



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.

Leveraging Cell-level Spatial Transcriptomics as Weak Supervision for Nuclei Classification

Kazuya Nishimura¹, Ryoma Bise², Haruka Hirose³, and Yasuhiro Kojima³

¹ The University of Osaka, Osaka, Japan

k.nishimura.d3c@osaka-u.ac.jp <https://is.d3c.osaka-u.ac.jp/en/>

² Kyushu University, Fukuoka, Japan

³ National Cancer Center Japan, Tokyo, Japan

Abstract. Deep learning-based nuclei analysis in histopathology relies heavily on large-scale, expert-curated datasets that are costly and difficult to scale across organs. To address this, we propose a scalable framework leveraging cell-level spatial transcriptomics (ST) as an alternative supervision source for nuclei classification. Cell-level ST provides both gene expression profiles and nucleus masks, enabling automatic generation of image-label pairs without manual annotation. We convert gene expression into cell-type labels and introduce an image-oriented classification strategy to effectively transfer these identities to histopathology images. Evaluated on unseen tissues, our framework achieves comparable segmentation and classification performance compared with conventional supervised approaches, despite using fewer manually annotated datasets. Code is publicly available at <https://github.com/naivete5656/CDGE>.

Keywords: Spatial transcriptomics · Digital pathology.

1 Introduction

Automated nuclei segmentation and classification from pathology images are fundamental tasks in computational pathology and quantitative precision medicine. Despite the remarkable progress of deep neural networks [15, 13, 6, 30], their performance critically depends on large-scale pixel-level annotations. In histopathology, generating such annotations is labor-intensive, costly, and subject to substantial inter-observer variability, thereby limiting scalability across diverse organs and clinical settings [12]. Reducing reliance on manually curated pixel-level dataset remains a central challenge.

To address this limitation, we propose a Cross-modal automatic supervision task, in which cell-level gene expression profiles obtained from spatial transcriptomics (ST) serve as supervisory signals for nuclei segmentation and classification. Emerging ST platforms, such as Xenium [18], simultaneously acquire H&E-stained images, DAPI-stained nuclear images, and spatially resolved gene expression measurements from the same tissue section.

Although generating new cell-level ST data remains costly, paired image-molecular datasets are increasingly released and aggregated in reusable public

resources. Driven by strict research data policies from publishers [23, 26] and massive tissue-mapping consortia such as the Human BioMolecular Atlas Program (HuBMAP) [27, 17], a vast amount of multimodal ST data is continuously deposited into public domains and curated bio-repositories [20, 3]. However, there is currently no established framework for deriving reliable image-level classification labels directly from ST data.

We propose a practical training framework that leverages ST data for pathology image analysis. DAPI-derived nuclear masks are used as segmentation labels, while class labels are inferred from gene expression profiles. Since transcriptomic taxonomy is often finer-grained than morphology-based categories, we define image-level classes according to clinically established pathology classifications. To handle ambiguity arising from measurement noise and the continuous nature of gene expression, we introduce an explicit *Unknown* category for cases where confident assignment is not feasible. Following pathological classification taxonomy, we adopt a two-stage classification strategy that explicitly models neoplastic identification.

We first evaluate segmentation robustness by assessing generalization to previously unseen organs and comparing with models trained on manually annotated datasets. We then evaluate classification performance and compare our labeling strategy with existing gene expression-based cell-typing approaches, demonstrating the effectiveness of the proposed task formulation.

The main contributions of this study are as follows:

- We formulate nuclear segmentation and classification using spatial transcriptomics as a cross-modal weakly supervised (cross-modal automatic) learning problem, and position ST as a reusable, potentially scalable supervision source.
- We propose a practical label generation framework with a hierarchical design and an explicit *Unknown* category to mitigate label noise.
- We demonstrate improved segmentation generalization to unseen organs and superior classification performance compared with existing transcriptomics-based labeling strategies.

