Bioinformatics, 2024, **40**, ii111–ii119
https://doi.org/10.1093/bioinformatics/btae383
Proceedings of ECCB2024

Predicting spatially resolved gene expression via tissue morphology using adaptive spatial GNNs

Tianci Song[1,2], Eric Cosatto[2], Gaoyuan Wang[3,4], Rui Kuang[1], Mark Gerstein[3,4,5,6,7], Martin Renqiang Min[2], Jonathan Warrell[2,3,4,]*

[1]Department of Computer Science and Engineering, University of Minnesota, Minneapolis, MN 55455, United States
[2]Machine Learning Department, NEC Laboratories America, Princeton, NJ 08540, United States
[3]Program in Computational Biology and Bioinformatics, Yale University, New Haven, CT 06520, United States
[4]Department of Molecular Biophysics and Biochemistry, Yale University, New Haven, CT 06520, United States
[5]Department of Computer Science, Yale University, New Haven, CT 06520, USA
[6]Department of Statistics and Data Science, Yale University, New Haven, CT 06520, USA
[7]Department of Biomedical Informatics and Data Science, Yale University, New Haven, CT 06520, USA

*Corresponding author. Machine Learning Department, NEC Laboratories America, Princeton, NJ 08540, United States, and Department of Molecular Biophysics and Biochemistry, Yale University, New Haven, CT 06520, United States. E-mail: jwarrell@nec-labs.com; jonathan.warrell@yale.edu (J.W.)

Abstract

Motivation: Spatial transcriptomics technologies, which generate a spatial map of gene activity, can deepen the understanding of tissue architecture and its molecular underpinnings in health and disease. However, the high cost makes these technologies difficult to use in practice. Histological images co-registered with targeted tissues are more affordable and routinely generated in many research and clinical studies. Hence, predicting spatial gene expression from the morphological clues embedded in tissue histological images provides a scalable alternative approach to decoding tissue complexity.

Results: Here, we present a graph neural network based framework to predict the spatial expression of highly expressed genes from tissue histological images. Extensive experiments on two separate breast cancer data cohorts demonstrate that our method improves the prediction performance compared to the state-of-the-art, and that our model can be used to better delineate spatial domains of biological interest.

Availability and implementation: https://github.com/song0309/asGNN/

1 Introduction

Dissecting the cellular and spatial heterogeneity of tissues is critical to characterizing cellular composition and organization, and ultimately their contribution to phenotype variation. Unlike traditional bulk and single-cell transcriptomics, spatial transcriptomics technologies enable spatially resolved gene expression profiling within intact tissues by using imaging or sequencing methods. Unlike imaging methods, which required targeted probes for a predetermined set of genes, sequencing-based methods perform RNA sequencing of the whole transcriptome with a positionally barcoded array of spots aligned to the histological image of the tissue (Asp *et al.* 2020). However, while sequencing-based spatial transcriptomics technologies have been widely used in biomedical research, their high cost still hinders their application in clinical studies. In contrast, histological images, such as hematoxylin and eosin (H&E) or immunofluorescence (IF) staining images, which are generated by most spatial transcriptomics technologies with ISC method, can be acquired cost-efficiently at high quality. However, while these histological images are commonly used to compensate for spatial gene expression in downstream analyses (Dries *et al.* 2021), the dependencies between spatial gene expression profiles and histological images have only been explored to a limited extent; using such dependencies may alleviate the reliance on spatial transcriptomics by estimating spatial gene expression directly from tissue morphology.

As spatial transcriptomics data continue to accumulate, an increasing number of computational methods (He *et al.* 2020, Dawood *et al.* 2021, Monjo *et al.* 2022) aim to establish a connection between spatial gene expression profiles and histological images based on existing spatial transcriptomics datasets. These approaches predict the gene expression of each capturing spot with the corresponding image patch from H&E staining image. However, all of these methods fail to model spatial proximity in gene expression, which is one of the essential properties in real spatial transcriptomics data. A few attempts have been made to circumvent this issue, which either apply Transformer (Pang *et al.* 2021, Yang *et al.* 2023) or GNN (Zeng *et al.* 2022, Mejia *et al.* 2023) approaches to incorporate the relations among capturing spots when predicting spatial gene expression.

Despite the fact that transformer-based methods naturally model global relations among image patches and capturing spots by exploiting the self-attention mechanism, some methods attempt to further refine these relations by either incorporating positional information [e.g. HisToGene (Pang *et al.* 2021)] or imposing locality in the image embedding space [e.g. EGN (Yang *et al.* 2023)]; they therefore lack the capability to distinguish compartments showing similar morphological features but distinct gene expressions in the tissues, such as tumors and their microenvironment. In addition, transformer-based methods may be prone to overfitting issue due to limited training data in the existing spatial transcriptomics datasets. Conversely, while GNN-based methods emphasize local relations among image patches and capturing spots in the graph, they may not retain global relations between spatially distant regions showing both similar morphological features and gene expression, particularly in the non-well-structured tissues, such as lymph nodes and tumors. Some methods allow both local and global relations to be encoded by considering both image and positional embeddings in the graph construction [e.g. SEPAL (Mejia *et al.* 2023)], but these relations, which are hard-coded in the graph prior to model training, might not be present in gene expression.

To overcome the aforementioned issues, we propose an adaptive spatial Graph Neural Network (asGNN) for spatial gene expression prediction, which builds on the smoothing-based GNN (SBGNN) framework of (Wang *et al.* 2024). The SBGNN framework was developed to predict liquid–liquid phase separation from 3D molecular graphs, by using graph structure to adaptively refine molecular graphs to remove task irrelevant edges to help perform graph classification. Similarly, we adaptively remove edges in our spatial graph to help accurately predict gene expression. Following Wang *et al.* (2024), during training we apply smoothing-based variational optimization (VO) (Leordeanu and Hebert 2008) to search for a graph that captures both local and global relations important to a given task. In our case, these are relations among the capturing spots for a given image, and we use Graph Transformer Networks (GTN) (Shi *et al.* 2020) as the backbone to better align these relations with actual proximity in gene expression among capturing spots. Our experiments demonstrate that the spatial graphs learned from asGNN not only improve the spatial gene expression prediction compared to the state-of-the-art methods but also help to detect biologically interpretable spatial domains. Furthermore, the prototype clustering analysis on breast cancer tissues suggests that asGNN can be used to study the homogeneity and heterogeneity of spatial organization across tissue sections from patients in different conditions. The predictions from our model thus have the potential to be translated into clinical diagnostics tools to inform personalized treatment decisions.

2 Materials and methods

The overall architecture of our model is summarized in Fig. 1. Below, we provide full details of the architecture and our end-to-end optimization algorithm, which build on the smoothing-based GNN framework of Wang *et al.* (2024).

