the extra experimental steps required to preserve the locations of cells, which brings noise into spatially resolved data [38], posing a great challenge for designing effective integrative algorithms. Current algorithms remove noise of SRT data by in the pre-processing procedure, which separates noise of data and feature learning, failing to fully characterize and model noise of data. Second, spatially resolved data are highly heterogeneous because of spatial and transcriptional information, and current approaches directly learn the low-dimensional features of cells, which neglects the heterogeneity of spatial locations and transcriptiomics. How to avoid heterogeneity of SRT data is still a great challenge for integrative analysis. Third, spatially resolved data consist of multiple modalities that are complementary to each other, and current algorithms fail to deeply fuse all modalities, thereby reducing the quality of cell features. How to learn compatible and discriminative features of cells is also vital for the integration of morphology and SRT data.

To address the aforementioned issues, we present a novel and flexible *Mu*lti-modal *C*ontrastive learning for the integration of *S*patially resolved *T*ranscriptomics (MuCST), including morphology, spatial coordinates and transcription profiles of cells. As shown in Fig. 1, MuCST consists of two major components, i.e., multi-modal feature learning, and consistent feature learning

(See figure on next page.)

Fig. 1 Overview of MuCST for integrating spatially resolved data with histology images, spatial coordinates and transcriptional information. **A** Spatially resolved multi-modal data include histology images, spatial coordinates of cells, and transcriptional profiles of cells. **B** MuCST learns the morphological feature of cells from histology images with available SimCLR [39], and learns the transcriptional features of cells by integrating spatial and transcriptional information, where an attributed cell network model is proposed. Graph convolution network (GCN) is employed to learn compatible transcriptional features with constrastive learning. **C** MuCST reconstructs histology images and expression profiles of cells with consistent cell features that are obtained with multi-modal contrast learning by fusing the morphological and transcriptional features of cells. **D** Down-stream analysis of fused cell features from the reconstructed image and expression profiles with principle component analysis (PCA) include spatial domain identification, tumor micro-environment, spatial transcriptomics data denoising, and so on

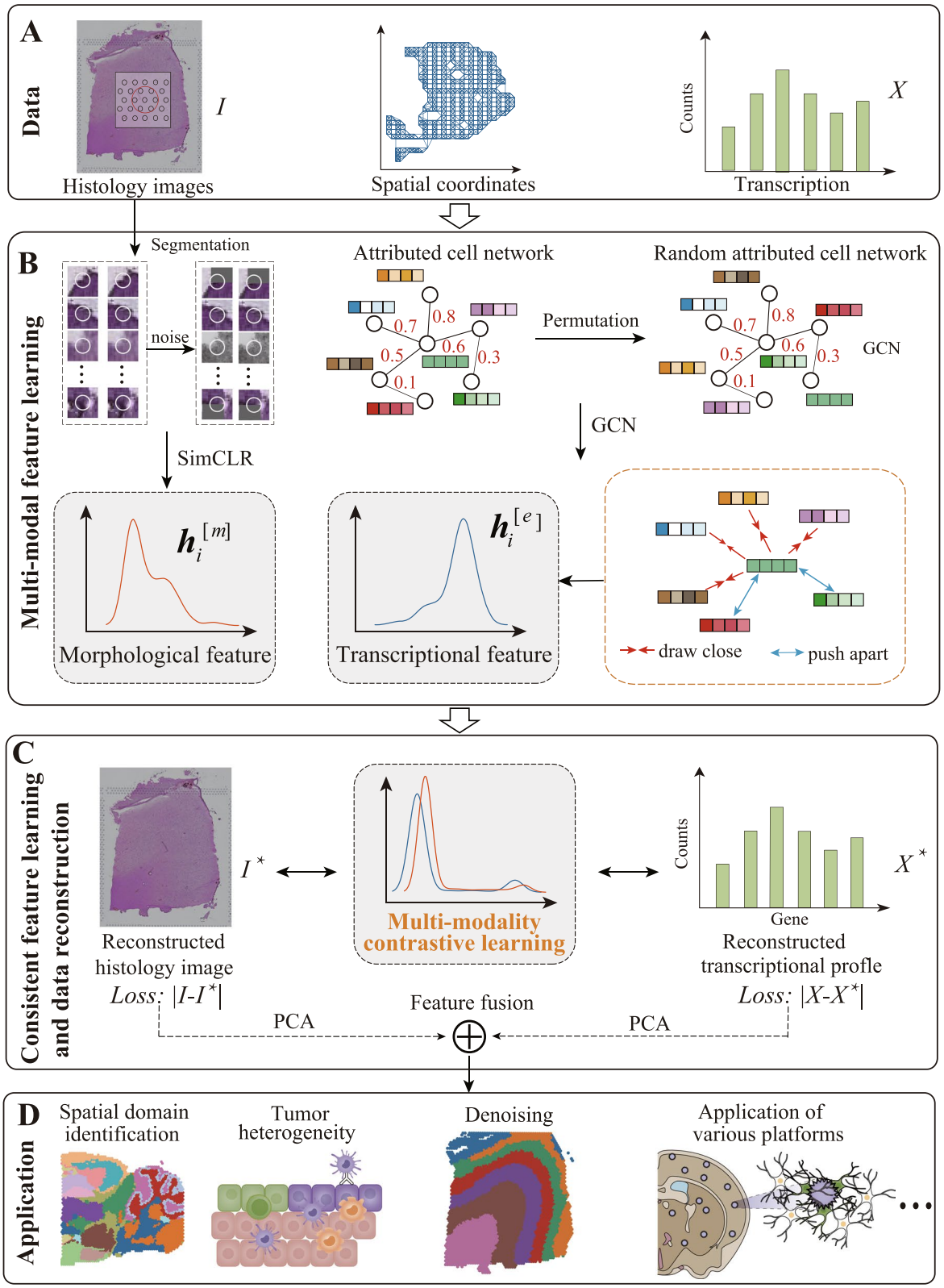

Fig. 1 (See legend on previous page.)

and data restoration, where the former procedure independently obtains features of cells from morphology and ST data, the morphological features of cell are learned with pre-trained image process techniques, and transcriptional features of cells are obtained with GNN. The latter procedure removes heterogeneity of morphological and transcriptional features of cells by learning consistent features of cells with multi-modal contrastive learning, where noise of SRT data is also removed with data restoration, thereby enhancing discriminative and compatibility of consistent features of cells. The experimental results on the simulated and SRT data demonstrate that MuCST not only precisely removes heterogeneity of morphology and SRT data, but also improves the dissection and interpretation of spatial patterns.

Methods

Data pre-processing and network construction

Eleven simulated and thirteen biological datasets (Additional file 1: Table S1) are employed to fully validate performance of MuCST. For all these datasets, spots (cells) outside of the main tissue regions are removed. Histology images are split into patches for each spot according to spatial coordinates, and morphological features of patches are learned with ResNet-50 [40] (denoted by $\mathbf{m}_i$). K-nearest neighborhood (KNN) is utilized to construct the cell network $G = (V, E)$ with euclidean distance of spatial coordinates of cells (k=6 according to Ref. [41]). Then, weights on edges are calculated with similarity of morphological features of corresponding cells, i.e., weight a_{ij} for the ith and jth cell is the cosine similarity of $\mathbf{m}_i$ and $\mathbf{m}_j$. The adjacent matrix of G is denoted by $A = (a_{ij})_{n\times n}$ with element a_{ij} as the weight on edge (i, j), where n is the number of cells. The raw expression profile of n cells $\widetilde{X} = [\widetilde{\mathbf{x}}_1, \ldots, \widetilde{\mathbf{x}}_n]$ is normalized, log-transformed and scaled according to library size with SCANPY [17]. By following Seurat [42], genes expressed in less than 10 cells are filtered. To overcome the low expression profile of cells, each cell is augmented with its neighbors. In details, given expression profile of the ith cell $\widetilde{\mathbf{x}}_i$, and its neighbors in G ($N_i(G) = \{j|(i, j) \in E\}$), the augmentation is performed by integrating morphological and expression of neighbors as

$$\mathbf{x}_i = \widetilde{\mathbf{x}}_i + \frac{\sum_{j\in N_i(G)} \widetilde{\mathbf{x}}_i a_{ij}}{|N_i(G)|}. \quad (1)$$

The attributed cell network $\mathcal{G} = (G, X)$ is constructed by setting G as topological structure of cells, and the augmented expression profile of cells X as attributes of vertices. The random attributed network $\widehat{\mathcal{G}} = (G, \widehat{X})$ of $\mathcal{G}$ is generated by preserving the topological structure G and randomly permutating attributes of cells.

Mathematical model for MuCST

As shown in Fig. 1, MuCST consists of two major procedures, i.e., multi-modal feature learning, and consistent feature learning and data restoration, where the former one procedure aims to obtain morphological and transcriptional features of cells, and the latter one focuses on learning consistent features of spots by fusing transcriptional and morphological features produced by the former procedure, and reconstructing the original data with the learned consistent features of spots.

On the multi-modal feature learning issue, MuCST first learns morphological feature of cells by splitting histology image I into patches for each spot, where these patches are randomly noised with the pre-trained ResNet [40], followed by multilayer perception (MLP) [43] (how to segment histology images is presented in Additional file 1: Section 1.1). To enhance quality of morphological features, SimCLR [39] is employed to discriminate the original and noised patches Then, MuCST employs graph neural network (GNN) to learn transcriptional features of cells by manipulating structure of attributed cell network $\mathcal{G}$. Specifically, MuCST utilizes graph convolution network (GCN) (ι layers) [44] to learn cell transcriptional with structure as

$$Z^{(l)} = \begin{cases} X, & \text{if } l = 0 \\ \sigma\left(\tilde{A}Z^{(l-1)}O^{(l-1)} + B^{(l-1)}\right), & \text{if } l \geq 1, \end{cases} \quad (2)$$

where $\tilde{A} = D^{-\frac{1}{2}}AD^{-\frac{1}{2}}$ is the normalized adjacent matrix of $\mathcal{G}$ (D is diagonal matrix with element as $d_{ii} = \sum_{j=1} a_{ij}$), O and B denote the trainable weight and bias matrix respectively, $\sigma(\cdot)$ is a non-linear activation function, $Z^{(l)}$ is the latent transcriptional feature at the lth layer, and $\mathbf{z}_i$ is the ith row of Z, i.e., low-dimensional representation of the ith spot (how to select the number and dimensions of layers of GCN can be referred in Additional file 1: Section 1.2). Analogously, MuCST learns the random transcriptional features, denoted by $\widehat{Z}$, by manipulating the random attributed network $\widehat{\mathcal{G}}$.

On the consistent feature learning and data reconstruction issue, MuCST aims to learn consistent features from morphological and transcriptional features of spots. To improve discriminative of features, we expect features of cells belonging to the same spatial domains are close, whereas these from different domains apart from each other. Specifically, for each cell $\mathbf{z}_i$, MuCST enforces it to be close to the center of its neighbors in the attributed cell network $\mathcal{G}$, i.e., $\|\mathbf{z}_i - \mathbf{z}_i^{[c]}\|^2$. According to Ref. [45], the loss of contrastive learning is formulated as

$$\mathcal{L}^{[g]} = -\frac{1}{2n}\left(\sum_{i=1}^{n}\left(\mathbb{E}_{Z,W}\left[log\phi\left(Z, Z^{[c]}\right)\right] + \mathbb{E}_{\hat{Z},W}\left[log\left(1-\phi\left(\hat{Z}, Z^{[c]}\right)\right)\right]\right)\right) \tag{3}$$

where $\phi(\cdot)$ and $\mathbb{E}(X)$ are the bilinear function and mathematical expectation of X, respectively. Equation (3) minimizes distance between features of each spot and its center and maximizes distance between random features of each spot and its center.