2 Related Work

Cell Nuclei Segmentation. Nuclei segmentation and classification in H&E-stained histopathology images is a core task in computational pathology. Early methods relied on handcrafted processing techniques [6, 30]. With the availability of large annotated datasets [22, 11, 14], deep learning has become the dominant paradigm. Instance-aware models such as HoVer-Net [13] and CellViT [15] address overlapping nuclei by jointly modeling segmentation and structural cues.

However, these approaches require extensive pixel-level expert-curated dataset. In contrast, we leverage spatial transcriptomics (ST) as weak supervision.

Nuclei Segmentation with Spatial Transcriptomics. Recent work has explored segmentation methods tailored to ST data. SCS [5] and RNA2seg [8] learn

segmentation directly from gene expression signals. Multimodal approaches, such as GeneSegNet [31] and BIDCell [24], integrate gene expression with histological images to improve segmentation accuracy.

While effective, these methods require gene expression measurements at inference time, which may limit generalizability across platforms.

Cell-Type Annotation in Spatial Transcriptomics. Cell-type annotation in ST largely builds on methods developed for single-cell RNA sequencing [7]. Marker-based approaches assign cell identities using predefined gene signatures [1, 2], but their performance depends on the robustness of selected markers. Reference-based methods, including cell2location [21], RCTD [4], and SPOTlight [9], leverage scRNA-seq data to infer cell-type composition in ST data. These approaches require appropriate reference datasets, whose availability and transferability may be limited across tissues and disease conditions.

Importantly, these methods are primarily designed for molecular analysis and are not designed for image classification.

3 Cross-modal Weak Supervised Learning

We formulate nuclear segmentation and classification in histopathology as a cross-modal weakly supervised learning problem, where spatial transcriptomics (ST) provides supervisory signals without manual pixel-level annotation.

Let $(S, \{(b_j, g_j)\}_{j=1}^{N_{\text{cell}}})$ denote the data acquired using a Xenium platform [18], where S is a high-resolution whole-slide image (WSI) obtained from tissue type t , b_j is nuclear boundary mask derived from DAPI staining, and g_j is the corresponding gene expression profile with $g_j \in \mathbb{R}^{N_g}$.

Since transcriptomic measurements provide molecular characterization rather than image-level class labels, we define a mapping $\phi : \mathbb{R}^{N_g} \rightarrow \mathcal{Y}$, that converts each gene expression vector g_j into a discrete cell-type label $y_j \in \mathcal{Y}$. This process yields cell-level labels $\{y_j\}_{j=1}^{N_{\text{cell}}}$ inferred from molecular information rather than manual image annotation.

To enable model training, we partition the WSI S into image patches $\{x_i\}_{i=1}^{N_{\text{patch}}}$. For each patch x_i , we define the index set $\mathcal{J}_i = \{j \mid b_j \cap x_i \neq \emptyset\}$, which contains the nuclei located within the spatial extent of x_i . The resulting training dataset is given by $\mathcal{D} = \left\{x_i, t, \{(b_j, y_j)\}_{j \in \mathcal{J}_i}\right\}_{i=1}^{N_{\text{patch}}}$, where each patch is associated with the tissue type t , nuclear segmentation masks $\{b_j\}_{j \in \mathcal{J}_i}$, transcriptomics-derived cell-type labels $\{y_j\}_{j \in \mathcal{J}_i}$. Given $\mathcal{D}$, the objective is to train a deep neural network $f_\theta : x_i \mapsto (\hat{M}_i, \hat{Y}_i)$, that jointly predicts a nuclear segmentation map $\hat{M}_i$, a set of nucleus-level class predictions $\hat{Y}_i$.

The supervision is weak in the sense that segmentation masks and class labels are not manually annotated from histopathology images, but instead derived automatically from ST measurements.

