

Gene expression prediction from histology images via hypergraph neural networks

Bo Li¹, Yong Zhang^{1,*}, Qing Wang², Chengyang Zhang¹, Mengran Li³, Guangyu Wang^{4,5}, Qianqian Song^{6,*}

¹Beijing Key Laboratory of Multimedia and Intelligent Software Technology, Faculty of Information Technology, Beijing Institute of Artificial Intelligence, Beijing University of Technology, Beijing, 100124, China

²Key Laboratory of Intelligent Education Technology and Application of Zhejiang Province, Zhejiang Normal University, Jinhua, Zhejiang 321004, China

³School of Intelligent Systems Engineering, Sun Yat-sen University, Guangzhou, Guangdong 510275, China

⁴Center for Bioinformatics and Computational Biology, Houston Methodist Research Institute, Houston, TX 77030, United States

⁵Department of Cardiothoracic Surgery, Weill Cornell Medicine, Cornell University, New York, NY 10065, United States

⁶Department of Health Outcomes and Biomedical Informatics, College of Medicine, University of Florida, FL 32611, United States

*Corresponding authors. Qianqian Song, E-mail: qsong1@ufl.edu; Yong Zhang, E-mail: zhangyong2010@bjut.edu.cn

Abstract

Spatial transcriptomics reveals the spatial distribution of genes in complex tissues, providing crucial insights into biological processes, disease mechanisms, and drug development. The prediction of gene expression based on cost-effective histology images is a promising yet challenging field of research. Existing methods for gene prediction from histology images exhibit two major limitations. First, they ignore the intricate relationship between cell morphological information and gene expression. Second, these methods do not fully utilize the different latent stages of features extracted from the images. To address these limitations, we propose a novel hypergraph neural network model, HGGEF, to predict gene expressions from histology images. HGGEF includes a gradient enhancement module to enhance the model's perception of cell morphological information. A lightweight backbone network extracts multiple latent stage features from the image, followed by attention mechanisms to refine the representation of features at each latent stage and capture their relations with nearby features. To explore higher-order associations among multiple latent stage features, we stack them and feed into the hypergraph to establish associations among features at different scales. Experimental results on multiple datasets from disease samples including cancers and tumor disease, demonstrate the superior performance of our HGGEF model than existing methods.

Keywords: spatial transcriptomics; gene expression prediction; histology image; gradient enhancement; attention mechanism; hypergraph

Introduction

Diverse cell types are intricately arranged both spatially and structurally within tissues to fulfill their specific functions. Revealing the intricate spatial architecture and cell activities within heterogeneous tissues holds considerable importance in comprehending the cellular mechanisms and functions associated with diseases [1–4]. Spatial Transcriptomics (ST) serves as an advanced technology that can be utilized to elucidate the spatial distribution of genes at both tissue and spot levels. This technology has significantly advanced our understanding of gene expressions in biological processes [5], playing a crucial role in exploring disease mechanisms [6] and revealing novel drug targets [7]. The rapid progress of ST technology allows for the simultaneous analysis of gene expression, cell or spot locations, and corresponding histology images. Currently, numerous researchers are actively engaged in related studies, covering spatial domain recognition [8–10], spatial transcriptomics deconvolution [11–13], and inference of spatial cellular interactions [14, 15].

However, the considerable cost associated with acquiring spatial transcriptomics data limits the widespread pursuit of research on ST technologies. In contrast, histology images of various disease tissues are more accessible. Recently, researchers have shifted their focus toward predicting gene expression from whole-slide image (WSI) data. Some methods, such as ST-Net [16], HisToGene [17], Hist2ST [18], DeepPT [19], BLEEP [20], and THItToGene [21], have emerged for this purpose. Initially, ST-Net pioneers the use of deep learning techniques to predict spatial gene expression from WSI, yielding promising results. HisToGene and Hist2ST improve the prediction performance by incorporating transformer models to capture global associations of image features across different spots in a WSI. Meanwhile, Hist2ST leverages graph neural networks [22] to enhance local associations of image features between spots. Recently, THItToGene further employs graph attention networks to explore the correlation between gene expression and spatial location. Unlike the direct prediction methods from images to gene expression mentioned above, BLEEP uses a contrastive learning approach to align images with gene expression. It's worth mentioning

Received: February 24, 2024. Revised: June 3, 2024. Accepted: September 30, 2024

© The Author(s) 2024. Published by Oxford University Press.

This is an Open Access article distributed under the terms of the Creative Commons Attribution Non-Commercial License (<https://creativecommons.org/licenses/by-nc/4.0/>), which permits non-commercial re-use, distribution, and reproduction in any medium, provided the original work is properly cited. For commercial re-use, please contact journals.permissions@oup.com

that Adam et al. [23] conduct a comprehensive benchmark in the current field. They provide multiple metrics to thoroughly evaluate various models, including the performance of predicted gene expression, model generalizability, translational potential, usability, and computational efficiency of each method.

Although the aforementioned work has made significant progress, several important aspects remain overlooked. For instance, ST-Net ignores the positional information of spots and does not explore the correlations between multiple spots. HisToGene directly feeds image patches into ViT, resulting in significant information loss. Hist2ST and THItToGene use graph neural networks and graph attention networks, respectively, to model spot-neighborhood relations within global features, but they overlook the associations between spots that are spatially distant yet closely related. Overall, these existing methods still face two primary limitations: (i) they ignore the intricate relationship between cell morphological information and gene expression; (ii) insufficient utilization of image-based features at multiple latent stages, coupled with the oversight of high-order associations among those features. A detailed comparison of existing methods is shown in [Supplementary Table 1](#).

Regarding the first limitation, existing methods based on traditional convolution primarily concentrate on semantic information, i.e. pixel values in the image. They do not sufficiently consider the gradient relationship between the current position and its neighboring positions, which leads to the difficulty of the model to perceive the cell morphological information related to gene expression. To address this limitation, our HGGEF model includes the gradient enhancement module to refine the extracted imaging features and generate latent feature maps with prominent cell morphological information. To address the second limitation and enhance the utilization of features at multiple latent stages within WSI, our HGGEF model employs a two-step strategy. Specifically, HGGEF first extracts multiple latent features from WSI through a lightweight backbone network [24], and subsequently refine the representation of features at each latent stage using the attention mechanism [25, 26]. To explore higher-order associations among multiple latent stage features, we innovatively introduce a hypergraph association module based on multiple metrics. Compared to traditional graph neural networks [27], hypergraph convolution [28, 29] can multiple nodes with a single hyperedge, enabling effective joint representation and modeling of higher-order associations among features. Collectively, we propose a novel HGGEF model that overcomes existing challenges and achieves superior performance in gene expression prediction from histology images.

Results

Overview of HGGEF model

Our hypergraph neural network-based model, HGGEF (Hyper-Graph Gene Expression Prediction), is depicted in [Fig. 1](#). The process initiates with the partitioning of a whole slide image into multiple patches centered around spots. Acknowledging the intricate relationship between cell morphology and gene expression, the Gradient Enhancement Module (GEM) is utilized to enhance the model's perception of cell morphology. The latent features preprocessed by GEM undergo processing through a lightweight backbone network [24] to extract features at multiple latent stages. These latent features are then fed into the Convolutional Block Attention Module (CBAM [25]) and Vision Transformer (ViT [26]), utilizing attention mechanisms to optimize the representation of features at each latent stage. To uncover high-order associations among features across

different stages, a Hypergraph Association Module (HAM) based on nearby positions and distance relations is employed for global associations modeling and local representation. Following this, the latent features output from the ViT and HAM are concatenated and fed into the Long Short-Term Memory (LSTM [30, 31]) module, enhancing information exchange among latent features. Ultimately, the latent features from the last layer are input into a gene regression head to yield the predicted gene expression on this histology image.

Performance evaluation using diverse spatial transcriptomics datasets

To evaluate the performance of HGGEF, we conducted experiments using a leave-one-out validation approach on a total of 44 sections from human HER2-positive breast tumor (HER2+ [32]) dataset (section A1-G3) and human cutaneous squamous cell carcinoma (cSCC [33]) dataset (section P2_ST_rep1-P10_ST_rep3). The performance of HGGEF is compared with existing methods including Hist2ST, HisToGene, and ST-Net. [Figure 2](#) illustrates the comparison results based on the Pearson Correlation Coefficient (PCC) index for each method. Specifically, on the HER2+ datasets, the HGGEF model exhibits an average PCC index and median PCC index $\sim 5\%$ higher than the second-best-performing Hist2ST model. Similarly, on the cSCC datasets, the HGGEF model outperforms the Hist2ST model by around 4%. [Figure 2](#) reveals that most of the compared models perform less favorably in sections A2-A6 and F1-F3, with PCC indices consistently below 0.07. In contrast, our HGGEF model successfully improves the performance by $\sim 5\%$. Notably, the most substantial improvement relative to the compared models occurs in sections C1-C5, with the PCC increase of $\sim 7\%$. These results demonstrate that compared with existing methods, our HGGEF model presents superior capabilities of predicting gene expressions from histology images. Additionally, we conduct experiments to evaluate the model's goodness-of-fit on the training dataset, as shown in [Fig. S4](#) of the supplementary materials. The results indicate that the proposed HGGEF model demonstrates excellent goodness-of-fit on the training dataset and significantly outperforms other models in overall performance.

