Cancer-Type-Agnostic Pan-Cancer Gene Expression Prediction from Histopathological Images

Yiyang Shen[1,⋆], Bochong Li[2,⋆], Huitao Wang[3,⋆], Zhiyu Chen[4], Zhe Li[1,✉], and Xiu Li[1,✉]

[1] Department of Control Science and Engineering, Tsinghua Shenzhen International Graduate School, Tsinghua University, Shenzhen, 518055, China
[2] National Cancer Institue, National Institutes of Health, 37 Convent Dr, Bethesda, MD 20814, United States
[3] Biomedical Engineering Department, Institute of Advanced Clinical Medicine, Peking University, Beijing, 100191, China
[4] School of Computer Science and Engineering, University of Electronic Science and Technology of China, Chengdu, 611731, China
li.zhe@sz.tsinghua.edu.cn, li.xiu@sz.tsinghua.edu.cn

Abstract. Gene expression profiles play a crucial role in understanding cancer pathogenesis, guiding personalized treatment, and predicting clinical outcomes. However, traditional gene expression detection relies on invasive tissue sequencing, which is time-consuming and costly. Existing image-based gene prediction methods are often limited to specific cancer types and fail to adapt to the pan-cancer scenario with diverse histological features. To address these challenges, we propose a novel pan-cancer image-based gene expression prediction method (PIGP) that enables gene expression prediction without the need for prior acquisition of cancer type information. Specifically, the model directly extracts discriminative features from pan-cancer histopathological images to preliminarily determine the cancer type, and further combines pan-cancer networks and corresponding cancer-type-specific networks to predict the corresponding gene expression levels. Extensive experiments on the pan-cancer dataset (HEST) have shown that the Pearson correlation coefficient for specific genes reaches up to 0.743. Moreover, heatmaps and distribution maps confirm the biological rationality and interpretability of the predicted results. This approach provides a cancer-type-agnostic computational framework for pan-cancer gene expression prediction and may support future translational studies. Our code is available at https://github.com/oceanflyfly/PIGP.

Keywords: Pan-cancer · Gene expression prediction · Histopathological image.

⋆ These authors contributed equally.

1 Introduction

Gene expression profiling is a cornerstone of precision oncology and translational medicine [16]. By characterizing molecular activity within tumor tissues, it enables mechanistic understanding of tumorigenesis, supports personalized therapeutic decisions, and facilitates prognosis assessment [11]. However, current gene expression measurement still primarily relies on invasive tissue sequencing, which is time-consuming, expensive, and difficult to scale. These limitations restrict large-scale pan-cancer studies and hinder longitudinal clinical monitoring [17,22]. Therefore, developing a non-invasive, cost-effective, and scalable alternative for gene expression prediction is of significant clinical and scientific importance.

Recent advances in computational pathology and spatial transcriptomics modeling have demonstrated that histopathological images encode rich molecular information [14]. By leveraging deep learning models to extract morphological and microenvironmental features from whole slide images, prior studies have shown encouraging results in predicting spatial gene expression or molecular alterations directly from tissue images [12,21]. Furthermore, multimodal integration frameworks have explored combining histology and spatial transcriptomics to enhance molecular inference [23].

Despite promising progress, existing models are typically trained and evaluated under closed-set settings, where training and testing data share identical or highly overlapping cancer types [6,5]. Such assumptions implicitly constrain models to operate within fixed distributional boundaries. Large-scale benchmarking studies suggest that these approaches often exhibit limited robustness when confronted with distribution shifts. Given the substantial inter-cancer heterogeneity in both histological morphology and molecular regulation, models optimized for specific cancer cohorts frequently fail to generalize to unseen cancer types. Consequently, pan-cancer gene expression prediction under unknown cancer types constitutes an open-set generalization problem that remains insufficiently addressed in current literature [8].

To tackle this challenge, we propose a novel Pan-cancer Image-based Gene Prediction framework (PIGP) that eliminates the requirement of prior cancer-type knowledge at inference time. The central idea is to decompose gene expression modeling into cross-cancer shared structural regularities and cancer-aware residual deviations within a unified dual-branch architecture. A global branch captures stable pan-cancer molecular patterns from aggregated data, while a specific branch models structured heterogeneity through residual learning. An adaptive fusion mechanism dynamically balances these components, enabling robust generalization across heterogeneous cancer distributions and improving predictive performance on unseen cancer types.

