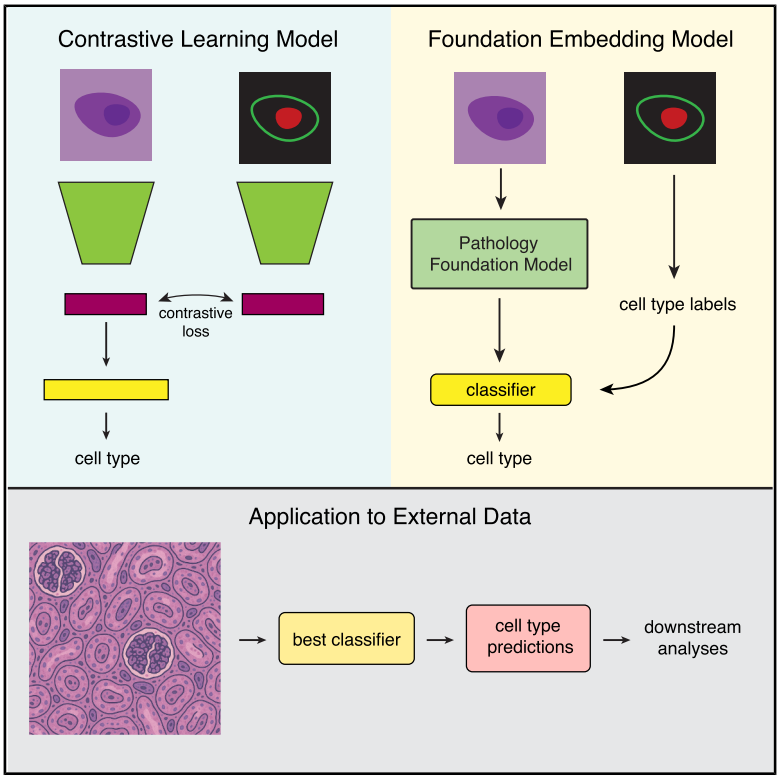
Cell Reports
Methods

Report

Using spatial proteomics to enhance cell type assignments in histology images

Graphical abstract



Authors

Monica T. Dayao, Aaron T. Mayer, Alexandro E. Trevino, Ziv Bar-Joseph

Correspondence

alex@enablemedicine.com (A.E.T.), zivbj@cs.cmu.edu (Z.B.-J.)

In brief

Dayao et al. present a deep learning approach to assign cell types in histology images by training on paired spatial proteomics data. This method enables detailed cell type identification from routine tissue slides, bridging accessible histological techniques and costly molecular imaging while providing a powerful tool for studying tissue organization.

Highlights

- Deep learning framework enables cell typing in H&E images using paired spatial proteomics
- Foundation models outperform custom approaches for cross-dataset generalization
- Application to kidney H&E images identifies disease-relevant cell type differences
- The method bridges clinical histology with spatial proteomics

Dayao et al., 2025, Cell Reports Methods 5, 101204
November 17, 2025
https://doi.org/10.1016/j.crmeth.2025.101204

CellPress

Cell Reports Methods

CellPress
OPEN ACCESS

Report

Using spatial proteomics to enhance cell type assignments in histology images

Monica T. Dayao,[1,2] Aaron T. Mayer,[3] Alexandro E. Trevino,[3,*] and Ziv Bar-Joseph[2,4,5,*]
[1]Joint Carnegie Mellon University-University of Pittsburgh PhD Program in Computational Biology, Pittsburgh, PA 15213, USA
[2]Computational Biology Department, School of Computer Science, Carnegie Mellon University, Pittsburgh, PA 15213, USA
[3]Enable Medicine, Menlo Park, CA 94063, USA
[4]Machine Learning Department, School of Computer Science, Carnegie Mellon University, Pittsburgh, PA 15213, USA
[5]Lead contact
*Correspondence: alex@enablemedicine.com (A.E.T.), zivbj@cs.cmu.edu (Z.B.-J.)
https://doi.org/10.1016/j.crmeth.2025.101204

MOTIVATION Despite its widespread clinical use, H&E staining provides only morphological information, limiting precise cell type identification in complex tissues. While spatial proteomics technologies such as CODEX can reveal molecular-level cellular identities, their high cost and technical complexity prevent routine clinical adoption. This creates a critical bottleneck: researchers need detailed cell type information for precision medicine applications but must rely on standard histology that lacks the molecular resolution to distinguish many cell types. A method that could transfer molecular knowledge from spatial proteomics to enhance standard H&E analysis would address this fundamental limitation.

SUMMARY

Hematoxylin and eosin (H&E) staining has been a standard in clinical histopathology for many decades but lacks molecular detail. Advances in multiplexed spatial proteomics imaging allow cell types and tissues to be annotated by their expression patterns as well as their morphological features. However, these technologies are at present unavailable in most clinical settings. In this work, we present a machine learning framework that leverages histopathology foundation models and paired H&E and spatial proteomic imaging data to enable enhanced cell type annotation on H&E-only datasets. We trained and evaluated our method on kidney datasets with paired H&E and spatial proteomic imaging data and found that models trained using our methods outperform models trained directly on the imaging data. We also show how our framework can be used to study biological differences between two major kidney diseases.

INTRODUCTION

The development of highly multiplexed spatial proteomics imaging technologies, such as CODEX/Phenocycler,[1] Cell DIVE,[2] and others, have enabled detailed profiling of spatially resolved single cells *in situ*. Such data offers deep insights into cell and tissue organization, cell-cell interactions impacting disease progression,[3] and tissue heterogeneity.[4] Furthermore, the high-plex nature of the data allows for detailed annotation of cell types in a given sample, as researchers can use a panel of tens of antibodies to distinguish many cell types of interest. Despite these advances, several challenges hinder the direct application of these technologies to patient samples. One significant barrier is the high cost associated with these techniques, often amounting to several hundred US dollars per tissue slide.[5] This expense can be prohibitive for large-scale clinical studies or routine diagnostic use. Additionally, the complexity and volume of data generated by high-plex spatial proteomics require substantial computational resources and expertise.[6–8] The throughput and turnaround times currently do not meet the demands of routine pathology, and variability in tissue processing and staining further hinders reproducibility.[6] Moreover, integrating these high-dimensional datasets into clinical decision-making poses substantial challenges due to a lack of harmonized data interpretation frameworks.[7]

On the other hand, pathologists have been using histopathology imaging, such as hematoxylin and eosin staining (H&E), for decades at a very low cost (tens of US dollars per slide). This technology is widely used and applied in diagnostics and histology. For example, The Cancer Genome Atlas (TCGA) program contains ~29,000 formalin-fixed paraffin-embedded and frozen H&E whole-slide images (WSIs). H&E images coupled with advanced computational methods have been used in several analysis and diagnostic applications. These include cancer subtyping, cancer staging, diagnostic prediction, and prognostic prediction.[9–12] However, to date, computational analysis

methods are only able to identify abroad categories of cell types can be distinguished in this type of imaging (such as immune cells, stroma vs. tumor cells in cancer tissues, glomerular cells in kidney, blood vessels, etc.).[13–16] This prevents the use of these images for predictions that are based on detailed cellular microenvironments within tissues.

To address this issue, and to enable cell-based analysis of H&E images, we developed and explored the use of a machine learning framework to enhance cell type annotation of H&E data. We tested two distinct approaches that utilize CODEX data in different ways. The first approach is based on contrastive learning,[17–20] where we directly leverage paired CODEX and H&E image patches as inputs during training. We used a multimodal self-supervised contrastive approach to map input CODEX and H&E images to a joint embedding space, encouraging embeddings of the same cell to be similar across modalities. We then fine-tuned the resulting model using a supervised cell type classification objective. At inference time, the resulting model takes an H&E image patch and outputs a cell type classification. The second approach utilizes CODEX data for label derivation, where we first use the high-plex proteomic information to generate detailed cell type annotations, and then train H&E models using these labels. Specifically, we tested a method based on a foundation model for histopathology (Phikon) that leverages self-supervised learning with masked image modeling for its training.[21] We used Phikon to generate embeddings for the H&E image data and trained a classifier on those embeddings for cell type prediction using the CODEX-derived labels. As baselines, we constructed a supervised cell type classifier based on H&E image features and a naive baseline model. We tested these approaches on two spatial proteomics datasets of kidney transplant rejection[22,23] and applied it to unannotated public H&E data from the Kidney Precision Medicine Project (KPMP).[24] Our method leads to both more accurate cell type assignment and relevant biological insights in kidney tissues, demonstrating the utility of this approach to studying and annotating disease relevant tissue samples.

RESULTS

We developed a machine learning framework that leverages paired spatial proteomic and H&E data for annotating cells in H&E image patches to enable higher resolution cell-based analysis of H&E-only datasets. We developed and tested two types of models: (1) a multi-modal contrastive model that uses high-plex proteomic (CODEX) and H&E image patches as input and maps them to a joint embedding space by optimizing a contrastive objective, followed by supervised fine-tuning (Figures 1A and 1C) and (2) a cell type classifier that utilizes embeddings generated from a histopathology foundation model (Figure 1B). We trained and evaluated our models on two paired H&E-CODEX kidney transplant rejection datasets. Once trained, the models can be used to assign cells in a given H&E image patch.

Contrastive model improves intra-dataset performance

The multi-modal contrastive model takes CODEX and H&E image patches as input and maps them to a joint embedding space by optimizing a contrastive objective (STAR Methods). After contrastive pre-training, the model is fine-tuned for cell type prediction. We trained the model on two paired H&E-CODEX datasets of kidney transplant rejection (TWMU-TR, Stanford-TR, and STAR Methods), and measured the performance a held-out validation set by computing the average weighted F1 score for the H&E predictions only. We also used a histology foundation model, namely Phikon,[21] for our H&E datasets, and trained classifiers on them, using labels derived from the paired CODEX data, to evaluate the "out-of-the-box" performance of these embeddings on the cell type prediction task.

