PEaRL: Pathway-Enhanced Representation Learning for Gene and Pathway Expression Prediction from Histology

— Supplementary Material —

In this supplementary material, we begin with the details of the dataset and experimental design in section 6. Next we provide implementation details in section 7. We also provide additional visualizations in terms of correlation plots, Leiden clustering, and pathway and gene heatmaps along with their descriptions in section 8.

6. Dataset and Experimental Design Details

We evaluated PEaRL across three spatial transcriptomics (ST) cohorts derived from three distinct cancer types: breast cancer (invasive ductal carcinoma), skin cancer (cutaneous squamous cell carcinoma), and renal-cell–associated lymphoid tissue.

For all spatial transcriptomics datasets, labels are assigned at the tissue-section (slide) level, not the spot level. Each slide corresponds to a single patient sample provided in the original study metadata, and we used these official sample identifiers for the 5 fold cross-validation splits. No artificial or derived labels were created by us; tissue-level labels (e.g., breast IDC, cSCC, lymphoid) come directly from the originating publications. Importantly, all spots from a given slide inherit the slide label and remain together in the same fold, ensuring that no slide contributes spots to both training and testing sets.

Breast ST datasets. The breast cancer ST cohort comprised 36 tissue sections, totaling 135,620 spatial spots with 19,717 detected genes. To ensure robust evaluation and reduce sample-level bias, we adopted a 5-fold cross-validation protocol at the tissue-sample level. In each fold, 29 samples were allocated for model training and 7 samples for held-out testing.

Skin ST datasets. The skin cancer dataset consisted of 12 tissue sections, encompassing 8,671 spots and 13,251 genes. We applied the same 5-fold cross-validation strategy, with each fold containing 9 training samples and 3 test samples.

Lymph ST datasets. The lymphoid dataset included 24 tissue sections, totaling 74,220 spots and 36,601 expressed genes. As with the other datasets, we implemented 5-fold cross-validation, using 18 samples for training and 5 samples for testing in each fold.

7. Implementation Details

PEaRL operates in two stages: contrastive multimodal pretraining followed by downstream prediction tasks. In the pretraining stage, ssGSEA is used to compute pathway activity scores for every spatial spot, producing pathway-level transcriptomic features aligned with corresponding 224×224 histology patches. A multimodal contrastive objective is then used to align image embeddings with pathway embeddings, enabling the model to learn a shared representation across modalities.

During downstream evaluation, the trained image encoder is applied to test histology patches, and the resulting embeddings are passed through a 3-layer gene MLP head (using the top 1,000 HVGs) and a 3-layer pathway MLP head (using ssGSEA pathway targets) to generate gene and pathway expression predictions. For survival analysis, the breast-trained image encoder influenced by pathway scores is further evaluated on the TCGA-BRCA cohort.

All pathway scores (ssGSEA) are computed within the training fold, using the gene expression from slides assigned to that fold. During contrastive pretraining, each training slide contributes both its histology patches and its own pathway vectors; no pathway information from test slides is used anywhere in training, hyperparameter tuning, or model selection. At inference time, pathway scores from test slides are never used—the model sees only histology images, and predictions are generated solely from the image encoder and prediction heads. This ensures strict separation of training and testing data and prevents leakage between modalities.

Setting for training stage. The pretraining architecture consists of two encoders: a UNI-based image encoder with a 3-layer projection head, and a 2-layer Transformer pathway encoder that ingests pathway scores along with spatial coordinates. Both encoders output 256-dimensional embeddings, which are aligned via contrastive learning by fine-tuning the last four layers of UNI while training the Transformer from scratch. After alignment, image embeddings are routed to the two MLP heads for gene and pathway prediction.

Training is conducted with a learning rate of 1×10^{-5}, weight decay of 1×10^{-4}, 8 attention heads (64-dimensional), a temperature of 0.8, a maximum of 150 epochs, and early stopping with a patience of 50. All experiments are performed on an NVIDIA Quadro RTX 8000 GPU with 48 GB memory.

Setting for testing stage. During testing, only the trained image encoder and prediction heads are used. Test histology patches are encoded to produce image embeddings, which are then processed by the gene and pathway MLP heads to obtain final predictions. We follow a 5-fold cross-validation scheme across all datasets, selecting the best checkpoint for each fold based on validation performance. Final results are computed using the five fold-specific models, and we report the mean and standard deviation across all folds to ensure robust, unbiased evaluation.

8. Additional Visualizations

Correlation Plots. Correlation plots illustrate how strongly genes or pathways co-vary across spatial spots, making them useful for visualizing underlying biological relationships. In the ground truth, these plots contain clear blocks and structured patterns that reflect coordinated biological processes. PEaRL's predictions best reproduce these structures, demonstrating that the model preserves the true co-variation patterns present in real tissue (Fig. 5 and Fig. 6). In contrast, baseline models tend to oversmooth, exaggerate, or entirely lose these patterns, indicating a weaker ability to capture meaningful biological relationships.

Leiden Clustering. Leiden clustering is an unsupervised method that groups spatial transcriptomics spots based on similarity in gene-expression profiles, revealing functional regions that may not always align with visible histology but correspond to underlying molecular states such as tumor, stroma, or immune niches. The Adjusted Rand Index (ARI) quantifies how well predicted clusters match the ground-truth transcriptomic clusters, making it a useful metric for evaluating whether a model preserves biologically meaningful spatial zones. In our visualizations (Fig. 7 and Fig. 8), the ground-truth Leiden clusters correspond to transcriptomically distinct tissue compartments that may not be visually obvious from morphology alone. PEaRL's predictions most accurately recover these molecularly defined regions and achieve the highest ARI scores, indicating that the model captures both subtle and large-scale gene-expression structure across the tissue.