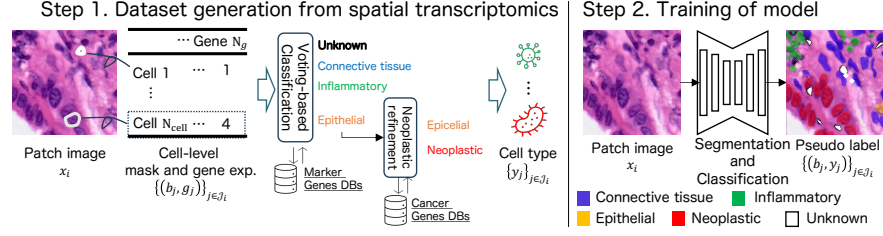


Fig. 1: Overview of proposed framework.

4 Spatial Transcriptomics–Guided Model Training

An overview of the proposed framework is shown in Fig. 1. Following the formulation above, the pipeline consists of two steps: 1) dataset generation from ST observations, and 2) training a joint segmentation and classification model. Instead of manually curated cell-type labels, pseudo-labels are derived in three stages: gene expression clustering, marker-based cluster classification, and neoplastic refinement.

In this study, we design the classification taxonomy with the objective of enabling multi-organ recognition based on Pannuke dataset [11]. Although not covered in this study, alternative classification taxonomy can also be adopted depending on the target application, such as organ-specific subtype classification designed by Lizard [14] or PUMA [25] datasets.

Unlike PanNuke, which defines a *Dead* category, we alternatively introduce an *Unknown* label in our classification scheme. This class is intended to include cells with unstable or unreliable observations, as well as samples for which confident assignment to a specific category is difficult.

This design is motivated by the nature of gene expression measurements: transcriptomic profiles form a continuous spectrum rather than strictly separable clusters, and the reliability of expression measurements is not uniform across all cells. By incorporating an explicit *Unknown* category, we aim to account for ambiguity and variability inherent in the observed molecular data.

4.1 Dataset Generation from Spatial Transcriptomics

Based on the defined classes, we adopt a hierarchical classification strategy following the PanNuke taxonomy.

Transcriptomic Clustering. For each ST sample, cells are clustered based on their gene expression profiles using the Leiden algorithm [29], applied to a nearest-neighbor graph constructed in principal component space. The resulting cluster assignment $c_j \in \{1, \dots, K\}$ is used as a proxy for cell identity.

Marker Gene-Based Cluster Classification. For each cluster k , we compute differentially expressed genes (DEGs) using the Wilcoxon rank-sum test. The top- N ranked genes $\{g_1^k, \dots, g_N^k\}$ are then matched against two marker gene

databases: PanglaoDB [10], CellMarker [16]. These databases provide mappings from individual genes to known cell types, which are further mapped to four coarse categories, $\mathcal{C} = \{\text{Epithelial}, \text{Inflammatory}, \text{Connective}, \text{Unknown}\}$, based on name matching.

For each gene g_l^k at rank l , a category vote vector $\mathbf{v}(g_l^k) \in \mathbb{R}^{|\mathcal{C}|}$ is retrieved from the databases. The cluster-level category score is computed as $\mathbf{s}^k = \sum_{l=1}^N (M-l) \mathbf{v}(g_l^k)$, where the weight $(M-l)$ implements rank-based down-weighting, giving higher-ranked DEGs greater influence. The predicted category for cluster k is $\hat{c}^k = \arg \max_{c \in \mathcal{C}} s_c^k$. The final cell-level label y_j is determined by propagating the cluster-level prediction to all cells belonging to that cluster.

If total votes s_c^k fall below τ_{vote} , fallback classification uses the full database. Clusters with insufficient marker evidence are labeled *Unknown* and treated as unlabeled during training.

Neoplastic Refinement. Clusters initially classified as *Epithelial* undergo a secondary evaluation to distinguish normal epithelial cells from neoplastic (cancer) cells. We construct two neoplastic-related gene sets from CellMarker and CancerSEA [32]. The fraction of the top- M DEGs overlapping with a curated set of cancer-associated genes is computed. If this ratio exceeds a threshold τ_{cancer} , the cluster is re-labeled as *Neoplastic*.

Unlabeled or unclassified cells are assigned the Unknown label and are excluded from the supervised loss during training.

