



This MICCAI paper is the Open Access version, provided by the MICCAI Society. It is identical to the accepted version, except for the format and this watermark; the final published version is available on SpringerLink.

Benchmarking Pathology Foundation Models for Spatial Domain Understanding

Bokai Zhao^{1,2,3,4,*}, Yiyang Zhang^{2,3,*}, Yuanchi Zhu^{2,3,5,*}, Hanqing Chao^{4,6}, Long Bai^{4,6}, Tai Ma^{4,6}, Minfeng Xu^{4,6}, Ming Song^{1,2,3}, and Tianzi Jiang^{1,2,3,5}✉

¹School of Artificial Intelligence, University of Chinese Academy of Sciences.

²Brainnetome Center, Institute of Automation, Chinese Academy of Sciences.

³Beijing Key Laboratory of Brainnetome and Brain-Computer Interface, Institute of Automation, Chinese Academy of Sciences.

correspondence author ✉: jiangtz@nlpr.ia.ac.cn

⁴DAMO Academy, Alibaba Group. ⁵ShanghaiTech University. ⁶Hupan Lab.

Abstract. Pathology foundation models (PFMs) have emerged as a core approach for learning transferable representations from whole slide images (WSIs), and they are typically benchmarked through downstream clinical endpoints. While such task level evaluations are indispensable, they offer limited insight into what the representations themselves encode, particularly whether PFM embeddings can distinguish meaningful tissue regions and capture their spatial relationships. We present SpaPath-Bench, a representation level benchmark designed to diagnose spatial representation capability in PFMs. SpaPath-Bench formulates spatial domain identification (SDI) on paired whole slide image and spatial transcriptomics (ST) data as a diagnostic task. It curates 42 public paired WSI and ST slides, enables large scale evaluation across 19 encoders and seven SDI methods, and measures partition quality using three complementary criteria: unsupervised spatial coherence, transcriptomics referenced agreement, and expert referenced agreement. Across 83K runs, SpaPath-Bench reveals that different pretraining paradigms capture distinct aspects of tissue spatial architecture, and it provides practical guidance for building the next generation of spatially aware computational pathology models. Code and data pipelines are publicly available at <https://bokai-zhao.github.io/SpaPath-benchboard/>.

Keywords: Pathology Foundation Model · Benchmark · Spatial Omics · Spatial Domain Identification.

1 Introduction

Computational pathology develops algorithms that interpret whole-slide images (WSIs) to support diagnosis, prognosis, and biomarker discovery at scale. Beyond recognizing cellular morphology, many diagnostic and prognostic cues arise from how multiple tissue components form coherent regions and interact across a section—for example, epithelial versus stromal areas, necrotic regions, or immune-rich neighborhoods. The ability to separate meaningful tissue regions and to

capture their spatial neighborhood structure is therefore a foundational capability that underpins a wide range of clinical and research applications [20].

Pathology foundation models (PFMs) have recently become the dominant approach for learning transferable WSI representations [27]. By self-supervised pre-training on large collections of unlabeled histology, models such as UNI [3], Virchow [22], and GigaPath [29] provide strong generic image embeddings. Vision-language approaches, including CONCH [13] and MUSK [26], further aim to enrich these embeddings by aligning histological patterns with medical text. Accordingly, current benchmarking efforts largely focus on downstream endpoints, e.g., mutation prediction, survival modeling, and cancer subtyping [31,2,15]. These task-level evaluations are indispensable as they reflect practical use cases. However, they do not directly characterize what the representation itself encodes[16]: before adding task-specific heads and training aggregators, to what extent do the embeddings themselves separate meaningful tissue regions and respect their spatial continuity within a section?

We propose to answer this question using paired WSI-spatial transcriptomics (ST) datasets, where gene expression is profiled at spatially indexed spots whose coordinates are registered to the matched pathology image from the same section[1]. A common first step in ST analysis is spatial domain identification (SDI), which partitions a section into contiguous tissue regions that share similar expression profiles[30,23,11]. SDI is widely used because the inferred regions often correspond to coherent anatomical or functional tissue regions, and many SDI methods further incorporate the matched pathology image as a morphological [17]. This makes SDI a natural diagnostic for pathology representations: if PFM image embeddings capture meaningful tissue semantics, then spot-level embeddings extracted from the matched WSI should support SDI-style partitioning into spatially coherent regions that are biologically or anatomically meaningful regions.