Visualization of the *HALLMARK MYC TARGETS V1* Pathway and its Corresponding Gene *SSBP1* in Breast Cancer (Fig. 9 and Fig. 10). The *HALLMARK_MYC_TARGETS_V1* pathway encompasses a core set of genes activated by the *MYC* oncogene, reflecting cellular programs involved in proliferation, metabolism, and protein synthesis. Within this pathway, *SSBP1* encodes a mitochondrial DNA–binding protein essential for maintaining mitochondrial genome stability and supporting cellular energy production (MSigDB [27]). Biologically, *MYC* activation and *SSBP1* function in concert to sustain the high metabolic and biosynthetic demands of rapidly dividing cancer cells, making this pathway a key marker of tumor aggressiveness. In breast cancer, elevated *MYC*-target activity is consistently associated with higher tumor grade, increased proliferation, endocrine resistance, and poorer patient survival [41]. *SSBP1* dysregulation also holds clinical relevance, having been linked to altered mitochondrial function, enhanced metastatic behavior—particularly in triple-negative breast cancer—and worse clinical outcomes [23].

Visualization of the *Reactome Eukaryotic Translation Initiation* Pathway and its Corresponding Gene *EIF4EBP1* in Skin Cancer (Fig. 9 and Fig. 10). The *Reactome_Eukaryotic_Translation_Initiation* pathway describes how cells begin synthesizing new proteins, a process that becomes highly active in rapidly growing or stressed cancer cells [10]. *EIF4EBP1* encodes 4E-BP1, a central regulator that can either inhibit or permit protein production depending on its phosphorylation state, which is controlled by growth-related signaling pathways such as mTOR. In skin cancer, increased activity of translation-initiation pathways is often associated with higher proliferation and more aggressive tumor behavior [9, 35]. Although direct studies in cutaneous squamous cell carcinoma are limited, dysregulation of *EIF4EBP1* has been linked to poorer outcomes in multiple cancer types, suggesting its potential relevance as a molecular indicator of tumor growth and progression.

Visualization of the *Reactome ABC-Family of Proteins Mediated Transport* Pathway and its Corresponding Gene *PKP2* in Lymphoid Cancer (Fig. 9 and Fig. 10). The *Reactome_ABC_Family_of_Proteins_Mediated_Transport* pathway characterizes how ATP-binding cassette (ABC) transporters move molecules across cellular membranes, including drugs, metabolites, and lipids—processes frequently altered in cancer biology [4]. *PKP2* encodes plakophilin-2, a scaffold protein involved in cell–cell adhesion and cytoskeletal organization, influencing cellular structure and communication [12]. In lymphoid cancers, ABC-mediated transport pathways are highly significant, as dysregulation of these transporters contributes to chemotherapy resistance, altered immune-cell behavior, and metabolic rewiring that supports tumor survival [40].