To comprehensively validate the gene expression prediction performance of our model, we also provide performance comparisons using other metrics, including Structural Similarity Index Measure (SSIM) and Root Mean Square Error (RMSE). SSIM helps evaluate the model's ability to preserve the spatial patterns and structural features of gene expression. RMSE quantifies the difference between the predicted gene expression values and the actual observations by calculating the square root of the average squared prediction errors. Therefore, a higher SSIM and a lower RMSE indicate better model performance. [Figure 3](#) presents the comparison results of various models based on SSIM and RMSE metrics. Clearly, HGGEF achieves the highest SSIM value and the lowest RMSE value, demonstrating that the HGGEF model proposed in this paper achieves the best gene expression prediction performance on these two metrics. Additionally, we included comparisons with THItToGene in the supplementary file ([Figs S1-S3](#)). The results still indicate that our method achieved the best performance.

To assess the contribution of each module's performance, ablation experiments were conducted on the HER2+ datasets, and the results are presented in [Fig. 4](#). In this figure, red color represents the performance of ablation experiments conducted on the HGGEF model, while the other three colors correspond to the compared methods (Hist2ST, HisToGene, ST-Net). The observations in [Fig. 4](#) lead to the following three conclusions: (i) As the number of

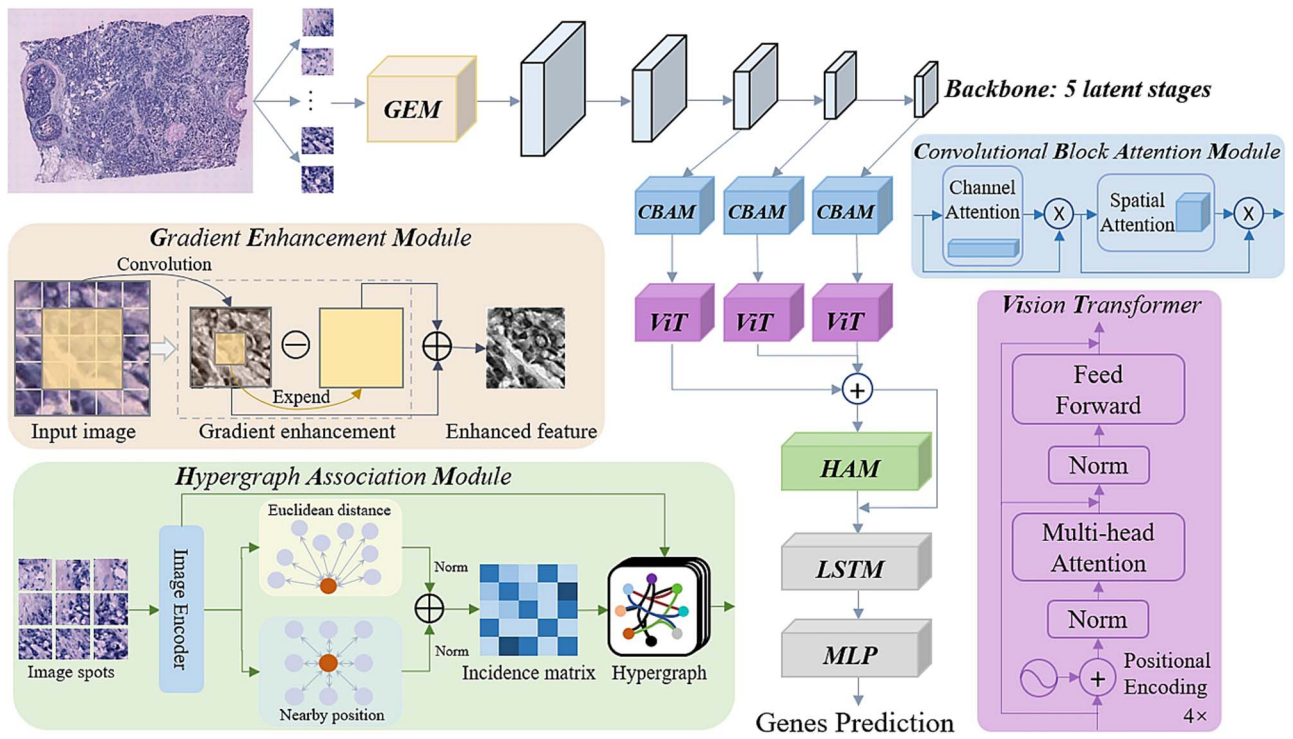


Figure 1. Overview of our HGGE model. This model consists of three pivotal components: GEM, designed to capture the intricate relationship between cell morphology and gene expression; CBAM and vision transformer modules, employed for extracting internal features at each latent stage; and a hypergraph association module (HAM), dedicated to revealing higher-order associations among multiple latent stage features.

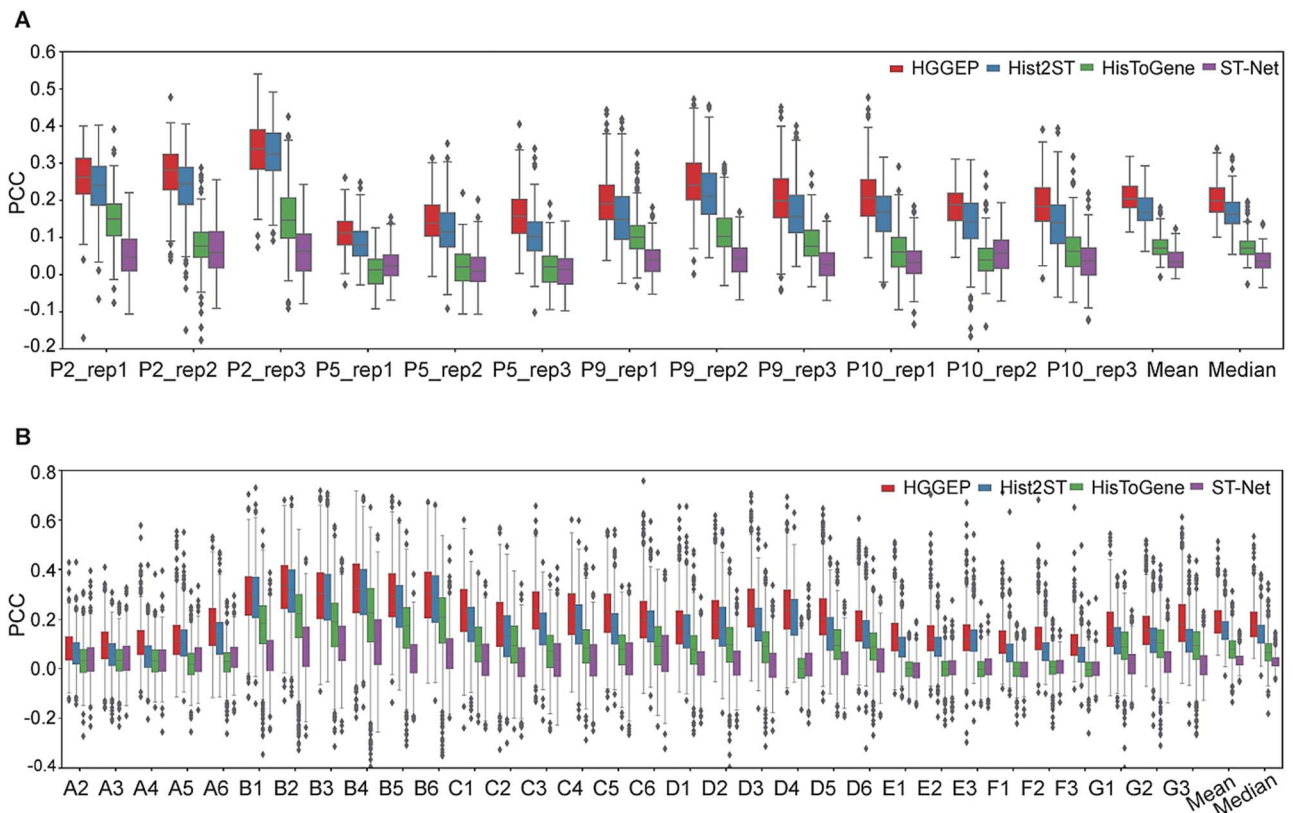


Figure 2. Benchmark of the gene expression prediction performance. Comparison results between our HGGE model and existing methods on the (A) HER2+ datasets and (B) cSCC datasets.