4.2 Training of Nuclei Segmentation and Classification Model

Given a patch image x_i , its nuclear masks and estimated cell-type, and the tissue label t , denoted as $\mathcal{Y}_i = \{(b_j, y_j)\}_{j \in \mathcal{J}_i}$, we train a nuclei segmentation and classification model. We use CellViT [15]. It has an encoder-decoder structure with three branches for nuclei-related predictions: (1) nuclei probability (binary segmentation), (2) horizontal and vertical distance maps, and (3) nuclei type classification, as well as one additional branch for tissue classification. For training, we convert the $\mathcal{Y}_i$ into a binary nuclei mask $B \in \mathbb{R}^{H \times W}$, horizontal and vertical distance maps $D \in \mathbb{R}^{H \times W \times 2}$, and a nuclei type mask $S \in \mathbb{R}^{H \times W \times C}$.

$$L_{\text{total}}(x_i, \mathcal{Y}_i, t) = L_{\text{NP}}(x_i, B) + L_{\text{HV}}(x_i, D) + L_{\text{NT}}(x_i, S) + L_{\text{TC}}(x_i, t), \quad (1)$$

where L_{NP} , L_{HV} , L_{NT} , and L_{TC} denote the loss functions for each branches. For the nuclei type classification loss L_{NT} , pixels annotated as Unknown are excluded from the loss computation.

5 Experiments

Implementation details. To ensure a fair comparison and reproducibility, all hyperparameters were set to the default CellViT configuration for training on the PanNuke dataset [15]. Specifically, we used the same optimizer, learning rate schedule, input patch size, loss functions, and number of training epochs

Table 1: Dataset statics. *: patches were randomly sampled from more than 20,000 whole-slide images (WSIs) for PanNuke.

Dataset	Split	Num. of tissues	Num. of slide	Patches	Nuclei
HEST1K (ST-based)	Train	9	51	373,214	6,643,746
PanNuke (Manual)	Validation	19	*200,000	7,901	196,468
Seg.-Eval. (HEST1K)	Test	4	8	73,091	3,704,652
Cls.-Eval.	Test	1	1	10,238	219,342

Table 2: Segmentation evaluation on Seg.-Eval. comparing supervised (trained on PanNuke dataset) and Ours.

Model	Dice $\uparrow$	Jaccard $\uparrow$	bPQ $\uparrow$	F1 $\uparrow$	Prec. $\uparrow$	Recall $\uparrow$
CellViT (PanNuke)	0.6350	0.5220	0.4148	0.7492	0.7155	0.7862
CellViT (HEST1K, Ours)	0.6585	0.5510	0.4018	0.7562	0.7028	0.8183

as described in the official implementation for PanNuke. The batch size was set to 256, as the number of patches in our dataset is larger than in PanNuke. We set the clustering resolution to $r = 4.0$. For gene voting, the top- $N = 10$ differentially expressed genes (DEGs) were used for classification with rank-based weighting enabled. For the neoplastic check, the top- $M = 20$ DEGs were used, with thresholds of $\tau_{\text{vote}} = 5$ and $\tau_{\text{cancer}} = 0.25$.

Dataset. Table 1 summarizes the dataset statistics: HEST1K [20] for training, PanNuke [11], Seg.-Eval, and Cls.-Eval [19, 18]. Model training was conducted on HEST1K (ST-based), which provides whole-slide H&E images, DAPI-derived nuclear masks, and cell-level gene expression profiles obtained via Xenium. Validation was performed on the PanNuke dataset, which contains expert-curated nucleus instance segmentation masks and annotations for five cell types (Neoplastic, Inflammatory, Connective, Dead, and Epithelial). To ensure a fair comparison, experiments were restricted to the tissues shared by HEST1K and PanNuke (Breast, Colon, Kidney, Liver, Lung, Ovarian, Pancreatic, Prostate, Skin, and Stomach), and the same PanNuke validation split was used for all methods.

Segmentation Evaluation (Seg.-Eval.). Segmentation generalization was evaluated on four organs from HEST1K that are not included in Training dataset.