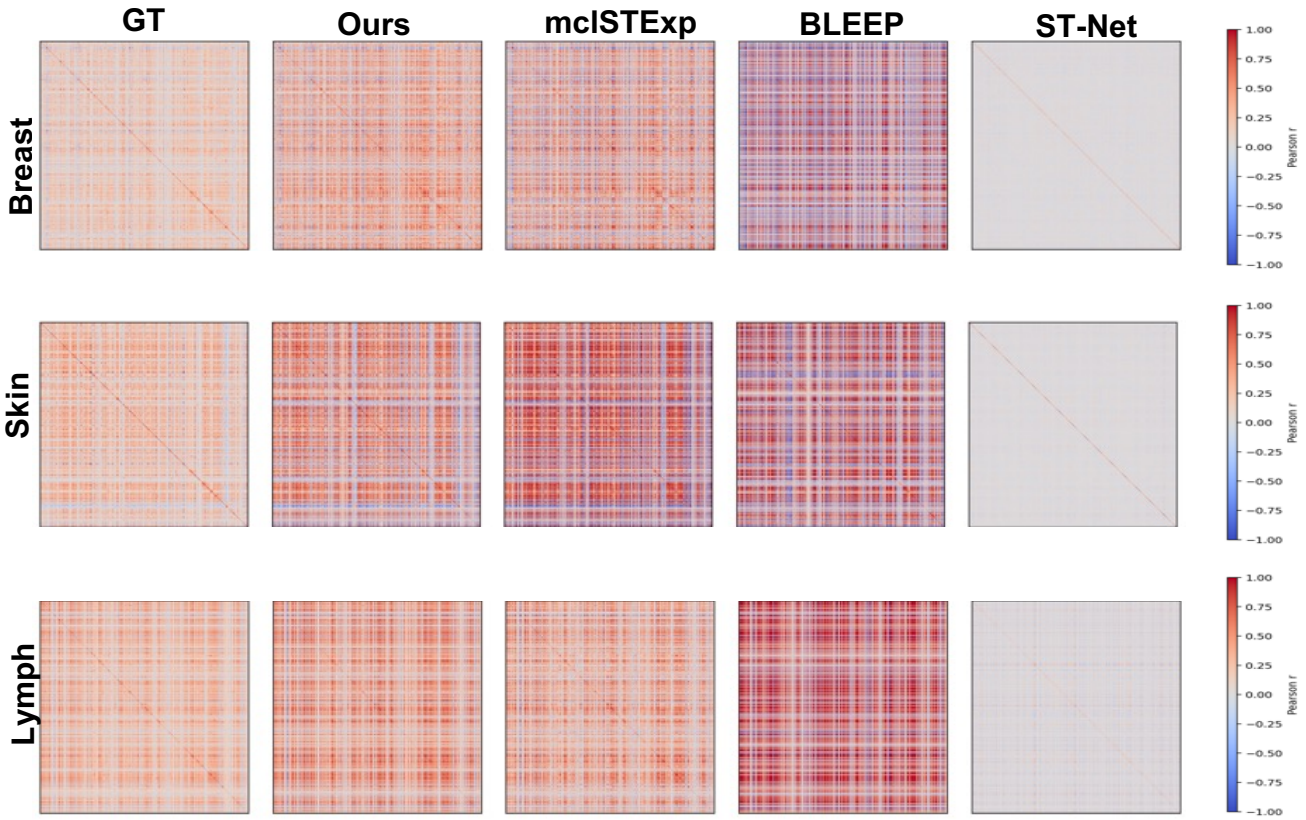

Figure 5. Visualization of the pathway-pathway correlation plots across the three datasets and comparison of PEaRL with baseline models

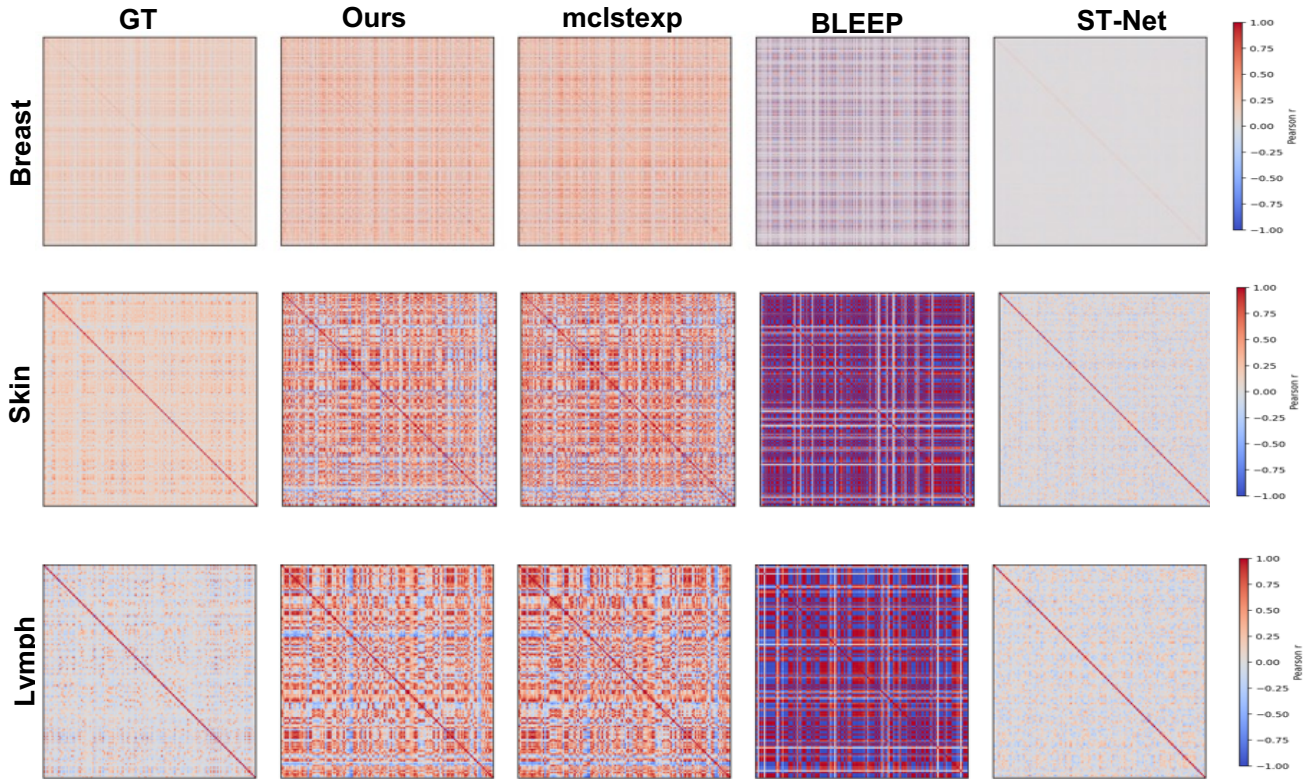

Figure 6. Visualization of the gene-gene correlation plots across the three datasets and comparison of PEaRL with baseline models

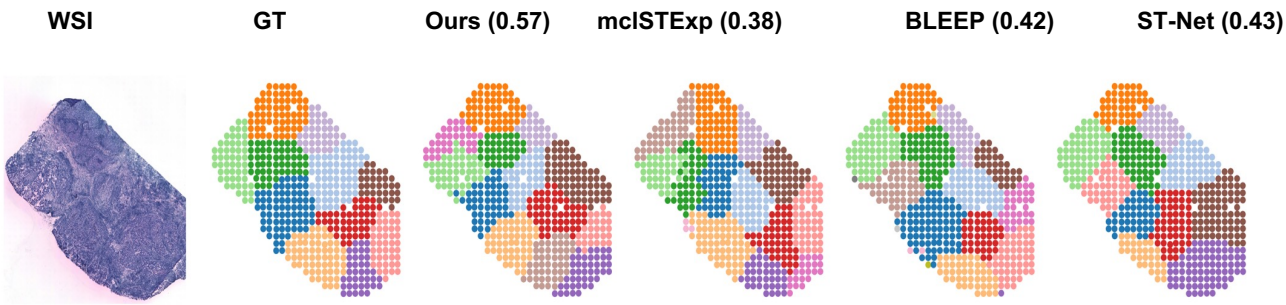

Figure 7. Visualization of leiden clusterings for ground truth and predicted gene expressions for the breast cancer dataset. ARI index shown in (.)

