Bioinformatics, 38(5), 2022, 1344–1352
https://doi.org/10.1093/bioinformatics/btab812
Advance Access Publication Date: 2 December 2021
Original Paper

Systems biology

Detecting spatially co-expressed gene clusters with functional coherence by graph-regularized convolutional neural network

Tianci Song[1], Kathleen K. Markham[2], Zhuliu Li[1], Kristen E. Muller[3], Kathleen Greenham [2] and Rui Kuang[1,*]

[1]Department of Computer Science and Engineering, University of Minnesota Twin Cities, Minneapolis, MN 55414, USA, [2]Department of Plant and Microbial Biology, University of Minnesota Twin Cities, Minneapolis, MN 55414, USA and [3]Department of Pathology and Laboratory Medicine, Dartmouth-Hitchcock Medical Center, Lebanon, NH 03756, USA

*To whom correspondence should be addressed.

Associate Editor: Pier Luigi Martelli

Received on July 4, 2021; revised on October 29, 2021; editorial decision on November 21, 2021; accepted on November 29, 2021

Abstract

Motivation: Clustering spatial-resolved gene expression is an essential analysis to reveal gene activities in the underlying morphological context by their functional roles. However, conventional clustering analysis does not consider gene expression co-localizations in tissue for detecting spatial expression patterns or functional relationships among the genes for biological interpretation in the spatial context. In this article, we present a convolutional neural network (CNN) regularized by the graph of protein–protein interaction (PPI) network to cluster spatially resolved gene expression. This method improves the coherence of spatial patterns and provides biological interpretation of the gene clusters in the spatial context by exploiting the spatial localization by convolution and gene functional relationships by graph-Laplacian regularization.

Results: In this study, we tested clustering the spatially variable genes or all expressed genes in the transcriptome in 22 Visium spatial transcriptomics datasets of different tissue sections publicly available from 10× Genomics and spatialLIBD. The results demonstrate that the PPI-regularized CNN constantly detects gene clusters with coherent spatial patterns and significantly enriched by gene functions with the state-of-the-art performance. Additional case studies on mouse kidney tissue and human breast cancer tissue suggest that the PPI-regularized CNN also detects spatially co-expressed genes to define the corresponding morphological context in the tissue with valuable insights.

Availability and implementation: Source code is available at https://github.com/kuanglab/CNN-PReg.

Contact: kuang@umn.edu

Supplementary information: Supplementary data are available at *Bioinformatics* online.

1 Introduction

Dissecting the heterogeneity and spatial organization of the transcriptome is key to understand how distinct cell types contribute to morphological and phenotypic diversity. For example, embryonic stem cell differentiation is spatially regulated during development to ensure proper organ formation (Combs and Eisen, 2013; Glen *et al.*, 2018). Sub-populations of cancer cells typically show distinct gene expression patterns and structural features, and co-evolve in tumor micro-environments to promote tumor development and progression (Meacham and Morrison, 2013; Schmidt and Efferth, 2016). Earlier methods based on Fluorescent In Situ Hybridization use fluorescently labeled probes hybridized to targeted RNA transcripts to measure and visualize gene expression at subcellular resolution (Raj *et al.*, 2008). More recent high-throughput technologies have emerged to unbiasedly profile spatially resolved gene expression at transcriptome-wide scale by retaining spatial localization information in intact tissue or cell culture. In situ capturing uses arrayed probes with positional barcodes covering a tissue or cell culture of interest followed by RNA sequencing. These technologies include spatial transcriptomics (ST) (Ståhl *et al.*, 2016), Visium (10× Genomics, 2019), Slide-seq (Rodriques *et al.*, 2019) and high-definition spatial transcriptomics (Vickovic *et al.*, 2019).

These new platforms for spatial genomics have revolutionized the studies of transcriptome landscape by providing spatially localized gene expression within a tissue micro-environment. The analyses of these datasets have focused primarily on clustering spatial locations to detect sub-regions in the tissue or detecting genes exhibiting spatial patterns (Bergenstråhle *et al.*, 2020; Dries *et al.*, 2019; Edsgärd *et al.*, 2018; Sun *et al.*, 2020; Svensson *et al.*, 2018). Since each individual spatially variable gene often plays multiple functional roles, these methods still rely on detecting spatial gene co-expression modules for reliable and specific functional analysis of their spatial characteristics. However, less attention has been paid on improving clustering of spatially co-expressed genes in the underlying tissue-specific micro-environment. For example, to group spatially variable genes with co-expressed patterns, STUtility (Bergenstråhle *et al.*, 2020) uses Non-negative Matrix Factorization, whereas SPARK (Sun *et al.*, 2020) and Giotto (Dries *et al.*, 2019) apply hierarchical clustering to detect co-expressed patterns. These conventional clustering methods have two limitations. First, these methods do not consider gene expression co-localization within the tissue when detecting spatial expression patterns. Ideally, the genes should be clustered based on similar overall spatial patterns by the tissue composition rather than their expression at each individual spot independently. Second, these methods do not incorporate functional relationships among the expressed genes for better biological interpretation. Moreover, the in situ RNA capture technology has a very low capture efficiency (5–10%) compared with up to 40% capture for single-cell RNAseq (Asp *et al.*, 2020). Thus, the prior knowledge on gene functional relationships has more potential to improve clustering the extremely sparse gene expression in the spatial transcriptomes.

To address these two limitations, we present a graph-regularized convolutional neural network (CNN) incorporating a protein–protein interaction (PPI) network to cluster spatially resolved gene expression (Fig. 1). The graph-regularized CNN models the expression of a gene over spatial locations as an image of a gene activity map, and naturally utilizes the spatial localization information by performing convolution operation to capture the nearby tissue textures. The model further exploits prior knowledge of gene relationships encoded in a PPI network as a regularization by graph Laplacian of the network to enhance biological interpretation of the detected gene clusters. PPI networks are curated knowledge bases containing experimental interactions, genetic interactions and predicted interactions based on complementary information to provide crucial information for deciphering gene functional relationships. It has been shown that neighboring genes in the PPI network are more likely to not only share similar biological functions but also common expression patterns (Li *et al.*, 2021; Zhang *et al.*, 2017), and thus we reasoned that the inclusion of PPI information would improve both the consistency of spatially co-expressed patterns and the significance of enriched biological functions over gene clusters. In this study, we comprehensively compared our graph-regularized CNN model with widely used clustering methods for clustering spatial gene expression across 22 ST datasets. Specifically, we tested clustering on spatially variable genes or all expressed genes in 22 Visium ST datasets of different tissue sections publicly available from 10× Genomics and spatialLIBD (Maynard *et al.*, 2021). We conducted functional analysis of spatial gene clusters on mouse kidney tissue and human breast cancer tissue, and our model resulted in clusters with significant enrichment for gene ontology processes that are consistent with the spatial morphology of the tissues.

2 Materials and methods

In this section, we introduce the end-to-end graph-regularized CNN model and two optimization algorithms for training the model. The mathematical notations used to describe the model are summarized in Supplementary Table S1.