Concretely, for each paired sample we extract spot-level PFM embeddings at the ST spot coordinates and treat embedding dimensions as “pseudo-expression” variables, forming a spot-by-feature matrix analogous to the gene-expression matrix used by SDI. We then apply established SDI pipelines on this pseudo-expression matrix to obtain a spatial partition. We evaluate the resulting partition using three parallel criteria: (i) unsupervised spatial coherence metrics that quantify whether inferred regions are spatially smooth and well-structured; (ii) transcriptomics-referenced agreement, comparing embedding-derived partitions against SDI partitions obtained from gene expression as a molecularly grounded target; and (iii) expert-referenced agreement, comparing against manual anatomical annotations when available. Together, these evaluations test whether PFM embeddings support both region separability and spatial neighborhood consistency, and whether the resulting partitions align with transcriptomics-derived structure or expert tissue labels.

Based on this view, we introduce *SpaPath-Bench*, a systematic benchmark for evaluating spatial organization encoded in pathology foundation models. Our contributions are threefold:

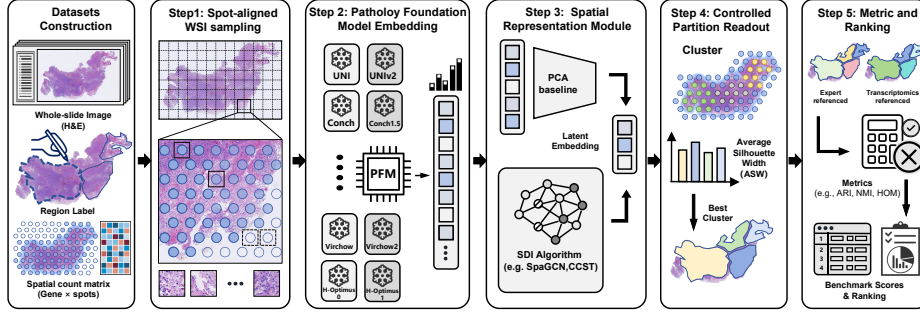


Fig. 1. Overview of the benchmark pipeline for spatial domain understanding.

- We formulate SDI on PFM image embeddings as a representation-level benchmark that complements conventional downstream task-level evaluation.
- We provide a three-way evaluation protocol: unsupervised spatial coherence, transcriptomics-referenced agreement, and expert-referenced agreement, capturing distinct yet meaningful notions of region validity.
- We conduct a large-scale factorial study on 42 public paired WSI-ST datasets across 19 encoders, 7 SDI methods, and multiple clustering algorithms, yielding 83k total runs with statistical reporting.

2 Benchmark Protocol

2.1 Benchmark Pipeline

Given a spatial transcriptomics (ST) slide with a paired H&E image, we denote the measured spots as $\mathcal{S} = \{s_i\}_{i=1}^N$, where each spot s_i is associated with a 2D physical coordinate $\mathbf{p}_i \in \mathbb{R}^2$ in the tissue plane. Our benchmark evaluates whether pathology foundation model (PFM) representations preserve information sufficient for recovering *spatial domain structures* by standardizing the pipeline from spot-level feature extraction to domain partitioning and metric computation.

Step 1: Spot-aligned WSI sampling. For each spot $s_i \in \mathcal{S}$, we extract an aligned histology patch x_i from the paired H&E image centered at $\mathbf{p}_i$ (using a fixed magnification and patch size). We apply only model-input standardization, including cropping and intensity normalization, to obtain the spot-level image inputs $\mathcal{X} = \{x_i\}_{i=1}^N$.

Step 2: Foundation model embedding. Given a pre-trained encoder f_θ (a PFM or a baseline vision model), we compute a D -dimensional embedding for each patch as $\mathbf{z}_i = f_\theta(x_i) \in \mathbb{R}^D$. Stacking these embeddings forms the feature matrix $\mathbf{Z} = [\mathbf{z}_1, \dots, \mathbf{z}_N]^\top \in \mathbb{R}^{N \times D}$. We conceptualize $\mathbf{Z}$ as a *pseudo-expression* matrix, where each latent dimension acts as a *virtual gene* measured at that spot, aligning histology-derived representations with standard ST analysis workflows.

