ProteinPNet: Prototypical Part Networks for Concept Learning in Spatial Proteomics

Louis McConnell
Lausanne University Hospital
Lausanne, CH
louie.mc@berkeley.edu

Jieran Sun
Lausanne University Hospital
Lausanne, CH
jieran.sun@chuv.ch

Theo Maffei
Lausanne University Hospital
Lausanne, CH
theo.maffei@chuv.ch

Raphael Gottardo
Lausanne University Hospital
Lausanne, CH
raphael.gottardo@chuv.ch

Marianna Rapsomaniki
Lausanne University Hospital
Lausanne, CH
marianna.rapsomaniki@chuv.ch

Abstract

Understanding the spatial architecture of the tumor microenvironment (TME) is critical to advance precision oncology. We present **ProteinPNet**, a novel framework based on prototypical part networks that discovers TME motifs from spatial proteomics data. Unlike traditional post-hoc explanability models, ProteinPNet directly learns discriminative, interpretable, faithful spatial prototypes through supervised training. We validate our approach on synthetic datasets with ground truth motifs, and further test it on a real-world lung cancer spatial proteomics dataset. ProteinPNet consistently identifies biologically meaningful prototypes aligned with different tumor subtypes. Through graphical and morphological analyses, we show that these prototypes capture interpretable features pointing to differences in immune infiltration and tissue modularity. Our results highlight the potential of prototype-based learning to reveal interpretable spatial biomarkers within the TME, with implications for mechanistic discovery in spatial omics[1].

1 Introduction

Tumors are complex ecosystems where diverse cell populations interact to form heterogeneous tumor microenvironments (TMEs) [1]. The spatial heterogeneity of the TME has been the focus of intensive research, enabled by a range of tumor profiling technologies. From hematoxylin and eosin (H&E) staining, routinely used to assess morphological tissue alterations, to single-cell technologies that capture full molecular profiles, emerging data are starting to uncover prognostic patterns within the TME [2], associated with different spatial patterns of immune cells [3, 4] or cancer-associated fibroblasts (CAFs) [5, 6]. Spatial heterogeneity is also associated with differences in tumor architecture, as different morphological structures can lead to diverse cancer invasion patterns [7]. Recent developments in single cell and spatial omics technologies now enable the deep molecular profiling of each individual cell within the TME [8], offering an unprecedented opportunity to address an outstanding question in cancer biology: *how can we discover recurrent spatial patterns within the TME that drive disease outcomes?* Identifying and targeting these spatial biomarkers could spearhead the development of more precise and personalized therapies.

Despite growing data availability, extracting biologically meaningful spatial patterns from spatial omics data remains a challenge. In histopathology, where traditional deep learning models trained

[1] Code available at `https://github.com/AI4SCR/ProteinPNet`.

39th Conference on Neural Information Processing Systems (NeurIPS 2025) Workshop: Imageomics Workshop.

on H&E images have demonstrated high predictive power on a number of clinical tasks [9, 10], attention-based multiple-instance learning (MIL) pipelines [11] and post-hoc explainers (e.g., Grad-CAM [12], layer-wise relevance propagation [13], saliency maps [14]), allow visualization of TME regions that influence the models' decisions. In the nascent field of spatial omics, current work on identifying spatial biomarkers is heavily based on well-known tumor properties [15], for example, hand-crafted spatial features of immune cells to predict immunotherapy response [16–18]. Early attempts to devise methods that automatically discover spatial patterns are emerging [19–21]. While useful, both histopathology and spatial omics methods for mechanistic discovery are based on diffuse, hard-to-interpret post-hoc explanations that do not necessarily reflect the true mechanistic basis of the models' predictions and generally have problems with faithfulness [22, 23].

An attractive alternative to post-hoc explainers is to build inherently interpretable models [24]. Prototypical part networks [25] are one type of interpretable model that learn prototype representations, where each prototype corresponds to a representative part training example, and predictions are made by comparing input images to prototypes. Because the prediction is a function of the computed similarity between prototypes and the input image, they are able to circumvent many of the faithfulness problems of post-hoc explanations. Although prototypical part networks have been used in the medical domain for interpretable, case-based deep learning in clinical applications [26–28], their potential for mechanistic discovery in the realm of spatial omics is underexplored. These networks are particularly fitting to the problem of identifying spatial biomarkers from spatial omics, as they resemble how oncologists or pathologists reason and use resemblance to learned spatial motifs in the data as a bottleneck layer in prediction. In this paper, we propose **ProteinPNet**, a novel prototypical part-based model tailored to spatial proteomics data. We show that ProteinPNet learns prototypes that align with distinct tumor subtypes and reflect biologically relevant patterns, as assessed via graphical and morphological analysis. Although preliminary, our findings suggest that prototype-based learning holds considerable promise in identifying spatial biomarkers in the TME and guiding future discoveries in oncology.

2 Methods

We developed **ProteinPNet**, a framework for prototypical part learning tailored to spatial proteomics data that enables the discovery of interpretable, morphologically-aware spatial biomarkers. The ProteinPNet framework consists of two principal stages, namely *prototype discovery* and *prototype interpretation* (Figure 1). During *prototype discovery*, a prototypical part network inspired by Chen et al. [25][2] is trained to directly learn prototypical concepts, hereafter referred to as prototypes, from spatial proteomics data. To account for the high dimensionality and variable channel structure of spatial proteomics compared to standard RGB images, we treat each protein measurement as an image channel and design a custom encoder architecture. While several options for analyzing data with a large number of input channels are available, we encountered a high amount of overfitting when taking in all protein channels (over 97% train accuracy and roughly random train performance), even when using small networks (Resnet18 [29]) and low numbers of prototypes (1 per class). Instead, we reduced the data to three principal components (PCs) and used a pretrained ResNet152 backbone. Similar to Chen et al. [25], ProteinPNet performs prediction by regressing on a set of instance-specific scores that measure the degree to which a prototype is a part of the instance. For a given spatial proteomics sample x and a convolutional head $\mathbf{z} = f(\mathbf{x})$ with convolutional output representations of dimension $H \times W \times D$, ProteinPNet learns m prototypes $\mathbf{P} = \{\mathbf{p}_j\}_{j=1}^{m}$ of shape $H_p \times W_p \times D$, where $H_p \leq H, W_p \leq W$. From the prototypes and a sample z, we compute the prototype scores $g_{p_j}(z) := \min_{\tilde{z} \in \text{patches}(z)} d(\tilde{\mathbf{z}}, p_j)$ for some distance metric d. Following [30], we use cosine distance as our distance metric to get our prototype scores $g_{p_j}(z)$. The scores are linearly combined to map to the output distribution used for prediction. To maintain interpretability, at every k epochs, each prototype p_j is assigned to the filter representation of the closest patch z. This ensures that each prototype has a concrete visual reference directly aligned with the prototype vector used to compute its scores (see [25] for details).

During *prototype interpretation*, these prototypical concepts are analyzed in terms of their morphological, topological, and protein-expression-level characteristics to reveal differences in the underlying biology and link to different TME topologies, pathways, or niches. Once learned, the prototypes can be further interrogated using custom downstream analyses to assess if they are driven by differences in

[2]We follow the notation from this paper as well. For more details, please see [25].

tumor morphology, cellular composition, or both. Prototype activation maps allow prototype isolation by selecting the top N-percentile most activated regions in the top-k most activated samples. Those regions can then be subjected to domain-specific analyses (e.g., using graph-based or morphological scores, or differential expression analysis), revealing key biological processes. This stage of analysis can be customized to the biological hypothesis and the dataset at hand.

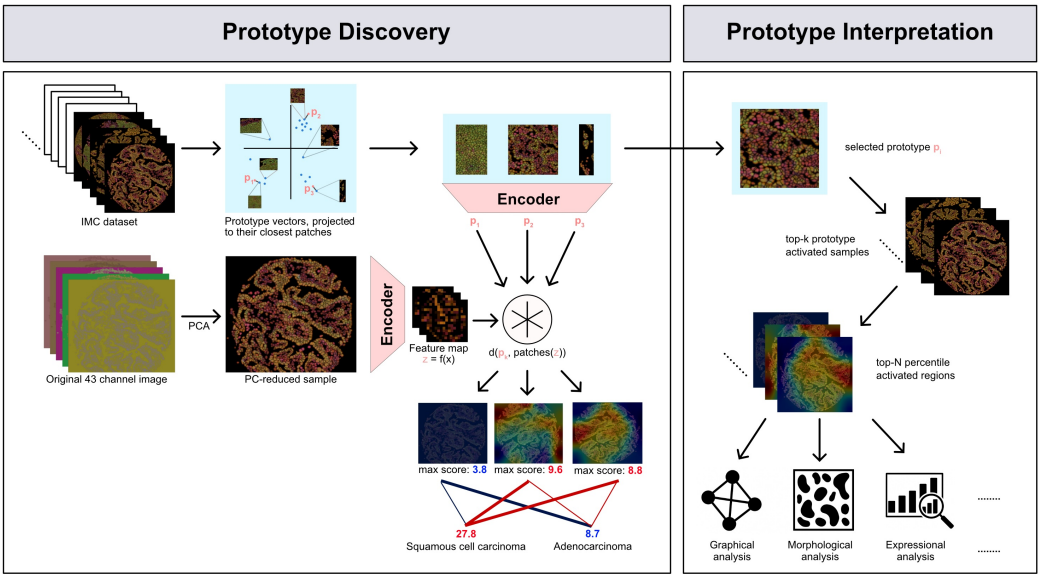

Figure 1: **The ProteinPNet workflow.** During *prototype discovery*, the prototype vectors are randomly initialized and projected onto the closest patch representation. The prototype representations are then convolved over the representation of the spatial proteomics image with a cosine similarity kernel to generate an activation heatmap, which generates a set of prototype scores that are linearly combined to make the final prediction. During *prototype interpretation*, prototypes that generated the highest accuracy are analyzed in terms of their morphological and compositional characteristics.

3 Results

Benchmarking on synthetic data We first evaluated the effectiveness of ProteinPNet in identifying prototypical motifs on synthetic data with ground-truth prototypes. We generated a synthetic dataset consisting of 3-channel images assigned to two different hypothetical classes, containing different types of randomly distributed three-node subgraphs (Figure S1), and injected a unique subgraph in each class (red circles). This proxy task is important, as it allows us to evaluate ProteinPNet on ground-truth prototypes which do not exist in real-world spatial omics data. It is thus essential that ProteinPNet can provably recover known class-discriminative synthetic prototypes while ignoring the neutral ones. We assessed ProteinPNet's performance both in terms of accuracy, as well as by checking whether the injected class-specific pattern was detected in the extracted prototypes. ProteinPNet consistently achieved a 100% classification accuracy (Table 1). In every run, prototype activation maps recovered the exact area containing the injected prototype for at least one class (see representative examples in Figure S2). Interestingly, in one seed, the model used white space to indicate the absence of a key prototype, while in all others, it identified class-specific prototypes in both classes (Figure S3). This demonstrates an interesting effect in which ProteinPNet is able to learn prototypes identifying the "absence" of key prototypes.