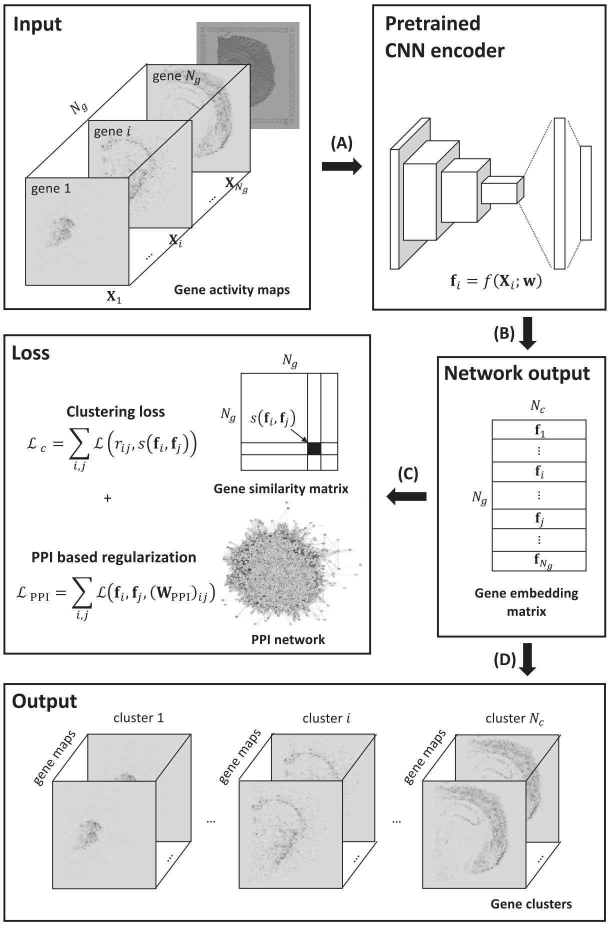


Fig. 1. The workflow of graph-regularized CNN for spatial gene expression clustering. (**A**) Feed gene expression into CNN with pretrained weights on MNIST, where gene expression is modeled as 2*D* gene activity map in the spatial coordinates. (**B**) Obtain gene embeddings from CNN encoder. (**C**) Construct the clustering loss with gene similarity matrix and PPI graph regularization based in gene embeddings to self-train the CNN encoder. (**D**) Infer gene cluster memberships based on gene embeddings once training with the combined loss is completed

2.1 Graph-regularized CNN model for gene clustering

Given ST data $\mathcal{X} = \{\mathbf{X}_i\}_{i=1}^{N_g}$, where the matrix $\mathbf{X}_i \in \mathbb{R}^{N_x \times N_y}$ denotes the activity map for gene i, N_g is the total number of genes in $\mathcal{G}$, N_x and N_y are the number of capturing spots along x-axis and y-axis. Given an undirected graph $G_{\text{PPI}} = (V, E)$ defined by PPI network, where each node $v_i \in V$ represents i-th gene in $\mathcal{G}$, and each edge $(v_i, v_j) \in E$ represents the interaction between the protein products of gene i and gene j. Let $\mathbf{W}_{\text{PPI}} \in \{0,1\}^{N_g \times N_g}$ denote the adjacency matrix of G with $(\mathbf{W}_{\text{PPI}})_{ij}$ specifying each edge (v_i, v_j).

A CNN encoder is defined to learn the embedded representation of each gene based on convolution among the capturing spots in the gene activity map as follows,

$$\mathbf{f}_i = f(\mathbf{X}_i; \mathbf{w}), \tag{1}$$

where the non-linear mapping function f is parameterized by $\mathbf{w}$ for CNN encoder on $\mathcal{X}$, and $\mathbf{f}_i$ represents the embedding of $\mathbf{X}_i$ for gene i. Let $\mathbf{F} \in \mathbb{R}^{N_g \times N_c}$ denote the gene embedding matrix with each row $\mathbf{f}_i = f(\mathbf{X}_i; \mathbf{w}) \in \mathbb{R}^{N_c}$, where N_c denotes the number of clusters, and $\mathbf{R} \in \mathbb{R}^{N_g \times N_g}$ denotes the cluster membership agreement matrix with each entry r_{ij} indicating cluster membership agreement ($r_{ij} = 1$ means genes i and j are in the same cluster and otherwise $r_{ij} = 0$). The total loss of graph-regularized CNN model combining clustering loss and graph regularization is defined as,

$$\mathcal{L}_{\text{total}} = \mathcal{L}_c(\mathbf{F}, \mathbf{R}) + \alpha \mathcal{L}_{\text{PPI}}(\mathbf{F}, \mathbf{W}_{\text{PPI}}), \tag{2}$$

where $\mathcal{L}_c$ is the cross-entropy-based clustering loss, $\mathcal{L}_{\text{PPI}}$ is the PPI-based graph regularization and $\alpha \in \mathbb{R}^+$ is a hyper-parameter modulating the impact of $\mathcal{L}_{\text{PPI}}$. $\mathcal{L}_c$ and $\mathcal{L}_{\text{PPI}}$ are defined in the following sections.

2.2 Cross-entropy loss for gene clustering

The cross-entropy-based clustering loss $\mathcal{L}_c(\mathbf{F}, \mathbf{R})$ was introduced by Chang (2017) as follows:

$$\begin{aligned}
&\min\nolimits_{\mathbf{w},l,u} \mathcal{L}_c = \min\nolimits_{\mathbf{w},l,u} \textstyle\sum_{i,j} 1_{\{0,1\}}(r_{ij} \in \{0,1\}) \mathcal{L}(s(\mathbf{f}_i, \mathbf{f}_j), r_{ij}) + g(l,u) \\
&\text{s.t.}\, l \le u,\ ||\mathbf{f}_i||_2^2 = 1,\ \mathbf{f}_i \ge 0,\ \forall i \\
&r_{ij} = \begin{cases} 0, \text{if}\, s(\mathbf{f}_i, \mathbf{f}_j) < l \\ 1, \text{if}\, s(\mathbf{f}_i, \mathbf{f}_j) \ge u \\ \text{None, otherwise,} \end{cases}
\end{aligned} \tag{3}$$

where $\mathcal{L}(s(\mathbf{f}_i, \mathbf{f}_j), r_{ij}) = -r_{ij} \log(s(\mathbf{f}_i, \mathbf{f}_j)) - (1 - r_{ij}) \log(1 - s(\mathbf{f}_i, \mathbf{f}_j))$ is the cross-entropy loss, which measures the divergence between embedding similarity and cluster membership agreement r_{ij} for gene pair (i, j), similar to binary classification. $s(\mathbf{f}_i, \mathbf{f}_j) = \mathbf{f}_i \cdot \mathbf{f}_j$ denotes the cosine function for measuring embedding similarity. $1_{\{0,1\}}(.)$ is an indicator function which returns 1 for true and 0 for false for gene pair selection in the clustering loss.

Since r_{ij} is totally unknown, a curriculum learning strategy (Bengio *et al.*, 2009) is used to estimate r_{ij}s by using embedding similarities between gene pair (i, j) iteratively during the clustering. Specifically, $u, l \in \mathbb{R}$ $(u \le l)$ are thresholds to select similar and dissimilar gene pairs. Only, gene pair (i, j) with relatively high $(r_{ij} = 1)$ or low similarity $(r_{ij} = 0)$ are involved in computing the clustering loss during each iteration. The in-between pairs with $r_{ij} =$ None are ignored. At the early stage of clustering, we assume the gene embeddings extracted from the CNN encoder with randomly initialized $\mathbf{w}$ are coarse-grained with less confidence, and thus, thresholds u and l are more strict to select only very few gene pairs with reliable r_{ij}s to construct the clustering loss. As clustering proceeds, the gene embeddings are fine-grained and more informative, and thus u and l are relaxed to involve more gene pairs for training. $g(l, u) \ge 0$ measures the gap between thresholds l and u, shrinking the gap $g(l, u)$ is equivalent to relaxing l and u simultaneously.

2.3 Graph-regularization with PPI network

To incorporate the prior knowledge of gene relationships in the PPI network, PPI-based graph regularization $\mathcal{L}_{\text{PPI}}$ is introduced on gene embeddings by the CNN encoder to encourage the consistency among the embeddings of neighboring genes in the PPI graph as follows,

$$\mathcal{L}_{\text{PPI}}(\mathbf{F}, \mathbf{W}_{\text{PPI}}) = \sum_{i,j=1}^{N_g} (\mathbf{W}_{\text{PPI}})_{ij} \left|\left| \frac{\mathbf{f}_i}{\sqrt{d_i}} - \frac{\mathbf{f}_j}{\sqrt{d_j}} \right|\right|_2^2 \tag{4a}$$

$$= ||\mathbf{F}^T \tilde{\mathbf{L}}_{\text{PPI}} \mathbf{F}||_F^2, \tag{4b}$$

where $\mathbf{D}_{\text{PPI}} = \text{diag}(d_1, \ldots, d_{N_g}) \in \mathbb{R}_{\ge 0}^{N_g \times N_g}$ denotes the node degree matrix of G_{PPI} with $d_i = (\mathbf{D}_{\text{PPI}})_{ii} = \sum_j (\mathbf{W}_{\text{PPI}})_{ij}$, and $\tilde{\mathbf{L}}_{\text{PPI}} = \mathbf{I} - (\mathbf{D}_{\text{PPI}})^{-\frac{1}{2}} \mathbf{W}_{\text{PPI}} (\mathbf{D}_{\text{PPI}})^{-\frac{1}{2}}$ denotes the normalized graph Laplacian of G_{PPI}, where the normalization corrects the bias introduced by node degree differences in learning the embedding. In this loss, the squared difference between the embeddings of a pair of connected genes $(\mathbf{f}_i, \mathbf{f}_j)$ are penalized such that the connected gene pairs have more similar embeddings than the unconnected gene pairs. The graph Laplacian loss is widely used to introduce PPI information for various learning problems in genomics and systems biology (Zhang *et al.*, 2017).