Our contributions are summarized as follows: (1) We formulate pan-cancer image-based gene expression prediction under unknown cancer types as an open-set generalization problem and highlight its clinical and translational significance; (2) We propose a shared–specific dual-branch framework with adaptive fusion to jointly model cross-cancer regularities and structured heterogeneity;

(3) We demonstrate through comprehensive experiments that explicitly disentangling shared structures and cancer-aware deviations substantially improves robustness and generalization in pan-cancer gene prediction.

2 Method

We present a robust framework for spatial transcriptomics prediction based on a "Spatial Enhancement - Decoupled Learning - Adaptive Fusion" paradigm. The core design philosophy is using a single unified model rather than by separately trained cancer-specific networks to address biological heterogeneity in cross-dataset scenarios. Specifically, the framework first utilizes a Spatial Context Aggregation Module to incorporate topological contexts, thereby correcting local visual biases. Subsequently, a Dual-Branch Decoupled Learning Module decouples the prediction task into a stable linear baseline and a non-linear residual correction driven by a Mixture of Experts (MoE) network to capture fine-grained expression fluctuations. Finally, an Adaptive Dual-Branch Fusion Mechanism achieves a dynamic balance between global consensus and cluster-specific knowledge. This hierarchical architecture establishes a robust mapping from high-dimensional histology to complex molecular profiles. Our learning pipeline is illustrated in Fig. 1.

2.1 Spatial Context Aggregation via Graph Convolutional Networks.

Spatial transcriptomics data encompasses not only independent morphological features but also profound spatial correlations within the local microenvironment. To incorporate spatial topological constraints into the prediction of gene expression, we construct a Graph Convolutional Network (GCN) based on K-Nearest Neighbors (K-NN) for feature aggregation.

Given a set of spots and their spatial coordinates, we define an undirected graph $G = (V, E)$. The layer-wise update rule for node features is formulated as:

$$H^{(l+1)} = \sigma\left(\tilde{D}^{-\frac{1}{2}}\tilde{A}\tilde{D}^{-\frac{1}{2}}H^{(l)}W^{(l)}\right) \tag{1}$$

where $\tilde{A}$ denotes the spatial adjacency matrix with self-loops, $\tilde{D}$ is its corresponding degree matrix, and $W^{(l)}$ represents the learnable transformation weights. Through this mechanism, the representation of each spot absorbs morphological context from its neighborhood. To mitigate the over-smoothing issue common in deep graph networks, we design a feature fusion layer that linearly weights the original high-frequency visual features X_{orig} and the spatially aggregated features $H^{(L)}$ to produce the enhanced representation $X_{enhanced}$:

$$X_{enhanced} = 0.9 \cdot X_{orig} + 0.1 \cdot H^{(L)} \tag{2}$$

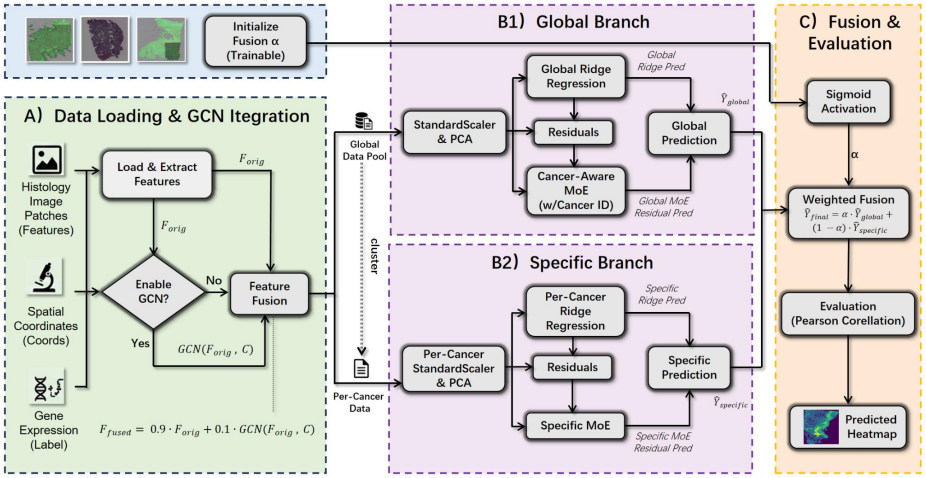

Fig. 1: The overview of PIGP. (A) Initial features are enriched with spatial context via a GCN, followed by global unsupervised clustering to partition the data and define distinct tissue microenvironments; (B) Dual-Branch Decoupled Learning use a Global Branch(B1) and a Specific Branch(B2) to concurrently model universal expression trends and cluster-specific heterogeneity; (C) Adaptive Dual-Branch Fusion Mechanism introduces a learnable parameter α dynamically weighs the outputs of both branches to adaptively balance consensus and specificity for the final prediction.