Step 3: Spatial representation module (PCA vs. SDI-based). To incorporate spatial context, we transform $\mathbf{Z}$ into a spatially informed representation $\mathbf{H} \in \mathbb{R}^{N \times d}$ via one of two routes: (i) **Non-spatial baseline (PCA):** $\mathbf{H} = \text{PCA}_d(\mathbf{Z})$ with $d \in \{32, 64, 128, 256\}$. (ii) **SDI-based spatial representation:** We construct a spatial neighborhood graph $G = (V, E)$ over $\mathcal{S}$ using k -nearest neighbors ($k = 6$) based on the coordinate matrix $\mathbf{P} = [\mathbf{p}_1, \dots, \mathbf{p}_N]^\top \in \mathbb{R}^{N \times 2}$. An SDI method g_ϕ then learns the spatial representations as $\mathbf{H} = g_\phi(\mathbf{Z}, \mathbf{P}, G)$, acting as a standardized *spatial representation learner* operating on the pseudo-expression space to promote neighborhood coherence.

Step 4: Controlled partition readout (clustering). Given $\mathbf{H}$, we generate a spatial domain partition $\hat{\mathbf{y}} \in \{1, \dots, K\}^N$ using a controlled clustering function $c \in \{\text{K-means, Louvain, Leiden}\}$, i.e., $\hat{\mathbf{y}} = c(\mathbf{H})$. For expert-annotated slides, K matches the number of reference domains; for unlabeled slides, K is determined via reference-free criteria (Sec. 2.2). Decoupling *representation learning* (PCA/SDI) from the *partition readout* (clustering) makes performance differences more directly attributable to the learned embeddings under consistent readout choices.

Step 5: Metric computation and ranking. We quantify spatial domain fidelity by comparing $\hat{\mathbf{y}}$ against available references (transcriptomics-referenced agreement and expert-referenced agreement), alongside reference-free spatial coherence metrics (Sec. 2.2). Results are aggregated across slides and three independent random seeds to report robust summary statistics and rankings.

2.2 Evaluation Metrics Protocol

We quantify spatial domain understanding via three complementary regimes: (i) structural alignment with expert knowledge, (ii) biological concordance with molecular profiles, and (iii) intrinsic spatial coherence.

Reference-free Metrics. To assess intrinsic spatial quality without any external labels, we compute: (i) CHAOS, to measure spatial dispersion; (ii) Patch Adjacency Score (PAS), to evaluate local neighborhood consistency; and (iii) Average Silhouette Width (ASW), to quantify cluster separation in the latent representation space.

Reference-based Metrics. Given a predicted partition $\hat{\mathbf{y}}$ and a reference $\mathbf{y}^*$, we compute Adjusted Rand Index (ARI), Normalized Mutual Information (NMI), homogeneity (HOM), and completeness (COM).

(1) *Expert-referenced (labeled slides):* For slides with manual annotations, we set the cluster number to the ground truth, $K = |\mathcal{Y}^{\text{expert}}|$, and compute agreement against expert labels.

(2) *Transcriptomics-referenced (unlabeled slides):* For slides lacking manual labels, we determine $\hat{\mathbf{y}}$ by maximizing the ASW over $K \in \{4, \dots, 12\}$. A proxy reference $\mathbf{y}^{\text{HVG}}$ is independently constructed by applying the same SDI and clustering pipeline to the top 2000 highly variable genes (HVGs), with its K similarly selected via maximum ASW. Agreement between $\hat{\mathbf{y}}$ and $\mathbf{y}^{\text{HVG}}$ quantifies the alignment between morphology-derived patterns and molecular-defined functional domains.

Rank-score and Statistical Robustness. To ensure robust comparison across 3 random seeds and 42 slides, we perform one-sided Wilcoxon signed-rank tests under matched configurations (same SDI method, clustering readout, and seed). A model m_i earns a rank-score $S(m_i, c) = \sum_{j \neq i} \mathbb{I}(p_{i,j,c} < 0.05)$, representing the number of competitors it significantly outperforms. Overall benchmark rankings are obtained by aggregating $S(m_i, c)$ across all configurations and slides.