Evaluation on real-world spatial proteomics data We then applied ProteinPNet on a publicly available spatial proteomics dataset [31], from a large non-small cell lung cancer (NSCLC) cohort containing both adenocarcinoma (LUAD) and squamous cell carcinoma (LUSC), profiled by imaging mass cytometry. The dataset contains a total of 1021 samples, with each sample corresponding to the simultaneous quantification of 43 different protein markers, resulting in a 43-channel image. We first trained ProteinPNet to classify between different NSCLC subtypes, i.e., LUAD vs. LUSC. We note that, across all 1021 samples, 618 belong to the LUAD and 403 to LUSC subtype, giving a naive

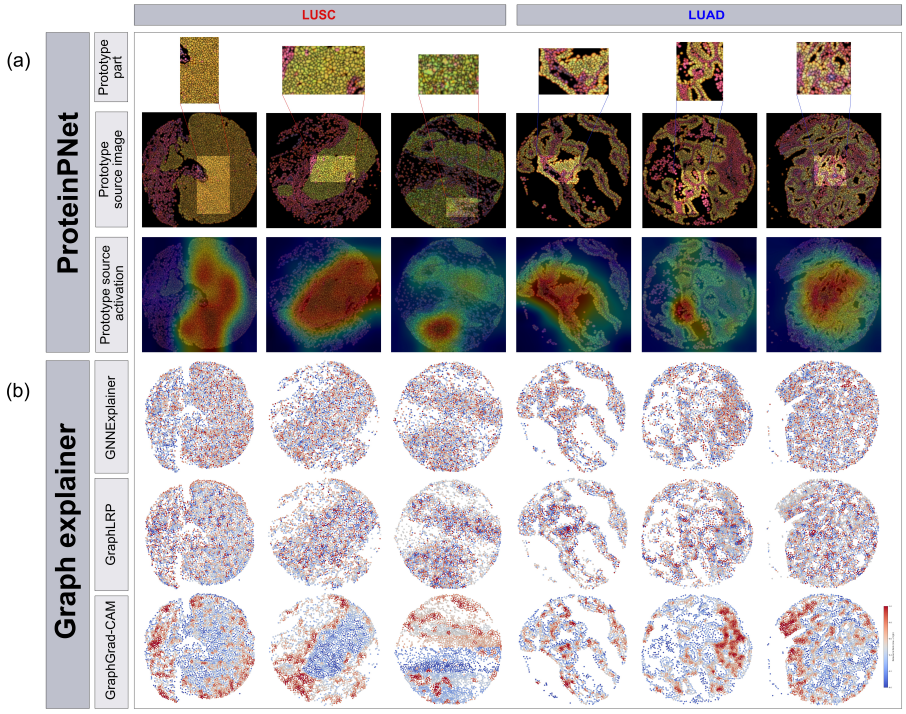

Figure 2: (a) Characteristic examples of LUSC and LUAD prototypes, collected across many runs with only one prototype per class, together with the source image and activation maps. (b) Performance of three graph explainers for the same example samples.

accuracy of 60.5% when predicting only the majority class. We used the predictive accuracy of the classifier as an initial validation of the predictive value of the learned prototypes: a low predictive loss in this task suggests that the prototype scores - and consequently the prototypes - contain important information needed to distinguish data between cancer classes. ProteinPNet reached an 80.7% accuracy on LUAD vs. LUSC prediction (Table 1). To probe what information the model relies on, we conducted two ablations. First, we evaluated whether learned prototypes outperform random crops; in other words, are the prototype scores indicative of unique prototype selection, or will a distance metric to any random patch have the same predictive power? To test this, we froze the model backbone after training beyond 80% accuracy and reinitialized the prototype vectors, pushing them onto random patches. This setup ensures that the backbone is capable of representing high quality prototypes while testing the quality of the learned vs. random prototypes. Accuracy dropped by 6.2% (Table 1), indicating that prototypes learned by the model encode meaningful information about the underlying tissue structure. Second, we removed the cell type information by setting all cell representations to a constant, leaving only morphological information. In this setting, we were only able to reach an accuracy of 71.2%, suggesting that both morphological features and protein expression are essential for prediction.

We then benchmarked ProteinPNet against graph-based explainers. We first applied different types of Graph Neural Networks, namely Graph Convolutional Networks (GCN) [32], Graph Isomorphism Networks (GIN) [33] and Graph Attention Networks (GAT)[34], for the same prediction task (LUAD vs. LUSC). All GNN-based models exceeded an accuracy of 70%, with GIN reaching 74.5% (Table 1), close to the ProteinPNet ablation with random prototypes. Next, we evaluated different post-hoc graph explainers, namely GNNExplainer [35], GNN-LRP [36] and GraphGrad-CAM [37] (Figure 2B using the trained GIN model. We observed that GNNExplainer and GNN-LRP highlighted cells and nodes that are randomly distributed in the tissue (Figure 2). Conversely, GraphGrad-CAM, resulted in spatially colocalized high/low importance regions, occasionally antithetical than those of ProteinPNet. We then estimated how well these graph explainers agree with each other and with a random explainer (Figure S6) and observed consistently low average scores, suggesting that pairwise

agreement between methods is on par with that of a random selection of cells. This result further highlights the limitations of post-hoc explainers: although GNN baselines can predict the disease subtype, the underlying high-importance cells are likely capturing spatially meaningful patterns.

Prototype interpretation We conducted an exploratory analysis as part of the workflow to identify biological concepts encapsulated in the prototypes. For each prototype, we selected the $m = 100$ most prototypically activated samples and isolated regions above the 80th percentile of activation as prototypical regions, with the remaining regions in the same samples serving as references, and investigated differences among prototypes and between prototypes and references. Qualitatively, the 100 LUAD and LUSC highest activation images differ in both tissue morphology and spatial heterogeneity (Figure S4 and Figure S5. LUSC prototypes appear as clumps of densely connected, small regions with minimal infiltration, whereas LUAD prototypes exhibit pronounced glandular morphology. These prototypes align with the typical LUAD and LUSC core patterns (see Fig. 2A of [31]). To quantify those differences, we estimated different heterogeneity metrics in prototypically activated regions using ATHENA [38]. To test differences in cell density and topology, we compared extent (the ratio of the tissue area to the area of its bounding box) and coreness (the maximal k-node subgraph with nodes of degree $\geq k$) between prototypes. Both extent ($p < 10^{-3}$) and coreness ($p < 1.03 \cdot 10^{-4}$) were higher in LUSC prototype regions, confirming our hypothesis on connectivity (Figure S7, Figure S8). To assess tumor infiltration, we used the infiltration score in ATHENA, defined as the ratio of tumor/non-tumor to tumor-tumor interaction. LUAD prototypes showed significantly higher infiltration of non-tumor cells (fibroblasts and immune cells) into tumor regions (Figure S9), consistent with existing literature [31]. Finally, when computing modularity [39] based on tumor-nontumor partition, both LUAD and LUSC prototype regions show significantly lower values than reference regions ($p < 10^{-3}$), suggesting that prototypes preferentially capture spatially heterogeneous tumor edges enriched with infiltration and tumor-non-tumor interaction (Figure S10).

Table 1: Results from all experiments. Averaged over 5 runs.

Architecture	Experiment	Accuracy (%)
ProteinPNet	Synthetic Dataset	100 ± 0
GCN	NSCLC	71.5 ± 0.7
GIN	NSCLC	74.6 ± 1.4
GAT	NSCLC	72.3 ± 1.3
ProteinPNet	NSCLC (no cell type information)	71.2 ± 0.4
ProteinPNet	NSCLC (randomized prototypes)	74.5 ± 1.5
ProteinPNet	NSCLC	80.7 ± 0.4

4 Conclusions

In this work, we present ProteinPNet, a prototypical part-network-based framework for learning and analyzing spatial motifs from spatial proteomics data. On synthetic data, ProteinPNet reliably recovered ground-truth class-specific prototypes, while on the NSCLC dataset, it achieved high predictive accuracy and identified prototypes consistent with known LUAD and LUSC patterns. Ablations confirmed that both the prototype learning process and protein expression contribute critically to performance. Downstream analyses further verified that prototypes capture biologically meaningful patterns of tissue connectivity and tumor infiltration. Together, these results demonstrate the potential of prototype-based learning to uncover recurrent organizational structures in TME and to provide interpretable insights into cancer biology. A main limitation of the current implementation of our model is the use of PCA to compress the 43-plex images to a pseudo-RGB representation. Although training the model on the PCA-reduced images led to high prediction accuracy and was able to capture high-level spatial motifs, it potentially caused some loss of information from the original data. We are currently investigating architectural modifications and more powerful encoding strategies that can better preserve the richness and complexity of multiplexed images. This is particularly important in view of future extension of ProteinPNet to spatial transcriptomics data that contain thousands of channels. In addition, we are exploring best practices for analyzing and interpreting spatial motifs to yield more refined biological insights, as the optimal way to interpret and use these prototypical spatial motifs is currently guided by visual inspections.

Disclosure of Funding This project has been made possible in part by grant numbers 202297, 215550 and 235972 from the Swiss National Science Foundation and grant number 2024-345909 from the Chan Zuckerberg Initiative DAF, an advised fund of Silicon Valley Community Foundation.

References

[1] Aditya Kashyap, Maria Anna Rapsomaniki, Vesna Barros, Anna Fomitcheva-Khartchenko, Adriano Luca Martinelli, Antonio Foncubierta Rodriguez, Maria Gabrani, Michal Rosen-Zvi, and Govind Kaigala. Quantification of tumor heterogeneity: from data acquisition to metric generation. *Trends in Biotechnology*, 40(6):647–676, 2022. ISSN 0167-7799.

[2] Frances R. Balkwill, Melania Capasso, and Thorsten Hagemann. The tumor microenvironment at a glance. *Journal of Cell Science*, 125(23):5591–5596, December 2012. ISSN 0021-9533. doi: 10.1242/jcs.116392. URL https://doi.org/10.1242/jcs.116392.