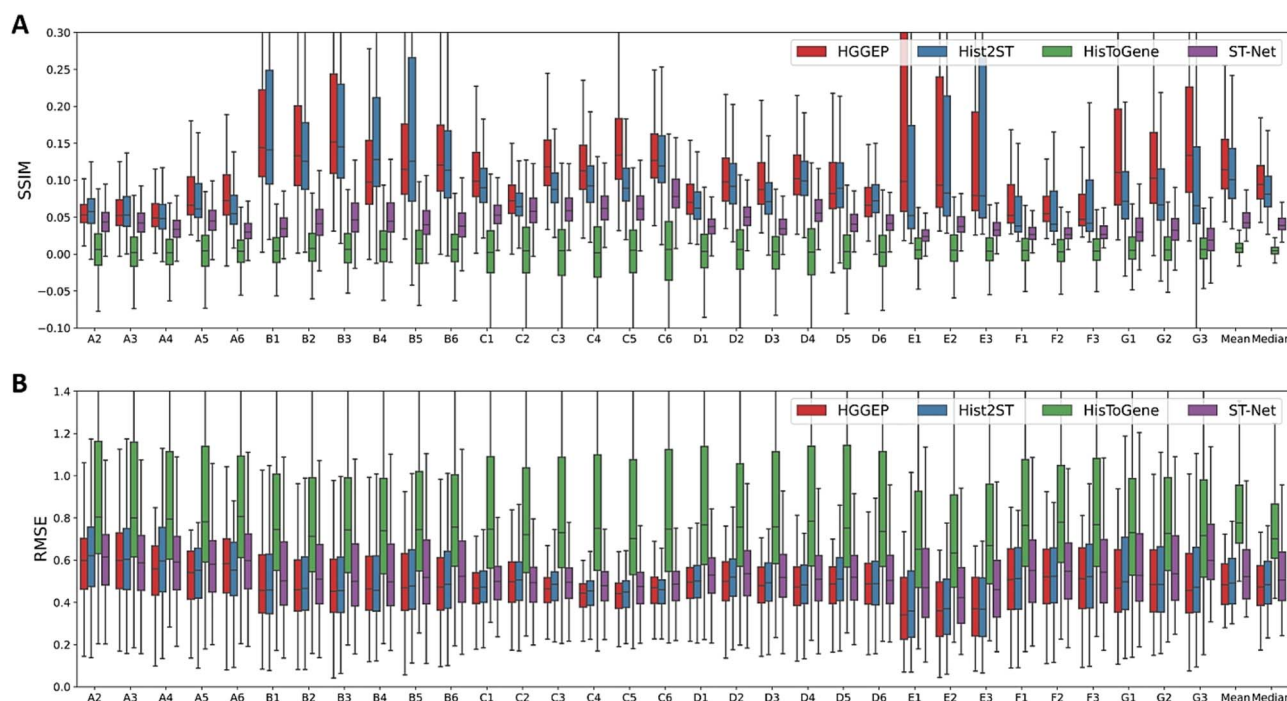


Figure 3. Benchmark of the gene expression prediction performance based on SSIM and RMSE. Comparison results between our HGGEp model and existing methods for SSIM (A) and RMSE (B) on the HER2+ datasets. Among them, a higher SSIM and a lower RMSE indicate better model performance.

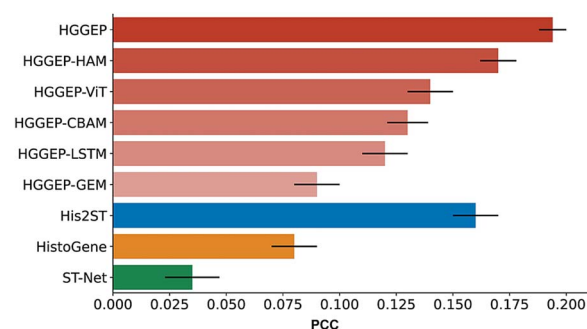


Figure 4. Ablation studies of our HGGEp model. As the number of modules removes, the mean PCC of the model gradually declines, underscoring the substantial contribution each module makes to the overall performance of the HGGEp model.

modules removes, the mean PCC of the model gradually declines, underscoring the substantial contribution each module makes to the overall performance of the HGGEp model. (ii) The results from the ablation experiments indicate that the HAM, ViT, and GEM modules significantly enhance performance. This finding underscores these modules' ability to capture cellular morphology information in tissue images, optimize feature representation at each latent stage, and capture higher-order associations between features across multiple latent stages.

Evaluation of predicted gene expression

Next we have compared the predicted gene expression between the HGGEp model, Hist2ST, and HisToGene. As shown in Fig. S4, most dots are positioned above the diagonal line, indicating that the HGGEp model achieves superior gene expression predictions. It is noteworthy that the top genes identified in the HER2+ dataset (GNAS, UBA52, MUCL1) and the cSCC dataset (EIF5, MSMO1, STMN1) indicate significant biological relevance. In the

HER2+ tumor microenvironment, GNAS is implicated with critical roles in cell signal transduction [34], potentially influencing the biological processes including the proliferation and metastasis of tumor cells [35]. UBA52, closely associated with ubiquitin-protein conjugation [36], may significantly impact the treatment response and prognosis of HER2+ tumors due to its aberrant expression [37]. Additionally, MUCL1, involved in cell adhesion and tumor microenvironment regulation, plays a crucial role in HER2+ tumors and offers potential insights for precision therapy [38]. On the other hand, in the cSCC environment, EIF5 is implicated in protein synthesis regulation and cell proliferation [39]. MSMO1 participates in cholesterol biosynthesis [40], and STMN1 may influence biological processes such as cell division and migration [41]. Therefore, accurate prediction of those genes contributes to exploring the molecular mechanisms driving cSCC development, thus gaining a better understanding of tumor cell activities in the microenvironment.

For a more intuitive comparison of top gene prediction performance across different models, we conducted the visualization of predicted genes. To ensure a fair comparison, we select four genes mentioned in Hist2ST (GNAS, FASN, MY12B, and SCD) and present their predicted gene expressions by each model in Fig. 5. This figure demonstrates our model achieves the best prediction performance. For example, our model achieves better prediction of the GNAS gene expression (PCC=0.637), compared with the competitors (Hist2ST, PCC=0.591; HisToGene, PCC=0.493). Similarly, for the genes FASN, MY12B, and SCD, our model demonstrates consistently better prediction performance. These results support that our model not only excels in overall performance but also demonstrates outstanding performance in predicting specific genes with biological significance.

Evaluation of spatial region detection

To further demonstrate the accurate gene expressions predicted by our model, we evaluate the spatial region detection