2.4 Mini-batch training algorithm and its fast approximation

To minimize $\mathcal{L}_{\text{total}}$ in eq. (2), we apply an alternating optimizing approach to minimizes $\mathcal{L}_{\text{total}}$ with respect to $\mathbf{w}$ fixing thresholds l and u and vice versa. When thresholds u and l are fixed, the term $g(l, u)$ is also fixed. Optimizing $\mathcal{L}_{\text{total}}$ can be done with regular forward and back propagation of the errors to update $\mathbf{w}$. In the alternating step, once $\mathbf{w}$ is learned and the cluster membership matrix $\mathbf{R}$ is computed, the optimization problem is simplified to finding l and u with gradient descent as follows,

$$\begin{aligned}
&\min\nolimits_{l,u} \mathcal{L}_{\text{total}} = \min\nolimits_{l,u} g(l,u)\ \text{s.t.}\, l \le u \\
&\nabla l = \eta \frac{\partial g(l,u)}{l} < 0, \nabla u = \eta \frac{\partial g(l,u)}{u} > 0,
\end{aligned} \tag{5}$$

where $\eta > 0$ is the step size. We choose the gap function $g(l, u) = u - l$ such that $\nabla l = -\eta$ and $\nabla u = \eta$, i.e. to reduce the gap by increasing l and decreasing u gradually size in each epoch until $g(l, u) \le 0$.

For efficient and scalable training, the graph-regularized CNN is trained with mini-batches of gene activity maps to compute the clustering loss $\mathcal{L}_c$ and updates $\mathbf{w}$ of the CNN encoder based on the genes in the mini-batch. But for the construction of graph regularization $\mathcal{L}_{\text{PPI}}$, we use different strategies depending on the size of the PPI network.

2.4.1 Graph regularization with entire PPI network

When the size of the PPI network is small, it is straightforward to exploit the entire PPI network to compute graph regularization $\mathcal{L}_{\text{PPI}}$ in the mini-batch training setting. Eq. 4b can be rewritten as follows:

$$\begin{aligned}
\mathcal{L}_{\text{PPI}} &= \left|\left| \begin{bmatrix} \mathbf{F}^{(B)} \\ \mathbf{F}^{(R)} \end{bmatrix}^T \begin{bmatrix} \tilde{\mathbf{L}}_{\text{PPI}}^{(BB)} & \tilde{\mathbf{L}}_{\text{PPI}}^{(BR)} \\ \tilde{\mathbf{L}}_{\text{PPI}}^{(RB)} & \tilde{\mathbf{L}}_{\text{PPI}}^{(RR)} \end{bmatrix} \begin{bmatrix} \mathbf{F}^{(B)} \\ \mathbf{F}^{(R)} \end{bmatrix} \right|\right|_F^2 \\
&= ||\mathbf{F}^{(B)^T} \tilde{\mathbf{L}}_{\text{PPI}}^{(BB)} \mathbf{F}^{(B)}||_F^2 + ||\mathbf{F}^{(R)^T} \tilde{\mathbf{L}}_{\text{PPI}}^{(RR)} \mathbf{F}^{(R)}||_F^2 \\
&+2||\mathbf{F}^{(B)^T} (\tilde{\mathbf{L}}_{\text{PPI}})^{(BR)} \mathbf{F}^{(R)}||_F^2,
\end{aligned} \tag{6}$$

where B denotes genes in the mini-batch while R denotes the genes in the complement set $\mathcal{G} \setminus \mathcal{B}$. And $\mathbf{F}^{(B)} \in \mathbb{R}^{N_b \times N_c}$ and $\mathbf{F}^{(R)} \in \mathbb{R}^{(N_g - N_b) \times N_c}$ are the embedding matrix for genes in $\mathcal{B}$ and $\mathcal{G} \setminus \mathcal{B}$, respectively. Note that, $\mathbf{F}^{(R)}$ is fixed in optimizing $\mathcal{L}_{\text{PPI}}$ and $\mathbf{F}$ is updated at the end of each iteration. $\tilde{\mathbf{L}}_{\text{PPI}}^{(BB)}$ and $\tilde{\mathbf{L}}_{\text{PPI}}^{(RR)}$, $\tilde{\mathbf{L}}_{\text{PPI}}^{(BR)}$ and $\tilde{\mathbf{L}}_{\text{PPI}}^{(RB)}$ are submatrices of graph Laplacian matrix $\tilde{\mathbf{L}}_{\text{PPI}}$, whose rows and columns correspond to the genes either in $\mathcal{B}$ or $\mathcal{G} \setminus \mathcal{B}$. The algorithm is described in Algorithm 1.

2.4.2 Approximating graph regularization with sampling

When the size of the PPI network is large for clustering a larger number of genes, computing gene embeddings $\mathbf{F}$ and graph regularization $\mathcal{L}_{\text{PPI}}$ with the entire PPI network is a heavy computational burden. Thus, we also adopted a sampling strategy to approximate $\mathcal{L}_{\text{PPI}}$ by aggregating neighbors from the PPI subnetwork consisting

Algorithm 1

```
Input: $\mathcal{X}$, $G_{\text{PPI}}$, $f$.
Parameters: $\mathbf{w}$, $l$, $u$, $N_b$, $\alpha$.
Output: $\mathbf{y}$ (cluster memberships)
Randomly initialize $\mathbf{w}$, and obtain $\mathbf{F}$ with forwarding $\mathbf{f}_i = f(\mathbf{X}_i, \mathbf{w})$.
while $l \le u$ do
    for $1, \ldots, \left\lfloor \frac{N_g}{N_b} \right\rfloor$-th batch do
        Sample $\mathcal{X}^{(B)}$ from $\mathcal{X}$.
        Select $\mathbf{F}^{(B)}$ and $\mathbf{F}^{(R)}$ from $\mathbf{F}$.
        Compute $\mathcal{L}_c$ and $\mathcal{L}_{\text{PPI}}$ according to eq. (3) and eq. (6).
        Minimize $\mathcal{L}_{\text{total}}$ to update $\mathbf{w}$.
        Update $\mathbf{F}$ with forwarding $\mathbf{f}_i = f(\mathbf{X}_i, \mathbf{w})$.
    end for
    Update $l$ and $u$ according to eq. (5).
end while
return $\mathbf{y}_i = argmax_c \mathbf{f}_i$, where $c$th entry in $\mathbf{f}_i$ is the maximum.
```

of genes within mini-batch rather than the entire PPI network, by which eq. 4a can be rewritten as

$$\mathcal{L}_{\mathrm{PPI}} = \sum_{i,j=1}^{N_b} (\mathbf{W}_{\mathrm{PPI}}^{(BB)})_{ij} || \frac{\mathbf{f}_i}{\sqrt{d_i}} - \frac{\mathbf{f}_j}{\sqrt{d_j}} ||_2^2 \tag{7}$$

where $\mathbf{W}_{\mathrm{PPI}}^{(BB)}$ is the submatrix of $\mathbf{W}_{\mathrm{PPI}}$ with rows and columns corresponding to genes in $\mathcal{B}$, and $d_i = (\mathbf{D}_{\mathrm{PPI}})_{ii} = \sum_{j=1}^{N_g} (\mathbf{W}_{\mathrm{PPI}})_{ij}$ is the same as in eq. 4a. The algorithm is described in Algorithm 2.