To evaluate these two methods, we compared their performances to a purely supervised model that only takes an H&E image patch as input (H&E convolutional neural network [CNN] and STAR Methods) and to a naive method that predicts cells based on the (known) background distribution. Results, presented in Table 1, show that the contrastive pre-training approach leads to higher prediction performance compared to all other methods for both datasets (7.6%, 7.6%, and 204.6% average increase over the foundation embedding classifier, H&E CNN, and the naive baseline, in weighted F1, respectively). The foundation embedding classifier was the second-best model, and slightly worse on the TWMU-TR, indicating that when using internal validation of a single dataset, the general foundation model does not provide much improvement compared to a CNN that is trained directly on image data. Figure 2 presents the average F1 score for each of the individual cell type classes for both the Stanford-TR (a) and TWMU-TR (b) validation results. For both datasets, the top performing cell types are the proximal and distal tubules, and the worst-performing cell types include B cells and glomerular endothelial cells (of which there are only 3.5k and 2k cells in the TWMU-TR training set, respectively, representing less than 1% of the cells) and the $CD4^+$ and $CD8^+$ T cells. These results are unsurprising as tubules have a distinct morphology, and $CD4^+$ and $CD8^+$ T cells cannot traditionally be distinguished by histology alone.

Using the predictions from our contrastive model, we visualized the predicted cell types on the TWMU-TR validation dataset (Figure 3). We found that the general spatial distribution of the predicted cell types matches well with the ground truth image. Specifically, the proximal and distal tubules are readily identified and the podocyte/lymphatic cells are co-located with the glomerular endothelial cells. Figure S1 shows the confusion matrix for these predictions.

Embeddings from foundation models enable cross-dataset generalization

We next performed cross-dataset experiments by training on the entire TWMU-TR dataset and testing on the Stanford-TR dataset. We excluded certain cell types that were only present in one of the two datasets or were inconsistently labeled in the two datasets (STAR Methods). Results, shown in Table 1 and Figure 2C, show that the contrastive loss model outperforms the H&E CNN and the random naive baseline (average F1 score of 0.229 for the contrastive model, 0.132 for the H&E CNN, and 0.2 for naive baseline). However, for this analysis, we observe that the foundation model embeddings followed by the logistic regression classifier led to the best results (5%, 137.9%, and 57% improvement over the contrastive model, H&E CNN, and naive baseline, respectively). This supports the use of the more

CellPress
OPEN ACCESS

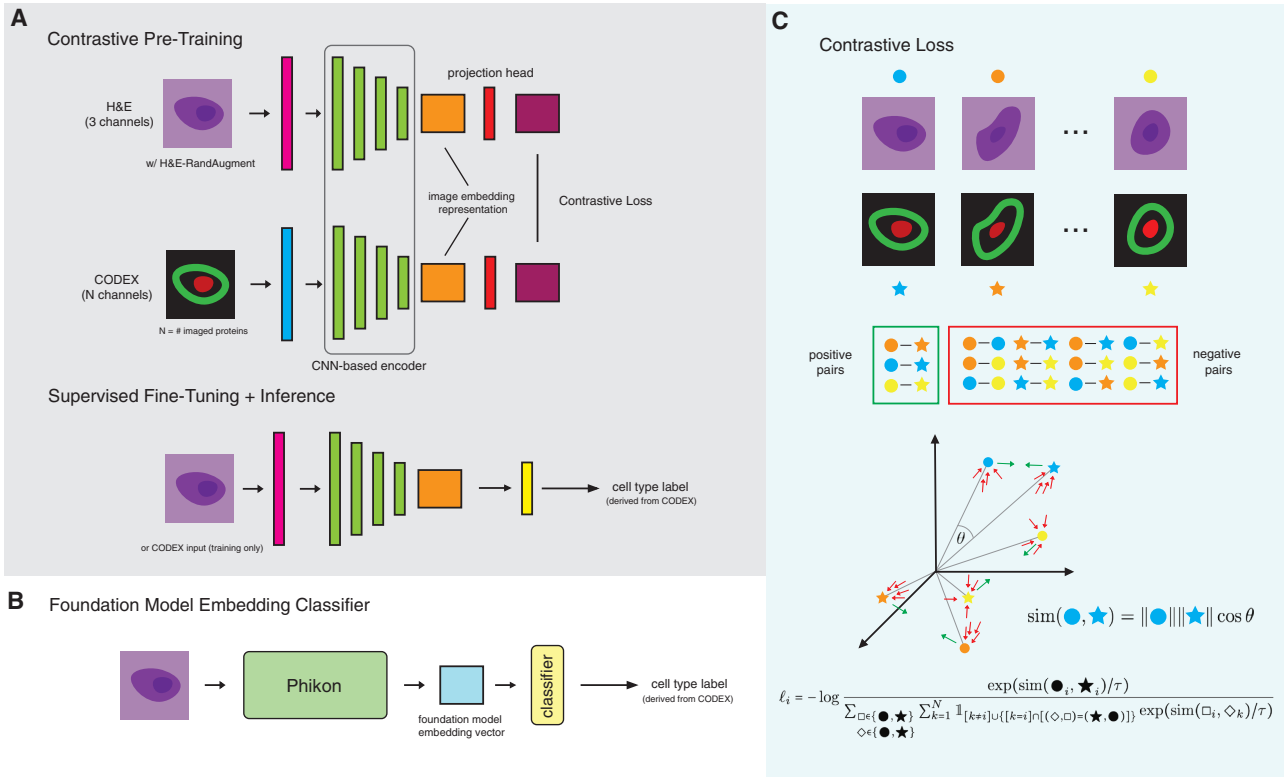


Figure 1. Method overview

(A) Multi-modal contrastive pre-training and supervised fine-tuning setup. The model is trained on paired H&E-CODEX cell-centered image patches using a contrastive loss function. The 3 image channels in the H&E images are red, green, and blue channels. Next, the model is fine-tuned using cross-entropy loss to predict cell types.

(B) Classification using foundation model embeddings. CODEX data are used to derive cell type labels for training, while image embeddings are generated using the histopathology foundation model Phikon on H&E images. A logistic regression model is trained on the resulting embeddings to predict cell type.

(C) Input setup for the contrastive loss learning process. Positive pairs are H&E-CODEX image pairs of the same cell. Negative pairs are all other pair combinations in the training mini-batch. The method attempts to embed positive pairs close to each other.

general foundation model when trying to generalize to an unseen dataset. We note that because the naive baseline method randomly predicts cell type based on the cell type proportion present in the training set, its performance improves in the cross-dataset setting, as the number of cell types decreases from 9 in the intra-dataset setting to 6.

Table 1. Average weighted F1 score

	Weighted F1		
Model type	TWMU-TR	Stanford-TR	Cross-dataset
Foundation embedding classifier	0.456	0.496	**0.314**
Multi-modal contrastive	**0.510**	**0.513**	0.229
H&E CNN	0.486	0.465	0.132
Naive baseline	0.178	0.159	0.200

"TWMU-TR" and "Stanford-TR" columns are results on the respective validation datasets. "Cross-dataset" shows the results on Stanford-TR after training on TWMU-TR. Bold font indicates top performing method for the corresponding dataset.

Application to clinical datasets show differences between disease states

Given its success in cross-dataset analysis, we next used the foundation embedding model and classifier to annotate two additional H&E datasets. For this, we used data from the Kidney Precision Medicine Project (KPMP). We processed H&E samples from two disease conditions, chronic kidney disease (CKD) and acute kidney injury (AKI), and used our classifier to predict cell types for these samples (Figure 4A; STAR Methods; Table S1). There were 138 samples for CKD and 53 samples for AKI. We computed the cell type proportions across all samples for each condition. Performing statistical analysis on these fraction values showed that there were significant differences in proportion of proximal tubules, distal tubules, and podocyte/lymphatic cells between the two conditions (Figure 4B; STAR Methods). Figures S2–S5 show random image patches for each cell type predicted by the foundation embedding model. The KPMP-AKI collection protocol primarily focused on patients who were affected by AKI as a result of acute tubular necrosis (ATN), which is characterized by tubular epithelial cell death and dysfunction in one or several tubular segments,[25] and the KPMP-CKD collection protocol primarily