2.2 Dual-Branch Decoupled Learning.

To address the high variance of gene expression in cross-cancer scenarios, we propose a "Base-Residual" decoupled prediction model. The model first captures the linear baseline of gene expression $\hat{y}_{ridge}$ through Ridge regression:

$$\hat{y}_{ridge} = X_{pca} W_{ridge} \tag{3}$$

Recognizing that linear models struggle to represent complex non-linear gene regulatory networks, we introduce a Cluster-Aware Mixture of Experts (MoE) to fit the residual component $y_{res} = y_{true} - \hat{y}_{ridge}$. The unsupervised clustering-based phenotypic label c is encoded and concatenated with the image-derived feature representation X_{pca} before being passed into the gating network G, allowing the MoE module to route residual correction according to tissue phenotype:

$$\hat{y}_{res} = \sum_{i=1}^{N_{experts}} G_i(X_{pca}, c) E_i(X_{pca}) \tag{4}$$

This design allows the model to dynamically invoke different expert weights for residual correction based on the tissue phenotype (e.g., tumor core vs. stroma). The final prediction for a single branch is the summation of both components:

$$\hat{y}_{branch} = 0.8 \cdot \hat{y}_{ridge} + 0.2 \cdot \hat{y}_{res} \tag{5}$$

2.3 Adaptive Dual-Branch Fusion Mechanism.

To balance the model's generalization across datasets with its specificity for particular phenotypes, we construct a joint inference architecture comprising a global consensus branch and a cluster-specific branch. The global branch is trained on the entire dataset to extract universal cross-cancer features, while specific branches are optimized independently for each cluster.

During the inference phase, the model dynamically adjusts the contribution of each branch via a learnable parameter α, enabling the adaptive synthesis of the final prediction y_{final}:

$$y_{final} = \text{Sigmoid}(w_\alpha) \cdot \hat{y}_{global} + (1 - \text{Sigmoid}(w_\alpha)) \cdot \hat{y}_{specific} \tag{6}$$

where w_α is a weight optimized through backpropagation. This mechanism empowers the model to prioritize the specific branch for in-distribution samples to ensure high precision, while relying on the robustness of the global branch when encountering long-tail distributions or novel phenotypes.

3 Experimental Results

3.1 Dataset and Implementation

Dataset Source and Preprocessing. Adapted from the Hest-1k dataset[7], our cohort includes paired spatial transcriptomics (ST) profiles, whole slide images stained with hematoxylin and eosin (H&E) and comprehensive metadata. Data were aggregated from ten public repositories, including 10x Genomics (TENX), Mendeley (MEND), Spatial Research (SPA), Zenodo, NCBI, GitHub, Human Cell Atlas, BioStudies, HTAN, and internal cohorts. We extracted 224×224 pixel patches centered at each ST spot at $20\times$ magnification. 72 samples from all ten cancer types were pooled into a unified pan-cancer cohort. For each cancer type, one quarter of the samples (rounded up) were randomly selected for validation, and the remaining samples were used for training.

We formulated ten histology-to-gene-expression prediction tasks across various cancer types: invasive ductal carcinoma (IDC), prostate adenocarcinoma (PRAD), pancreatic adenocarcinoma (PAAD), skin cutaneous melanoma (SKCM), colon adenocarcinoma (COAD), rectum adenocarcinoma (READ), clear cell renal cell carcinoma (ccRCC), hepatocellular carcinoma (HCC), lung adenocarcinoma (LUNG), and axillary lymph node metastasis (Lymph-IDC). For each task, we predicted the expression of 13 highly variable genes shared across all ten cancers from 112×112 μm H&E regions, which is equivalent to 224×224 pixels at $20\times$ magnification.

Evaluation Metrics. Model performance was rigorously evaluated using the Pearson correlation coefficient (R) between the predicted and ground-truth gene expression values.

Table 1: Average Pearson coefficients for gene expression prediction across ten histological tasks using thirteen different feature extraction foundation models.