2.1 Adaptive spatial GNN architecture

We assume we have input data of the form $\mathcal{X} = \{G_i = (X_i, E_i) | i = 1, \dots, N\}$, where G_i is the image graph for the ith data point (a whole slide image), with X_i the matrix of node features for data point i, and E_i the edge set for the spatial connectivity of graph G_i (in our images, the spots are positioned in a regular grid, and we use an 8-connected neighborhood). X_i has dimensions $N_i \times D_X$, where N_i is the number of nodes (spots) in the image graph G_i, and D_X is the dimensionality of the image features. Our task is then to predict the output spatial gene expression data, $\mathcal{Y} = \{Y_i | i = 1, \dots, N\}$, where Y_i is the expression matrix for image i, whose dimensionality is $N_i \times D_Y$, where D_Y is the number of predicted genes.

2.1.1 Adaptive graph refinement

We adapt the graph refinement procedure introduced in Wang *et al.* (2024). During training, we learn to predict a matrix of latent meta-features, Z_i for each image, with dimensionality $N_i \times D_Z$, where D_Z is the meta-feature dimensionality. The latent meta-features in our model are learned as a linear transformation of the image features, i.e. $Z_i = X_i W_0$. We then define a distance function on the nodes of graph i:

$$d(n_1, n_2) = \alpha d_1(\mathbf{z}_{n_1}, \mathbf{z}_{n_2}) + \beta d_2(n_1, n_2), \tag{1}$$

where $\mathbf{z}_n$ are the meta-features associated with node n (i.e. the nth row of Z_i), d_1 is the Euclidean distance, $d_2(n_1, n_2)$ is the shortest path distance between nodes n_1 and n_2 in G_i, and $\{\alpha, \beta\}$ are hyperparameters. Our distance function here is adapted from Wang *et al.* (2024), where we exclude their final "degree consistency" term, due to the regular topology of our initial spatial graph. We use the distances in Eq. (1) to define a distance matrix D_i between each pair of nodes in graph i, and use an arbitrary clustering algorithm to map this to a vector C_i of cluster indices; here $C_i \in \{1, \dots, K_i\}^{N_i}$, where K_i is the maximum cluster index for image i, and $C_i(n) = k$ indicates that node n belongs to cluster k. The clusters $\{1, \dots, K_i\}$ represent potentially meaningful spatial domains in image i (e.g. tumor microenvironment regions). We choose Affinity Propagation (AP) as our clustering algorithm (Frey and Dueck 2007), hence, $C_i = \text{AP}(D_i)$, where AP(.) denotes the application of the AP algorithm. Finally, using the learned cluster vectors, we form the refined spatial graphs for each training instance, $G_i' = \{X_i, E_i'\}$, where $E_i' = \{(n,m) \in E | C_n = C_m\}$, hence restricting the graph so that information is shared *via* message passing only within spatial domains (clusters).

2.1.2 Predicting spatial gene expression

We use the refined graphs to predict the spatial gene expression matrices as output; we thus adapt the framework of Wang *et al.* (2024) (designed for graph classification tasks) to perform multivariate graph regression. Our network outputs the predicted matrix by performing message passing on the refined graphs, G_i'. The network is parameterized by weight matrices $W_{1\dots L}$, where L is the number of layers in the GNN, with W_l having dimensionality $D_{l-1} \times D_l$, such that D_l is the number of hidden units per node in layer l, and $D_0 = D_X$, $D_L = D_Y$. We treat $\{L, D_{1\dots L-1}\}$ as additional hyperparameters. The message-passing updates can be written as:

$$\mathbf{x}_n^l = \sigma\left(\sum_{\{m | (m,n) \in E'\}} \frac{\mathbf{x}_m^{l-1} W_l}{\sqrt{\deg(n)\deg(m)}} \right), \tag{2}$$

for levels $l < L$, where $\sigma(x) = \max(0, x)$ is the RELU function, and $\deg(n)$ is the degree of node n. For level L, a final linear is used, which is applied to each node independently; hence: $\mathbf{x}_n^L = \mathbf{x}_n^{L-1} W_L$.

For our training loss, we use the function:

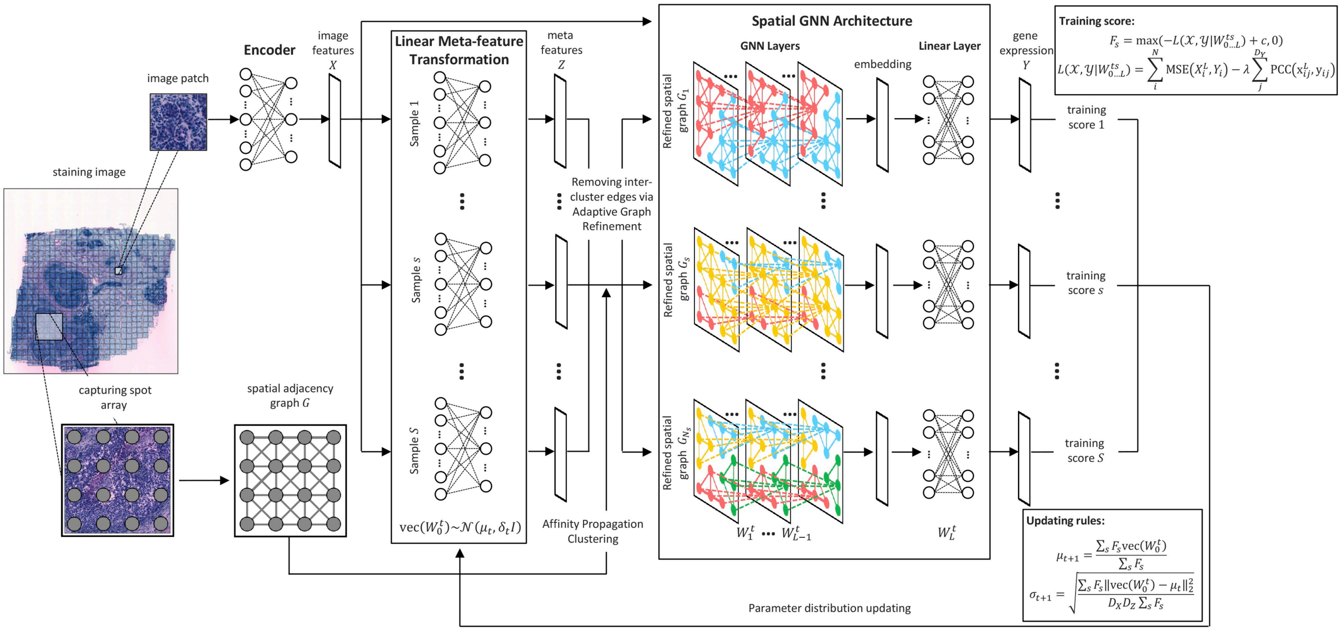
Encoder
image features X
image patch
staining image
capturing spot array
spatial adjacency graph G
Linear Meta-feature Transformation
Sample 1
Sample s
Sample S
$\mathrm{vec}(W_0^t) \sim \mathcal{N}(\mu_t, \delta_t I)$
meta features Z
Removing inter-cluster edges via Adaptive Graph Refinement
Affinity Propagation Clustering
Spatial GNN Architecture
GNN Layers
Linear Layer
embedding
Refined spatial graph G_1
Refined spatial graph G_s
Refined spatial graph G_{N_s}
W_1^t ••• W_{L-1}^t
W_L^t
gene expression Y
Training score:
$F_s = \max(-L(\mathcal{X}, \mathcal{Y}|W_{0\ldots L}^{ts}) + c, 0)$
$L(\mathcal{X}, \mathcal{Y}|W_{0\ldots L}^{ts}) = \sum_i^N \mathrm{MSE}(X_i^L, Y_i) - \lambda \sum_j^{D_Y} \mathrm{PCC}(\mathrm{x}_{ij}^L, \mathrm{y}_{ij})$
training score 1
training score s
training score S
Updating rules:
$\mu_{t+1} = \frac{\sum_s F_s \mathrm{vec}(W_0^t)}{\sum_s F_s}$
$\sigma_{t+1} = \sqrt{\frac{\sum_s F_s \|\mathrm{vec}(W_0^t) - \mu_t\|_2^2}{D_X D_Z \sum_s F_s}}$
Parameter distribution updating

