A framework for gene representation on spatial transcriptomics

Shenghao Cao[1], Ye Yuan[1,2,3,*]

Affiliations:

[1] Institute of Image Processing and Pattern Recognition, Shanghai Jiao Tong University, and Key Laboratory of System Control and Information Processing, Ministry of Education of China, Shanghai 200240, China.

[2] Key Laboratory of Biopharmaceutical Preparation and Delivery, Chinese Academy of Sciences, Beijing 100190, China.

[3] State Key Laboratory of Biochemical Engineering, Institute of Process Engineering, Chinese Academy of Sciences, Beijing 100190, China.

[*] Corresponding author, e-mail: yyuan@ipe.ac.cn; yuanye_auto@sjtu.edu.cn

Abstract

Recent foundation models for single-cell transcriptomics data generate informative, context-aware gene representations. The spatially resolved transcriptomics data offer extra positional insights, yet corresponding gene representation methods that integrate both intracellular and spatial context are still lacking. Here, we introduce a gene representation framework tailored for spatial transcriptomics data. It incorporates ligand genes within the spatial niche into the transformer encoder of single-cell transcriptomics. We further propose a biased cross-attention method to enable the framework to do learning with single-cell resolution on low-resolution, whole-transcriptome Visium data. We implemented our framework on a hybrid Visium dataset derived from two human tissue types with distinct developmental and disease states, and tested on various downstream applications. Compared with the latest foundation model for single-cell transcriptomics, our spatially informed gene representations could identify cell groups and gene functions more accurately, and could predict the perturbation effects of cell-cell ligand-receptor interactions on downstream targets.

Main

The large language model based on the generative pretrained transformer (GPT) architecture has made groundbreaking progress in the field of natural language processing[1,2]. Recently, the transformer pretraining method was applied to building foundation models for single-cell transcriptomics data[3-5]. The foundation models generated contextual gene representations within single cells. These gene representations served as input for diverse downstream applications, including cell type annotation, perturbation response prediction, gene network inference, and achieved state-of-the-art performances.

Notably, the intracellular transcription is not only regulated by gene networks, but also mediated by extracellular signals. However, the single-cell transcriptomics data do not provide information in the extracellular microenvironment. Fortunately, the spatially resolved transcriptomics (ST) technologies offer the opportunity to profile the gene expressions in their native locations. Here, we adapt the transformer model for ST data to integrate the effects of cell-cell communication via ligand secretion and reception into gene representations (**Fig. 1a and Methods**). The key improvement over the previous foundation models is that we insert a cross-attention module after the self-attention module in the transformer block to compute attentions with ligand genes from niche cells (**Fig. 1a**).

A variety of ST technologies are available with a trade-off between resolution and gene coverage. One of the most commonly used platforms is the 10x Genomics Visium[6]. It captures whole transcriptomes at the arranged spots in the tissue section. However, the spot diameter, i.e. the resolution, is 55 μm, which is several-fold larger than a typical cell size, so each spot may contain multiple cells. Deconvolution methods such as cell2location[7] can estimate the cell-type proportions and cell-type specific expressions at each spot, but the cell number and cell specific expressions remain unknown. To address this problem, we propose the cell-type-wise biased cross-attention method, which enables the ST framework to do single-cell resolution learning on spot-resolution Visium data (**Fig. 1b**). The detailed theory and deductions can be found in the **Methods** section.

We pretrained our framework on a combined corpus of two human Visium datasets (**Supplementary**

Table S1), one comprising 8 colon and terminal ileum tissue sections from adult and fetal[8], the other comprising 28 heart tissue sections from normal myocardium and multiple physiological zones at distinct timepoints after myocardial infarction[9]. The entire corpus includes a total of about 580,000 spatially resolved cells. After pretraining, we visualized the cell embeddings on the heart tissue sections from different physiological zones using the UMAP visualization[10]. We present two of them on **Fig. 2a** and the remaining ones on **Supplementary Figure S1**. The derived clusters were highly consistent with the annotated cell types, achieving higher normalized mutual information than scFoundation[5]. The results indicated that pooling our gene representations could generate cell embeddings that better encode the characteristics of cell states.