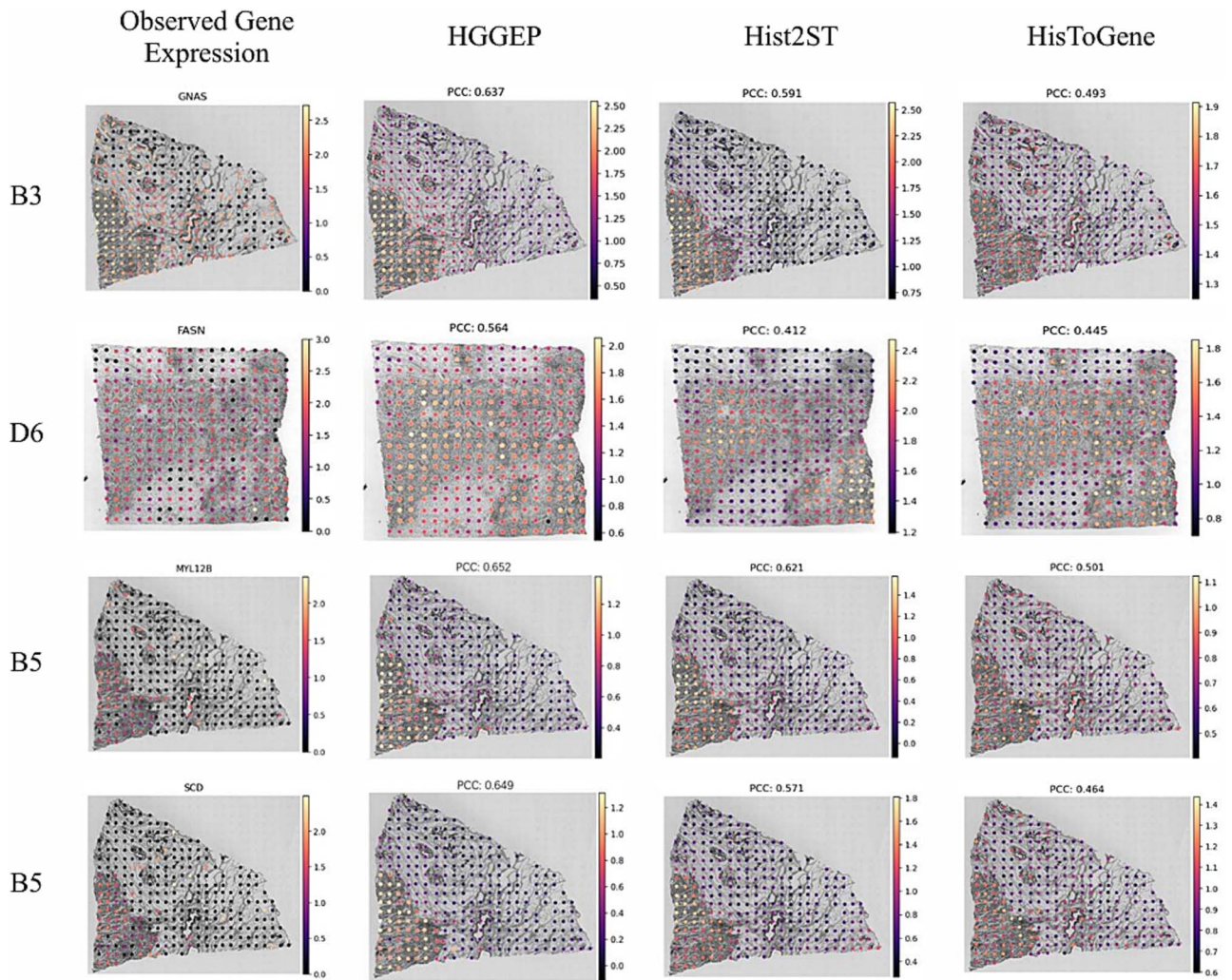


Figure 5. Visualization of predicted genes. The top predicted genes across all tissue sections by HisToGene in the HER2+ dataset, where the p-value for each tissue section was obtained in the association test between the predicted and observed gene expression.

capability across the entire histology image accordingly. We performed Leiden clustering analysis on the predicted gene expressions of each method, as illustrated in Fig. 6. Here we utilized six sections (B1, C1, D1, E1, F1, and G2) of the HER2+ dataset, with available annotations from expert professionals. As depicted in Fig. 6, our model consistently outperforms the latest Hist2ST and HisToGene models, achieving nearly optimal detection performance. Across the six sections, our model surpasses the runner-up model Hist2ST by ~11% in average performance. Particularly noteworthy is the outstanding performance in section C1, where our model significantly improves the ARI index by 30% compared to Hist2ST, demonstrating the superior precision of our HGGEp model in region detection.

Discussion

Gene expression prediction is a pivotal focus in current research, with its extensive applications making it a central theme in scientific exploration. However, many existing methods face limitations due to their reliance on costly data, which constrains further research progress. In response to this challenge, some researchers have turned to a more cost-effective strategy, utilizing histology images to unveil gene expression. In this work, we have introduced HGGEp, a cutting-edge hypergraph neural network

model tailored for the precise prediction of gene expression from histology images. HGGEp has been applied to a total of 44 sections, showcasing robust performance in predicting gene expression and underscoring the pivotal role of histology images in this process.

Our HGGEp model outperforms existing methods in three aspects. First, we enhance the model's ability to perceive morphological information through the gradient enhancement module, thus enabling the model to capture the relationship between cell morphology and gene expression. Second, features at different stages in the image have different information, usually high-level features have more semantic information, while low-level features carry more detailed features. Therefore, we use a lightweight backbone to extract different levels of information and then use this attention mechanism to optimize the features at each stage. Lastly, we fuse features from various stages, establishing global associations among spots to achieve comprehensive information integration.

Though HGGEp presents superior performance, we anticipate key areas for future exploration and improvement. First, HGGEp still relies on spot-level predictions for the gene expression of corresponding images and does not explore the strong correlation between individual cell information and gene expression within the images. In the future, we will focus on

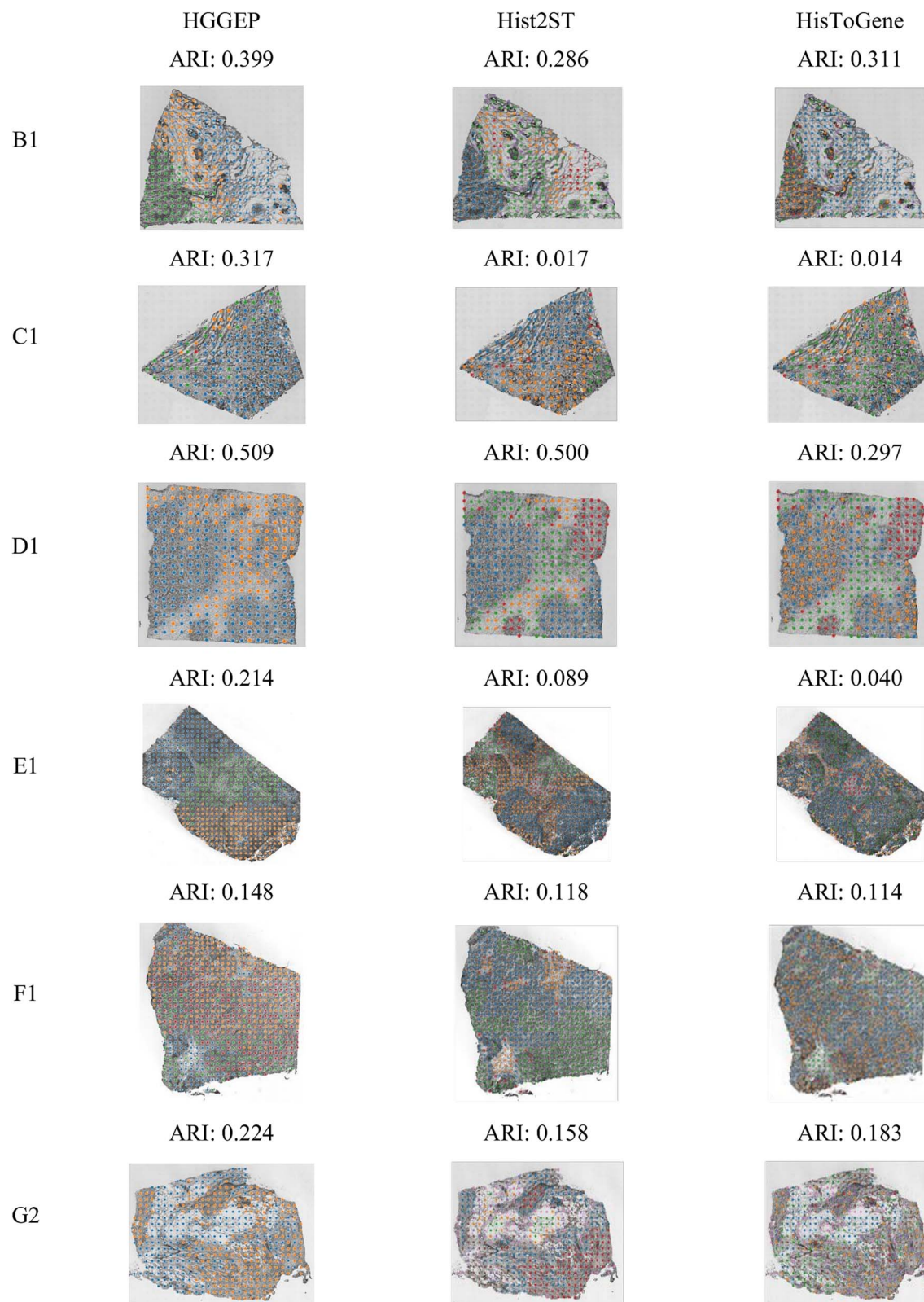


Figure 6. Spatial domain detection based on predicted gene expressions. The first column presents the observed gene expression clustering results, while the last three columns show the clustering outcomes for gene expression as predicted by different methods (HGGEp, Hist2ST, and HisToGene).

predicting gene expression based on cell-level or even pixel-level data. Second, most existing methods rely on large amounts of labeled data for learning, thus limiting the ability of the model to generalize to small sample domains. To overcome this limitation, future work could consider fine-tuning based on large vision

models in the biomedical field [42, 43]. By leveraging the wealth of knowledge that already exists in large vision models, we can enhance the generalizability of the models to a wider range of application scenarios. Third, existing methods typically use histology images as input for predicting gene expression. Future

research will consider introducing more modal information as input to help establish mapping relationships between images and gene expression, such as molecular biology data or pathology data. By integrating multimodal information and leveraging advanced hypergraph neural network methods for modeling, we aim to enhance the accuracy and comprehensiveness of gene expression prediction. In conclusion, the experiments in this paper demonstrate the power of HGGEp in predicting gene expression based on histology images. Moreover, future work could focus on enhancing model generalization capabilities, including the use of large vision models and multimodal information, to better accommodate the diversity of biomedical data and scenarios.