Figure 1. Adaptive spatial graph neural network (asGNN) architecture. We summarize the key components of our asGNN architecture. asGNN begins by extracting image features from image patches over capturing spots arranged in an 8-connected spatial graph with an encoder. Adaptive graph refinement is then applied, as introduced in Wang *et al.* (2024); for each meta-epoch, asGNN samples N_s sets of parameters for a linear meta-feature transformation, that projects the image features to a set of meta-features for each spot, and graph refinement is performed by applying Affinity Propagation (AP) clustering to the meta-features to sparsify the input graph. We use a GTN as the backbone model in asGNN, and train the ensemble of GTNs with image features on the sparsified spatial graphs to predict the gene expression separately. Lastly, we average the training scores over all the samples using the score function shown, which combines the mean squared error (MSE) and per-gene Pearson correlation coefficient (PCC) of the predicted gene expression values, and apply variational updates to parameters of the linear meta-feature transformation layer

$$
\begin{aligned}
L(\mathcal{X}, \mathcal{Y}|W_{0\ldots L}) &= \textstyle\sum_i L_i(G_i, Y_i|W_{0\ldots L}) \\
L_i(G_i, Y_i|W_{0\ldots L}) &= \mathrm{MSE}(X_i^L, Y_i) - \lambda \textstyle\sum_{j=1\ldots D_Y} \mathrm{PCC}(\mathbf{x}_{ij}^L, \mathbf{y}_{ij}).
\end{aligned} \tag{3}
$$

Here, $\mathrm{MSE}(X, Y)$ is the mean squared error between matrices X and Y (summed across all elements), $\mathrm{PCC}(\mathbf{x}, \mathbf{y})$ is the Pearson correlation coefficient (PCC) between vectors $\mathbf{x}$ and $\mathbf{y}$ (each being a vector of expression values across the nodes of the final layer), and λ is a tradeoff parameter (which we set to 0 when considering the MSE loss only).

2.2 End-to-end training

Since, in our architecture, the projection W_0 determines the meta-feature matrix Z, which in turn determines the graph structure of the adapted (refined) spatial graph G' used for message passing, we have a complex interaction between a discrete optimization over the space of refined spatial graphs (implicitly parameterized by W_0) and the continuous predictions of the network (determined by $W_{1\ldots L}$). The underlying objective (Eq. (3)) is therefore discontinuous at points where changing W_0 changes the refined graphs; however, if W_0 is held constant, the objective is continuous over the remaining parameters, and can be handled by gradient descent.

Following Wang *et al.* (2024), we thus use a modified form of VO (Leordeanu and Hebert 2008), which allows us to convert an objective with discontinuities into a continuous objective. This is done by introducing a variational distribution over the parameters W_0, which we take to be a Gaussian with a symmetric covariance matrix. At a given meta-epoch t, this variational distribution has the form:

$$\mathrm{vec}(W_0^t) \sim \mathcal{N}(.|\mu_t, \sigma_t I), \tag{4}$$

where $\mathrm{vec}(W_0)$ is the vectorization of matrix W_0, μ_t is a vector of mean values, σ_t is a scalar, and I is the identity matrix. At meta-epoch t, we draw S samples from Eq. (4), $W_0^1, \ldots, W_0^S$, and for each we optimize the remaining parameters using gradient descent, to find $W_{1\ldots L}^s$ for $s = 1, \ldots, S$ (where we reserve a portion of the training data as a validation set to perform early stopping). Hence, we can calculate the sample training loss at meta-epoch t:

$$L_s^t = L(\mathcal{X}, \mathcal{Y}|W_{0\ldots L}^{ts}). \tag{5}$$

We then apply the smoothing-based optimization (SMO) updates from Leordeanu and Hebert (2008), to update μ and σ:

$$
\begin{aligned}
\mu_{t+1} &= \frac{\sum_s F_s \mathrm{vec}(W_0^t)}{\sum_s F_s}, \\
\sigma_{t+1} &= \sqrt{\frac{\sum_s F_s |\mathrm{vec}(W_0^t) - \mu_t|_2^2}{D_X D_Z \sum_s F_s}},
\end{aligned} \tag{6}
$$

where $F_s = \max(-L_s + c, 0)$ is the score for sample s, with the offset $c > 0$ set to a positive constant (we treat c as an additional hyperparameter, which is set empirically to ensure that $-L_s + c > 0$ for observed values of L_s). The updates in Eq. (6) can be shown to improve the value of F (i.e. the inverse loss) in expectation [as shown in Wang *et al.* (2024) and Leordeanu and Hebert (2008)]; hence:

$$\mathbb{E}_{W_0\sim Q_{t+1}}[F(\mathcal{X},\mathcal{Y}|W_{0\ldots L})] \geq \mathbb{E}_{W_0\sim Q_t}[F(\mathcal{X},\mathcal{Y}|W_{0\ldots L})], \qquad (7)$$

where $Q_t = \mathcal{N}(.|\mu_t,\sigma_t)$ is the variational distribution over W_0 at meta-epoch t, and $\mathbb{E}[.]$ is the expectation operator.