Model Name	Cancer Types (Pearson Correlation r)									
	IDC	PRAD	PAAD	SKCM	COAD	READ	CCRCC	HCC	LUNG	LYMPH IDC
uni_v1[2]	0.6558	0.0875	0.6646	0.6516	0.6012	0.1590	0.1525	0.0361	0.7086	0.1855
resnet50[10]	0.5839	0.0451	0.5991	0.5038	0.5132	-0.0262	0.0857	0.0211	0.6549	0.1752
ctranspath[19]	0.6069	0.0896	0.6102	0.5676	0.5083	0.0036	0.1960	0.0384	0.6567	0.1750
phikon[3]	0.6111	**0.0957**	0.6199	0.6338	0.6324	0.1250	**0.2325**	-0.0005	0.7053	0.1858
phikon_v2[3]	0.6308	0.0511	0.6409	0.5975	0.6251	0.1745	0.1358	0.0461	0.7112	0.1892
h0_mini[4]	0.6363	0.0595	0.6604	**0.6603**	0.5892	0.1448	0.1626	0.0204	0.7088	**0.2089**
hibou_large[13]	0.6559	0.0893	**0.6749**	0.6294	0.6235	0.1511	0.1744	0.0389	0.7283	0.1782
kaiko_base_8[1]	0.6438	0.0613	0.6318	0.6322	0.6118	0.1632	0.1799	0.0364	0.6630	0.1660
conch_v1[9]	0.6369	0.0831	0.6413	0.5958	0.6149	0.1516	0.0909	0.0023	0.7129	0.1950
virchow[18]	0.6736	0.0751	0.6642	0.6599	0.6376	0.1366	0.1245	0.0548	0.7312	0.1833
gigapath[20]	0.6125	0.0523	0.6297	0.6432	0.5988	0.1536	0.0925	0.0504	0.7151	0.1810
virchow2[24]	0.6675	0.0645	0.6509	0.6260	0.5394	**0.1967**	0.1765	**0.0621**	**0.7430**	0.1921
hoptimus0[15]	**0.6798**	0.0572	0.6448	0.6239	**0.6705**	0.1504	0.1409	0.0436	0.7178	0.2077

Implementation Details. We employed thirteen pathology foundation models as feature extractors: Uni_v1, ResNet50, CTransPath, Phikon, Phikon_v2, H0_mini, Hibou_Large, Conch, Virchow, GigaPath, Virchow 2, H-Optimus-0. Model-specific patch embeddings, ranging from 512 to 2048 dimensions, were mapped to the log1p-normalized expression values of the 13 highly variable genes. All experiments were conducted on a single NVIDIA RTX A100 GPU.

3.2 Results

Evaluation of Generation Performance. Table 1 shows the excellent result of different tasks. When using Virchow2 as the feature extraction backbone for predicting gene expression in lung cancer, the Pearson correlation coefficient reached 0.7430. Because the original HEST-1k benchmark [7] trains cancer-specific models and evaluates larger high-expression gene sets, direct numerical comparison with our unified pan-cancer setting should be interpreted cautiously. Our experiments instead aim to evaluate whether a single cancer-type-agnostic framework can maintain robust prediction across heterogeneous cancer cohorts. Under this objective, PIGP shows consistent gains over the corresponding baseline configurations, suggesting improved modeling of morphology-expression associations in the pan-cancer scenario. The best-performing backbone varied across cancer types. In particular, H-Optimus-0 achieved the highest correlations for IDC ($r = 0.6798$) and COAD ($r = 0.6705$). Comparative analysis shows that the performance gain is most significant in the PAAD (pancreatic cancer) task, where the r-value improved from 0.5010 to 0.6749. PRAD and HCC remain challenging ($r < 0.12$) due to tissue heterogeneity and the cross-cancer gene selection strategy, which may yield low expression and limited variance in these cancer types, thereby constraining correlation-based evaluation. This

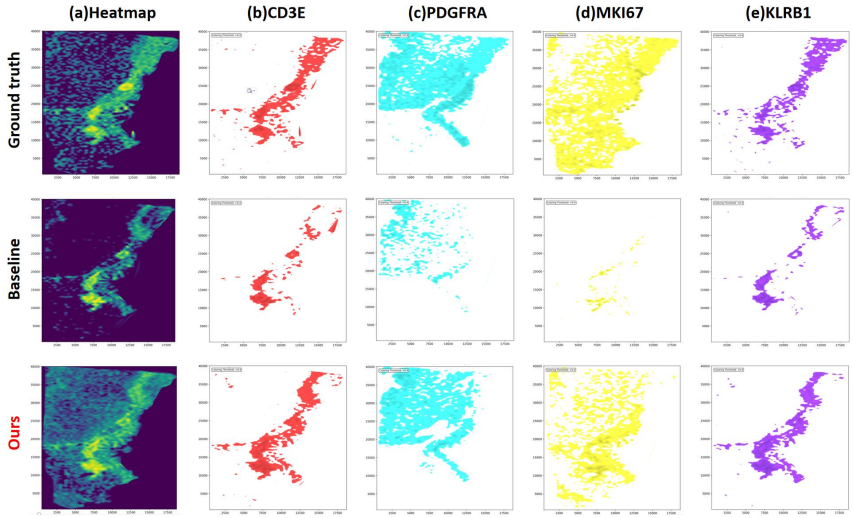

Fig. 2: A series of visual results. Baseline represents the model using in HEST [7] (a) Gene Expression Heatmap of CD3E. (b)-(e) Distribution map of gene expression above the threshold = 0.4, (b)-(e) represent genes CD3E, PDGFRA, MKI67, and KLRB1, respectively.