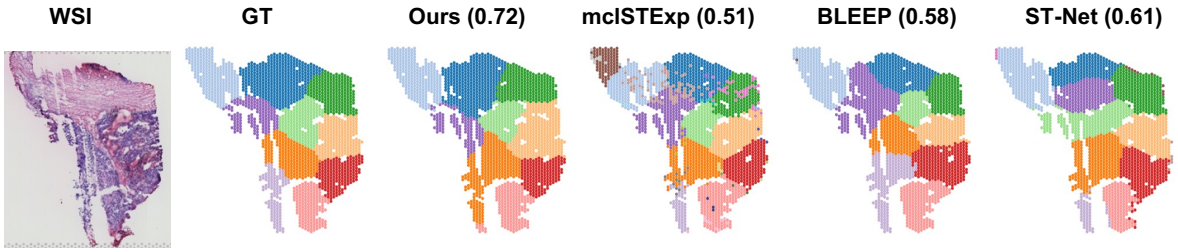

Figure 8. Visualization of leiden clusterings for ground truth and predicted gene expressions for the lymph cancer sample. ARI index shown in (.)

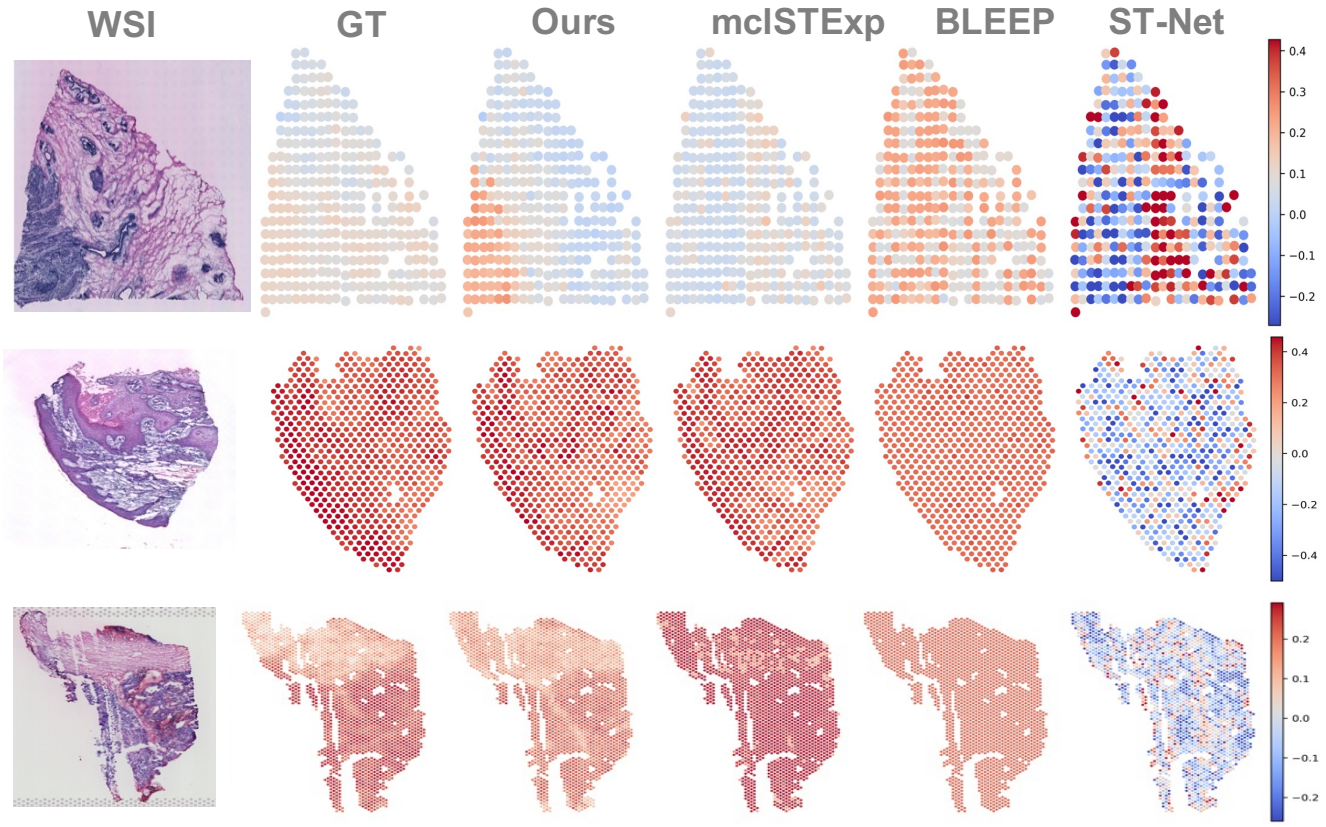

Figure 9. Visualization of pathway heatmaps across the three datasets.We show the `Hallmark_Myc_targets_v1` pathway for the breast sample, `Reactome_Eukaryotic_Translation_Initiation` for the skin sample, and `Reactome_ABC_family_of_proteins_mediated_transport` for the lymph cancer sample.

