

---

# STimage-1K4M: A histopathology image-gene expression dataset for spatial transcriptomics

---

Jiawen Chen   Muqing Zhou   Wenrong Wu   Jinwei Zhang   Yun Li\*   Didong Li\*

University of North Carolina at Chapel Hill  
{jiawenn, muqingz, wenrong, jinweizh, yun\_li, didongli}@unc.edu

## Abstract

Recent advances in multi-modal algorithms have driven and been driven by the increasing availability of large image-text datasets, leading to significant strides in various fields, including computational pathology. However, in most existing medical image-text datasets, the text typically provides high-level summaries that may not sufficiently describe sub-tile regions within a large pathology image. For example, an image might cover an extensive tissue area containing cancerous and healthy regions, but the accompanying text might only specify that this image is a cancer slide, lacking the nuanced details needed for in-depth analysis. In this study, we introduce STimage-1K4M, a novel dataset designed to bridge this gap by providing genomic features for sub-tile images. STimage-1K4M contains 1,149 images derived from spatial transcriptomics data, which captures gene expression information at the level of individual spatial spots within a pathology image. Specifically, each image in the dataset is broken down into smaller sub-image tiles, with each tile paired with 15,000 – 30,000 dimensional gene expressions. With 4,293,195 pairs of sub-tile images and gene expressions, STimage-1K4M offers unprecedented granularity, paving the way for a wide range of advanced research in multi-modal data analysis and innovative applications in computational pathology, and beyond.<sup>1</sup>

## 1 Introduction

Multi-modal data, especially image-text pairs, has gained significant importance and popularity (Srinivasan et al., 2021; Schuhmann et al., 2022), driven by the recent success of multi-modal models such as Contrastive Language-Image Pre-Training (CLIP, Radford et al. (2021)). Researchers have been leveraging these models and data across various fields due to their versatility. Initially, many image models were trained to predict a fixed set of predetermined object categories, limiting their generalizability to identify other visual objects or concepts. Learning directly from text descriptions about images provides a complementary and broader source of supervision, expanding the range of potential applications.

Histopathology plays a crucial role in medical diagnostics, focusing on the microscopic examination of tissue samples to detect diseases and guide treatment decisions (Bera et al., 2019). It helps identify cellular abnormalities, including cancerous cells, inflammation, and tissue degeneration (Reddy, 1996; Bera et al., 2019). The collection of image-text pair data for histopathology requires careful annotation of whole-slide images (Huang et al., 2023; Ikezogwo et al., 2024) to create large-scale

---

\*Corresponding authors

<sup>1</sup>The data and code are publicly available at <https://github.com/JiawenChenn/STimage-1K4M> and <https://huggingface.co/datasets/jiawennnn/STimage-1K4M>

datasets suitable for research, training, and diagnostic tool development. Recent efforts to collect and annotate histopathology slides have opened up new opportunities in this domain (Gamper et al., 2019; Graham et al., 2021; Amgad et al., 2019; Huang et al., 2023; Ikezogwo et al., 2024). These annotations vary from simple single labels, such as cell/nuclei types in PanNuke (Gamper et al., 2019) and Lizard datasets (Graham et al., 2021), and cancer regions in NuCLS (Amgad et al., 2019). They also extend to more complex natural language descriptions derived from social media sources such as Twitter or YouTube, as seen in datasets such as OpenPath (Huang et al., 2023) and Quilt-1M (Ikezogwo et al., 2024). Fine-tuning multi-modal models like CLIP with these diverse datasets has shown improved performance in various tasks, including tissue structure classification and image/text retrieval (Huang et al., 2023; Ikezogwo et al., 2024). By combining the capabilities of multi-modal models with detailed histopathology annotations, researchers can achieve greater accuracy and flexibility in medical image analysis. This advancement not only enhances the diagnostic process but also opens the door to new applications in the development of automated pathology tools and, more generally, in personalized medicine (Liu et al., 2019; Nikolov et al., 2021).

While advancements in image annotation have shifted from single labels to natural language descriptions, histopathology slides remain complex and contain a wealth of information that can be challenging to encapsulate in a limited amount of text (Radford et al., 2021; Chen and Zou, 2023). These large tissue slides often feature diverse tissue structures, making it difficult to accurately describe all aspects within a confined-length text. This complexity is further compounded in slides depicting certain diseases, where the focus tends to be on diseased regions, potentially overlooking healthy tissue areas. Randomly cropping images from these slides can lead to misinterpretation and incorrect annotations (Ciga et al., 2021). Histopathology slides are commonly stained with Hematoxylin and Eosin (H&E), revealing details like nuclei and stroma. However, much more biological information exists in these tissue samples, such as gene expression changes and cell-cell communication, which cannot be discerned through staining alone.

Gene expression, the process through which mRNA molecules are generated from the information encoded by the DNA of a gene, is pivotal in studying biological processes. Gene expression data can significantly enhance the annotation of histopathology images. For instance, cancer regions can be identified by the over-expression of specific genes like *ERBB2* in human epidermal growth factor receptor 2 (HER2)-positive breast cancer (Andersson et al., 2021). Moreover, gene expression data can support various downstream analyses, such as deconvolution (Chen et al., 2022, 2023; Luo et al., 2024), which infers the proportion of different cell types in a sample, or clustering (Yuan et al., 2024; Hu et al., 2021; Luecken et al., 2022), which can reveal distinct cell/tissue types/states. The potential applications of gene expression data mark the potential benefit of such paired image and gene expression data.

Gene expression can be measured through several technologies. Bulk RNA-sequencing provides an average expression across large cell populations (Kukurba and Montgomery, 2015). Single-cell RNA sequencing allows for analysis at the individual cell level, enabling more detailed insights into cellular heterogeneity (Kukurba and Montgomery, 2015). However, it still loses the spatial context within the tissue, which is crucial for integrating gene expression data with pathology images to utilize multi-modal methods effectively. To address this need, we highlight spatial transcriptomics (ST) (Ståhl et al., 2016; Moses and Pachter, 2022), a technology that uniquely measures gene expression while preserving spatial information within the tissue (Figure 1a,b). To be more specific, ST can provide gene expression measurement for individual sub-tiles that altogether make up the whole tissue slide. ST has gained significant attention and popularity in recent years due to this unique ability to measure gene expression within spatial context (Ståhl et al., 2016). These ST technologies have revolutionized the way researchers study tissue, allowing for more in-depth analysis of spatial interactions within the tissue and insights into tissue organization and disease mechanisms (Moses and Pachter, 2022; Tian et al., 2023). A key advantage of ST is its ability to provide both high-resolution histopathology images and detailed whole-transcriptome data for each spatial coordinate within a large tissue image (Ståhl et al., 2016). This makes ST a perfect source for paired medical image and text datasets, offering a richer, more accurate annotation that addresses the limitations of over-simplified textual descriptions that typically focus solely on broad categories like cancer or non-cancer regions. By providing high-dimensional annotations for each sub-tile, ST enables a more comprehensive understanding of tissue granularity, facilitating studies of cell-cell communication, tissue architecture, and disease progression (Ståhl et al., 2016; Tian et al., 2023).

Despite these advantages, existing datasets that combine pathology images with gene expression data are often limited in size and scope (Fan et al., 2020; Xu et al., 2022; Fan et al., 2023; Yuan et al., 2023; Zheng et al., 2023; Moses and Pachter, 2022; Zhou et al., 2024; Li et al., 2022), preventing the full utilization of advanced multi-modal models. To bridge this gap, we curated the STImage-1K4M dataset (Figure 1), a comprehensive collection of 1,149 ST slides from three leading ST technologies that provide histopathology images: Spatial Transcriptomics (Stahl et al., 2016) (note that we use the full name to indicate this particular technology, and ST for general ST technologies), Visium, and VisiumHD. These slides were further subdivided into smaller sub-tiles, resulting in a total of 4,293,195 images, each with corresponding high-dimensional gene expression data. This extensive dataset spans 10 different species and includes 50 distinct tissue types. STImage-1K4M represents a significant advancement in multi-modal datasets for computational pathology and related fields. By providing a large and diverse collection of image sub-tiles paired with detailed gene expression data, the dataset offers researchers a unique resource for exploring the spatial organization of tissues and understanding the complex relationships between cellular structures, gene activity, and disease and health related outcomes. This dataset paves the way for advanced research in multi-modal analysis and innovative applications in computational pathology and personalized medicine.

## 2 Related work

This work builds on existing research in vision-language datasets for histopathology, spatial omics datasets, and representation learning in medical image analysis.

**Vision-Language Pairs in Histopathology.** Multiple histopathology image-text pair datasets have emerged, serving as a foundational resource for studying medical images. The ARCH dataset consists of 8,617 figure-caption pairs with histology or immunohistochemistry (IHC) images, curated from research publications (Gamper and Rajpoot, 2021). The OpenPath dataset offers a broader perspective, featuring 116,504 image-text pairs from Twitter posts across 32 pathology subspecialties, along with 59,869 image-text pairs from replies to popular tweets, and 32,041 additional image-text pairs scraped from the LAION dataset (Huang et al., 2023; Schuhmann et al., 2022). Quilt-1M, a combination of Quilt with datasets from other sources, represents one of the largest vision-language histopathology datasets to date, with over 1 million image-text samples (Ikezogwo et al., 2024). These datasets prove to be valuable resources for training and evaluating models that can understand and correlate textual information with histopathology images.

**Spatial Omics Datasets.** The rise of ST and spatial omics data has spurred the development of various datasets that focus on transcriptomics or other omics data in tissue samples. Notable databases include SpatialDB (Fan et al., 2020), STOmicsDB (Xu et al., 2022), SPASCER (Fan et al., 2023), SODB (Yuan et al., 2023), Aquila (Zheng et al., 2023), Museum of Spatial Transcriptomics (Moses and Pachter, 2022), SORC Zhou et al. (2024), and SOAR (Li et al., 2022). These datasets focus primarily on gene expression data, providing researchers with a wealth of information about the spatial distribution of gene expression in tissue samples. However, there is currently a lack of datasets that provide paired image and gene expression data, which is crucial for bridging the gap between visual information and underlying transcriptomic profiles.

**Representation Learning in Medical Imaging.** Representation learning has made significant strides in medical imaging. Early models focused on predicting single values such as gene expression (He et al., 2020a) or survival outcome (Chen et al., 2021), while more recent approaches employ self-supervised learning (SSL) techniques to learn from unlabeled image data (Ikezogwo et al., 2022). Contrastive SSL models including PLIP (Huang et al., 2023), Quilt-Net (Ikezogwo et al., 2024) and CONCH (Lu et al., 2024), which use image and label annotation, have gained popularity, with models successfully trained on image-text pairs. However, text encoders are limited by token length, making it challenging to incorporate gene expression data. In the ST field, researchers have explored contrastive SSL for image-gene expression data or other modalities like gene expression paired with protein abundance (Zeng et al., 2023; Long et al., 2023; Yao et al., 2024). These models are typically trained on a single slide, constrained by the lack of large datasets that pair histopathology images with gene expression data, and the challenge of integrating gene expression across different datasets.

### 3 Curating STimage-1K4M: Overview

ST technologies can be broadly categorized into two main types: sequencing-based and imaging-based. Sequencing-based ST technology typically involves capturing spatial information using unique barcodes that correspond to specific regions which are usually called "spots" within a tissue sample (see Figure 1a middle panel and Figure 1b for example). This approach enables researchers to capture the entire transcriptome while retaining the spatial context through the barcodes. Imaging-based ST technology, on the other hand, uses fluorescence or other imaging techniques to visualize gene expression directly in the tissue context, and can reach cellular and even sub-cellular resolution. However, imaging-based ST technology has a limitation: the number of genes it can measure is restricted, due to the complexity of multiple rounds of fluorescence of many genes. To be more specific, sequencing-based technologies like Spatial Transcriptomics (Ståhl et al., 2016), Visium and VisiumHD (10x Genomics) can measure  $\sim 15\text{k}-30\text{k}$  genes (Figure 1b) while imaging-based technologies like MERFISH (Chen et al., 2015) and STARmap (Wang et al., 2018) can only measure hundreds of genes (median number of genes around 300 in the SOAR database (Li et al., 2022)).

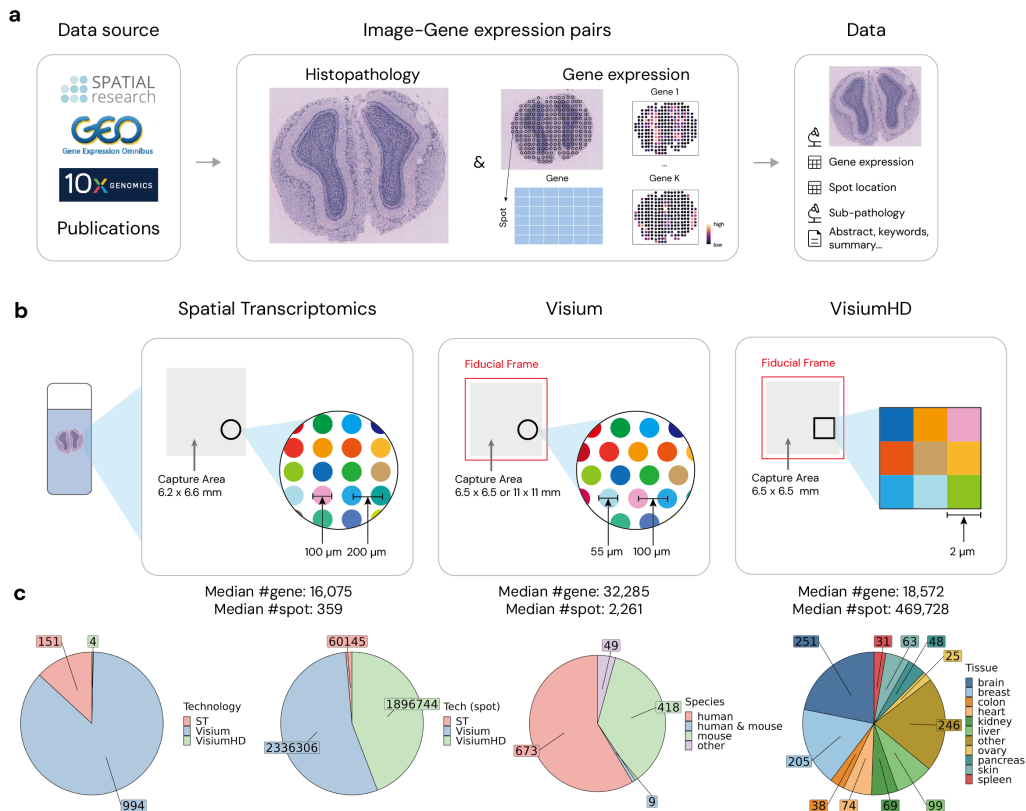


Figure 1: Overview of STimage-1K4M. (a) Curation overview. (b) ST technologies resolution. (c) Breakdown of technologies, species and tissue types in STimage-1K4M.