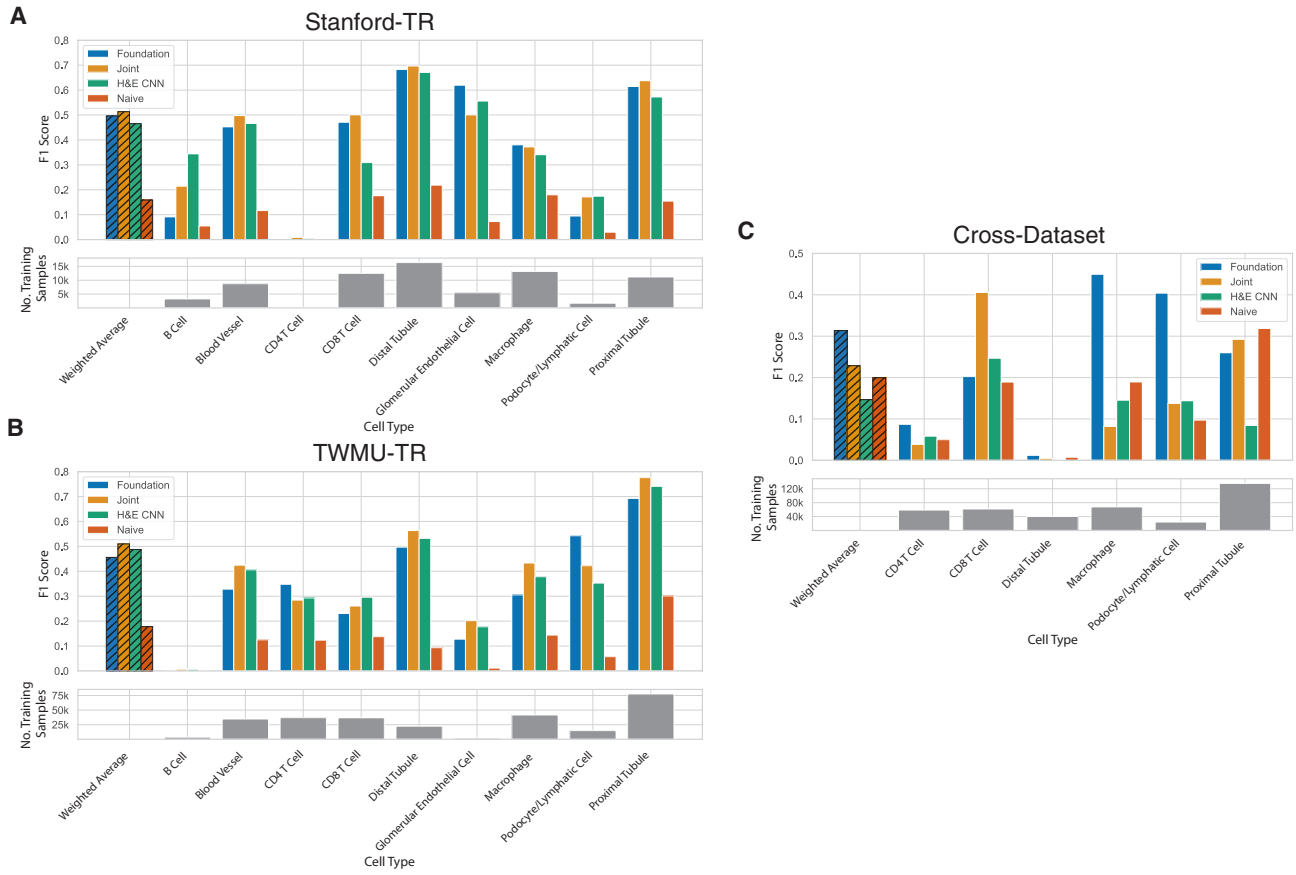


Figure 2. Average F1 scores by cell type

(A) On validation set of Stanford-TR.

(B) On validation set of TWMU-TR.

(C) Cross-dataset results (trained on TWMU-TR, tested on Stanford-TR).

See also Figure S1.

focused on patients with diabetic kidney disease.[24] ATN-related AKI has been shown to cause death of both proximal and distal tubules.[25–28] The progression of DKD-related CKD has a strong link to proximal tubular injury.[29,30] These can explain why our model predicted a lower proportion of distal tubules for the AKI samples and a lower proportion of proximal tubules for the CKD samples. The higher proportion of lymphatic cells in CKD samples could be related to lymphangiogenesis, or new lymphatic growth, during the progression of CKD.[31] While there is still very little known about lymphangiogenesis in the pathophysiology of kidney disease, there have been several studies linking lymphangiogenesis to CKD, and specifically DKD-related CKD.[32–34]

While we initially explored both contrastive and foundation embedding-based approaches to this task, we chose to apply the foundation embedding model due to its superior performance in cross-dataset evaluation (Table 1). This model demonstrated better generalization across datasets than the contrastive method, which performed well in-distribution but failed to generalize effectively. Thus, we focused on the most robust and interpretable model.

DISCUSSION

Highly multiplexed spatial proteomics imaging give researchers the opportunity to study human tissue with unprecedented complexity by combining spatial context with rich molecular information. These technologies, however, are often prohibitively expensive and remain difficult to scale for broad clinical or cohort-level use. Their application is constrained by high per-slide costs, technical complexity, and the need for specialized infrastructure and expertise.[6,7] As such, they are typically limited to a small number of samples, restricting their use in large-scale studies of human disease. In contrast, histopathology imaging modalities such as H&E are several orders of magnitude cheaper to acquire and are routinely used by pathologists for both diagnostics and research. However, H&E images provide only morphological information, making them inherently more limited than spatial proteomics data. This limitation makes identifying individual cell types particularly challenging. Given the large number of cells in a typical image (often in the hundreds of thousands), human annotators can only label a small fraction of these cells, limiting the feasibility of detailed, cell-level analysis.

CellPress
OPEN ACCESS

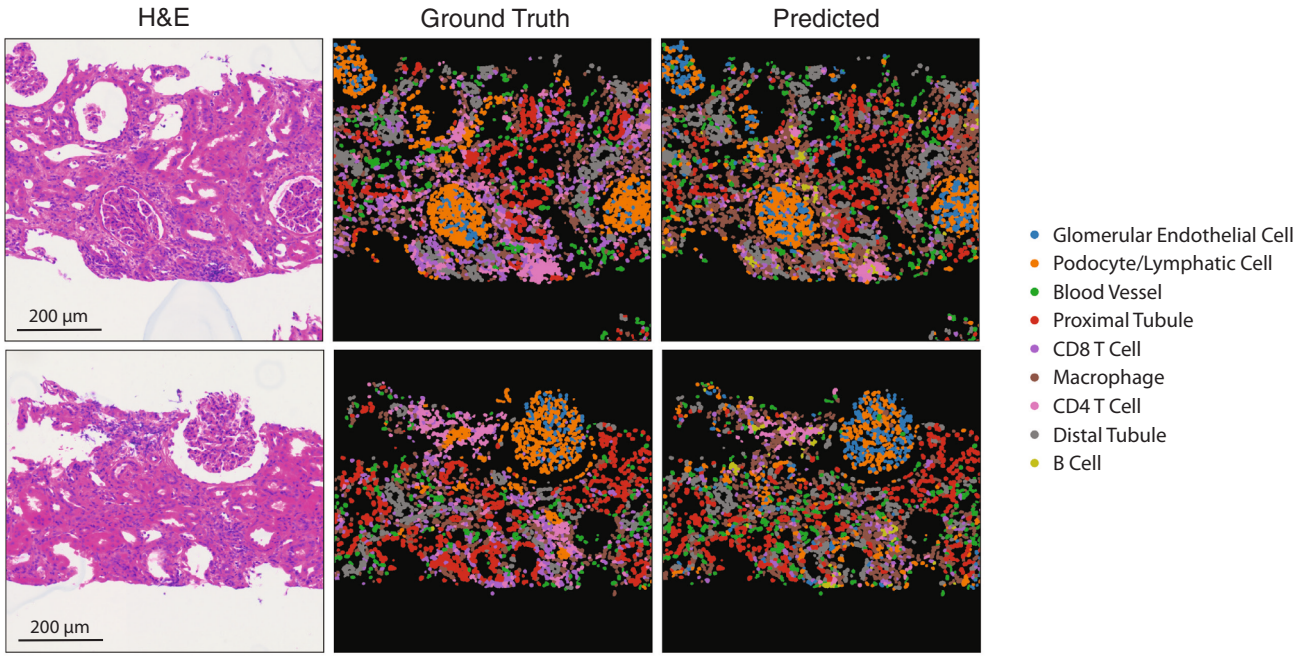


Figure 3. Examples of predicted cell types from the contrastive model on the validation set of TWMU-TR

Left: H&E image; middle: cell segmentation masks colored by ground truth cell type labels; right: cell segmentation masks colored by predicted cell type. Scale bar: 200 µm.

See also Figure S1.

As a result, researchers typically annotate broader tissue structures, which may obscure finer differences in cellular organization. These differences can be critical for understanding disease mechanisms and pathological states.[35–37] To address this gap, our method leverages paired H&E and spatial proteomics data to transfer molecular-level annotations onto H&E images,

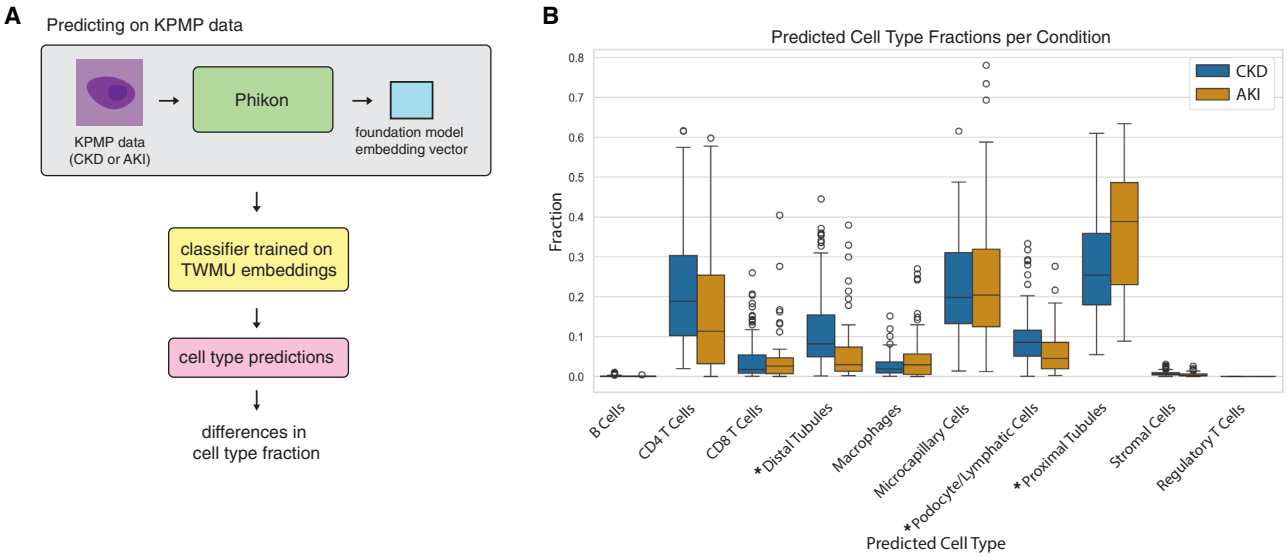


Figure 4. Application to external KPMP data

(A) Predicting on KPMP data using foundation embedding classifier.

(B) Differences in predicted cell type fraction between KPMP-AKI and KPMP-CKD samples. Asterisk indicates p value <0.001 (two-sample t test).

See also Figures S2–S5.

enabling more detailed and scalable cell type analysis in settings where spatial proteomics is impractical.