[3] Khalid AbdulJabbar, Shan E. Ahmed Raza, Rachel Rosenthal, Mariam Jamal-Hanjani, Selvaraju Veeriah, Ayse Akarca, Tom Lund, David A. Moore, Roberto Salgado, Maise Al Bakir, Luis Zapata, Crispin T. Hiley, Leah Officer, Marco Sereno, Claire Rachel Smith, Sherene Loi, Allan Hackshaw, Teresa Marafioti, Sergio A. Quezada, Nicholas McGranahan, John Le Quesne, Charles Swanton, and Yinyin Yuan. Geospatial immune variability illuminates differential evolution of lung adenocarcinoma. *Nature Medicine*, 26(7):1054–1062, July 2020. ISSN 1546-170X. doi: 10.1038/s41591-020-0900-x. URL https://www.nature.com/articles/s41591-020-0900-x. Number: 7 Publisher: Nature Publishing Group.

[4] Chang Gong, Robert A. Anders, Qingfeng Zhu, Janis M. Taube, Benjamin Green, Wenting Cheng, Imke H. Bartelink, Paolo Vicini, Bing Wang, and Aleksander S. Popel. Quantitative Characterization of CD8+ T Cell Clustering and Spatial Heterogeneity in Solid Tumors. *Frontiers in Oncology*, 8, 2019. ISSN 2234-943X. doi: 10.3389/fonc.2018.00649. URL https://www.frontiersin.org/articles/10.3389/fonc.2018.00649/full. Publisher: Frontiers.

[5] Rana Mhaidly and Fatima Mechta-Grigoriou. Fibroblast heterogeneity in tumor microenvironment: Role in immunosuppression and new therapies. *Seminars in Immunology*, 48:101417, April 2020. ISSN 1044-5323. doi: 10.1016/j.smim.2020.101417. URL https://www.sciencedirect.com/science/article/pii/S1044532320300336.

[6] Berna C. Özdemir, Tsvetelina Pentcheva-Hoang, Julienne L. Carstens, Xiaofeng Zheng, Chia-Chin Wu, Tyler R. Simpson, Hanane Laklai, Hikaru Sugimoto, Christoph Kahlert, Sergey V. Novitskiy, Ana De Jesus-Acosta, Padmanee Sharma, Pedram Heidari, Umar Mahmood, Lynda Chin, Harold L. Moses, Valerie M. Weaver, Anirban Maitra, James P. Allison, Valerie S. LeBleu, and Raghu Kalluri. Depletion of carcinoma-associated fibroblasts and fibrosis induces immunosuppression and accelerates pancreas cancer with reduced survival. *Cancer Cell*, 25(6): 719–734, June 2014. ISSN 1878-3686. doi: 10.1016/j.ccr.2014.04.005.

[7] Robert Noble, Dominik Burri, Cécile Le Sueur, Jeanne Lemant, Yannick Viossat, Jakob Nikolas Kather, and Niko Beerenwinkel. Spatial structure governs the mode of tumour evolution. *Nature Ecology & Evolution*, pages 1–11, December 2021. ISSN 2397-334X. doi: 10.1038/s41559-021-01615-9. URL https://www.nature.com/articles/s41559-021-01615-9. Bandiera_abtest: a Cc_license_type: cc_by Cg_type: Nature Research Journals Primary_atype: Research Publisher: Nature Publishing Group Subject_term: Cancer genetics;Population genetics Subject_term_id: cancer-genetics;population-genetics.

[8] Sabrina M Lewis, Marie-Liesse Asselin-Labat, Quan Nguyen, Jean Berthelet, Xiao Tan, Verena C Wimmer, Delphine Merino, Kelly L Rogers, and Shalin H Naik. Spatial omics and multiplexed imaging to explore cancer biology. *Nature methods*, 18(9):997–1012, 2021. Publisher: Nature Publishing Group US New York.

[9] Amelie Echle, Niklas Timon Rindtorff, Titus Josef Brinker, Tom Luedde, Alexander Thomas Pearson, and Jakob Nikolas Kather. Deep learning in cancer pathology: a new generation of clinical biomarkers. *British Journal of Cancer*, 124(4):686–696, February 2021. ISSN 1532-1827. doi: 10.1038/s41416-020-01122-x. URL https://www.nature.com/articles/s41416-020-01122-x. Publisher: Nature Publishing Group.

[10] Cagla Deniz Bahadir, Mohamed Omar, Jacob Rosenthal, Luigi Marchionni, Benjamin Liechty, David J. Pisapia, and Mert R. Sabuncu. Artificial intelligence applications in histopathology. *Nature Reviews Electrical Engineering*, 1(2):93–108, February 2024. ISSN 2948-1201. doi: 10.1038/s44287-023-00012-7. URL https://www.nature.com/articles/s44287-023-00012-7. Publisher: Nature Publishing Group.

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

[12] Ramprasaath R Selvaraju, Michael Cogswell, Abhishek Das, Ramakrishna Vedantam, Devi Parikh, and Dhruv Batra. Grad-cam: Visual explanations from deep networks via gradient-based localization. In *Proceedings of the IEEE international conference on computer vision*, pages 618–626, 2017.

[13] Grégoire Montavon, Alexander Binder, Sebastian Lapuschkin, Wojciech Samek, and Klaus-Robert Müller. Layer-wise relevance propagation: an overview. *Explainable AI: interpreting, explaining and visualizing deep learning*, pages 193–209, 2019.

[14] Meng-Hao Guo, Tian-Xing Xu, Jiang-Jiang Liu, Zheng-Ning Liu, Peng-Tao Jiang, Tai-Jiang Mu, Song-Hai Zhang, Ralph R Martin, Ming-Ming Cheng, and Shi-Min Hu. Attention mechanisms in computer vision: A survey. *Computational visual media*, 8(3):331–368, 2022.

[15] Jiahui Jiang, Yunhe Liu, Jiangjiang Qin, Jianfeng Chen, Jingjing Wu, Melissa P. Pizzi, Rossana Lazcano, Kohei Yamashita, Zhiyuan Xu, Guangsheng Pei, Kyung Serk Cho, Yanshuo Chu, Ansam Sinjab, Fuduan Peng, Xinmiao Yan, Guangchun Han, Ruiping Wang, Enyu Dai, Yibo Dai, Bogdan A. Czerniak, Andrew Futreal, Anirban Maitra, Alexander Lazar, Humam Kadara, Amir A. Jazaeri, Xiangdong Cheng, Jaffer Ajani, Jianjun Gao, Jian Hu, and Linghua Wang. METI: deep profiling of tumor ecosystems by integrating cell morphology and spatial transcriptomics. *Nature Communications*, 15(1):7312, August 2024. ISSN 2041-1723. doi: 10.1038/s41467-024-51708-9. URL https://www.nature.com/articles/s41467-024-51708-9. Publisher: Nature Publishing Group.

[16] Hannah L. Williams, Ana Leni Frei, Thibaud Koessler, Martin D. Berger, Heather Dawson, Olivier Michielin, and Inti Zlobec. The current landscape of spatial biomarkers for prediction of response to immune checkpoint inhibition. *npj Precision Oncology*, 8(1):1–18, August 2024. ISSN 2397-768X. doi: 10.1038/s41698-024-00671-1. URL https://www.nature.com/articles/s41698-024-00671-1. Publisher: Nature Publishing Group.

[17] Tobias Hoch, Daniel Schulz, Nils Eling, Julia Martínez Gómez, Mitchell P. Levesque, and Bernd Bodenmiller. Multiplexed imaging mass cytometry of the chemokine milieus in melanoma characterizes features of the response to immunotherapy. *Science Immunology*, 7(70):eabk1692, April 2022. doi: 10.1126/sciimmunol.abk1692. URL https://www.science.org/doi/abs/10.1126/sciimmunol.abk1692. Publisher: American Association for the Advancement of Science.

[18] Xiao Qian Wang, Esther Danenberg, Chiun-Sheng Huang, Daniel Egle, Maurizio Callari, Begoña Bermejo, Matteo Dugo, Claudio Zamagni, Marc Thill, Anton Anton, Stefania Zambelli, Stefania Russo, Eva Maria Ciruelos, Richard Greil, Balázs Győrffy, Vladimir Semiglazov, Marco Colleoni, Catherine M. Kelly, Gabriella Mariani, Lucia Del Mastro, Olivia Biasi, Robert S. Seitz, Pinuccia Valagussa, Giuseppe Viale, Luca Gianni, Giampaolo Bianchini, and H. Raza Ali. Spatial predictors of immunotherapy response in triple-negative breast cancer. *Nature*, 621 (7980):868–876, September 2023. ISSN 1476-4687. doi: 10.1038/s41586-023-06498-3. URL https://www.nature.com/articles/s41586-023-06498-3. Publisher: Nature Publishing Group.

[19] Zhenqin Wu, Alexandro E. Trevino, Eric Wu, Kyle Swanson, Honesty J. Kim, H. Blaize D'Angio, Ryan Preska, Gregory W. Charville, Piero D. Dalerba, Ann Marie Egloff, Ravindra Uppaluri, Umamaheswar Duvvuri, Aaron T. Mayer, and James Zou. Graph deep learning for the characterization of tumour microenvironments from spatial protein profiles in tissue specimens. *Nature Biomedical Engineering*, pages 1–14, November 2022. ISSN 2157-846X. doi: 10.1038/s41551-022-00951-w. URL https://www.nature.com/articles/s41551-022-00951-w. Publisher: Nature Publishing Group.

[20] Jovan Tanevski, Ricardo Omar Ramirez Flores, Attila Gabor, Denis Schapiro, and Julio Saez-Rodriguez. Explainable multiview framework for dissecting spatial relationships from highly multiplexed data. *Genome Biology*, 23(1):97, April 2022. ISSN 1474-760X. doi: 10.1186/s13059-022-02663-5. URL https://doi.org/10.1186/s13059-022-02663-5.

[21] Chunman Zuo, Junjie Xia, and Luonan Chen. Dissecting tumor microenvironment from spatially resolved transcriptomics data by heterogeneous graph learning. *Nature Communications*, 15(1): 5057, June 2024. ISSN 2041-1723. doi: 10.1038/s41467-024-49171-7. URL https://www.nature.com/articles/s41467-024-49171-7. Publisher: Nature Publishing Group.

[22] Cynthia Rudin, Chaofan Chen, Zhi Chen, Haiyang Huang, Lesia Semenova, and Chudi Zhong. Interpretable machine learning: Fundamental principles and 10 grand challenges. *Statistic Surveys*, 16:1–85, 2022.

[23] Thibault Laugel, Marie-Jeanne Lesot, Christophe Marsala, Xavier Renard, and Marcin Detyniecki. The dangers of post-hoc interpretability: Unjustified counterfactual explanations. *arXiv preprint arXiv:1907.09294*, 2019.

[24] Cynthia Rudin. Stop explaining black box machine learning models for high stakes decisions and use interpretable models instead, 2019. URL https://arxiv.org/abs/1811.10154.