NOTCH1 functions as a transmembrane receptor. The NOTCH1-dependent gene network governing cardiac valve disease has been mapped[11]. It was reported that the single-cell transcriptomics foundation model, Geneformer, could be fine-tuned to classify central versus peripheral factors within the NOTCH1-dependent gene network, and boosted predictive accuracy compared to alternative methods[3]. Here, we trained a three-layer perceptron classifier for the same task, using as input the gene embeddings generated through our pretrained model without any fine-tuning. The training and validation were on a normal heart tissue section 10X001. The final performance was comparable to Geneformer with fine-tuning (AUROC 0.81 vs 0.81)[3]. We also trained the classifier using the gene embeddings from another single-cell transcriptomics foundation model, scFoundation[5], and its fine-tuned version (Methods). As can be seen in **Fig. 2b**, our method consistently outperformed them in most values of false positive rate. In the task of classifying the NOTCH1 downstream targets versus non-targets, similarly we trained another three-layer perceptron classifier. Our method also outperformed the scFoundation methods (**Fig. 2c**), and even surpassed the fine-tuning Geneformer (AUROC 0.83 vs 0.81)[3]. These results suggested that our spatially informed gene representations encoded the hierarchy in the receptor NOTCH1-dependent gene network more accurately.

We further evaluated if our gene representation could more accurately encode the gene set of the entire receptor-dependent gene network. We took the TGF-beta signaling pathway as an example, which exhibited increased activity in the fibrotic samples[9]. The TGF-beta signaling pathway is

initiated by the binding of TGF-beta ligands to their receptors. We trained a three-layer perceptron network to classify the member versus non-member genes of the GSEA[12] gene set "signaling by TGFB family members". The training and validation were on a fibrotic heart tissue section ACH005. The results showed that our method indeed outperformed the scFoundation methods (AUROC 0.74 vs 0.69 and 0.68) (**Fig. 2d**).

We next evaluated if our gene representation encoded the ligand-receptor connections. Specifically, we concatenated the pretrained gene embeddings of ligand-receptor pairs and fed into a three-layer perceptron to classify the interacting versus non-interacting pairs (**Methods**). The train and test sets were completely separated in the fivefold cross-validation to avoid any information leakage (**Methods**). In a randomly sampled dataset of ligand-receptor interaction pairs, our method showed a remarkable boost (>30%) over the scFoundation methods, which nearly failed to complete the task with an AUROC around 0.5 (**Fig. 2e**). Furthermore, in another randomly sampled dataset, the average AUROC of our method is 30% higher than those of the scFoundation methods (**Supplementary Figure S2**).

Finally, we tested whether our gene representation captured the response of cell-cell signaling via ligand-receptor interactions. It was reported that the WNT signaling pathway plays a pivotal role in cardiac remodeling[13]. Thus, we chose an ischemic heart tissue section ACH0014 and in silico downregulated the gene expressions of ligands and receptors in the WNT signaling pathway (Methods). The deleterious effects were predicted by the alterations of gene embeddings as shown in **Fig. 2f**. The WNT targets were impacted more significantly than the NOTCH1 targets and housekeeping genes (**Supplementary Table S2**). In silico downregulation of WNT ligands or receptors respectively also impacted WNT targets significantly more than housekeeping genes, but significantly less than joint downregulation of both ligands and receptors (**Supplementary Table S2**). We tested on another ischemic heart tissue section and obtained consistent results (**Supplementary Figure S3**).

We have demonstrated that incorporating ligand expression information from the spatial niche into transformer architecture enhances the performance of gene representations in a series of

downstream tasks. Although the data size of our pretraining corpus was lower than existing foundation models of single-cell data, our gene representations were still superior on these tasks. Our framework can also be utilized for conducting in silico perturbation analysis in the context of ligand-receptor mediated cell-cell interactions. As a result, we speculate that implementing our framework on a large-scale dataset could establish a foundation model for spatial transcriptomics data.