Algorithm 2

```
Input: $\mathcal{X}$, $G_{\mathrm{PPI}}$, $f$.
Parameters: $\mathbf{w}$, $l$, $u$, $\alpha$, $N_b$, $N_{\mathrm{iter}}$.
Output: $\mathbf{y}$ (cluster memberships)
Randomly initialize $\mathbf{w}$.
while $l \leq u$ do
    for $1, \ldots, N_{\mathrm{iter}}$-th batch do
        Sample $\mathcal{X}^{(B)}$ from $\mathcal{X}$.
        Obtain $\mathbf{F}^{(B)}$ by forwarding $\mathbf{f}_i = f(\mathbf{X}_i, \mathbf{w})$ for $i \in \mathcal{X}^{(B)}$.
        Compute $\mathcal{L}_c$ and $\mathcal{L}_{\mathrm{PPI}}$ according to eq. (3) and eq. (7).
        Minimize $\mathcal{L}_{\mathrm{total}}$ to update $\mathbf{w}$.
    end for
    Update $l$ and $u$ according to eq. (5).
end while
Obtain $\mathbf{F}$ with forwarding $\mathbf{f}_i = f(\mathbf{X}_i, \mathbf{w})$
return $\mathbf{y}_i = argmax_c \mathbf{f}_i$, where $c$th entry in $\mathbf{f}_i$ is the maximum.
```

3 Experiments

3.1 Data preparation

We downloaded 10 ST datasets from 10× Genomics containing 5 mouse brain tissues (MAB, MBSA1, MBSA2, MBSP1 and MBSP2), 1 mouse kidney tissue (MK), 2 human breast cancer tissues (HBC1 and HBC2), 1 human heart tissue (HH) and 1 human lymph node tissue (HLN) and 12 additional datasets from spatialLIBD, which are all human dorsolateral prefrontal cortex tissues (HDLPFC1-12). All datasets were generated by 10× Genomics Visium spatial protocol (v1 chemistry), which profiles tissue in a high-density hexagonal array with 4992 spots at a resolution of $55\,\mu\mathrm{m}$ ($1 - 10$ cells per spot). Each gene activity map was reshaped into a $N_x = 78$ and $N_y = 64$ matrix by shifting the spots at the odd-numbered rows by half of a spot to align all spots into a grid. Note that, this slight shift does not change the order and the spatial relationships among the spots by definition. Genes with total Unique Molecular Identifier (UMI) counts less than 100 in their activity maps were removed to eliminate technical artifacts. The UMI counts were set to 0 for capturing spots outside the tissue region.

Human and mouse PPI networks were obtained from the STRINGdb. We retained all the physical interactions and additional highly confident interactions with a confidence score larger than 0.8. We then removed genes with very few neighbors in the PPI network and intersected the ST data with pruned PPI network. Supplementary Table S2 summarizes the details of the ST data and the PPI networks used in the experiments.

3.2 Performance measures

To evaluate the clustering performance for the ST datasets, we applied several measures to evaluate gene co-expression, spatial coherence and functional enrichment of the spatial gene clusters.

Gene co-expressions: A Davies–Bouldin (DB) index (Davies and Bouldin, 1979) based on Pearson distance was used to evaluate the co-expression of gene clusters,

$$\frac{1}{N_c} \sum_{p=1}^{N_c} \max_{p \neq q} \frac{d_p + d_q}{d_{(p,q)}}, \tag{8}$$

where $d_p = \frac{2}{N_p(N_p - 1)} \sum_{i,j \in \mathcal{G}_p, i \neq j} \rho_{ij}$ denotes intra-cluster distance for gene cluster p, N_p represents the number of genes in cluster p, and $\mathcal{G}_p$ represents the gene set in cluster p. Here, $\rho_{ij} = 1-$ Pearson-Correlation $(\mathbf{x}_i, \mathbf{x}_j)$ is the Pearson distance between gene i and gene j. Similarly, $d_{(p,q)} = \rho_{pq}$ denotes inter-cluster distance between clusters p and q, where ρ_{pq} is the Pearson distance between the centroids of clusters p and q. Smaller DB indices indicate better clustering performance.

Spatial coherence: Beyond co-expression, we also defined spatial Euclidean distance (Wang *et al.*, 2005) between genes involving spatial relationships among capturing spots to replace the Pearson distance for calculating the DB index as follows,

$$\begin{aligned} \rho_{ij} &= \sqrt{\sum_{m,n=1}^{N_x N_y} (\mathbf{W}_{\mathrm{spot}})_{mn} \Big((\mathbf{x}_i)_m - (\mathbf{x}_j)_m \Big) \Big((\mathbf{x}_i)_n - (\mathbf{x}_j)_n \Big)} \\ &= \Big((\mathbf{x}_i - \mathbf{x}_j)^T \mathbf{W}_{\mathrm{spot}} (\mathbf{x}_i - \mathbf{x}_j) \Big)^{\frac{1}{2}}, \end{aligned} \tag{9}$$

where $\mathbf{W}_{\mathrm{spot}} \in \mathbb{R}^{N_x N_y \times N_x N_y}$ is the weight matrix of closeness graph G_{spot} of the capturing spots, where $(G_{\mathrm{spot}})_{mn}$ reflects the closeness between capturing spots m and n based on their distance in the slide.

Functional enrichment: We also conducted functional enrichment analysis of the genes in each cluster with 10,185 Gene Ontology (GO) terms from the C5 collection in the Molecular Signatures Database (MSigDB). *P*-values with false discovery control adjusted by Benjamini–Hochberg procedure were reported. We computed the average of the minimum log *P*-value over gene clusters to evaluate clustering performance.

3.3 Experimental design

We compared the CNN and CNN with PPI-graph regularization (CNN-PReg) against six baseline methods, including K-Means (KM), Spectral Clustering (SC), Non-negative Matrix Factorization (NMF) (Bergenstråhle *et al.*, 2020), Hierarchical Clustering (HC) (Sun *et al.*, 2020), Hierarchical Correlation Clustering (HCC) (Dries *et al.*, 2019) and Automatic Expression Histology (AEH) (Svensson *et al.*, 2018). Note that, all the baseline methods except AEH are designed to explore gene co-expression patterns without considering spatial relationships in expression, and none of them consider functional relationships among genes or spatial relationships among spots. AEH uses Gaussian models with a covariance matrix built on the distance among the capturing spots for clustering. Each clustering was repeated five times for reporting the mean and the standard error of performance measures. Moreover, two clustering settings were assessed in the experiments.

Clustering spatially variable genes: Spatially variable genes can be further clustered to show the collective spatial co-expressions for functional analysis. To evaluate the spatial patterns in the spatially variable genes, we first used trendSceek, SpatialDE, SPARK, spatial auto-correlation in STUtility, binSpect and SilhouetteRank in Giotto to detect a list of spatially variable genes in each dataset by overlapping the top 2000 genes identified by each of the six methods (shown in Supplementary Table S2). The identified spatially variable genes were clustered into 10 clusters in the experiments. CNN and CNN-PReg were benchmarked against all the baseline methods. We also performed similar experiments with 10 clusters but on fewer spatially variable genes by overlapping the top 1000 genes instead.

Clustering all expressed genes: Since there are different definitions of spatially variable genes, the list of the spatially variable genes for clustering often depends on the specific methods used for the detection. This also leads to discrepancies in the interpretation of the clustering results. Thus, it is also interesting to cluster all the expressed genes to find general spatial expression patterns beyond those in the spatially variable genes (shown in Supplementary Table S2). We clustered all the genes in each dataset into 100 or 200 clusters. CNN and CNN-PReg were compared with all the baseline methods except AEH, which does not scale to all the genes due to low computational efficiency. We used a similar generalized mixture model as a replacement of AEH for the comparison.