Here, we developed and applied a machine learning framework to improve computational annotation of H&E data by integration with spatial proteomics. We tested and extended two types of models: a multi-modal contrastive model that directly incorporates both CODEX and H&E image patches as inputs during training to learn joint embeddings across modalities, and a cell type classifier that uses CODEX data for label derivation for training H&E models using embeddings from a histopathology foundation model. These models were used on two separate kidney datasets with paired H&E and CODEX images of the same tissue slice. We note that individual image patches may include more than one cell, but we view this as a strength rather than a limitation. Surrounding cellular context, such as partially visible neighboring cells, can provide informative morphological cues that support more accurate classification. In our intra-dataset analysis, we found that the contrastive model outperforms the baseline model trained on H&E images alone as well as the foundation embedding classifier. In our cross-dataset analysis, on the other hand, the foundation embedding classifier performs best. We believe that the H&E CNN performed poorly in this experiment because it was unable to generalize to overcome batch effects. This is likely because it was trained on a single dataset and image type. We believe the joint model generalizes better because it leverages both CODEX and H&E data. Foundation model embeddings performed best during the cross-dataset experiment since those representations were a result of training on millions of H&E images from many different datasets, enabling better generalization. Our results indicate that learning cell type assignments using CODEX data can help us assign cell types to general H&E images. However, there is likely a limit to the level of granularity that could be achieved. For example, the method distinguishes well between tubules and podocytes/lymphatic cell types but has a hard time distinguishing between cell types that are visually very similar, including $CD4^+$ and $CD8^+$ cells.

As the foundation embedding classifier performed best on the cross-dataset evaluation, we next applied this model on unannotated clinical datasets from the KPMP.[24] These included samples from patients with CKD and AKI. The model successfully annotated both datasets and identified significant differences in the proportions of several cell types. Specifically, we observed differences in the proportions of proximal tubules, distal tubules, and podocyte/lymphatic cells between the two conditions. These findings are consistent with prior studies linking lymphangiogenesis to CKD, including in patient samples and in experimental models of renal fibrosis.[33,34] Tubular injury has also been strongly associated with both AKI and CKD; ATN is a common cause of AKI,[25] and failure of tubular repair following injury is known to drive CKD progression.[30] While the method works reasonably well when annotating cells types that are relatively common, it struggles with annotating rarer cell types (those representing less than 1% of the cells). Overall, accuracies vary across cell types, as we show in Figure 2. Nevertheless, the method accurately annotates several cell types and can be used to identify biologically relevant differences between samples.

While we explored both multi-modal contrastive and foundation model approaches, we selected the foundation embedding model for application to clinical datasets due to its superior generalization performance across datasets. Rather than solely proposing new models, our aim was to assess their suitability for real-world use, such as clinical H&E annotation. The contrastive multi-modal model, while promising in single-dataset settings, showed limited robustness across domains. In contrast, the foundation embedding model offered a more stable and transferable representation by leveraging large-scale H&E pre-training. We believe this highlights an important consideration for future work in multi-modal learning: that simpler, well-generalized models combined with linear probing may currently offer more practical value in translational settings than more complex architectures trained from scratch.

The work presented here uses paired spatial proteomic and H&E datasets for predicting higher resolution cell types on H&E datasets. We hope that our methods open the door to additional, more accurate, analysis of H&E data using foundation models at even a single-cell level.

Limitations of the study

First, the accuracy of the method depends on the quality and representativeness of the paired spatial proteomics and H&E datasets used for training. Variations in tissue preparation, imaging protocols, and staining quality can introduce batch effects that negatively impact model performance and generalization across different datasets or clinical environments. Second, the resolution of cell type annotation achieved by our models is limited. Lastly, the foundation embedding classifier's effectiveness depends on the relevance and coverage of pre-trained embeddings to specific tissues and conditions.

RESOURCE AVAILABILITY

Lead contact

Further information and requests for resources should be directed to and will be fulfilled by the lead contact, Ziv Bar-Joseph (zivbj@cs.cmu.edu).

Materials availability

This study did not generate new unique reagents.

Data and code availability

- Spatial proteomics and H&E datasets used in this study (TWMU-TR dataset, Stanford-TR dataset, and KPMP datasets) are publicly available. Links to processed image data are available at Zenodo: https://doi.org/10.5281/zenodo.16938687 (TWMU-TR), Zenodo: https://doi.org/10.5281/zenodo.16938814 (Stanford-TR), and Zenodo: https://doi.org/10.5281/zenodo.16970433 (KPMP). Raw data and images are available from lead contact upon request.
- All original code has been deposited at Zenodo: https://doi.org/10.5281/zenodo.16937845 and github.com/mdayao/hist-prot-integration.
- Any additional information required to reanalyze the data reported in this paper is available from the lead contact upon request.

ACKNOWLEDGMENTS

This work was supported by the National Institutes of Health (F31LM014194 and T32EB009403 to M.T.D. and OT2OD026682, 1U54AG075931, and 1U24CA268108 to Z.B.-J.) and the National Science Foundation (CBET-2134999 to Z.B.-J.).

The results here are in part based upon data generated by the Kidney Precision Medicine Project (KPMP https://www.kpmp.org). The KPMP is supported by the National Institute of Diabetes and Digestive and Kidney Diseases (NIDDK) through the following grants: U01DK133081, U01DK133091, U01DK133092, U01DK133093, U01DK133095, U01DK133097, U01DK114866, U01DK114908, U01DK133090, U01DK133113, U01DK133766, U01DK133768, U01DK114907, U01DK114920, U01DK114923, U01DK114933, U24DK114886, UH3DK 114926, UH3DK114861, UH3DK114915, and UH3DK114937. We gratefully acknowledge the essential contributions of our patient participants and the support of the American public through their tax dollars.

AUTHOR CONTRIBUTIONS

Conceptualization and methodology, M.T.D., A.E.T., and Z.B.-J.; data curation, M.T.D., A.T.M., and A.E.T.; formal analysis, investigation, software, visualization, writing – original draft, and validation, M.T.D.; supervision and project administration, A.E.T. and Z.B.-J.; resources, A.T.M., A.E.T., and Z.B.-J.; funding acquisition, M.T.D. and Z.B.-J.; writing – review & editing, all authors.

DECLARATION OF INTERESTS

A.T.M. and A.E.T. are employees at Enable Medicine. Z.B.-J. is an employee at Sanofi.

STAR★METHODS

Detailed methods are provided in the online version of this paper and include the following:

- KEY RESOURCES TABLE
- METHOD DETAILS
  - Description of datasets and pre-processing steps
  - Refining cell types with Spatial Cellular Graph Partitioning
  - Overlapping cell types for training and evaluation
  - Removing cell types for cross-dataset evaluation
  - Random data augmentation
  - Multi-modal contrastive model
  - Using foundation model embeddings
  - Cell type assignment using foundation model embeddings
  - Comparison models
- QUANTIFICATION AND STATISTICAL ANALYSIS
  - Cell expression normalization
  - Statistical analysis on KPMP data

SUPPLEMENTAL INFORMATION

Supplemental information can be found online at https://doi.org/10.1016/j.crmeth.2025.101204.

Received: January 14, 2025
Revised: July 26, 2025
Accepted: September 23, 2025
Published: October 15, 2025

REFERENCES

1. Goltsev, Y., Samusik, N., Kennedy-Darling, J., Bhate, S., Hale, M., Vazquez, G., Black, S., and Nolan, G.P. (2018). Deep profiling of mouse splenic architecture with CODEX multiplexed imaging. Cell *174*, 968–981.e15. https://doi.org/10.1016/j.cell.2018.07.010.
2. Gerdes, M.J., Sevinsky, C.J., Sood, A., Adak, S., Bello, M.O., Bordwell, A., Can, A., Corwin, A., Dinn, S., Filkins, R.J., et al. (2013). Highly multiplexed single-cell analysis of formalin-fixed, paraffin-embedded cancer tissue. Proc. Natl. Acad. Sci. USA *110*, 11982–11987. https://doi.org/10.1073/pnas.1300136110.
3. Strasser, M.K., Gibbs, D.L., Gascard, P., Bons, J., Hickey, J.W., Schürch, C.M., Tan, Y., Black, S., Chu, P., Ozkan, A., et al. (2023). Concerted epithelial and stromal changes during progression of Barrett's Esophagus to invasive adenocarcinoma exposed by multi-scale, multi-omics analysis. Preprint at bioRxiv. https://doi.org/10.1101/2023.06.08.544265.
4. Walsh, L.A., and Quail, D.F. (2023). Decoding the tumor microenvironment with spatial technologies. Nat. Immunol. *24*, 1982–1993. https://doi.org/10.1038/s41590-023-01678-9.
5. PhenoCycler (formerly known as CODEX) – SpITR. en-US. https://spitr.ccr.cancer.gov/technologies/codex-co-detection-by-indexing/.
6. Mi, H., Sivagnanam, S., Ho, W.J., Zhang, S., Bergman, D., Deshpande, A., Baras, A.S., Jaffee, E.M., Coussens, L.M., Fertig, E.J., and Popel, A.S. (2024). Computational methods and biomarker discovery strategies for spatial proteomics: a review in immuno-oncology. Brief. Bioinform. *25*, bbae421. https://doi.org/10.1093/bib/bbae421.
7. Mulholland, E.J., and Leedham, S.J. (2024). Redefining clinical practice through spatial profiling: a revolution in tissue analysis. Ann. R. Coll. Surg. Engl. *106*, 305–312. https://doi.org/10.1308/rcsann.2023.0091.
8. Lee, Y., Lee, M., Shin, Y., Kim, K., and Kim, T. (2025). Spatial Omics in Clinical Research: A Comprehensive Review of Technologies and Guidelines for Applications. Int. J. Mol. Sci. *26*, 3949. https://doi.org/10.3390/ijms26093949.
9. Brancati, N., Anniciello, A.M., Pati, P., Riccio, D., Scognamiglio, G., Jaume, G., De Pietro, G., Di Bonito, M., Foncubierta, A., Botti, G., et al. (2022). Bracs: A dataset for breast carcinoma subtyping in h&e histology images. Database *2022*, baac093. https://doi.org/10.1093/database/baac093.
10. Chen, R.J., Ding, T., Lu, M.Y., Williamson, D.F.K., Jaume, G., Song, A.H., Chen, B., Zhang, A., Shao, D., Shaban, M., et al. (2024). Towards a general-purpose foundation model for computational pathology. Nat. Med. *30*, 850–862. https://doi.org/10.1038/s41591-024-02857-3.
11. Fu, Y., Jung, A.W., Torne, R.V., Gonzalez, S., Vöhringer, H., Shmatko, A., Yates, L.R., Jimenez-Linan, M., Moore, L., and Gerstung, M. (2020). Pan-cancer computational histopathology reveals mutations, tumor composition and prognosis. Nat. Cancer *1*, 800–810. https://doi.org/10.1038/s43018-020-0085-8.
12. Luo, X., Zang, X., Yang, L., Huang, J., Liang, F., Rodriguez-Canales, J., Wistuba, I.I., Gazdar, A., Xie, Y., and Xiao, G. (2017). Comprehensive computational pathological image analysis predicts lung cancer prognosis. J. Thorac. Oncol. *12*, 501–509. https://doi.org/10.1016/j.jtho.2016.10.017.
13. Al-Milaji, Z., Ersoy, I., Hafiane, A., Palaniappan, K., and Bunyak, F. (2019). Integrating segmentation with deep learning for enhanced classification of epithelial and stromal tissues in H&E images. Pattern Recognit. Lett. *119*, 214–221. https://doi.org/10.1016/j.patrec.2017.09.015.
14. Nguyen, L., Tosun, A.B., Fine, J.L., Lee, A.V., Taylor, D.L., and Chennubhotla, S.C. (2017). Spatial statistics for segmenting histological structures in H&E stained tissue images. IEEE Trans. Med. Imaging *36*, 1522–1532. https://doi.org/10.1109/TMI.2017.2681519.
15. Jayapandian, C.P., Chen, Y., Janowczyk, A.R., Palmer, M.B., Cassol, C.A., Sekulic, M., Hodgin, J.B., Zee, J., Hewitt, S.M., O'Toole, J., et al. (2021). Development and evaluation of deep learning–based segmentation of histologic structures in the kidney cortex with multiple histologic stains. Kidney Int. *99*, 86–101. https://doi.org/10.1016/j.kint.2020.07.044.
16. Hermsen, M., de Bel, T., Den Boer, M., Steenbergen, E.J., Kers, J., Florquin, S., Roelofs, J.J.T.H., Stegall, M.D., Alexander, M.P., Smith, B.H., et al. (2019). Deep learning–based histopathologic assessment of kidney tissue. J. Am. Soc. Nephrol. *30*, 1968–1979. https://doi.org/10.1681/ASN.2019020144.
17. Chen, T., Kornblith, S., Norouzi, M., and Hinton, G. (2020). A simple framework for contrastive learning of visual representations. Proceedings of the 37th International Conference on Machine Learning, 1597–1607.