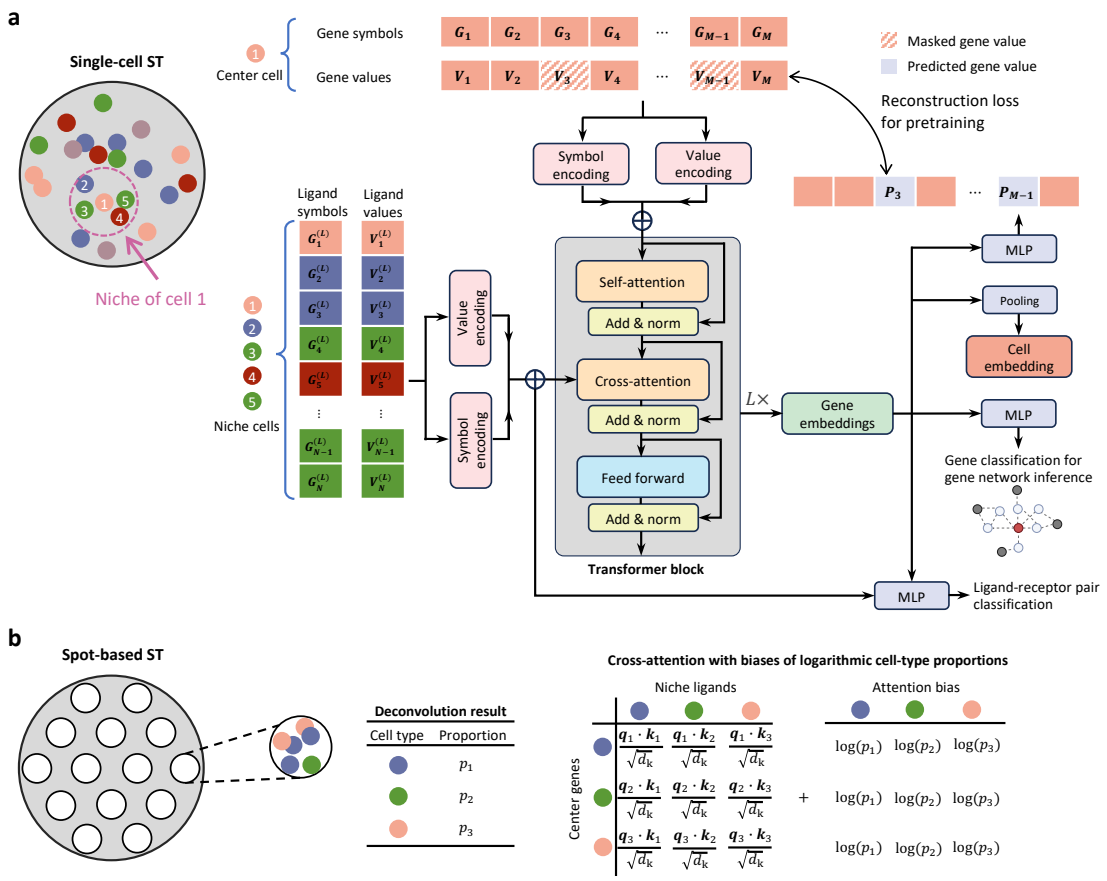

Fig.1 | Architecture of the spatially informed gene representation framework. a, For the spatial transcriptomics (ST) data with single-cell resolution, we regard each cell as a centric node and the neighborhood cells within a preset radius as its niche. An input sample comprises gene symbols and values of the center cell as well as ligand gene symbols and values of the niche cells. The samples are fed into a stack of $L = 6$ identical transformer block layers and output the embeddings of genes in the center cell. The model parameters are pretrained by minimizing the reconstruction loss of the masked gene values. Then, the gene embeddings can serve as input for multiple downstream tasks. **b,** For the ST data based on spot RNA sequencing, we deconvolve the cell-type composition at each spot first. Then, adding a bias of logarithmic cell-type proportion to each attention score in cross-attention can unify the case within the single-cell framework in **a**. See **Methods** for details.