To obtain comprehensive gene expression information across the entire transcriptome, we focus on sequencing-based ST technologies, specifically those that also offer histopathology images: Spatial Transcriptomics, Visium, and VisiumHD (Figure 1b). Spatial Transcriptomics Ståhl et al. (2016) is one of the first sequencing-based ST methods that allows spatial mapping of gene expression in tissue sections. This technology uses unique barcodes to capture gene expression at specific spots in a grid pattern, with a spot diameter of 100  $\mu\text{m}$  and a center-to-center distance of 200  $\mu\text{m}$  (Figure 1b left panel). Visium evolves from Spatial Transcriptomics, offering improved resolution and higher throughput with a spot diameter of 55  $\mu\text{m}$  and a refined center-to-center distance of 100

$\mu\text{m}$  (Figure 1b middle panel). VisiumHD further extends the capabilities of Visium by offering even finer resolution and data density. This advanced technology introduces a continuous measurement system, utilizing a  $2\mu\text{m} \times 2\mu\text{m}$  grid of bins for gene expression analysis (Figure 1b right panel). In our work, we employed an  $8\mu\text{m} \times 8\mu\text{m}$  bin structure, as instructed by 10X Genomics, achieved by aggregating the gene expression data from smaller  $2\mu\text{m} \times 2\mu\text{m}$  bins to match the desired resolution.

Public available sources for ST data include Gene Expression Omnibus (GEO), 10X Genomics datasets, Spatial Research datasets, and various publications. We queried the GEO website using keywords "spatial transcriptomics", specifically targeting supplementary files in JPG, PNG, or TIFF formats. This search resulted in 856 datasets from 121 unique GEO studies. Additionally, we gathered 58 Visium and 4 VisiumHD datasets from 10X Genomics, complementing these with 233 slides manually collected from 10 additional studies (see Appendix A for a full list of references).

A significant challenge in this process was the inconsistent sharing standards for ST data, particularly for the image components. Many datasets lack corresponding images, making it difficult to analyze the gene expression data in its proper spatial context. For Visium data, the standard format typically includes at least one image, which can be of full-resolution, high-resolution, and low-resolution. In this work, we used the highest resolution images available for each dataset. Spatial Transcriptomics data posed additional hurdles. This kind of data requires CytAssist images to map the coordinates to the image, but these images are rarely publicly available, making it challenging to link gene expression data to histopathology images. Only datasets with mapped and unmapped coordinates could be included in the study for the calculation of spot diameter following ST pipeline in the SpatialTranscriptomicsResearch GitHub repository. Given the various sharing formats and the common absence of key data, it's particularly challenging for researchers unfamiliar with ST to align gene expression data with histopathology images. To address this, we manually processed and verified every dataset to ensure accurate coordinate mapping, allowing precise linking of gene expression data to histopathology images. Furthermore, we calculated and included the corresponding spot radius to indicate the area of measurement. These manual efforts underscore our commitment to providing a reliable and comprehensive dataset, facilitating easier integration of ST data with histopathology images for researchers across various disciplines.

In summary, we systematically collected a diverse collection of 1,149 ST slides, encompassing 4,293,195 spots with paired gene expression information. For each dataset, we provide histopathology images, spot center coordinates and radius, as well as the associated gene expression data. Our STImage-1K4M dataset comprises of data from Spatial Transcriptomics, Visium, and VisiumHD platforms. At the slide level, STImage-1K4M has 13.1% from Spatial Transcriptomics, 86.5% from Visium, and 0.3% from VisiumHD. At the spot level, due to the resolution difference, the composition shifts to 1.4% from Spatial Transcriptomics, 54.4% from Visium, and 44.2% from VisiumHD. STImage-1K4M predominantly includes data from human and mouse, encompasses 50 tissues with the largest proportion of images from brain, accounting for 21.8% (251 slides), followed by the breast tissue at 17.8% (205 slides). Given a major focus on cancer in the field of ST, it's noteworthy that 39.7% of the slides (456 slides) are from studies related to cancer.

In addition to the paired image and gene expression data, we also included pathologist annotations for the slides (Figure 6). Spatial domain detection or clustering is a popular topic in ST data analysis. However, due to the lack of organized datasets, evaluations in most ST clustering methods utilizing image data rely on limited samples (Andersson et al., 2021; Maynard et al., 2021). We manually reviewed relevant publications and extracted annotations from 9 studies including 71 slides to enrich our STImage-1K4M dataset. These pathologist annotations are anticipated to substantially reduce efforts required for collecting labeled data with "ground truth" in the ST field and to provide researchers with a more comprehensive resource for evaluating clustering methods and dimension reduction techniques.

As a comprehensive and meticulously curated dataset, STImage-1K4M aims to facilitate research in ST, computational pathology, and related fields. This dataset can significantly streamline the data collection process, allowing researchers to focus on developing innovative methods and gaining deeper insights into tissue structure and gene expression patterns.

## 4 Popular tasks using ST images

Within the field of traditional computational biology with no gene expression involved, commonly performed tasks such as tissue type classification and image-text retrieval have well-established solutions. However, ST introduces new complexity and opportunity with additional gene expression information. ST data allows researchers to engage in a variety of specialized tasks that are particularly suited to the strengths of this new type of technology.

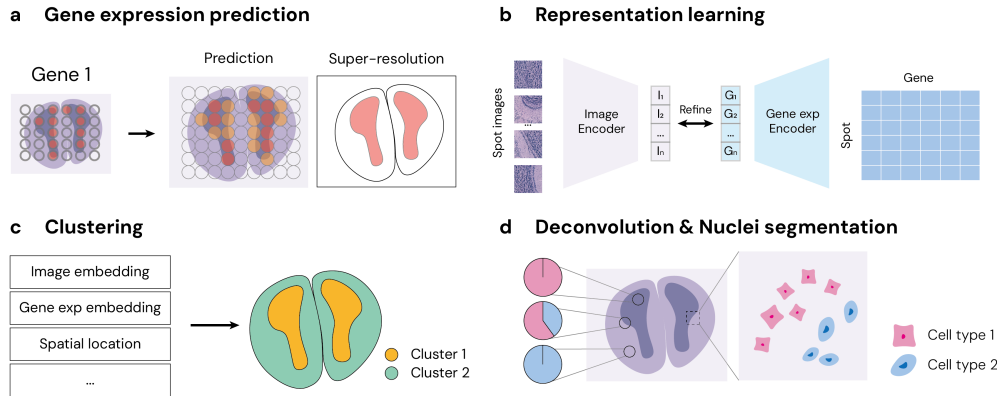


Figure 2: Popular tasks in ST data analysis.

**Gene Expression Prediction and Resolution Enhancement.** One key usage of images in ST is predicting gene expression (Figure 2a) from histopathology images (Xie et al., 2024; He et al., 2020a). This approach allows researchers to infer gene expression levels from visual data, potentially reducing the need for expensive and time-consuming library preparation and sequencing. Additionally, increasing the resolution of gene expression data through high-quality imaging techniques offers a more detailed understanding of spatial patterns within tissue samples, leading to improved accuracy in analyzing gene expression spatial distributions (Hu et al., 2023; Zhang et al., 2024).

**Representation Learning and Clustering.** Similar as in image-based computational biology, learning image embeddings is also a popular task in ST (Figure 2b). This process involves transforming high-resolution tissue images into compact, informative representations that capture the essential features of the underlying biological processes. A key application of these embeddings is spatial clustering (Figure 2c), where similar tissue regions are grouped based on shared characteristics captured in the embeddings (Hu et al., 2021). Clustering allows researchers to explore tissue heterogeneity and identify distinct spatial clusters that may correspond to different cellular functions or disease states.

**Deconvolution and Cell Segmentation.** Deconvolution and cell segmentation are valuable computational methods that enhance our understanding of tissue composition at cellular level (Figure 2d). Deconvolution specifically focuses on deciphering mixed signals within spot-level gene expression data to accurately estimate the proportions of contributing cell types present in a tissue sample. Histology images are particularly valuable in this context because common staining methods inherently highlight nuclei information, providing a clear visual representation for cellular structures (Biancalani et al., 2021; Chen et al., 2023). This visual clarity allows deconvolution via computational methods, as spots that appear similar in the images are likely to have similar cell type compositions. Additionally, the integration of image analysis with deconvolution facilitates the application of trained models to new images or to areas within images where spots were not initially measured, potentially increasing analysis resolution. Furthermore, by employing cell segmentation techniques alongside these images, researchers can precisely identify and categorize individual nucleus, which allows accurate assignment of specific cell types to these identified nucleus, thereby enriching the gene expression data with detailed cellular annotations (Biancalani et al., 2021; Zhang et al., 2024).

## 5 Experiment training with STImage-1K4M

The STImage-1K4M dataset presents unprecedented opportunities for benchmarking, discovery, and training new models. In this section, we provide several examples illustrating its potential uses.

## 5.1 Gene expression prediction

We initially utilize the STImage-1K4M dataset to benchmark the task of gene expression prediction (Figure 2a). With each spot image paired with high-dimensional gene expression data, STImage-1K4M serves as an ideal resource for both learning gene expression and benchmarking large vision models. We assess the performance of image encoders from the pre-trained CLIP, PLIP, and UNI models using the HER2+ human breast cancer dataset. *ERBB2*, a marker gene for HER2+ breast cancer, was selected for this analysis (Figure 3a,b). We trained a simple regression model to predict *ERBB2* gene expression using the image embeddings from each model. The PLIP model demonstrated superior performance in this context, exhibiting higher correlation and lower mean squared error (MSE) (Figure 3).

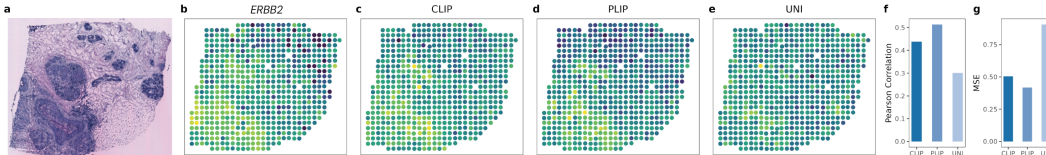


Figure 3: Gene expression prediction evaluation. (a) H2 histology images. (b) *ERBB2* gene expression. (c) Predicted *ERBB2* expression using zero-shot CLIP image embedding. (d) Predicted *ERBB2* expression using zero-shot PLIP image embedding. (e) Predicted *ERBB2* expression using zero-shot UNI image embedding. (f) Pearson Correlation of three models. (g) MSE of three models.

## 5.2 Representation learning and clustering

To demonstrate the effectiveness of our STImage-1K4M dataset, we employed contrastive learning to fine-tune the image encoders of pre-trained CLIP and PLIP models using STImage-1K4M, to enhance the models’ performance in integrating pathology images with corresponding gene expressions. To effectively incorporate gene expressions, we replaced the text encoder in these models with fully connected neural networks, as shown in Figure 2b and Appendix B. The objective of our contrastive learning remains consistent with the original CLIP framework, aiming to increase the cosine similarity between embeddings of aligned pairs while minimizing similarity for unaligned pairs.

Given the challenges of different genes measured across datasets and prevailing batch effects, we limited our analyses to samples from Maynard et al. (2021), which includes 12 human dorsolateral prefrontal cortex (DLPFC) slides encompassing 47,681 spots. To manage the high dimensionality of gene expression data, we explored two strategies: highly variable genes (HVG) selected separately from each slide, and HVGs selected from overlapping genes across slides (overlap-HVG). Once fine-tuned, we conducted experiments for image classification using linear probing and analyzed the image embeddings through t-Distributed Stochastic Neighbor Embedding (t-SNE) (Van der Maaten and Hinton, 2008). See Appendix for experiment details.

**Evaluation using linear probing.** We evaluated the performance of the fine-tuned models via linear probing. As shown in Figure 4a, the fine-tuned CLIP and PLIP with HVG achieved higher mean F1 scores than zero-shot CLIP and PLIP models, indicating that fine-tuning on our STImage-1K4M improves the performance. While we did not fine-tune the larger UNI model due to computational constraints, the results suggest that both fine-tuning with our dataset and using a more effective pre-trained model contribute to better performance. We conjecture that fine-tuning UNI on our STImage-1K4M could further enhance its performance, combining the benefits of both advanced model architecture and tailored training data.

**Image representation learning.** To evaluate the enhancement in image representations achieved by the fine-tuned models, we utilized pathologist-annotated brain layers (Figure 4c) as benchmarks to calculate several cluster quality metrics (Figure 4b), including the Silhouette score (Rousseeuw, 1987), the Calinski-Harabasz index (Caliński and Harabasz, 1974), and the Davies-Bouldin index (Davies and Bouldin, 1979). Additionally, we applied t-SNE (Van der Maaten and Hinton, 2008) for visualization to further analyze the clustering patterns (Figure 4d). Our findings indicate that, compared to zero-shot embeddings, the fine-tuned embeddings more effectively distinguish between various tissue subtypes, notably between white matter (WM) and other layers (L1–L6) in the brain

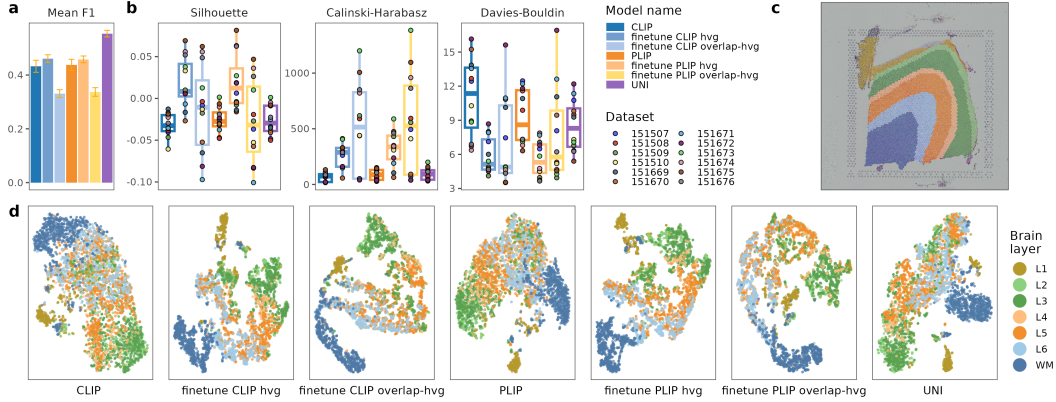


Figure 4: Evaluation results. (a) Linear probing results, denoted by average macro F1 (error bars indicate standard deviations). (b) Silhouette, Calinski-Harabasz and Davies-Bouldin scores for image embeddings. (c) Histopathology image of brain sample 151675 colored by pathologist annotation. (d) t-SNE embeddings of sample 151675, colored by the same layer annotations as in (c).

(Figure 4b,d). In particular, the image embeddings from the fine-tuned models outperform all zero-shot image embeddings. This enhancement suggests that incorporating gene expression data into the training process helps the model capture more nuanced differences within the tissue slides, which highlights the potential of integrating genetic and image information to learn more precise and informative interpretations of tissue structure and function.

### 5.3 Deconvolution and cell segmentation

In ST, a conventional approach for the nuclei segmentation and deconvolution task involves spot-level data. This includes gene expression, spatial coordinates, and image data, often supplemented by an external scRNA-seq reference to estimate cell type proportions for each spot. These inferred cell type proportions could serve as a benchmark, offering an alternative to traditional nuclei and cell type annotations which typically require extensive manual effort. The STImage-1K4M dataset allows for the benchmarking of cell segmentation methods against these inferred cell type proportions.

We evaluated the performance of the CellViT-256 and CellViT-SAM (Hörst et al., 2024) models in segmenting Neoplastic cells using HER2+ breast cancer data, comparing these results with the cancer epithelial cell proportions inferred by a computational deconvolution method RCTD (Cable et al., 2022). Visual inspection supports the models' comparable effectiveness in cell segmentation. Notably, the CellViT-SAM model tends to classify more cells as Neoplastic. We employed Pearson correlation to measure the agreement between the segmented and inferred cell type proportions, with the CellViT-256 model demonstrating a higher correlation (0.37 compared to 0.34 for the CellViT-SAM) (Figure 5).