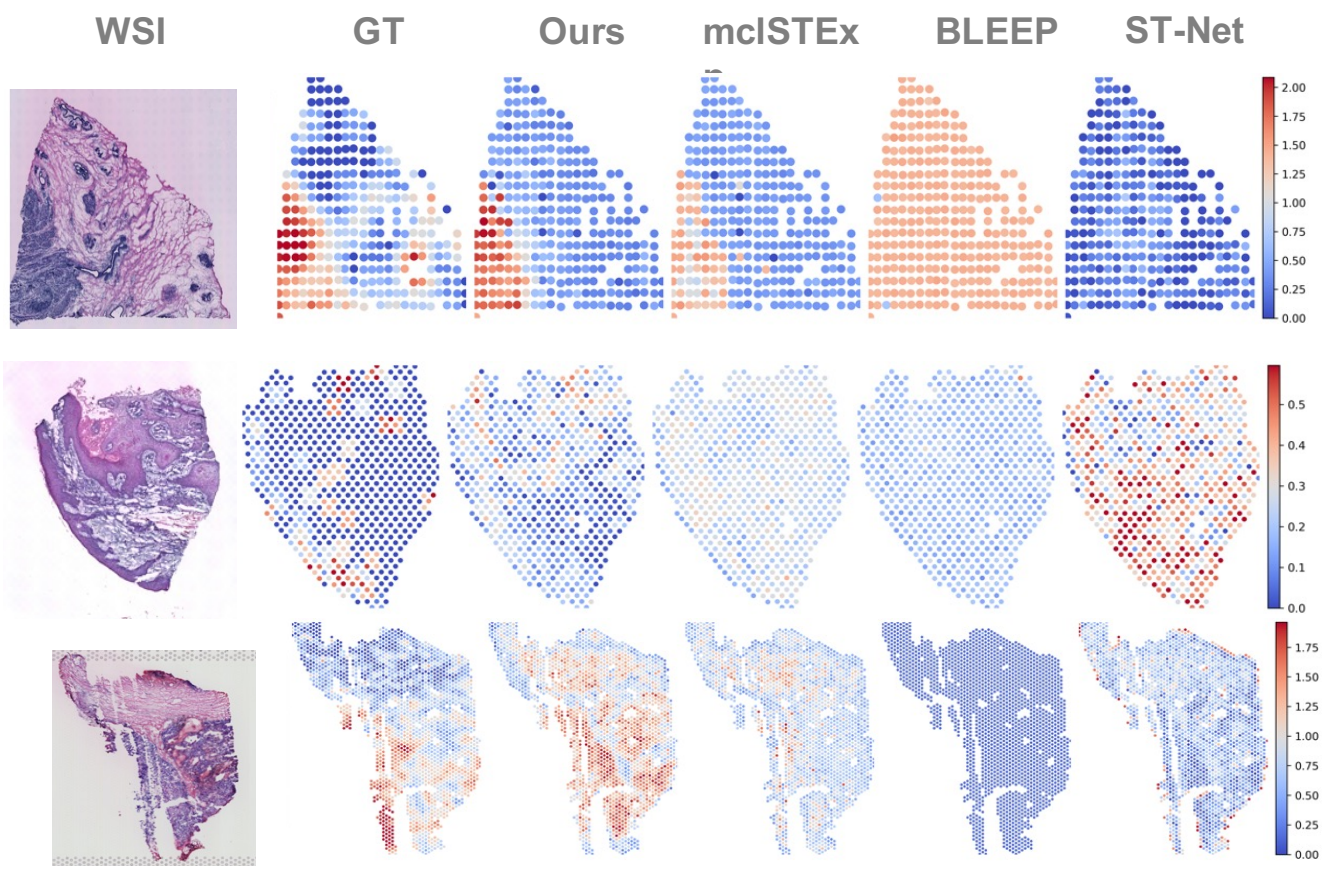

Figure 10. Visualization of the gene heatmaps across the three datasets.We show genes SSBP1, E1F4EBP1, and PKP2 for the breast, skin, and the lymph cancer samples respectively.

References

[1] Luca Abatangelo, Rosalia Maglietta, Angela Distaso, Annarita D'Addabbo, Teresa Maria Creanza, Sayan Mukherjee, and Nicola Ancona. Comparative study of gene set enrichment methods. *BMC bioinformatics*, 10:1–12, 2009. 3, 5

[2] Alma Andersson, Ludvig Larsson, Linnea Stenbeck, Fredrik Salmén, Anna Ehinger, Sunny Z Wu, Ghamdan Al-Eryani, Daniel Roden, Alex Swarbrick, Åke Borg, et al. Spatial deconvolution of her2-positive breast cancer delineates tumor-associated cell type interactions. *Nature communications*, 12 (1):6012, 2021. 2, 5

[3] Siti Haryanti Hairol Anuar, Zuraida Abal Abas, Norhazwani Mohd Yunos, Nurul Hafizah Mohd Zaki, Nurul Akmal Hashim, Mohd Fariddudin Mokhtar, Siti Azirah Asmai, Zaheera Zainal Abidin, and Ahmad Fadzli Nizam. Comparison between louvain and leiden algorithm for network structure: a review. In *Journal of Physics: Conference Series*, page 012028. IOP Publishing, 2021. 8

[4] Keith Baldwin, Surena Namdari, Derek Donegan, Atul F Kamath, and Samir Mehta. Early effects of resident work-hour restrictions on patient safety: a systematic review and plea for improved studies. *JBJS*, 93(2):e5, 2011. 11

[5] Richard J Chen, Ming Y Lu, Wei-Hung Weng, Tiffany Y Chen, Drew FK Williamson, Trevor Manz, Maha Shady, and Faisal Mahmood. Multimodal co-attention transformer for survival prediction in gigapixel whole slide images. In *Proceedings of the IEEE/CVF international conference on computer vision*, pages 4015–4025, 2021. 6