Generally, the gene activity maps in each dataset alone are not sufficient to train a well-performing CNN. Moreover, data augmentation of gene activity maps is not useful since the spots in the activity maps are already aligned. Thus, for CNN and CNN-PReg in all the experiments, a 9-layer CNN with pretrained weights for all convolutional layers trained on MNIST handwritten digit dataset was used. The CNN consists of six convolutional layers, followed by three fully connected layers. This specific architecture shows the best performance in clustering MNIST handwritten images. An additional convolutional layer is added at the front to project gene activity maps to the size of MNIST handwritten images. We performed grid search in the range $\{0.01, 0.05, 0.1, 0.5\}$ to find the optimal hyperparameter α on PPI graph regularization in CNN-PReg across the datasets for the same tissue type. CNN-PReg was implemented with machine learning library Tensorflow (Abadi *et al.*, 2015). Note that pretrained CNNs for MNIST handwritten images are well trained with a large number of images and supervised information. Pretrained CNNs with well-studied image databases such as MNIST and ImageNet have been used in many applications for biological and biomedical image analysis (Shen *et al.*, 2017). In this study, we found that the architecture and the pretrained weights are also helpful for learning a better CNN for the gene activity maps.

We normalized UMI counts per gene and did log transformation after adding 1, and then standardized the ST data as z-scores for projection with principal component analysis. We found these preprocessing steps generally improve the clustering performance of all the baseline methods. Euclidean distance and Ward criterion were used in both HC and HCC. The gene similarity graph in SC was constructed by 10-nearest neighbors, and the gene dissimilarity graph was built using a Pearson distance in HCC. The hyper-parameter indicating the distance scale that a gene changes expression over was set to its default value in AEH.

3.4 Improved gene co-expression and spatial coherence

The clustering performance was evaluated for gene co-expression using the DB index with Pearson distance and spatial coherence using the DB index with spatial Euclidean distance. The clustering results were compared for CNN, CNN-PReg and the baseline methods across 10 tissues from 10× Genomics (Fig. 2) and 12 tissues from spatialLIBD (Supplementary Fig. S1). In both figures, results are shown for clustering spatially variable genes into 10 clusters in panel (a) and all expressed genes into 100 clusters in panel (b).

In Figure 2, it is clear that CNN and CNN-PReg consistently achieve a better DB index with either Pearson distance or spatial Euclidean distance for clustering both spatially variable genes and all expressed genes in all 10 datasets from 10× Genomics. CNN-PReg also exhibits better DB indices than CNN in all the 10 datasets by integration of the PPI network. Similar results are also observed in the experiments on the 12 datasets from spatialLIBD shown in Supplementary Figure S1. CNN-PReg consistently improved the DB index with either Pearson distance or spatial Euclidean distance over the CNN and the other baseline methods as shown in Supplementary Figure S1a for clustering spatially variable genes and Supplementary Figure S1b for clustering all expressed genes. Additionally, we introduced four additional measures [Calinski Harabasz index (Calinski and Harabasz, 1974), Dunn index (Dunn, 1974), PBM index (Pakhira *et al.*, 2004) and Xie Beni index (Xie and Beni, 1991)] to comprehensively evaluate spatial coherence. These addition clustering indices evaluate the cluster compactness and separation with measures of a wide range of properties, which cannot be directly minimized by any clustering loss $\mathcal{L}_c$. The results shown in Supplementary Figure S2 are very similar to those by DB index-based measures. These highly consistent results support our hypotheses that (a) convolution over the spatial gene expression leads to better representation of the ST for clustering genes and (b) integration of the PPI networks improves clustering in both gene co-expression and spatial coherence.

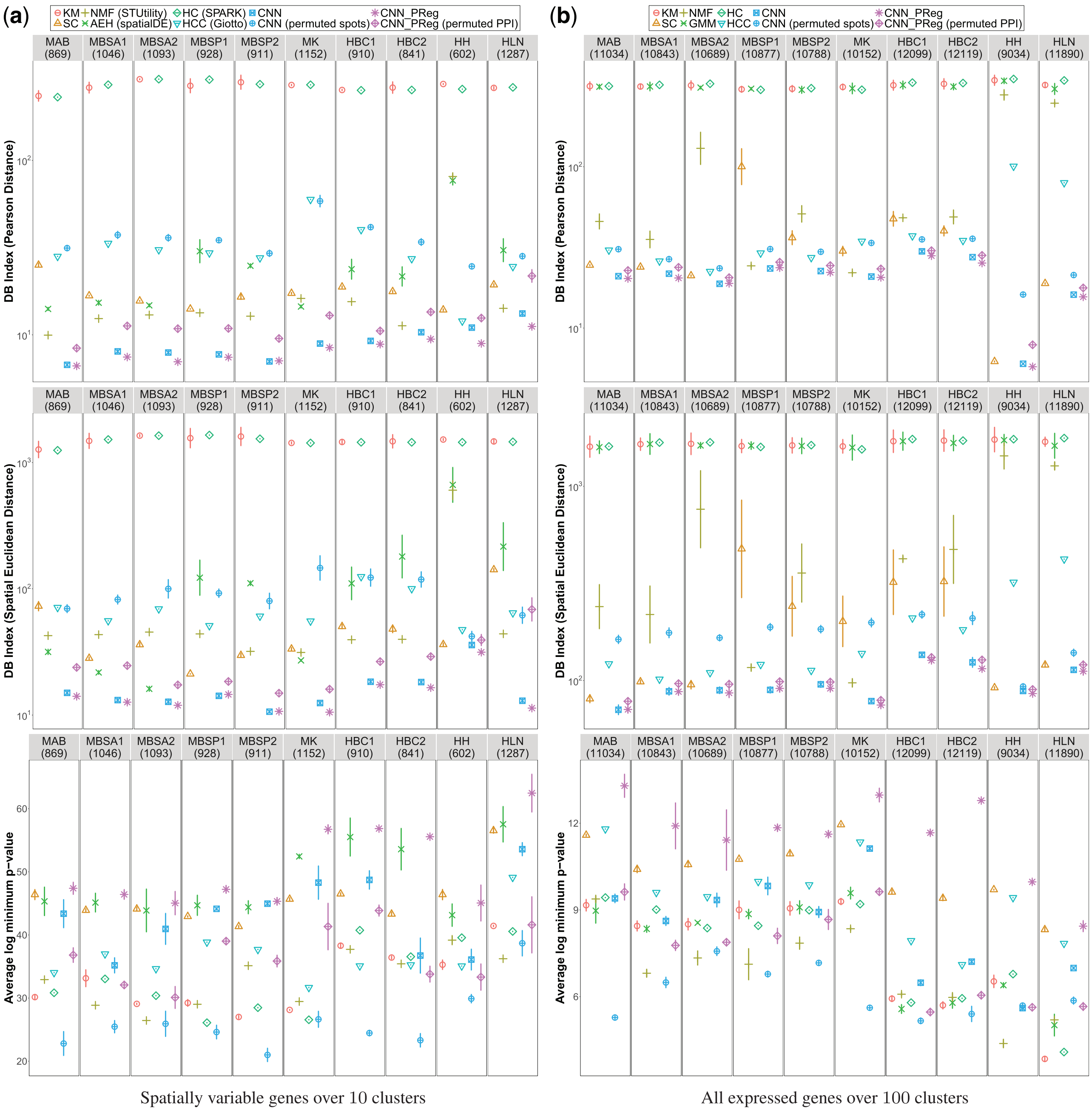
3.5 Enhanced biological interpretation of spatial patterns

The biological interpretability of the detected clusters was evaluated by the log of the minimum (most significant) P-value of each cluster in GO functional enrichment analysis. The average and standard deviation of log min P-values over the clusters of both spatially variable genes and all expressed genes are reported in the bottom panels of Figure 2 for the 10 tissues from 10× Genomics and Supplementary Figure S1 for the 12 tissues from spatialLIBD.