[25] Chaofan Chen, Oscar Li, Chaofan Tao, Alina Jade Barnett, Jonathan Su, and Cynthia Rudin. This Looks Like That: Deep Learning for Interpretable Image Recognition, December 2019. URL http://arxiv.org/abs/1806.10574. arXiv:1806.10574 [cs].

[26] A case-based interpretable deep learning model for classification of mass lesions in digital mammography | Nature Machine Intelligence. URL https://www.nature.com/articles/s42256-021-00423-x.

[27] Yuanyuan Wei, Roger Tam, and Xiaoying Tang. MProtoNet: A Case-Based Interpretable Model for Brain Tumor Classification with 3D Multi-parametric Magnetic Resonance Imaging, April 2023. URL http://arxiv.org/abs/2304.06258. arXiv:2304.06258 [cs].

[28] Eunji Kim, Siwon Kim, Minji Seo, and Sungroh Yoon. XProtoNet: Diagnosis in Chest Radiography With Global and Local Explanations. pages 15719–15728, 2021. URL https://openaccess.thecvf.com/content/CVPR2021/html/Kim_XProtoNet_Diagnosis_in_Chest_Radiography_With_Global_and_Local_Explanations_CVPR_2021_paper.html.

[29] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Deep Residual Learning for Image Recognition, 2015. URL https://arxiv.org/abs/1512.03385.

[30] Frank Willard, Luke Moffett, Emmanuel Mokel, Jon Donnelly, Stark Guo, Julia Yang, Giyoung Kim, Alina Jade Barnett, and Cynthia Rudin. This looks better than that: Better interpretable models with protopnext, 2024. URL https://arxiv.org/abs/2406.14675.

[31] Lena Cords, Stefanie Engler, Martina Haberecker, Jan Hendrik Rüschoff, Holger Moch, Natalie de Souza, and Bernd Bodenmiller. Cancer-associated fibroblast phenotypes are associated with patient outcome in non-small cell lung cancer. *Cancer Cell*, 42(3):396–412.e5, March 2024. ISSN 1535-6108, 1878-3686. doi: 10.1016/j.ccell.2023.12.021. URL https://www.cell.com/cancer-cell/abstract/S1535-6108(23)00449-X. Publisher: Elsevier.

[32] Thomas N. Kipf and Max Welling. Semi-supervised classification with graph convolutional networks. In *Proceedings of the International Conference on Learning Representations (ICLR)*, 2017. URL https://arxiv.org/abs/1609.02907.

[33] Keyulu Xu, Weihua Hu, Jure Leskovec, and Stefanie Jegelka. How powerful are graph neural networks? In *International Conference on Learning Representations (ICLR)*, 2019. URL https://openreview.net/forum?id=ryGs6iA5Km.

[34] Petar Veličković, Guillem Cucurull, Arantxa Casanova, Adriana Romero, Pietro Liò, and Yoshua Bengio. Graph attention networks. *arXiv preprint arXiv:1710.10903*, 2018. URL https://doi.org/10.48550/arXiv.1710.10903. Preprint.

[35] Zhitao Ying, Dylan Bourgeois, Jiaxuan You, Marinka Zitnik, and Jure Leskovec. Gnnexplainer: Generating explanations for graph neural networks. *Advances in neural information processing systems*, 32, 2019.

[36] Robert Schwarzenberg, Marc Hübner, David Harbecke, Christoph Alt, and Leonhard Hennig. Layerwise relevance visualization in convolutional text graph classifiers. In *Proceedings of the 13th Workshop on Graph-Based Methods for Natural Language Processing (TextGraphs-13)*, pages 58–62, Hong Kong, 2019. Association for Computational Linguistics.

[37] Phillip E. Pope, Soheil Kolouri, Mohammad Rostami, Charles E. Martin, and Heiko Hoffmann. Explainability methods for graph convolutional neural networks. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 10764–10773. IEEE, 2019. doi: 10.1109/CVPR.2019.01099.

[38] Adriano Luca Martinelli and Maria Anna Rapsomaniki. Athena: analysis of tumor heterogeneity from spatial omics measurements. *Bioinformatics*, 38(11):3151–3153, 04 2022. ISSN 1367-4803. doi: 10.1093/bioinformatics/btac303. URL `https://doi.org/10.1093/bioinformatics/btac303`.

[39] Aric A. Hagberg, Daniel A. Schult, Pieter J. Swart, and NetworkX Developers. NetworkX: Modularity function (networkx.algorithms.community.quality.modularity). `https://networkx.org/documentation/stable/reference/algorithms/generated/networkx.algorithms.community.quality.modularity.html`, 2008–2025. In: Exploring Network Structure, Dynamics, and Function Using NetworkX, Proceedings of the 7th Python in Science Conference (SciPy 2008), pp. 11–15. Accessed: 2025-08-28.

A Supplementary Figures

A.1 Synthetic Data Supplementary Figures

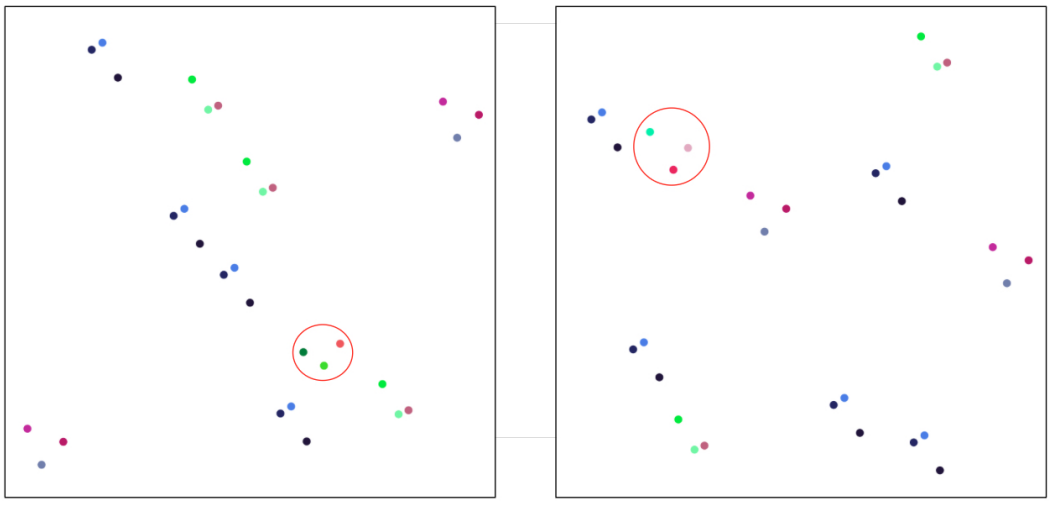

Figure S1: The two classes present in the synthetic dataset, with the red circle outlining the class-defining prototypes. Class independent prototypes can be seen in both samples.

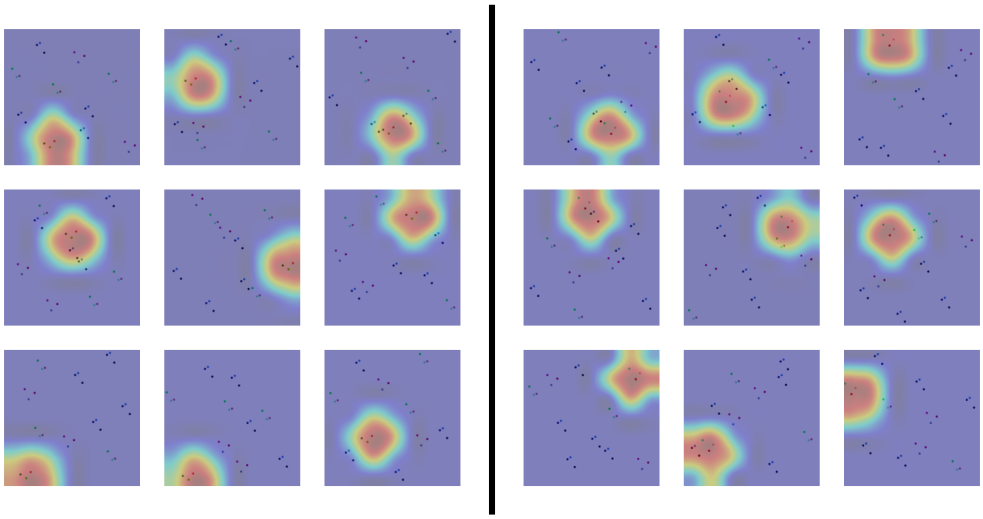

Figure S2: Synthetic data activations.

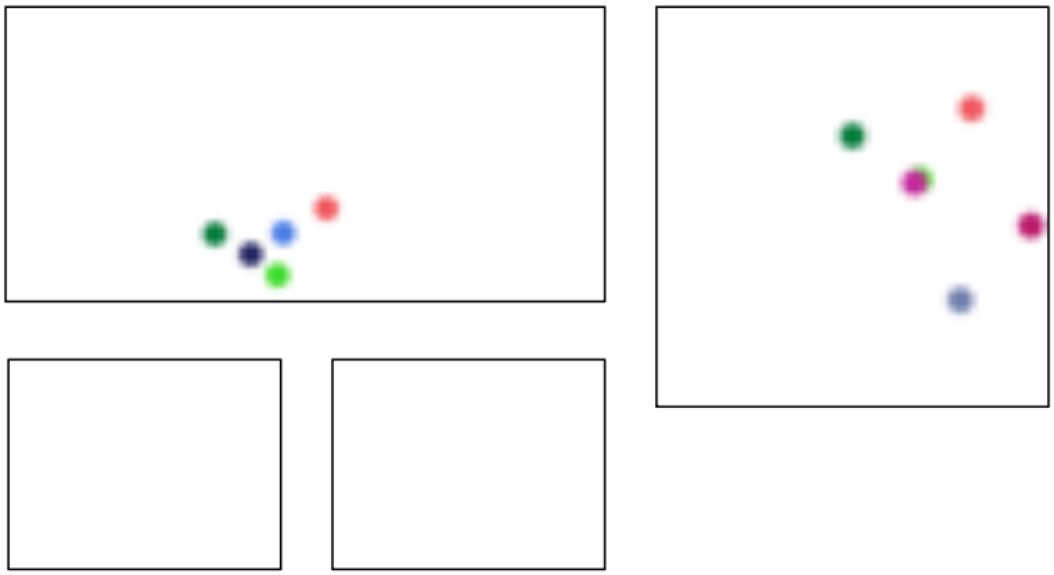

Figure S3: Example of prototypes extracted from the synthetic dataset as above. These are class specific prototypes, indicating that each of the top two prototypes belong to the first class and the bottom two belong to the second class. One can see that the second prototype contains an occlusion of a neutral prototype over the classifying prototype. In order to demonstrate the lack of the classifying prototype in the other class, the model focuses on the white space present in the model.