widespread performance gain suggests that the unsupervised clustering-guided dual-branch architecture can effectively break feature barriers between cancer types, precisely capturing tissue-specific molecular signatures while maintaining global generalization, providing strong technological support for cross-cancer spatial transcriptomics prediction.

Visual Analysis. The predicted results for PIGP maintain an exceptionally high level of consistency with the Ground Truth in terms of both spatial distribution and expression intensity gradients in typical gene expression heatmaps, such as CD3E, shown in Fig. 2. In contrast, the predictions generated by the original model exhibit a fragmented distribution, with significant signal loss in several high-expression regions. For genes like CD3E and KLRB1, which feature highly organized high-expression zones, our model achieves accurate and continuous reconstruction. Furthermore, for gene types such as PDGFRA and MKI67, which involve large blocks requiring precise high-expression boundaries—our model accurately defines the demarcation lines. These visualization results demonstrate that the PIGP framework effectively compensates for the lack of spatial prediction sensitivity found in existing models.

Ablation Study. We conducted ablation experiments on Dual-Branch Decoupled Learning (DBDL) framework across 13 pathological feature extraction models in gene expression tasks, as the global branch, specific branch, and adap-

Table 2: Performance Comparison: PIGP with Dual-Branch Decoupled Learning(w/) vs. PIGP without Dual-Branch Decoupled Learning(w/o) across 13 Foundation Models. Positive Δ values indicate substantial performance gains.