3 Experiments

3.1 Benchmark Datasets Construction

We curate a benchmark of 42 publicly available spatial transcriptomics (ST) slides with paired H&E images, comprising one expert-annotated cohort and one unlabeled cohort.

- **Expert-annotated cohort.** We include 12 human dorsolateral prefrontal cortex (DLPFC) slides [9] with spot-level expert annotations into 7 anatomical regions (Layers 1-6 and White Matter). This cohort enables reference-based evaluation of domain recovery against human-interpretable anatomical boundaries.

- **Unlabeled HE-ST cohort.** We collect 30 additional slides without expert annotations, spanning multiple tissues and both human and mouse samples. For these slides, we quantify morphology-molecule consistency using the omics-guided proxy evaluation protocol (Sec. 2.2), comparing morphology-derived partitions with transcriptomics-derived HVG reference partitions.

Table 1. Summary of evaluated foundation models and pre-training details.

Modality	PFM	Arch	Input (px)	Dim D	Params.	Data		Pretrain
						#Slides	#Patches	Methods
Pure Vision	H-Optimus-1	ViT-G/14	224	1536	1.1B	1M	2.0B	–
	H-Optimus-0	ViT-G/14	224	1536	1.1B	500K	–	DINOv2
	GigaPath	ViT-G/14	224	1536	1.1B	171K	1.4B	DINOv2
	UNI2	ViT-H/14	224	1536	681M	350K	200M	DINOv2
	Virchow2	ViT-H/14	224	1280	631M	3134K	2.0B	DINOv2
	Virchow	ViT-H/14	224	1280	631M	1488K	2.0B	DINOv2
	UNI	ViT-L/16	224	1024	304M	100K	100M	DINOv2
	Phikon2	ViT-L/16	224	1024	304M	58K	456M	DINOv2
	Hibou-L	ViT-L/16	224	1024	304M	1100K	1.2B	DINOv2
	Hibou-B	ViT-B/16	224	768	86M	1100K	1.2B	DINOv2
Vision- Language	Phikon	ViT-B/16	224	768	86M	6,093	43.4M	iBOT
	CtransPath	Swin-T/14	224	768	28M	60,530	15.6M	MoCov3
	MUSK	ViT-L/16	384	2048	304M	–	1M	BEiT-3
	CONCH-V1.5	ViT-L/16	512	768	304M	–	1.26M	CoCa
	CONCH	ViT-B/16	224	512	86M	–	1.17M	CoCa
V-Omics	PLIP	ViT-B/32	224	512	86M	–	208K	CLIP
	OmicsCLIP	ViT-L/14	224	768	304M	1,007	2.2M	CoCa
Natural Image	DINOv3-ViT-l	ViT-L/16	256	1024	304M	–	–	DINOv3
	DINOv3-ViT-b	ViT-B/16	256	768	86M	–	–	DINOv3

3.2 Experimental Setup

Foundation Models (Encoders). We systematically evaluate 19 foundation models, categorizing them into four distinct pre-training paradigms: (i) *Pure Vision* PFMs (e.g., UNI/UNI2[3], Virchow/Virchow2[22], H-Optimus-0/1, GigaPath[29], Phikon/Phikon2[6], Hibou-B/L[14], CtransPath[24]); (ii) *Vision-Language* PFMs (CONCH/ CONCH-V1.5[13], PLIP[8], MUSK[26]), which align morphological features with medical text; (iii) *Vision-Omics* models (OmicsCLIP[4]), bridging histology and molecular profiles; and (iv) *Natural Image* baselines (DINOv3-ViT-B/L[19]). Detailed specifications of each model, including architecture, parameter count, and pre-training corpus, are summarized in Table 1.