[6] Richard J Chen, Ming Y Lu, Wei-Hung Weng, Tiffany Y Chen, Drew FK Williamson, Trevor Manz, Maha Shady, and Faisal Mahmood. Multimodal co-attention transformer for survival prediction in gigapixel whole slide images. In *Proceedings of the IEEE/CVF international conference on computer vision*, pages 4015–4025, 2021. 1, 3

[7] Richard J Chen, Chengkuan Chen, Yicong Li, Tiffany Y Chen, Andrew D Trister, Rahul G Krishnan, and Faisal Mahmood. Scaling vision transformers to gigapixel images via hierarchical self-supervised learning. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 16144–16155, 2022. 3

[8] Richard J Chen, Tong Ding, Ming Y Lu, Drew FK Williamson, Guillaume Jaume, Bowen Chen, Andrew Zhang, Daniel Shao, Andrew H Song, Muhammad Shaban, et al. A general-purpose self-supervised model for computational pathology. *arXiv preprint arXiv:2308.15474*, 2023. 3, 5

[9] Kunpeng Du, Jingwen Zou, Chunshan Liu, Muhammad Khan, Tao Xie, Xiaoting Huang, Ke Zhang, Yawei Yuan, and Baiyao Wang. A multi-omics pan-cancer analysis of 4ebp1 in cancer prognosis and cancer-associated fibroblasts infiltration. *Frontiers in Genetics*, 13:845751, 2022. 11

[10] Antonio Fabregat, Steven Jupe, Lisa Matthews, Konstantinos Sidiropoulos, Marc Gillespie, Phani Garapati, Robin Haw, Bijay Jassal, Florian Korninger, Bruce May, et al. The reactome pathway knowledgebase. *Nucleic acids research*, 46 (D1):D649–D655, 2018. 5, 11

[11] Simon Graham, Quoc Dang Vu, Shan E Ahmed Raza, Ayesha Azam, Yee Wah Tsang, Jin Tae Kwak, and Nasir Rajpoot. Hover-net: Simultaneous segmentation and classification of nuclei in multi-tissue histology images. *Medical image analysis*, 58:101563, 2019. 1

[12] Mechthild Hatzfeld, Annika Wolf, and René Keil. Plakophilins in desmosomal adhesion and signaling. *Cell communication & adhesion*, 21(1):25–42, 2014. 11

[13] Bryan He, Ludvig Bergenstråhle, Linnea Stenbeck, Abubakar Abid, Alma Andersson, Åke Borg, Jonas Maaskola, Joakim Lundeberg, and James Zou. Integrating spatial gene expression and breast tumour morphology via deep learning. *Nature biomedical engineering*, 4(8): 827–834, 2020. 2, 3

[14] Bryan He, Ludvig Bergenstråhle, Linnea Stenbeck, Abubakar Abid, Alma Andersson, Åke Borg, Jonas Maaskola, Joakim Lundeberg, and James Zou. Integrating spatial gene expression and breast tumour morphology via deep learning. *Nature biomedical engineering*, 4(8): 827–834, 2020. 5, 6

[15] Fabian Hörst, Moritz Rempe, Lukas Heine, Constantin Seibold, Julius Keyl, Giulia Baldini, Selma Ugurel, Jens Siveke, Barbara Grünwald, Jan Egger, et al. Cellvit: Vision transformers for precise cell segmentation and classification. *Medical Image Analysis*, 94:103143, 2024. 1

[16] Zhi Huang, Federico Bianchi, Mert Yuksekgonul, Thomas J Montine, and James Zou. A visual–language foundation model for pathology image analysis using medical twitter. *Nature medicine*, 29(9):2307–2316, 2023. 3

[17] Maximilian Ilse, Jakub Tomczak, and Max Welling. Attention-based deep multiple instance learning. In *International conference on machine learning*, pages 2127–2136. PMLR, 2018. 6

[18] Guillaume Jaume, Anurag Vaidya, Richard J Chen, Drew FK Williamson, Paul Pu Liang, and Faisal Mahmood. Modeling dense multimodal interactions between biological pathways and histology for survival prediction. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 11579–11590, 2024. 3

[19] Guillaume Jaume, Anurag Vaidya, Richard J Chen, Drew FK Williamson, Paul Pu Liang, and Faisal Mahmood. Modeling dense multimodal interactions between biological pathways and histology for survival prediction. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 11579–11590, 2024. 1

[20] Guillaume Jaume, Paul Doucet, Andrew Song, Ming Yang Lu, Cristina Almagro Pérez, Sophia Wagner, Anurag Vaidya, Richard Chen, Drew Williamson, Ahrong Kim, et al. Hest-1k: A dataset for spatial transcriptomics and histology image analysis. *Advances in Neural Information Processing Systems*, 37:53798–53833, 2025. 5

[21] Andrew L Ji, Adam J Rubin, Kim Thrane, Sizun Jiang, David L Reynolds, Robin M Meyers, Margaret G Guo, Benson M George, Annelie Mollbrink, Joseph Bergenstråhle, et al. Multimodal analysis of composition and spatial architecture in human squamous cell carcinoma. *cell*, 182(2): 497–514, 2020. 2, 5

[22] Yuran Jia, Junliang Liu, Li Chen, Tianyi Zhao, and Yadong Wang. Thitogene: a deep learning method for predicting spatial transcriptomics from histological images. *Briefings in Bioinformatics*, 25(1):bbad464, 2024. 2, 3