3. Experiments

3.1 Data preparation

In this work, we focus on spatial transcriptomics data generated by ST using the ST1K protocol (Shah *et al.* 2016), which measures the spatial expression of 26 949 genes by placing an array of 1007 capturing spots arranged in a 33 × 35 grid onto the tissue section and provides a tissue image stained with H&E. ST data were collected from two cohorts for breast cancer studies: one cohort (He *et al.* 2020) contains 68 tissue sections from 23 breast cancer patients with five different molecular subtypes, and the other cohort (Andersson *et al.* 2021) consists of 36 tissue sections from eight HER2-positive breast cancer patients. In the latter, each patient has one tissue section that was manually annotated with up to five tissue types based on the morphological features of the associated H&E staining image. We extracted image patches of 224 × 224 pixels centered on the corresponding capturing spots for each H&E staining image, where 224 × 224 pixels approximately cover each spot and is the standard input size for convolutional neural networks (CNNs) to derive a convenient feature set. For the spatial gene expression data, we followed the experimental setting in ST-Net (He *et al.* 2020), and first preprocessed the unique molecular identifier (UMI) counts in the raw data by normalizing them to sum to one after adding a pseudo count one for each capturing spot, and then transformed the normalized counts onto a log scale. Many of the genes are either lowly or sparsely expressed, and thus may not be essential for latent representation learning. Therefore, we followed the ST-Net setting (He *et al.* 2020) and filtered the top 250 genes with the highest gene expression across all tissue sections from two data cohorts for model training and prediction.

3.2 Experimental design

For spatial gene expression prediction, we benchmarked asGNN against four main baseline methods, including ST-Net (He *et al.* 2020), HisToGene (Pang *et al.* 2021), a basic (non-adaptive) GTN, and AP-Clustering + GTN (AP-GTN). The basic GTN was coupled with different spatial adjacency graphs by randomly dropping edges among capturing spots at ratios ranging from 0% (full) to 100% (empty), and the AP-GTN was combined with a spatial graph determined by AP clustering with similar distance function as Eq. (1) but defined on the untransformed image features. Note that other methods we mentioned in the Introduction were excluded from the comparison since either their codes were not publicly available at the time of investigation or they showed unreasonably poor performance on the data in our experimental setting. We used two different image features, including morphological and convolutional features, with all the methods tested, except for ST-Net and HisToGene. The morphological features were calculated as a 142-dimensional vector concatenating morphological statistics and nuclei type proportion, derived from the nuclei segmentation produced by HoVerNet (Cosatto *et al.* 2013) on each image patch around a capturing spot. The convolutional features were calculated as a 1024-dimensional vector extracted from pre-trained DenseNet-121 on ImageNet for each image patch around a capturing spot. Lastly, we assessed the performance of spatial gene expression prediction for all compared methods with both holdout and external validation sets. In the holdout validation, we stratified the 68 tissue sections from the first data cohort based on their molecular subtypes into training, validation, and test sets, consisting of 38, 15, and 15 sections, respectively, where the validation set was used to prevent overfitting in model training and select the best hyperparameters while the test set was used to evaluate the performance of the methods. In the external validation set, we employed 24 tissue sections from the second data cohort to evaluate the generalization of methods.

For spatial domain detection, we applied AP clustering to raw convolutional features and convolution-associated meta-features acquired from asGNN ($\lambda = 0$) and asGNN for spot clustering separately. Without the need to specify the number of clusters, the performance of AP clustering may be degraded when an intact spatial domain is partitioned into separate regions. We thus first attempted to merge AP clusters by hierarchical clustering based on the averaged features per cluster, to find the best alignment of the detected spatial domains with the annotations. However, depending on the complexity of the tissue section, the clustering performance may still be suboptimal as delicate spatial domains can be merged with their surrounding domains. Therefore, we utilized the AP generated clusters to first sparsify the spatial adjacency graph, and then identified connected components (CC) in the sparsified graph as fine-grained clusters, before finally applying hierarchical clustering to merge these fine-grained clusters according to their averaged features to find the best alignment between the merged spatial domains and the annotations.

In the last experiment, we performed a prototype clustering analysis to investigate the spatial organization across tissue sections. We pooled spatial domains detected by the asGNN for all tissue sections and computed the average of the nuclei type proportions over the image patches within each spatial domain based on the nuclei segmentation results derived from HoVerNet. We then applied k-means to group spatial domains into either $k = 5$ or $k = 10$ prototype clusters according to their averaged nuclei type proportion. Furthermore, we used the Wilcoxon rank sum test to identify differentially expressed genes for each prototype cluster based on the corresponding spatial gene expression, and then performed gene ontology (GO) enrichment analysis on prototype clusters to explore their associated biological functions.

Finally, all the experiments were conducted on a cluster using 30 CPUs and 256 GB RAM. In this environment, running the asGNN with $S = 30$ required roughly 1 h of wall time on the training data for each meta-epoch. Despite the best-performing model typically being discovered in the earlier epochs, we ran 60 meta-epochs in total to ensure comprehensive exploration.

3.3 Spatial expression prediction

To evaluate the spatial gene expression prediction performance, we initially applied asGNN, along with two state-of-the-art methods, ST-Net and HisToGene, for the holdout and external validations on the spatial transcriptomics data from 68 and 36 breast cancer tissue sections, respectively, where the validation performance was measured by two metrics, mean squared error (MSE) and PCC. It is evident that

Table 1. Spatial gene expression prediction performance comparison of state-of-the-art models.

Method	Spatial graph	Morphological features				Convolutional features			
		Holdout		External		Holdout		External	
		MSE	PCC	MSE	PCC	MSE	PCC	MSE	PCC
ST-Net[a]	N/A	–	–	–	–	0.712	0.065	1.081	0.292
HisToGene[a]	N/A	–	–	–	–	0.723	0.024	1.297	0.204
GTN[a]	Full	0.719	0.063	1.246	0.199	0.736	0.065	1.071	0.280
AP-GTN[a]	Pre-clustered	0.716	0.051	1.361	0.125	0.733	0.074	0.986	0.235
asGNN[a]	Adaptive	0.701	0.069	1.213	0.210	0.705	0.083	0.990	0.288
GTN	Full	0.710	0.090	1.240	0.204	0.711	0.101	0.961	0.297
AP-GTN	Pre-clustered	0.713	0.073	1.302	0.193	0.721	0.098	0.973	0.242
asGNN	Adaptive	0.703	0.103	1.208	0.212	0.696	0.113	0.932	0.312

[a] The models only optimize mean square error in their loss function.

ST-Net and HisToGene, basic GTN models with full (8-connected) spatial adjacency graphs, GTN model with sparsified spatial graph by AP clustering (AP-GTN), and the asGNN model, using both morphological and convolutional features from HoVerNet and DenseNet-121 as image features respectively, and performing both holdout and external validation on two data cohorts consisting of 68 and 32 breast cancer tissue sections separately. The basic GTN and AP-GTN models are as outlined in Wang *et al.* (2024). The best performance in terms of mean square error (MSE) and Pearson correlation coefficient (PCC) are underlined in each column of the table (see complete results in Supplementary Table S1).

asGNN consistently achieved the best spatial gene expression prediction performance with the lowest MSE (0.696 and 0.932) and the highest PCC (0.113 and 0.312) across all tissue sections in both holdout and external validation sets, compared to the other baseline methods, as shown in Table 1 and Supplementary Table S1. ST-Net and HisToGene exhibit competitive prediction performance in terms of MSE (ST-Net: 0.712; HisToGene: 0.723) in the holdout validation, but demonstrate worse performance (ST-Net: 1.081; HisToGene: 1.297) in the external validation, which suggests these methods might be prone to overfitting, due to the limited training data in our experimental setting, and may be less generalizable to external data. The significance levels of improvement achieved by asGNN compared to baseline models are presented in Supplementary Table S2; we note that a large majority of the comparisons (across diverse feature sets, test sets, performance metrics and training objectives), have high statistical significance, suggesting our approach provides a robust improvement over baseline methods.