SDI Methods and Clustering. The high-dimensional representations extracted by the frozen PFMs are processed through a unified spatial module. We compare conventional dimensionality reduction (PCA) against 7 mainstream Spatial Domain Identification (SDI) algorithms, systematically categorized by their core learning paradigms: (i) *Graph Clustering* (SpaGCN[7], CCST[10]); (ii) *Graph Autoencoders* (SEDR[28], STAGATE[5]); (iii) *Contrastive Learning* (conST[32], GraphST[12]); and (iv) *Manifold Learning* (SpaceFlow[18]). Unlike PCA, these graph-based methods explicitly model tissue topology by constructing spatial adjacency graphs from spot coordinates. The resulting spatially-aware embeddings are partitioned into functional domains using three clustering strategies: K-means, Louvain[25], and Leiden[21]. To ensure a fair and reproducible benchmark, we strictly adhere to the default hyperparameters (e.g., learning rates, latent dimensions, training epochs) recommended by the respective original publications. All detailed configurations are comprehensively summarized in Table 2, and all evaluations are rigorously repeated across 3 independent random seeds.

Table 2. Summary of evaluated SDI methods and default hyperparameters.

Category	Graph Clustering		Graph Autoencoder		Contrastive		Manifold
Method	SpaGCN	CCST	SEDR	STAGATE	conST	GraphST	SpaceFlow
Dims	50	256	32	30	28	28	50
Epochs	2000	5000	200	1000	200	500	1000
LR	0.05	10^{-6}	0.01	10^{-3}	0.01	5×10^{-4}	10^{-3}

4 Results

4.1 Overall Benchmark Summary

Pipeline Stability (Fig.2A). We evaluated the stability of our pipeline by comparing metric distributions across three independent random seeds. Paired Wilcoxon signed-rank tests revealed no statistically significant differences ($p >$

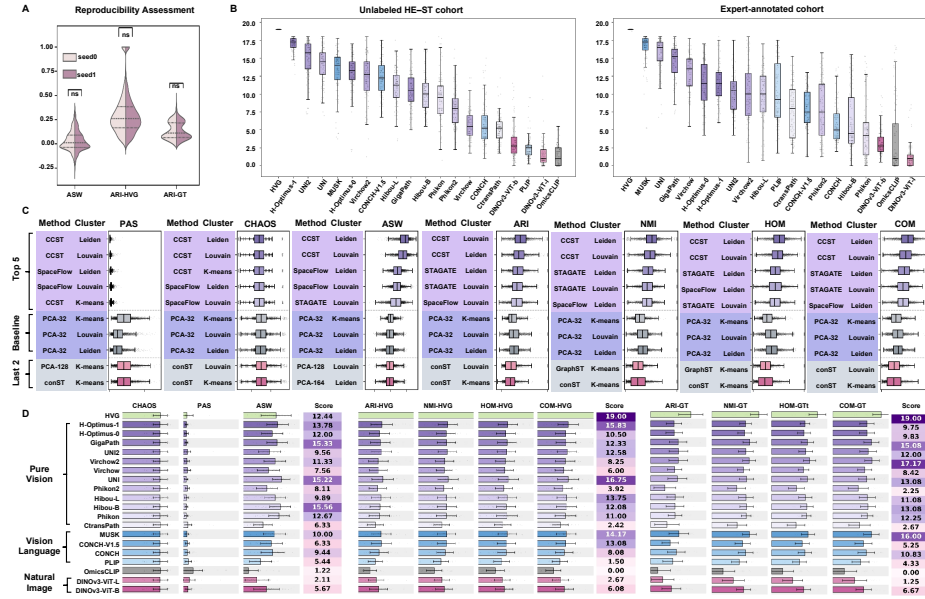


Fig. 2. Overall benchmark summary evaluating PFM spatial domain understanding.

0.05) across all evaluation metrics, indicating that the benchmark outcomes are not driven by stochastic initialization.

Global Model Rankings (Fig.2B). Aggregating evaluation metrics across all 42 ST slides and all clustering methods, we established a global ranking for the 19 evaluated models. When assessed against the transcriptomic proxy (HVG-guided reference), **H-Optimus-1** achieved the highest overall rank. Conversely, when evaluated against expert manual annotations (the DLPFC subset), **MUSK** recorded the highest performance metrics.