## 6 Discussion

In this work, we introduced STImage-1K4M, a groundbreaking open-source dataset that pairs histopathology images with gene expression data. Our empirical results demonstrate the effectiveness of pre-training using STImage-1K4M, which has shown to outperform larger state-of-the-art models such as CLIP and PLIP. This success highlights the significant potential of integrating image and gene expression data to enhance model performance and provide new opportunities for advancing research in ST and computational pathology. For further reading on the innovative computational techniques being explored in this field, readers may refer to Szałata et al. (2024); Coleman et al. (2024). Despite these promising results, this emerging field also presents significant challenges that require innovative approaches to overcome. Next, we discuss the potentials and challenges associated with this integration.



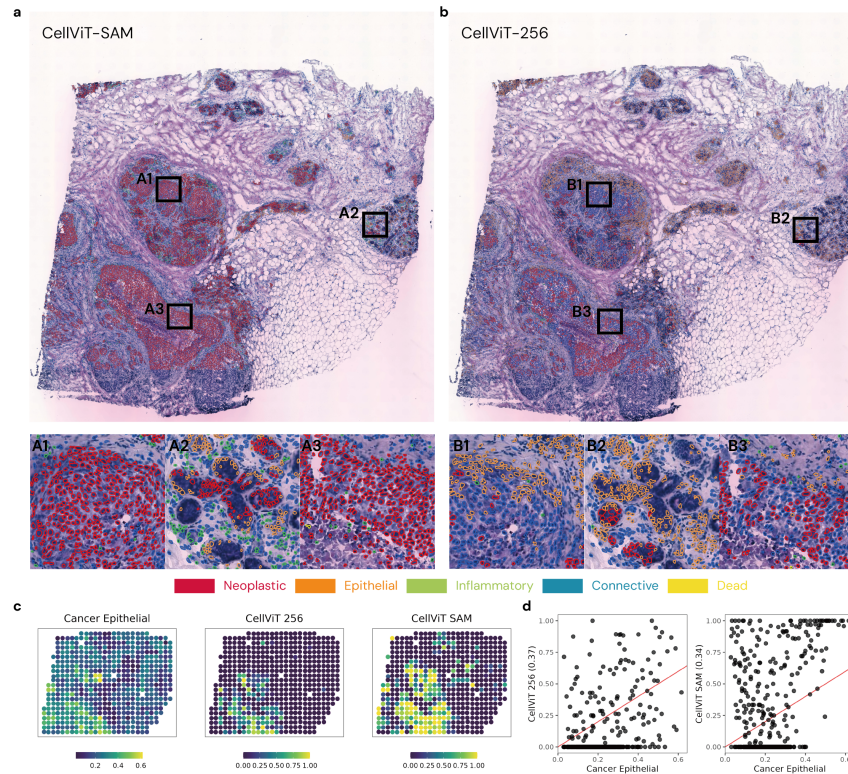


Figure 5: Cell segmentation evaluation. (a) CellViT-SAM model segmented cells. (b) CellViT-256 model segmented cells. (c) left: RCTD inferred cancer epithelial cell proportion in spot level, middle: aggregated CellViT-256 segmented neoplastic cell proportion. Right: aggregated CellViT-SAM segmented neoplastic cell proportion. (d) scatter plot of RCTD inferred cell type proportion and aggregated segmented cell proportion. Ylab: model name (Pearson correlation).

**High-dimensional image.** A typical histopathology image consists of three primary color channels — red, green, and blue (RGB). These channels represent the standard visualization used in most imaging technologies to capture the visual structure and patterns in tissue samples. When histopathology images are paired with gene expression data, the data dimension increases enormously, presenting both challenges and opportunities for analysis. If gene expression data is treated as a separate set of "channels", where each gene's expression level is represented as a gray-scale image channel, the entire histopathology image transforms into a high-dimensional data structure. Instead of having just three RGB channels, the transformed image would now have around  $\sim 20,000$  channels, each representing the expression of a different gene. This high-dimensionality adds molecular information to the visual data, offering insights far beyond what can be revealed by staining methods.

Given this expanded data structure, a crucial question arises: How can this high-dimensional data be effectively analyzed and utilized? One of the central challenges is to strike an optimal balance between sample size and resolution. When focusing on spot-level images, there's a risk of losing spatial connections between the spots. On the other hand, slide-level information provides a broader context but at the cost of reduced sample size, which could limit the scope of analysis. This challenge leads to further questions: How can slide-level information improve the image embeddings for spot-level images? Can the data be augmented by pairing it with datasets containing image-text or purely image-based information?

For spot-level data, several approaches have been attempted, including contrastive SSL (Long et al., 2023; Yao et al., 2024), but these approaches typically concentrate on spot-level images from a single slide, limiting their generalizability. Determining the optimal approach for analyzing multi-dimensional datasets remains an open question. Should researchers employ contrastive SSL, where models learn from paired image-gene expression data, or treat the dataset as a multi-channel image, where traditional image-processing techniques can be applied? These questions are central

to the ongoing evolution of computational pathology, as they determine the effectiveness of latent embedding extraction and ultimately influence models' performance in real-world applications.

**Position encoding.** In traditional vision transformer models (Dosovitskiy et al., 2020), positional encoding is used to provide context about the relative or absolute positions of input image patches within a sequence. This is crucial because transformers, unlike convolutional neural networks, do not inherently retain information about the order or spatial arrangement of their inputs. Positional encoding typically involves adding a set of coordinates or numerical values to the model's inputs, enabling the model to understand spatial relationships and preserve structure during analysis. In the context of integrating histopathology images with gene expression data, gene expression data could potentially serve as a unique form of positional encoding. By linking specific regions within an image to their corresponding transcriptomic information, researchers can create spatially-aware models that can learn from both visual and transcriptomic cues.

**Gene expression annotation.** Gene expression data has become an indispensable resource in the annotation of complex biological datasets. It has been widely used for various downstream analysis including clustering, which classify cells/spots into distinct groups based on their gene expression profiles, and deconvolution. Recent advancements include integrating large language models (LLMs) to extract meaningful gene expression embeddings (Chen et al., 2023; Schaar et al., 2024), which utilize gene names and text descriptions to enhance data interpretation. However, several significant challenges remain. Variation in genome structures across different species complicate cross-species analysis, and batch effects introduce systematic biases in gene expression measurements. Additionally, using gene names with rankings (Chen and Zou, 2023; Schaar et al., 2024) lacks the precision of quantitative values, and high-dimensionality necessitates effective dimension reduction techniques. Current methods, such as using PCs or HVGs, often fall short in multi-slides analysis across different tissue and species.

Our STimage-1K4M dataset has the potential to address these challenges by providing a large, diverse collection of paired histopathology images and gene expression data across multiple species and tissue types. This dataset may facilitate the development of robust annotation methods that manage high-dimensional data and mitigate batch effects. Additionally, integrating STimage-1K4M with other large-scale datasets, such as those containing only images or only gene expression data (Schaar et al., 2024), could further enhance multi-modal learning and improve our understanding of complex biological processes and disease mechanisms.

**Limitations.** In this work, although we have shown that the integration of gene expression data has enhanced the performance of the pre-trained CLIP and PLIP image encoders, the fine-tuned models are still inferior to the UNI model, suggesting that employing a more powerful foundational model could potentially yield further improvements. However, due to limited computational resources, we were unable to fine-tune the UNI model. Additionally, we only utilized a maximum of 128 dimensions, compressed into a 32-dimensional latent layer, to incorporate gene expression data. This simplistic implementation may not fully capture the complexity and richness of gene expression information. Efforts to fine-tune the models using data from other tissue types, resulted in sub-optimal performance. This suggests that batch effects across datasets introduce noise and variability, significantly impacting model performance. For instance, the tSNE embedding of samples from 12 human brain slides in the Maynard et al. study illustrates significant batch effects. In this visualization, points are clustered by data source rather than by biological layers, clearly highlighting the presence of batch effects (Figure 10). Furthermore, our study has not utilized the slide-level abstract and experimental design text information available in our dataset, which represents another area for potential exploration and application.

**Data collection and societal biases.** STimage-1K4M may exhibit inherent biases due to existing ST publications focusing disproportionately focus on brain tissues and breast cancer samples. Such ascertainment biases introduced by published ST work can influence the outcomes of analyses conducted using our STimage-1K4M dataset, potentially leading to models that are better attuned to recognizing patterns specific to these tissues.

## Acknowledgments and Disclosure of Funding

This work is supported by R01 AG079291, R01 MH125236, U01 HG011720, P50 HD103573, R56 LM013784, R01 HL149683, and UM1 TR004406. We thank Dr. Hongtu Zhu for the computational resource support.

## References

- Akiyama, T., T. Yasuda, T. Uchihara, N. Yasuda-Yoshihara, B. J. Y. Tan, A. Yonemura, T. Semba, J. Yamasaki, Y. Komohara, K. Ohnishi, F. Wei, L. Fu, J. Zhang, F. Kitamura, K. Yamashita, K. Eto, S. Iwagami, H. Tsukamoto, T. Umemoto, M. Masuda, O. Nagano, Y. Satou, H. Saya, P. Tan, H. Baba, and T. Ishimoto (2023, March). Stromal reprogramming through dual PDGFR $\alpha/\beta$  blockade boosts the efficacy of anti-PD-1 immunotherapy in fibrotic tumors. *Cancer Res.* 83(5), 753–770.
- Amgad, M., H. Elfandy, H. Hussein, L. A. Atteya, M. A. Elsebaie, L. S. Abo Elnasr, R. A. Sakr, H. S. Salem, A. F. Ismail, A. M. Saad, et al. (2019). Structured crowdsourcing enables convolutional segmentation of histology images. *Bioinformatics* 35(18), 3461–3467.
- Andersson, A., L. Larsson, L. Stenbeck, F. Salmén, A. Ehinger, S. Z. Wu, G. Al-Eryani, D. Roden, A. Swarbrick, Å. Borg, J. Frisén, C. Engblom, and J. Lundeberg (2021, October). Spatial deconvolution of HER2-positive breast cancer delineates tumor-associated cell type interactions. *Nat. Commun.* 12(1), 6012.
- Andrews, T. S., D. Nakib, C. T. Perciani, X. Z. Ma, L. Liu, E. Winter, D. Camat, S. W. Chung, P. Lumanto, J. Manuel, S. Mangroo, B. Hansen, B. Arpinder, C. Thoeni, B. Sayed, J. Feld, A. Gehring, A. Gulamhusein, G. M. Hirschfield, A. Ricciuto, G. D. Bader, I. D. McGilvray, and S. MacParland (2024, January). Single-cell, single-nucleus, and spatial transcriptomics characterization of the immunological landscape in the healthy and PSC human liver. *J. Hepatol.*
- Arora, R., C. Cao, M. Kumar, S. Sinha, A. Chanda, R. McNeil, D. Samuel, R. K. Arora, T. W. Matthews, S. Chandarana, R. Hart, J. C. Dort, J. Biernaskie, P. Neri, M. D. Hycza, and P. Bose (2023, August). Spatial transcriptomics reveals distinct and conserved tumor core and edge architectures that predict survival and targeted therapy response. *Nat. Commun.* 14(1), 5029.
- Asp, M., S. Giacomello, L. Larsson, C. Wu, D. Fürth, X. Qian, E. Wärdell, J. Custodio, J. Reimegård, F. Salmén, C. Österholm, P. L. Ståhl, E. Sundström, E. Åkesson, O. Bergmann, M. Bienko, A. Månsson-Broberg, M. Nilsson, C. Sylvén, and J. Lundeberg (2019, December). A spatiotemporal organ-wide gene expression and cell atlas of the developing human heart. *Cell* 179(7), 1647–1660.e19.
- Ballester Roig, M. N., T. Leduc, J. Dufort-Gervais, Y. Maghmoul, O. Tastet, and V. Mongrain (2023, May). Probing pathways by which rhynchophylline modifies sleep using spatial transcriptomics. *Biol. Direct* 18(1), 21.
- Barkley, D., R. Moncada, M. Pour, D. A. Liberman, I. Dryg, G. Werba, W. Wang, M. Baron, A. Rao, B. Xia, G. S. França, A. Weil, D. F. Delair, C. Hajdu, A. W. Lund, I. Osman, and I. Yanai (2022, August). Cancer cell states recur across tumor types and form specific interactions with the tumor microenvironment. *Nat. Genet.* 54(8), 1192–1201.
- Bashkirova, E. V., N. Klimpert, K. Monahan, C. E. Campbell, J. M. Osinski, L. Tan, I. Schieren, A. Pourmorady, B. Stecky, G. Barnea, X. S. Xie, I. Abdus-Saboor, B. Shykind, B. Jones-Marlin, R. M. Gronostajski, A. Fleischmann, and S. Lomvardas (2023, September). Opposing, spatially-determined epigenetic forces impose restrictions on stochastic olfactory receptor choice. *bioRxiv*.
- Bassiouni, R., M. O. Idowu, L. D. Gibbs, V. Robila, P. J. Grizzard, M. G. Webb, J. Song, A. Noriega, D. W. Craig, and J. D. Carpten (2023, January). Spatial transcriptomic analysis of a diverse patient cohort reveals a conserved architecture in triple-negative breast cancer. *Cancer Res.* 83(1), 34–48.