18. He, K., Fan, H., Wu, Y., Xie, S., and Girshick, R. (2020). Momentum Contrast for Unsupervised Visual Representation Learning. Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). https://doi.org/10.1109/CVPR42600.2020.00975.

19. Huang, Z., Bianchi, F., Yuksekgonul, M., Montine, T.J., and Zou, J. (2023). A visual–language foundation model for pathology image analysis using medical twitter. Nat. Med. *29*, 2307–2316. https://doi.org/10.1038/s41591-023-02504-3.

20. Wang, T., and Isola, P. (2020). Understanding contrastive representation learning through alignment and uniformity on the hypersphere. Proceedings of the 37th International Conference on Machine Learning, 9929–9939.

21. Filiot, A., Ghermi, R., Olivier, A., Jacob, P., Fidon, L., Camara, A., Mac Kain, A., Saillard, C., and Schiratti, J.-B. (2023). Scaling self-supervised learning for histopathology with masked image modeling. Preprint at medRxiv. https://doi.org/10.1101/2023.07.21.23292757.

22. Hirai, T., Kondo, A., Shimizu, T., Fukuda, H., Tokita, D., Takagi, T., Mayer, A. T., and Ishida, H. (2024). Unveiling spatial immune cell profile in kidney allograft rejections using 36-plex immunofluorescence imaging. Transplantation *108*, 2446–2457. https://doi.org/10.1097/TP.0000000000005107.

23. Zarkhin, V., Kambham, N., Li, L., Kwok, S., Hsieh, S.-C., Salvatierra, O., and Sarwal, M.M. (2008). Characterization of intra-graft B cells during renal allograft rejection. Kidney Int. *74*, 664–673. https://doi.org/10.1038/ki.2008.249.

24. De Boer, I.H., Alpers, C.E., Azeloglu, E.U., Balis, U.G.J., Barasch, J.M., Barisoni, L., Blank, K.N., Bomback, A.S., Brown, K., Dagher, P.C., et al. (2021). Rationale and design of the kidney precision medicine project. Kidney Int. *99*, 498–510. https://doi.org/10.1016/j.kint.2020.08.039.

25. Sancho-Martínez, S.M., López-Novoa, J.M., and López-Hernández, F.J. (2015). Pathophysiological role of different tubular epithelial cell death modes in acute kidney injury. Clin. Kidney J. *8*, 548–559. https://doi.org/10.1093/ckj/sfv069.

26. Amdur, R.L., Chawla, L.S., Amodeo, S., Kimmel, P.L., and Palant, C.E. (2009). Outcomes following diagnosis of acute renal failure in US veterans: focus on acute tubular necrosis. Kidney Int. *76*, 1089–1097. https://doi.org/10.1038/ki.2009.332.

27. Allory, Y., Audard, V., Fontanges, P., Ronco, P., and Debiec, H. (2008). The L1 cell adhesion molecule is a potential biomarker of human distal nephron injury in acute tubular necrosis. Kidney Int. *73*, 751–758. https://doi.org/10.1038/sj.ki.5002640.

28. Wen, Y., Yang, C., Menez, S.P., Rosenberg, A.Z., and Parikh, C.R. (2020). A systematic review of clinical characteristics and histologic descriptions of acute tubular injury. Kidney Int. Rep. *5*, 1993–2001. https://doi.org/10.1016/j.ekir.2020.08.026.

29. Liu, B.-C., Tang, T.-T., Lv, L.-L., and Lan, H.-Y. (2018). Renal tubule injury: a driving force toward chronic kidney disease. Kidney Int. *93*, 568–579. https://doi.org/10.1016/j.kint.2017.09.033.

30. Takaori, K., Nakamura, J., Yamamoto, S., Nakata, H., Sato, Y., Takase, M., Nameta, M., Yamamoto, T., Economides, A.N., Kohno, K., et al. (2016). Severity and frequency of proximal tubule injury determines renal prognosis. J. Am. Soc. Nephrol. *27*, 2393–2406. https://doi.org/10.1681/ASN.2015060647.

31. Donnan, M.D., Kenig-Kozlovsky, Y., and Quaggin, S.E. (2021). The lymphatics in kidney health and disease. Nat. Rev. Nephrol. *17*, 655–675. https://doi.org/10.1038/s41581-021-00438-y.

32. Zarjou, A., Black, L.M., Bolisetty, S., Traylor, A.M., Bowhay, S.A., Zhang, M.-Z., Harris, R.C., and Agarwal, A. (2019). Dynamic signature of lymphangiogenesis during acute kidney injury and chronic kidney disease. Lab. Invest. *99*, 1376–1388. https://doi.org/10.1038/s41374-019-0259-0.

33. Yazdani, S., Navis, G., Hillebrands, J.-L., van Goor, H., and van den Born, J. (2014). Lymphangiogenesis in renal diseases: passive bystander or active participant? Expert Rev. Mol. Med. *16*, e15. https://doi.org/10.1017/erm.2014.18.

34. Tanabe, K., Wada, J., and Sato, Y. (2020). Targeting angiogenesis and lymphangiogenesis in kidney disease. Nat. Rev. Nephrol. *16*, 289–303. https://doi.org/10.1038/s41581-020-0260-2.

35. Dayao, M.T., Trevino, A., Kim, H., Ruffalo, M., D'Angio, H.B., Preska, R., Duvvuri, U., Mayer, A.T., and Bar-Joseph, Z. (2023). Deriving spatial features from in situ proteomics imaging to enhance cancer survival analysis. Bioinformatics *39*, i140–i148. https://doi.org/10.1093/bioinformatics/btad245.

36. Arnol, D., Schapiro, D., Bodenmiller, B., Saez-Rodriguez, J., and Stegle, O. (2019). Modeling cell-cell interactions from spatial molecular data with spatial variance component analysis. Cell Rep. *29*, 202–211.e6. https://doi.org/10.1016/j.celrep.2019.08.077.

37. Peng, H., Wu, X., Liu, S., He, M., Xie, C., Zhong, R., Liu, J., Tang, C., Li, C., Xiong, S., et al. (2023). Multiplex immunofluorescence and single-cell transcriptomic profiling reveal the spatial cell interaction networks in the non-small cell lung cancer microenvironment. Clin. Transl. Med. *13*, e1155. https://doi.org/10.1002/ctm2.1155.

38. Wu, Z., Kondo, A., McGrady, M., Baker, E.A.G., Chidester, B., Wu, E., Rahim, M.K., Bracey, N.A., Charu, V., Cho, R.J., et al. (2024). Discovery and generalization of tissue structures from spatial omics data. Cell Rep. Methods *4*, 100838. https://doi.org/10.1016/j.crmeth.2024.100838.

39. Faryna, K., van der Laak, J., and Litjens, G. (2021). Tailoring automated data augmentation to H&E-stained histopathology. In Proceedings of the Fourth Conference on Medical Imaging with Deep Learning, M. Heinrich, Q. Dou, M. de Bruijne, J. Lellmann, A. Schläfer, and F. Ernst, eds. (PMLR), p. 168.

40. Paszke, A., Gross, S., Massa, F., Lerer, A., Bradbury, J., Chanan, G., Killeen, T., Lin, Z., Gimelshein, N., Antiga, L., et al. (2019). Pytorch: An imperative style, high-performance deep learning library. In Proceedings of the 33rd International Conference on Neural Information Processing Systems, pp. 8026–8037. https://doi.org/10.5555/3454287.3455008.

41. Greenwald, N.F., Miller, G., Moen, E., Kong, A., Kagel, A., Dougherty, T., Fullaway, C.C., McIntosh, B.J., Leow, K.X., Schwartz, M.S., et al. (2022). Whole-cell segmentation of tissue images with human-level performance using large-scale data annotation and deep learning. Nat. Biotechnol. *40*, 555–565. https://doi.org/10.1038/s41587-021-01094-0.