In Figure 2, it is evident that CNN-PReg outperforms the baseline methods regarding enrichment significance on both spatially variable genes and all expressed genes across all 10× Genomics datasets except for losing only to SC on the Human Heart (HH) dataset in clustering spatially variable genes. Interestingly, while CNN detects clusters with less significant functional enrichment than SC and several other baselines, CNN-PReg is able to significantly improve the convolution-based clustering with better functional enrichment and better gene co-expression and spatial coherence. This strongly suggests that the PPI graph regularization substantially improves the biological interpretability of the spatially co-expressed patterns in the clusters. The enrichment results on the spatially variable genes from the 12 human dorsolateral prefrontal cortex (HDLPFC) tissues from the spatialLIBD are depicted in Supplementary Figure S1a. The significance of the co-expressed clusters from the CNN-PReg is higher than the CNN and the baseline methods on 9 of the 12 datasets but lower than SC on 2 datasets (HDLPFC4 and HDLPFC5) and AEH on 1 dataset (HDLPFC8). Similarly, the enrichment results on all expressed genes for the same tissues depicted in Supplementary Figure S1b shows the significance of spatially co-expressed clusters from the CNN-PReg is higher than the CNN and the baseline methods but similar to HCC on the 3 HDLPFC datasets (HDLPFC1-3). Moreover, we further compared the number of significantly enriched clusters detected by CNN-PReg or CNN and the baselines in clustering all expressed genes into 100 clusters on the 10 datasets from 10× Genomics in Supplementary Figure S3. CNN-PReg detects about 60% of the clusters in mouse tissues and 50% in human tissues that are significantly enriched by at least one GO term (P-value ≤ 0.01), while more than half of the clusters detected by CNN and the baseline methods (except SC) in most tissues are not significantly enriched by any GO terms. At all significance levels, the CNN-PReg consistently detects more enriched clusters. Note that, even though in most of the datasets that SC enriches a similar number of significant clusters as CNN-PReg, the P-values of the most significantly enriched clusters by CNN-PReg are more significant than SC as reported by log min P-values in Figure 2. Overall, the results suggest substantial improvement in functional enrichment across all tissues when the PPI graph regularization is introduced in the CNN-PReg. Although a few baseline methods show similar or slightly better enrichment significance in a few datasets, these methods consistently show reduced spatial pattern coherence over co-expressed gene clusters.

3.6 Statistical analysis and parameter tuning

To further validate the importance of convolution and PPI graph regularization for the spatial gene clustering, we also performed two additional experiments with permutation tests, including CNN on gene activity maps with permuted capturing spots (‘CNN (permuted spots)’), where capturing spots were shuffled to distort spatial co-localization information within the gene activity map; and CNN with graph regularization on a permuted PPI network (‘CNN-PReg (permuted PPI)’), where genes with the same node degrees were shuffled to distort the functional relationships in the PPI network. Note that, the permuted PPI network is isomorphic of the original graph and still maintains the same graph structure with permuted gene identities such that the test focuses on the roles of the genes rather than the variations in graph structure. The results of the two permutation experiments are also shown in Figure 2. It is clear that



Spatially variable genes over 10 clusters

All expressed genes over 100 clusters

Fig. 2. Comparison of clustering performance on 10 ST datasets from 10× Genomics. (**a**) Comparison of clustering overlapped spatially variable genes (top 2000) into 10 clusters. (**b**) Comparison of clustering all expressed genes into 100 clusters. In the plots, the error bars are shown for mean and variances of each of the three measures, DB index with Pearson distance for gene co-expression, DB index with spatial Euclidean distance for spatial coherence and average log minimum *P*-value for gene ontology enrichment significance. Each column shows the results on one dataset with the number of genes in the clustering shown below the name of the dataset

CNN (permuted spots) and CNN-PReg (permuted PPI) show significantly inferior performance compared to the CNN and the CNN-PReg, respectively, with regard to co-expression, spatial coherence and enrichment significance for clustering both spatially variable genes and all expressed genes across allten 10× Genomics datasets. The results confirm that both spatial co-localization information captured by convolution and gene functional relationships encoded by PPI graph regularization indeed play pivotal roles in detecting better spatially co-expressed patterns and improving biological interpretation over the gene clusters.

To investigate how the weight on the PPI graph regularization affects the clustering performance, we also compared the performance of CNN-PReg with varying weights for clustering both spatially variable genes and all expressed genes across 10 datasets from 10× Genomics. The results are shown in Supplementary Figures S4 and S5. By varying the weight in the range $\{0, 0.01, 0.05, 0.1, 0.5\}$, there are clear trends that clustering performance regarding co-expression, spatial coherence and enrichment significance gradually improves to an optimum and then sharply degrades as too much confidence are put on gene functional relationships in the clustering. The weights achieving the optimal clustering performance across different tissues are quite consistent. Based on the results, 0.05 and 0.1 were selected for spatially variable genes and all expressed genes on the 10 tissue types. These observations suggest that the clustering performance of CNN-PReg indeed improves by leveraging a certain amount of information of gene functional relationships encoded in the PPI graph regularization, and the weight on the PPI graph regularization can be chosen by either the experimental evaluation of clustering to determine the optimal weights or as the same weight achieving the best clustering performance on similar types of tissues.

Moreover, we performed additional experiments by intersecting the top 1000 spatially variable genes detected using the 6 methods and clustering all expressed genes into 200 clusters. The results are similar to the previous experiments as shown in Supplementary Figure S6 for the 10 datasets from 10× Genomics and Supplementary Figure S7 for the 12 datasets from spatialLIBD. These additional experiments further show that the clustering performance regarding co-expression, spatial coherence and enrichment significance is not sensitive to either the number of spatially variable genes or the number of clusters.

Finally, to understand the role of different types of PPIs for spatial gene clustering, we also show the results of the experiments to explore clustering performance of CNN-PReg using the PPI network constructed with different confidence thresholds and PPIs from different sources including co-expression, database, experiments, genomics and text-mining on both spatially variable genes and all expressed genes across the 10 datasets from 10× genomics. The results are shown in Supplementary Figures S8 and S9. In Supplementary Figure S8, CNN-PReg with PPI constructed using confidence threshold 0.8 outperforms other thresholds regarding co-expression, spatial coherence and enrichment significance on both spatially variable genes and all expressed genes across most tissues from the 10× Genomics datasets. In Supplementary Figure S9, CNN-PReg with PPI combining different types of PPIs achieves the best clustering performance regarding co-expression, spatial coherence and enrichment significance on both spatially variable genes and all expressed genes across most tissues from 10× Genomics datasets. Interestingly, we noticed that the CNN-PReg with the PPI consisting of co-expression PPIs showed better co-expression and spatial coherence but worse enrichment significance compared to other types of PPIs on most tissues. This observation indicates co-expression PPIs could improve clustering performance regarding co-expression and spatial coherence while other types of PPIs could complement functional interpretation over clusters.

3.7 Spatial co-expression patterns reveal biological functions underlying tissue anatomical structures

To study the functional relationship between spatial gene expression patterns and tissue-specific anatomical structures, 2 case studies were conducted to match the 10 clusters in spatially variable genes with the anatomical structure in mouse kidney (MK) and human breast cancer tissue (HBC1). We manually annotated the H&E image of the HBC1 tissue section with regions of Ductal Carcinoma In Situ (DCIS), Invasive Ductal Carcinoma (IDC) and surrounding benign stroma.