- Bera, K., K. A. Schalper, D. L. Rimm, V. Velcheti, and A. Madabhushi (2019). Artificial intelligence in digital pathology—new tools for diagnosis and precision oncology. *Nature reviews Clinical oncology* 16(11), 703–715.
- Biancalani, T., G. Scalia, L. Buffoni, R. Avasthi, Z. Lu, A. Sanger, N. Tokcan, C. R. Vanderburg, Å. Segerstolpe, M. Zhang, et al. (2021). Deep learning and alignment of spatially resolved single-cell transcriptomes with tangram. *Nature methods* 18(11), 1352–1362.
- Cable, D. M., E. Murray, L. S. Zou, A. Goeva, E. Z. Macosko, F. Chen, and R. A. Irizarry (2022). Robust decomposition of cell type mixtures in spatial transcriptomics. *Nature biotechnology* 40(4), 517–526.
- Caetano, A. J., Y. Redhead, F. Karim, P. Dhimi, S. Kannambath, R. Nuamah, A. A. Volponi, L. Nibali, V. Booth, E. M. D’Agostino, and P. T. Sharpe (2023, January). Spatially resolved transcriptomics reveals pro-inflammatory fibroblast involved in lymphocyte recruitment through CXCL8 and CXCL10. *Elife* 12.
- Caliński, T. and J. Harabasz (1974). A dendrite method for cluster analysis. *Communications in Statistics-theory and Methods* 3(1), 1–27.
- Canela, V. H., W. S. Bowen, R. M. Ferreira, F. Syed, J. E. Lingeman, A. R. Sabo, D. Barwinska, S. Winfree, B. B. Lake, Y.-H. Cheng, J. P. Gaut, M. Ferkowicz, K. A. LaFavers, K. Zhang, F. L. Coe, E. Worcester, Kidney Precision Medicine Project, S. Jain, M. T. Eadon, J. C. Williams, Jr, and T. M. El-Achkar (2023, July). A spatially anchored transcriptomic atlas of the human kidney papilla identifies significant immune injury in patients with stone disease. *Nat. Commun.* 14(1), 4140.
- Caronni, N., F. La Terza, F. M. Vittoria, G. Barbiera, L. Mezzananza, V. Cuzzola, S. Barresi, M. Pellegatta, P. Canevazzi, G. Dunsmore, C. Leonardi, E. Montaldo, E. Lusito, E. Dugnani, A. Citro, M. S. F. Ng, M. Schiavo Lena, D. Drago, A. Andolfo, S. Brugiapaglia, A. Scagliotti, A. Mortellaro, V. Corbo, Z. Liu, A. Mondino, P. Dellabona, L. Piemonti, C. Taveggia, C. Doglioni, P. Cappello, F. Novelli, M. Iannacone, L. G. Ng, F. Ginhoux, S. Crippa, M. Falconi, C. Bonini, L. Naldini, M. Genua, and R. Ostuni (2023, November). IL-1 $\beta$ + macrophages fuel pathogenic inflammation in pancreatic cancer. *Nature* 623(7986), 415–422.
- Castranio, E. L., P. Hasel, J.-V. Haure-Mirande, A. V. Ramirez Jimenez, B. W. Hamilton, R. D. Kim, C. G. Glabe, M. Wang, B. Zhang, S. Gandy, S. A. Liddelow, and M. E. Ehrlich (2023, June). Microglial INPP5D limits plaque formation and glial reactivity in the PSAPP mouse model of alzheimer’s disease. *Alzheimers. Dement.* 19(6), 2239–2252.
- Chaker, Z., C. Segalada, J. A. Kretz, I. E. Acar, A. C. Delgado, V. Crotet, A. E. Moor, and F. Doetsch (2023, November). Pregnancy-responsive pools of adult neural stem cells for transient neurogenesis in mothers. *Science* 382(6673), 958–963.
- Chen, J., Y. Katada, K. Okimura, T. Yamaguchi, Y.-J. Guh, T. Nakayama, M. Maruyama, Y. Furukawa, Y. Nakane, N. Yamamoto, Y. Sato, H. Ando, A. Sugimura, K. Tabata, A. Sato, and T. Yoshimura (2022, November). Prostaglandin E2 synchronizes lunar-regulated beach spawning in grass puffers. *Curr. Biol.* 32(22), 4881–4889.e5.
- Chen, J., W. Liu, T. Luo, Z. Yu, M. Jiang, J. Wen, G. P. Gupta, P. Giusti, H. Zhu, Y. Yang, et al. (2022). A comprehensive comparison on cell-type composition inference for spatial transcriptomics data. *Briefings in Bioinformatics* 23(4), bbac245.
- Chen, J., T. Luo, M. Jiang, J. Liu, G. P. Gupta, and Y. Li (2023). Cell composition inference and identification of layer-specific spatial transcriptional profiles with polaris. *Science Advances* 9(9), eadd9818.
- Chen, K. H., A. N. Boettiger, J. R. Moffitt, S. Wang, and X. Zhuang (2015). Spatially resolved, highly multiplexed rna profiling in single cells. *Science* 348(6233), aaa6090.
- Chen, K. S., M. H. Noureldein, L. M. McGinley, J. M. Hayes, D. M. Rigan, J. F. Kwentus, S. N. Mason, F. E. Mendelson, M. G. Savelieff, and E. L. Feldman (2023, December). Human neural stem cells restore spatial memory in a transgenic alzheimer’s disease mouse model by an immunomodulating mechanism. *Front. Aging Neurosci.* 15, 1306004.

- Chen, L., D. Chang, B. Tandukar, D. Deivendran, J. Pozniak, N. Cruz-Pacheco, R. J. Cho, J. Cheng, I. Yeh, C. Marine, B. C. Bastian, A. L. Ji, and A. H. Shain (2023, November). STmut: a framework for visualizing somatic alterations in spatial transcriptomics data of cancer. *Genome Biol.* 24(1), 273.
- Chen, R. J., M. Y. Lu, W.-H. Weng, T. Y. Chen, D. F. Williamson, T. Manz, M. Shady, and F. Mahmood (2021). Multimodal co-attention transformer for survival prediction in gigapixel whole slide images. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pp. 4015–4025.
- Chen, Y. T. and J. Zou (2023). Genept: A simple but hard-to-beat foundation model for genes and cells built from chatgpt. *bioRxiv*.
- Cherief, M., J. Xu, Z. Li, R. J. Tower, S. Ramesh, Q. Qin, M. Gomez-Salazar, J.-H. Yea, S. Lee, S. Negri, M. Xu, T. Price, A. R. Kendal, C.-M. Fan, T. L. Clemens, B. Levi, and A. W. James (2023, December). TrkA-mediated sensory innervation of injured mouse tendon supports tendon sheath progenitor cell expansion and tendon repair. *Sci. Transl. Med.* 15(727), eade4619.
- Chitturi, P., S. Xu, B. Ahmed Abdi, J. Nguyen, D. E. Carter, S. Sinha, R. Arora, J. Biernaskie, R. J. Stratton, and A. Leask (2023, September). Tripterygium wilfordii derivative celastrol, a YAP inhibitor, has antifibrotic effects in systemic sclerosis. *Ann. Rheum. Dis.* 82(9), 1191–1204.
- Ciga, O., T. Xu, S. Nofech-Mozes, S. Noy, F.-I. Lu, and A. L. Martel (2021). Overcoming the limitations of patch-based learning to detect cancer in whole slide images. *Scientific Reports* 11(1), 8894.
- Coleman, K., A. Schroeder, and M. Li (2024). Unlocking the power of spatial omics with ai. *Nature Methods* 21(8), 1378–1381.
- Cortese, N., R. Carriero, M. Barbagallo, A. R. Putignano, G. Costa, F. Giavazzi, F. Grizzi, F. Pasqualini, C. Peano, G. Basso, S. Marchini, F. S. Colombo, C. Soldani, B. Franceschini, L. Di Tommaso, L. Terracciano, M. Donadon, G. Torzilli, P. Kunderfranco, A. Mantovani, and F. Marchesi (2023, April). High-resolution analysis of mononuclear phagocytes reveals GPNMB as a prognostic marker in human colorectal liver metastasis. *Cancer Immunol. Res.* 11(4), 405–420.
- Coutant, A., V. Cockenpot, L. Muller, C. Degletagne, R. Pommier, L. Tonon, M. Ardin, M.-C. Michallet, C. Caux, M. Laurent, A.-P. Morel, P. Saintigny, A. Puisieux, M. Ouzounova, and P. Martinez (2023, December). Spatial transcriptomics reveal pitfalls and opportunities for the detection of rare high-plasticity breast cancer subtypes. *Lab. Invest.* 103(12), 100258.
- Davies, D. L. and D. W. Bouldin (1979). A cluster separation measure. *IEEE transactions on pattern analysis and machine intelligence* (2), 224–227.
- Deshpande, A., M. Loth, D. N. Sidiropoulos, S. Zhang, L. Yuan, A. T. F. Bell, Q. Zhu, W. J. Ho, C. Santa-Maria, D. M. Gilkes, S. R. Williams, C. R. Uytingco, J. Chew, A. Hartnett, Z. W. Bent, A. V. Favorov, A. S. Popel, M. Yarchoan, A. Kiemen, P.-H. Wu, K. Fujikura, D. Wirtz, L. D. Wood, L. Zheng, E. M. Jaffee, R. A. Anders, L. Danilova, G. Stein-O’Brien, L. T. Kagohara, and E. J. Fertig (2023a, April). Uncovering the spatial landscape of molecular interactions within the tumor microenvironment through latent spaces. *Cell Syst.* 14(4), 285–301.e4.
- Deshpande, A., M. Loth, D. N. Sidiropoulos, S. Zhang, L. Yuan, A. T. F. Bell, Q. Zhu, W. J. Ho, C. Santa-Maria, D. M. Gilkes, S. R. Williams, C. R. Uytingco, J. Chew, A. Hartnett, Z. W. Bent, A. V. Favorov, A. S. Popel, M. Yarchoan, A. Kiemen, P.-H. Wu, K. Fujikura, D. Wirtz, L. D. Wood, L. Zheng, E. M. Jaffee, R. A. Anders, L. Danilova, G. Stein-O’Brien, L. T. Kagohara, and E. J. Fertig (2023b, August). Uncovering the spatial landscape of molecular interactions within the tumor microenvironment through latent spaces. *Cell Syst.* 14(8), 722.
- Dhainaut, M., S. A. Rose, G. Akturk, A. Wroblewska, S. R. Nielsen, E. S. Park, M. Backup, V. Roudko, L. Pia, R. Sweeney, J. Le Berichel, C. M. Wilk, A. Bektesevic, B. H. Lee, N. Bhardwaj, A. H. Rahman, A. Baccarini, S. Gnjjatic, D. Pe’er, M. Merad, and B. D. Brown (2022, March). Spatial CRISPR genomics identifies regulators of the tumor microenvironment. *Cell* 185(7), 1223–1239.e20.

- Di Marco, B., J. Vázquez-Marín, H. Monyer, L. Centanin, and J. Alfonso (2023, October). Spatial transcriptomics map of the embryonic mouse brain - a tool to explore neurogenesis. *Biol. Open* 12(10).
- Dixon, E. E., H. Wu, Y. Muto, P. C. Wilson, and B. D. Humphreys (2022, February). Spatially resolved transcriptomic analysis of acute kidney injury in a female murine model. *J. Am. Soc. Nephrol.* 33(2), 279–289.
- Dosovitskiy, A., L. Beyer, A. Kolesnikov, D. Weissenborn, X. Zhai, T. Unterthiner, M. Dehghani, M. Minderer, G. Heigold, S. Gelly, et al. (2020). An image is worth 16x16 words: Transformers for image recognition at scale. *arXiv preprint arXiv:2010.11929*.
- Erickson, A., M. He, E. Berglund, M. Marklund, R. Mirzazadeh, N. Schultz, L. Kvastad, A. Andersson, L. Bergenstråhle, J. Bergenstråhle, L. Larsson, L. Alonso Galicia, A. Shamikh, E. Basmaci, T. Díaz De Ståhl, T. Rajakumar, D. Doultinos, K. Thrane, A. L. Ji, P. A. Khavari, F. Tarish, A. Tanoglidis, J. Maaskola, R. Colling, T. Mirtti, F. C. Hamdy, D. J. Woodcock, T. Helleday, I. G. Mills, A. D. Lamb, and J. Lundeberg (2022, August). Spatially resolved clonal copy number alterations in benign and malignant tissue. *Nature* 608(7922), 360–367.
- Eum, H. H., D. Jeong, N. Kim, A. Jo, M. Na, H. Kang, Y. Hong, J.-S. Kong, G. H. Jeong, S.-A. Yoo, and H.-O. Lee (2024, February). Single-cell RNA sequencing reveals myeloid and T cell co-stimulation mediated by IL-7 anti-cancer immunotherapy. *Br. J. Cancer*.
- Fan, Z., R. Chen, and X. Chen (2020). Spatialdb: a database for spatially resolved transcriptomes. *Nucleic acids research* 48(D1), D233–D237.
- Fan, Z., Y. Luo, H. Lu, T. Wang, Y. Feng, W. Zhao, P. Kim, and X. Zhou (2023). Spascer: spatial transcriptomics annotation at single-cell resolution. *Nucleic Acids Research* 51(D1), D1138–D1149.
- Filipescu, D., S. Carcamo, A. Agarwal, N. Tung, É. Humblin, M. S. Goldberg, N. S. Vyas, K. G. Beaumont, D. Demircioglu, S. Sridhar, F. G. Ghiraldini, C. Capparelli, A. E. Aplin, H. Salmon, R. Sebra, A. O. Kamphorst, M. Merad, D. Hasson, and E. Bernstein (2023, September). MacroH2A restricts inflammatory gene expression in melanoma cancer-associated fibroblasts by coordinating chromatin looping. *Nat. Cell Biol.* 25(9), 1332–1345.
- Foster, D. S., M. Januszyk, D. Delitto, K. E. Yost, M. Griffin, J. Guo, N. Guardino, A. E. Delitto, M. Chinta, A. R. Burcham, A. T. Nguyen, K. E. Bauer-Rowe, A. L. Titan, A. Salhotra, R. E. Jones, O. da Silva, H. G. Lindsay, C. E. Berry, K. Chen, D. Henn, S. Mascharak, H. E. Talbot, A. Kim, F. Nosrati, D. Sivaraj, R. C. Ransom, M. Matthews, A. Khan, D. Wagh, J. Collier, G. C. Gurtner, D. C. Wan, I. L. Wapnir, H. Y. Chang, J. A. Norton, and M. T. Longaker (2022, November). Multiomic analysis reveals conservation of cancer-associated fibroblast phenotypes across species and tissue of origin. *Cancer Cell* 40(11), 1392–1406.e7.
- Foster, D. S., M. Januszyk, K. E. Yost, M. S. Chinta, G. S. Gulati, A. T. Nguyen, A. R. Burcham, A. Salhotra, R. C. Ransom, D. Henn, K. Chen, S. Mascharak, K. Tolentino, A. L. Titan, R. E. Jones, O. da Silva, W. T. Leavitt, C. D. Marshall, H. E. des Jardins-Park, M. S. Hu, D. C. Wan, G. Wernig, D. Wagh, J. Collier, J. A. Norton, G. C. Gurtner, A. M. Newman, H. Y. Chang, and M. T. Longaker (2021, October). Integrated spatial multiomics reveals fibroblast fate during tissue repair. *Proc. Natl. Acad. Sci. U. S. A.* 118(41), e2110025118.
- Gamper, J., N. Alemi Koohbanani, K. Benet, A. Khuram, and N. Rajpoot (2019). Pannuke: an open pan-cancer histology dataset for nuclei instance segmentation and classification. In *Digital Pathology: 15th European Congress, ECDP 2019, Warwick, UK, April 10–13, 2019, Proceedings 15*, pp. 11–19. Springer.
- Gamper, J. and N. Rajpoot (2021). Multiple instance captioning: Learning representations from histopathology textbooks and articles. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pp. 16549–16559.
- Garbarino, O., L. Lambroia, G. Basso, V. Marrella, B. Franceschini, C. Soldani, F. Pasqualini, D. Giuliano, G. Costa, C. Peano, D. Barbarossa, D. Annarita, A. Salvati, L. Terracciano, G. Torzilli, M. Donadon, and F. Faggioli (2023, July). Spatial resolution of cellular senescence dynamics in human colorectal liver metastasis. *Aging Cell* 22(7), e13853.