[23] Hong-Lin Jiang, He-Fen Sun, Shui-Ping Gao, Liang-Dong Li, Sheng Huang, Xin Hu, Sheng Liu, Jiong Wu, Zhi-Ming Shao, and Wei Jin. Ssbp1 suppresses tgfβ-driven epithelial-to-mesenchymal transition and metastasis in triple-negative breast cancer by regulating mitochondrial retrograde signaling. *Cancer research*, 76(4):952–964, 2016. 11

[24] Osamu Kimura, Yasuteru Kondo, and Tooru Shimosegawa. Ppar could contribute to the pathogenesis of hepatocellular carcinoma. *PPAR research*, 2012(1):574180, 2012. 3

[25] Brent M Kuenzi and Trey Ideker. A census of pathway maps in cancer systems biology. *Nature Reviews Cancer*, 20(4): 233–246, 2020. 3

[26] Alain Le Moine, Michel Goldman, and Daniel Abramowicz. Multiple pathways to allograft rejection. *Transplantation*, 73 (9):1373–1381, 2002. 8

[27] Arthur Liberzon, Aravind Subramanian, Reid Pinchback, Helga Thorvaldsdóttir, Pablo Tamayo, and Jill P Mesirov. Molecular signatures database (msigdb) 3.0. *Bioinformatics*, 27(12):1739–1740, 2011. 5, 11

[28] Ming Y Lu, Bowen Chen, Drew F K Williamson, Richard J Chen, Ivy Liang, Tong Ding, Guillaume Jaume, Igor Odintsov, Long Phi Le, Georg Gerber, Anil V Parwani, Andrew Zhang, and Faisal Mahmood. A visual-language foundation model for computational pathology. *Nat. Med.*, 30(3): 863–874, 2024. 3

[29] Teng Ma, Haochen Zhao, Qichang Zhao, and Jianxin Wang. Cox-path: Biological pathway-informed graph neural network for cancer survival prediction. In *Proceedings of the 15th ACM International Conference on Bioinformatics, Computational Biology and Health Informatics*, pages 1–6, 2024. 3

[30] Jose Marino, Joshua Paster, and Gilles Benichou. Allorecognition by t lymphocytes and allograft rejection. *Frontiers in immunology*, 7:582, 2016. 8

[31] Maxime Meylan, Florent Petitprez, Etienne Becht, Antoine Bougoüin, Guilhem Pupier, Anne Calvez, Ilenia Giglioli, Virginie Verkarre, Guillaume Lacroix, Johanna Verneau, et al. Tertiary lymphoid structures generate and propagate anti-tumor antibody-producing plasma cells in renal cell cancer. *Immunity*, 55(3):527–541, 2022. 2, 5

[32] Weiming Mi, Junjie Li, Yucheng Guo, Xinyu Ren, Zhiyong Liang, Tao Zhang, and Hao Zou. Deep learning-based multiclass classification of breast digital pathology images. *Cancer Management and Research*, pages 4605–4617, 2021. 1

[33] Wenwen Min, Zhiceng Shi, Jun Zhang, Jun Wan, and Changmiao Wang. Multimodal contrastive learning for spatial gene expression prediction using histology images. *Briefings in Bioinformatics*, 25(6):bbae551, 2024. 2, 3, 6

[34] Ali Mortazavi, Brian A Williams, Kenneth McCue, Lorian Schaeffer, and Barbara Wold. Mapping and quantifying mammalian transcriptomes by rna-seq. *Nature methods*, 5 (7):621–628, 2008. 1

[35] J Musa, MF Orth, M Dallmayer, M Baldauf, C Pardo, B Rotblat, T Kirchner, G Leprivier, and TGP Grünewald. Eukaryotic initiation factor 4e-binding protein 1 (4e-bp1): a master regulator of mrna translation involved in tumorigenesis. *Oncogene*, 35(36):4675–4688, 2016. 11

[36] M Pang, K Su, and M Li. Leveraging information in spatial transcriptomics to predict super-resolution gene expression from histology images in tumors. biorxiv. 2021; 2021–11. *doi. org/10.1101/2021.11*, 28, 2021. 2, 3

[37] Alec Radford, Jong Wook Kim, Christopher Hallacy, Aditya Ramesh, Gabriel Goh, Sandhini Agarwal, Girish Sastry, Amanda Askell, Pamela Mishkin, Jack Clark, et al. Learning transferable visual models from natural language supervision. In *International Conference on Machine Learning*, pages 8748–8763. PMLR, 2021. 3

[38] U Raghavendra, U Rajendra Acharya, and Hojjat Adeli. Artificial intelligence techniques for automated diagnosis of neurological disorders. *European neurology*, 82(1-3):41–64, 2020. 1

[39] Anjali Rao, Dalia Barkley, Gustavo S França, and Itai Yanai. Exploring tissue architecture using spatial transcriptomics. *Nature*, 596(7871):211–220, 2021. 1

[40] Robert W Robey, Kristen M Pluchino, Matthew D Hall, Antonio T Fojo, Susan E Bates, and Michael M Gottesman. Revisiting the role of abc transporters in multidrug-resistant cancer. *Nature Reviews Cancer*, 18(7):452–464, 2018. 11

[41] Amy Schulze, Masanori Oshi, Itaru Endo, and Kazuaki Takabe. Myc targets scores are associated with cancer aggressiveness and poor survival in er-positive primary and metastatic breast cancer. *International journal of molecular sciences*, 21(21):8127, 2020. 11

[42] Patrik L Ståhl, Fredrik Salmén, Sanja Vickovic, Anna Lundmark, José Fernández Navarro, Jens Magnusson, Stefania Giacomello, Michaela Asp, Jakub O Westholm, Mikael Huss, et al. Visualization and analysis of gene expression in tissue sections by spatial transcriptomics. *Science*, 353 (6294):78–82, 2016. 1