The average expression of the genes in the 10 clusters are visualized with the annotated H&E image of mouse kidney in Figure 3a. The spatially co-expressed patterns primarily highlight three well-studied anatomical structures, including cortex, inner stripe of the outer medulla (ISOM) and outer stripe of the outer medulla (OSOM). The enriched functions of the clusters for the three anatomical structures are listed in Supplementary Table S3 along with references that support the association of the biological processes with these structures. We found relevant biological processes for the anatomical structures. The spatial patterns in clusters 1, 2, 3 and 4 overlapping with ISOM enriched nucleotide metabolisms, regulation of vasodilatation and renal system process (Lemley and Kriz, 1991; Ren *et al.*, 2014). The spatial patterns in clusters 5, 6, 7 and 8 overlapping with OSOM enriched catabolic processes of organic and inorganic molecules, which are active in renal proximal tubule across cortex and OSOM (Anzai *et al.*, 2005; Zalups, 1995). The spatial patterns in clusters 9 and 10 resembling cortex mainly enriched biological functions for the regulation of blood pressure, transport/homeostasis of inorganic molecules and the transport of cellular metabolites, which are related to the regulation of kidney function

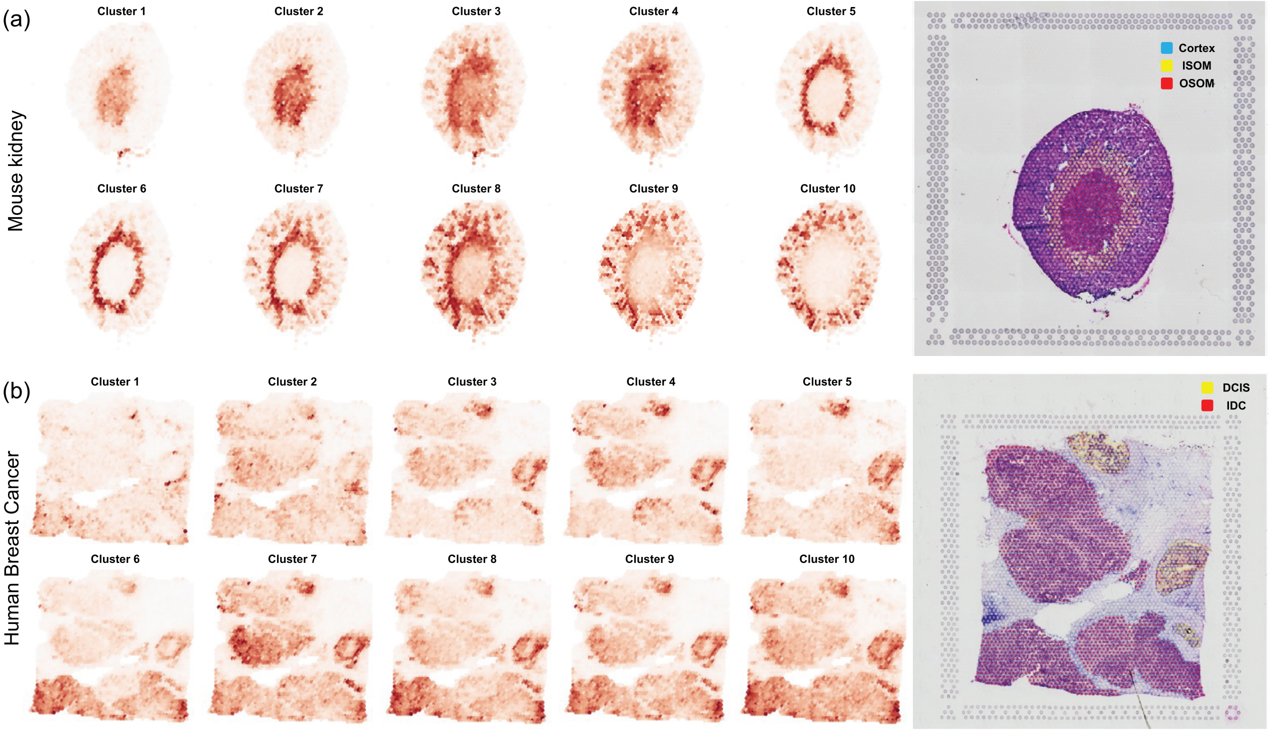


Fig. 3. Spatial gene expression patterns in mouse kidney tissue and human breast cancer tissue. (a) The annotated H&E image of mouse kidney is shown on the right with different colors denoting three anatomical structures, inner stripe of the outer medulla (ISOM), the outer stripe of the outer medulla (OSOM) and Cortex. Spatial patterns of the 10 clusters are matched to ISOM (1, 2, 3 and 4), OSOM (5, 6, 7 and 8) and Cortex (9 and 10). (b) The annotated H&E image of breast cancer tissue is shown on the right with different colors denoting two tumor types, Ductal Carcinoma In Situ (DCIS), Invasive Ductal Carcinoma (IDC) and the surrounding benign stroma is shown in background color. Spatial patterns of the 10 clusters are matched to surrounding benign stroma (1 and 2), DCIS (3, 4 and 5), IDC (6, 7) and the mixture of all 3 cell types (8, 9 and 10).

reported in previous studies (Brown and Wagner, 2012; Crowley *et al.*, 2005).

The average expressions of the genes in the 10 clusters are visualized with the annotated H&E image of human breast cancer tissue in Figure 3b. The spatially co-expressed patterns match anatomical structures of the tissues, in this case the Ductal Carcinoma In Situ (DCIS), Invasive Ductal Carcinoma (IDC) and surrounding benign stroma. The enriched functions of the clusters for the three anatomical structures are listed in Supplementary Table S3 along with supporting citations. The biological functions of the significant GO terms are relevant to benign stroma, DCIS and IDC. The spatial patterns in clusters 1 and 2 covering surrounding benign stroma are enriched in biological functions of extracellular matrix (ECM) construction, including glycosaminoglycan, proteoglycan, biosynthetic and metabolic processes and the regulation of immune cells, including lymphocytes and macrophages proliferation and activation. ECM and immune cells are two main components constituting a tumor micro-environment that closely interacts with tumor cells and further promotes tumor progression (Place *et al.*, 2011; Soysal *et al.*, 2015). The spatial expression patterns in clusters 3, 4 and 5 highlighting DCIS are enriched by biological functions for interferon signaling and chemokine signaling, where interferon signaling may contribute to tumor dissemination, and chemokine signaling promotes tumor invasion and survival (Critchley-Thorne *et al.*, 2009; Gao and Fish, 2018). The spatial patterns in clusters 6 and 7 highlighting IDC are enriched by biological functions for cholesterol and sphingolipid metabolism, which are highly connected to tumor metastasis (Ehmsen *et al.*, 2019; Nagahashi *et al.*, 2016), and aerobic glycolysis energy metabolism, which are proven to be the main energy source for tumor cells and regarded as an indication of malignancy (Liberti and Locasale, 2016; Wu *et al.*, 2020). Finally, the spatial patterns in clusters 8, 9 and 10 overlapping both DCIS and IDC are mainly enriched in biological functions for aerobic respiration energy metabolism, which are rewired to modulate tumor progression (Avagliano *et al.*, 2019; Jia *et al.*, 2018), and interleukin signaling, which are reported as identifiers of tumor proliferation and invasion (Fasoulakis *et al.*, 2018).

In addition, we also compared the enriched GO terms by AEH and CNN on the MK and HBC1 datasets by taking the most significant enrichment among all the clusters for each of the GO terms in Supplementary Tables S3 and S4. As expected, even though clusters identified by AEH and CNN could display some similar spatially co-expressed patterns, the genes in the clusters do not have or produce less significant enrichment on the same biological functions related to tissue anatomical structures (Supplementary Tables S3 and S4). Specifically, 28/36 and 30/36 GO terms are missing or less significantly enriched on MK, and 44/48 and 40/48 GO terms on HBC1 by AEH and CNN, respectively. All these observations corroborate that CNN-PReg provides better biological interpretation in understanding the functional relationship between spatial gene expression patterns and tissue anatomical structures.

4 Conclusion

In this study, we propose a CNN model coupled with PPI graph regularization to discover gene co-expression clusters with spatial patterns in ST data. To our knowledge, this is the first gene clustering method integrating both spatial and functional information for clustering spatial transcriptomes. The experimental results show that spatial relationships among captured spots and functional relationships among genes can significantly improve the coherence of spatial co-expression patterns and the biological interpretation of the clusters. Further enrichment analysis also shows that the spatial co-expression patterns in the gene clusters correspond to distinct anatomical structures within tissue suggesting more biologically informative clusters.

Deep neural networks have been shown to be useful for various challenging learning problems in the analysis of single-cell and ST data, such as removing batch effects in clustering scRNAseq data (Li *et al.*, 2020), and integrating H&E image with spatial gene expression for cell-type identification (Tan *et al.*, 2020) or correlating local gene expression with H&E image features (He *et al.*, 2020). Our current model relies on using a pre-trained neural network and integration of PPI-based graph-regularization to avoid complex model tuning with a relative small sample size. In the future, we will focus on improving the neural network architecture for these more advanced data integration approaches to be applied to detecting and interpreting spatial co-expression patterns.