Model Name		IDC	PRAD	PAAD	SKCM	COAD	READ	CCRCC	HCC	LUNG	LYMPH IDC
uni_v1	w/	0.6558	0.0875	0.6646	0.6516	0.6012	0.1590	0.1525	0.0361	0.7086	0.1855
	w/o	0.6109	0.0666	0.5391	0.6172	0.5481	0.1095	0.1350	0.0249	0.6647	0.1998
	Δ	**+0.0449**	**+0.0209**	**+0.1255**	**+0.0344**	**+0.0531**	**+0.0495**	**+0.0175**	**+0.0112**	**+0.0439**	-0.0143
resnet50	w/	0.5839	0.0451	0.5991	0.5038	0.5132	-0.0262	0.0857	0.0211	0.6549	0.1752
	w/o	0.5296	0.0553	0.4646	0.5030	0.3710	0.0239	0.0817	0.0113	0.6026	0.1759
	Δ	**+0.0543**	-0.0102	**+0.1345**	**+0.0008**	**+0.1422**	-0.0501	**+0.0040**	**+0.0098**	**+0.0523**	-0.0007
ctranspath	w/	0.6069	0.0896	0.6102	0.5676	0.5083	0.0036	0.1960	0.0384	0.6567	0.1750
	w/o	0.5660	0.0475	0.4977	0.5705	0.4516	0.0626	0.1303	0.0271	0.6256	0.1850
	Δ	**+0.0409**	**+0.0421**	**+0.1125**	-0.0029	**+0.0567**	-0.0590	**+0.0657**	**+0.0113**	**+0.0311**	-0.0100
phikon	w/	0.6111	0.0957	0.6199	0.6338	0.6324	0.1250	0.2325	-0.0005	0.7053	0.1858
	w/o	0.5611	0.0561	0.4980	0.5911	0.5211	0.0673	0.1349	0.0247	0.6683	0.1854
	Δ	**+0.0500**	**+0.0396**	**+0.1219**	**+0.0427**	**+0.1113**	**+0.0577**	**+0.0976**	-0.0252	**+0.0370**	**+0.0004**
phikon_v2	w/	0.6308	0.0511	0.6409	0.5975	0.6251	0.1745	0.1358	0.0461	0.7112	0.1892
	w/o	0.5821	0.0556	0.4996	0.5672	0.5048	0.0813	0.1117	0.0244	0.6593	0.2132
	Δ	**+0.0487**	-0.0045	**+0.1413**	**+0.0303**	**+0.1203**	**+0.0932**	**+0.0241**	**+0.0217**	**+0.0519**	-0.0240
h0_mini	w/	0.6363	0.0595	0.6604	0.6603	0.5892	0.1448	0.1626	0.0204	0.7088	0.2089
	w/o	0.6277	0.0864	0.5620	0.6366	0.5212	0.0912	0.1413	0.0201	0.6893	0.2094
	Δ	**+0.0086**	-0.0269	**+0.0984**	**+0.0237**	**+0.0680**	**+0.0536**	**+0.0213**	**+0.0003**	**+0.0195**	-0.0005
hibou_large	w/	0.6559	0.0893	0.6749	0.6294	0.6235	0.1511	0.1744	0.0389	0.7283	0.1782
	w/o	0.6164	0.0621	0.5010	0.6255	0.5458	0.1126	0.1185	0.0202	0.6870	0.1805
	Δ	**+0.0395**	**+0.0272**	**+0.1739**	**+0.0039**	**+0.0777**	**+0.0385**	**+0.0559**	**+0.0187**	**+0.0413**	-0.0023
kaiko_base_8	w/	0.6438	0.0613	0.6318	0.6322	0.6118	0.1632	0.1799	0.0364	0.6630	0.1660
	w/o	0.6002	0.0726	0.5176	0.6272	0.5201	0.0776	0.1194	0.0304	0.6449	0.1865
	Δ	**+0.0436**	-0.0113	**+0.1142**	**+0.0050**	**+0.0917**	**+0.0856**	**+0.0605**	**+0.0060**	**+0.0181**	-0.0205
conch_v1	w/	0.6369	0.0831	0.6413	0.5958	0.6149	0.1516	0.0909	0.0023	0.7129	0.1950
	w/o	0.5746	0.0575	0.5181	0.6549	0.4917	0.0851	0.1047	0.0011	0.6500	0.2043
	Δ	**+0.0623**	**+0.0256**	**+0.1232**	-0.0591	**+0.1232**	**+0.0665**	-0.0138	**+0.0012**	**+0.0629**	-0.0093
virchow	w/	0.6736	0.0751	0.6642	0.6599	0.6376	0.1366	0.1245	0.0548	0.7312	0.1833
	w/o	0.5952	0.0597	0.5605	0.5968	0.5483	0.0723	0.1260	0.0317	0.6812	0.1819
	Δ	**+0.0784**	**+0.0154**	**+0.1037**	**+0.0631**	**+0.0893**	**+0.0643**	-0.0015	**+0.0231**	**+0.0500**	**+0.0014**
gigapath	w/	0.6125	0.0523	0.6297	0.6432	0.5988	0.1536	0.0925	0.0504	0.7151	0.1810
	w/o	0.5972	0.0464	0.5295	0.5991	0.5283	0.0768	0.1004	0.0269	0.6786	0.2029
	Δ	**+0.0153**	**+0.0059**	**+0.1002**	**+0.0441**	**+0.0705**	**+0.0768**	-0.0079	**+0.0235**	**+0.0365**	-0.0219
virchow2	w/	0.6675	0.0645	0.6509	0.6260	0.5394	0.1967	0.1765	0.0621	0.7430	0.1921
	w/o	0.6166	0.0674	0.5150	0.6427	0.5287	0.0866	0.1205	0.0274	0.6815	0.1836
	Δ	**+0.0509**	-0.0029	**+0.1359**	-0.0167	**+0.0107**	**+0.1101**	**+0.0560**	**+0.0347**	**+0.0615**	**+0.0085**
hoptimus0	w/	0.6798	0.0572	0.6448	0.6239	0.6705	0.1504	0.1409	0.0436	0.7178	0.2077
	w/o	0.6230	0.0699	0.5581	0.6679	0.5810	0.1171	0.1298	0.0142	0.6987	0.1872
	Δ	**+0.0568**	-0.0127	**+0.0867**	-0.0440	**+0.0895**	**+0.0333**	**+0.0111**	**+0.0294**	**+0.0191**	**+0.0205**

tive fusion module are functionally interdependent. As shown in Table 2, the experimental results indicate that in the majority of test cases, the Δ values are positive, demonstrating that DBDL consistently improves the prediction accuracy of base models across different architectures. In the PAAD task, the performance gain reached approximately 0.13. Even for low-correlation prediction tasks such as HCC and PRAD, the inclusion of the dual-branch structure yielded positive improvements. However, we observed a negative growth in correlation coefficients for the LYMPH_IDC prediction task when using the `uni_v1` and `kaiko_base_8` models. This may be attributed to the highly dense immune microenvironment in lymphoid tissues, suggesting that the cluster-guided strategy in decoupled learning requires more refined hyperparameter tuning for such contexts.