To better understand how the adaptive spatial graph learned from asGNN informs spatial gene expression prediction, we introduced GTN models with varying levels of edge removal (dropout) applied to the full spatial graph as baseline methods; following the baseline comparisons used in Wang *et al.* (2024). The comparison between GTN models with different spatial adjacency graphs confirms that spatial gene expression prediction benefits from local relations among the capturing spots. The observation that asGNN outperforms all basic GTN models indicates that the local relations might be redundant and even misleading, and that pruning the spatial adjacency graph with guidance from the spatial gene expression prediction significantly improves the prediction performance. Further comparison between asGNN and AP-GTN suggests that hard-coded local relations are not always reflected in the actual spatial gene expression, resulting in overfitting to training data.

Furthermore, we assessed the spatial gene expression prediction performance of asGNN, AP-GTN, and GTN models with both morphological and convolutional features in both validation settings to explore the predictive power of different image features. Our results revealed that the performance improved significantly only when convolutional features were coupled with the asGNN, which indicates that while morphological features generally show high predictive performance, convolutional features have the potential to show higher concordance with spatial gene expression by establishing better local relations.

To investigate the importance of modeling gene correlation in spatial gene expression prediction, we conducted ablation studies by setting λ to 0 in the loss for asGNN and GTN models. The asGNN clearly shows better prediction performance compared to its variant with $\lambda = 0$, which indicates that modeling gene correlation in GTNs improves the exploitation of local relations for spatial gene expression prediction.

Finally, we visualized the gene expression patterns generated by asGNN and the alternative methods (the baseline ST-Net and HisToGene) on breast tissue sections in the test set from the holdout validation, as shown in Fig. 2. We selected the top predicted gene for each section from the asGNN, including COL1A2, MYL9, C4B, IGLL5, and GAS5, which are all spatially variable genes [assessed using SPARK (Sun *et al.* 2020), P-value $<.05$] and related to breast cancer. We then compared the spatial expression patterns of the selected genes with their ground-truth from associated spatial transcriptomics data and counterparts from ST-Net and HisToGene. It is clear that spatial expression patterns predicted by asGNN are highly correlated with their ground-truth patterns and exhibit appropriate continuity over neighboring spots, which demonstrates that asGNN is capable of capturing local relations in the spatial expression. Supplementary Table S4 and Figs S7–S9 further compare the top predicted genes by asGNN, ST-Net and HisToGene; as shown, there is only minimal overlap between the top genes predicted by each method dataset at tissue-specific levels, while out of the 15 selected tissue-specific genes (across Fig. 2 and Supplementary Figs S8 and S9), asGNN performs best according to PCC on nine of these (compared to two and four for ST-Net and HisToGene, respectively).

3.4 Spatial domain detection

One of the key advantages of asGNN is that it performs spot clustering implicitly in refining the spatial adjacency graph, where clusters obtained from either AP clustering on the latent meta-features or the sparsified spatial graph can be interpreted as spatial domains within the tissue section. To provide a quantitative measure of the spatial domains detected by asGNN, we evaluated spot clustering