[43] Aravind Subramanian, Pablo Tamayo, Vamsi K Mootha, Sayan Mukherjee, Benjamin L Ebert, Michael A Gillette, Amanda Paulovich, Scott L Pomeroy, Todd R Golub, Eric S Lander, et al. Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. *Proceedings of the National Academy of Sciences*, 102(43):15545–15550, 2005. 3

[44] PJ Sudharshan, Caroline Petitjean, Fabio Spanhol, Luiz Eduardo Oliveira, Laurent Heutte, and Paul Honeine. Multiple instance learning for histopathological breast cancer image classification. *Expert Systems with Applications*, 117:103–111, 2019. 3

[45] Valentine Svensson, Roser Vento-Tormo, and Sarah A Teichmann. Exponential scaling of single-cell rna-seq in the last decade. *Nature protocols*, 13(4):599–604, 2018. 1

[46] Fuchou Tang, Catalin Barbacioru, Yangzhou Wang, Ellen Nordman, Clarence Lee, Nanlan Xu, Xiaohui Wang, John Bodeau, Brian B Tuch, Asim Siddiqui, et al. mrna-seq whole-transcriptome analysis of a single cell. *Nature methods*, 6(5):377–382, 2009. 1

[47] Vincent A Traag, Ludo Waltman, and Nees Jan Van Eck. From louvain to leiden: guaranteeing well-connected communities. *Scientific reports*, 9(1):1–12, 2019. 8

[48] Quoc Dang Vu, Simon Graham, Tahsin Kurc, Minh Nguyen Nhat To, Muhammad Shaban, Talha Qaiser, Navid Alemi Koohbanani, Syed Ali Khurram, Jayashree Kalpathy-Cramer, Tianhao Zhao, et al. Methods for segmentation and classification of digital microscopy tissue images. *Frontiers in bioengineering and biotechnology*, 7:53, 2019. 1

[49] Xiyue Wang, Sen Yang, Jun Zhang, Minghui Wang, Jing Zhang, Wei Yang, Junzhou Huang, and Xiao Han. Transformer-based unsupervised contrastive learning for histopathological image classification. *Medical image analysis*, 81:102559, 2022. 3

[50] Zhong Wang, Mark Gerstein, and Michael Snyder. Rna-seq: a revolutionary tool for transcriptomics. *Nature reviews genetics*, 10(1):57–63, 2009. 6

[51] F Alexander Wolf, Philipp Angerer, and Fabian J Theis. Scanpy: large-scale single-cell gene expression data analysis. *Genome biology*, 19(1):15, 2018. 5

[52] Ellery Wulczyn, David F Steiner, Zhaoyang Xu, Apaar Sadhwani, Hongwu Wang, Isabelle Flament-Auvigne, Craig H Mermel, Po-Hsuan Cameron Chen, Yun Liu, and Martin C Stumpe. Deep learning-based survival prediction for multiple cancer types using histopathology images. *PloS one*, 15 (6):e0233678, 2020. 1, 3

[53] Ronald Xie, Kuan Pang, Sai Chung, Catia Perciani, Sonya MacParland, Bo Wang, and Gary Bader. Spatially resolved gene expression prediction from histology images via bimodal contrastive learning. *Advances in Neural Information Processing Systems*, 36:70626–70637, 2023. 2, 3, 6

[54] Hanwen Xu, Naoto Usuyama, Jaspreet Bagga, Sheng Zhang, Rajesh Rao, Tristan Naumann, Cliff Wong, Zelalem Gero, Javier González, Yu Gu, Yanbo Xu, Mu Wei, Wenhui Wang, Shuming Ma, Furu Wei, Jianwei Yang, Chunyuan Li, Jianfeng Gao, Jaylen Rosemon, Tucker Bower, Soohee Lee, Roshanthi Weerasinghe, Bill J Wright, Ari Robicsek, Brian Piening, Carlo Bifulco, Sheng Wang, and Hoifung Poon. A whole-slide foundation model for digital pathology from real-world data. *Nature*, 630(8015):181–188, 2024. 3

[55] Yingxue Xu and Hao Chen. Multimodal optimal transport-based co-attention transformer with global structure consistency for survival prediction. In *Proceedings of the IEEE/CVF international conference on computer vision*, pages 21241–21251, 2023. 6

[56] Yan Xu, Jun-Yan Zhu, Eric Chang, and Zhuowen Tu. Multiple clustered instance learning for histopathology cancer image classification, segmentation and clustering. In *2012 IEEE Conference on Computer Vision and Pattern Recognition*, pages 964–971. IEEE, 2012. 3

[57] Yuansong Zeng, Zhuoyi Wei, Weijiang Yu, Rui Yin, Yuchen Yuan, Bingling Li, Zhonghui Tang, Yutong Lu, and Yuedong Yang. Spatial transcriptomics prediction from histology jointly through transformer and graph neural networks. *Brief. Bioinform.*, 23(5), 2022. 2, 3

[58] Yuansong Zeng, Zhuoyi Wei, Weijiang Yu, Rui Yin, Yuchen Yuan, Bingling Li, Zhonghui Tang, Yutong Lu, and Yuedong Yang. Spatial transcriptomics prediction from histology jointly through transformer and graph neural networks. *Briefings in Bioinformatics*, 23(5), 2022. 5, 6

[59] Yilan Zhang, Yingxue Xu, Jianqi Chen, Fengying Xie, and Hao Chen. Prototypical information bottlenecking and disentangling for multimodal cancer survival prediction. *arXiv preprint arXiv:2401.01646*, 2024. 3