Classification Evaluation (Cls.-Eval.). For classification evaluation, we used a manually annotated breast cancer Xenium dataset [19, 18], which also includes whole-slide images, masks, and gene expression data. Since the manual annotation included 14 fine-grained labels, we mapped them to four cell types based on cell type: Neoplastic, Inflammatory, Connective, and Epithelial. As there was no label corresponding to dead cells, the “Dead” category was excluded.

Evaluation of Segmentation Performance. We evaluated the segmentation performance on unseen organs by comparing a model trained on the PanNuke dataset [11] with our model trained using spatial transcriptomics data. For a rigorous comparison, both models utilized the identical CellViT architecture,

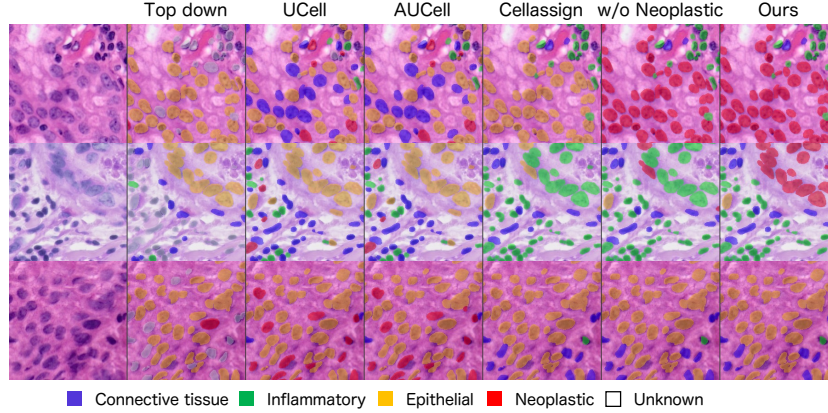


Fig. 2: Example of generated masks.

with the PanNuke baseline trained according to official data splits. Validation was conducted on the corresponding folds as detailed in the Dataset section.

Tables 1 and 2 summarize the dataset statistics and segmentation results, respectively. While the PanNuke baseline was fully supervised across 19 organs, the proposed HEST1K model was trained on 9 organs. Despite the smaller number of slides and organs, HEST1K leverages substantially more nuclear instances per slide via DAPI-derived masks, thereby facilitating dense supervision.

As demonstrated in Table 2, the proposed method achieves performance comparable to the PanNuke baseline. Notably, the most significant improvement is observed in Recall, which increased from 0.7862 to 0.8183 (+0.0321). This suggests enhanced robustness in detecting true nuclei within unseen organs.

Furthermore, DAPI-derived nuclear masks from Xenium data provide dense, spatially aligned supervision that improves detection accuracy. Given that these annotations are automatically generated, the framework is inherently scalable and stands to benefit from the incorporation of additional data.

Evaluation of Classification Performance. For cell-type classification, we compared our method with four reference-free approaches. Reference-based methods were excluded because they require organ-specific scRNA-seq datasets, whose availability varies across tissues. We evaluated: (1) a top-down approach implemented [28], which assigns cell identities based on the average expression of predefined gene sets; (2) UCell [2], a rank-based gene signature enrichment method; (3) AUCell [1], which computes enrichment using the area under the recovery curve; and (4) CellAssign [33], a Bayesian model that assigns cell types based on marker genes. The same marker gene databases were used across all methods. As an ablation study, we also evaluated a variant without hierarchical classification of neoplastic cells (w/o Hier). To compare with conventional supervised approaches, we additionally conducted experiments on the PanNuke dataset following the standard segmentation setup.