A.2 Top-100 Prototypically Activated Images Supplementary Figures

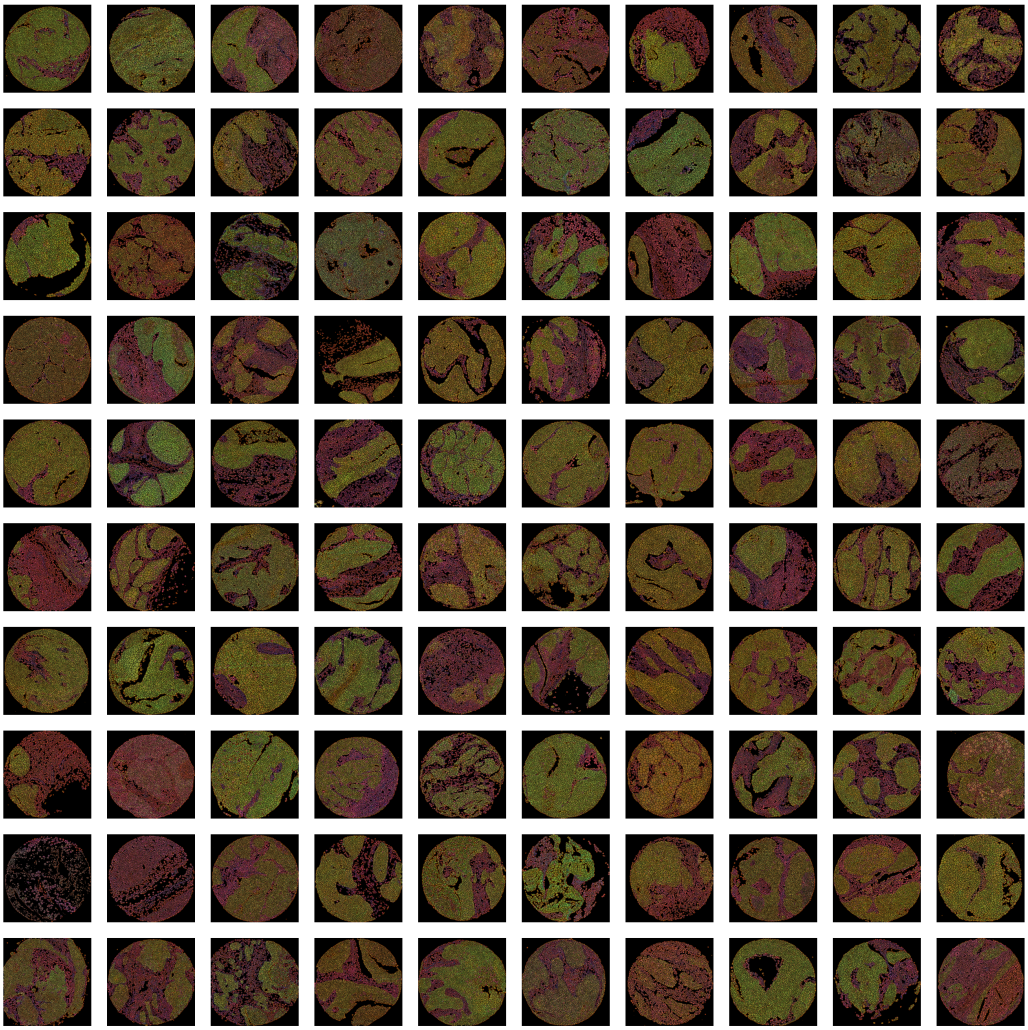

Figure S4: 100 images with highest activation to LUSC prototype in model with highest accuracy.

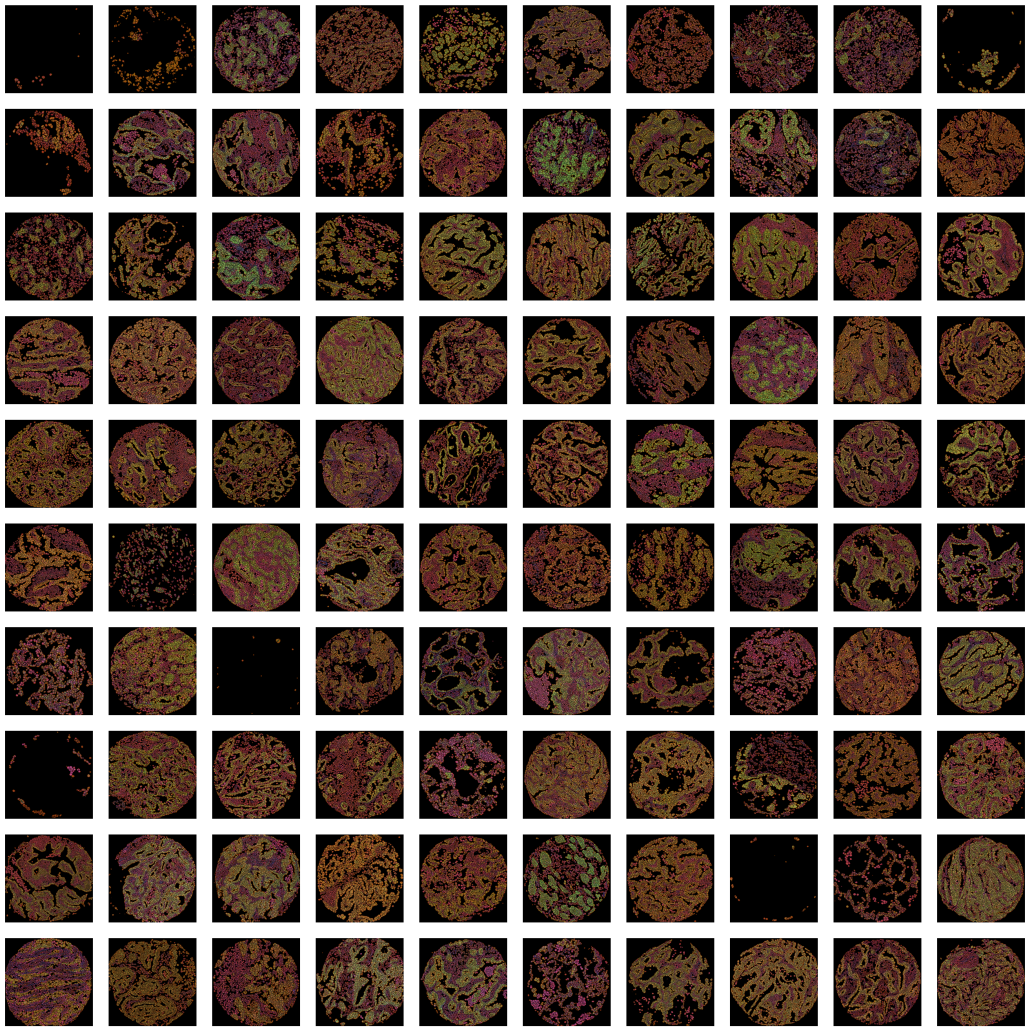

Figure S5: 100 images with highest activation to LUAD prototype in model with highest accuracy.

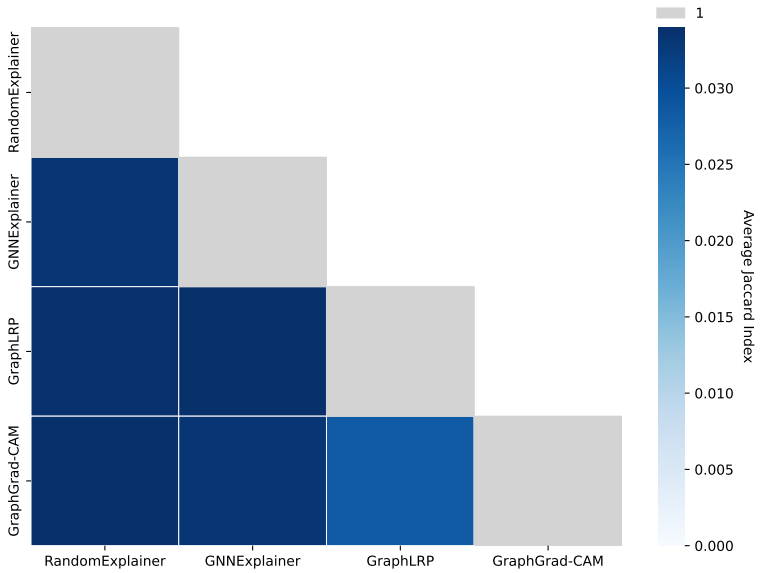

Figure S6: Jaccard index between RandomExplainer, GNNExplainer, GraphLRP, GraphGrad-CAM explainers computed for the top 100 most relevant cells. Jaccard index is defined as follows: $J(A_i, A_j) = \frac{|A_i \cap A_j|}{|A_i \cup A_j|}$, where A_i and A_j are set of the top 100 most relevant cells for explainer i and j.

A.3 Graphical and Morphological Metrics: Full Results

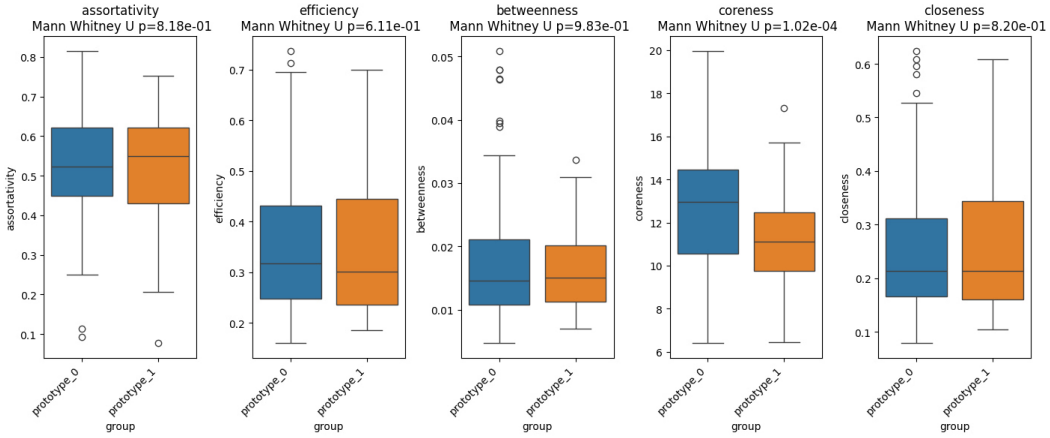

Figure S7: Graphical metrics from ATHENA. We can see a statistically significant difference in coreness higher in prototype 0, corresponding to LUSC.