To obtain consistent features of cells from the morphological and transcriptional features, we employ a two-layer neural network Υ to project them into a shared subspace to reduce the heterogeneity of multimodal features as

$$\begin{cases} \mathbf{h}_i^{[m]} = \Upsilon(\mathbf{m}_i), \\ \mathbf{h}_i^{[e]} = \Upsilon(\mathbf{z}_i). \end{cases} \tag{4}$$

where the relations between morphology and spatial expression are implicitly exploited. To comprehensively fuse these two types of cell features, we expect the distributions of morphological and transcriptional features are consistent. Specifically, MuCST enforces the morphological features $\mathbf{h}_i^{[m]}$ is close to the average of transcriptional features of its neighbors (denoted by $\mathbf{h}_i^{[ce]}$), where can be fulfilled with contrastive learning [46] as

$$\mathcal{L}^{[c]} = -\frac{1}{n}\sum_{i=1}^{n} log\frac{exp\left(\left\|\mathbf{h}_i^{[ce]} - \mathbf{h}_i^{[m]}\right\|^2\right)}{\sum_{j=1}^{n} exp\left(\left\|\mathbf{h}_i^{[ce]} - \mathbf{h}_j^{[m]}\right\|^2\right)}. \tag{5}$$

After obtaining consistent features of cells, MuCST also expects them to preserve structure of morphology and transcriptional profiles. To reach a tradeoff between consistence and specificity, we adopt the restoration strategy to ensure the learned consistent features of spots can also reconstruct the original morphology and transcriptional profiles. In details, we employs GCN to reconstruct expression profile of cells by using the $H^{[e]}$ with structure as

$$H^{(t)} = \begin{cases} H^{[e]}, & \text{if } t = 0 \\ \sigma\left(\tilde{A}H^{(t-1)}O^{(t-1)} + B^{(t-1)}\right), & \text{if } t \geq 1, \end{cases} \tag{6}$$

where $H^{(t)}$ denotes the reconstructed expression profiles at the t-th layer of GCN. Therefore, the loss function for expression reconstruction is defined as

$$\mathcal{L}^{[e]} = \left\|X - H^{(\tau)}\right\|^2, \tag{7}$$

where parameter τ is the number of layers for decoder. Then, histology image is reconstructed $I^* = H^{[m]}(H^{[m]})'$, where $(H^{[m]})'$ is the transpose of $H^{[m]}$. MuCST minimizes the approximation, i.e.,

$$\mathcal{L}^{[m]} = \left\|I - I^*\right\|^2. \tag{8}$$

By combining Eqs. (3), (5), (8), and (7), the overall objective function of MuCST is formulated as

$$\mathcal{L} = \mathcal{L}^{[e]} + \lambda_1\mathcal{L}^{[m]} + \lambda_2\mathcal{L}^{[g]} + \lambda_3\mathcal{L}^{[c]}, \tag{9}$$

where parameter λ_1, λ_2, and λ_3 control the relative importance of morphological information reconstruction, transcriptional feature consistence, and multi-modal fusion, respectively. And, in case when the histology image is absent, we set $\lambda_1{=}\lambda_3{=}0$, e.t. the objective function of MuCST is reformulated as

$$\mathcal{L} = \mathcal{L}^{[e]} + \lambda_2\mathcal{L}^{[g]}. \tag{10}$$

The optimization, parameter selection, and termination of MuCST are presented (Additional file 1: Section 1.3, 1.4 and 1.5).

After learning consistent features and restoring data, MuCST obtains morphological and transcriptional feature of cells, denoted by $\mathbf{f}_i^{[e]}$ and $\mathbf{f}_i^{[m]}$, with PCA (principle component analysis) from the reconstructed data, and then combine them via a linear function as

$$\mathbf{f}_i = \mathbf{f}_i^{[e]} + \lambda_1\mathbf{f}_i^{[m]} \tag{11}$$

where parameter λ_1 controls importance of morphological features of cells. Feature $F = [\mathbf{f}_1, \ldots, \mathbf{f}_n]$ is utilized for the downstream analysis. For example, MuCST identifies spatial domains by manipulating F with Mclust [47] if the number of domains is known, PhenoGraph [48] otherwise.

Identification and functional analysis of differentially expressed genes

MuCST performs differential expression analysis of genes for each spatial domain by using Wilcoxon rank-sum test implemented in SCANPY package [17]. Genes expressed in more than 80% of cells/spots in each domain, with a fold change ≥ 1 and an adjusted FDR ≤ 0.05, are selected as differentially expressed genes (DEGs). The filtered DEGs serve as input for gene ontology enrichment analysis, which is conducted with clusterProfiler [49]. Enriched functional terms with −log10(adjusted P-value) are plotted.

Clustering criteria

When manual annotation of spatial domains is absent, two extensively adopted clustering criteria, Silhouette Coefficient (SC) and Davies-Bouldin (DB) scores, are selected

to validate the performance of clustering in terms of computation. Specifically, SC takes into account compactness within and separation across clusters as $(b-a)/max(a,b)$, where a is the mean intra-cluster distance, and b is the mean nearest-cluster distance. It ranges between −1 and 1, where a higher score refers to more coherent clusters. SC=0 means that the sample is on or close to the boundary of neighboring clusters, whereas negative values denote potentially wrong clusters. DB score is the average ratio of within-cluster distances to between-cluster distances, favoring farther apart and less dispersed clusters with low values.

Overlap ratio

If the ground truth spatial domains are unknown, we employ ratio of overlap between spatial domains identified by various algorithms and bio-marker genes as a metric to evaluate performance of algorithms. Specifically, we use the minimum expression value of the identified differential genes within the identified spatial domain as a threshold, then calculate the R_{gene} for all spots with expression values exceeding this threshold and measure its overlap ratio with the $R_{spatial}$ of the identified spatial domain. The overlap ratio is formulated as

$$OR(R_{gene}, R_{spatial}) = \frac{|R_{gene} \cap R_{spatial}|}{|R_{gene} \cup R_{spatial}|}. \tag{12}$$

Benchmarking

To comprehensively evaluate performance of MuCST, we conduct extensive experiments to demonstrate its superiority over existing state-of-the-art methods, including the non-spatial method SCANPY [17] and spatial methods Giotto [20], stLearn [31], SEDR [25], BayesSpace [21], SpaGCN [30], STAGATE [22], SpatialPCA [27], DeepST [32], Spatial-MGCN [24], MUSE [35], ConGI [36], conST [37], stMVC [33], and stMGATF [34]. All these algorithms are executed to achieve the best performance for a fair comparison. When the ground truth spatial domains are known, performance of algorithms is measured with adjusted rank index (ARI) [50] as

$$ARI(P^*, P) = \frac{\sum_{ij}\binom{n_{ij}}{2} - \left[\sum_i \binom{n_i}{2} + \sum_j \binom{n_j}{2}\right] / \binom{n}{2}}{\frac{1}{2}\left[\sum_i \binom{n_i}{2} + \sum_j \binom{n_j}{2}\right] - \left[\sum_i \binom{n_i}{2} + \sum_j \binom{n_j}{2}\right] / \binom{n}{2}}, \tag{13}$$

where n is the number of cells, n_{ij} is the number of cells of class label $C^* \in P^*$ assigned to cluster C_i in partition P, and n_i / n_j is the number of cells in cluster C_i / C_j of partition P. Additional measurements, such as normalized mutual information [51], and $F1$-score are also adopted.

Data cohorts

Human dorsolateral prefrontal cortex dataset

The human dorsolateral prefrontal cortex (DLPFC) dataset [52] was sequenced using 10× Visium, which includes 12 slices with each containing 3460∼4789 spots and 33,538 genes. Spots of each slice of DLPFC were manually annotated into seven different layers based on marker genes, i.e., from Layer 1 to Layer 6, and white matter (WM).

Mouse brain datasets

Several mouse brain datasets sequenced by using different technologies are included in this study. Specifically, the normal mouse brain posterior slice, normal mouse brain coronal slices, and transgenic mouse (TgCRND8) brain slice were sequenced using the 10× Visium and are publicly available in the 10× Genomics Data Repository (https://www.10xgenomics.com/resources/datasets) [53]. The normal posterior slice contains 3353 spots and 31,053 genes, the normal coronal slice contains 2702 spots and 32,285 genes, and the TgCRND8 slice includes 3063 spots and 19,465 genes, respectively. Furthermore, to evaluate performance of MuCST on imaging-based ST platform, we also include the mouse brain cortex dataset sequenced by using osmFISH [54], which contains 5328 cells and 33 genes.

Human intestine dataset

The human intestine dataset [55] were collected from the colon of a male patient aged 66 years sequenced by using 10× Visium. Slice labeled as "A1" in the original publication is involved in this study, which contains 2807 spots and 33,538 genes.

Human cancer datasets

To benchmark MuCST on cancer-related datasets, we employ several human cancer related datasets. The human breast cancer slice [56] and human invasive ductal carcinoma slice were sequenced using 10× Visium [53], where the breast cancer slice contains 3798 spots and 36,601 genes, and invasive ductal carcinoma contains 4727 spots and 36,601 genes. The human pancreatic ductal adenocarcinoma (PDAC) slice [57], human prostate cancer slice [58], and human HER2 breast tumor slice [59] were sequenced using Legacy ST [12], where the PDAC slice contains 428 spots and 19,738 genes, the prostate cancer slice contains 501 spots and 17,335 genes, the HER2 breast tumor slice contains 603 spots and 14,907 genes.

Mouse visual cortex dataset

The mouse visual cortex dataset [11] sequenced by using STARmap is also selected, which contains 817 cells and 1020 genes. By following stMVC [33], we employ ClusterMap [60] to annotate cells into seven distinct layers from the raw fluorescence data based on watershed segmentation.

Mouse olfactory bulb and hippocampus datasets

To evaluate performance of MuCST on high-resolution spatial transcriptomics data, we additionally include the mouse olfactory bulb dataset [61] sequenced using Stereo-seq, which contains 19,109 spots and 14,376 genes. And, the mouse hippocampus dataset [14] sequenced using Slide-seq V2 is also selected, which contains 41,786 spots and 23,264 genes.

Results

Overview of MuCST

To facilitate the understanding of this study, the rationale and procedures of MuCST are briefly presented in this section (technical details can be referred to section of methods). For clarity, we utilize "cell" to denote the basic measurement units for imaging-based ST technologies, and "spot" to denote the basic measurement units for barcode-based ST technologies.

Spatially resolved transcriptomics data comprehensively cover histology images, spatial coordinates and transcription of spots (Fig. 1A), posing a great challenge for integrative analysis of them because of extra-ordinary heterogeneity of multi-modal data. Available algorithms directly fuse various types of features of spots, failing to appropriately address heterogeneity and intrinsic structure of data, resulting in the undesirable performance. To address these issues, we propose a novel and flexible algorithm for the integration of histological images and spatial transcriptomics with contrastive learning, which consists of three components, i.e., multi-modal feature learning, consistent feature learning and data restoration, and downstream analysis (Fig. 1). The underlying assumption is that spatial resolved data characterize tissues from different perspectives and levels, and integrative analysis of these data with a refined ordering according to their roles is promising for analyzing heterogeneous multi-modal data. Specifically, spatial and expression profiles of spots characterize tissues from micro-level, whereas histological images depict tissues from macro-level. Thus, MuCST integrates micro- and macro-level information to characterize and model intrinsic structure of patterns.

On the multi-modal feature learning issue, MuCST independently learns the morphological and transcriptional features with different strategies (Fig. 1B). Specifically, MuCST splits the morphological image I into patches for each spot. And, the pre-trained deep neural network model ResNet [40], followed by multi-layer perception (MLP) [43], is adopted to learn the morphological features of spots. To enhance quality of morphological features of spots, SimCLR [39] is employed to discriminate the original and noised patches (Fig. 1B). And, MuCST learns the transcriptional features of spots with graph convolution network (GCN), where the indirect topological structure is exploited to fully characterize spatial and transcriptional information. Specifically, MuCST first constructs an attributed network by integrating histological images, spatial coordinates and expression profiles X of spots ("Methods" section). Then, MuCST learns the transcriptional features of spots by discriminating the attributed network and random one generated with permutation of gene expression profiles, where neighbors of spots are also in close proximity to each other in the transcriptional feature space.

On the consistent feature learning and data reconstruction issue, MuCST considers the morphological and transcriptional features of spots as complementary modalities, which fuses these heterogeneous features of spots with multi-modal contrast learning (Fig. 1C). In detail, the morphological and transcriptional features of spots are projected into a shared subspace, where MuCST aligns the distributions of two types of features such that they are subjected to the identical distribution. In this case, the heterogeneity of multi-modal features is removed at the feature level, facilitating the down-stream analysis. Then, MuCST restores histology image I^* and transcriptional profiles of spots X^* by minimizing the reconstruction errors, i.e., $\|I - I^*\|$ and $\|X - X^*\|$, thereby preserving capability of consistent features to represent the original morphology and transcriptomics, balancing consistence and specificity of features of spots. Finally, MuCST employs principle component analysis (PCA) to independently learn the transcriptional and morphological features of spots from the reconstructed data for down-stream analysis. In experiments, MuCST facilitates the critical applications of spatially resolved data, including spatial domain identification, tumor micro-environment, and denoising spatial transcriptomics (Fig. 1D).

In all, MuCST is a flexible network-based model for integrating spatially resolved data, which jointly learns the compatible features of spots with multi-modality contrast learning. It attempts to address heterogeneity of spatial omic data with network models, where heterogeneity of data is modeled and removed at feature level. Furthermore, MuCST not only facilitates users for down-stream analysis, but also serves as the pre-processing step for spatially resolved data, such as denoising and restoring SRT data.