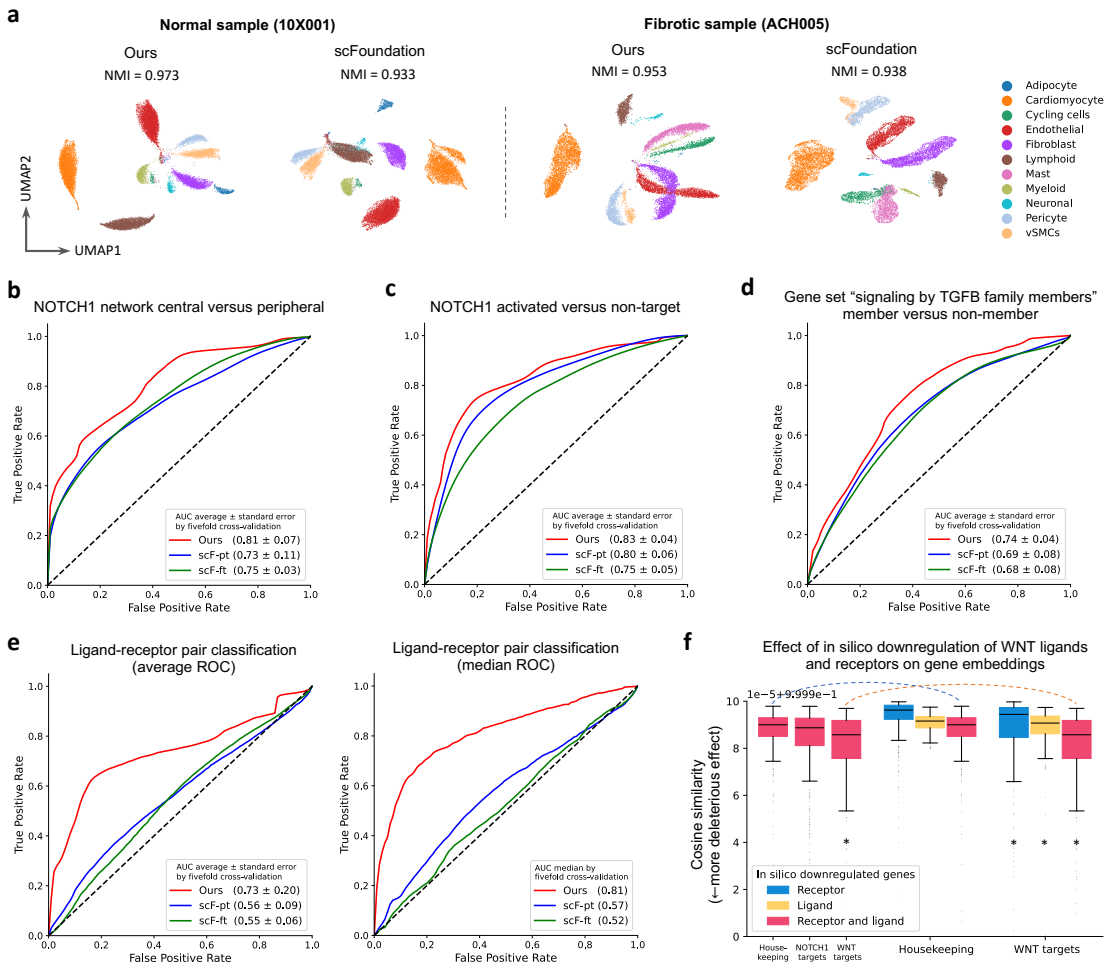

Fig.2 | Results of downstream tasks based on the gene representations. a, Clustering of the cell embeddings, which are max-pooling of gene embeddings, compared to the results of scFoundation. The cell type annotation is indicated by color. Normal heart tissue section 10X001 and fibrotic heart tissue section ACH005 were analyzed. NMI, normalized mutual information. **b,** Comparison on the task of classifying central versus peripheral genes within the NOTCH1-dependent gene network. The comparison was performed on a normal heart tissue section 10X001. scF-pt, pretrained scFoundation; scF-ft, fine-tuned scFoundation. **c,** Comparison on the task of classifying NOTCH1 activated versus non-target genes. The comparison was performed on a normal heart tissue section 10X001. **d,** Comparison on the task of classifying member versus non-member genes of the gene set "signaling by TGFB family members". The comparison was performed on a fibrotic heart tissue section ACH005. **e,** Comparison on the task of classifying interacting versus non-interacting ligand-receptor pairs. The comparison was performed on a normal heart tissue section 10X001. **f,** In silico downregulation of WNT ligands and receptors impacted significantly more on WNT targets than on housekeeping genes and NOTCH1 targets. In silico downregulation of WNT ligands or receptors respectively impacted significantly more on WNT targets than on housekeeping genes. In silico downregulation of WNT ligands and receptors together impacted significantly more on WNT targets than separate downregulation of ligands or receptors. The test was performed on an ischemic heart tissue section ACH0014. *P<0.05 Wilcoxon, FDR-corrected.

Methods

Framework architecture

Assumptions. Suppose that the ST data we are analyzing are resolved at the single-cell level. Otherwise, if the ST data are generated through spot RNA sequencing, such as with the 10x Genomics Visium platform, we can follow the methods described in the next subsection to convert them to single-cell scenarios. Centered on each cell in the ST data, the neighborhood cells within a preset radius are defined as the cell's niche. By default, the radius is set to half of 55μm, the spot diameter of the Visium platform.

Layers. The schematic overview of the framework architecture is illustrated in Fig. 1a. It is composed of six identical transformer blocks, each composed of a self-attention layer, cross-attention layer, and feed-forward neural network layer with the following parameters: 768 embedding dimensions, 12 attention heads per layer, and feed-forward size of 3072.