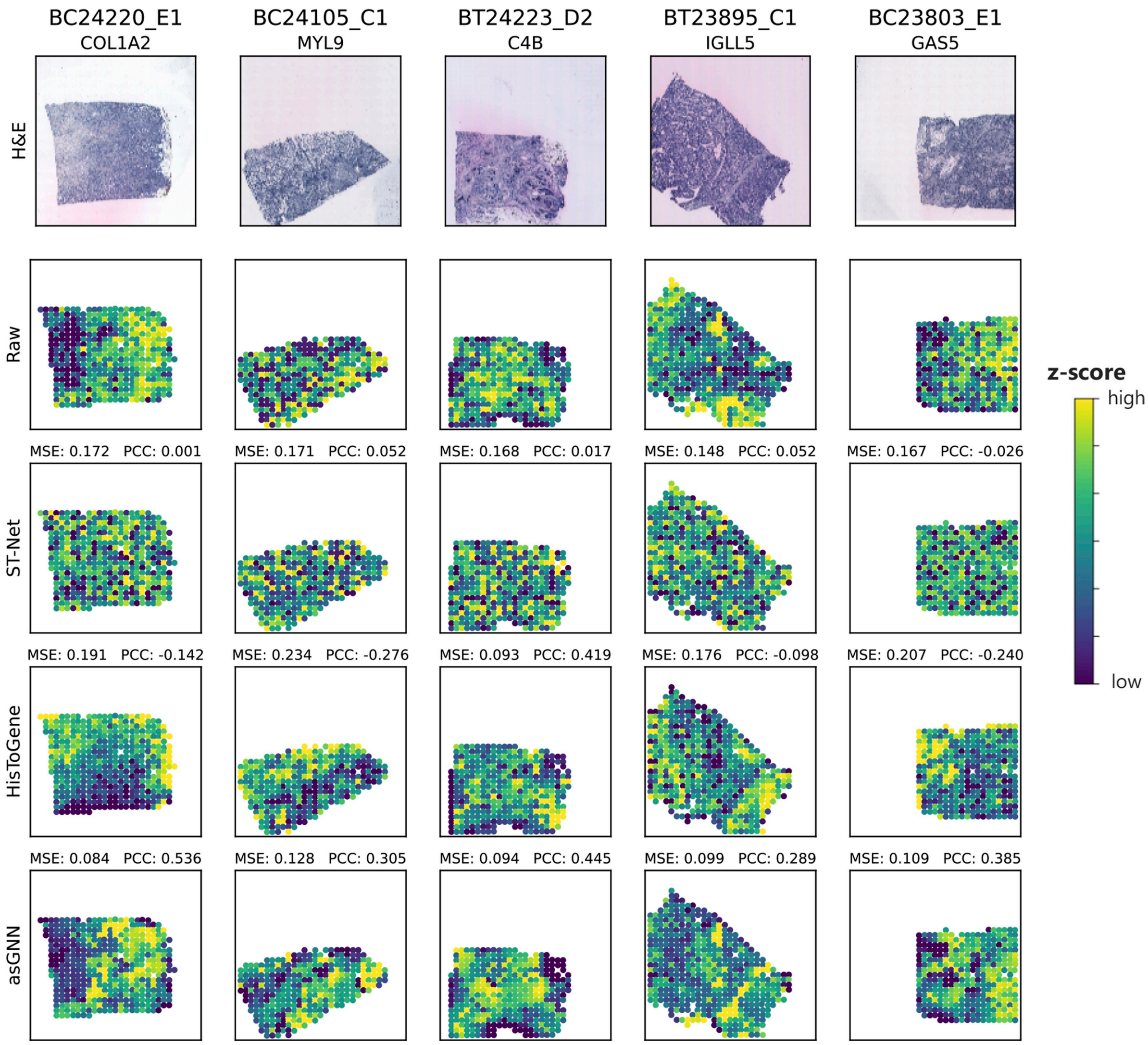

Figure 2. Expression pattern visualization of the top predicted genes by asGNN on five breast cancer tissue sections. Visualization of the raw and predicted expression patterns from ST-Net, HisToGene, and asGNN for five spatially variable genes. Genes were selected by ranking prediction performance for each breast cancer tissue sections in the holdout validation set. Both mean squared error (MSE) and Pearson correlation coefficient (PCC) between raw and predicted expression are reported for each gene. Further visualizations of the top predicted genes by ST-Net and HisToGenes are shown in Supplementary Figs S8 and S9

performance by computing the adjusted rand index (ARI) between spot clusters and annotations on 8 breast cancer tissue sections from the external validation set. ARIs were further optimized by merging either AP or CC clusters based on latent meta-features with hierarchical clustering, as described above in the Experimental Design. To better understand the importance of latent meta-features in spatial domain detection, we introduced a naive method in the comparison, which directly applied AP clustering to raw image features to generate AP and CC clusters. Note that we only focus on the convolutional features and their latent meta-features in this experiment, due to their superior performance in spatial gene expression prediction.

As shown in Supplementary Fig. S1, asGNN outperforms other baseline methods regardless of merging strategies and achieves the overall best ARI (median ARI = 0.423) by merging CC clusters for spatial domain detection, which indicates fine-grained CC clusters with convolution-associated meta-features learned from asGNN have the potential to accurately delineate spatial domains. We observed that convolution-associated meta-features generally improve the spot clustering performance compared to raw convolutional features in both merging strategies, and noticed a substantial improvement by merging CC clusters (*P*-value <.05), which implies that convolution-associated meta-features are more informative in capturing location relations. Interestingly, asGNN shows slightly worse ARI (median ARI = 0.322) than those from asGNN ($\lambda = 0$) (median ARI = 0.363) when merging AP clusters, which might be attributed to delicate spatial domains being masked by the coarse-grained clusters obtained from AP clustering, possibly as a result of using non-optimal clustering hyperparameters.

To illustrate how the CC merging strategy and convolution-associated meta-features contribute to spatial domain detection intuitively, we further visualized the spot clustering results on annotated breast cancer tissue sections, as depicted in Fig. 3. We found that the spatial domains from asGNN matched well with the annotated tissue regions and even exhibited high agreement with fine-grained structures (e.g. breast gland and immune infiltrate), whereas the spatial

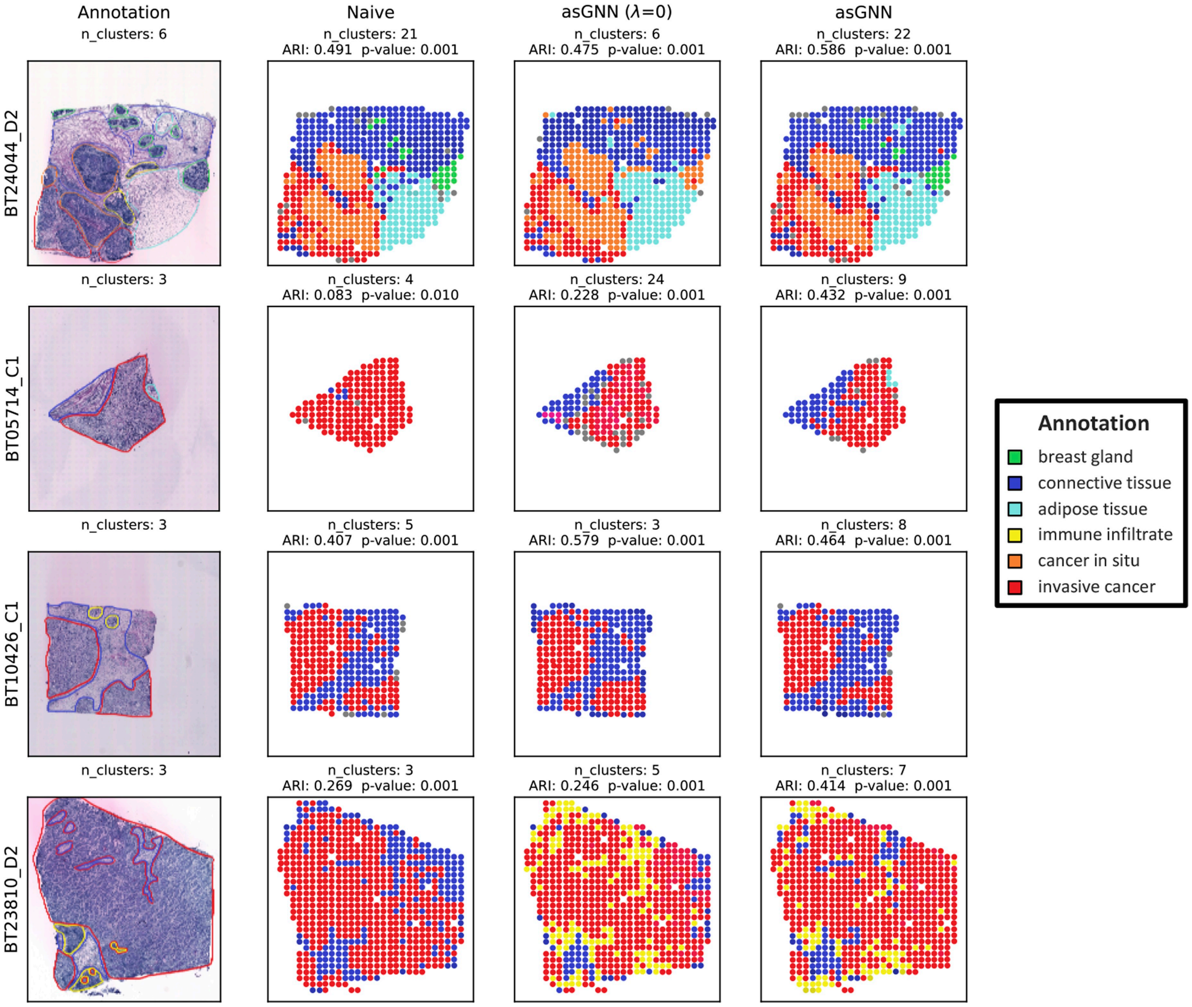

Figure 3. Spatial domain visualization on four breast cancer tissues. Visualization of spatial regions annotated by pathologists and spatial domains optimized by merging connected-component (CC) clusters from naive, asGNN ($\lambda = 0$) and asGNN on four breast cancer tissue sections. Adjusted rand index (ARI) between detected spatial domains and annotated spatial regions, along with the corresponding *P*-values from permutation test, were reported for each method across different tissues. Note that the spatial domains are labeled similarly if they were recognized as the same region in the annotation. Singleton clusters were excluded for better visualization

domains from naive methods appeared over-smoothed and failed to distinguish spatial domains correctly. asGNN and asGNN ($\lambda = 0$) demonstrated superior performance in spatial domain detection by merging AP clusters (as shown in Supplementary Fig. S2); asGNN could identify continuous spatial domains with clear boundaries and asGNN ($\lambda = 0$) could even recognize narrow structures consisting of a few spots (e.g. immune infiltrate). We note, however, that asGNN tends to produce fewer AP clusters than the actual number of annotated regions for most tissues, which might obfuscate some fine-grained structures.