Table 3: Classification performance for each model

Model	Macro F1	Acc
TopDown	0.1459	0.1261
UCell	0.1297	0.1595
AUCell	0.2309	0.2325
CellAssign	0.1919	0.2560
w/o Hier.	0.3034	0.3041
Ours	0.3492	0.5359
PanNuke	0.3569	0.5696

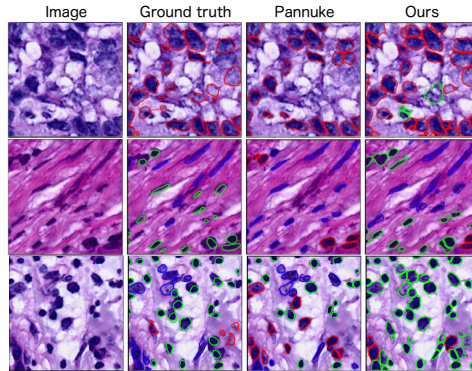


Fig. 3: Example segmentation results.

Fig. 2 shows examples of labels generated by comparisons. As shown in the top two rows of the results, the conventional method misclassifies neoplastic cells as epithelial. In contrast, the proposed method correctly classifies neoplastic cells through the proposed two-stage refinements.

Table 3 reports classification accuracy and macro F1 scores. The proposed method outperforms the reference-free baselines and improves upon the w/o Hier variant, demonstrating the effectiveness of hierarchical refinement. Performance remains slightly below the PanNuke-trained model, likely due to noise in gene expression-based automatic labels. Incorporating manually annotated datasets may further improve robustness.

Fig. 3 presents qualitative comparisons between the PanNuke-trained model and the proposed method. While their classification tendencies differ, the proposed model correctly identifies some nuclei misclassified by the PanNuke-trained model. It indicates the feasibility of assigning classification labels that can support and enhance manual supervision in clinical image analysis workflows.

6 Conclusion

In this paper, we proposed a novel problem formulation for nuclei segmentation and classification in digital pathology that leverages spatial transcriptomics (ST) as an alternative to large-scale manual annotation. By constructing a practical framework, we trained segmentation and classification models without conventional pixel-level annotations. Experimental validation on a public dataset demonstrated the effectiveness of the proposed approach. The results indicate that ST-derived supervision can help for scalable training.

Acknowledgements: This work was supported by JSPS KAKEN JP24KJ2205, 23H03862, and JP23K16991, JST ASPIRE JPMJAP2403, AMED grant JP24ama221609h0001 (PPROMOTE) (to YK) and JP25gm7010014h0001 (PRIME), National Cancer Center Research and Development Fund 2024-A-6 (to YK).