4 Conclusion

In this study, we propose PIGP, a pan-cancer gene expression prediction model. PIGP first uses GCN to enhance extracted features, followed by data partitioning through global unsupervised clustering. Subsequently, a dual-branch decoupled learning framework is employed to simultaneously universal expression trends and cluster-specific heterogeneity. Finally, an adaptive fusion mechanism employs a learnable parameter α to dynamically weight the dual-branch outputs for the final prediction. Extensive experiments and spatial visualization results demonstrate the superior performance of PIGP, highlighting its capability to enhance the morphology-to-expression predictive power of pathology foundation models. Our results suggest that PIGP can help characterize morphology-expression associations across cancer types and may provide useful computational support for future screening of unknown primary lesions, while further validation is required before clinical deployment. Future work will therefore explore cancer-aware gene list refinement and dynamic adjustment of the weights according to round-wise prediction behavior, so that the model can better exploit spatial feature cues in these difficult cohorts.

Acknowledgments. This research received no external funding.

Disclosure of Interests. The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

References

1. kaiko. ai, Aben, N., de Jong, E.D., Gatopoulos, I., Känzig, N., Karasikov, M., Lagré, A., Moser, R., van Doorn, J., Tang, F.: Towards large-scale training of pathology foundation models (2024), https://arxiv.org/abs/2404.15217
2. Chen, R.J., Ding, T., Lu, M.Y., Williamson, D.F., Jaume, G., Song, A.H., Chen, B., Zhang, A., Shao, D., Shaban, M., et al.: Towards a general-purpose foundation model for computational pathology. Nature medicine **30**(3), 850–862 (2024)
3. Filiot, A., Ghermi, R., Olivier, A., Jacob, P., Fidon, L., Mac Kain, A., Saillard, C., Schiratti, J.: Scaling self-supervised learning for histopathology with masked image modeling. medrxiv 2023. Google Scholar (2023)
4. Filiot, A., Dop, N., Tchita, O., Riou, A., Dubois, R., Peeters, T., Valter, D., Scalbert, M., Saillard, C., Robin, G., et al.: Distilling foundation models for robust and efficient models in digital pathology. In: International Conference on Medical Image Computing and Computer-Assisted Intervention. pp. 162–172. Springer (2025)
5. Hu, Y., Li, X., Yi, Y., Huang, Y., Wang, G., Wang, D.: Deep learning-driven survival prediction in pan-cancer studies by integrating multimodal histology-genomic data. Briefings in Bioinformatics **26**(2), bbaf121 (03 2025). https://doi.org/10.1093/bib/bbaf121, https://doi.org/10.1093/bib/bbaf121