3.5 Prototype clustering and enrichment analysis

To further chart the spatial organization under various tissue contexts, we employed *k*-means clustering to discover prototypes for the spatial domains across all tissue sections from breast cancer patients. Instead of relying on latent metafeatures, which might be biased toward spatial expression of particular genes, we used nuclei type composition derived from nuclei segmentation in the prototype clustering. As depicted in Fig. 4, the spatial organization shows visual consistency across replicated tissue sections, implying the robustness of asGNN in spatial domain detection. While prototype clusters from different granularities of clustering exhibit high correspondence, the prototype clustering with $k = 10$ allows finer-grained characterization spatial organization in breast cancer, such as tumor regions with different subtypes (clusters 0, 1, and 6).

To evaluate the stability of the cluster prototypes, we measured the stability score (SS) for each prototype cluster by calculating the bootstrapped ARIs of the clustering results after masking out each prototype cluster in turn. Subsequently, to further investigate the biological interpretation of prototype clusters, we conducted GO enrichment analysis on the differentially expressed genes from all the genes in the original spatial gene expression data (not only the predicted genes) for each prototype cluster for both $k = 5$ and $k = 10$ clustering settings, and the top enriched GO terms along with the SSs for each prototype cluster are present in Supplementary Table S3. We identified prototype clusters with SS >0.7 and found their enriched biological processes to be highly relevant to the corresponding regions in the annotated tissue sections; for example, clusters 1 and 6 in the $k = 10$ setting are enriched for tumor development and tumor-associated immune responses (de Visser and Joyce 2023), respectively, closely resembling the tumor core and surrounding regions in

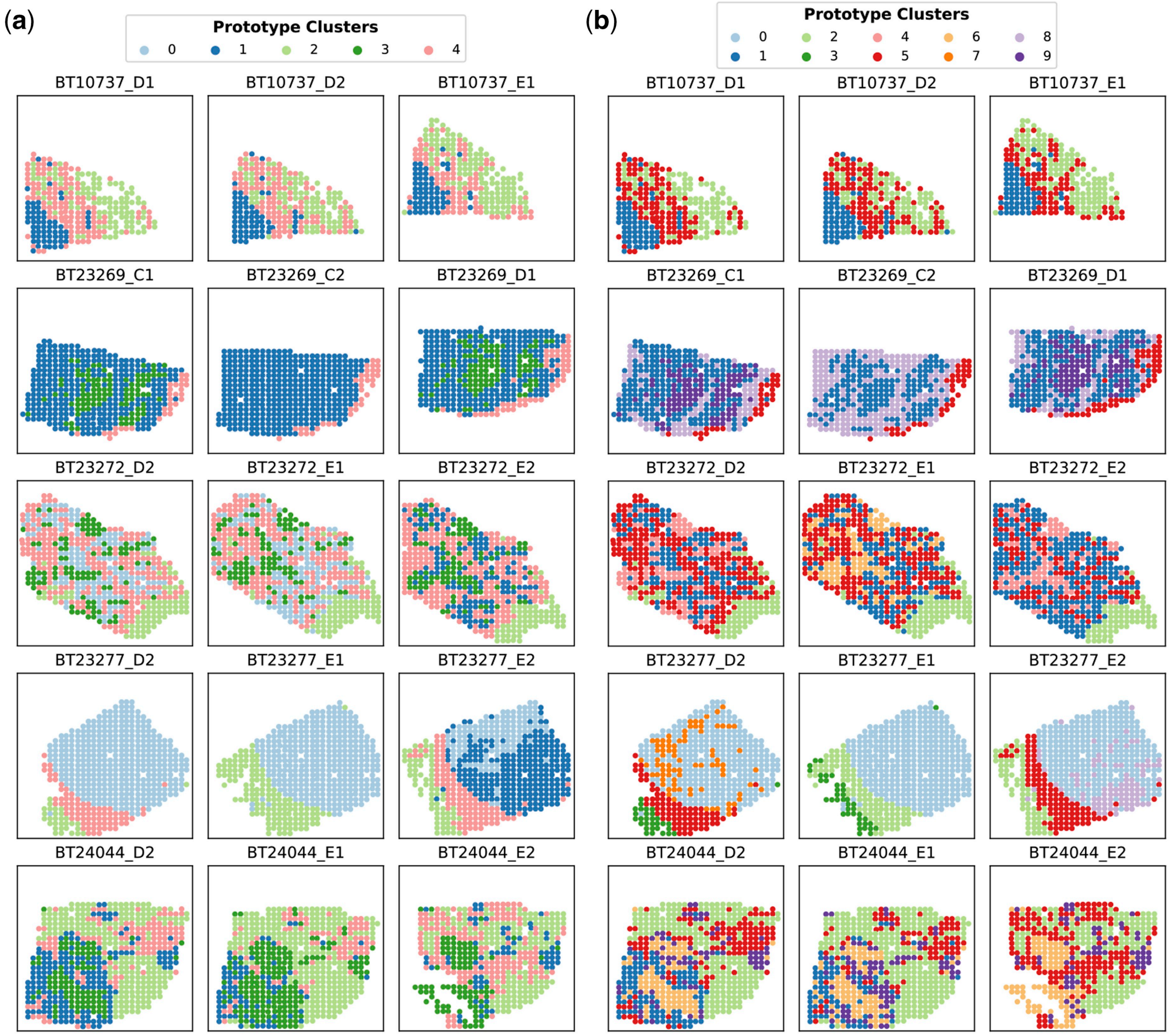

Figure 4. Prototype cluster visualization on 15 breast cancer tissue sections. Discovering prototype clusters across different breast cancer tissue sections by applying *k*-means to nuclei type composition of their AP clusters from the asGNN method (nuclei type composition determined using the nuclei segmentation results from HoVerNet). (a) Visualization of prototype clusters (*k* = 5). (b) Visualization of prototype clusters (*k* = 10). Further examples are shown in Supplementary Figs S5 and S6

a few tissue sections. Interestingly, we also discovered that several prototype clusters, which are enriched for biological processes related to the microenvironment, are spatially adjacent to each other in some breast cancer tissue sections, which aligns well with the concept of a multilayered microenvironment (Laplane *et al.* 2018).