- Gharaie, S., K. Lee, K. Noller, E. K. Lo, B. Miller, H. J. Jung, A. M. Newman-Rivera, J. T. Kurzhagen, N. Singla, P. A. Welling, J. Fan, P. Cahan, S. Noel, and H. Rabb (2023, November). Single cell and spatial transcriptomics analysis of kidney double negative T lymphocytes in normal and ischemic mouse kidneys. *Sci. Rep.* 13(1), 20888.
- Graham, S., M. Jahanifar, A. Azam, M. Nimir, Y.-W. Tsang, K. Dodd, E. Hero, H. Sahota, A. Tank, K. Benes, et al. (2021). Lizard: a large-scale dataset for colonic nuclear instance segmentation and classification. In *Proceedings of the IEEE/CVF international conference on computer vision*, pp. 684–693.
- Gu, Y., Y. Zhou, S. Ju, X. Liu, Z. Zhang, J. Guo, J. Gao, J. Zang, H. Sun, Q. Chen, J. Wang, J. Xu, Y. Xu, Y. Chen, Y. Guo, J. Dai, H. Ma, C. Wang, G. Jin, C. Li, Y. Xia, H. Shen, Y. Yang, X. Guo, and Z. Hu (2022, December). Multi-omics profiling visualizes dynamics of cardiac development and functions. *Cell Rep.* 41(13), 111891.
- Guilliams, M., J. Bonnardel, B. Haest, B. Vanderborght, C. Wagner, A. Remmerie, A. Bujko, L. Martens, T. Thoné, R. Browaeys, F. F. De Ponti, B. Vanneste, C. Zwicker, F. R. Svedberg, T. Vanhalewyn, A. Gonçalves, S. Lippens, B. Devriendt, E. Cox, G. Ferrero, V. Wittamer, A. Willaert, S. J. F. Kaptein, J. Neyts, K. Dallmeier, P. Geldhof, S. Casaert, B. Deplancke, P. Ten Dijke, A. Hoorens, A. Vanlander, F. Berrevoet, Y. Van Nieuwenhove, Y. Saeys, W. Saelens, H. Van Vlierberghe, L. Devisscher, and C. L. Scott (2022, January). Spatial proteogenomics reveals distinct and evolutionarily conserved hepatic macrophage niches. *Cell* 185(2), 379–396.e38.
- Habenicht, L. K. L., Z. Wang, X. Zhang, Y. Li, C. Mogler, J. S. Huspenina, R. M. Schmid, C. Weber, S. K. Mohanta, Z. Ma, and C. Yin (2022, October). The C1q-ApoE complex: A new hallmark pathology of viral hepatitis and nonalcoholic fatty liver disease. *Front. Immunol.* 13, 970938.
- Hasel, P., I. V. L. Rose, J. S. Sadick, R. D. Kim, and S. A. Liddelow (2021, October). Neuroinflammatory astrocyte subtypes in the mouse brain. *Nat. Neurosci.* 24(10), 1475–1487.
- He, B., L. Bergenstråhle, L. Stenbeck, A. Abid, A. Andersson, Å. Borg, J. Maaskola, J. Lundeberg, and J. Zou (2020a). Integrating spatial gene expression and breast tumour morphology via deep learning. *Nature biomedical engineering* 4(8), 827–834.
- He, B., L. Bergenstråhle, L. Stenbeck, A. Abid, A. Andersson, Å. Borg, J. Maaskola, J. Lundeberg, and J. Zou (2020b, August). Integrating spatial gene expression and breast tumour morphology via deep learning. *Nat. Biomed. Eng.* 4(8), 827–834.
- Heezen, L. G. M., T. Abdelaal, M. van Putten, A. Aartsma-Rus, A. Mahfouz, and P. Spitali (2023, August). Spatial transcriptomics reveal markers of histopathological changes in duchenne muscular dystrophy mouse models. *Nat. Commun.* 14(1), 4909.
- Heimli, M., S. T. Flåm, H. S. Hjorthaug, D. Trinh, M. Frisk, K.-A. Dumont, T. Ribarska, X. Tekpli, M. Saare, and B. A. Lie (2022). Multimodal human thymic profiling reveals trajectories and cellular milieu for T agonist selection. *Front. Immunol.* 13, 1092028.
- Hörst, F., M. Rempe, L. Heine, C. Seibold, J. Keyl, G. Baldini, S. Ugurel, J. Siveke, B. Grünwald, J. Egger, et al. (2024). Cellvit: Vision transformers for precise cell segmentation and classification. *Medical Image Analysis* 94, 103143.
- Hu, J., K. Coleman, D. Zhang, E. B. Lee, H. Kadara, L. Wang, and M. Li (2023). Deciphering tumor ecosystems at super resolution from spatial transcriptomics with tesla. *Cell systems* 14(5), 404–417.
- Hu, J., X. Li, K. Coleman, A. Schroeder, N. Ma, D. J. Irwin, E. B. Lee, R. T. Shinohara, and M. Li (2021). Spagcn: Integrating gene expression, spatial location and histology to identify spatial domains and spatially variable genes by graph convolutional network. *Nature methods* 18(11), 1342–1351.
- Huang, Z., F. Bianchi, M. Yuksekogonul, T. J. Montine, and J. Zou (2023). A visual–language foundation model for pathology image analysis using medical twitter. *Nature medicine* 29(9), 2307–2316.

- Hudson, W. H. and L. J. Sudmeier (2022, June). Localization of T cell clonotypes using the visium spatial transcriptomics platform. *STAR Protoc.* 3(2), 101391.
- Huuki-Myers, L., A. Spangler, N. Eagles, K. D. Montgomery, S. H. Kwon, B. Guo, M. Grant-Peters, H. R. Divecha, M. Tippani, C. Sriworarat, A. B. Nguyen, P. Ravichandran, M. N. Tran, A. Seyedian, PsychENCODE consortium, T. M. Hyde, J. E. Kleinman, A. Battle, S. C. Page, M. Ryten, S. C. Hicks, K. Martinowich, L. Collado-Torres, and K. R. Maynard (2023, February). Integrated single cell and unsupervised spatial transcriptomic analysis defines molecular anatomy of the human dorsolateral prefrontal cortex. *bioRxiv.org*.
- Ikezogwo, W., S. Seyfioglu, F. Ghezloo, D. Geva, F. Sheikh Mohammed, P. K. Anand, R. Krishna, and L. Shapiro (2024). Quilt-1m: One million image-text pairs for histopathology. *Advances in Neural Information Processing Systems* 36.
- Ikezogwo, W. O., M. S. Seyfioglu, and L. Shapiro (2022). Multi-modal masked autoencoders learn compositional histopathological representations. *arXiv preprint arXiv:2209.01534*.
- Ji, A. L., A. J. Rubin, K. Thrane, S. Jiang, D. L. Reynolds, R. M. Meyers, M. G. Guo, B. M. George, A. Mollbrink, J. Bergenstr hle, L. Larsson, Y. Bai, B. Zhu, A. Bhaduri, J. M. Meyers, X. Rovira-Clav , S. T. Hollmig, S. Z. Aasi, G. P. Nolan, J. Lundeberg, and P. A. Khavari (2020, July). Multimodal analysis of composition and spatial architecture in human squamous cell carcinoma. *Cell* 182(2), 497–514.e22.
- Joglekar, A., A. Prjibelski, A. Mahfouz, P. Collier, S. Lin, A. K. Schlusche, J. Marrocco, S. R. Williams, B. Haase, A. Hayes, J. G. Chew, N. I. Weisenfeld, M. Y. Wong, A. N. Stein, S. A. Hardwick, T. Hunt, Q. Wang, C. Dieterich, Z. Bent, O. Fedrigo, S. A. Sloan, D. Risso, E. D. Jarvis, P. Flicek, W. Luo, G. S. Pitt, A. Frankish, A. B. Smit, M. E. Ross, and H. U. Tilgner (2021, January). A spatially resolved brain region- and cell type-specific isoform atlas of the postnatal mouse brain. *Nat. Commun.* 12(1), 463.
- Kadur Lakshminarasimha Murthy, P., V. Sontake, A. Tata, Y. Kobayashi, L. Macadlo, K. Okuda, A. S. Conchola, S. Nakano, S. Gregory, L. A. Miller, J. R. Spence, J. F. Engelhardt, R. C. Boucher, J. R. Rock, S. H. Randell, and P. R. Tata (2022, April). Human distal lung maps and lineage hierarchies reveal a bipotent progenitor. *Nature* 604(7904), 111–119.
- Kasmani, M. Y., P. Topchyan, A. K. Brown, R. J. Brown, X. Wu, Y. Chen, A. Khatun, D. Alson, Y. Wu, R. Burns, C.-W. Lin, M. R. Kudek, J. Sun, and W. Cui (2023, October). A spatial sequencing atlas of age-induced changes in the lung during influenza infection. *Nat. Commun.* 14(1), 6597.
- Kawai, K., A. Sakamoto, M. Mokry, S. K. B. Ghosh, R. Kawakami, W. Xu, L. Guo, D. T. Fuller, T. Tanaka, P. Shah, A. Cornelissen, Y. Sato, M. Mori, T. Konishi, A. E. Vozenilek, R. Dhingra, R. Virmani, G. Pasterkamp, and A. V. Finn (2023, December). Clonal proliferation within smooth muscle cells in unstable human atherosclerotic lesions. *Arterioscler. Thromb. Vasc. Biol.* 43(12), 2333–2347.
- Kenney, H. M., J. Rangel-Moreno, Y. Peng, K. L. Chen, J. Bruno, A. Embong, E. Pritchett, J. I. Fox, E. Becerril-Villanueva, A. Gamboa-Dom nguez, S. Quataert, G. Muthukrishnan, R. W. Wood, B. D. Korman, J. H. Anolik, L. Xing, C. T. Ritchlin, E. M. Schwarz, and C.-L. Wu (2023, August). Multi-omics analysis identifies IgG2b class-switching with ALCAM-CD6 co-stimulation in joint-draining lymph nodes during advanced inflammatory-erosive arthritis. *Front. Immunol.* 14, 1237498.
- Kerzel, T., G. Giacca, S. Beretta, C. Bresesti, M. Notaro, G. M. Scotti, C. Balestrieri, T. Canu, M. Redegalli, F. Pedica, M. Genua, R. Ostuni, A. Kajaste-Rudnitski, M. Oshima, G. Tonon, I. Merelli, L. Aldrighetti, P. Dellabona, N. Coltella, C. Doglioni, P. M. V. Rancoita, F. Sanvito, L. Naldini, and M. L. Squadrito (2023, November). In vivo macrophage engineering reshapes the tumor microenvironment leading to eradication of liver metastases. *Cancer Cell* 41(11), 1892–1910.e10.
- Kukurba, K. R. and S. B. Montgomery (2015). Rna sequencing and analysis. *Cold Spring Harbor Protocols* 2015(11), pdb-top084970.



- Kuppe, C., R. O. Ramirez Flores, Z. Li, S. Hayat, R. T. Levinson, X. Liao, M. T. Hannani, J. Tanevski, F. Wünnemann, J. S. Nagai, M. Halder, D. Schumacher, S. Menzel, G. Schäfer, K. Hoefl, M. Cheng, S. Ziegler, X. Zhang, F. Peisker, N. Kaesler, T. Saritas, Y. Xu, A. Kassner, J. Gummert, M. Morshuis, J. Amrute, R. J. A. Veltrop, P. Boor, K. Klingel, L. W. Van Laake, A. Vink, R. M. Hoogenboezem, E. M. J. Bindels, L. Schurgers, S. Sattler, D. Schapiro, R. K. Schneider, K. Lavine, H. Milting, I. G. Costa, J. Saez-Rodriguez, and R. Kramann (2022, August). Spatial multi-omic map of human myocardial infarction. *Nature* 608(7924), 766–777.
- Lake, B. B., R. Menon, S. Winfree, Q. Hu, R. Melo Ferreira, K. Kalhor, D. Barwinska, E. A. Otto, M. Ferkowicz, D. Diep, N. Plongthongkum, A. Knoten, S. Urata, L. H. Mariani, A. S. Naik, S. Eddy, B. Zhang, Y. Wu, D. Salamon, J. C. Williams, X. Wang, K. S. Balderrama, P. J. Hoover, E. Murray, J. L. Marshall, T. Noel, A. Vijayan, A. Hartman, F. Chen, S. S. Waikar, S. E. Rosas, F. P. Wilson, P. M. Palevsky, K. Kiryluk, J. R. Sedor, R. D. Toto, C. R. Parikh, E. H. Kim, R. Satija, A. Greka, E. Z. Macosko, P. V. Kharchenko, J. P. Gaut, J. B. Hodgkin, KPMP Consortium, M. T. Eadon, P. C. Dagher, T. M. El-Achkar, K. Zhang, M. Kretzler, and S. Jain (2023, July). An atlas of healthy and injured cell states and niches in the human kidney. *Nature* 619(7970), 585–594.
- Lause, J., P. Berens, and D. Kobak (2021). Analytic pearson residuals for normalization of single-cell rna-seq umi data. *Genome biology* 22, 1–20.
- Lebrigand, K., J. Bergenstråhle, K. Thrane, A. Mollbrink, K. Meletis, P. Barbry, R. Waldmann, and J. Lundeberg (2023, May). The spatial landscape of gene expression isoforms in tissue sections. *Nucleic Acids Res.* 51(8), e47.
- Lee, S., N. A. Devanney, L. R. Golden, C. T. Smith, J. L. Schwartz, A. E. Walsh, H. A. Clarke, D. S. Goulding, E. J. Allenger, G. Morillo-Segovia, C. M. Friday, A. A. Gorman, T. R. Hawkinson, S. M. MacLean, H. C. Williams, R. C. Sun, J. M. Morganti, and L. A. Johnson (2023, March). APOE modulates microglial immunometabolism in response to age, amyloid pathology, and inflammatory challenge. *Cell Rep.* 42(3), 112196.
- Lequain, H., C. Déglétagne, N. Streichenberger, J. Valantin, T. Simonet, L. Schaeffer, P. Sève, and P. Leblanc (2023, November). Spatial transcriptomics reveals signatures of histopathological changes in muscular sarcoidosis. *Cells* 12(23).
- Li, Y., S. Dennis, M. R. Hutch, Y. Li, M. S. Broad, Z. Zeng, and Y. Luo (2022). Soar: a spatial transcriptomics analysis resource to model spatial variability and cell type interactions. *bioRxiv*, 2022–04.
- Liu, S., H. Li, Y. Shen, W. Zhu, Y. Wang, J. Wang, N. Zhang, C. Li, L. Xie, and Q. Wu (2023, February). Moxibustion improves hypothalamus aqp4 polarization in APP/PS1 mice: Evidence from spatial transcriptomics. *Front. Aging Neurosci.* 15, 1069155.
- Liu, Y., T. Kohlberger, M. Norouzi, G. E. Dahl, J. L. Smith, A. Mohtashamian, N. Olson, L. H. Peng, J. D. Hipp, and M. C. Stumpe (2019). Artificial intelligence–based breast cancer nodal metastasis detection: insights into the black box for pathologists. *Archives of pathology & laboratory medicine* 143(7), 859–868.
- Long, Y., K. S. Ang, M. Li, K. L. K. Chong, R. Sethi, C. Zhong, H. Xu, Z. Ong, K. Sachaphibulkij, A. Chen, et al. (2023). Spatially informed clustering, integration, and deconvolution of spatial transcriptomics with graphst. *Nature Communications* 14(1), 1155.
- Lowe, M. M., J. N. Cohen, M. I. Moss, S. Clancy, J. P. Adler, A. E. Yates, H. B. Naik, R. Yadav, M. Pauli, I. Taylor, A. McKay, H. Harris, E. Kim, S. L. Hansen, M. D. Rosenblum, and J. M. Moreau (2024, February). Tertiary lymphoid structures sustain cutaneous B cell activity in hidradenitis suppurativa. *JCI Insight* 9(3).
- Lu, M. Y., B. Chen, D. F. Williamson, R. J. Chen, I. Liang, T. Ding, G. Jaume, I. Odintsov, L. P. Le, G. Gerber, et al. (2024). A visual-language foundation model for computational pathology. *Nature Medicine* 30(3), 863–874.
- Luecken, M. D., M. Büttner, K. Chaichoompu, A. Danese, M. Interlandi, M. F. Müller, D. C. Strobl, L. Zappia, M. Dugas, M. Colomé-Tatché, et al. (2022). Benchmarking atlas-level data integration in single-cell genomics. *Nature methods* 19(1), 41–50.