Input. The input sample consists of the gene symbols and values of the center cell, as well as the ligand gene symbols and values of the niche cells. We constructed a ligand gene library by integrating seven knowledge databases[14-20] and only used the ligand genes included in at least two databases. We limit the input to genes with non-zero expressions. For each gene, the symbol embedding is retrieved from a learnable lookup table with a vocabulary size of 19,264. The embeddings in the lookup table are initialized by those from the scFoundation[5] model. The gene value is transformed into a learnable value embedding through a scalar encoding module borrowed from scFoundation[5]. The final input embedding for each gene is the sum of its symbol embedding and value embedding.

Output. The transformer block first calculates the self-attention among all gene embeddings within the center cell, then computes the cross-attention between all gene embeddings of the center cell and ligand gene embeddings of the niche cells, and finally propagates the gene embeddings through a two-layer feed-forward neural network. The model ultimately outputs gene embeddings specific to both the center cell's intracellular context and spatial niche.

Cell-type-wise biased cross-attention method on Visium data

Visium data consists of spot-wise gene expression profiles, with each spot covering multiple cells. We initially used the deconvolution method cell2location[7] with default settings to disentangle the mixed RNA measurements at each spot. Cell2location outputs the proportions of different cell types and cell-type specific expression profiles at each spot. To utilize the framework described in the previous subsection, we developed the cell-type-wise biased cross-attention method based on the following two assumptions.

Assumption 1. Cells of the same type within a spot approximately have the same expression profiles.

Assumption 2. The niche of each cell in a spot approximates to the cellular composition of that spot.

The assumptions are reasonable since the spot size 55μm is relatively small, several-fold larger than a typical cell size, and *Assumption 1* is also adopted by most cell-type deconvolution methods. Under these two assumptions, calculating cross-attention using cell-wise ligand gene embeddings is equivalent to using cell-type-wise ligand gene embeddings with biases of logarithmic cell-type proportions, so that our framework can do learning with single-cell resolution on the low-resolution Visium data. The details are described as follows.

Suppose that a total of *n* cells from *t* different cell types are located in a given spot, and the cell-type proportions estimated by cell2location are $p_1, p_2, \cdots, p_t$ respectively. For the sake of simplicity and without loss of generality, we suppose only one ligand gene l is detected. According to *Assumption 1*, the np_i cells of the *i*th cell type approximately have identical ligand gene expression, which is estimated by cell2location, resulting in the same ligand gene embedding, denoted as $\boldsymbol{E}_i^l \in \mathbb{R}^{d_e}$ $(i = 1,2,\cdots,t)$.

For any cell **c** in the spot, according to *Assumption 2*, the niche cells centered on cell **c** are approximately the *n* cells in the spot. Denote an arbitrary gene embedding of cell **c** as $\boldsymbol{E}^c \in \mathbb{R}^{d_e}$. Then, the cross-attention between $\boldsymbol{E}^c$ and the ligand gene embeddings of its niche cells $\boldsymbol{E}_i^l$ $(i = 1,2,\cdots,t)$ can be calculated as follows.

First, the gene embeddings are projected to the query, key, and value spaces respectively using the

learnable parameter matrices $W^Q \in \mathbb{R}^{d_k \times d_e}$, $W^K \in \mathbb{R}^{d_k \times d_e}$, and $W^V \in \mathbb{R}^{d_v \times d_e}$.

$$\boldsymbol{q} = W^Q \boldsymbol{E}^c$$
$$\boldsymbol{k}_i = W^K \boldsymbol{E}_i^l \quad (i = 1,2,\cdots,t)$$
$$\boldsymbol{v}_i = W^V \boldsymbol{E}_i^l \quad (i = 1,2,\cdots,t)$$

Then, the keys and values are packed together into matrices K and V, where $\mathbf{1}_{np_i}$ is an np_i-dimensional vector with all elements being 1 and the prime symbol denotes matrix transpose.

$$K = \left(\boldsymbol{k}_1 \mathbf{1}'_{np_1}, \cdots, \boldsymbol{k}_t \mathbf{1}'_{np_t}\right)$$
$$V = \left(\boldsymbol{v}_1 \mathbf{1}'_{np_1}, \cdots, \boldsymbol{v}_t \mathbf{1}'_{np_t}\right)$$