Efficacy of Spatial Representations (Fig.2C). We systematically compared diverse combinations of SDI methods and clustering algorithms. For each metric, we visualize the top five configurations alongside three PCA baselines and the two lowest-ranked methods ; among all evaluated settings, **CCST+Leiden** yields the best aggregate performance.

Model Rankings under Optimal Setting (Fig.2D). As a specific case study utilizing the optimal **CCST+Leiden** configuration, we summarized the evaluation metrics across all 42 ST slides to present the detailed performance ranking of the 19 PFMs under this high-performing setting.

4.2 Key Findings

Reference-dependent representations. **H-Optimus-1** shows strong performance when evaluated against the transcriptomic proxy (HVG), implying self-supervised learning can capture fine-grained molecular heterogeneity. Conversely,

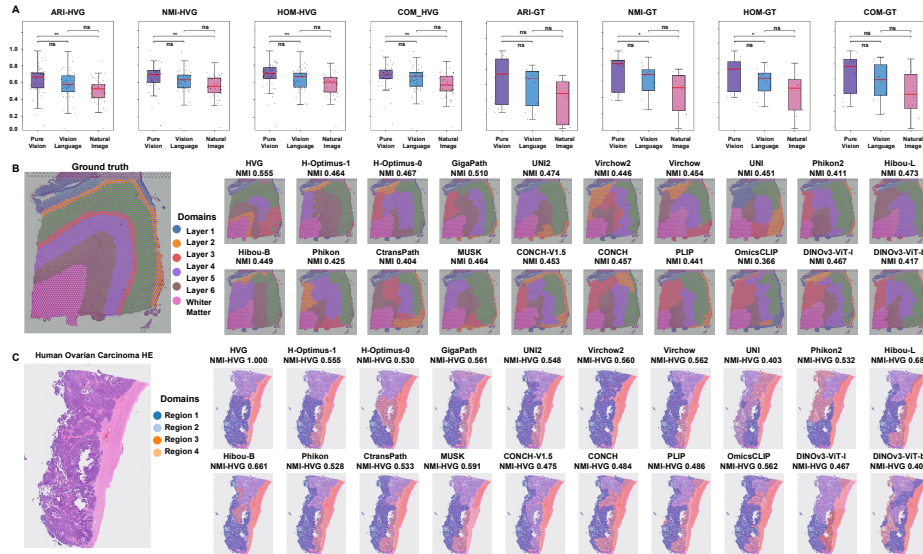


Fig. 3. Pre-training paradigm impacts (A) and qualitative spatial domain visualizations on DLPFC (B) and ovarian carcinoma (C) slides.

MUSK achieves top results on expert annotations, suggesting vision-language pre-training may better align with human-defined macro-architectural semantics. **WSI-level spatial contrastive learning.** CCST generally outperforms baselines, indicating that integrating whole-slide contrastive learning helps aggregate regional context for holistic tissue comprehension, improving upon isolated patch features. **Pre-training paradigm impacts.** (Fig.3A) While pathology models tend to outperform natural-image baselines, vision-language models do not consistently show a distinct advantage over pure-vision counterparts. Strict patch-level text-image alignment might potentially limit the macroscopic spatial coherence beneficial for tissue-level clustering.

4.3 Qualitative Visual Assessment

Qualitative visualizations on two representative slides illustrate how PFM-derived embeddings translate into spatial domains under our pipeline. **Expert-annotated case (DLPFC, Fig.3B):** Because cortical layers exhibit subtle and gradual morphological transitions, histology-driven partitions show only partial agreement with expert annotations. This is consistent with the fact that expert layer labels encode high-level neuroanatomical context that may not be fully determined by local H&E appearance alone. **Unlabeled case (Ovarian carcinoma, Fig.3C):** In contrast, for tissues with pronounced structural heterogeneity, PFM-based domains delineate distinct microenvironments and align well with visually apparent macro-scale boundaries.

5 Summary

We introduced *SpaPath-Bench*, a standardized benchmark to assess spatial organization encoded in PFMs using SDI on spot-aligned embeddings. With a unified pipeline and three evaluation regimes (expert-referenced, transcriptomics-referenced, and reference-free spatial coherence), we systematically compare 19 encoders on 42 paired WSI-ST slides and identify strengths and bottlenecks in region separability and neighborhood coherence. We envision *SpaPath-Bench* as a testbed for developing more robust, context-aware pathology representations that better integrate morphology with spatial omics.