- Luo, T., J. Chen, W. Wu, J. Zhao, H. Yao, H. Zhu, and Y. Li (2024). Mast-decon: Smooth cell-type deconvolution method for spatial transcriptomics data. *bioRxiv*, 2024–05.
- Lyubetskaya, A., B. Rabe, A. Fisher, A. Lewin, I. Neuhaus, C. Brett, T. Brett, E. Pereira, R. Golhar, S. Kebede, A. Font-Tello, K. Mosure, N. Van Wittenberghe, K. J. Mavrakis, K. MacIsaac, B. J. Chen, and E. Drokhyansky (2022, November). Assessment of spatial transcriptomics for oncology discovery. *Cell Rep. Methods* 2(11), 100340.
- Mauduit, O., V. Delcroix, T. Umazume, C. S. de Paiva, D. A. Dartt, and H. P. Makarenkova (2022, October). Spatial transcriptomics of the lacrimal gland features macrophage activity and epithelium metabolism as key alterations during chronic inflammation. *Front. Immunol.* 13, 1011125.
- Maynard, K. R., L. Collado-Torres, L. M. Weber, C. Uyttingco, B. K. Barry, S. R. Williams, J. L. Catallini, 2nd, M. N. Tran, Z. Besich, M. Tippani, J. Chew, Y. Yin, J. E. Kleinman, T. M. Hyde, N. Rao, S. C. Hicks, K. Martinowich, and A. E. Jaffe (2021, March). Transcriptome-scale spatial gene expression in the human dorsolateral prefrontal cortex. *Nat. Neurosci.* 24(3), 425–436.
- Mei, Y., X. Wang, J. Zhang, D. Liu, J. He, C. Huang, J. Liao, Y. Wang, Y. Feng, H. Li, X. Liu, L. Chen, W. Yi, X. Chen, H.-M. Bai, X. Wang, Y. Li, L. Wang, Z. Liang, X. Ren, L. Qiu, Y. Hui, Q. Zhang, Q. Leng, J. Chen, and G. Jia (2023, September). Siglec-9 acts as an immune-checkpoint molecule on macrophages in glioblastoma, restricting t-cell priming and immunotherapy response. *Nat. Cancer* 4(9), 1273–1291.
- Meylan, M., F. Petitprez, E. Becht, A. Bougouiin, G. Pupier, A. Calvez, I. Giglioli, V. Verkarre, G. Lacroix, J. Verneau, C.-M. Sun, P. Laurent-Puig, Y.-A. Vano, R. Elaïdi, A. Méjean, R. Sanchez-Salas, E. Barret, X. Cathelineau, S. Oudard, C.-A. Reynaud, A. de Reyniès, C. Sautès-Fridman, and W. H. Fridman (2022, March). Tertiary lymphoid structures generate and propagate anti-tumor antibody-producing plasma cells in renal cell cancer. *Immunity* 55(3), 527–541.e5.
- Mikheenko, A., A. D. Prjibelski, A. Joglekar, and H. U. Tilgner (2022, April). Sequencing of individual barcoded cDNAs using pacific biosciences and oxford nanopore technologies reveals platform-specific error patterns. *Genome Res.* 32(4), 726–737.
- Mirzazadeh, R., Z. Andrusivova, L. Larsson, P. T. Newton, L. A. Galicia, X. M. Abalo, M. Avijgan, L. Kvastad, A. Denadai-Souza, N. Stakenborg, A. B. Firsova, A. Shamikh, A. Jurek, N. Schultz, M. Nistér, C. Samakovlis, G. Boeckxstaens, and J. Lundeberg (2023, January). Spatially resolved transcriptomic profiling of degraded and challenging fresh frozen samples. *Nat. Commun.* 14(1), 509.
- Misra, A., C. D. Baker, E. M. Pritchett, K. N. Burgos Villar, J. M. Ashton, and E. M. Small (2021, December). Characterizing neonatal heart maturation, regeneration, and scar resolution using spatial transcriptomics. *J. Cardiovasc. Dev. Dis.* 9(1), 1.
- Mitamura, Y., M. Reiger, J. Kim, Y. Xiao, D. Zhakparov, G. Tan, B. Rückert, A. O. Rinaldi, K. Baerenfaller, M. Akdis, M.-C. Brüggem, K. C. Nadeau, P. M. Brunner, D. Roqueiro, C. Traidl-Hoffmann, and C. A. Akdis (2023, August). Spatial transcriptomics combined with single-cell RNA-sequencing unravels the complex inflammatory cell network in atopic dermatitis. *Allergy* 78(8), 2215–2231.
- Moeyersoms, A. H. M., R. A. Gallo, M. G. Zhang, V. Stathias, M. M. Maeng, D. Owens, R. Abou Khzam, Y. Sayegh, C. Maza, S. R. Dubovy, D. T. Tse, and D. Pelaez (2023, June). Spatial transcriptomics identifies expression signatures specific to lacrimal gland adenoid cystic carcinoma cells. *Cancers (Basel)* 15(12).
- Mohammadi, E., K. Chojnowska, M. Bieñkowski, A. Kostecka, M. Koczkowska, M. A. Żmijewski, M. Jąkalski, M. Ingelsson, N. Filipowicz, P. Olszewski, H. Davies, J. M. Wierzbicka, B. T. Hyman, J. P. Dumanski, A. Piotrowski, and J. Mieczkowski (2023, April). Size matters: the impact of nucleus size on results from spatial transcriptomics. *J. Transl. Med.* 21(1), 270.
- Moses, L. and L. Pachter (2022). Museum of spatial transcriptomics. *Nature methods* 19(5), 534–546.

- Nakata, T., C. Li, T. Mayassi, H. Lin, K. Ghosh, Å. Segerstolpe, E. L. Diamond, P. Herbst, T. Biancalani, S. Gaddam, S. Parkar, Z. Lu, A. Jaiswal, B. Li, E. A. Creasey, A. Lefkovith, M. J. Daly, D. B. Graham, and R. J. Xavier (2023, October). Genetic vulnerability to crohn’s disease reveals a spatially resolved epithelial restitution program. *Sci. Transl. Med.* 15(719), eadg5252.
- Ng, M. S. F., I. Kwok, L. Tan, C. Shi, D. Cerezo-Wallis, Y. Tan, K. Leong, G. F. Calvo, K. Yang, Y. Zhang, J. Jin, K. H. Liong, D. Wu, R. He, D. Liu, Y. C. Teh, C. Bleriot, N. Caronni, Z. Liu, K. Duan, V. Narang, I. Ballesteros, F. Moalli, M. Li, J. Chen, Y. Liu, L. Liu, J. Qi, Y. Liu, L. Jiang, B. Shen, H. Cheng, T. Cheng, V. Angeli, A. Sharma, Y.-H. Loh, H. L. Tey, S. Z. Chong, M. Iannacone, R. Ostuni, A. Hidalgo, F. Ginhoux, and L. G. Ng (2024, January). Deterministic reprogramming of neutrophils within tumors. *Science* 383(6679), eadf6493.
- Ni, Z., A. Prasad, S. Chen, R. B. Halberg, L. M. Arkin, B. A. Drolet, M. A. Newton, and C. Kendzierski (2022, May). SpotClean adjusts for spot swapping in spatial transcriptomics data. *Nat. Commun.* 13(1), 2971.
- Nikolov, S., S. Blackwell, A. Zverovitch, R. Mendes, M. Livne, J. De Fauw, Y. Patel, C. Meyer, H. Askham, B. Romera-Paredes, et al. (2021). Clinically applicable segmentation of head and neck anatomy for radiotherapy: deep learning algorithm development and validation study. *Journal of medical Internet research* 23(7), e26151.
- Olaniru, O. E., U. Kadolsky, S. Kannambath, H. Vaikkinen, K. Fung, P. Dhimi, and S. J. Persaud (2023, January). Single-cell transcriptomic and spatial landscapes of the developing human pancreas. *Cell Metab.* 35(1), 184–199.e5.
- Parigi, S. M., L. Larsson, S. Das, R. O. Ramirez Flores, A. Frede, K. P. Tripathi, O. E. Diaz, K. Selin, R. A. Morales, X. Luo, G. Monasterio, C. Engblom, N. Gagliani, J. Saez-Rodriguez, J. Lundeberg, and E. J. Villablanca (2022, February). The spatial transcriptomic landscape of the healing mouse intestine following damage. *Nat. Commun.* 13(1), 828.
- Pavel, I., L. Irina, G. Tatiana, P. Denis, K. Philipp, K. Sergei, and D. Evgeny (2023, March). Comparison of the illumina NextSeq 2000 and GeneMind genolab M sequencing platforms for spatial transcriptomics. *BMC Genomics* 24(1), 102.
- Pedregosa, F., G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, and E. Duchesnay (2011). Scikit-learn: Machine learning in Python. *Journal of Machine Learning Research* 12, 2825–2830.
- Pham, D., X. Tan, B. Balderson, J. Xu, L. F. Grice, S. Yoon, E. F. Willis, M. Tran, P. Y. Lam, A. Raghobar, P. Kalita-de Croft, S. Lakhani, J. Vukovic, M. J. Ruitenber, and Q. H. Nguyen (2023, November). Robust mapping of spatiotemporal trajectories and cell-cell interactions in healthy and diseased tissues. *Nat. Commun.* 14(1), 7739.
- Prjibelski, A. D., A. Mikheenko, A. Joglekar, A. Smetanin, J. Jarroux, A. L. Lapidus, and H. U. Tilgner (2023, July). Accurate isoform discovery with IsoQuant using long reads. *Nat. Biotechnol.* 41(7), 915–918.
- Radford, A., J. W. Kim, C. Hallacy, A. Ramesh, G. Goh, S. Agarwal, G. Sastry, A. Askell, P. Mishkin, J. Clark, et al. (2021). Learning transferable visual models from natural language supervision. In *International conference on machine learning*, pp. 8748–8763. PMLR.
- Rahimikollu, J., H. Xiao, A. Rosengart, A. B. I. Rosen, T. Tabib, P. M. Zdinak, K. He, X. Bing, F. Bunea, M. Wegkamp, A. C. Poholek, A. V. Joglekar, R. A. Lafyatis, and J. Das (2024, February). SLIDE: Significant latent factor interaction discovery and exploration across biological domains. *Nat. Methods*.
- Ratz, M., L. von Berlin, L. Larsson, M. Martin, J. O. Westholm, G. La Manno, J. Lundeberg, and J. Frisén (2022, March). Clonal relations in the mouse brain revealed by single-cell and spatial transcriptomics. *Nat. Neurosci.* 25(3), 285–294.

- Rauber, S., H. Mohammadian, C. Schmidkonz, A. Atzinger, A. Soare, C. Treutlein, S. Kemble, C. B. Mahony, M. Geisthoff, M. R. Angeli, M. G. Raimondo, C. Xu, K.-T. Yang, L. Lu, H. Labinsky, M. S. A. Saad, C. A. Gwellem, J. Chang, K. Huang, E. Kampylafka, J. Knitza, R. Bilyy, J. H. W. Distler, M. M. Hanlon, U. Fearon, D. J. Veale, F. W. Roemer, T. Bäuerle, H. M. Maric, S. Maschauer, A. B. Ekici, C. D. Buckley, A. P. Croft, T. Kuwert, O. Prante, J. D. Cañete, G. Schett, and A. Ramming (2024, April). CD200+ fibroblasts form a pro-resolving mesenchymal network in arthritis. *Nat. Immunol.* 25(4), 682–692.
- Reddy, P. (1996). Diagnostic histopathology of tumours. *Medical Journal, Armed Forces India* 52(3), 209.
- Ren, Y., Z. Huang, L. Zhou, P. Xiao, J. Song, P. He, C. Xie, R. Zhou, M. Li, X. Dong, Q. Mao, C. You, J. Xu, Y. Liu, Z. Lan, T. Zhang, Q. Gan, Y. Yang, T. Chen, B. Huang, X. Yang, A. Xiao, Y. Ou, Z. Su, L. Chen, Y. Zhang, Y. Ju, Y. Zhang, and Y. Wang (2023, February). Spatial transcriptomics reveals niche-specific enrichment and vulnerabilities of radial glial stem-like cells in malignant gliomas. *Nat. Commun.* 14(1), 1028.
- Rousseeuw, P. J. (1987). Silhouettes: a graphical aid to the interpretation and validation of cluster analysis. *Journal of computational and applied mathematics* 20, 53–65.
- Russ, J. E., M. E. Haywood, S. L. Lane, W. B. Schoolcraft, and M. G. Katz-Jaffe (2022, August). Spatially resolved transcriptomic profiling of ovarian aging in mice. *iScience* 25(8), 104819.
- Rustagi, Y., A. S. Abouhashem, P. Verma, S. S. Verma, E. Hernandez, S. Liu, M. Kumar, P. R. Guda, R. Srivastava, S. K. Mohanty, S. Kacar, S. Mahajan, K. E. Wanczyk, S. Khanna, M. P. Murphy, G. M. Gordillo, S. Roy, J. Wan, C. K. Sen, and K. Singh (2022, May). Endothelial phospholipase C $\gamma$ 2 improves outcomes of diabetic ischemic limb rescue following VEGF therapy. *Diabetes* 71(5), 1149–1165.
- Sanchez-Ferras, O., A. Pacis, M. Sotiropoulou, Y. Zhang, Y. C. Wang, M. Bourgey, G. Bourque, J. Ragoussis, and M. Bouchard (2021, May). A coordinated progression of progenitor cell states initiates urinary tract development. *Nat. Commun.* 12(1), 2627.
- Sanders, B. E., R. Wolsky, E. S. Doughty, K. L. Wells, D. Ghosh, L. Ku, J. G. Pressey, B. B. Bitler, and L. W. Brubaker (2022, December). Small cell carcinoma of the ovary hypercalcemic type (SCCOHT): A review and novel case with dual germline SMARCA4 and BRCA2 mutations. *Gynecol. Oncol. Rep.* 44(101077), 101077.
- Sans, M., Y. Makino, J. Min, K. I. Rajapakshe, M. Yip-Schneider, C. M. Schmidt, M. W. Hurd, J. K. Burks, J. A. Gomez, F. I. Thege, J. F. Fahrman, R. A. Wolff, M. P. Kim, P. A. Guerrero, and A. Maitra (2023, August). Spatial transcriptomics of intraductal papillary mucinous neoplasms of the pancreas identifies NKX6-2 as a driver of gastric differentiation and indolent biological potential. *Cancer Discov.* 13(8), 1844–1861.
- Satija, R., J. A. Farrell, D. Gennert, A. F. Schier, and A. Regev (2015). Spatial reconstruction of single-cell gene expression data. *Nature biotechnology* 33(5), 495–502.
- Schaar, A. C., A. Tejada-Lapueta, G. Palla, R. Gutgesell, L. Halle, M. Minaeva, L. Vornholz, L. Dony, F. Drummer, M. Bahrami, et al. (2024). Nicheformer: a foundation model for single-cell and spatial omics. *bioRxiv*, 2024–04.
- Schuhmann, C., R. Beaumont, R. Vencu, C. Gordon, R. Wightman, M. Cherti, T. Coombes, A. Katta, C. Mullis, M. Wortsman, et al. (2022). Laion-5b: An open large-scale dataset for training next generation image-text models. *Advances in Neural Information Processing Systems* 35, 25278–25294.
- Srinivasan, K., K. Raman, J. Chen, M. Bendersky, and M. Najork (2021). Wit: Wikipedia-based image text dataset for multimodal multilingual machine learning. In *Proceedings of the 44th International ACM SIGIR Conference on Research and Development in Information Retrieval*, pp. 2443–2449.
- Ståhl, P. L., F. Salmén, S. Vickovic, A. Lundmark, J. F. Navarro, J. Magnusson, S. Giacomello, M. Asp, J. O. Westholm, M. Huss, et al. (2016). Visualization and analysis of gene expression in tissue sections by spatial transcriptomics. *Science* 353(6294), 78–82.