Benchmarking MuCST with simulated spatially resolved data

To evaluate performance of MuCST, we first utilize simulated spatially resolved data, including the morphological information, spatial coordinate and transcription profiles of cells, where is originated from as MUSE [35] (generation and visualization of simulated datasets are shown in Additional file 1: Section 1.6 and Additional file 1: Fig. S1A, respectively). The typical algorithms, such as CCA (with PCA for features) [62], AE (auto-encoder) [63], MUSE [35], as well as the concatenation of various features (also called feature concat.), and other multi-modal integrative algorithms such as conST [37], ConGI [36], stMVC [33], and stMGATF [34], are selected as baselines, where SpiceMix [28] and CellCharter [29] are excluded since they are designed for integrating multiple slices. In details, CCA and ConGI learn features by maximizing cross-modality correlation, and AE integrates multi-modal data by reconstructing the original data. conST directly concatenates heterogeneity multi-modal features as attributes of cell network, and stMVC and stMGATF employ semi-supervised strategy for integrative analysis of multi-modal data. We select the Adjusted Rand Index (ARI) [50], NMI and $F1$-score to measure performance of various algorithms for identifying domains in the simulated dataset.

We first validate capacity of algorithms to learn discriminative features from each modality, where the number of domains in the full multi-modal space is fixed by randomly merging different clusters for each modality. As the number of clusters decreases, CCA and feature concatenation-based algorithms are similar to the single-modality approaches, whereas MUSE and MuCST are significantly superior to others, demonstrating that these algorithms capture the discriminative features of multi-modal data (Fig. 2A, Additional file 1: Fig. S1B). Furthermore, MuCST outperforms MUSE in all these cases. Specifically, ARI of MuCST is 0.880 ± 0.024 (for 15 clusters), 0.887 ± 0.020 (for 10 clusters), and 0.883 ± 0.018 (for 6 clusters) respectively, which is 2%∼5% higher than MUSE. Visualization of features learned by various algorithms with t-SNE [64] demonstrates that MuCST learns the compatible and discriminative features of cells that precisely characterize and model structure of domains (Additional file 1: Fig. S1B). However, AE and conST are even worse than single-modality approaches, indicating that heterogeneity of modalities greatly affects performance of algorithms. And, ConGI neglects spatial information, which results in an undesired performance. stMVC and stMGATF achieve an excellent performance since they take 50% annotations as prior information, and performance of them also decreases with the number of clusters increases.

Next, we validate performance of MuCST by degrading quality of one modality, where two persistent strategies, i.e., dropouts and noise, are selected to perturb transcription data. By varying dropout rate, average ARI of morphology-alone method is ∼0.6 (Fig. 2B, center horizontal dashed lines). As the quality of transcript degrades (from right to left along with x-axis), performance of all these multi-modal methods drops dramatically. In all, MUSE, stMVC, stMGATF, and MuCST are much more robust than others (Fig. 2B, region between "min" and "max" of morphology-alone). Furthermore, MuCST is inferior to stMVC and stMGATF, but is superior to others for all the dropout rates. The reason is that stMVC and stMGATF make use of 50% labels as prior, whereas MuCST requires no prior information. These results demonstrate that MuCST automatically discriminates the high- and low-quality modality, thereby improving performance of algorithms. Visualization of features learned by various algorithms demonstrates that MuCST precisely models structure of ten domains regardless of dropout rate (Fig. 2C and Additional file 1: Fig. S1C). Notice that CCA, conST, ConGI, and feature concatenation are dramatically affected by degradation of data quality because these algorithms solely focus on deriving relations among various modalities, thereby resulting in high sensitivity to data perturbation. Furthermore, by replacing ARI with NMI and $F1$-score, performance of MuCST is robust for simulated datasets by varying the number of clusters and dropout rates (Additional file 1: Fig. S2A and S2B).

Moreover, simulated dataset is simultaneously contaminated for transcript and morphology modalities by using additive Gaussian random noise with various variances. Performance of these algorithms drops as variance of noise increases, and MuCST achieves the best performance (Additional file 1: Fig. S2C left panel). In detail, AE, CCA and feature concatenation are inferior to single modality approaches, failing to learn compatible features from noised heterogeneous multi-modal data, whereas MUSE and MuCST precisely characterize noise and learn discriminative features. Furthermore, performance gap between MuCST and MUSE dramatically enlarges as the variance of noise increases from 0.1 to 2, demonstrating that MuCST is more precise and robust than MUSE. Three reasons explain why MuCST is superior to baselines. First, MuCST employs attributed cell network model for multi-modal data, which provides a better and comprehensive way to characterize intrinsic structure of domains. Second, MuCST makes use of contrast learning to remove heterogeneity of multi-modal data, thereby improving quality of features. Third, MuCST reconstructs the original multi-modal data to preserves specificity of each modality, where features of cells reaches a good balance between consistence and specificity.

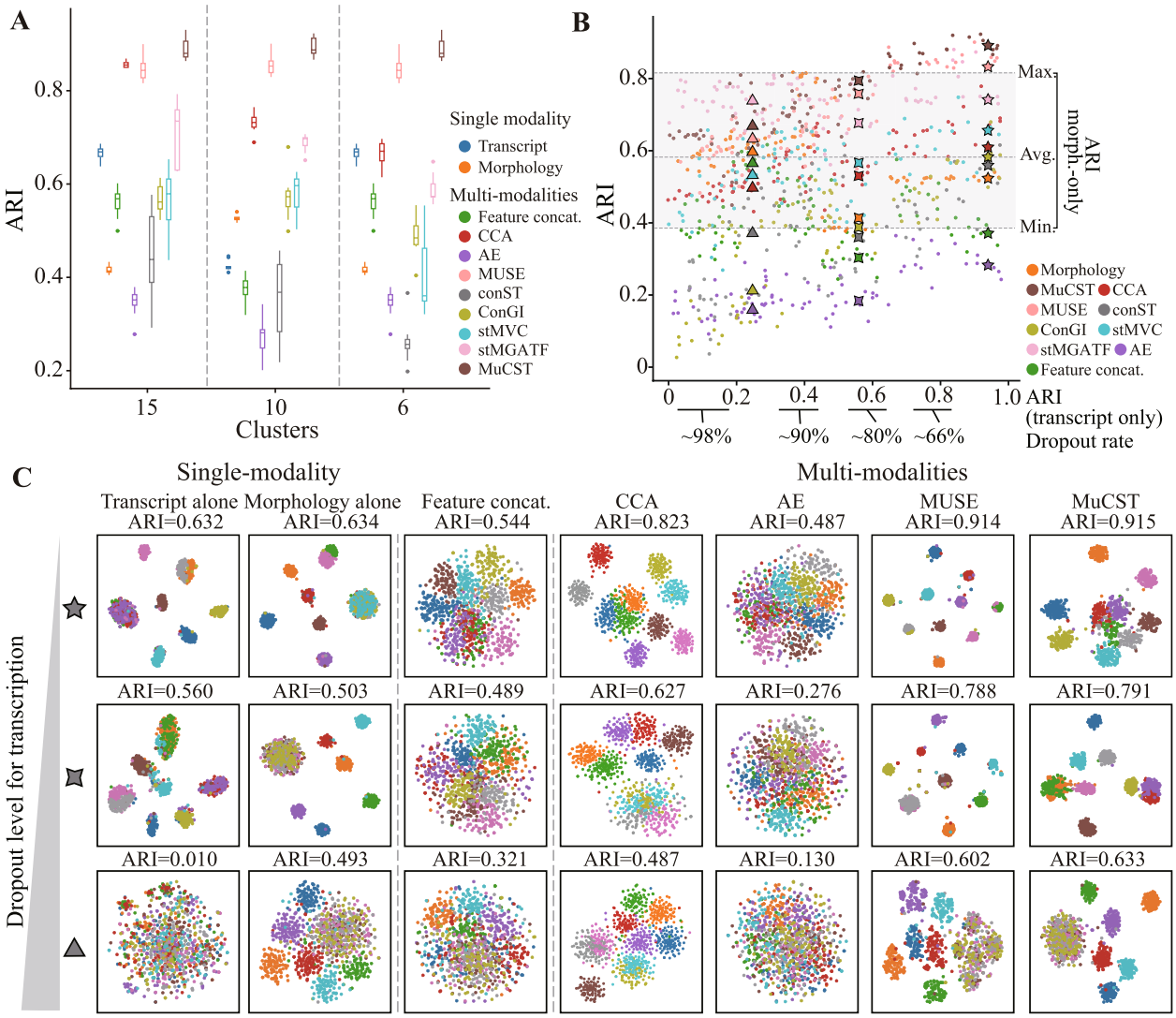

Fig. 2 Performance of various algorithms on the simulated data. **A** ARI of identifying ground truth high-resolution subpopulations from lower-resolution single-modality subpopulations ($k = 15$, 10, or 6), where 1000 cells with transcriptional and morphological profiles are simulated. And, box plot is based on 10 replicates with median (center line), interquartile range (box) and data range (whiskers). **B** ARI of identifying ground truth clusters by varying the range of dropout levels from the transcriptional modality, where dashed lines denote minimum, average and maximum ARI of morphology modality alone, x axis represent ARI of PCA on transcript modality alone, y axis denotes ARI of combined-modality methods. **C** tSNE visualizations of latent representations from single- and combined-modality methods, where ground truth subpopulation is labeled with various colors in simulation

To check whether strategy for clustering effect performance of algorithms, we employ PheoGraph, Hierarchical and K-means to identify domains in simulated dataset with features of cells learned by various algorithms, where performance of these algorithms is consistent with that of the original ones, indicating that they are insensitive to the selection of clustering methods (Additional file 1: Fig. S2C right panel). Furthermore, MuCST achieves a good balance between efficiency and accuracy, where it saves 50% running time of MUSE with even higher performance (Additional file 1: Fig. S2D). Parameter analysis demonstrates that MuCST is quite stable (Additional file 1: Section 1.4, Additional file 1: Fig. S2E and S3). And, running time of various algorithms on biological datasets demonstrate MuCST achieves an balance between efficiency and accuracy, indicating that it is applicable to large-scale datasets (Additional file 1: Fig. S4, Additional file 1: Table S2 and Table S3). Overall, MuCST not only captures discriminative features in multi-modal data, but also effectively avoids being confounded by data quality of different modalities.

MuCST significantly enhances performance of identifying spatial domains for various tissues

Spatial domains play a crucial role for investigating structure and functions of tissues [65], and we validate performance of MuCST for identifying spatial domains by using the LIBD human dorsolateral prefrontal cortex (DLPFC) dataset [52], 10 × Visium dataset of mouse brain tissue, and the human intestine dataset [55]. Fifteen state-of-the-art clustering algorithms, including SCANPY [17], Giotto [20], stLearn [31], SEDR [25], BayesSpace [21], SpaGCN [30], STAGATE [22], SpatialPCA [27], DeepST [32], ConGI [36], conST [37], stMVC [33], Spatial-MGCN [24], stMGATF [34], and MUSE [35], are selected as baselines.

DLPFC dataset consists of 12 slices obtained from human brain that are manually annotated as six layers of dorsolateral prefrontal cortex (Layer1 ∼ Layer6) and white matter (WM) on the basis of histology image and gene markers (Fig. 3A). MuCST outperforms baselines on the identification of spatial domains in slice 151673 with ARI 0.641, while that of the best baseline is 0.620 (Fig. 3A, Additional file 1: Fig. S5–S8). MUSE is criticized for ignoring spatial information that is critical factor for spatial domains. These results demonstrate that MuCST captures discriminative features of spots. Performance of various algorithms for all 12 slices of DLPFC in terms of ARI, NMI, and F1-score is presented (Fig. 3B and Additional file 1: Fig. S9) where MuCST outperforms baselines except for stMVC. In details, ARI of MuCST is 0.584 ± 0.060, whereas that of Spatial-MGCN is 0.498 ± 0.097, conST 0.437 ± 0.052, DeepST 0.501 ± 0.077, and BayesSpace 0.432 ± 0.104 (mean ± standard deviation, Fig. 3B). stLearn and SCANPY are inferior to others because these algorithms either utilize one of modalities, or fail to fully integrate multi-modalities, which is consistent with assertion in simulated data. Furthermore, MuCST is more robust than others since its variance is much less than baselines, demonstrating that MuCST learns discriminative features of spots for all slices (Additional file 1: Fig. S5–S8). The reason why stMVC outperforms others is that it takes 50% annotations as prior, whereas MuCST requires no prior information (stMVC: 0.636 ± 0.099 vs MuCST: 0.584 ± 0.060). By replacing ARI with NMI and F1-score, performance of MuCST is consistent with that of ARI, demonstrating MuCST is insensitive to measurements (Additional file 1: Fig. S9). To investigate whether stMVC outperforms MuCST is due to prior information, we also incorporate prior information into MuCST, and find that MuCST is much better than stMVC with the same prior information, indicating the superiority of MuCST for spatial domain identification (Additional file 1: Fig. S10A).