Finally, the attention function between $\boldsymbol{q}$ and K, V are computed.

$$\begin{aligned}
\text{Attention}(\boldsymbol{q}, K, V) &= \text{softmax}\left(\frac{\boldsymbol{q}'K}{\sqrt{d_k}}\right)V' \\
&= \text{softmax}\left(\frac{\boldsymbol{q}'\boldsymbol{k}_1\mathbf{1}'_{np_1}}{\sqrt{d_k}}, \cdots, \frac{\boldsymbol{q}'\boldsymbol{k}_t\mathbf{1}'_{np_t}}{\sqrt{d_k}}\right) \cdot \left(\boldsymbol{v}_1\mathbf{1}'_{np_1}, \cdots, \boldsymbol{v}_t\mathbf{1}'_{np_t}\right)' \\
&= \text{softmax}\left(\frac{\boldsymbol{q}'\boldsymbol{k}_1}{\sqrt{d_k}} + \log(p_1), \cdots, \frac{\boldsymbol{q}'\boldsymbol{k}_t}{\sqrt{d_k}} + \log(p_t)\right) \cdot (\boldsymbol{v}_1, \cdots, \boldsymbol{v}_t)' \\
&= \text{softmax}\left(\frac{\boldsymbol{q}'\widetilde{K}}{\sqrt{d_k}} + bias\right)\tilde{V}'
\end{aligned}$$

where $\widetilde{K} = (\boldsymbol{k}_1, \cdots, \boldsymbol{k}_t)$, $\tilde{V} = (\boldsymbol{v}_1, \cdots, \boldsymbol{v}_t)$ are the cell-type-wise keys and values; $bias = (\log(p_1), \cdots, \log(p_t))$ is a bias term composed of logarithmic cell-type proportions. The detailed deductions of the above equations can be found in **Supplementary Note S1**.

Pretraining

The pretraining was accomplished using a masked learning objective. Specifically, 15% of the genes in the center cells were masked. Their value embeddings were replaced with mask embeddings and then added with the symbol embeddings to form the input embeddings. A three-layer perceptron block was added after the framework architecture to predict the expression values of the masked genes according to their output gene embeddings. The pretraining objective was to minimize the

mean squared error (MSE) loss between the predicted and true values of the masked genes.

The pretraining data corpus was comprised of two human Visium datasets. One was from developmental intestine and adult colon tissue[8]; the other was from myocardial tissue after infarction and controls[9] (details provided in **Supplementary Table S1**). The total cell number was about 0.58 million. The model was pretrained with the following hyperparameters: batch size, 8; max learning rate 1×10^{-4}; learning scheduler, decay by 0.9 every epoch; optimizer, Adam. The train and valid loss of pretraining are shown in **Supplementary Figure S4**.

Downstream tasks

Clustering of cell embeddings. The cell embedding for each cell was generated by max-pooling its output gene embeddings of the sample centered on the cell. For comparison, we also computed the cell embeddings from the scFoundation model. To evaluate the consistency between clustering results of the cell embeddings and annotated cell types, the normalized mutual information (NMI) metrics were calculated using the Python implementation of scib.metrics[21]. The evaluations were conducted on heart tissue sections from different physiological zones in the myocardial infarction Visium dataset[9].

Classification of hierarchy and membership within receptor-dependent gene networks. We used a separate three-layer perceptron network to perform the classification tasks. The input was the output gene embeddings of single-cell samples by the pretrained models. The objective was to minimize the cross-entropy loss between predicted gene type probabilities and ground truth labels.

We evaluated on three datasets respectively using fivefold cross-validation, and the results were reported as AUROC $\pm$ standard deviation. The first two datasets were based on the NOTCH1 receptor-dependent gene network, with genes labeled as central or peripheral, and as activated or non-target[3]. The evaluation ST samples were obtained from a normal tissue section (10X001) of the myocardial infarction Visium dataset[9]. The third dataset focused on the TGF-beta receptor-dependent gene network, with genes labeled as member or non-member. The member genes were derived from the GSEA[12] gene set "signaling by TGFB family members" with ligand genes excluded.

The evaluation samples were obtained from a fibrotic tissue section (ACH005) of the myocardial infarction Visium dataset[9].

For comparison, we also evaluated the classification performance using gene embeddings generated by a single-cell transcriptomics foundation model, scFoundation[5], and its fine-tuned version. The fine-tuned version was obtained by initializing the model with the pretrained weights and training on the myocardial infarction Visium dataset for three epochs using the masked learning objective of pretraining. The train and valid loss are shown in **Supplementary Figure S5**.