- Stahl, P. L., F. Salmén, S. Vickovic, A. Lundmark, J. F. Navarro, J. Magnusson, S. Giacomello, M. Asp, J. O. Westholm, M. Huss, A. Mollbrink, S. Linnarsson, S. Codeluppi, Å. Borg, F. Pontén, P. I. Costea, P. Sahlén, J. Mulder, O. Bergmann, J. Lundeberg, and J. Frisén (2016, July). Visualization and analysis of gene expression in tissue sections by spatial transcriptomics. *Science* 353(6294), 78–82.
- Stec, M. J., Q. Su, C. Adler, L. Zhang, D. R. Golann, N. P. Khan, L. Panagis, S. A. Villalta, M. Ni, Y. Wei, J. R. Walls, A. J. Murphy, G. D. Yancopoulos, G. S. Atwal, S. Kleiner, G. Halasz, and M. W. Sleeman (2023, July). A cellular and molecular spatial atlas of dystrophic muscle. *Proc. Natl. Acad. Sci. U. S. A.* 120(29), e2221249120.
- Stein, A. N., A. Joglekar, C.-L. Poon, and H. U. Tilgner (2022, June). ScisorWiz: visualizing differential isoform expression in single-cell long-read data. *Bioinformatics* 38(13), 3474–3476.
- Subramanian, A., N. Nemat-Gorgani, T. J. Ellis-Caleo, D. G. P. van IJzendoorn, T. J. Sears, A. Somani, B. A. Luca, M. Y. Zhou, M. Bradic, I. A. Torres, E. Oladipo, C. New, D. E. Kenney, R. S. Avedian, R. J. Steffner, M. S. Binkley, D. G. Mohler, W. D. Tap, S. P. D’Angelo, M. van de Rijn, K. N. Ganjoo, N. Q. Bui, G. W. Charville, A. M. Newman, and E. J. Moding (2024, March). Sarcoma microenvironment cell states and ecosystems are associated with prognosis and predict response to immunotherapy. *Nat. Cancer*.
- Sudmeier, L. J., K. B. Hoang, E. K. Nduom, A. Wieland, S. G. Neill, M. J. Schniederjan, S. S. Ramalingam, J. J. Olson, R. Ahmed, and W. H. Hudson (2022, May). Distinct phenotypic states and spatial distribution of CD8+ T cell clonotypes in human brain metastases. *Cell Rep. Med.* 3(5), 100620.
- Sukhanov, S., Y. Higashi, T. Yoshida, S. Danchuk, M. Alfortish, T. Goodchild, A. Scarborough, T. Sharp, J. S. Jenkins, D. Garcia, J. Ivey, D. L. Tharp, J. Schumacher, Z. Rozenbaum, J. K. Kolls, D. Bowles, D. Lefer, and P. Delafontaine (2023, February). Insulin-like growth factor 1 reduces coronary atherosclerosis in pigs with familial hypercholesterolemia. *JCI Insight* 8(4).
- Szałata, A., K. Hrovatin, S. Becker, A. Tejada-Lapuerta, H. Cui, B. Wang, and F. J. Theis (2024). Transformers in single-cell omics: a review and new perspectives. *Nature Methods* 21(8), 1430–1443.
- Tian, L., F. Chen, and E. Z. Macosko (2023). The expanding vistas of spatial transcriptomics. *Nature Biotechnology* 41(6), 773–782.
- Topchyan, P., R. Zander, M. Y. Kasmani, C. Nguyen, A. Brown, S. Lin, R. Burns, and W. Cui (2022, November). Spatial transcriptomics demonstrates the role of CD4 T cells in effector CD8 T cell differentiation during chronic viral infection. *Cell Rep.* 41(9), 111736.
- Tower, R. J., E. Busse, J. Jaramillo, M. Lacey, K. Hoffseth, A. R. Guntur, J. Simkin, and M. C. Sammarco (2022, May). Spatial transcriptomics reveals metabolic changes underlying age-dependent declines in digit regeneration. *Elife* 11.
- Tower, R. J., Z. Li, Y.-H. Cheng, X.-W. Wang, L. Rajbhandari, Q. Zhang, S. Negri, C. R. Uyttingco, A. Venkatesan, F.-Q. Zhou, P. Cahan, A. W. James, and T. L. Clemens (2021, October). Spatial transcriptomics reveals a role for sensory nerves in preserving cranial suture patency through modulation of BMP/TGF- $\beta$  signaling. *Proc. Natl. Acad. Sci. U. S. A.* 118(42), e2103087118.
- Tung, L. W., E. Groppa, H. Soliman, B. Lin, C. Chang, C. W. Cheung, M. Ritso, D. Guo, L. Rempel, S. Sinha, C. Eisner, J. Brassard, K. McNagny, J. Biernaskie, and F. Rossi (2023, December). Spatiotemporal signaling underlies progressive vascular rarefaction in myocardial infarction. *Nat. Commun.* 14(1), 8498.
- Van der Maaten, L. and G. Hinton (2008). Visualizing data using t-sne. *Journal of machine learning research* 9(11).
- Vanrobaeys, Y., U. Mukherjee, L. Langmack, S. E. Beyer, E. Bahl, L.-C. Lin, J. J. Michaelson, T. Abel, and S. Chatterjee (2023, September). Mapping the spatial transcriptomic signature of the hippocampus during memory consolidation. *Nat. Commun.* 14(1), 6100.

- Vanrobaeys, Y., Z. J. Peterson, E. N. Walsh, S. Chatterjee, L.-C. Lin, L. C. Lyons, T. Nickl-Jockschat, and T. Abel (2023, November). Spatial transcriptomics reveals unique gene expression changes in different brain regions after sleep deprivation. *Nat. Commun.* 14(1), 7095.
- Villemin, J.-P., L. Bassaganyas, D. Pourquier, F. Boissière, S. Cabello-Aguilar, E. Crapez, R. Tanos, E. Cornillot, A. Turtoi, and J. Colinge (2023, May). Inferring ligand-receptor cellular networks from bulk and spatial transcriptomic datasets with BulkSignalR. *Nucleic Acids Res.* 51(10), 4726–4744.
- Wang, X., W. E. Allen, M. A. Wright, E. L. Sylwestrak, N. Samusik, S. Vesuna, K. Evans, C. Liu, C. Ramakrishnan, J. Liu, et al. (2018). Three-dimensional intact-tissue sequencing of single-cell transcriptional states. *Science* 361(6400), eaat5691.
- Wolf, F. A., P. Angerer, and F. J. Theis (2018). Scanpy: large-scale single-cell gene expression data analysis. *Genome biology* 19, 1–5.
- Wu, H., E. E. Dixon, Q. Xuanyuan, J. Guo, Y. Yoshimura, C. Debashish, A. Niesnerova, H. Xu, M. Rouault, and B. D. Humphreys (2024, February). High resolution spatial profiling of kidney injury and repair using RNA hybridization-based in situ sequencing. *Nat. Commun.* 15(1), 1396.
- Wu, S. Z., G. Al-Eryani, D. L. Roden, S. Junankar, K. Harvey, A. Andersson, A. Thennavan, C. Wang, J. R. Torpy, N. Bartonicek, T. Wang, L. Larsson, D. Kaczorowski, N. I. Weisenfeld, C. R. Uyttingco, J. G. Chew, Z. W. Bent, C.-L. Chan, V. Gnanasambandapillai, C.-A. Dutertre, L. Gluch, M. N. Hui, J. Beith, A. Parker, E. Robbins, D. Segara, C. Cooper, C. Mak, B. Chan, S. Warriar, F. Ginhoux, E. Millar, J. E. Powell, S. R. Williams, X. S. Liu, S. O’Toole, E. Lim, J. Lundeberg, C. M. Perou, and A. Swarbrick (2021, September). A single-cell and spatially resolved atlas of human breast cancers. *Nat. Genet.* 53(9), 1334–1347.
- Xie, R., K. Pang, S. Chung, C. Perciani, S. MacParland, B. Wang, and G. Bader (2024). Spatially resolved gene expression prediction from histology images via bi-modal contrastive learning. *Advances in Neural Information Processing Systems* 36.
- Xu, Z., W. Wang, T. Yang, J. Chen, Y. Huang, J. Gould, W. Du, F. Yang, L. Li, T. Lai, et al. (2022). Stomicsdb: a database of spatial transcriptomic data. *bioRxiv*, 2022–03.
- Yamasaki, J., Y. Hirata, Y. Otsuki, K. Suina, Y. Saito, K. Masuda, S. Okazaki, T. Ishimoto, H. Saya, and O. Nagano (2022, March). MEK inhibition suppresses metastatic progression of KRAS-mutated gastric cancer. *Cancer Sci.* 113(3), 916–925.
- Yao, J., J. Yu, B. Caffo, S. C. Page, K. Martinowich, and S. C. Hicks (2024). Spatial domain detection using contrastive self-supervised learning for spatial multi-omics technologies. *bioRxiv*.
- Yoshitake, R., H. Mori, D. Ha, X. Wu, J. Wang, X. Wang, K. Saeki, G. Chang, H. J. Shim, Y. Chan, and S. Chen (2024, January). Molecular features of luminal breast cancer defined through spatial and single-cell transcriptomics. *Clin. Transl. Med.* 14(1), e1548.
- Yu, X., Q. Gong, D. Yu, Y. Chen, Y. Jing, F. Zoulim, and X. Zhang (2024, April). Spatial transcriptomics reveals a low extent of transcriptionally active hepatitis B virus integration in patients with HBsAg loss. *Gut* 73(5), 797–809.
- Yuan, Z., W. Pan, X. Zhao, F. Zhao, Z. Xu, X. Li, Y. Zhao, M. Q. Zhang, and J. Yao (2023). Sodb facilitates comprehensive exploration of spatial omics data. *Nature Methods* 20(3), 387–399.
- Yuan, Z., F. Zhao, S. Lin, Y. Zhao, J. Yao, Y. Cui, X.-Y. Zhang, and Y. Zhao (2024). Benchmarking spatial clustering methods with spatially resolved transcriptomics data. *Nature Methods*, 1–11.
- Zeng, Y., R. Yin, M. Luo, J. Chen, Z. Pan, Y. Lu, W. Yu, and Y. Yang (2023). Identifying spatial domain by adapting transcriptomics with histology through contrastive learning. *Briefings in Bioinformatics* 24(2), bbad048.
- Zhang, D., A. Schroeder, H. Yan, H. Yang, J. Hu, M. Y. Lee, K. S. Cho, K. Susztak, G. X. Xu, M. D. Feldman, et al. (2024). Inferring super-resolution tissue architecture by integrating spatial transcriptomics with histology. *Nature Biotechnology*, 1–6.

- Zheng, Y., Y. Chen, X. Ding, K. H. Wong, and E. Cheung (2023). Aquila: a spatial omics database and analysis platform. *Nucleic Acids Research* 51(D1), D827–D834.
- Zhi, Y., Q. Wang, M. Zi, S. Zhang, J. Ge, K. Liu, L. Lu, C. Fan, Q. Yan, L. Shi, P. Chen, S. Fan, Q. Liao, C. Guo, F. Wang, Z. Gong, W. Xiong, and Z. Zeng (2024, March). Spatial transcriptomic and metabolomic landscapes of oral submucous fibrosis-derived oral squamous cell carcinoma and its tumor microenvironment. *Adv. Sci. (Weinh.)* 11(12), e2306515.
- Zhong, J., R. Aires, G. Tsissios, E. Skoufa, K. Brandt, T. Sandoval-Guzmán, and C. Aztekin (2023, October). Multi-species atlas resolves an axolotl limb development and regeneration paradox. *Nat. Commun.* 14(1), 6346.
- Zhou, W., M. Su, T. Jiang, Q. Yang, Q. Sun, K. Xu, J. Shi, C. Yang, N. Ding, Y. Li, et al. (2024). Sorc: an integrated spatial omics resource in cancer. *Nucleic Acids Research* 52(D1), D1429–D1437.

## Checklist

1. For all authors...
  - (a) Do the main claims made in the abstract and introduction accurately reflect the paper's contributions and scope? [Yes]
  - (b) Did you describe the limitations of your work? [Yes] See Section 6.
  - (c) Did you discuss any potential negative societal impacts of your work? [Yes] See Section 6.
  - (d) Have you read the ethics review guidelines and ensured that your paper conforms to them? [Yes]
2. If you are including theoretical results...
  - (a) Did you state the full set of assumptions of all theoretical results? [N/A]
  - (b) Did you include complete proofs of all theoretical results? [N/A]
3. If you ran experiments (e.g. for benchmarks)...
  - (a) Did you include the code, data, and instructions needed to reproduce the main experimental results (either in the supplemental material or as a URL)? [Yes] The URL is provided in Supplementary S1 and S2.
  - (b) Did you specify all the training details (e.g., data splits, hyperparameters, how they were chosen)? [Yes] The details are provided in the Appendix B.
  - (c) Did you report error bars (e.g., with respect to the random seed after running experiments multiple times)? [Yes] The error bars are provided in Figure 4a.
  - (d) Did you include the total amount of compute and the type of resources used (e.g., type of GPUs, internal cluster, or cloud provider)? [Yes] The details are provided in the Appendix B.
4. If you are using existing assets (e.g., code, data, models) or curating/releasing new assets...
  - (a) If your work uses existing assets, did you cite the creators? [Yes] All studies are cited in the main paper and in Appendix A.
  - (b) Did you mention the license of the assets? [Yes] See Supplementary S1 and S2.
  - (c) Did you include any new assets either in the supplemental material or as a URL? [No]
  - (d) Did you discuss whether and how consent was obtained from people whose data you're using/curating? [Yes] See Supplementary S1.
  - (e) Did you discuss whether the data you are using/curating contains personally identifiable information or offensive content? [Yes] See Supplementary S1.
5. If you used crowdsourcing or conducted research with human subjects...
  - (a) Did you include the full text of instructions given to participants and screenshots, if applicable? [N/A] No direct interact with subjects.
  - (b) Did you describe any potential participant risks, with links to Institutional Review Board (IRB) approvals, if applicable? [N/A] No direct interact with subjects.
  - (c) Did you include the estimated hourly wage paid to participants and the total amount spent on participant compensation? [N/A] No direct interact with subjects.