42. Weigert, M., and Schmidt, U. (2022). Nuclei Instance Segmentation and Classification in Histopathology Images with Stardist. The IEEE International Symposium on Biomedical Imaging Challenges (ISBIC). https://doi.org/10.1109/ISBIC56247.2022.9854534.

43. Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., et al. (2011). Scikit-learn: Machine Learning in Python. J. Mach. Learn. Res. *12*, 2825–2830.

44. Virtanen, P., Gommers, R., Oliphant, T.E., Haberland, M., Reddy, T., Cournapeau, D., Burovski, E., Peterson, P., Weckesser, W., Bright, J., et al. (2020). SciPy 1.0: Fundamental Algorithms for Scientific Computing in Python. Nat. Methods *17*, 261–272. https://doi.org/10.1038/s41592-019-0686-2.

45. Jaume, G., Pati, P., Anklin, V., Foncubierta, A., and Gabrani, M. (2021). Histocartography: A toolkit for graph analytics in digital pathology. Proceedings of Machine Learning Research *156*, 117–128.

46. Traag, V.A., Waltman, L., and Van Eck, N.J. (2019). From Louvain to Leiden: guaranteeing well-connected communities. Sci. Rep. *9*, 5233. https://doi.org/10.1038/s41598-019-41695-z.

47. Wu, Z., Trevino, A.E., Wu, E., Swanson, K., Kim, H.J., D'Angio, H.B., Preska, R., Charville, G.W., Dalerba, P.D., Egloff, A.M., et al. (2022). Graph deep learning for the characterization of tumour microenvironments from spatial protein profiles in tissue specimens. Nat. Biomed. Eng. *6*, 1435–1448. https://doi.org/10.1038/s41551-022-00951-w.

48. Schmidt, U., Weigert, M., Broaddus, C., and Myers, G. (2018). Cell Detection with Star-Convex Polygons. Medical Image Computing and Computer Assisted Intervention - MICCAI 2018, 265–273. https://doi.org/10.1007/978-3-030-00934-2_30.

49. Weigert, M., Schmidt, U., Haase, R., Sugawara, K., and Myers, G. (Mar. 2020). Star-convex Polyhedra for 3D Object Detection and Segmentation in Microscopy. The IEEE Winter Conference on Applications of Computer Vision (WACV). https://doi.org/10.1109/WACV45572.2020.9093435.

50. Tellez, D., Balkenhol, M., Karssemeijer, N., Litjens, G., van der Laak, J., and Ciompi, F. (2018). H and E stain augmentation improves generalization of convolutional networks for histopathological mitosis detection. Medical Imaging 2018: Digital Pathology *10581*, 264–270. https://doi.org/10.1117/12.2293048.

51. Tellez, D., Litjens, G., Bándi, P., Bulten, W., Bokhorst, J.-M., Ciompi, F., and Van Der Laak, J. (2019). Quantifying the effects of data augmentation and stain color normalization in convolutional neural networks for computational pathology. Med. Image Anal. *58*, 101544. https://doi.org/10.1016/j.media.2019.101544.

52. Cubuk, E.D., Zoph, B., Shlens, J., and Le, Q.V. (2020). Randaugment: Practical automated data augmentation with a reduced search space. 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW), 702–703. https://doi.org/10.1109/CVPRW50498.2020.00359.

53. Maas, A.L., Hannun, A.Y., and Ng, A.Y. (2013). Rectifier nonlinearities improve neural network acoustic models. Proc. icml. *30*, 3.

54. Kingma, D.P. (2014). Adam: A method for stochastic optimization. Preprint at arXiv. https://doi.org/10.48550/arXiv.1412.6980.

55. Xu, H., Usuyama, N., Bagga, J., Zhang, S., Rao, R., Naumann, T., Wong, C., Gero, Z., González, J., Gu, Y., et al. (2024). A whole-slide foundation model for digital pathology from real-world data. Nature *630*, 181–188. https://doi.org/10.1038/s41586-024-07441-w.

56. Lu, M.Y., Chen, B., Williamson, D.F.K., Chen, R.J., Liang, I., Ding, T., Jaume, G., Odintsov, I., Le, L.P., Gerber, G., et al. (2024). A visual-language foundation model for computational pathology. Nat. Med. *30*, 863–874. https://doi.org/10.1038/s41591-024-02856-4.

57. Vorontsov, E., Bozkurt, A., Casson, A., Shaikovski, G., Zelechowski, M., Severson, K., Zimmermann, E., Hall, J., Tenenholtz, N., Fusi, N., et al. (2024). A foundation model for clinical-grade computational pathology and rare cancers detection. Nat. Med. *30*, 2924–2935. https://doi.org/10.1038/s41591-024-03141-0.

58. Rives, A., Meier, J., Sercu, T., Goyal, S., Lin, Z., Liu, J., Guo, D., Ott, M., Zitnick, C.L., Ma, J., and Fergus, R. (2021). Biological structure and function emerge from scaling unsupervised learning to 250 million protein sequences. Proc. Natl. Acad. Sci. USA *118*, e2016239118. https://doi.org/10.1073/pnas.2016239118.

59. Cui, H., Wang, C., Maan, H., Pang, K., Luo, F., Duan, N., and Wang, B. (2024). scGPT: toward building a foundation model for single-cell multi-omics using generative AI. Nat. Methods *21*, 1470–1480. https://doi.org/10.1038/s41592-024-02201-0.

60. Schaar, A.C., Tejada-Lapuerta, A., Palla, G., Gutgesell, R., Halle, L., Minaeva, M., Vornholz, L., Dony, L., Drummer, F., Bahrami, M., and Theis, F.J. (2024). Nicheformer: a foundation model for single-cell and spatial omics. Preprint at bioRxiv. https://doi.org/10.1101/2024.04.15.589472.

61. Zhou, J., Wei, C., Wang, H., Shen, W., Xie, C., Yuille, A., and Kong, T. (2021). ibot: Image bert pre-training with online tokenizer. Preprint at arXiv. https://doi.org/10.48550/arXiv.2111.07832.

STAR★METHODS

KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Deposited data		
TWMU-TR processed data	This paper	Zenodo: https://doi.org/10.5281/zenodo.16938687
TWMU-TR original data	Hirai et al.[22]	https://app.enablemedicine.com/portal/atlas-library/studies/066f2976-f42d-4e70-b88a-6f97c421659f?sid=314
Stanford-TR processed data	This paper	Zenodo: https://doi.org/10.5281/zenodo.16938814
Stanford-TR original data	Wu et al.[38]	Zenodo: https://doi.org/10.5281/zenodo.12515410
KPMP processed data	This paper	Zenodo: https://doi.org/10.5281/zenodo.16970433
KPMP original data	De Boer et al.[24]	https://atlas.kpmp.org/
Software and algorithms		
Analysis and modeling code	This paper	https://doi.org/10.5281/zenodo.16937845 github.com/mdayao/hist-prot-integration
SCGP	Wu et al.[38]	https://gitlab.com/enable-medicine-public/scgp
H&E RandAugment	Faryna et al.[39]	https://github.com/DIAGNijmegen/pathology-he-auto-augment
PyTorch	Paszke et al.[40]	https://pytorch.org/
Python	N/A	https://www.python.org/
Enable cloud platform	N/A	https://www.enablemedicine.com/
Deepcell	Greenwald et al.[41]	https://www.deepcell.org/
StarDist	Weigert et al.[42]	https://stardist.net/
scikit-learn	Pedregosa et al.[43]	https://scikit-learn.org/stable/index.html
Phikon	Filiot et al.[21]	https://huggingface.co/owkin/phikon
SciPy	Virtanen et al.[44]	https://scipy.org/
NT-Xent loss	Chen et al.[17]	https://dl.acm.org/doi/abs/10.5555/3524938.3525087
HistoCartography	Jaume et al.[45]	https://github.com/BiomedSciAI/histocartography
Leiden clustering	Traag et al.[46]	https://github.com/vtraag/leidenalg

METHOD DETAILS

Description of datasets and pre-processing steps

TWMU-TR

The first dataset used in this work was obtained from 11 T cell-mediated rejection (TCMR) and 12 antibody-mediated rejection (AMR) samples of the kidney, as detailed previously from the Tokyo Women's Medical University (TWMU).[22] We used 17 of the samples (9 TCMR, 8 AMR), after filtering out samples with significant image artifacts and keeping only the samples that included both the 36-plex spatial proteomic (CODEX) image and H&E stains of the same tissue slice. These samples were segmented into 570,828 single cells using DeepCell's Mesmer algorithm[41] Specifically, we used the nuclear segmentation model on the DAPI channel of the of the data, and then dilated to obtain whole cell segmentation, following the same procedure previously done.[47] After cell expression normalization, the cells were grouped into 12 cell types using Leiden clustering.[46] We additionally used Spatial Cellular Graph Partitioning (SCGP)[38] to annotate tissue structures, resulting in 6 tissue structure labels assigned to the cells. 64x64 pixel image patches centered around each cell were saved. Any cells at the edge of the image, resulting in a patch size of less than 64x64, were disregarded. Pixel intensity values were normalized between 0 and 1 before input to the models. We refined and merged annotations using both the Leiden clustering and SCGP label sets, resulting in 11 cell types across 482,975 cells. The dataset was divided into training and validation sets intra-dataset; the training set consisted of 11 samples (289,433 cells), and the validation set consisted of 6 samples (193,542 cells).