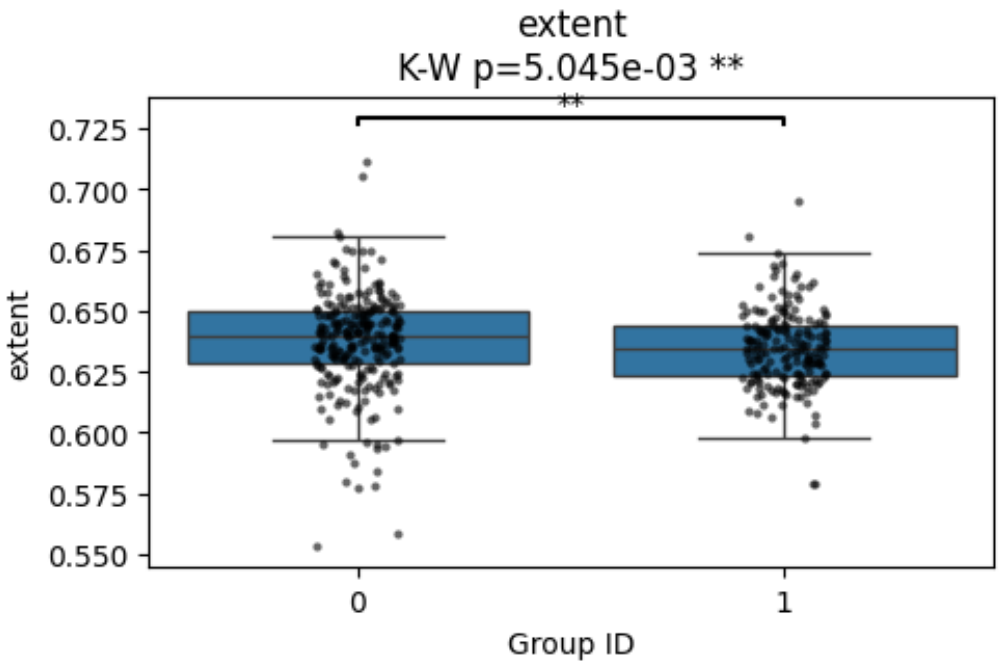

Figure S8: Extent by prototype. We can see a statistically significant difference in extent higher in prototype 0, corresponding to LUSC, corresponding to a lower amount of empty space morphologically.

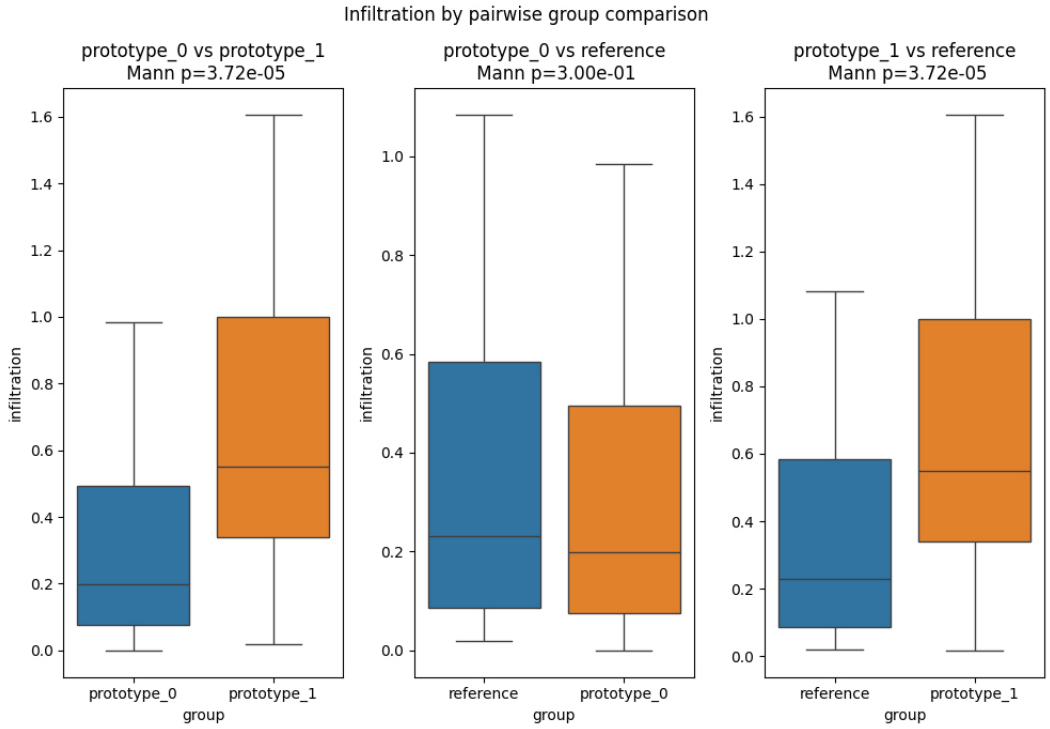

Figure S9: Infiltration by prototype. Prototype 0 corresponds to LUSC and prototype 1 corresponds to LUAD. We can see a clear increase in tumor infiltration in prototypes corresponding to LUAD as observed qualitatively.

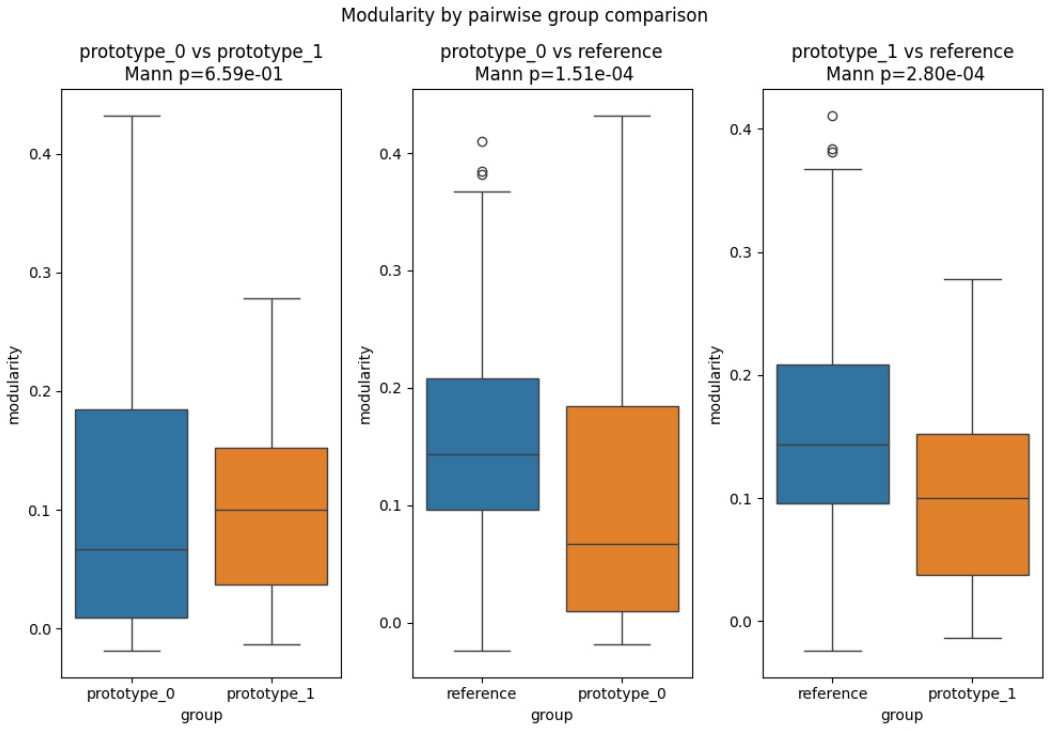

Figure S10: Modularity by prototype. Prototype 0 corresponds to LUSC and prototype 1 corresponds to LUAD. We can see that both prototypes express a lower modularity than the reference population, suggesting that heterogeneous regions in the tumor microenvironment are particularly relevant for NSCLC classification.

B Training Details

All experiments were conducted using a ResNet152 backbone with a 60/20/20 train/test/val split on PCA reduced data. Hyperparameters were borrowed directly from the ProtoPNet paper. In all experiments 1 prototype was used per class (2 total). All models are trained with ADAM and a StepLR scheduler. More details about training setups can be seen in the code.

C Dataset Preprocessing Details

We have followed the preprocessing of [31] for the NSCLC data. We first filter for only samples corresponding to LUAD or LUSC. Of the 43 proteins measured, we first remove both Iridium channels, leaving 41 remaining proteins.

For normalization, the samples are arcsinh transformed before being clipped at the 99th percentile and 0-1 transformed using the global minimum and maximum per channel. Following this, the PCA per pixel is taken over all samples in the train set. This PCA reduction is then applied to the entire dataset, and the dataset is then 0-1 normalized again.

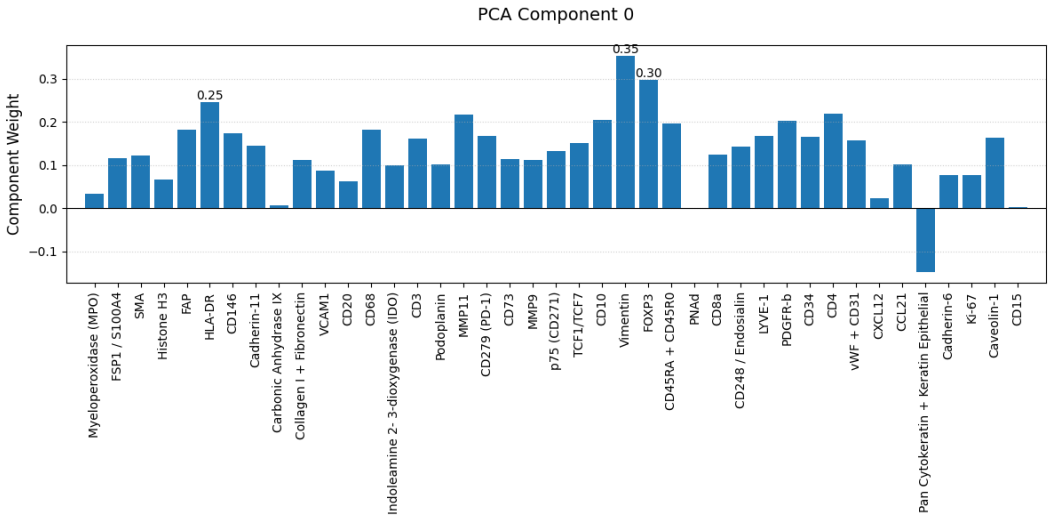

Figure S11: PCA Component 0 (Stromal Cells)