Materials and methods

Data processing

To validate the efficacy of our proposed method, we utilized the spatial transcriptomics datasets comprising histology images and gene expression data at spot locations. We primarily leveraged the HER2+ and cSCC datasets profiled from 32 and 12 tissue slides, respectively, with a total of 9612 spots and 6630 spots respectively. The tissue images in the Her2+ and cSCC datasets correspond to 785 and 171 highly variable genes, respectively. For each histology image in the datasets, we segment it into multiple sub-images based on the positions of spots. Each sub-image is cropped by 112×112 pixels around the spot's center. The input sub-image features are annotated as $\mathbf{x}_{in} \in \mathbb{R}^{N \times 3 \times 112 \times 112}$, where N is the number of spots within the histology image.

During the validation phase, we adopt a consistent leave-one-out cross-validation strategy. Taking the HER2+ dataset as an example, for each section, we train the model on the remaining 31 sections and validate it on that leave-out section.

HGGEp model

In contrast to prior methods, HGGEp significantly improves the prediction accuracy of gene expression from histology images. The detailed structure of HGGEp model is illustrated below.

Gradient enhancement module

To enrich the cell morphological information closely associated with gene expression, this paper proposes the GEM, which enhances the model's perception of cell morphology through difference convolution [44, 45]. This module comprises two key components: 1) the convolution process that convolves the sub-images; 2) the gradient enhancement process that enhances the cell morphology information by difference convolution.

Figure 1 illustrates the implementation of GEM with the steps of convolution and difference operations. Specifically, a 3×3 convolutional kernel is applied to a 5×5 input split histology image, resulting in a downsampled feature map of size 3×3 . Subsequently, a gradient enhancement operation is performed on the resulting feature map, aiming to utilize the differences between neighboring pixels in the feature map to enhance its gradient information. Traditional convolution process can be expressed as:

$$\text{TC}(p_0) = \sum_{p_n \in R(p_0)} \mathbf{w}_{p_n} \cdot \mathbf{x}_{p_n}, \quad (1)$$

where p_0 represents the central position of the local receptive field $R(p_0)$, p_n denotes the relative position of other pixels within the receptive field $R(p_0) = \{(-1, -1), (-1, 0), \dots, (0, 1), (1, 1)\}$, \mathbf{x}_{p_n} denotes the pixel value at position p_n in the input feature map,

\mathbf{w}_{p_n} is a learnable parameter. To achieve gradient enhancement, we deform the traditional convolution $\text{TC}(p_0)$ to difference convolution $\text{DC}(p_0)$:

$$\text{DC}(p_0) = \sum_{p_n \in R(p_0)} (\mathbf{w}_{p_n} \cdot \mathbf{x}_{p_n}) + \theta \left(-\mathbf{x}_{p_0} \cdot \sum_{p_n \in R(p_0)} \mathbf{w}_{p_n} \right), \quad (2)$$

where θ is a hyperparameter that regulates the balance between semantic and gradient information, and we set it to 0.7 default. When $\theta = 0$, the difference convolution is the same as the traditional convolution operation. Given the input sub-images $\mathbf{x}_{in} \in \mathbb{R}^{N \times 3 \times 112 \times 112}$, each of the convolution (F_{TC} and F_{DC}) exports the $F_{TC} \in \mathbb{R}^{N \times 6 \times 56 \times 56}$ and $F_{DC} \in \mathbb{R}^{N \times 6 \times 56 \times 56}$. The latent features from GEM module are $\mathbf{z}^{GEM} \in \mathbb{R}^{N \times 3 \times 56 \times 56}$. The entire GEM operates as follows:

$$F_{TC} = \text{BN}(\sigma(\text{TC}(\mathbf{x}_{in}))), \quad (3)$$

$$F_{DC} = \text{BN}(\sigma(\text{DC}(\mathbf{x}_{in}))), \quad (4)$$

$$\mathbf{z}^{GEM} = \text{BN}(\sigma(\text{MLP}([F_{TC}, F_{DC}]))), \quad (5)$$

where \mathbf{z}^{GEM} is the enhanced latent feature map, σ is ReLU activation function, BN is Batch Norm. Here the Multi-Layer Perceptron (MLP) layer is used to adjust the number of feature channels, which then serve as input into the subsequent backbone network.

Backbone and multiple latent stage feature

With the GEM-enhanced feature map \mathbf{z}^{GEM} , the backbone network, CBAM [25], and ViT [26] are employed to extract and optimize multiple latent stage feature information from the \mathbf{z}^{GEM} , enhance the model's prowess in latent feature modeling, and capture global relationships among spots at each latent stage.

For the backbone network, we chose a lightweight shufflenet V2 [24] to optimize computing efficiency. This network produces output across five stages, with diminishing latent feature map sizes as the network deepens. For subsequent processing, we leverage features from the last three stages from the backbone, feeding them into the CBAM. As shown in Fig. 1, CBAM utilizes both Channel Attention (CA) and Spatial Attention (SA) for adjusting the weight of each channel/spot by considering global information. With these attention mechanisms, CBAM enhances the network's comprehension of vital features in the image and improves the model's performance and generalization capabilities.

Specifically, the GEM output features $\mathbf{z}^{GEM} \in \mathbb{R}^{N \times 3 \times 56 \times 56}$ are first fed into the shufflenet V2 backbone network to get the embeddings at five latent stages [$\mathbf{S}_1 \in \mathbb{R}^{N \times 24 \times 29 \times 29}$, $\mathbf{S}_2 \in \mathbb{R}^{N \times 48 \times 15 \times 15}$, $\mathbf{S}_3 \in \mathbb{R}^{N \times 96 \times 8 \times 8}$, $\mathbf{S}_4 \in \mathbb{R}^{N \times 192 \times 4 \times 4}$, $\mathbf{S}_5 \in \mathbb{R}^{N \times 1024 \times 4 \times 4}$]. The embeddings of the last three latent stages are selected for subsequent analysis. Subsequently, we input these embedding features into the CBAM to optimize the features of each latent stage separately. Specifically, for \mathbf{S}_i , $i \in \{3, 4, 5\}$, the CBAM module performs as below:

$$\text{CA}(\mathbf{S}_i) =$$

$$\sigma \left(\text{MLP}_{\text{channel}} \left(\text{AvgPool}_{\text{spatial}}(\mathbf{S}_i) \oplus \text{MaxPool}_{\text{spatial}}(\mathbf{S}_i) \right) \right), \quad (6)$$

$$\mathbf{h}_i^{\text{tmp}} = \mathbf{S}_i \odot \text{CA}(\mathbf{S}_i), \quad (7)$$

$$\text{SA}(\mathbf{h}_i^{\text{tmp}}) =$$

$$\sigma \left(\text{Conv}_{7 \times 7} \left(\left[\text{AvgPool}_{\text{channel}}(\mathbf{h}_i^{\text{tmp}}), \text{MaxPool}_{\text{channel}}(\mathbf{h}_i^{\text{tmp}}) \right] \right) \right), \quad (8)$$

$$\mathbf{h}_i = \mathbf{h}_i^{\text{tmp}} \odot \text{SA}(\mathbf{h}_i^{\text{tmp}}), \quad (9)$$

where the symbol \oplus denotes element-wise addition, \odot represents element-wise multiplication, and σ is the ReLU activation function. $\text{AvgPool}_{\text{spatial}}(\mathbf{S}_i)$ and $\text{MaxPool}_{\text{spatial}}(\mathbf{S}_i)$ denote the average and maximum pooling operations on the embedding feature \mathbf{S}_i , respectively, and the output of both is $\mathbb{R}^{N \times 192 \times 1 \times 1}$. As shown in equation (6), the outputs of the two pooling operations are element-wise added. They subsequently enter the $\text{MLP}_{\text{channel}}$ for channel scaling, with the scaling factor set to 16 by default (equation 6). For spatial attention $\text{SA}(h_i^{\text{tmp}})$, $\text{AvgPool}_{\text{channel}}(h_i^{\text{tmp}})$ and $\text{MaxPool}_{\text{channel}}(h_i^{\text{tmp}})$ perform average and maximum pooling for the channels, respectively, followed by the 7×7 convolution. It is worth noting that CBAM-optimized hidden features maintain their original input dimension, i.e. the hidden features h_i after CBAM optimization are $h_3 \in \mathbb{R}^{N \times 96 \times 8 \times 8}$, $h_4 \in \mathbb{R}^{N \times 192 \times 4 \times 4}$, $h_5 \in \mathbb{R}^{N \times 1024 \times 4 \times 4}$.