Since MuCST integrates morphological and transcriptional features with contrast learning to remove heterogeneity of spatially resolved data, it is natural to validate quality of features. Layer 6 and WM are critical spatial domains in brain, and evidence demonstrates that integrating all slices of DLPFC dataset can precisely discriminate these two domains [28, 29]. Interestingly, among these single slice based algorithms, only MuCST and STAGATE precisely discriminate Layer 6 and WM. We compare distribution density of the normalized features learned by various algorithms for the ground truth Layer 6 and WM (Fig. 3C). Surprisingly, features learned by MuCST significantly discriminate these two domains, whereas all these baselines fail to discriminate them (Fig. 3C, Additional file 1: Fig. S10B). For example, deviation of distribution of raw features learned by MuCST is 1.69 and 7.62 for Layer 6 and WM respectively (p = 7.6E−3, two-sided Kolmogorov-Smirnov (KS) test), whereas that of MUSE is 0.56 (Layer 6) and 0.53 (WM) respectively (p = 0.70, two-sided KS test). Furthermore, either transcript or morphology also fails to discriminate Layer 6 and WM (transcript: p = 0.72, morphology: p=0.67, two-sided KS test). These results demonstrate that morphology is critical complementary information for characterizing and modeling spatial domains in spatial transcriptomics data, providing an alternative for integrating multiple slices of SRT data as Ref. [28, 29].

Trajectory of spatial domains is fundamental for revealing mechanisms of biological evolution [66], and PAGA [67] is employed to infer relations of spatial domains identified by various algorithms. MuCST and STAGATE precisely identify the organization of the cortical layers derives from L1 to L6 and WM with high PAGA score. But, the other baselines mistakenly draw connections among various spatial domains whose PAGA score is much less than them (Additional file 1: Section 1.8, Additional file 1: Fig. S11A). These results demonstrate that MuCST accurately captures intrinsic structure and evolutionary relations of spatial domains. We further perform a comprehensive parameter analysis of MuCST on the DLPFC slice, demonstrating that MuCST is robust, and is capable of fast convergence during training (Additional file 1: Fig. S11B and S11C).

Two additional 10× Visium datasets from mouse posterior and coronal brain are selected to further validate performance of MuCST, where the anatomical reference annotations are from the Allen Mouse Atlas. Figure 3D visualizes annotation of posterior tissue and spatial domains identified by various algorithms, where MuCST outperforms baselines (Additional file 1: Fig. S12). Specifically, transcript or morphology solely identifies Cerebellum (CB) area, but fails to distinguish Hippocampal Formation (HPF) and Brain Stem (BS) (Additional file 1:

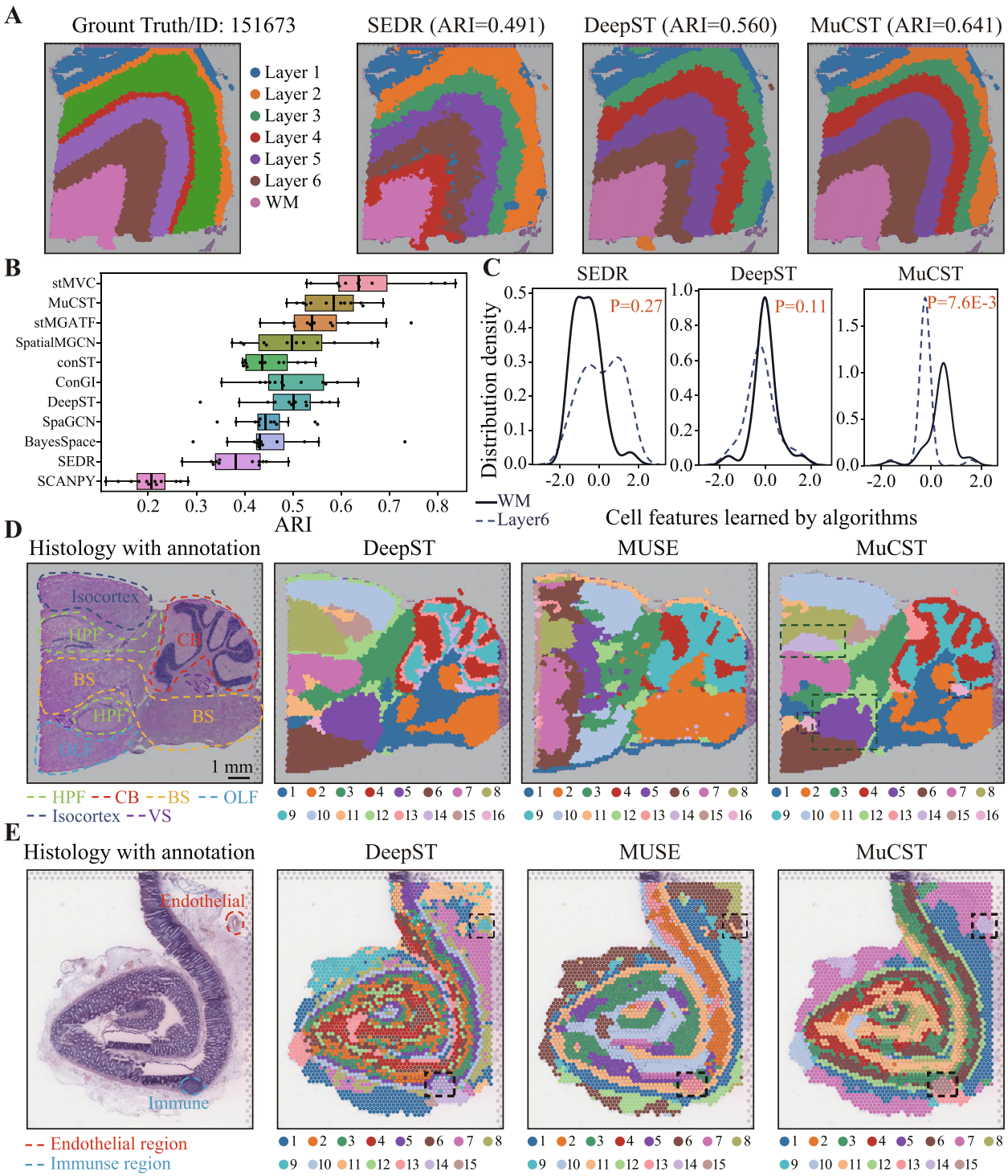

Fig. 3 MuCST significantly enhances performance of spatial domain identification in normal tissues. **A** Ground truth segmentation of cortical layers and white matter (WM) in slice 151673 of DLPFC data, and visualization of spatial domains identified by SEDR, DeepST, and MuCST in slice 151673. **B** Boxplot of ARIs of various algorithms for spatial domains in all 12 slices of DLPFC, where *x*-axis denotes ARI, and the center line, box limits, and whiskers denote the median, upper and lower quartiles, and 1.5 × interquartile range, respectively. **C** Distribution density of cell features learned by SEDR, DeepST, and MuCST for Layer 6 and WM, where two-sided KS test is employed for significance. **D** Annotated histology image of mouse brain posterior (left), and spatial domains obtained by different methods in delineating different structures of posterior brain. **E** Annotated histology image of human intestine (left), and spatial domains obtained by different methods in delineating spatial structures of human intestine

Fig. S12A and S12B). DeepST cannot identify Cornu Ammonis (CA) and Dentate Gyrus (DG) areas, whereas MuCST precisely identifies the CA and DG areas within the HPF areas in the mouse brain (surrounded by the dashed squares, Fig. 3D), as well as the Cerebellar Cortex and Dorsal Gyrus areas in the sagittal posterior mouse brain (surrounded by the dashed squares, Additional file 1: Fig. S12B), which are consistent with the reference annotations. Since no spot-level annotation is available, we employs the Silhouette Coefficient (SC) and Davies-Bouldin Index (DB) to measure compactness and separation of spatial domains, where MuCST achieves higher SC and lower DB score than baselines, indicating that these domains identified by MuCST are more precise from perspective of computation (Additional file 1: Fig. S13A). Furthermore, MuCST also has higher overlap ratio (Additional file 1: Fig. S13B), demonstrating that MuCST outperforms MUSE. We finally examine the expression levels of known bio-marker genes within the corresponding identified domains (Additional file 1: Fig. S13C), it is evident that these bio-marker genes have the highest expression levels in the spatial domains identified by MuCST than those identified by others. Moreover, MuCST also effectively detects the Cornu Ammonis and Dentate Gyrus in HPF regions, demonstrating that MuCST delineates the spatial domain in more details (surrounded by the dashed squares, Additional file 1: Fig. S14A and S14B). Furthermore, spatial domains identified by MuCST are with stronger regional continuity and fewer noise points than others (Additional file 1: Fig. S14B and S14C). These results indicate that MuCST is also promising for characterizing complicated spatial domains in mouse brain. To validate contribution of DAPI staining to MuCST, we further adopt the mouse corona brain dataset with histology image stained by antibodies (Alexa Fluor 488 anti-NeuN) and DAPI, where MuCST precisely identifies the Ammon's horn as well as the dentate gyrus structure in the hippocampus, demonstrating that MuCST can effectively integrate morphological information extracted from DAPI staining images (Additional file 1: Fig. S14D), region surrounded by square with white border).

The human intestine dataset [55] with four major spatial domains, such as epithelium, muscle, immune and endothelium region is shown in Fig. 3E. MUSE and MuCST precisely identify these four spatial domains, whereas baselines fail to discriminate them (Fig. 3E, Additional file 1: Fig. S15A). Furthermore, morphology is much more precise than transcript, SEDR, SpaGCN, STAGATE, conST, ConGI, and Spatial-MGCN because morphology dominates transcript in the intestine dataset. Interestingly, MuCST precisely identifies all these four spatial domains, which are highly consistent with annotation (Additional file 1: Fig. S15B–S15E), demonstrating that MuCST reaches a good balance between transcript and morphology. Furthermore, MuCST also achieves a better performance in terms of SC and DB scores, overlap ratio, and expression of known bio-marker genes within the corresponding identified domains (Additional file 1: Fig. S16A–S16C). These results clearly demonstrate that MuCST achieves accurate spatial domain identification even on morphology-dominant human intestine data.

Overall, all these results demonstrate that MuCST enhances the identification of spatial domains for spatially resolved data from various tissues.

MuCST precisely reveals tumor heterogeneity from cancer spatially resolved data

Spatial transcriptomics technologies are widely applied to cancers, and it is natural to investigate the generalization power of MuCST for revealing tumor heterogeneity. Six typical datasets, including 10 × Visium human breast cancer, humanpancreatic ductal adenocarcinoma (PDAC) [57], human invasive ductal carcinoma (IDC) [21], human HER2 breast cancer [59], human prostate cancer [58], and Alzheimer's disease mouse brain datasets, are selected. The human breast cancer data consist of 3798 spots and 36,601 genes, which is manually annotated by pathologists [25], including 20 regions and 4 main morphotypes, i.e., ductal carcinoma in situ/lobular carcinoma in situ (DCIS/LCIS), healthy tissue (Healthy), invasive ductal carcinoma (IDC), and tumor edge regions (Fig. 4A left panel and Additional file 1: Fig. S17A).

Performance of MuCST on the human breast cancer dataset with the number of domains as 20 is shown in Fig. 4A (performance with various numbers of clusters is also investigated in Additional file 1: Fig. S17B). By comparing MuCST to baselines, MuCST achieves similar performance to stMVC and stMGATF, and outperforms other baselines (Fig. 4A and Additional file 1: Fig. S17C). In details, cancer-related spatial domains identified by MuCST are highly consistent with the manual annotations (ARI = 0.586), whereas domains obtained by baselines (except for stMVC and stMGATF) with less regional continuity and more outliers, implying that MuCST is also promising for characterizing and identifying cancer spatial domains. Furthermore, either transcript or morphology alone is insufficient to fully characterize cancer-related spatial domains since ARI of them is 0.444 and 0.260, respectively (Additional file 1: Fig. S17C), indicating that morphological and transcriptional information is complement for the characterization of tumor heterogeneity. Comparison among stMVC, stMGATF, and MuCST demonstrates that MuCST is much better than others if the same prior information is utilized, indicating

superiority of integrating morphology and SRT data (Additional file 1: Fig. S17D).