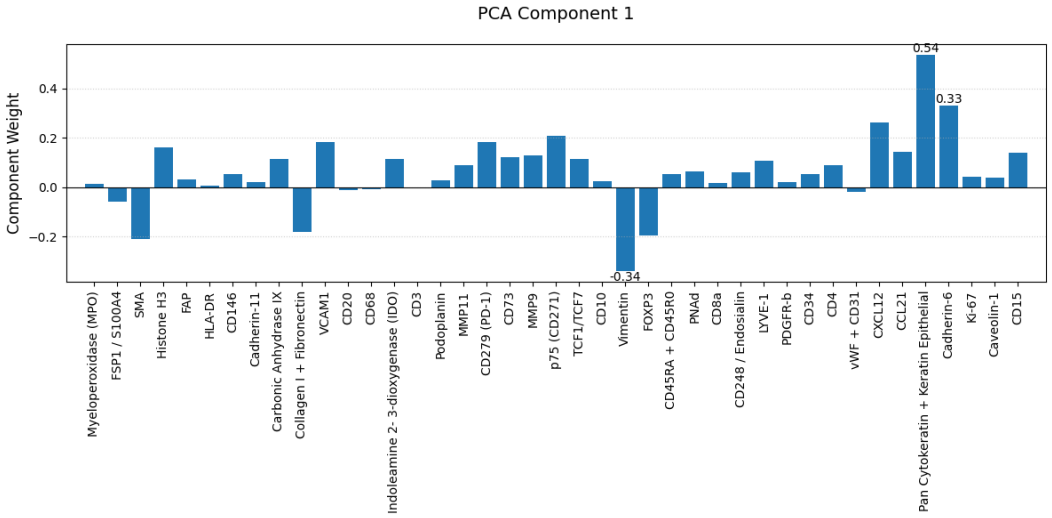

Figure S12: PCA Component 1 (Epithelial Cells)

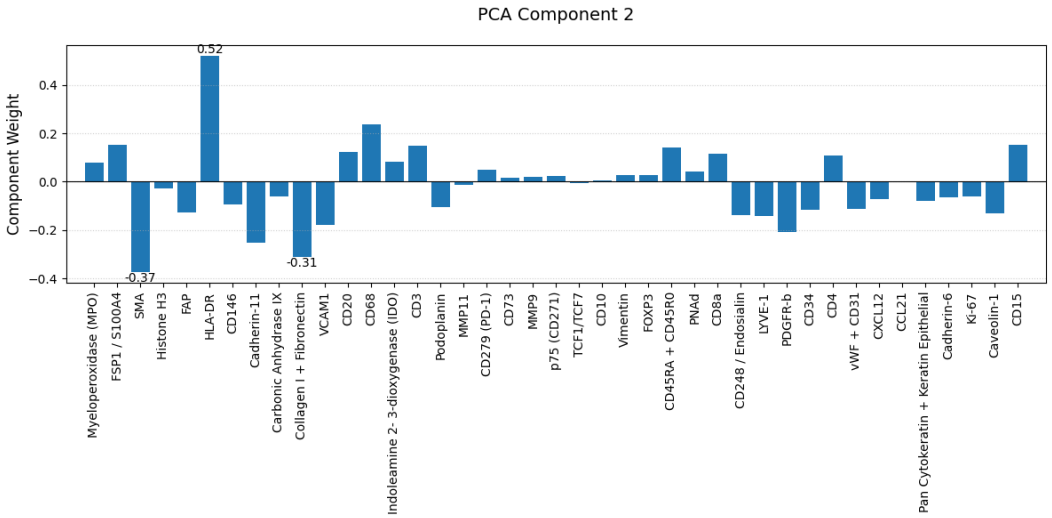

Figure S13: PCA Component 2 (Immune Cells)

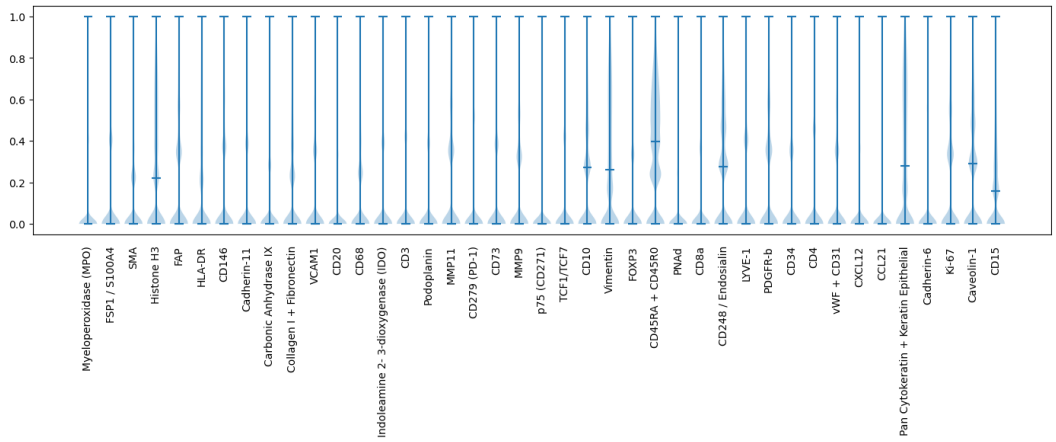

Figure S14: Violin plot of protein expression distributions over all cells in the NSCLC data.

NeurIPS Paper Checklist

1. **Claims**

   Question: Do the main claims made in the abstract and introduction accurately reflect the paper's contributions and scope?

   Answer: [Yes]

   Justification: We discuss the limitations of post-hoc interpretability in spatial omics and identify a literature gap in interpretable methods for concept discovery in spatial omics, all of which are addressed in the main paper.

   Guidelines:

   - The answer NA means that the abstract and introduction do not include the claims made in the paper.
   - The abstract and/or introduction should clearly state the claims made, including the contributions made in the paper and important assumptions and limitations. A No or NA answer to this question will not be perceived well by the reviewers.
   - The claims made should match theoretical and experimental results, and reflect how much the results can be expected to generalize to other settings.
   - It is fine to include aspirational goals as motivation as long as it is clear that these goals are not attained by the paper.

2. **Limitations**

   Question: Does the paper discuss the limitations of the work performed by the authors?

   Answer: [Yes]

   Justification: We contain a limitations section in the conclusion addressing future work and limitations of this existing, preliminary approach to prototype discovery in spatial omics.

   Guidelines:

   - The answer NA means that the paper has no limitation while the answer No means that the paper has limitations, but those are not discussed in the paper.
   - The authors are encouraged to create a separate "Limitations" section in their paper.
   - The paper should point out any strong assumptions and how robust the results are to violations of these assumptions (e.g., independence assumptions, noiseless settings, model well-specification, asymptotic approximations only holding locally). The authors should reflect on how these assumptions might be violated in practice and what the implications would be.

- The authors should reflect on the scope of the claims made, e.g., if the approach was only tested on a few datasets or with a few runs. In general, empirical results often depend on implicit assumptions, which should be articulated.
- The authors should reflect on the factors that influence the performance of the approach. For example, a facial recognition algorithm may perform poorly when image resolution is low or images are taken in low lighting. Or a speech-to-text system might not be used reliably to provide closed captions for online lectures because it fails to handle technical jargon.
- The authors should discuss the computational efficiency of the proposed algorithms and how they scale with dataset size.
- If applicable, the authors should discuss possible limitations of their approach to address problems of privacy and fairness.
- While the authors might fear that complete honesty about limitations might be used by reviewers as grounds for rejection, a worse outcome might be that reviewers discover limitations that aren't acknowledged in the paper. The authors should use their best judgment and recognize that individual actions in favor of transparency play an important role in developing norms that preserve the integrity of the community. Reviewers will be specifically instructed to not penalize honesty concerning limitations.

3. **Theory assumptions and proofs**

   Question: For each theoretical result, does the paper provide the full set of assumptions and a complete (and correct) proof?

   Answer: [NA]

   Justification: there are no theoretical results.

   Guidelines:

   - The answer NA means that the paper does not include theoretical results.
   - All the theorems, formulas, and proofs in the paper should be numbered and cross-referenced.
   - All assumptions should be clearly stated or referenced in the statement of any theorems.
   - The proofs can either appear in the main paper or the supplemental material, but if they appear in the supplemental material, the authors are encouraged to provide a short proof sketch to provide intuition.
   - Inversely, any informal proof provided in the core of the paper should be complemented by formal proofs provided in appendix or supplemental material.
   - Theorems and Lemmas that the proof relies upon should be properly referenced.

4. **Experimental result reproducibility**

   Question: Does the paper fully disclose all the information needed to reproduce the main experimental results of the paper to the extent that it affects the main claims and/or conclusions of the paper (regardless of whether the code and data are provided or not)?

   Answer [Yes]

   Justification: All codes for reproducing the results are published with the paper, and all data can be downloaded online in the link and reference provided in the paper.

   Guidelines:

   - The answer NA means that the paper does not include experiments.
   - If the paper includes experiments, a No answer to this question will not be perceived well by the reviewers: Making the paper reproducible is important, regardless of whether the code and data are provided or not.
   - If the contribution is a dataset and/or model, the authors should describe the steps taken to make their results reproducible or verifiable.
   - Depending on the contribution, reproducibility can be accomplished in various ways. For example, if the contribution is a novel architecture, describing the architecture fully might suffice, or if the contribution is a specific model and empirical evaluation, it may be necessary to either make it possible for others to replicate the model with the same dataset, or provide access to the model. In general. releasing code and data is often

one good way to accomplish this, but reproducibility can also be provided via detailed instructions for how to replicate the results, access to a hosted model (e.g., in the case of a large language model), releasing of a model checkpoint, or other means that are appropriate to the research performed.

- While NeurIPS does not require releasing code, the conference does require all submissions to provide some reasonable avenue for reproducibility, which may depend on the nature of the contribution. For example
  (a) If the contribution is primarily a new algorithm, the paper should make it clear how to reproduce that algorithm.
  (b) If the contribution is primarily a new model architecture, the paper should describe the architecture clearly and fully.
  (c) If the contribution is a new model (e.g., a large language model), then there should either be a way to access this model for reproducing the results or a way to reproduce the model (e.g., with an open-source dataset or instructions for how to construct the dataset).
  (d) We recognize that reproducibility may be tricky in some cases, in which case authors are welcome to describe the particular way they provide for reproducibility. In the case of closed-source models, it may be that access to the model is limited in some way (e.g., to registered users), but it should be possible for other researchers to have some path to reproducing or verifying the results.

5. **Open access to data and code**

   Question: Does the paper provide open access to the data and code, with sufficient instructions to faithfully reproduce the main experimental results, as described in supplemental material?