Stanford-TR

Cross-dataset generalization was evaluated using another kidney H&E-CODEX dataset, as described previously.[23,38] The dataset consists of samples from patients who underwent allograft nephrectomy. Briefly, a tissue microarray (TMA) was constructed using

2mm cores of cortical tissue. This dataset consists of 63 samples of 51-plex CODEX and H&E stains of the same tissue slice. These samples were segmented into 266,589 cells using DeepCell's Mesmer algorithm,[41] following the same procedure done previously.[47] 64x64 pixel image patches centered around each cell were saved. Any cells at the edge of the image, resulting in a patch size of less than 64x64, were disregarded. Pixel intensity values were normalized between 0 and 1 before input to the models. After cell expression normalization, the cells were grouped into 17 cell types using Leiden clustering.[46] As with the TWMU-TR dataset, we used SCGP[38] for annotation of tissue structures, resulting in 8 tissue structure labels assigned to the cells. We refined and merged annotations using both of these label sets, resulting in 14 cell types across 178,365 cells. The dataset was divided into training and validation sets for the intra-dataset experiments; the training set consisted of 45 samples (110,153 cells), and the validation set consisted of 22 samples (64,250 cells).

Finally, to train and evaluate our models for cross-dataset generalization, we kept only the overlapping cell types from the TWMU-TR and Stanford-TR datasets, resulting in 9 cell types across 451,888 and 112,354 cells, respectively. Figure S1 shows heatmaps of protein expression by cell type for each dataset.

KPMP-AKI and KPMP-CKD

We also used two datasets from the KPMP Kidney Tissue Atlas.[24] Within the atlas, we filtered samples under the "KPMP Main Protocol", "Light Microscopic WSIs", and "H&E stain" tags. The KPMP-AKI dataset consisted of the samples in the "AKI" (AKI) category, and the KPMP-CKD dataset consisted of the samples in the "CKD" (CKD) category. We took one WSI per patient for processing. Each of the WSIs were processed using the following workflow.

(1) Using the `get_tissue_mask` function (with sigma = 20) from the HistoCartography toolkit,[45] we found the bounding boxes for the tissue within the image. In the case of multiple bounding boxes, the one with the largest area was used. If the largest bounding box had an area of <10000 pixels, we disregarded that sample completely. Bounding boxes with aspect ratio greater than 100 or less than 1/100 were disregarded as those bounding boxes captured artifacts at the edges of the slides.
(2) Cells were segmented using the `2d_versatile_he` from StarDist.[42,48,49] If the region had less than 5000 cells, we disregarded the region, and tried again with then next largest bounding box. If after 3 tries, we were not able to find a suitable bounding box, we disregard that sample completely.
(3) 64x64 pixel image patches centered around each cell were saved. Any cells at the edge of the image, resulting in a patch size of less than 64x64, were disregarded. Pixel intensity values were normalized between 0 and 1 before input to the models.

The final KPMP-AKI dataset consisted of 53 samples and 889,680 cells; and the KPMP-CKD dataset consisted of 138 samples and 2,931,898 cells. A list of ids for the WSIs used can be found in Table S1.

Refining cell types with Spatial Cellular Graph Partitioning

We used SCGP[38] to annotate tissue structures in our CODEX datasets, and used these annotations to refine and merge cell types found through Leiden clustering. Below details how these clusters were refined.

TWMU-TR

The 12 cell types labeled after Leiden clustering were as follows: "B cells", "CD4 T cells", "CD8 T cells", "regulatory T cells", "macrophages", "microcapillary cells", "stromal cells", "distal tubules", "proximal tubules", "proximal tubules activated", "proximal tubules atrophic", and "podocytes/lymphatic cells". The 6 structures found by SCGP included "basement membrane/intersitium", "glomeruli ($CD31^+$ $CD34^+$ Podoplanin+)", "immune cells ($CD45^+$)", "tubules (CD183+ PanCK+)", "tubules (PGP9.5+ PanCK+)", and "artifact/edge/undefined".

We merged all of the proximal tubule clusters into a single cell type "proximal tubules", and separated the "microcapillary cells" into two clusters, depending on whether they belonged to the SCGP "glomeruli ($CD31^+$ $CD34^+$ Podoplanin+)" group. These clusters were labeled "microcapillary (glomeruli)" and "microcapillary (other)". The resulting set of 11 cell types was "B cells", "CD4+ T cells", "CD8+ T cells", "Tregs", "distal tubules", "macrophages", "microcapillary (glomeruli)", "microcapillary (other)", "podocytes/lymphatic cells", "proximal tubules", and "stromal cells".

Stanford-TR

The 17 cell types labeled after Leiden clustering were as follows: "B cell", "B cell (PCNA)", "B cell (VISTA+)", "CD4 T cell", "CD8 T cell", "Dendritic cell", "Endothelial cell", "Endothelial cell (CollagenIV)", "Endothelial cell (Podo)", "Endothelial cell (aSMA)", "Immune cell", "Macrophage", "Memory CD4 T cell", "Neutrophil", "T reg", "Tubule", and "Unassigned". The 7 structures found by SCGP included "Basement membrane and low-signal tubules", "Basement membrane with immune activities", "Blood vessels or fibrosis (aSMA+ Caveolin1+)", "Glomeruli ($CD31^+$ $CD34^+$)", "Low signal regions", "Tubules (ECad+ EpCAM+ Keratin+)", "Tubules (Keratin+)", and "Undefined".

We merged the 3 B cell clusters into a single cluster, "B cell". We merged the "CD4 T cell" and "Memory CD4 T cell" into a single cluster. We removed the "Unassigned" cells from the dataset. The cells in the "Tubule" cluster were separated into three groups, "distal tubules", "proximal tubules", and "other tubules", if they were assigned to the SCGP group "Tubules (ECad+ EpCAM+ Keratin+)", "Tubules (Keratin+)", or a different SCGP group, resepectively. Upon further inspect of the overlap of the "T reg" cluster with SCGP labels, we found that it overlapped heavily with the SCGP "Tubules (Keratin+)" group, and merged that cluster with the "proximal tubules" group. The "Endothelial cell (Podo)" cluster was relabeled "podocytes". The "Endothelial cell (aSMA)" cluster was

relabeled "blood vessel" due to its overlap with the SCGP group "Blood vessels or fibrosis (aSMA+ Caveolin1+)". The "Endothelial cell" cluster was relabeled "endothelial cells (glomeruli)" as it overlapped with the SCGP group "Glomeruli ($CD31^+$ $CD34^+$)". The "Endothelial cell (Collagen IV)" cluster was relabeled "endothelial cells (other)".

The resulting set of 14 cell types was "B cell", "CD4 T cell", "CD8 T cell", "Dendritic cell", "Immune cell", "Macrophage", "Neutrophil", "blood vessels", "distal tubules", "endothelial cells (glomeruli)", "endothelial cells (other)", "other tubules", "podocytes", and "proximal tubules".

Overlapping cell types for training and evaluation

For training and evaluation of our models, we used only the overlapping cell types between the two datasets. For the TWMU-TR dataset, we excluded the "Tregs" and "stromal cells" cell types. The "microcapillary (other)" cell type was renamed "blood vessel", and "microcapillary (glomeruli)" was renamed "glomerular endothelial cell". For the Stanford-TR dataset, we excluded the "Dendritic cell", "immune cell", "Neutrophil", "endothelial cells (other)", and "other tubules" cell types. The "endothelial cells (glomeruli)" cell type was renamed "glomerular endothelial cell", and "podocytes" was renamed "podocyte/lymphatic cell".

The final set of cell types were "b cell", "blood vessel", "cd4 T cell", "cd8 T cell", "distal tubules", "glomerular endothelial cell", "macrophage", "podocyte/lymphatic cell", and "proximal tubules".

Removing cell types for cross-dataset evaluation

For our cross-dataset experiments, we removed 3 cell types from the datasets. We removed the "B cell" and "glomerular endothelial cell" cells because they only made up less than 1% of the training data. These cell types also performed the worst in the intra-dataset analysis. We also removed the "blood vessel" cell type from consideration because there is a discrepancy in protein expression for this cell type between the two datasets (Figure S1).

Random data augmentation

Stain variation in histopathology imaging, such as H&E, is extremely common as data acquisition conditions can vary massively among imaging centers. Data augmentation is a widely used technique to generate additional training data in an effort to increase the robustness of neural network models to domain shifts in data, and this has been shown to create more generalizable models for H&E imaging.[39,50,51] In this work, we use the method from Faryna et al.,[39] which we refer to as H&E-RandAugment, to augment training data for our models (multi-modal contrastive model, H&E CNN, see below). RandAugment[52] is an image data augmentation method that randomly applies image transformations to the training data such as rotation, translation, contrast changes, and color adjustments. H&E-RandAugment builds upon this method by adding two histopathology-specific transformations: random shifts in hematoxylin-eosin-DAB (HED)[50] and hue-saturation-value (HSV) color spaces.

The full set of transforms used by the method, along with their corresponding sets of magnitude ranges are detailed in Faryna et al.[39] For each of the parameterized transforms, a linear discretization $m = \{0{:}15\}$ is defined between the minimum and maximum values of the range. n is the number of sequentially applied transforms per sample. Each time a transform is applied, its magnitude M is applied using $M \sim U(0, m_{opt})$. As in Faryna et al.,[39] we use $m_{opt} = 5$ and $n = 3$. We apply H&E-RandAugment to the TWMU-TR and Stanford-TR datasets when training our image-based models.

Multi-modal contrastive model

The multi-modal contrastive model used in this work was inspired by SimCLR and other contrastive learning frameworks.[17–20] As illustrated in Figure 1A, the model comprises the following components.

(1) A random data augmentation module for the H&E images that applies random image transformations the H&E sample $\mathbf{x}_i^{he}$[39], as described above.
(2) Modality-specific neural networks that take the input image patches and transforms them to the same size. The H&E network takes an H&E image patch $\mathbf{x}_i^{he}$ as input and consists of a 2D convolutional layer with 3 input channels and 64 output channels, followed by a LeakyReLU[53] as the activation layer. The CODEX network takes a CODEX image patch $\mathbf{x}_i^{codex}$ as input and consists of a 2D convolutional layer with 52 input channels and 64 output channels, also followed by a LeakyReLU activation layer. Samples $\mathbf{x}_i^{he}$ and $\mathbf{x}_i^{codex}$ are of the same cell i, and we consider this a positive pair. Both of these convolutional layers have `kernel_size = 3`, `stride = 1`, and `padding = 1`. During training, image patches are randomly flipped and rotated and normalized between 0 and 1 before input to these networks.
(3) A CNN-based encoder that extracts representation vectors from the transformed input samples. The CNN consists of 4 convolutional layers, each followed by a LeakyReLU activation function[53] and a max-pooling layer, followed by a single fully connected layer, resulting in an output with size 128.
(4) A small neural network projection head that maps representations to the space where we apply the contrastive loss. We use a 2-layer fully connected network, with a ReLU[53] in between. The output size is 128.