Identification of ligand-receptor interaction pairs. We used a separate three-layer perceptron network to classify the interacting and non-interacting ligand-receptor pairs. For a specific ligand-receptor pair, the ligand gene embedding of a single-cell sample was derived by averaging the ligand gene embeddings of all cells within the niche. For Visium data, under the two assumptions we propose in the subsection "Cell-type-wise biased cross-attention method on Visium data", it is equivalent to calculating a weighted average of ligand gene embeddings across cell types, using the cell-type proportions as weights (see deductions in **Supplementary Note S2**). We concatenated the ligand gene embedding with the receptor gene embedding output through the pretrained model and fed into the perceptron classifier. The objective was to minimize the cross-entropy loss between the predicted interacting probabilities and ground truth labels.

We randomly sampled 150 ligand-receptor interaction pairs from the CellTalkDB database[17], and evaluated the classification performance using fivefold cross-validation. Specifically, we respectively split the ligand genes and receptor genes of the 150 interaction pairs into five equal parts. In each of the five iterations, one fold of ligand genes and one fold of receptor genes were picked out as test data, and the interaction pairs between them were labeled as positive pairs. The sample number of negative pairs were randomly selected between non-interacting ligand and receptor genes. The remaining four folds of ligand genes and receptor genes were treated as train data. The corresponding positive and negative pairs were generated in the same way. As a result, we ensured a complete separation of train and test data, with no intersection of ligand genes or receptor genes between them.

The evaluations were performed on single-cell samples obtained from a normal tissue section (10X001) of the myocardial infarction Visium dataset[9]. For comparison, we also evaluated the classification performance using gene embeddings generated by scFoundation[5] and its fine-tuned edition. Since the input samples used by scFoundation did not include ligand genes of niche cells, the classifier input used intracellular ligand and receptor gene embeddings generated by the model.

In silico perturbation analysis of ligand-receptor interaction. We attempted to employ our pretrained model to simulate the perturbation effects of ligand-receptor interaction. Specifically, we reduced the expression values of ligand and receptor genes in the WNT signaling pathway to the lowest level (0.0004 and 0.001) in ischemic tissue sections (ACH0014 and ACH0010) of the myocardial infarction Visium dataset[9]. Since we restricted the input to genes with non-zero expressions, the downregulated ligand genes were WNT5B and the downregulated receptor genes were FZD4, FZD5, FZD6. We analyzed the effects on various genes, including direct targets of the WNT signaling pathway, housekeeping genes, and NOTCH1 targets, by calculating the cosine similarity of their output gene embeddings before and after in silico perturbation. The in silico perturbation was performed on ligand genes and receptor genes either individually or simultaneously.

Data availability

The spatial Visium dataset of human intestine is publicly available from the GEO database under accession number GSE158328, and the paired single-cell RNA-seq data used for deconvolution reference was downloaded from the CELLxGENE database under website https://cellxgene.cziscience.com/collections/60358420-6055-411d-ba4f-e8ac80682a2e. The spatial Visium dataset of human myocardial infarction is publicly accessible from the Zenodo data repository via record number 6580069, and the paired single-cell RNA-seq data used for deconvolution reference was downloaded from the CELLxGENE database under website https://cellxgene.cziscience.com/collections/8191c283-0816-424b-9b61-c3e1d6258a77. The final deconvolved data were deposited on the Zenodo data repository under record number 13895591.

The gene lists of "NOTCH1 activated versus non-target genes", "NOTCH1 network central versus peripheral genes", and housekeeping genes were retrieved from the "example_input_files" folder in the Genecorpus-30M dataset (https://huggingface.co/datasets/ctheodoris/Genecorpus-30M). The gene set "signaling by TGFB family members" was accessed through the GSEA Molecular Signatures Database under website https://www.gsea-msigdb.org/gsea/msigdb/human/geneset/REACTOME_SIGNALING_BY_TGFB_FAMILY_MEMBERS.html. The direct target genes of the WNT signaling pathway were retrieved from Nusse Lab's website https://web.stanford.edu/group/nusselab/cgi-bin/wnt/target_genes. The processed gene lists for downstream tasks can be accessed on the Github repository at https://github.com/csh3/stFormer.

Code availability

The model and pretrained weights, as well as the codes for pretraining and producing results of the downstream tasks are available on the Github repository at https://github.com/csh3/stFormer.