To facilitate subsequent processing, we uniformly adjust the above three hidden features $[h_3, h_4, h_5]$ to $\mathbb{R}^{N \times 1024}$ by linear and dimension transformation. These hidden features are fed into the ViT module separately and its self-attention mechanism is utilized to capture the associations among the spots in WSI. For h_i , $i \in \{3, 4, 5\}$, the ViT module is implemented as follows:

$$\mathbf{z}_i^{\text{tmp}} = \text{Multihead}(\mathbf{h}_i) = \text{Concat}(\text{head}_1(\mathbf{h}_i), \dots, \text{head}_8(\mathbf{h}_i)) \mathbf{W}^O, \quad (10)$$

For each head $t \in \{1, \dots, 8\}$, $\text{head}_t(\mathbf{h}_i) =$

$$\text{Attention}(\mathbf{Q}_{h_i}, \mathbf{K}_{h_i}, \mathbf{V}_{h_i}) = \text{Softmax}\left(\frac{\mathbf{Q}_{h_i} \mathbf{K}_{h_i}^T}{\sqrt{d_k^{h_i}}}\right) \mathbf{V}_{h_i}, \quad (11)$$

where $\mathbf{Q}_{h_i} = \mathbf{h}_i \mathbf{W}_q^h$, $\mathbf{K}_{h_i} = \mathbf{h}_i \mathbf{W}_k^h$, $\mathbf{V}_{h_i} = \mathbf{h}_i \mathbf{W}_v^h$, \mathbf{h}_i represents one of the input features $[h_3, h_4, h_5]$, \mathbf{W}^O , \mathbf{W}_q^h , \mathbf{W}_k^h , \mathbf{W}_v^h denote the learnable weight matrices, \mathbf{Q}_{h_i} , \mathbf{K}_{h_i} , \mathbf{V}_{h_i} are the matrices of queries, keys, and values obtained by linear transformation of \mathbf{h}_i , respectively. The parameter $d_k^{h_i}$ is used to scale the denominator in the dot-product attention, controlling the scaling of attention weights. These operations enable the model to dynamically model the associations among spots, leading to an optimal allocation of attention. After the optimization of the multi-head attention mechanism, the dimension of the ViT output $[\mathbf{z}_3^{\text{tmp}}, \mathbf{z}_4^{\text{tmp}}, \mathbf{z}_5^{\text{tmp}}]$ remain constant at $\mathbb{R}^{N \times 1024}$.

Finally, a Feed-Forward Network (FFN) is applied for additional non-linear transformations and representation learning of the hidden features from each stage:

$$\mathbf{z}_i^{\text{ViT}} = \text{FFN}(\mathbf{z}_i^{\text{tmp}}) = \text{ReLU}(\mathbf{z}_i^{\text{tmp}} \mathbf{W}^1 + b^1) \mathbf{W}^2 + b^2, \quad (12)$$

where \mathbf{W}^i and b^i are the weight matrix and bias vector, respectively. After ViT optimization, it is obtained $[\mathbf{z}_3^{\text{ViT}}, \mathbf{z}_4^{\text{ViT}}, \mathbf{z}_5^{\text{ViT}}] \in \mathbb{R}^{N \times 1024}$.

Hypergraph association module

The feature representations of spots at each latent stage are derived by encoding histology images through the model's front-end image encoder, which encompasses the GEM, Backbone, CBAM, and ViT modules. However, the above modules are limited to treating the features of each stage independently and do not explore the association among the multiple latent stage features $[\mathbf{z}_3^{\text{ViT}}, \mathbf{z}_4^{\text{ViT}}, \mathbf{z}_5^{\text{ViT}}]$. Herein, we first fuse the multiple latent feature maps described above, and then introduce a HAM aimed at capturing high-order association among spots. To comprehensively model feature associations at varying distances, we adopt a global modeling approach based on Euclidean distance and a local modeling approach based on nearby positions.

Specifically, with the summed element by element of $[\mathbf{z}_3^{\text{ViT}}, \mathbf{z}_4^{\text{ViT}}, \mathbf{z}_5^{\text{ViT}}]$, $\mathbf{H}_{\text{in}} \in \mathbb{R}^{N \times 1024}$ is obtained and fed into the HAM. We use \mathbf{v}_i as the node and the attributes of the node \mathbf{v}_i is $\mathbf{m}_i \in \mathbb{R}^{1 \times 1024}$. To establish hyperedges among nodes, we propose a method that combines Euclidean distance and nearby positions metrics to generate an incidence matrix. Initially, we measure the Euclidean distance between the current node \mathbf{v}_i and other nodes. This effectively models relationships between distant nodes \mathbf{v}_i and \mathbf{v}_j :

$$\text{Dis}(\mathbf{v}_i, \mathbf{v}_j) = \sqrt{\sum_{k=1}^{1024} (\mathbf{m}_{ik} - \mathbf{m}_{jk})^2}. \quad (13)$$

Simultaneously, considering that neighboring nodes on the coordinates usually have similar genetic phenotypes, we also build hyperedges based on the positional relationship between nodes. Assume that the coordinates of node \mathbf{v}_i are (x_i, y_i) and the coordinates of node \mathbf{v}_j are (x_j, y_j) , the positional relationship is encoded as:

$$\text{Pos}(\mathbf{v}_i, \mathbf{v}_j) = \sqrt{(x_i - x_j)^2 + (y_i - y_j)^2}. \quad (14)$$

By combining the two aforementioned metrics, we establish the relationship between the current node and other nodes. To ensure a balanced contribution of the two metrics to the final incidence matrix, normalization method is applied:

$$\text{Inc_Mat}(\mathbf{v}_i, \mathbf{v}_j) = \text{Norm}(\text{Dis}(\mathbf{v}_i, \mathbf{v}_j)) + \text{Norm}(\text{Pos}(\mathbf{v}_i, \mathbf{v}_j)). \quad (15)$$

Next, we construct the hyperedges based on the incidence matrix. For each node, we find the top K nearest nodes based on the incidence matrix, and then construct a hyperedge from these K nodes. The hyperedges, along with their attributes and features $\mathbf{H}_{\text{in}} \in \mathbb{R}^{N \times 1024}$, are input into a hypergraph convolutional network. Throughout the convolution process, the model enhances its understanding of inter-node dependencies, encompassing relationships among nodes as well as between nodes and hyperedges. For the concrete implementation of hypergraph convolution [28, 29], this study leverages the HypergraphConv module from the torch_geometric library:

$$\mathbf{H}_{\text{tmp}} = \text{HypergraphConv}(\mathbf{H}_{\text{in}}, \mathbf{v}_i, \mathbf{m}_i), \quad (16)$$

$$\mathbf{H}_{\text{out}} = \text{Dropout}(\text{Norm}(\text{ReLU}(\mathbf{H}_{\text{tmp}}))). \quad (17)$$

The output feature obtained after the HAM module is $\mathbf{H}_{\text{out}} \in \mathbb{R}^{N \times 1024}$.

Finally, to obtain the final gene prediction, a LSTM module and MLP module are introduced in this paper. LSTM models the association among features by treating the input multiple latent stage features as time series. The features of the four stages $[\mathbf{z}_3^{\text{ViT}}, \mathbf{z}_4^{\text{ViT}}, \mathbf{z}_5^{\text{ViT}}, \mathbf{H}_{\text{out}}]$ are concatenated together to obtain $\mathbf{L}_{\text{in}} \in \mathbb{R}^{4 \times N \times 1024}$, which is used as an input to the LSTM. Subsequently, the first dimension of the LSTM output features is average pooled to obtain the average vector $\mathbf{L}_{\text{out}} \in \mathbb{R}^{N \times 1024}$. Finally, MLP is used to map the latent feature $\mathbf{L}_{\text{out}} \in \mathbb{R}^{N \times 1024}$ to the number of final predicted genes, which is $\mathbb{R}^{N \times 785}$ for HER2+ and $\mathbb{R}^{N \times 171}$ for cSCC dataset.