Acknowledgments. This work was supported by Brain Science and Brain-like Intelligence Technology - National Science and Technology Major Project 2021ZD0200200.

Disclosure of Interests. The authors declare that they have no competing interests.

References

1. Bressan, D., Battistoni, G., Hannon, G.J.: The dawn of spatial omics. *Science* **381**(6657), eabq4964 (2023)
2. Campanella, G., Chen, S., Singh, M., Verma, R., Muehlstedt, S., Zeng, J., Stock, A., Croken, M., Veremis, B., Elmas, A., et al.: A clinical benchmark of public self-supervised pathology foundation models. *Nature Communications* **16**(1), 3640 (2025)
3. 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)
4. Chen, W., Zhang, P., Tran, T.N., Xiao, Y., Li, S., Shah, V.V., Cheng, H., Brannan, K.W., Youker, K., Lai, L., et al.: A visual-omics foundation model to bridge histopathology with spatial transcriptomics. *Nature Methods* **22**(7), 1568–1582 (2025)
5. Dong, K., Zhang, S.: Deciphering spatial domains from spatially resolved transcriptomics with an adaptive graph attention auto-encoder. *Nature communications* **13**(1), 1739 (2022)
6. Filiot, A., Jacob, P., Kain, A.M., Saillard, C.: Phikon-v2, a large and public feature extractor for biomarker prediction (2024), <https://arxiv.org/abs/2409.09173>
7. Hu, J., Li, X., Coleman, K., Schroeder, A., Ma, N., Irwin, D.J., Lee, E.B., Shinohara, R.T., Li, M.: Spagcn: Integrating gene expression, spatial location and histology to identify spatial domains and spatially variable genes by graph convolutional network. *Nature methods* **18**(11), 1342–1351 (2021)
8. Huang, Z., Bianchi, F., Yuksekgonul, M., Montine, T.J., Zou, J.: A visual-language foundation model for pathology image analysis using medical twitter. *Nature medicine* **29**(9), 2307–2316 (2023)
9. Huuki-Myers, L.A., Spangler, A., Eagles, N.J., Montgomery, K.D., Kwon, S.H., Guo, B., Grant-Peters, M., Divecha, H.R., Tippani, M., Sriworarat, C., et al.: A data-driven single-cell and spatial transcriptomic map of the human prefrontal cortex. *Science* **384**(6698), eadh1938 (2024)