Then, we investigate quality of features learned by various algorithms for characterizing tumur heterogeneity of breast cancer by discriminating these 4 major morphotypes. Figure 4B describes distribution density of features learned by MUSE (left) and MuCST (right), where only MuCST precisely discriminate IDC, DCIS/LCIS, tumor edge and healthy morphotype (two-sided KS test for significance). However, all these baselines, except for DeepST, fail to discriminate these four major morphotypes (Additional file 1: Fig. S18A). Interestingly, distribution density of features learned by MuCST not only discriminates these major morphotypes, but also characterizes evolutionary of morphotypes of breast cancer from healthy to tumor edge, and then to IDC (Fig. 4B, healthy: 0.67 ± 1.25 vs tumor edge: 0.82 ± 1.31, $p = 1.3E{-}105$; tumor edge: 0.82 ± 1.31 vs IDC: 0.94 ± 1.38, $p = 3.2E{-}52$, two-sided KS test), which cannot be fulfilled by current baselines. These results demonstrate the proposed multi-modal contrastive learning strategy captures intrinsic structure of complicated cancer-related domains, providing an insight into mechanisms of tumors. Moreover, spatial domains identified by MuCST are divided into two categories with hierarchical clustering in terms of Pearson correlation coefficient, i.e., tumor and non-tumor group, where the latter one include tumor edge and healthy (Additional file 1: Fig. S18B). These results demonstrate that MuCST is also promising for characterizing and modeling tumor heterogeneity by exploiting meta-structure of spatial domains. And, to further dissect tumor heterogeneity, differentially expressed genes (DEGs) among these four major morphotypes are obtained, which are highly associated with breast cancers. For example, *APOE* and *C1Q1* in tumor edge regions are associated with the differential abundance of tumor-associated infiltration of macrophages (TAM) that is critical for survival outcomes of patients due to its role in promoting tumor angiogenesis [68, 69]. And, the up-regulated genes are involved in immune and signal pathway, and down-regulated ones are associated with cell cycle process (Additional file 1: Fig. S18C). Moreover, tumor heterogeneity results in hierarchical structure of spatial domains, i.e., annotation of IDC domain [32] is further divided into two sub-domains (domains 6 and 15, Fig. 4C). MuCST also precisely identifies them (Fig. 4C), where domain bio-marker genes, such as *ABCC11* and *TFF1*, are differentially expressed, where ABCC11 is a known marker and multi-drug resistance gene in human breast cancer [70], and TFF1 is associated with tumor differentiation [71]. Notice that only MuCST significantly discriminates these two domains at feature level ($p = 1.3E{-}2$, Kolmogorov-Smirnov test, Additional file 1: Fig. S18D). These results demonstrate that MuCST reveals tumor heterogeneity from various levels, i.e., from macro-level (spatial domains) to micro-level (feature).

The human pancreatic ductal adenocarcinoma (PDAC) dataset sequenced by Legacy platform [57] is also adopted, which is manually annotated with the normal, cancer, duct epithelium and stroma regions (Fig. 4D left panel, and Additional file 1: Fig. S19A). Spatial domains identified by MuCST are highly consistent with the manually annotated areas (Fig. 4D right panel), whereas baselines fail to identify pancreatic cancer related domains (Additional file 1: Fig. S19B). Specifically, almost all these baselines mix the cancer and non-cancer domains, demonstrating superiority of MuCST for modeling tumor heterogeneity in complex tissues. We further conduct differential expression analysis between cancer and normal region (Fig. 4E left), where four of the top 5 DEGs, i.e., *KRT17*, *LAMC2*, *S100A14*, and *TM4SF1*, are bio-markers for PDAC [72, 73]. Furthermore, these genes are significantly associated with survival time of patients, further substantiating the accuracy of spatial domains identified by MuCST (log-rank test for significance, Additional file 1: Fig. S19C). Overlapping ratio of domains identified by MuCST are more consistent with annotation than baselines (Fig. 4E right panel, two sided Student's test). The additional human HER2 breast cancer dataset, where the zoomed-in regions are manually annotated

(See figure on next page.)

Fig. 4 MuCST accurately identifies cancer-related domains for revealing tumor heterogeneity. **A** Visualization of annotation of human breast cancer data with healthy, tumor edge, IDC and DCIS/LCIS morphotype (left), and spatial domains identified by DeepST, MUSE, and MuCST, respectively (right). **B** Distribution density of spot features of breast cancer data learned by MUSE (left) and MuCST (right) for four morphotypes, respectively, where *x*-axis denotes features and *y*-axis denotes estimated distribution density (two-sided KS test for significance). **C** Visualization of expression of *TFF1* and *ABCC11* between domain 6 and 15 (top), and violin plots of expression of these two genes (bottom). **D** Visualization of region-level manual annotation of the human pancreatic ductal adenocarcinoma dataset with the normal, cancer, duct epithelium and stroma regions [57] (left), and spatial domains identified by MuCST (right). **E** Overexpressed genes in cancer regions through differential expression analysis between tumor and non-tumor regions characterized by MuCST (left), and overlap ratio of tumor region and marker genes identified by different algorithms (right). **F** Histology images of brain tissue sections from transgenic mice at the middle stage, where the zoomed in regions correspond to the primary region of amyloid plaque deposition. **G** Disease-related spatial domains identified by MUSE and MuCST, respectively, where the zoomed in regions correspond to the primary region of amyloid plaque deposition

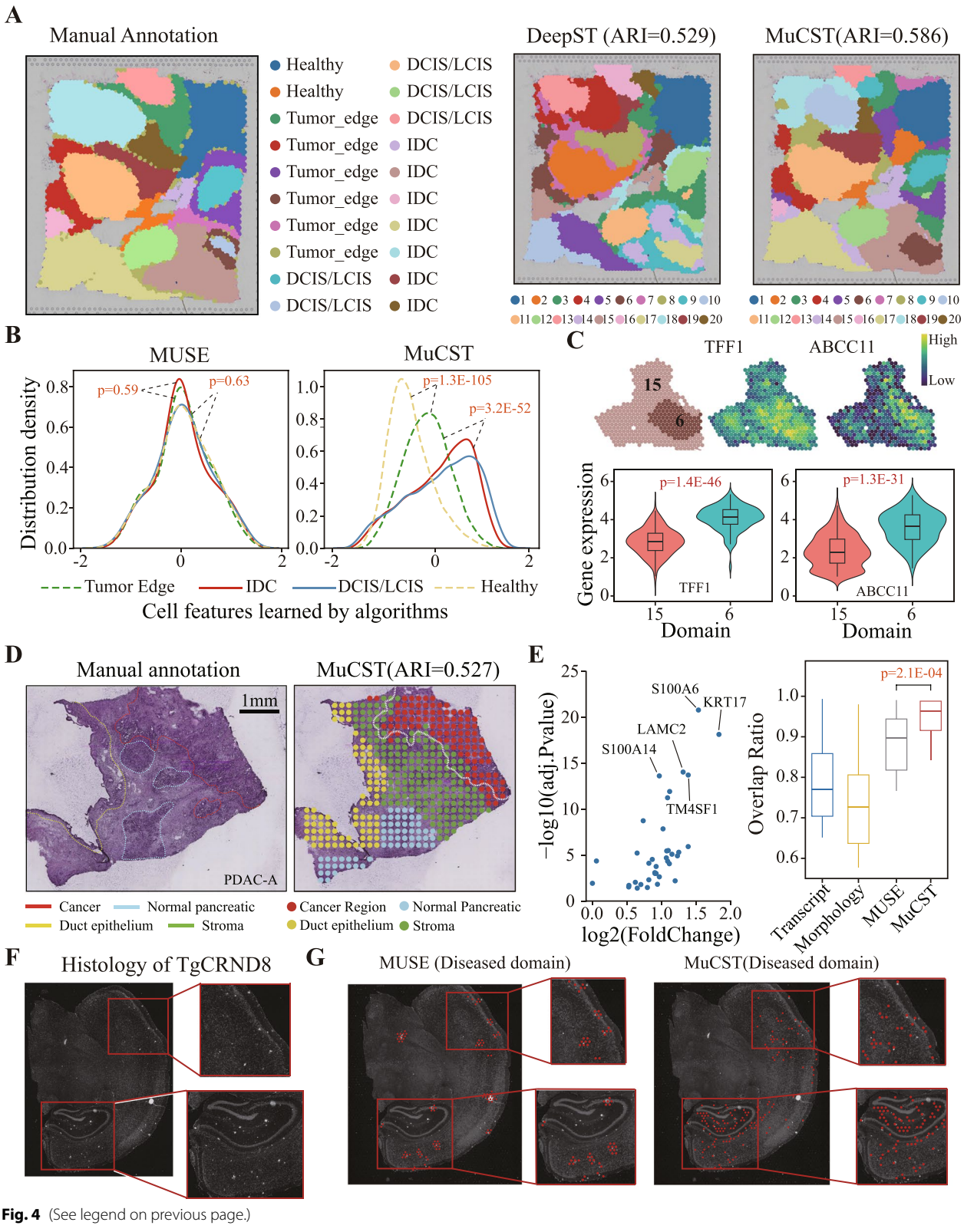

Fig. 4 (See legend on previous page.)

as rare ones associated with breast cancer, which was ignored by pathologist (Additional file 1: Fig. S20A and S20B). The transcript-alone approach misclassifies it as *In situ cancer*, rather than the annotated *Breast glands*, whereas the morphology-alone approach fails to identify these rare spatial domains (Additional file 1: Fig. S20**C**). The two semi-supervised algorithms, stMVC and stMGATF, along with MuCST, accurately identify these rare cancer domains, highlighted by white solid squares. Differential gene analysis identify bio-marker genes of the marker genes of *Breast glands* and *In situ cancer* (Additional file 1: Fig. S20D–E). Furthermore, we also validate performance of MuCST with the human prostate cancer and human invasive ductal carcinoma (IDC) dataset (Additional file 1: Fig. S21), where spatial domains identified by MuCST are consistent with both region-level and spot-level annotations. For example, 4 regions correspond to the annotated regions of predominantly IC (2, 6, 7, and 9), carcinoma in situ (5), benign hyperplasia (1), and predominantly non-tumor areas (3, 4, 8, and 10). By replacing ARI with NMI and *F*1-score, performance of MuCST on PDAC, HER2, IDC, and Prostate dataset are quit stable, showing robust of the proposed algorithm (Additional file 1: Fig. S22).

Finally, the Alzheimer's disease (AD) mouse brain dataset is selected because it is morphology-dominated, where algorithms without integrating morphology are invalid. Fig. 4F visualizes immunofluorescence image of brain slice from a 5.7-month-old transgenic mouse, where white amyloid plaques highlight the primary areas of amyloid-beta ($A\beta$) deposition. MUSE and MuCST identify discrete spatial domains associated with amyloid plaque accumulation, whereas transcript- or morphology-alone approach only recognizes coherent spatial domains of mouse brain tissue (Fig. 4G, Additional file 1: Fig. S23A). Furthermore, MuCST identifies the accumulation of amyloid plaques in the hippocampal region that cannot be accomplished by other baselines. Moreover, MuCST also identifies the plaque-covered areas by preserving the healthy brain regions on late-stage AD mouse brain slices (Additional file 1: Fig. S23B–S23D). Analogously, differential expression analysis between disease and normal domain identify DEGs that are significantly enriched with gliogenesis and glial cell differentiation, which are associated with neuro-degenerative diseases in brain regions [74] (gene-ontology enrichment analysis, hypergeometric test, Additional file 1: Fig. S23E). Spatial distribution of expression of *Gfap* and *Cst3*) is localized in the $A\beta$ accumulation, which are highly expressed in AD brain slices, exhibiting low expression in healthy brain slices (Additional file 1: Fig. S23F and S23G).

In summary, MuCST is more accurate to model and extract cancer-related domains, facilitating the understanding of tumor heterogeneity at various levels, which provides an alternative for integrative analysis of spatially resolved data.

MuCST precisely characterizes and removes noise in spatially resolved data

Evidence demonstrates that spatially resolved data suffer from noise because of dedicated procedures to preserve transcriptional and spatial information. Therefore, denoising is critical pre-processing for down-stream analysis [38, 75–77]. MuCST restores the original data with the compatible features, providing an alternative for denoising of spatially resolved transcriptomics.

To fully validate quality of restored data, the third-party algorithm SCANPY is selected to perform spatial domain identifications on the original and restored data, respectively. Figure 5A illustrates the identified spatial domains from the original (left) and restored (right) data of slice 151673 in DLPFC, where ARI dramatically improves from 0.181 to 0.480. Obviously, domains in the original data are suffer from high level of noise with mixed boundary, whereas these domains from the reconstructed data are clear and accurate. Specifically, WM and Layer 6 are clearly classified in the restored data, demonstrating that MuCST captures intrinsic structure of spatially resolved data by removing noise. Furthermore, we compare distribution of features learned by SCANPY between the ground truth WM and Layer 6 for the original and restored transcript, where difference between these domains is non-significant in the original data (Fig. 5B left), but it is significant in the restored data (Fig. 5B right). In detail, the standard deviation of features from the original data is 1.03 and 0.76 for WM

(See figure on next page.)