Funding

This research work was supported by a grant from the National Science Foundations, USA [NSF BIO DBI-IIBR 2042159].

Conflict of Interest: none declared.

References

10× Genomics (2020) Visium spatial transcriptomics. www.10xgenomics.com/products/spatial-gene-expression (28 January 2020, date last accessed).

Abadi,M. *et al.* (2015) TensorFlow: large-scale machine learning on heterogeneous systems. tensorflow.org.

Anzai,N. *et al.* (2005) Functional characterization of rat organic anion transporter 5 (slc22a19) at the apical membrane of renal proximal tubules. *J. Pharmacol. Exp. Therap.*, **315**, 534–544.

Asp,M. *et al.* (2020) Spatially resolved transcriptomes—next generation tools for tissue exploration. *BioEssays*, **42**, 1900221.

Avagliano,A. *et al.* (2019) Mitochondrial flexibility of breast cancers: a growth advantage and a therapeutic opportunity. *Cells*, **8**, 401.

Bengio,Y. *et al.* (2009) Curriculum learning. In: *Proceedings of the 26th Annual International Conference on Machine Learning*, Montreal, Canada, pp. 41–48.

Bergenstråhle,J. *et al.* (2020) Seamless integration of image and molecular analysis for spatial transcriptomics workflows. *BMC Genomics*, **21**, 1–7.

Brown,D. and Wagner,C.A. (2012) Molecular mechanisms of acid–base sensing by the kidney. *J. Am. Soc. Nephrol.*, **23**, 774–780.

Calinski,T. and Harabasz,J. (1974) A dendrite method for cluster analysis. *Commun. Stat. Theory Methods*, **3**, 1–27.

Chang,J. (2017) Deep adaptive image clustering. In: *Proceedings of the IEEE International Conference on Computer Vision*, Venice, Italy, pp. 5879–5887.

Combs,P.A. and Eisen,M.B. (2013) Sequencing mRNA from cryo-sliced drosophila embryos to determine genome-wide spatial patterns of gene expression. *PLoS One*, **8**, e71820.

Critchley-Thorne,R.J. *et al.* (2009) Impaired interferon signaling is a common immune defect in human cancer. *Proc. Natl. Acad. Sci. USA*, **106**, 9010–9015.

Crowley,S.D. *et al.* (2005) Distinct roles for the kidney and systemic tissues in blood pressure regulation by the renin-angiotensin system. *J. Clin. Investig.*, **115**, 1092–1099.

Davies,D.L. and Bouldin,D.W. (1979) A cluster separation measure. *IEEE Trans. Pattern Anal. Mach. Intell.*, **1**, 224–227.

Dries, R. et al. (2021) Giotto: a toolbox for integrative analysis and visualization of spatial expression data. Genome biology, 22, 1–31.

Dunn,J.C. (1974) Well-separated clusters and optimal fuzzy partitions. *J. Cybern.*, **4**, 95–104.

Edsgärd,D. *et al.* (2018) Identification of spatial expression trends in single-cell gene expression data. *Nat. Methods*, **15**, 339–342.

Ehmsen,S. *et al.* (2019) Increased cholesterol biosynthesis is a key characteristic of breast cancer stem cells influencing patient outcome. *Cell Rep.*, **27**, 3927–3938.

Fasoulakis,Z. *et al.* (2018) Interleukins associated with breast cancer. *Cureus*, **10**, e3549.

Gao,D. and Fish,E.N. (2018) Chemokines in breast cancer: regulating metabolism. *Cytokine*, **109**, 57–64.

Glen,C.M. *et al.* (2018) Dynamic intercellular transport modulates the spatial patterning of differentiation during early neural commitment. *Nat. Commun.*, **9**, 1–13.

He,B. *et al.* (2020) Integrating spatial gene expression and breast tumour morphology via deep learning. *Nat. Biomed. Eng.*, **4**, 827–828.

Jia,D. *et al.* (2018) Elucidating the metabolic plasticity of cancer: mitochondrial reprogramming and hybrid metabolic states. *Cells*, **7**, 21.

Lemley,K.V. and Kriz,W. (1991) Anatomy of the renal interstitium. *Kidney Int.*, **39**, 370–381.

Li,X. *et al.* (2020) Deep learning enables accurate clustering with batch effect removal in single-cell RNA-seq analysis. *Nat. Commun.*, **11**, 1–14.

Li,Z. *et al.* (2021) Imputation of spatially-resolved transcriptomes by graph-regularized tensor completion. *PLoS Comput. Biol.*, **17**, e1008218.

Liberti,M.V. and Locasale,J.W. (2016) The Warburg effect: how does it benefit cancer cells? *Trends Biochem. Sci.*, **41**, 211–218.

Maynard,K.R. *et al.* (2021) Transcriptome-scale spatial gene expression in the human dorsolateral prefrontal cortex. *Nat. Neurosci.*, **24**, 425–436.

Meacham,C.E. and Morrison,S.J. (2013) Tumour heterogeneity and cancer cell plasticity. *Nature*, **501**, 328–337.

Nagahashi,M. *et al.* (2016) High levels of sphingolipids in human breast cancer. *J. Surg. Res.*, **204**, 435–444.

Pakhira,M.K. *et al.* (2004) Validity index for crisp and fuzzy clusters. *Pattern Recogn.*, **37**, 487–501.

Place,A.E. *et al.* (2011) The microenvironment in breast cancer progression: biology and implications for treatment. *Breast Cancer Res.*, **13**, 1–11.

Raj,A. *et al.* (2008) Imaging individual mRNA molecules using multiple singly labeled probes. *Nat. Methods*, **5**, 877–879.

Ren,H. *et al.* (2014) Spatial organization of the vascular bundle and the interbundle region: three-dimensional reconstruction at the inner stripe of the outer medulla in the mouse kidney. *Am. J. Physiol. Renal Physiol.*, **306**, F321–F326.

Rodriques,S.G. *et al.* (2019) Slide-seq: a scalable technology for measuring genome-wide expression at high spatial resolution. *Science*, **363**, 1463–1467.

Schmidt,F. and Efferth,T. (2016) Tumor heterogeneity, single-cell sequencing, and drug resistance. *Pharmaceuticals*, **9**, 33.

Shen,D. *et al.* (2017) Deep learning in medical image analysis. *Annu. Rev. Biomed. Eng.*, **19**, 221–248.

Soysal,S.D. *et al.* (2015) Role of the tumor microenvironment in breast cancer. *Pathobiology*, **82**, 142–152.

Ståhl,P.L. *et al.* (2016) Visualization and analysis of gene expression in tissue sections by spatial transcriptomics. *Science*, **353**, 78–82.

Sun,S. *et al.* (2020) Statistical analysis of spatial expression patterns for spatially resolved transcriptomic studies. *Nat. Methods*, **17**, 193–200.

Svensson,V. *et al.* (2018) Spatialde: identification of spatially variable genes. *Nat. Methods*, **15**, 343–346.

Tan,X. *et al.* (2020) Spacell: integrating tissue morphology and spatial gene expression to predict disease cells. *Bioinformatics*, **36**, 2293–2294.

Vickovic,S. *et al.* (2019) High-definition spatial transcriptomics for in situ tissue profiling. *Nat. Methods*, **16**, 987–990.

Wang,L. *et al.* (2005) On the Euclidean distance of images. *IEEE Trans. Pattern Anal. Mach. Intell.*, **27**, 1334–1339.

Wu,Z. *et al.* (2020) Emerging roles of aerobic glycolysis in breast cancer. *Clin. Transl. Oncol.*, **22**, 631–646.

Xie,X.L. and Beni,G. (1991) A validity measure for fuzzy clustering. *IEEE Trans. Pattern Anal. Mach. Intell.*, **13**, 841–847.

Zalups,R.K. (1995) Organic anion transport and action of γ-glutamyl transpeptidase in kidney linked mechanistically to renal tubular uptake of inorganic mercury. *Toxicol. Appl. Pharmacol.*, **132**, 289–298.

Zhang,W. *et al.* (2017) Network-based machine learning and graph theory algorithms for precision oncology. *NPJ Precision Oncol.*, **1**, 1–15.