10. Li, J., Chen, S., Pan, X., Yuan, Y., Shen, H.B.: Cell clustering for spatial transcriptomics data with graph neural networks. *Nature Computational Science* **2**(6), 399–408 (2022)
11. Liu, Y., Wang, C., Wang, Z., Chen, L., Li, Z., Song, J., Zou, Q., Gao, R., Qian, B.Z., Feng, X., et al.: High-parameter spatial multi-omics through histology-anchored integration. *Nature Methods* pp. 1–14 (2025)
12. Long, Y., Ang, K.S., Li, M., Chong, K.L.K., Sethi, R., Zhong, C., Xu, H., Ong, Z., Sachaphibulkij, K., Chen, A., et al.: Spatially informed clustering, integration, and deconvolution of spatial transcriptomics with graphst. *Nature communications* **14**(1), 1155 (2023)
13. 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)
14. Nechaev, D., Pchelnikov, A., Ivanova, E.: Hibou: A family of foundational vision transformers for pathology. *arXiv preprint arXiv:2406.05074* (2024)
15. Neidlinger, P., El Nahhas, O.S., Muti, H.S., Lenz, T., Hoffmeister, M., Brenner, H., van Treeck, M., Langer, R., Dislich, B., Behrens, H.M., et al.: Benchmarking foundation models as feature extractors for weakly supervised computational pathology. *Nature biomedical engineering* pp. 1–11 (2025)
16. Pham, D., Tan, X., Balderson, B., Xu, J., Grice, L.F., Yoon, S., Willis, E.F., Tran, M., Lam, P.Y., Raghobar, A., et al.: Robust mapping of spatiotemporal trajectories and cell–cell interactions in healthy and diseased tissues. *Nature communications* **14**(1), 7739 (2023)
17. Rao, A., Barkley, D., França, G.S., Yanai, I.: Exploring tissue architecture using spatial transcriptomics. *Nature* **596**(7871), 211–220 (2021)
18. Ren, H., Walker, B.L., Cang, Z., Nie, Q.: Identifying multicellular spatiotemporal organization of cells with spaceflow. *Nature communications* **13**(1), 4076 (2022)
19. Siméoni, O., Vo, H.V., Seitzer, M., Baldassarre, F., Oquab, M., Jose, C., Khalidov, V., Szafraniec, M., Yi, S., Ramamonjisoa, M., Massa, F., Haziza, D., Wehrstedt, L., Wang, J., Darcet, T., Moutakanni, T., Sentana, L., Roberts, C., Vedaldi, A., Tolan, J., Brandt, J., Couprie, C., Mairal, J., Jégou, H., Labatut, P., Bojanowski, P.: DINOv3 (2025), <https://arxiv.org/abs/2508.10104>
20. Song, A.H., Jaume, G., Williamson, D.F., Lu, M.Y., Vaidya, A., Miller, T.R., Mahmood, F.: Artificial intelligence for digital and computational pathology. *Nature Reviews Bioengineering* **1**(12), 930–949 (2023)
21. Traag, V.A., Waltman, L., Van Eck, N.J.: From louvain to leiden: guaranteeing well-connected communities. *Scientific reports* **9**(1), 5233 (2019)
22. 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)
23. Wang, C., Chan, A.S., Fu, X., Ghazanfar, S., Kim, J., Patrick, E., Yang, J.Y.: Benchmarking the translational potential of spatial gene expression prediction from histology. *Nature Communications* **16**(1), 1544 (2025)
24. Wang, X., Zhao, J., Marostica, E., Yuan, W., Jin, J., Zhang, J., Li, R., Tang, H., Wang, K., Li, Y., et al.: A pathology foundation model for cancer diagnosis and prognosis prediction. *Nature* **634**(8035), 970–978 (2024)
25. Wolf, F.A., Angerer, P., Theis, F.J.: Scanpy: large-scale single-cell gene expression data analysis. *Genome biology* **19**(1), 15 (2018)

26. Xiang, J., Wang, X., Zhang, X., Xi, Y., Eweje, F., Chen, Y., Li, Y., Bergstrom, C., Gopaulchan, M., Kim, T., et al.: A vision–language foundation model for precision oncology. *Nature* **638**(8051), 769–778 (2025)
27. Xiong, C., Chen, H., Sung, J.J.: A survey of pathology foundation model: Progress and future directions. *arXiv preprint arXiv:2504.04045* (2025)
28. Xu, H., Fu, H., Long, Y., Ang, K.S., Sethi, R., Chong, K., Li, M., Uddamvathanak, R., Lee, H.K., Ling, J., et al.: Unsupervised spatially embedded deep representation of spatial transcriptomics. *Genome Medicine* **16**(1), 12 (2024)
29. Xu, H., Usuyama, N., Bagga, J., Zhang, S., Rao, R., Naumann, T., Wong, C., Gero, Z., González, J., Gu, Y., et al.: A whole-slide foundation model for digital pathology from real-world data. *Nature* **630**(8015), 181–188 (2024)
30. Yuan, Z., Zhao, F., Lin, S., Zhao, Y., Yao, J., Cui, Y., Zhang, X.Y., Zhao, Y.: Benchmarking spatial clustering methods with spatially resolved transcriptomics data. *Nature Methods* **21**(4), 712–722 (2024)
31. Zhang, A., Jaume, G., Vaidya, A., Ding, T., Mahmood, F.: Accelerating data processing and benchmarking of ai models for pathology. *arXiv preprint arXiv:2502.06750* (2025)
32. Zong, Y., Yu, T., Wang, X., Wang, Y., Hu, Z., Li, Y.: const: an interpretable multi-modal contrastive learning framework for spatial transcriptomics. *BioRxiv* pp. 2022–01 (2022)