Fig. 5 MuCST precisely removes noise in spatially resolved data to facilitate down-stream analysis. **A** Visualization of spatial domains in slice 151673 of DLPFC identified by SCANPY based on the raw (left) and restored (right) spatial transcriptomics data, respectively. **B** Distribution density of spot features of slice 151673 learned by SCANPY for Layer 6 and WM from the original (left) and restored (right) spatial transcriptomics data respectively, where *x*-axis denotes features and *y*-axis denotes estimated distribution density (two-sided KS test for significance). **C** Distributions of ARIs of various algorithms for identifying spatial domains with the original and reconstructed DLPFC data respectively, where *y*-axis denotes ARI, and one-sided Student's *t* test for significance. **D** Visualizations of the original (up), reconstructed data (middle) and expression of layer-marker genes (bottom) in slice 151673, where each column corresponds to one layer (two-sided Student's *t*-test for significance)

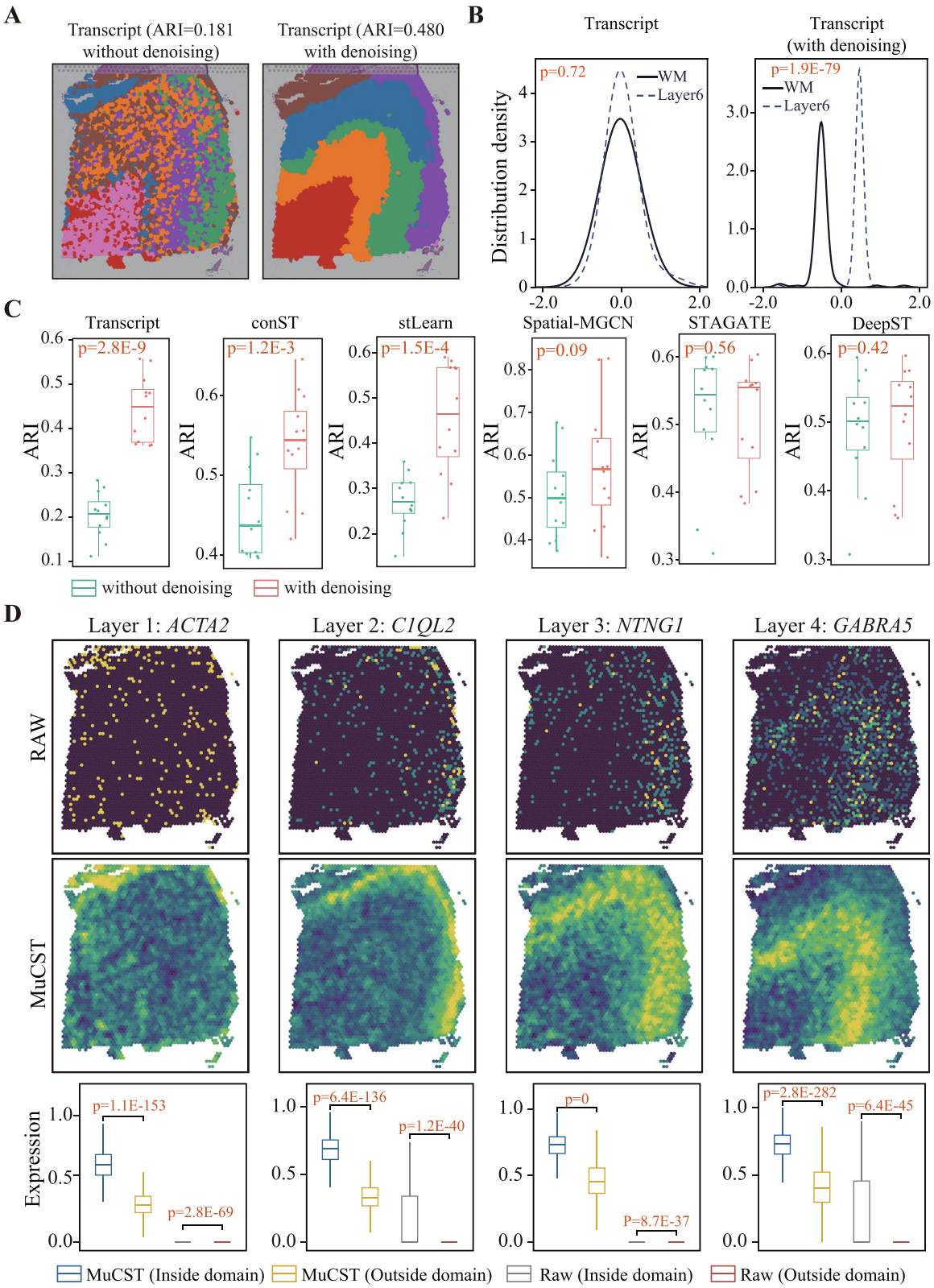

Fig. 5 (See legend on previous page.)

and Layer 6, respectively (p = 0.72, two-sided KS test), whereas that of the restored data is 2.43 and 0.67, respectively (p = 1.9E−79, two-sided KS test). These results show that MuCST precisely removes noise in spatially resolved data by exploiting the topological and multi-modality relations among them, thereby enhancing the discriminative of features.

To check whether improvement of denoising is co-factored by algorithms and data, we apply all these baselines to the original and restored DLPFC data, where distributions of ARIs of all these algorithms on DLPFC data are described in Fig. 5C. Surprisingly, all these algorithms improve performance on the restored data than the original one, proving that improvement of performance is not co-factored by the algorithms and data. Moreover, SCANPY, stLearn, and conST significantly improve performance of identifying spatial domains, and the other baselines also enhance performance with the restored data. For example, ARI of SCANPY (Transcript) increases from 0.205 ± 0.062 to 0.448 ± 0.071 (p = 5.0E−9, one-side Student's t-test), whereas that of conST and stLearn soars from 0.271 ± 0.057 and 0.437 ± 0.052 to 0.465 ± 0.118 and 0.544 ± 0.068 (p = 1.3E−9, one-side Student's t-test). Even though improvement for Spatial-MGCN and DeepST is non-significant (Fig. 5C), the restored data lead to 14.98% and 4.6% improvement, respectively. The possible reason why STAGATE enlarges deviation of ARIs on the restored data is that it also performs denoising with an auto-encoder strategy, i.e., double denoising procedures result in an undesirable performance. These results demonstrate that multi-modality fusion is promising for characterizing and modeling noise in spatially resolved data, and MuCST can also serve as a pre-processing tool for down-stream analysis.

Since bio-marker genes are critical for spatial domains [55, 78], we then compare the expression of layer-marker genes for each layer between the original and restored data for slice 151673. Figure 5D visualizes expression of bio-markers for each layers, such as *ACTA2* (Layer 1), *C1QL2* (Layer 2), *NTNG1* (Layer 4), and *GABRA5* (Layer 4) [52], for the original (up), and restored data (middle), where each column corresponds to a layer. It is easily observed that the bio-marker genes are not consistent with the structure of layers because of noise in the original data, while all these bio-marker genes are located in the corresponding domains. Then, we compare the expression of layer bio-marker genes within and outside of the corresponding layer in the restored data, where all these bio-marker genes are significantly expressed within domain than outsides (two-sided Student's t-test, Fig. 5D bottom). Even though difference of these layer bio-marker genes is also significant, the difference is not as large as these in the restored data. These results demonstrate that MuCST precisely removes noise in spatially resolved data by augmenting expression of layer bio-marker genes, thereby improving quality of data.

The human breast cancer dataset is also selected validate performance of MuCST for denoising (Fig. 4A), where all these algorithms achieves higher accuracy on the restored data than the original one (Additional file 1: Fig. S24A). Particularly, transcript-alone approach also enhances ARI from 0.444 to 0.546, and morphology-alone method increases ARI from 0.263 to 0.283, proving that improvement of accuracy is not co-factored by algorithms. Moreover, we also compare distribution density of cell features learned by various algorithms for IDC and DCIS/LCIS, where the difference of features for these two layers is significant on the restored data, and is non-significant on the original one (Additional file 1: Fig. S24B, two-sided KS test), demonstrating that MuCST also precisely removes noise in breast cancer dataset. Finally, we also validate that the restored data also facilitate identification of spatial domains with bio-marker genes (Additional file 1: Fig. S24C). In summary, MuCST precisely characterizes and removes noise with multi-modal contrastive learning, which can serve as critical pre-processing step for analyzing spatially resolved data.

MuCST is applicable for spatial omics data with various platforms

Here, we investigate the applicability of MuCST with spatially resolved data generated with different platforms, such as STARmap [11], osmFISH [54], Slide-seq V2 [14], and Stereo-seq [15]. The mouse primary visual cortex dataset generated by STARmap (Fig. 6A) is selected, which is annotated as seven distinct layers from raw fluorescence data (Fig. 6B). And, either transcript or morphology cannot fully characterize spatial domains with ARI 0.262 and 0.059 respectively because STARmap data is dominated by transcript. MuCST achieves the best performance among these unsupervised baselines with ARI 0.652, whereas that of STAGATE, SpaGCN, MUSE, ConGI, and conST is 0.586, 0.492, 0.057, 0.510, and 0.400, respectively (Fig. 6B, Additional file 1: Fig. S25A). stMVC and stMGATF achieve an excellent performance because they make use of 50% labels for spatial domains to guide feature learning, which is invalid if the balance of transcript and morphology loses. These results demonstrate that MuCST is also promising for integrating spatially resolved data generated with STARmap platform.

We then validate the applicability of MuCST with three additional datasets from various platforms without morphological information, including the mouse brain cortex data with osmFISH [54], mouse hippocampus tissue

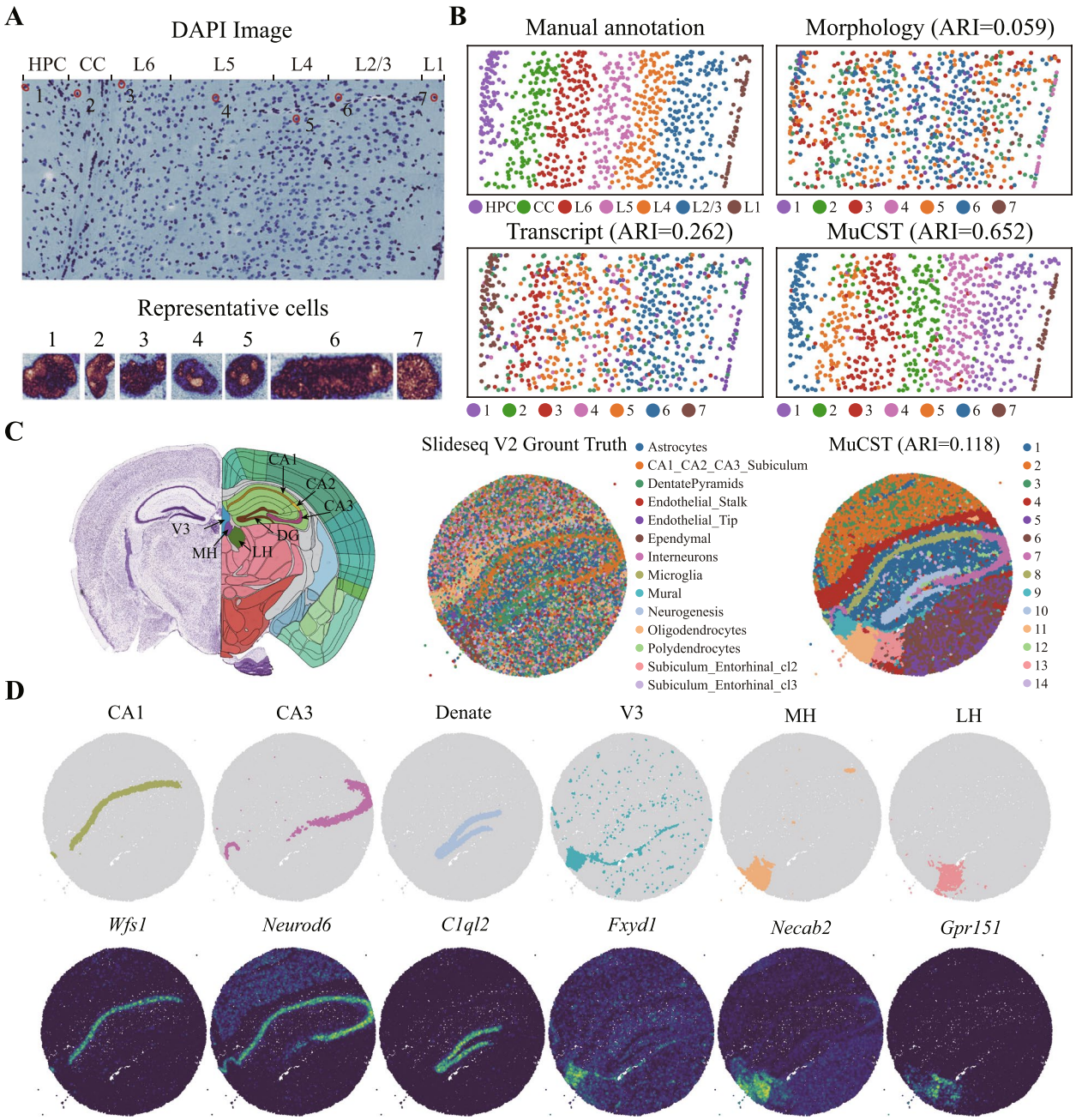

Fig. 6 MuCST is applicable for spatially resolved data with various platforms. **A** Raw DAPI image of the V1 tissue annotated with seven functionally distinct layers (upper panel), and seven representative cells from different layers (bottom panel). **B** Manual annotation of seven layers, and spatial domains identified by single modality and MuCST. **C** Visualization of mouse hippocampus tissue from Allen Mouse Brain Atlas (left), visualization of mouse hippocampus tissue annotated by [41] (middle), and spatial domains identified by MuCST (right). **D** Visualization of the spatial domains identified by MuCST, and the corresponding marker spatial gene expressions. The identified domains are aligned with the annotated hippocampus region of the Allen Mouse Brain Atlas

with Slide-seq V2 [14], and mouse olfactory bulb tissue with Stereo-seq [15]. The mouse brain cortex is a non-lattice-shaped spatially resolved transcriptomics dataset generated by osmFISH, where spatial domains are labeled with different colors (Additional file 1: Fig. S25B). Compared to the state-of-the-art algorithms, MuCST and STAGATE achieve the best performance with ARI ~0.500, demonstrating that the proposed model also works well for osmFISH data. The mouse hippocampus dataset generated with Slide-seq V2 are annotated based

on the Allen Brain Atlas [41] (Fig. 6C left and middle panel). MuCST successfully identifies annotated spatial domains, including the dentate gyrus (DG) and the pyramidal layers within Ammon horn, which are further separated into fields CA1, CA2, and CA3. And, it outperforms DeepST on the delineation of CA3 and DG (Additional file 1: Fig. S25C). Moreover, the spatial distribution of expression of domain bio-marker genes are consistent with annotation of hippocampus regions, where each column corresponds to a domain identified by MuCST (Fig. 6D). In contrast, baselines mix some domains, for example, DeepST merges the MH and LH (Additional file 1: Fig. S25C).