Model hyperparameter configuration

The hyperparameters used in the HGGE model are listed specifically. (i) GEM: the parameter θ in Equation 2 is set to 0.7 and the feature channel variations progress from $3 \rightarrow 6 \rightarrow 3$. (ii) The backbone network: we utilize the shufflenet_v2_x0_5 model, pre-trained on the ImageNet dataset, to extract features from the

final three stages for subsequent modules. (iii) CBAM: the scaling factor in the channel attention in Equation 6 is configured to shrink by $16\times$. Subsequently, for ease of subsequent processing, we reduce these features from the last three stages to $\mathbb{R}^{n\times 1024}$, where n represents the number of spots. (iv) ViT: the whole ViT module is iterated four times. The number of attention heads is set to 8, and the FeedForward component includes two MLP layers. (v) HAM: when constructing the incidence matrix, a fixed number of adjacent nodes is set to 3, and nodes retain their own indices. The convolution process involves two layers of hypergraph convolution, with the hidden layer being half the size of the input. (vi) LSTM: This encompasses input and hidden state dimensions of 1024 each, along with a layer count of 4. (vii) MLP: After output normalization, a single MLP layer is directly employed to predict the final gene output.

Loss function

Mean Squared Error (MSE) loss is utilized to measure the average squared difference between the predicted gene expression values and the ground truth for each gene. Meanwhile, given the prevalent zeros in the spatial transcriptomics data, we also include the Zero-Inflated Negative Binomial (ZINB [46]) loss. This loss is rooted in the zero-inflated negative binomial distribution, amalgamating the negative binomial distribution with a zero-inflation component. The total loss of the HGGE model is a fusion of MSE and ZINB loss, for the task of gene prediction from histology images.

Evaluation metrics

To comprehensively evaluate the model's performance, this paper utilizes three metrics: PCC, SSIM, and RMSE.

PCC is used to assess the performance of benchmarking models. This coefficient, a statistical measure for assessing the linear relationship between two continuous variables, is widely employed to quantify the strength and direction of the linear correlation between these variables. The PCC ranges from -1 to 1 , with the progression from -1 to 1 indicating a shift from negative to positive correlation. The specific calculation method is as follows:

$$\text{PCC} = \frac{\sum_{i=1}^n (\mathbf{x}_i - \bar{\mathbf{x}})(\mathbf{y}_i - \bar{\mathbf{y}})}{\sqrt{\sum_{i=1}^n (\mathbf{x}_i - \bar{\mathbf{x}})^2 \sum_{i=1}^n (\mathbf{y}_i - \bar{\mathbf{y}})^2}}, \quad (18)$$

where $\bar{\mathbf{x}}$ and $\bar{\mathbf{y}}$ represent the means of variables \mathbf{x} and \mathbf{y} , respectively, and \mathbf{x}_i and \mathbf{y}_i denote the i_{th} observations.

SSIM is a perceptual metric that quantifies data quality degradation caused by processing such as data compression or transmission. Unlike traditional metrics like Mean Squared Error (MSE), SSIM considers changes in structural information, luminance, and contrast. SSIM values range from -1 to 1 , where a value closer to 1 indicates that the two images are highly similar. The SSIM calculation involves three components: luminance (l), contrast (c), and structure (s). The formula is given by:

$$\text{SSIM}(\mathbf{x}, \mathbf{y}) = [l(\mathbf{x}, \mathbf{y})]^\alpha \cdot [c(\mathbf{x}, \mathbf{y})]^\beta \cdot [s(\mathbf{x}, \mathbf{y})]^\gamma \quad (19)$$

Typically, the values for α , β , and γ are set to 1 , leading to the simplified formula:

$$\text{SSIM}(\mathbf{x}, \mathbf{y}) = \frac{(2\mu_x\mu_y + C_1)(2\sigma_{xy} + C_2)}{(\mu_x^2 + \mu_y^2 + C_1)(\sigma_x^2 + \sigma_y^2 + C_2)} \quad (20)$$

where μ_x and μ_y are the mean intensities of images \mathbf{x} and \mathbf{y} , σ_x^2 and σ_y^2 are the variances, σ_{xy} is the covariance, and C_1 and C_2 are small constants to stabilize the division.

RMSE is used to evaluate the performance of benchmarking models by measuring the differences between predicted and actual values. RMSE is a widely-used statistical metric that quantifies the magnitude of prediction errors. The RMSE value is always non-negative, with a lower RMSE indicating better model performance. The RMSE is calculated as follows:

$$\text{RMSE} = \sqrt{\frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2} \quad (21)$$

Where y_i represents the actual observed values, \hat{y}_i represents the predicted values, and n is the number of observations. The formula takes the square root of the average of the squared differences between the predicted and actual values, providing a measure of the model's predictive accuracy.

These metrics, PCC, SSIM, and RMSE, together provide a comprehensive evaluation of the model's performance, capturing different aspects of prediction accuracy and data quality.

Training parameters

During training, the model utilizes a learning rate of 0.00001 , runs for a maximum of 400 epochs, and undergoes testing every five epochs. The experiments are conducted on an Ubuntu 20.04 system equipped with 128GB of RAM and an A6000 GPU featuring 48GB of memory.

Key Points

- We develop a novel histology image-based gene prediction model named HGGE, which demonstrates high accuracy and robust performance.
- To reveal the intricate relationship between cell morphology and gene expression in images, we propose a gradient enhancement module, which effectively improves the model's capability in perceiving cell morphology in images.
- HGGE includes a hypergraph module that efficiently models higher-order associations among features across multiple latent stages, resulting in significant performance improvement.

Supplementary data

Supplementary data is available at *Briefings in Bioinformatics* online.

Conflict of interest: The authors declare no competing interests.

Funding

G.W. is supported by the National Institute of General Medical Sciences of the National Institutes of Health (1R35GM150460).

Data availability

The spatial transcriptomics datasets used in this study include the [1] HER2-positive breast tumor ST datasets, which are available at <https://github.com/almaan/her2st/>; [2] 10x Visium data of

human cutaneous squamous cell carcinoma are publicly available in the Gene Expression Omnibus (GSE144240).

Code availability

All source codes and trained models in our experiments have been deposited at <https://github.com/QSong-github/HGGEP>.

Materials and correspondence

Correspondence and requests for materials should be addressed to QS or YZ.