(5) A fully connected neural network layer for cell type prediction. This layer is used in the supervised fine-tuning phase (see below).

Contrastive loss function

We use a contrastive loss function to encourage the embeddings between image patches of different modalities (H&E, CODEX) to be similar if they are of the same cell, and different if they are of different cells. We use the NT-Xent loss (normalized temperatured-scaled cross entropy loss),[17] which is defined for a positive pair of samples (i^{he},i^{codex}) as,

$$\ell_i = -\log \frac{\exp\left(\mathrm{sim}\left(\mathbf{z}_i^{he}, \mathbf{z}_i^{codex}\right)/\tau\right)}{Z} \quad \text{(Equation 1)}$$

$$Z = \sum_{k=1}^{N} \sum_{m \in \{he,codex\}} 1_{[k \neq i] \cup \{[k=i] \cap [m=codex]\}} \exp\left(\mathrm{sim}\left(\mathbf{z}_i^{he}, \mathbf{z}_k^{m}\right)/\tau\right) + \sum_{k=1}^{N} \sum_{m \in \{he,codex\}} 1_{[k \neq i]} \exp\left(\mathrm{sim}\left(\mathbf{z}_i^{codex}, \mathbf{z}_k^{m}\right)/\tau\right) \quad \text{(Equation 2)}$$

where $1_{[k \neq i] \cup \{[k=i] \cap [m=codex]\}} \in \{0,1\}$ is an indicator function that evaluates to 1 if $k \neq i$, or, if $k = 1$ AND m = codex. N is the number of cells represented in the mini-batch ($2N$ image samples), $\mathrm{sim}(\mathbf{u},\mathbf{v}) = \mathbf{u}^T\mathbf{v}/\|\mathbf{u}\|\|\mathbf{v}\|$ is the cosine similarity between $\mathbf{u}$ and $\mathbf{v}$, and τ is a temperature parameter. For this work, we use $\tau = 0.1$. Figure 1C shows a visualization of the contrastive loss function.

Contrastive model training

We trained the multi-modal contrastive model in two phases. First, we performed self-supervised contrastive pre-training on the H&E-CODEX pairs of cell-centered image patches using the contrastive loss function defined above. Positive pairs are image pairs of the same cell. We use the Adam optimizer[54] with a learning rate of 0.0001 and train our models for 100 training epochs in this phase. After the contrastive pre-training phase, we fine-tune our model using a supervised cross-entropy loss to predict cell type. This is implemented by the `nn.NLLLoss` in PyTorch.[40] We compute this loss separately for the CODEX and H&E images, resulting in $\mathcal{L}_{supervised}^{CODEX}$ and $\mathcal{L}_{supervised}^{H\&E}$. We combined both the supervised losses and the contrastive loss, $\mathcal{L}_{contrastive}$, as

$$\mathcal{L} = (1-w)\left(\mathcal{L}_{supervised}^{CODEX} + \mathcal{L}_{supervised}^{H\&E}\right) + w\mathcal{L}_{contrastive} \quad \text{(Equation 3)}$$

for training, with $w = 0.5$. We use the Adam optimizer with learning rate of 0.001 for 125 training epochs in this phase.

For each of our models, we train on the same training datasets for both the contrastive pre-training and supervised fine-tuning phases to ensure that the performance on the validation datasets is not affected by information leakage in the pre-training phase. For the intra-dataset experiments, contrastive pre-training and supervised fine-tuning is only trained on the training sets defined in Section, and evaluation is performed on the defined validation sets. For the cross-dataset experiments, the contrastive pre-training and supervised fine-tuning is trained on the entire TWMU-TR dataset, and evaluation is performed on the full Stanford-TR dataset.

Using foundation model embeddings

In addition to the contrastive loss algorithms, we tested the use of a foundation model for H&E data. Recently, there have been several efforts to build general-purpose, self-supervised vision models for computational pathology trained on massive amounts of histopathology data.[10,21,55–57] Such models are also described as "foundation models" for histopathology, due to their ability to a range of downstream tasks, such as tissue classification, cancer subtyping, protein structure prediction, gene network inference, and more.[57–60] The primary objective of these models is to develop robust and meaningful "off-the-shelf" representations, or embeddings, of pathology image data, most commonly H&E staining. By leveraging large datasets for pre-training from diverse sources, they have been shown to generalize well to a wide array of prediction tasks and can transfer to real-world clinical settings on new datasets. In this work, we used embeddings generated from a pathology foundation model and evaluated their use on the cell type prediction tasks.

As shown in Figure 1B, foundation embeddings were generated using Phikon, a self-supervised learning model for histopathology pre-trained on image tiles from TCGA Program (TCGA).[21] Phikon is based on the image BERT pre-training with Online Tokenizer (iBOT) framework, which combines masked image modeling (MIM) and contrastive learning with a vision transformer architecture.[61] We generated embeddings for the cell-centered H&E image patches for each of our datasets using the publicly available weights for the Phikon model trained on 40 million pan-cancer histology tiles from TCGA. The resulting embedding from the model is a 768-length vector for each image patch.

Cell type assignment using foundation model embeddings

We used multinomial logistic regression to classify cell types from the foundation model embedding vectors. Let $y_i \in 1, \ldots, K$ be the categorical target variable for observation i. Multinomial logistic regression uses a linear predictor function $f(k,i)$ to predict the probability that observation i has on the outcome/class k with the form

$$f(k,i) = W_{0,k} + W_{1,k}x_{1,i} + W_{2,k}x_{2,i} + \cdots + W_{M,k}x_{M,i} \quad \text{(Equation 4)}$$

$$= \mathbf{W}_k \cdot \mathbf{x}_i \quad \text{(Equation 5)}$$

where $\mathbf{W}_k$ is the set of regression coefficients associated with class k, and $\mathbf{x}_i$ is the set of variables associated with observation i (row vector with dimension M+1, where the first element is 1 for the bias variable of $\mathbf{W}_k$).

For the internal-dataset and cross-dataset evaluations, models were trained using the overlapping set of cell types ($K = 9$) between TWMU-TR and Stanford-TR. For prediction on the KPMP-AKI and KPMP-CKD datasets, we trained a logistic regression model on the the full 12 cell types from TWMU-TR. For these results, we merge the three proximal tubule subtypes.

Log-linear formulation of logistic regression

Multinomial linear regression can be formulated this as a log-linear model, where we aim to predict the class probabilities $P(y_i = k|\mathbf{x}_i)$ as,

$$P(y_i = k|\mathbf{x}_i) = \widehat{p}_k(\mathbf{x}_i) = \frac{\exp(\mathbf{W}_k \cdot \mathbf{x}_i)}{\sum_{l=0}^{K-1} \exp(\mathbf{W}_l \cdot \mathbf{x}_i)} \quad \text{(Equation 6)}$$

The optimization objective is,

$$\min_W -\sum_{i=1}^{n}\sum_{k=0}^{K-1} \log(\widehat{p}_k(\mathbf{x}_i)) + r(W) \quad \text{(Equation 7)}$$

where $r(W)$ is a regularization term. In this work, we use ℓ_2 regularization, $r(W) = \frac{1}{2}\|W\|_F^2 = \frac{1}{2}\sum_{i=1}^{m}\sum_{j=1}^{K} W_{ij}^2$.

We trained logistic regression models on the foundation embeddings using the logistic regression classifier from `scikit-learn`.[43]

Comparison models

H&E convolutional neural network

The "H&E CNN" model is a CNN that takes only the H&E cell image patch as input and predicts cell type. Before input to the CNN, each of the RGB channels of the image patch are Z-scaled separately and are randomly rotated and flipped (training only). The model consists of 3 convolutional layers, each followed by a LeakyReLU and max-pooling layer, and 2 fully connected layers with a LeakyReLU in between. It is trained using the cell type labels derived from the CODEX data with a cross entropy loss, implemented by the `nn.NLLLoss` in PyTorch.[40] It is trained using the Adam optimizer with learning rate 0.001 for 300 epochs.

Naive baseline

The naive baseline model randomly predicts cell type for each cell based on the proportions of cell types present in the training set. For example, if 20% of the cells in the training set are labeled cell type A, then the model would label a given test sample with cell type A with 20% probability.

QUANTIFICATION AND STATISTICAL ANALYSIS

Cell expression normalization

We perform the following steps to normalize cell biomarker expression.

(1) Compute the mean expression value across pixels within the cell segmentation mask. Denote the mean expression value of cell i as $x_i^{(j)}$, and denote the array of all expression values $\{x_1^{(j)}, x_2^{(j)}, \ldots\}$ as $X^{(j)}$.

(2) Normalize the expression value using quantile normalization and inverse sine transformation:

$$f\left(x_i^{(j)}\right) = \operatorname{arcsinh}\left(\frac{x_i^{(j)}}{5Q\left(0.2; X^{(j)}\right)}\right) \quad \text{(Equation 8)}$$

where $Q(0.2;X^{(j)})$ represents the 20th quantile of $X^{(j)}$, and arcsinh is the inverse hyperbolic sine function. We can denote the array of normalized expression values for biomarker j as $\{f(x_1^{(j)}), f(x_2^{(j)}), \ldots\}$ as $f(X^{(j)})$.

(1) Calculate the Z score of the normalized expression value:

$$z\left(x_i^{(j)}\right) = \frac{x_i^{(j)} - \mu}{\sigma} \quad \text{(Equation 9)}$$

where μ and σ are the mean and standard deviation of $f(X^{(j)})$, respectively.

These normalized values are used for clustering and assigning cell types.

Statistical analysis on KPMP data

After using the foundation embedding classifier to predict cell types on the KPMP-AKI and KPMP-CKD datasets, we computed the average cell type proportions for each sample. Significant differences were computed using the Python SciPy package[44] with an independent 2-sample t-test using a p-value <0.001.