References

1. Aibar, S., et al.: Scenic: single-cell regulatory network inference and clustering. *Nat. Methods* **14**(11), 1083–1086 (2017)
2. Andreatta, M., Carmona, S.J.: Ucell: Robust and scalable single-cell gene signature scoring. *Comput. Struct. Biotechnol. J.* **19**, 3796–3798 (2021)
3. Barrett, T., et al.: NCBI GEO: archive for functional genomics data sets—update. *Nucleic Acids Research* **41**(D1), D991–D995 (2013)
4. Cable, D.M., et al.: Robust decomposition of cell type mixtures in spatial transcriptomics. *Nat. biotechnol.* **40**(4), 517–526 (2022)
5. Chen, H., et al.: Scs: cell segmentation for high-resolution spatial transcriptomics. *Nat. Methods* **20**(8), 1237–1243 (2023)
6. Cheng, J., et al.: Segmentation of clustered nuclei with shape markers and marking function. *IEEE Trans. Biomed. Eng.* **56**(3), 741–748 (2008)
7. Cheng, J., et al.: Benchmarking cell type annotation methods for 10x xenium spatial transcriptomics data. *BMC bioinformatics* **26**(1), 22 (2025)
8. Defard, T., et al.: Rna2seg: a generalist model for cell segmentation in image-based spatial transcriptomics. *bioRxiv* pp. 2025–03 (2025)
9. Elosua-Bayes, M., et al.: Spotlight: seeded nmf regression to deconvolute spatial transcriptomics spots with single-cell transcriptomes. *Nucleic Acids Res.* **49**(9), e50–e50 (2021)
10. Franzén, O., et al.: Panglaodb: a web server for exploration of mouse and human single-cell rna sequencing data. *Database* **2019**, baz046 (2019)
11. Gamper, J., et al.: Pannuke: an open pan-cancer histology dataset for nuclei instance segmentation and classification. In: *Eur. Congr. Digit. Pathol.* pp. 11–19. Springer (2019)
12. Gamper, J., et al.: Pannuke dataset extension, insights and baselines. *arXiv preprint arXiv:2003.10778* (2020)
13. Graham, S., et al.: Hover-net: Simultaneous segmentation and classification of nuclei in multi-tissue histology images. *Med. Image Anal.* **58**, 101563 (2019)
14. Graham, S., et al.: Lizard: a large-scale dataset for colonic nuclear instance segmentation and classification. In: *ICCV.* pp. 684–693 (2021)
15. Hörst, F., et al.: Cellvit: Vision transformers for precise cell segmentation and classification. *Med. Image Anal.* **94**, 103143 (2024)
16. Hu, C., et al.: Cellmarker 2.0: an updated database of manually curated cell markers in human/mouse and web tools based on scrna-seq data. *Nucleic Acids Res.* **51**(D1), D870–D876 (2023)
17. HuBMAP Consortium: Human BioMolecular Atlas Program Data Portal, <https://portal.hubmapconsortium.org/>, accessed: 2026-06-22
18. Janesick, A., et al.: High resolution mapping of the tumor microenvironment using integrated single-cell, spatial and in situ analysis. *Nat. Commun.* **14**(1), 8353 (2023)
19. Janesick, A., et al.: Biomarker quantification in breast cancer using xenium in situ. *bioRxiv* pp. 2025–12 (2025)
20. Jaume, G., et al.: Hest-1k: A dataset for spatial transcriptomics and histology image analysis. *NeurIPS* **37**, 53798–53833 (2024)
21. Kleshchevnikov, V., Shmatko, A., Dann, E., Aivazidis, A., King, H.W., et al.: Cell2location maps fine-grained cell types in spatial transcriptomics. *Nat. biotechnol.* **40**(5), 661–671 (2022)

22. Kumar, N., et al.: A multi-organ nucleus segmentation challenge. *IEEE transactions on medical imaging* **39**(5), 1380–1391 (2019)
23. Nature Portfolio: Reporting standards and availability of data, materials, code and protocols, <https://www.nature.com/nature-portfolio/editorial-policies/reporting-standards>, accessed: 2026-06-22
24. Petukhov, V., et al.: Cell segmentation in imaging-based spatial transcriptomics. *Nat. biotechnol.* **40**(3), 345–354 (2022)
25. Schuiveling, M., et al.: A novel dataset for nuclei and tissue segmentation in melanoma with baseline nuclei segmentation and tissue segmentation benchmarks. *GigaScience* **14**, giaf011 (2025)
26. Springer Nature: Research data policy, <https://www.springernature.com/gp/authors/research-data-policy>, accessed: 2026-06-22
27. The HuBMAP Consortium: The human biomolecular atlas program (hubmap). *Nature* **574**(7777), 187–192 (2019)
28. Tirosh, I., et al.: Dissecting the multicellular ecosystem of metastatic melanoma by single-cell rna-seq. *Science* **352**(6282), 189–196 (2016)
29. Traag, V.A., et al.: From louvain to leiden: guaranteeing well-connected communities. *Scientific reports* **9**(1), 5233 (2019)
30. Veta, M., et al.: Automatic nuclei segmentation in h&e stained breast cancer histopathology images. *PloS one* **8**(7), e70221 (2013)
31. Wang, Y., et al.: Genesegnet: a deep learning framework for cell segmentation by integrating gene expression and imaging. *Genome Biol.* **24**(1), 235 (2023)
32. Yuan, H., et al.: Cancersea: a cancer single-cell state atlas. *Nucleic Acids Res.* **47**(D1), D900–D908 (2019)
33. Zhang, A.W., et al.: Probabilistic cell-type assignment of single-cell rna-seq for tumor microenvironment profiling. *Nat. Methods* **16**(10), 1007–1015 (2019)