References

1. Asp M, Giacomello S, Larsson L. et al. A spatiotemporal organ-wide gene expression and cell atlas of the developing human heart. *Cell* 2019;**179**:1647–1660.e19. <https://doi.org/10.1016/j.cell.2019.11.025>.
2. Maynard KR, Collado-Torres L, Weber LM, et al. Transcriptome-scale spatial gene expression in the human dorsolateral prefrontal cortex. *Nature Neuroscience* 2021;**24**:425–436.
3. Moncada R, Barkley D, Wagner F. et al. Integrating microarray-based spatial transcriptomics and single-cell RNA-seq reveals tissue architecture in pancreatic ductal adenocarcinomas. *Nat Biotechnol* 2020;**38**:333–42. <https://doi.org/10.1038/s41587-019-0392-8>.
4. Maniatis S, Åijö T, Vickovic S. et al. Spatiotemporal dynamics of molecular pathology in amyotrophic lateral sclerosis. *Science* 2019;**364**:89–93. <https://doi.org/10.1126/science.aav9776>.
5. Ståhl PL, Salmén F, Vickovic S. et al. Visualization and analysis of gene expression in tissue sections by spatial transcriptomics. *Science* 2016;**353**:78–82. <https://doi.org/10.1126/science.aaf2403>.
6. Schmauch B, Romagnoni A, Pronier E. et al. A deep learning model to predict RNA-Seq expression of tumours from whole slide images. *Nat Commun* 2020;**11**:3877. <https://doi.org/10.1038/s41467-020-17678-4>.
7. Tolios A, de Las Rivas J, Hovig E. et al. Computational approaches in cancer multidrug resistance research: identification of potential biomarkers, drug targets and drug-target interactions. *Drug Resist Updat* 2020;**48**:100662. <https://doi.org/10.1016/j.drug.2019.100662>.
8. Zhao E, Stone MR, Ren X. et al. Spatial transcriptomics at sub-spot resolution with BayesSpace. *Nat Biotechnol* 2021;**39**:1375–84. <https://doi.org/10.1038/s41587-021-00935-2>.
9. Hu J, Li X, Coleman K. et al. SpaGCN: integrating gene expression, spatial location and histology to identify spatial domains and spatially variable genes by graph convolutional network. *Nat Methods* 2021;**18**:1342–51. <https://doi.org/10.1038/s41592-021-01255-8>.
10. Xu C, Jin X, Wei S. et al. DeepST: identifying spatial domains in spatial transcriptomics by deep learning. *Nucleic Acids Res* 2022;**50**:e131–1. <https://doi.org/10.1093/nar/gkac901>.
11. Song Q, Su J. DSTG: Deconvoluting spatial transcriptomics data through graph-based artificial intelligence. *Brief Bioinform* 2021;**22**:bbaa414. <https://doi.org/10.1093/bib/bbaa414>.
12. Kleshchevnikov V, Shmatko A, Dann E. et al. Cell2location maps fine-grained cell types in spatial transcriptomics. *Nat Biotechnol* 2022;**40**:661–71. <https://doi.org/10.1038/s41587-021-01139-4>.
13. Biancalani T, Scalia G, Buffoni L. et al. Deep learning and alignment of spatially resolved single-cell transcriptomes with tangram. *Nat Methods* 2021;**18**:1352–62. <https://doi.org/10.1038/s41592-021-01264-7>.
14. Cang Z, Zhao Y, Almet AA. et al. Screening cell-cell communication in spatial transcriptomics via collective optimal transport. *Nat Methods* 2023;**20**:218–28. <https://doi.org/10.1038/s41592-022-01728-4>.
15. Tang Z, Zhang T, Yang B. et al. spaCI: deciphering spatial cellular communications through adaptive graph model. *Brief Bioinform* 2022;**24**:1–13. <https://doi.org/10.1093/bib/bbac563>.
16. He B, Bergensträhle L, Stenbeck L. et al. Integrating spatial gene expression and breast tumour morphology via deep learning. *Nat Biomed Eng* 2020;**4**:827–34. <https://doi.org/10.1038/s41551-020-0578-x>.
17. Pang M, Su K, Li M. Leveraging information in spatial transcriptomics to predict super-resolution gene expression from histology images in tumors. *bioRxiv*. 2021;2021.11.28.470212:1–25.
18. Zeng Y, Wei Z, Yu W. et al. Spatial transcriptomics prediction from histology jointly through transformer and graph neural networks. *Brief Bioinform* 2022;**23**:1–12. <https://doi.org/10.1093/bib/bbac297>.
19. Hoang DT, Dinstag G, Hermida LC. et al. Prediction of cancer treatment response from histopathology images through imputed transcriptomics. *Res Sq* 2023;**41**:1551. https://doi.org/10.1200/JCO.2023.41.16_suppl.1551.
20. Xie R, Pang K, Chung SW. et al. Spatially resolved gene expression prediction from histology images via Bi-modal contrastive learning. *Advances in Neural Information Processing Systems* 2024;**36**:1234–1245.
21. Jia Y, Liu J, Chen L. et al. THItogene: a deep learning method for predicting spatial transcriptomics from histological images. *Brief Bioinform* 2024;**25**:bbad464. <https://doi.org/10.1093/bib/bbad464>.
22. Wu Z, Pan S, Chen F. et al. A comprehensive survey on graph neural networks. *IEEE transactions on neural networks learning systems* 2020;**32**:4–24. <https://doi.org/10.1109/TNNLS.2020.2978386>.
23. Chan AS, Wang C, Fu X. et al. Benchmarking the translational potential of spatial gene expression prediction from histology. *bioRxiv*. 2023;2012.571251:1–25.
24. Ma N, Zhang X, Zheng H-T, Sun Js In: *Proceedings of the European conference on computer vision (ECCV)*. p. 116–31. 2018.
25. Woo S, Park J, Lee J-Y, Kweon IS In: *Proceedings of the European conference on computer vision (ECCV)*. p. 3–19. 2018.
26. Dosovitskiy A. An image is worth 16x16 words: transformers for image recognition at scale. *International Conference on Learning Representations (ICLR)*. 2021;**36**:1–25.
27. Zhou J, Cui G, Hu S. et al. Graph neural networks: a review of methods and applications. *AI open* 2020;**1**:57–81. <https://doi.org/10.1016/j.aiopen.2021.01.001>.
28. Bai S, Zhang F, Torr PH. Hypergraph convolution and hypergraph attention. *Pattern Recognition* 2021;**110**:107637. <https://doi.org/10.1016/j.patcog.2020.107637>.
29. Li B, Zhang Y, Zhang C. et al. Multi-scale hypergraph-based feature alignment network for cell localization. *Pattern Recognition* 2024;**149**:110260. <https://doi.org/10.1016/j.patcog.2024.110260>.
30. Shi X, Chen Z, Wang H. et al. Convolutional LSTM network: a machine learning approach for precipitation nowcasting. *Advances in neural information processing systems* 2015;**28**:802–810.
31. Memory LS-T. Long short-term memory. *Neural Comput* 2010;**9**:1735–80.
32. 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. <https://doi.org/10.1038/s41467-021-26271-2>.
33. Ji AL, Rubin AJ, Thrane K. et al. Multimodal analysis of composition and spatial architecture in human squamous cell

- carcinoma. *Cell* 2020;**182**:497–514. e422. <https://doi.org/10.1016/j.cell.2020.05.039>.
34. Jin X, Zhu L, Cui Z. et al. Elevated expression of GNAS promotes breast cancer cell proliferation and migration via the PI3K/AKT/Snail1/E-cadherin axis. *Clinical and Translational Oncology* 2019;**21**:1207–19. <https://doi.org/10.1007/s12094-019-02042-w>.
 35. Ding H, Zhang X, Su Y. et al. GNAS promotes inflammation-related hepatocellular carcinoma progression by promoting STAT3 activation. *Cell Mol Biol Lett* 2020;**25**:1–17. <https://doi.org/10.1186/s11658-020-00204-1>.
 36. Yang S, Wang-Su S-T, Cai H. et al. Changes in three types of ubiquitin mRNA and ubiquitin-protein conjugate levels during lens development. *Exp Eye Res* 2002;**74**:595–604. <https://doi.org/10.1006/exer.2001.1149>.
 37. Zhang L, Huang Y, Zhuo W. et al. Identification and characterization of biomarkers and their functions for Lapatinib-resistant breast cancer. *Med Oncol* 2017;**34**:1–8. <https://doi.org/10.1007/s12032-017-0953-y>.
 38. Kim J, Choi H. The mucin protein MUCL1 regulates melanogenesis and melanoma genes in a manner dependent on threonine content. *British Journal of Dermatology* 2022;**186**:532–43. <https://doi.org/10.1111/bjd.20761>.
 39. Mémin E, Hoque M, Jain MR. et al. Blocking eIF5A modification in cervical cancer cells alters the expression of cancer-related genes and suppresses cell proliferation. *Cancer Res* 2014;**74**:552–62. <https://doi.org/10.1158/0008-5472.CAN-13-0474>.
 40. Simigdala N, Gao Q, Pancholi S. et al. Cholesterol biosynthesis pathway as a novel mechanism of resistance to estrogen deprivation in estrogen receptor-positive breast cancer. *Breast Cancer Res* 2016;**18**:1–14. <https://doi.org/10.1186/s13058-016-0713-5>.
 41. Ni P-Z, He JZ, Wu ZY. et al. Overexpression of Stathmin 1 correlates with poor prognosis and promotes cell migration and proliferation in oesophageal squamous cell carcinoma. *Oncol Rep* 2017;**38**:3608–18. <https://doi.org/10.3892/or.2017.6039>.
 42. Qiu J, Li L, Sun J. et al. Large ai models in health informatics: applications, challenges, and the future. *IEEE Journal of Biomedical Health Informatics* 2023;**27**:6074–87. <https://doi.org/10.1109/JBHI.2023.3316750>.
 43. Yang Z, Li L, Lin K. et al. The dawn of lmms: preliminary explorations with gpt-4v (ision). arXiv preprint. 2023;2309.17421: 1–25.
 44. Yu Z, Zhao C, Wang Z. et al. Searching Central Difference Convolutional Networks for Face Anti-Spoofing. *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 2020:5295–305.
 45. Yu Z, Qin Y, Zhao H. et al. Dual-cross central difference network for face anti-spoofing. *Proceedings of the Thirtieth International Joint Conference on Artificial Intelligence (IJCAI)*. 2021: 1281–1287.
 46. Eraslan G, Simon LM, Mircea M. et al. Single-cell RNA-seq denoising using a deep count autoencoder. *Nat Commun* 2019;**10**:390. <https://doi.org/10.1038/s41467-018-07931-2>.