Finally, we apply MuCST to the coronal mouse olfactory bulb tissue acquired with Stereo-seq [15], which is annotated with the DAPI image, including the olfactory nerve layer (ONL), glomerular layer (GL), external plexiform layer (EPL), mitral cell layer (MCL), internal plexiform layer (IPL), granule cell layer (GCL), and rostral migratory stream (RMS) (Additional file 1: Fig. S25D). SCANPY, DeepST, and MuCST precisely identify domains in the outer layers, i.e., ONL, GL, and EPL, while SCANPY mixes GCL with the outer IPL region in inner structure (Additional file 1: Fig. S25E). We further adopt the spatial distribution of marker genes of each anatomical region to validate MuCST-identified domains, where a well match is observed, demonstrating domains identified by MuCST are consistent with annotations (Additional file 1: Fig. S25F). Overall, MuCST effectively leverage the whole transcript and spatial information to discern the relevant anatomical regions.

Discussion

The spatial transcriptomics measures gene expression at the cell level by retaining the associated spatial context, and integrating the gene expression, spatial coordinates, and morphological information of cells facilitates the identification of coherent cell patterns to understand the structure, functions and organization of tissues. However, it is highly non-trivial to integrate morphology and spatial transcriptomics because of noise and heterogeneity of data. In this study, we propose a novel and efficient algorithm (MuCST) to address this issue with contrastive learning.

In spatially resolved transcriptomics data, characterization of cellular heterogeneity is critical for revealing the structure and functions of tissues in health and disease. We first demonstrate that MuCST precisely reveals structure of domains by manipulating noise and quality of each modality in simulated datasets (Fig. 2). Then, we testify MuCST with morphology and SRT data from normal tissues, where MuCST discriminates critical domains with a single slice that previously only be separated by integrating multiple slices (Fig. 3). Furthermore, we validate that MuCST also precisely reveals tumor heterogeneity from five cancer spatially resolved datasets, where it effectively dissects tumor heterogeneity from the macro- and micro-level (Fig. 4). We further prove that MuCST also accurately models noise of spatially resolved data under the guidance of feature learning, which serves as pre-processing step for down-stream analysis (Fig. 5). We also validate that MuCST is also applicable to spatial omic data generated with various platforms (Fig. 6).

MuCST integrates morphology images and SRT data with contrastive learning to obtain discriminative and compatible representations of spots, providing a better way to characterize structure of tissues that cannot be fulfilled with single modality approaches. Here, we show that MuCST precisely identifies spatial domains from the normal as well as tumor tissues, covering different species, diseases, and platforms. The major difference between MuCST and state-of-the-art approaches lies in network-based multi-modality contrastive learning. Additionally, contributions of each component in MuCST is also investigated with comprehensive ablation studies on various datasets (Additional file 1: Section: 1.8, Additional file 1: Fig. S26–S29), where each component of MuCST is indispensable, further confirming the effectiveness of MuCST.

Pathology is fundamental for diagnosis and therapy of cancers, thereby enhancing capability of MuCST for addressing pathological features of cells is promising for the clinical pratice. Recently, three typical histology-specific models, such as Virchow [79], UNI [80], and Prov-GigaPath [81], are proposed to establish general-purpose foundation models for computational pathology. We exploit the possibility of incorporating histology-specific models into MuCST by replacing ResNet+SimCLR with either of Virchow, UNI and Prov-GigaPath. Figure S30 and Fig. S31 depict performance of variants of MuCST on the human breast cancer, DLPFC, and intestine dataset, respectively. In details, Fig. S30A describes performance of MuCST with various models for human breast cancer dataset, where histology-specific models improve performance of algorithms for H&E stained images, i.e., morphological features identified by histology-specific models are more accurate than ResNet+SimCLR. In details, ARI of ResNet+SimCLR is 0.263, whereas that of Virchow, UNI and Prov-GigaPath is 0.304, 0.449, and 0.430, respectively. Figure S30A2 shows that performance of MuCST enhances by replacing histology-specific models with ResNet+SimCLR, where ARI of MuCST increases from 0.586 to 0.633.

However, histology-specific models are not suitable for normal tissues (Additional file 1: Fig. S30B). For example, ARI of ResNet+SimCLR for slice 151673 of DLPFC is 0.283, whereas that of Virchow, UNI and Prov-GigaPath is 0.260, 0.270 and 0.217, respectively (Additional file 1: Fig. S30B1). And, ARI of MuCST decreases from 0.645 to 0.621 (Additional file 1: Fig. S30B2). Performance of these algorithms on all 12 slices of DLPFC dataset is presented in Fig. S30C, where consistence occurs. Moreover, performance of various methods on the human intestine dataset demonstrates ResNet+SimCLR achieves similar performance with histology-specific models (Additional file 1: Fig. S31A–S31D). The reason is that morphology images from normal and cancer tissues differ greatly, where histology-specific models are more precise to characterize tumor heterogeneity. These results demonstrate that MuCST flexibly integrates histology-specific models, showing its potential for clinical practice.

We see ample opportunities to extend potential clinical applications of MuCST. For example, MuCST provides a flexible framework to integrate pathological images and spatial transcriptomics data, which likely help clinicians to make decision by utilizing macro-level features of images and micro-level features of genes. Furthermore, we will also investigate whether spatial distribution of cancer-related domains facilitates clinicians to determine surgical plans in future.

Conclusions

In this work, we introduce a novel multi-modal contrastive learning algorithm, which is designed for addressing challenges inherent in integrating histology images with spatial transcriptomics data. MuCST not only facilitates the integration of multi-modal spatially resolved data, but also mitigates the impacts of noise and heterogeneity of multi-modal data.

Even though MuCST performs well on spatially transcriptomics datasets generated by 10× Visium, Legacy ST, STARmap, and Slide-seq, its ability to integrate morphology images with other spatial omics data is not fully investigated, such as Spatial-ATAC-seq data, which measures the chromatin accessibility landscape. Integrating these data bridges spatial domains with non-coding regulation in genome, facilitating the understanding of functions and structure of genes, which provides a more comprehensive way to investigate structure of tissues. In the future research, we will focus on integrating multi-omics data, and extending its clinical applications.

Supplementary Information

The online version contains supplementary material available at https://doi.org/10.1186/s13073-025-01449-1.

Additional file 1: Supplementary information, including supplementary notes for describing the detailed derivations of the MuCST algorithm, as well as supplementary tables and figures

Acknowledgements
The authors thank the members of the Ma Lab for helpful discussion and appreciate the researchers who provide us with source code for a comparison. The authors thank reviewers for their time and suggestions.

Authors' contributions
X.M. conceived and designed the study. X.M., Z.L., and Y.W. performed the research. X.M. and Y.W. implemented the model and performed simulation studies and benchmark evaluation, and released the source code on GitHub. Y.W. and Z.L. completed the downstream analysis. Z.L. manual annotate the cancer-related data based on marker gene and histology image. All authors read and approved the manuscript.

Funding
This work was supported by the National Science and Technology Major Project of China (No. 2024ZD0531100), Joint Funds of the National Natural Science Foundation of China (No. U22A20345), and National Natural Science Foundation of China (No. 62272361).

Data availability
Data is provided within the manuscript and supplementary information files [82, 83].

Declarations

Ethics approval and consent to participate
No ethical approval was required for this study. All utilized public datasets were generated by other organizations that obtained ethical approval.

Consent for publication
Not applicable.

Competing interests
The authors declare no competing interests.

Received: 17 November 2024 Accepted: 7 March 2025
Published online: 13 March 2025

References

1. Potter SS. Single-cell RNA sequencing for the study of development, physiology and disease. Nat Rev Nephrol. 2018;14:479–92.
2. Perlman ZE, Slack MD, Feng Y, Mitchison TJ, Wu LF, Altschuler SJ. Multidimensional drug profiling by automated microscopy. Science. 2004;306:1194–8.
3. Feldman D, Singh A, Schmid-Burgk JL, Carlson RJ, Mezger A, Garrity AJ, et al. Optical pooled screens in human cells. Cell. 2019;179:787–99.
4. Buettner F, Natarajan KN, Casale FP, Proserpio V, Scialdone A, Theis FJ, et al. Computational analysis of cell-to-cell heterogeneity in single-cell RNA-sequencing data reveals hidden subpopulations of cells. Nat Biotechnol. 2015;33:155–60.
5. Papalexi E, Satija R. Single-cell RNA sequencing to explore immune cell heterogeneity. Nat Rev Immunol. 2018;18:35–45.
6. Longo SK, Guo MG, Ji AL, Khavari PA. Integrating single-cell and spatial transcriptomics to elucidate intercellular tissue dynamics. Nat Rev Genet. 2021;22:627–44.
7. Moses L, Pachter L. Museum of spatial transcriptomics. Nat Methods. 2022;19:534–46.