References

1 Vaswani, A. Attention is all you need. *Advances in Neural Information Processing Systems* (2017).

2 Brown, T. B. Language models are few-shot learners. *arXiv preprint arXiv:2005.14165* (2020).

3 Theodoris, C. V. *et al.* Transfer learning enables predictions in network biology. *Nature* **618**, 616-624 (2023). https://doi.org:10.1038/s41586-023-06139-9

4 Cui, H. *et al.* scGPT: toward building a foundation model for single-cell multi-omics using generative AI. *Nat Methods* **21**, 1470-1480 (2024). https://doi.org:10.1038/s41592-024-02201-0

5 Hao, M. *et al.* Large-scale foundation model on single-cell transcriptomics. *Nat Methods* **21**, 1481-1491 (2024). https://doi.org:10.1038/s41592-024-02305-7

6 Du, J. *et al.* Advances in spatial transcriptomics and related data analysis strategies. *J Transl Med* **21**, 330 (2023). https://doi.org:10.1186/s12967-023-04150-2

7 Kleshchevnikov, V. *et al.* Cell2location maps fine-grained cell types in spatial transcriptomics. *Nat Biotechnol* **40**, 661-671 (2022). https://doi.org:10.1038/s41587-021-01139-4

8 Fawkner-Corbett, D. *et al.* Spatiotemporal analysis of human intestinal development at single-cell resolution. *Cell* **184**, 810-826.e823 (2021). https://doi.org:10.1016/j.cell.2020.12.016

9 Kuppe, C. *et al.* Spatial multi-omic map of human myocardial infarction. *Nature* **608**, 766-777 (2022). https://doi.org:10.1038/s41586-022-05060-x

10 McInnes, L., Healy, J. & Melville, J. Umap: Uniform manifold approximation and projection for dimension reduction. *arXiv preprint arXiv:1802.03426* (2018).

11 Zeisberg, E. M. *et al.* Endothelial-to-mesenchymal transition contributes to cardiac fibrosis. *Nature Medicine* **13**, 952-961 (2007). https://doi.org:10.1038/nm1613

12 Theodoris, C. V. *et al.* Human disease modeling reveals integrated transcriptional and epigenetic mechanisms of NOTCH1 haploinsufficiency. *Cell* **160**, 1072-1086 (2015). https://doi.org:10.1016/j.cell.2015.02.035

13 Subramanian, A. *et al.* Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. *Proceedings of the National Academy of Sciences* **102**, 15545-15550 (2005).

14 Bergmann, M. W. WNT signaling in adult cardiac hypertrophy and remodeling: lessons learned from cardiac development. *Circulation research* **107**, 1198-1208 (2010).

15 Jin, S. *et al.* Inference and analysis of cell-cell communication using CellChat. *Nature Communications* **12**, 1088 (2021). https://doi.org:10.1038/s41467-021-21246-9

16 Hou, R., Denisenko, E., Ong, H. T., Ramilowski, J. A. & Forrest, A. R. R. Predicting cell-to-cell communication networks using NATMI. *Nature Communications* **11**, 5011 (2020). https://doi.org:10.1038/s41467-020-18873-z

17 Zhang, Y. *et al.* Cellinker: a platform of ligand-receptor interactions for intercellular communication analysis. *Bioinformatics* (2021). https://doi.org:10.1093/bioinformatics/btab036

18 Shao, X. *et al.* CellTalkDB: a manually curated database of ligand-receptor interactions in humans and mice. *Brief Bioinform* **22** (2021). https://doi.org:10.1093/bib/bbaa269

19 Massenet-Regad, L. & Soumelis, V. ICELLNET v2: a versatile method for cell-cell communication analysis from human transcriptomic data. *Bioinformatics* **40** (2024). https://doi.org:10.1093/bioinformatics/btae089

20 Shao, X. *et al.* Knowledge-graph-based cell-cell communication inference for spatially resolved transcriptomic data with SpaTalk. *Nature Communications* **13**, 4429 (2022). https://doi.org:10.1038/s41467-022-32111-8

21 Troulé, K. *et al.* CellPhoneDB v5: inferring cell-cell communication from single-cell multiomics data. *arXiv preprint arXiv:2311.04567* (2023).

22 Luecken, M. D. *et al.* Benchmarking atlas-level data integration in single-cell genomics. *Nat Methods* **19**, 41-50 (2022). https://doi.org:10.1038/s41592-021-01336-8