   Answer: [Yes]

   Justification: We open source all code and data

   Guidelines:

   - The answer NA means that paper does not include experiments requiring code.
   - Please see the NeurIPS code and data submission guidelines (`https://nips.cc/public/guides/CodeSubmissionPolicy`) for more details.
   - While we encourage the release of code and data, we understand that this might not be possible, so "No" is an acceptable answer. Papers cannot be rejected simply for not including code, unless this is central to the contribution (e.g., for a new open-source benchmark).
   - The instructions should contain the exact command and environment needed to run to reproduce the results. See the NeurIPS code and data submission guidelines (`https://nips.cc/public/guides/CodeSubmissionPolicy`) for more details.
   - The authors should provide instructions on data access and preparation, including how to access the raw data, preprocessed data, intermediate data, and generated data, etc.
   - The authors should provide scripts to reproduce all experimental results for the new proposed method and baselines. If only a subset of experiments are reproducible, they should state which ones are omitted from the script and why.
   - At submission time, to preserve anonymity, the authors should release anonymized versions (if applicable).
   - Providing as much information as possible in supplemental material (appended to the paper) is recommended, but including URLs to data and code is permitted.

6. **Experimental setting/details**

   Question: Does the paper specify all the training and test details (e.g., data splits, hyperparameters, how they were chosen, type of optimizer, etc.) necessary to understand the results?

   Answer: [Yes]

   Justification: All preprocessing information and reproducibility information can be found in the methods section, appendix, or in the code.

   Guidelines:

- The answer NA means that the paper does not include experiments.
- The experimental setting should be presented in the core of the paper to a level of detail that is necessary to appreciate the results and make sense of them.
- The full details can be provided either with the code, in appendix, or as supplemental material.

7. **Experiment statistical significance**

   Question: Does the paper report error bars suitably and correctly defined or other appropriate information about the statistical significance of the experiments?

   Answer: [Yes]

   Justification: All accuracy metrics have standard error, and box plots and p values are provided for all graphical and morphological analyses.

   Guidelines:

   - The answer NA means that the paper does not include experiments.
   - The authors should answer "Yes" if the results are accompanied by error bars, confidence intervals, or statistical significance tests, at least for the experiments that support the main claims of the paper.
   - The factors of variability that the error bars are capturing should be clearly stated (for example, train/test split, initialization, random drawing of some parameter, or overall run with given experimental conditions).
   - The method for calculating the error bars should be explained (closed form formula, call to a library function, bootstrap, etc.)
   - The assumptions made should be given (e.g., Normally distributed errors).
   - It should be clear whether the error bar is the standard deviation or the standard error of the mean.
   - It is OK to report 1-sigma error bars, but one should state it. The authors should preferably report a 2-sigma error bar than state that they have a 96% CI, if the hypothesis of Normality of errors is not verified.
   - For asymmetric distributions, the authors should be careful not to show in tables or figures symmetric error bars that would yield results that are out of range (e.g. negative error rates).
   - If error bars are reported in tables or plots, The authors should explain in the text how they were calculated and reference the corresponding figures or tables in the text.

8. **Experiments compute resources**

   Question: For each experiment, does the paper provide sufficient information on the computer resources (type of compute workers, memory, time of execution) needed to reproduce the experiments?

   Answer: [Yes]

   Justification: The work uses publicly available datasets [31] and synthetic data, with no personally identifiable information. All experiments comply with standard practices in computational biology and machine learning, and there are no ethical concerns related to data collection, usage, or reporting. The research was conducted in accordance with the NeurIPS Code of Ethics.

   Guidelines:

   - The answer NA means that the paper does not include experiments.
   - The paper should indicate the type of compute workers CPU or GPU, internal cluster, or cloud provider, including relevant memory and storage.
   - The paper should provide the amount of compute required for each of the individual experimental runs as well as estimate the total compute.
   - The paper should disclose whether the full research project required more compute than the experiments reported in the paper (e.g., preliminary or failed experiments that didn't make it into the paper).

9. **Code of ethics**

Question: Does the research conducted in the paper conform, in every respect, with the NeurIPS Code of Ethics `https://neurips.cc/public/EthicsGuidelines`?

Answer: [Yes]

Justification: We conform to the code of ethics.

Guidelines:

- The answer NA means that the authors have not reviewed the NeurIPS Code of Ethics.
- If the authors answer No, they should explain the special circumstances that require a deviation from the Code of Ethics.
- The authors should make sure to preserve anonymity (e.g., if there is a special consideration due to laws or regulations in their jurisdiction).

10. **Broader impacts**

Question: Does the paper discuss both potential positive societal impacts and negative societal impacts of the work performed?

Answer: [No]

Justification: Our paper focuses on methodological contributions and empirical validation. While in the abstract and conclusion we highlight potential benefits for precision oncology and interpretable spatial biomarkers, there is no explicit consideration of risks (e.g., misuse, bias, privacy) or mitigation strategies.

Guidelines:

- The answer NA means that there is no societal impact of the work performed.
- If the authors answer NA or No, they should explain why their work has no societal impact or why the paper does not address societal impact.
- Examples of negative societal impacts include potential malicious or unintended uses (e.g., disinformation, generating fake profiles, surveillance), fairness considerations (e.g., deployment of technologies that could make decisions that unfairly impact specific groups), privacy considerations, and security considerations.
- The conference expects that many papers will be foundational research and not tied to particular applications, let alone deployments. However, if there is a direct path to any negative applications, the authors should point it out. For example, it is legitimate to point out that an improvement in the quality of generative models could be used to generate deepfakes for disinformation. On the other hand, it is not needed to point out that a generic algorithm for optimizing neural networks could enable people to train models that generate Deepfakes faster.
- The authors should consider possible harms that could arise when the technology is being used as intended and functioning correctly, harms that could arise when the technology is being used as intended but gives incorrect results, and harms following from (intentional or unintentional) misuse of the technology.
- If there are negative societal impacts, the authors could also discuss possible mitigation strategies (e.g., gated release of models, providing defenses in addition to attacks, mechanisms for monitoring misuse, mechanisms to monitor how a system learns from feedback over time, improving the efficiency and accessibility of ML).

11. **Safeguards**

Question: Does the paper describe safeguards that have been put in place for responsible release of data or models that have a high risk for misuse (e.g., pretrained language models, image generators, or scraped datasets)?

Answer: [NA]

Justification: There are no suck risks for responsible release of data.

Guidelines:

- The answer NA means that the paper poses no such risks.
- Released models that have a high risk for misuse or dual-use should be released with necessary safeguards to allow for controlled use of the model, for example by requiring that users adhere to usage guidelines or restrictions to access the model or implementing safety filters.

- Datasets that have been scraped from the Internet could pose safety risks. The authors should describe how they avoided releasing unsafe images.
- We recognize that providing effective safeguards is challenging, and many papers do not require this, but we encourage authors to take this into account and make a best faith effort.

12. **Licenses for existing assets**

    Question: Are the creators or original owners of assets (e.g., code, data, models), used in the paper, properly credited and are the license and terms of use explicitly mentioned and properly respected?

    Answer: [Yes]

    Justification: The only code used is the ProtoPNet code, and credit is explicitly given (MIT License). We also use the CORDS dataset for the lung cancer data, which is publicly available deposited in zenodo under a Creative Commons Attribution 4.0 International license.

    Guidelines:

    - The answer NA means that the paper does not use existing assets.
    - The authors should cite the original paper that produced the code package or dataset.
    - The authors should state which version of the asset is used and, if possible, include a URL.
    - The name of the license (e.g., CC-BY 4.0) should be included for each asset.
    - For scraped data from a particular source (e.g., website), the copyright and terms of service of that source should be provided.
    - If assets are released, the license, copyright information, and terms of use in the package should be provided. For popular datasets, `paperswithcode.com/datasets` has curated licenses for some datasets. Their licensing guide can help determine the license of a dataset.
    - For existing datasets that are re-packaged, both the original license and the license of the derived asset (if it has changed) should be provided.
    - If this information is not available online, the authors are encouraged to reach out to the asset's creators.

13. **New assets**

    Question: Are new assets introduced in the paper well documented and is the documentation provided alongside the assets?

    Answer: [Yes]

    Justification: Documentation for synthetic data, preprocessing, etc. is all available in the supplementary zip file in the code.

    Guidelines:

    - The answer NA means that the paper does not release new assets.
    - Researchers should communicate the details of the dataset/code/model as part of their submissions via structured templates. This includes details about training, license, limitations, etc.
    - The paper should discuss whether and how consent was obtained from people whose asset is used.
    - At submission time, remember to anonymize your assets (if applicable). You can either create an anonymized URL or include an anonymized zip file.

14. **Crowdsourcing and research with human subjects**

    Question: For crowdsourcing experiments and research with human subjects, does the paper include the full text of instructions given to participants and screenshots, if applicable, as well as details about compensation (if any)?

    Answer: [NA] .

    Justification: The paper does not involve crowdsourcing nor research with human subjects.

Guidelines:

- The answer NA means that the paper does not involve crowdsourcing nor research with human subjects.
- Including this information in the supplemental material is fine, but if the main contribution of the paper involves human subjects, then as much detail as possible should be included in the main paper.
- According to the NeurIPS Code of Ethics, workers involved in data collection, curation, or other labor should be paid at least the minimum wage in the country of the data collector.

15. **Institutional review board (IRB) approvals or equivalent for research with human subjects**

    Question: Does the paper describe potential risks incurred by study participants, whether such risks were disclosed to the subjects, and whether Institutional Review Board (IRB) approvals (or an equivalent approval/review based on the requirements of your country or institution) were obtained?

    Answer: [NA]

    Justification: no data was collected in this paper.

    Guidelines:

    - The answer NA means that the paper does not involve crowdsourcing nor research with human subjects.
    - Depending on the country in which research is conducted, IRB approval (or equivalent) may be required for any human subjects research. If you obtained IRB approval, you should clearly state this in the paper.
    - We recognize that the procedures for this may vary significantly between institutions and locations, and we expect authors to adhere to the NeurIPS Code of Ethics and the guidelines for their institution.
    - For initial submissions, do not include any information that would break anonymity (if applicable), such as the institution conducting the review.

16. **Declaration of LLM usage**

    Question: Does the paper describe the usage of LLMs if it is an important, original, or non-standard component of the core methods in this research? Note that if the LLM is used only for writing, editing, or formatting purposes and does not impact the core methodology, scientific rigorousness, or originality of the research, declaration is not required.

    Answer: [NA]

    Justification: No LLMs used.

    Guidelines:

    - The answer NA means that the core method development in this research does not involve LLMs as any important, original, or non-standard components.
    - Please refer to our LLM policy (`https://neurips.cc/Conferences/2025/LLM`) for what should or should not be described.