6. Huang, T., Liu, T., Babadi, M., Jin, W., Ying, R.: Scalable generation of spatial transcriptomics from histology images via whole-slide flow matching. In: Forty-second International Conference on Machine Learning (2025), https://openreview.net/forum?id=Ossg1IbHDT
7. Jaume, G., Doucet, P., Song, A., Lu, M.Y., Almagro Pérez, C., Wagner, S., Vaidya, A., Chen, R., Williamson, D., Kim, A., et al.: Hest-1k: A dataset for spatial transcriptomics and histology image analysis. Advances in Neural Information Processing Systems **37**, 53798–53833 (2024)
8. Li, B., Xiao, X., Zhang, C., Xiao, M., Zhang, L.: Dghnn: a deep graph and hypergraph neural network for pan-cancer related gene prediction. Bioinformatics **41**(7), btaf379 (06 2025). https://doi.org/10.1093/bioinformatics/btaf379, https://doi.org/10.1093/bioinformatics/btaf379
9. Lu, M.Y., Chen, B., Williamson, D.F., Chen, R.J., Liang, I., Ding, T., Jaume, G., Odintsov, I., Le, L.P., Gerber, G., et al.: A visual-language foundation model for computational pathology. Nature medicine **30**(3), 863–874 (2024)
10. Lu, M.Y., Williamson, D.F., Chen, T.Y., Chen, R.J., Barbieri, M., Mahmood, F.: Data-efficient and weakly supervised computational pathology on whole-slide images. Nature biomedical engineering **5**(6), 555–570 (2021)
11. Luo, Y., Zhao, X., Ye, K., Meng, D.: Stinr: Deciphering spatial transcriptomics via implicit neural representation. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). pp. 25930–25939 (June 2025)
12. Mejia, G., Ruiz, D., Cárdenas, P., Manrique, L., Vega, D., Arbeláez, P.: Enhancing gene expression prediction from histology images with spatial transcriptomics completion. In: Linguraru, M.G., Dou, Q., Feragen, A., Giannarou, S., Glocker, B., Lekadir, K., Schnabel, J.A. (eds.) Medical Image Computing and Computer Assisted Intervention – MICCAI 2024. pp. 91–101. Springer Nature Switzerland, Cham (2024)
13. Nechaev, D., Pchelnikov, A., Ivanova, E.: Hibou: A family of foundational vision transformers for pathology (2024), https://arxiv.org/abs/2406.05074
14. Que, N., Wang, X., Chen, J., Jiang, Y., Li, C.: Adaptive spatial transcriptomics interpolation via cross-modal cross-slice modeling. In: Gee, J.C., Alexander, D.C., Hong, J., Iglesias, J.E., Sudre, C.H., Venkataraman, A., Golland, P., Kim, J.H., Park, J. (eds.) Medical Image Computing and Computer Assisted Intervention – MICCAI 2025. pp. 45–54. Springer Nature Switzerland, Cham (2026)
15. Saillard, C., Jenatton, R., Llinares-López, F., Mariet, Z., Cahané, D., Durand, E., Vert, J.P.: H-optimus-0 (2024), https://github.com/bioptimus/releases/tree/main/models/h-optimus/v0
16. Shi, H., Chi, C., Wan, P., Zhang, D., Shao, W.: Multi-modal topology-embedded graph learning for spatially resolved genes prediction from pathology images with prior gene similarity information. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). pp. 20810–20819 (June 2025)
17. Su, X., Yu, H., Zhi, D., Ji, S.: Learning to discover regulatory elements for gene expression prediction. In: The Thirteenth International Conference on Learning Representations (2025), https://openreview.net/forum?id=Mfnh1Sqdwf
18. Vorontsov, E., Bozkurt, A., Casson, A., Shaikovski, G., Zelechowski, M., Severson, K., Zimmermann, E., Hall, J., Tenenholtz, N., Fusi, N., et al.: A foundation model for clinical-grade computational pathology and rare cancers detection. Nature medicine **30**(10), 2924–2935 (2024)
19. Wang, X., Yang, S., Zhang, J., Wang, M., Zhang, J., Huang, J., Yang, W., Han, X.: Transpath: Transformer-based self-supervised learning for histopathological im-

age classification. In: International Conference on Medical Image Computing and Computer-Assisted Intervention. pp. 186–195. Springer (2021)
20. Xu, H., Usuyama, N., Bagga, J., Zhang, S., Rao, R., Naumann, T., Wong, C., Gero, Z., González, J., Gu, Y., Xu, Y., Wei, M., Wang, W., Ma, S., Wei, F., Yang, J., Li, C., Gao, J., Rosemon, J., Bower, T., Lee, S., Weerasinghe, R., Wright, B.J., Robicsek, A., Piening, B., Bifulco, C., Wang, S., Poon, H.: A whole-slide foundation model for digital pathology from real-world data. Nature (2024)
21. Xue, S., Wang, C., Fan, X., Min, W.: Inferring super-resolved gene expression by integrating histology images and spatial transcriptomics with histex. In: Gee, J.C., Alexander, D.C., Hong, J., Iglesias, J.E., Sudre, C.H., Venkataraman, A., Golland, P., Kim, J.H., Park, J. (eds.) Medical Image Computing and Computer Assisted Intervention – MICCAI 2025. pp. 296–306. Springer Nature Switzerland, Cham (2026)
22. Yang, Z., Chen, Z., Liu, X., Kotoge, R., Chen, P., Matsubara, Y., Sakurai, Y., Sun, J.: Gesubnet: Gene interaction inference for disease subtype network generation. In: The Thirteenth International Conference on Learning Representations (2025), https://openreview.net/forum?id=ja4rpheN2n
23. Zhang, D., Schroeder, A., Yan, H., Yang, H., Hu, J., Lee, M.Y.Y., Cho, K.S., Susztak, K., Xu, G.X., Feldman, M.D., Lee, E.B., Furth, E.E., Wang, L., Li, M.: Inferring super-resolution tissue architecture by integrating spatial transcriptomics with histology. Nature Biotechnology **42**(9), 1372–1377 (Sep 2024). https://doi.org/10.1038/s41587-023-02019-9, https://doi.org/10.1038/s41587-023-02019-9
24. Zimmermann, E., Vorontsov, E., Viret, J., Casson, A., Zelechowski, M., Shaikovski, G., Tenenholtz, N., Hall, J., Klimstra, D., Yousfi, R., et al.: Virchow2: Scaling self-supervised mixed magnification models in pathology. arXiv preprint arXiv:2408.00738 (2024)