8. Bressan D, Battistoni G, Hannon GJ. The dawn of spatial omics. Science. 2023;381:eabq4964.
9. Moffitt JR, Hao J, Wang G, Chen KH, Babcock HP, Zhuang X. High-throughput single-cell gene-expression profiling with multiplexed error-robust fluorescence in situ hybridization. Proc Natl Acad Sci. 2016;113:11046–51.
10. Eng CHL, Lawson M, Zhu Q, Dries R, Koulena N, Takei Y, et al. Transcriptome-scale super-resolved imaging in tissues by RNA seqFISH+. Nature. 2019;568:235–9.
11. Wang X, Allen WE, Wright MA, Sylwestrak EL, Samusik N, Vesuna S, et al. Three-dimensional intact-tissue sequencing of single-cell transcriptional states. Science. 2018;361:eaat5691.
12. Ståhl PL, Salmén F, Vickovic S, Lundmark A, Navarro JF, Magnusson J, et al. Visualization and analysis of gene expression in tissue sections by spatial transcriptomics. Science. 2016;353:78–82.
13. Vickovic S, Eraslan G, Salmén F, Klughammer J, Stenbeck L, Schapiro D, et al. High-definition spatial transcriptomics for in situ tissue profiling. Nat Methods. 2019;16:987–90.
14. Stickels RR, Murray E, Kumar P, Li J, Marshall JL, Di Bella DJ, et al. Highly sensitive spatial transcriptomics at near-cellular resolution with Slide-seqV2. Nat Biotechnol. 2021;39:313–9.
15. Wei X, Fu S, Li H, Liu Y, Wang S, Feng W, et al. Single-cell Stereo-seq reveals induced progenitor cells involved in axolotl brain regeneration. Science. 2022;377:eabp9444.
16. Yuan Z, Pan W, Zhao X, Zhao F, Xu Z, Li X, et al. SODB facilitates comprehensive exploration of spatial omics data. Nat Methods. 2023;20:387–99.
17. Wolf FA, Angerer P, Theis FJ. SCANPY: large-scale single-cell gene expression data analysis. Genome Biol. 2018;19:1–5.
18. Wu W, Ma X. Joint learning dimension reduction and clustering of single-cell RNA-sequencing data. Bioinformatics. 2020;36:3825–32.
19. Wu W, Liu Z, Ma X. jSRC: a flexible and accurate joint learning algorithm for clustering of single-cell RNA-sequencing data. Brief Bioinform. 2021;22:bbaa433.
20. Dries R, Zhu Q, Dong R, Eng CHL, Li H, Liu K, et al. Giotto: a toolbox for integrative analysis and visualization of spatial expression data. Genome Biol. 2021;22:1–31.
21. Zhao E, Stone MR, Ren X, Guenthoer J, Smythe KS, Pulliam T, et al. Spatial transcriptomics at subspot resolution with BayesSpace. Nat Biotechnol. 2021;39:1375–84.
22. Dong K, Zhang S. Deciphering spatial domains from spatially resolved transcriptomics with an adaptive graph attention auto-encoder. Nat Commun. 2022;13:1–12.
23. Long Y, Ang KS, Li M, Chong KLK, Sethi R, Zhong C, et al. Spatially informed clustering, integration, and deconvolution of spatial transcriptomics with GraphST. Nat Commun. 2023;14:1155.
24. Wang B, Luo J, Liu Y, Shi W, Xiong Z, Shen C, et al. Spatial-MGCN: a novel multi-view graph convolutional network for identifying spatial domains with attention mechanism. Brief Bioinform. 2023;24:bbad262.
25. Xu H, Fu H, Long Y, Ang KS, Sethi R, Chong K, et al. Unsupervised spatially embedded deep representation of spatial transcriptomics. Genome Med. 2024;16:12.
26. Liu W, Liao X, Yang Y, Lin H, Yeong J, Zhou X, et al. Joint dimension reduction and clustering analysis of single-cell RNA-seq and spatial transcriptomics data. Nucleic Acids Res. 2022;50:e72.
27. Shang L, Zhou X. Spatially aware dimension reduction for spatial transcriptomics. Nat Commun. 2022;13:7203.
28. Chidester B, Zhou T, Alam S, Ma J. SPICEMIX enables integrative single-cell spatial modeling of cell identity. Nat Genet. 2023;55:78–88.
29. Varrone M, Tavernari D, Santamaria-Martínez A, Walsh LA, Ciriello G. Cell Charter reveals spatial cell niches associated with tissue remodeling and cell plasticity. Nat Genet. 2024;56:74–84.
30. Hu J, Li X, Coleman K, Schroeder A, Ma N, Irwin DJ, et al. SpaGCN: Integrating gene expression, spatial location and histology to identify spatial domains and spatially variable genes by graph convolutional network. Nat Methods. 2021;18:1342–51.
31. Pham D, Tan X, Balderson B, Xu J, Grice LF, Yoon S, et al. Robust mapping of spatiotemporal trajectories and cell-cell interactions in healthy and diseased tissues. Nat Commun. 2023;14:7739.
32. Xu C, Jin X, Wei S, Wang P, Luo M, Xu Z, et al. DeepST: identifying spatial domains in spatial transcriptomics by deep learning. Nucleic Acids Res. 2022;50:e131.
33. Zuo C, Zhang Y, Cao C, Feng J, Jiao M, Chen L. Elucidating tumor heterogeneity from spatially resolved transcriptomics data by multi-view graph collaborative learning. Nat Commun. 2022;13:5962.
34. Li Y, Lu Y, Kang C, Li P, Chen L. Revealing Tissue Heterogeneity and Spatial Dark Genes from Spatially Resolved Transcriptomics by Multiview Graph Networks. Research. 2023;6:0228.
35. Bao F, Deng Y, Wan S, Shen SQ, Wang B, Dai Q, et al. Integrative spatial analysis of cell morphologies and transcriptional states with MUSE. Nat Biotechnol. 2022;40:1200–9.
36. Zeng Y, Yin R, Luo M, Chen J, Pan Z, Lu Y, et al. Identifying spatial domain by adapting transcriptomics with histology through contrastive learning. Brief Bioinforma. 2023;24:bbad048.
37. Zong Y, Yu T, Wang X, Wang Y, Hu Z, Li Y. conST: an interpretable multimodal contrastive learning framework for spatial transcriptomics. bioRxiv. 2022;2022.01.14.476408.
38. Wang Y, Song B, Wang S, Chen M, Xie Y, Xiao G, et al. Sprod for denoising spatially resolved transcriptomics data based on position and image information. Nat Methods. 2022;19:950–8.
39. Chen T, Kornblith S, Norouzi M, Hinton G. A simple framework for contrastive learning of visual representations. In: International conference on machine learning. Virtual, PMLR; 2020. p. 1597–607.
40. He K, Zhang X, Ren S, Sun J. Deep residual learning for image recognition. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Las Vegas: IEEE; 2016. p. 770–8.
41. Palla G, Spitzer H, Klein M, Fischer D, Schaar AC, Kuemmerle LB, et al. Squidpy: a scalable framework for spatial omics analysis. Nat Methods. 2022;19:171–8.
42. Satija R, Farrell JA, Gennert D, Schier AF, Regev A. Spatial reconstruction of single-cell gene expression data. Nat Biotechnol. 2015;33:495–502.
43. Tang J, Deng C, Huang GB. Extreme learning machine for multilayer perceptron. IEEE Trans Neural Netw Learn Syst. 2015;27:809–21.
44. Kipf TN, Welling M. Semi-supervised classification with graph convolutional networks. 2016. arXiv preprint arXiv:160902907.
45. Veličković P, Fedus W, Hamilton WL, Lió P, Bengio Y, Hjelm RD. Deep Graph Infomax. In: International Conference on Learning Representations. 2019. https://openreview.net/forum?id=rklz9iAcKQ. Accessed 12 Mar 2025.
46. Jiang Q, Chen C, Zhao H, Chen L, Ping Q, Tran SD, et al. Understanding and constructing latent modality structures in multi-modal representation learning. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Vancouver: IEEE; 2023. p. 7661–71.
47. Scrucca L, Fop M, Murphy TB, Raftery AE. mclust 5: clustering, classification and density estimation using Gaussian finite mixture models. R J. 2016;8:289–317.
48. Levine JH, Simonds EF, Bendall SC, Davis KL, El-ad DA, Tadmor MD, et al. Data-driven phenotypic dissection of AML reveals progenitor-like cells that correlate with prognosis. Cell. 2015;162:184–97.
49. Wu T, Hu E, Xu S, Chen M, Guo P, Dai Z, et al. clusterProfiler 4.0: A universal enrichment tool for interpreting omics data. Innovation. 2021;2:100141.
50. Hubert L, Arabie P. Comparing partitions. J Classif. 1985;2:193–218.
51. Estévez PA, Tesmer M, Perez CA, Zurada JM. Normalized mutual information feature selection. IEEE Trans Neural Netw. 2009;20:189–201.
52. Maynard KR, Collado-Torres L, Weber LM, Uytingco C, Barry BK, Williams SR, et al. Transcriptome-scale spatial gene expression in the human dorsolateral prefrontal cortex. Nat Neurosci. 2021;24:425–36.
53. 10x Genomics. 10x Genomics Datasets. 2025. https://www.10xgenomics.com/resources/datasets. Accessed 26 Feb 2025.
54. Codeluppi S, Borm LE, Zeisel A, La Manno G, van Lunteren JA, Svensson CI, et al. Spatial organization of the somatosensory cortex revealed by osmFISH. Nat Methods. 2018;15:932–5.
55. Fawkner-Corbett D, Antanaviciute A, Parikh K, Jagielowicz M, Gerós AS, Gupta T, et al. Spatiotemporal analysis of human intestinal development at single-cell resolution. Cell. 2021;184:810–26.
56. 10x Genomics. Human Breast Cancer Block A, Section 1 Dataset. 2025. https://www.10xgenomics.com/datasets/human-breast-cancer-block-a-section-1-1-standard-1-1-0. Accessed 26 Feb 2025.
57. Moncada R, Barkley D, Wagner F, Chiodin M, Devlin JC, Baron M, et al. Integrating microarray-based spatial transcriptomics and single-cell RNA-seq reveals tissue architecture in pancreatic ductal adenocarcinomas. Nat Biotechnol. 2020;38:333–42.

58. Berglund E, Maaskola J, Schultz N, Friedrich S, Marklund M, Bergenstråhle J, et al. Spatial maps of prostate cancer transcriptomes reveal an unexplored landscape of heterogeneity. Nat Commun. 2018;9:2419.
59. Andersson A, Larsson L, Stenbeck L, Salmén F, Ehinger A, Wu SZ, et al. Spatial deconvolution of HER2-positive breast cancer delineates tumor-associated cell type interactions. Nat Commun. 2021;12:6012.
60. He Y, Tang X, Huang J, Ren J, Zhou H, Chen K, et al. ClusterMap for multi-scale clustering analysis of spatial gene expression. Nat Commun. 2021;12:5909.
61. Sampath Kumar A, Tian L, Bolondi A, Hernández AA, Stickels R, Kretzmer H, et al. Spatiotemporal transcriptomic maps of whole mouse embryos at the onset of organogenesis. Nat Genet. 2023;55:1176–85.
62. Thompson B. Canonical correlation analysis: Uses and interpretation. Thousand Oaks: Sage Publications; 1989.
63. Zhang C, Geng Y, Han Z, Liu Y, Fu H, Hu Q. Autoencoder in autoencoder networks. IEEE Trans Neural Netw Learn Syst. 2022;35:2263–75.
64. Van der Maaten L, Hinton G. Visualizing data using t-SNE. J Mach Learn Res. 2008;9:2579–605.
65. Williams CG, Lee HJ, Asatsuma T, Vento-Tormo R, Haque A. An introduction to spatial transcriptomics for biomedical research. Genome Med. 2022;14:1–18.
66. Saelens W, Cannoodt R, Todorov H, Saeys Y. A comparison of single-cell trajectory inference methods. Nat Biotechnol. 2019;37:547–54.
67. Wolf FA, Hamey FK, Plass M, Solana J, Dahlin JS, Göttgens B, et al. PAGA: graph abstraction reconciles clustering with trajectory inference through a topology preserving map of single cells. Genome Biol. 2019;20:1–9.
68. Ramos RN, Missolo-Koussou Y, Gerber-Ferder Y, Bromley CP, Bugatti M, Núñez NG, et al. Tissue-resident FOLR2+ macrophages associate with CD8+ T cell infiltration in human breast cancer. Cell. 2022;185:1189–207.
69. Wu L, Yan J, Bai Y, Chen F, Zou X, Xu J, et al. An invasive zone in human liver cancer identified by Stereo-seq promotes hepatocyte-tumor cell crosstalk, local immunosuppression and tumor progression. Cell Res. 2023;33:585–603.
70. Park S, Shimizu C, Shimoyama T, Takeda M, Ando M, Kohno T, et al. Gene expression profiling of ATP-binding cassette (ABC) transporters as a predictor of the pathologic response to neoadjuvant chemotherapy in breast cancer patients. Breast Cancer Res Treat. 2006;99:9–17.
71. Buache E, Etique N, Alpy F, Stoll I, Muckensturm M, Reina-San-Martin B, et al. Deficiency in trefoil factor 1 (TFF1) increases tumorigenicity of human breast cancer cells and mammary tumor development in TFF1-knockout mice. Oncogene. 2011;30:3261–73.
72. Whiteman HJ, Weeks ME, Dowen SE, Barry S, Timms JF, Lemoine NR, et al. The role of S100P in the invasion of pancreatic cancer cells is mediated through cytoskeletal changes and regulation of cathepsin D. Cancer Res. 2007;67:8633–42.
73. Cao J, Ramachandran V, Arumugam T, Nast F, Li Z, Logsdon CD. 475q Tm4sf1 is Highly Expressed in Pancreatic Cancer and Promotes Cancer Cell Migration, Invasion and Survival. Gastroenterology. 2010;138:S–66.
74. Hong S, Beja-Glasser VF, Nfonoyim BM, Frouin A, Li S, Ramakrishnan S, et al. Complement and microglia mediate early synapse loss in Alzheimer mouse models. Science. 2016;352:712–6.
75. Li WV, Li JJ. An accurate and robust imputation method scImpute for single-cell RNA-seq data. Nat Commun. 2018;9:997.
76. Huang M, Wang J, Torre E, Dueck H, Shaffer S, Bonasio R, et al. SAVER: gene expression recovery for single-cell RNA sequencing. Nat Methods. 2018;15:539–42.
77. Wang L, Maletic-Savatic M, Liu Z. Region-specific denoising identifies spatial co-expression patterns and intra-tissue heterogeneity in spatially resolved transcriptomics data. Nat Commun. 2022;13:6912.
78. Zhang C, Dong K, Aihara K, Chen L, Zhang S. STAMarker: determining spatial domain-specific variable genes with saliency maps in deep learning. Nucleic Acids Res. 2023;51:e103.
79. Vorontsov E, Bozkurt A, Casson A, Shaikovski G, Zelechowski M, Severson K, et al. A foundation model for clinical-grade computational pathology and rare cancers detection. Nat Med. 2024;30:2924–35.
80. Chen RJ, Ding T, Lu MY, Williamson DF, Jaume G, Song AH, et al. Towards a general-purpose foundation model for computational pathology. Nat Med. 2024;30:850–62.
81. Xu H, Usuyama N, Bagga J, Zhang S, Rao R, Naumann T, et al. A whole-slide foundation model for digital pathology from real-world data. Nature. 2024;630:181–8.
82. Yu W, Zai L, Xiao M. MuCST: restoring and integrating heterogeneous morphology images and spatial transcriptomics data with contrastive learning. Zendo. 2024. https://doi.org/10.5281/zenodo.10627683.
83. Yu W, Zai L, Xiao M. MuCST: restoring and integrating heterogeneous morphology images and spatial transcriptomics data with contrastive learning. Github. 2024. https://github.com/xkmaxidian/MuCST. Accessed 12 Mar 2025.

Publisher's Note