# Appendix

## A Studies included

Here we include the full list of reference of studies included in STImage-1K4M: (Ji et al., 2020; Chen et al., 2023; Andersson et al., 2021; He et al., 2020b; Ståhl et al., 2016; Asp et al., 2019; Hasel et al., 2021; Ratz et al., 2022; Lebrigand et al., 2023; Joglekar et al., 2021; Mikheenko et al., 2022; Stein et al., 2022; Prjibelski et al., 2023; Bashkirova et al., 2023; Sanchez-Ferras et al., 2021; Parigi et al., 2022; Tower et al., 2021; Meylan et al., 2022; Ni et al., 2022; Kadur Lakshminarasimha Murthy et al., 2022; Foster et al., 2021; Sudmeier et al., 2022; Hudson and Sudmeier, 2022; Tower et al., 2022; Rustagi et al., 2022; Dixon et al., 2022; Lake et al., 2023; Mohammadi et al., 2023; Yamasaki et al., 2022; Chen et al., 2022; Russ et al., 2022; Misra et al., 2021; Williams et al., 2022; Habenicht et al., 2022; Dhainaut et al., 2022; Ren et al., 2023; Kenney et al., 2023; Mitamura et al., 2023; Olaniru et al., 2023; Heezen et al., 2023; Topchyan et al., 2022; Filipescu et al., 2023; Vanrobaeys et al., 2023; Kasmani et al., 2023; Castranio et al., 2023; Barkley et al., 2022; Eum et al., 2024; Canela et al., 2023; Garbarino et al., 2023; Caetano et al., 2023; Tung et al., 2023; Heimli et al., 2022; Arora et al., 2023; Chen et al., 2023; Mauduit et al., 2022; Bassiouni et al., 2023; Lyubetskaya et al., 2022; Lee et al., 2023; Foster et al., 2022; Subramanian et al., 2024; Gu et al., 2022; Coutant et al., 2023; Sanders et al., 2022; Akiyama et al., 2023; Yoshitake et al., 2024; Ballester Roig et al., 2023; Villemain et al., 2023; Caronni et al., 2023; Sukhanov et al., 2023; Zhi et al., 2024; Kerzel et al., 2023; Mirzazadeh et al., 2023; Vanrobaeys et al., 2023; Liu et al., 2023; Chaker et al., 2023; Deshpande et al., 2023a,b; Stec et al., 2023; Chitturi et al., 2023; Pavel et al., 2023; Wu et al., 2024; Moeyersoms et al., 2023; Rauber et al., 2024; Sans et al., 2023; Gharaie et al., 2023; Mei et al., 2023; Nakata et al., 2023; Pham et al., 2023; Andrews et al., 2024; Di Marco et al., 2023; Kawai et al., 2023; Zhong et al., 2023; Lequain et al., 2023; Yu et al., 2024; Ng et al., 2024; Cherief et al., 2023; Rahimikollu et al., 2024; Lowe et al., 2024; Cortese et al., 2023; Huuki-Myers et al., 2023; Maynard et al., 2021; Wu et al., 2021; Kuppe et al., 2022; Erickson et al., 2022). We note here that there are some studies with no valid pmid. The full list of study names will be available at GitHub repository.

We also include pathologist annotation for 71 slides from 9 datasets involving human brain, human breast, human prostate, human kidney, and mouse brain (Figure 6).

## B CLIP and PLIP finetuning details

To evaluate the potential of STImage-1K4M, we fine-tuned the image encoder part of CLIP (Radford et al., 2021) and PLIP (Huang et al., 2023) model. All model implementations are built upon the training code (issue #83) posted in the CLIP GitHub repository. The hyperparameters are chose to be the same of CLIP training as detailed in Table 1. All the parameters are transformed into fp32. For CLIP, we loaded the pretrained parameters (ViT-B/32) from *openai/clip-vit-base-patch32* from hugging face. For PLIP, we loaded the pretrained parameters (ViT-L/14) from *vinid/plip* from hugging face. In our model architecture, the image encoder feeds into a fully connected layer that reduces its output to a 32-dimensional latent space. Similarly, the gene expression encoder also consists of a single fully connected layer that compresses its high-dimensional input down to a 32-dimensional embedding (Figure 7). These 32-dimensional representations from both the image and gene expression encoders are then utilized as the basis for our contrastive loss function. We fine-tuned the models for 15 epochs. All experiments are performed on single NVIDIA A100 GPU.

To compare the choice of gene sets, we employed two methods.

1. **HVG** For each dataset, we selected the top 128 HVGs. Then the HVGs of all dataset are sorted by highly variable rank and combined regardless of gene names as training data for fine-tuning. The HVG are selected using python scanpy package `scanpy.pp.highly_variable_genes` default settings (Lause et al., 2021; Wolf et al., 2018).
2. **Overlap HVG** For each training data, we first combined the gene expression with respected to gene names and only keep the overlapping genes. Then we select the top 100 HVGs for the combined gene expression as training data. The HVG are selected using python scanpy package `scanpy.experimental.pp.highly_variable_genes` default settings (Satija et al., 2015; Wolf et al., 2018).

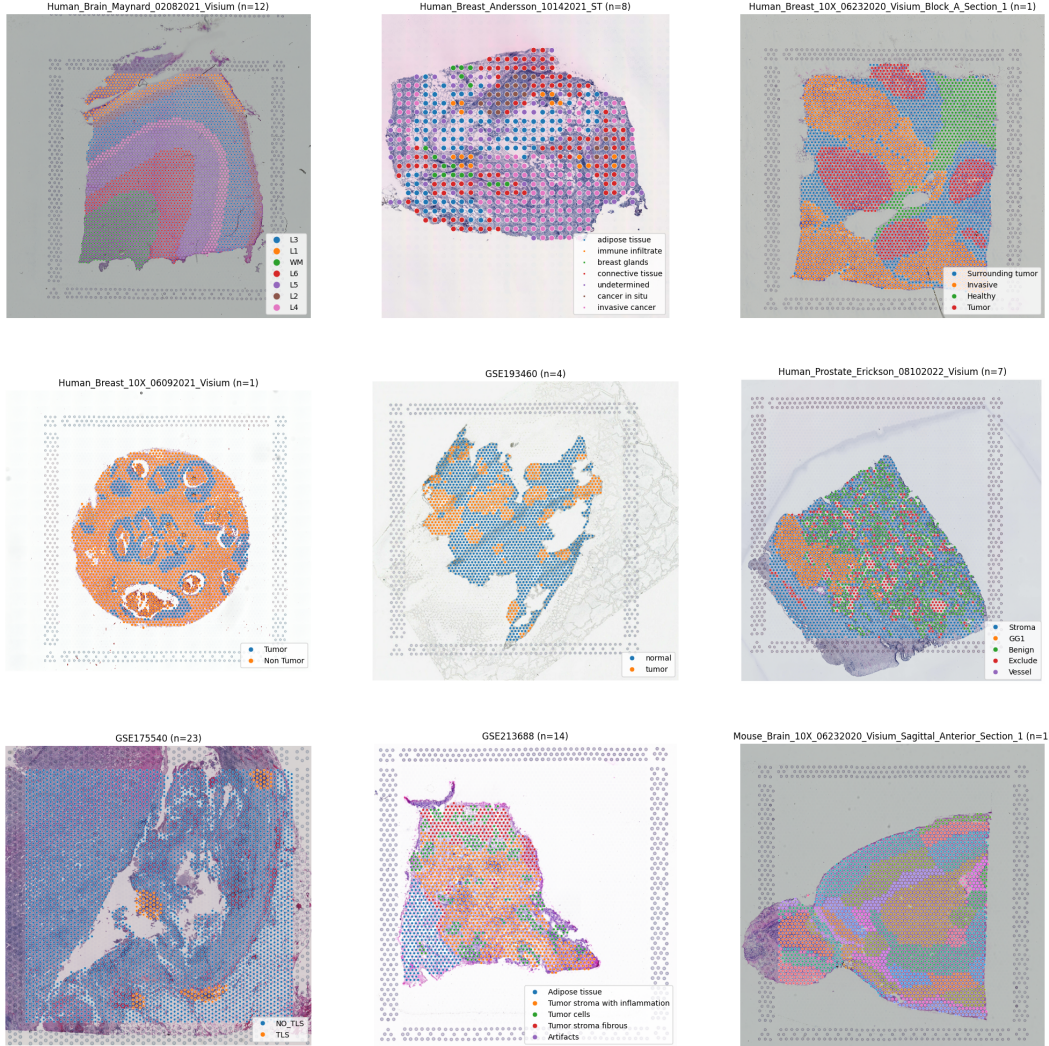


Figure 6: Datasets with pathologist annotation. The points are colored by annotation in each dataset. The legend for mouse brain data (bottom right) are omitted for visualization.

As discussed in Section 6, the gene names of different species are not overlapping. Thus, in order to use combined gene expression from multiple datasets, we subset STImage-1K4M by species. In this study, we fine-tuned models using all data from (Maynard et al., 2021) (human brain) with different gene sets.

Table 1: CLIP and PLIP fine-tuning hyperparameters.

Hyperparameter	Value
Batch size	1024
Training epochs	15
Weight decay	0.2
Adam $\beta_1$	0.9
Adam $\beta_2$	0.98
Adam $\epsilon$	1e-06

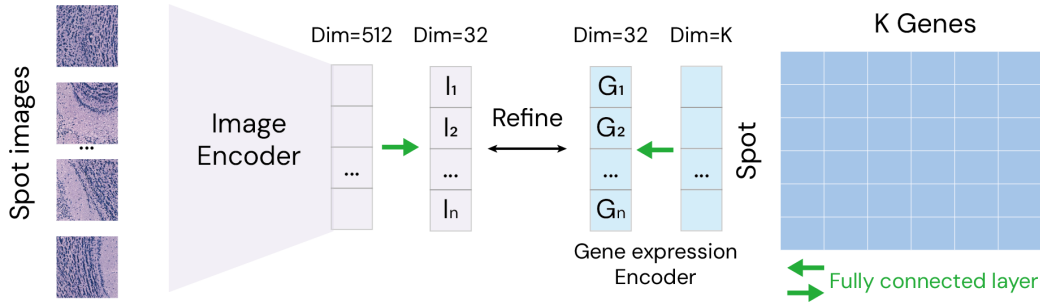


Figure 7: Fine-tune model structure. We replaced the text encoder in PLIP and CLIP with a single fully connected layer (gene expression encoder). We also add a fully connected layer after the image encoder to make the dimension same as the gene expression encoder, which in our case is 32.

### C Linear probing details

For linear probing, we follow the procedure in (Huang et al., 2023). We employ the stochastic gradient descent classifier (SGDClassifier) module for logistic regression classifier from the sklearn (Pedregosa et al., 2011) Python package. For fine-tuned models, we used 32-dimension image embeddings. For zero-shot models, we used 512-dimension embeddings. The performance of the trained linear classifier using L2 regularization with different regularization multipliers ( $\alpha = 1, 0.1, 0.01, 0.001, 0.0001$ ) are evaluated on the validation splits for all models (training:validation:test = 8 : 1 : 1). The best-performing linear classifier was selected based on the average macro F1 performance from the results trained on training splits sampled by five different random seeds (seed = 1, 2, 3, 4, 5). We note here that we fine-tune the models using data from other tissue types, resulted in sub-optimal performance (Figure 8). We also evaluated the performance of the fine-tuned model using the second-last layer image embeddings (512-dimension), the results remain similar as the performance of using 32-dimension image embedding (Figure 9).

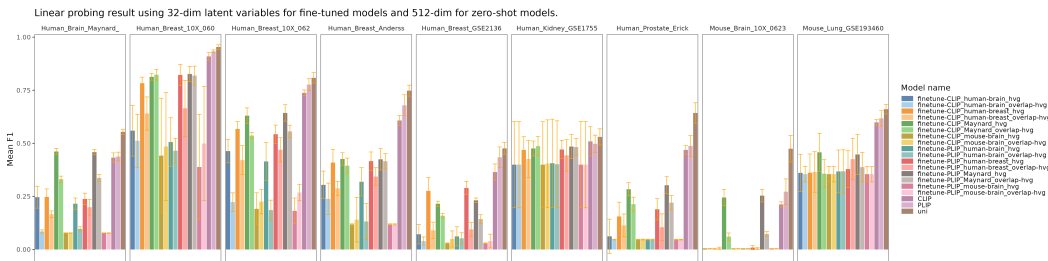


Figure 8: Classification results evaluated on multiple dataset for model fine-tuned with various training data. We fine-tuned the CLIP and PLIP models with human brain, human breast, Maynard et al., mouse brain. For fine-tuned models, we evaluate the 32-dimensional image embedding in a classification task. For zero-shot models, we evaluate the 512-dimensional image embedding. The classification performance of these models are tested on 9 datasets with annotation. The bars are colored by model type. The model name of the fine-tuned models is in “fine-tune model-type training-datatype gene-set” format.

### D t-SNE details

We applied t-SNE (Van der Maaten and Hinton, 2008) to each slide individually using sklearn.manifold.TSNE function from the sklearn (Pedregosa et al., 2011) Python package. For the fine-tuned models, we utilized 32-dimensional embeddings, whereas for the zero-shot models, we employed 512-dimensional embeddings. We also used this setting in our calculations of the



Figure 9: Classification results evaluated on multiple dataset for model fine-tuned with various training data. We fine-tuned the CLIP and PLIP models with human brain, human breast, Maynard et al., mouse brain. For both fine-tuned models and zero-shot models, we evaluate the 512-dimensional image embedding. The classification performance of these models are tested on 9 datasets with annotation. The bars are colored by model type. The model name of the fine-tuned models is in “fine-tune model-type training-datatype gene-set” format.

Silhouette, Calinski-Harabasz, and Davies-Bouldin scores using the sklearn (Pedregosa et al., 2011) Python package. We also included a t-SNE embedding of all samples from the Maynard et al. study, which comprises data from 12 human brain slides and demonstrates notable batch effects.

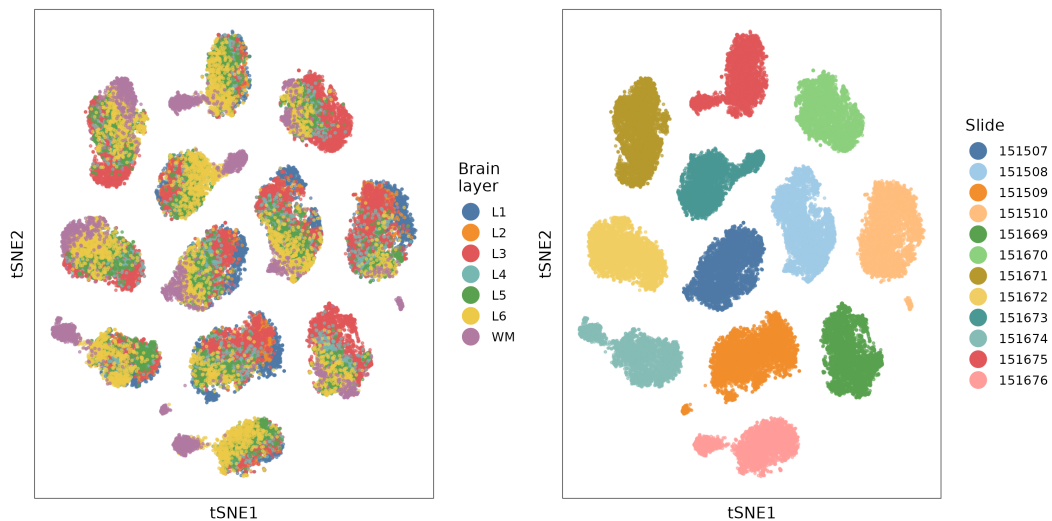


Figure 10: Batch effect in Maynard et al. human brain gene expression with hvg gene set. Left: gene expression tSNE embedding colored by brain layers. Right: gene expression tSNE embedding colored by slide name.