4 Discussion

We have introduced an asGNN architecture for spatial gene expression prediction, which builds on the adaptive graph refinement framework of Wang *et al.* (2024). We have shown that our model generates state-of-the-art performance for predicting spatial gene expression from histological image data. Our method learns to adapt the spatial graph structure on an image-by-image basis, so that information is only shared between spots in coherent spatial domains, defined by the learned meta-features. Our model can be trained in an end-to-end fashion, using a smoothing-based VO approach (Leordeanu and Hebert 2008, Wang et al. 2024). Further, we have shown that the spatial domains identified by our method achieve a high degree of alignment with pathologist annotations, and can be readily interpreted biologically through our prototype analysis as corresponding to layered tumor and tumor microenvironment regions.

As future work, we intend to investigate both the biological and clinical potential of our method. The ready availability of large quantities of histology images [for instance, in TCGA (Weinstein *et al.* 2013)] suggests that we may be able to improve the predictive performance or fine-tune our model to different tumors using a pseudo-labeling (semi-supervised) approach, by augmenting our training set with instances where only image or image and bulk expression data are available. Further, the ability of our approach to generate putative spatial domains, suggests that we may be able to identify *de novo* tumor-type specific tumor or microenvironment domains through such analysis. We further plan to test the potential of our approach to handle spatial expression data with higher spatial resolution (for instance, subcellular), and explore the potential for using cluster identity to influence graph refinement (to model inter-cluster dependencies). Finally, we intend to investigate the potential of our method to identify spatial biomarkers for patient stratification for clinical diagnostics and personalized treatment, where the spatial expression patterns predicted by our model can be used both as biomarkers themselves, and, in a semi-supervised setting, to help learn novel biomarkers. Code available at: https://github.com/song0309/asGNN/.

Supplementary data

Supplementary data are available at *Bioinformatics* online.

Conflict of interest

None declared.

Funding

This work was supported by NEC Laboratories America. T.S. and R.K. were partially supported by a grant from the National Science Foundation, USA (NSF BIO DBI-IIBR #2042159). This paper was published as part of a supplement financially supported by ECCB2024.

Data availability

All processed image and spatial transcriptomics data used in this article are publically available at https://github.com/song0309/asGNN/, and raw image, spatial transcriptomics data, and manual annotations for breast cancer tissues are available in Mendeley at https://data.mendeley.com/datasets/29ntw7sh4r/5 and Zenodo at https://zenodo.org/records/4751624.

References

Andersson A, Larsson L, Stenbeck L *et al.* Spatial deconvolution of HER2-positive breast cancer delineates tumor-associated cell type interactions. *Nat Commun* 2021;**12**:6012.

Asp M, Bergenstråhle J, Lundeberg J. Spatially resolved transcriptomes-next generation tools for tissue exploration. *Bioessays* 2020;**42**:e1900221.

Cosatto E, Laquerre P-F, Malon C *et al.* Automated gastric cancer diagnosis on H&E-stained sections; training a classifier on a large scale with multiple instance machine learning. In: *Medical Imaging 2013: Digital Pathology, SPIE 8676*, Lake Buena Vista, FL. Bellingham, WA: SPIE, 2013, 51–9.

Dawood M, Branson K, Rajpoot NM *et al.* All you need is color: image based spatial gene expression prediction using neural stain learning. In: *Joint European Conference on Machine Learning and Knowledge Discovery in Databases*, Virtual. Berlin, Germany: Springer, 2021, 437–50.

de Visser KE, Joyce JA. The evolving tumor microenvironment: from cancer initiation to metastatic outgrowth. *Cancer Cell* 2023;**41**:374–403.

Dries R, Chen J, Del Rossi N *et al.* Advances in spatial transcriptomic data analysis. *Genome Res* 2021;**31**:1706–18.

Frey B, Dueck D. Clustering by passing messages between data points. *Science* 2007;**315**:972–6.

He B, Bergenstråhle L, Stenbeck L *et al.* Integrating spatial gene expression and breast tumour morphology via deep learning. *Nat Biomed Eng* 2020;**4**:827–34.

Laplane L, Duluc D, Larmonier N *et al.* The multiple layers of the tumor environment. *Trends Cancer* 2018;**4**:802–9.

Leordeanu M, Hebert M. Smoothing-based optimization. In: *2008 IEEE Conference on Computer Vision and Pattern Recognition*, Anchorage, AK. New York, NY: IEEE, 2008, 1–8.

Mejia G, Cárdenas P, Ruiz D *et al.* SEPAL: spatial gene expression prediction from local graphs. . In: *Proceedings of the IEEE/CVF International Conference on Computer Vision*, Paris, France. New York, NY: IEEE, 2023, 2294–303.

Monjo T, Koido M, Nagasawa S *et al.* Efficient prediction of a spatial transcriptomics profile better characterizes breast cancer tissue sections without costly experimentation. *Sci Rep* 2022;**12**:4133.

Pang M, Su K, Li M. Leveraging information in spatial transcriptomics to predict super-resolution gene expression from histology images in tumors. bioRxiv, https://doi.org/10.1101/2021.11.28.470212, 2021, preprint: not peer reviewed.

Shah S, Lubeck E, Zhou W *et al.* In situ transcription profiling of single cells reveals spatial organization of cells in the mouse hippocampus. *Neuron* 2016;**92**:342–57.

Shi Y, Huang Z, Feng S *et al.* Masked label prediction: unified message passing model for semi-supervised classification. arXiv, arXiv:2009.03509, 2020, preprint: not peer reviewed.

Sun S, Zhu J, Zhou X. Statistical analysis of spatial expression patterns for spatially resolved transcriptomic studies. *Nat Methods* 2020;**17**:193–200.

Wang G, Warrell J, Zheng S *et al.* A variational graph partitioning approach to modeling protein liquid-liquid phase separation. bioRxiv, https://doi.org/10.1101/2024.01.20.576375, 2024.

Weinstein JN, Collisson EA, Mills GB *et al.* The cancer genome atlas pan-cancer analysis project. *Nat Genet* 2013;**45**:1113–20.

Yang Y, Hossain MZ, Stone EA *et al.* Exemplar guided deep neural network for spatial transcriptomics analysis of gene expression prediction. In: *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision*, Waikoloa, Hawaii. New York, NY: IEEE, 2023, 5039–48.

Zeng Y, Wei Z, Yu W *et al.* Spatial transcriptomics prediction from histology jointly through transformer and graph neural networks. *Brief Bioinform* 2022;**23**:bbac297.

Bioinformatics, 2024, 40, 111–119
https://doi.org/10.1093/bioinformatics/btae383
Proceedings of ECCB2024