Article in Press

mViSE: A visual search engine for analyzing multiplex IHC brain tissue images (spatial proteomics)

Liqiang Huang, Rachel Mills, Saikiran Mandula, Lin Bai, Mahtab Jeyhani, John Redell, Hien Nguyen, Saurabh Prasad, Dragan Maric & Badrinath Roysam

We are providing an unedited version of this manuscript to give early access to its findings. Before final publication, the manuscript will undergo further editing. Please note there may be errors present which affect the content, and all legal disclaimers apply.

If this paper is publishing under a Transparent Peer Review model then Peer Review reports will publish with the final article.

mViSE: A Visual Search Engine for Analyzing Multiplex IHC Brain Tissue Images (Spatial Proteomics)

Liqiang Huang[1], Rachel Mills[1], Saikiran Mandula[1], Lin Bai[1], Mahtab Jeyhani[1], John Redell[2], Hien Nguyen[1], Saurabh Prasad[1], Dragan Maric[3], Badrinath Roysam[1*]

[1] *University of Houston, Houston, Texas 77204, USA*
[2] *The University of Texas McGovern Medical School, Houston, Texas 77030, USA*
[3] *National Institute of Neurological Disorders and Stroke, Bethesda, Maryland 20892, USA*
*Corresponding Author: Badrinath Roysam (broysam@central.uh.edu)

Abstract: Whole-slide multiplex brain tissue images for spatial proteomics are massive, information-dense, and challenging to analyze. We present mVISE, an interactive multiplex visual search engine that offers an alternative programming-free query-driven analysis method based on retrieving and profiling communities of similar cells, proximal cell pairs, and multicellular niches. The retrievals can be used for exploratory cell and tissue analysis, delineating brain regions and cortical layers, profiling and comparing brain regions/sub-regions/sub-layers, etc. mVISE is enabled by multiplex encoders that seamlessly integrate visual cues across imaging channels overcoming the limitations of current foundation models. We train separate encoders to learn each facet of tissue including cell morphologies, spatial protein expression (chemoarchitecture), cell arrangements (cytoarchitecture), and wiring patterns (myeloarchitecture) from a set of user-defined molecular marker panels, without the need for human annotations or intervention, and with visual confirmation of successful learning. Multiple encoders can be combined logically to drive specialized searches. We validated mViSE's ability to retrieve single cells, proximal cell pairs, tissue patches, delineate cortical layers, brain regions, and sub-regions. mVISE is disseminated as an open-source QuPath plug-in tool.

Introduction

Advances in whole-slide high-resolution multispectral imaging and cyclic protocols for immunohistochemical staining are enabling the collection of large, complex, and information-dense images of whole brain slices (MP-WSI) that capture multiple cell types, cell states, tissue arrangements, and biological processes simultaneously in their cyto-histological context, e.g., the intended and side effects of a drug candidate[1]. Multiplex images are challenging to analyze due to their complexity, the

need for custom interpretations for each channel, complex process for establishing cell types/sub-types and phenotypic status, and large data size. Mammalian Brain tissue is especially challenging due to its scale and cell-molecular-spatial complexity[2-6], containing a heterogeneous mix of neuronal[7], glial[8], and vascular cell types[9], organized in elaborate region-specific cellular arrangements (cytoarchitectures)[10], protein distribution patterns (chemoarchitecture)[6], wiring patterns (myeloarchitecture)[6], and cyto-vascular (niche) patterns[11].

We earlier presented cyclic multiplex immunofluorescence protocols and computational methods for systematic and comprehensive profiling of brain tissue[12]. The list of markers included multiple identifiers of all major cell types resident to the normal rat brain: (1) ubiquitous nucleus markers DAPI and histones; (2) neuronal myeloarchitectonic markers NeuN, microtubule-associated protein (MAP2), neurofilament-H (NF-H), neurofilament-M (NF-M) and synaptophysin; (3) neuronal phenotyping markers Tbr1, Eomes, glutaminase (GLUT), tyrosine hydroxylase (TH), choline acetyltransferase (ChAT), parvalbumin (PARVALB), GAD67, calbindin and calretinin; (4) astrocyte differentiation markers glial fibrillary acidic protein (GFAP), aquaporin-4, S100β and GLAST; (5) oligodendrocyte markers Olig2, CNPase and myelin basic protein (MBP), vascular associated markers endogenous GFP, rat endothelial cell antigen (RECA1), smooth muscle actin (SMA) and tomato lectin (T-lectin), rat blood brain barrier (BBB) antigen and PDGFRβ a marker of pericytes; (6) microglial marker IBA1; (7) immature stem/progenitor cell markers Sox2, nestin, vimentin and doublecortin; (8) actively proliferating cells PCNA; and (9) actively apoptotic cells cleaved caspase-3 (CC-3). The multiplex image analysis was implemented traditionally, as a pipeline of steps, starting with image pre-processing, cell detection, cell segmentation, identification of cell types and phenotypic states based on protein expression patterns, and aggregation of the histocytometric data at multiple levels of tissue organization, ranging from individual cells and niches to atlas-defined brain regions[4,13]. Although this pipeline-based image analysis strategy is powerful and comprehensive, it is rigid and requires programming to address emergent needs. There is a need for more flexible and programming-free image analysis strategies.

Here, we present an alternative *visual query-driven* strategy for multiplex brain tissue IHC image analysis, based on a multiplex Visual Search Engine (mViSE)[14,15] (**Figure 1**). It directly serves the core need in most image analysis tasks – identifying one or more cell populations, multi-cellular units, or tissue regions of interest based on their morphological and spatio-molecular characteristics and profiling them via expression maps and feature tables. mViSE operates in two phases. In the learning phase

(**Figure 2A**), it learns brain cell morphologies, spatial protein expression patterns (chemoarchitecture), cell arrangements (cytoarchitecture), and wiring patterns (myeloarchitecture) in an integrated manner fully automatically in an unsupervised manner, without the need for human annotations or intervention.

Brain tissue has a multifaceted architecture. With this in mind, the learning process is conducted in a 'divide and conquer' manner. The multiplex data is organized into a group of panels of functionally related user-defined molecular markers (**Figure 2C**), one panel per facet. For example, a panel can consist of markers that describe the myeloarchitecture, from which mViSE learns the wiring patterns of brain tissue. Another panel can enable learning the distribution of neuronal subtypes, and so on. Collectively, these panels capture the multifaceted tissue architecture and can be used individually or in combination. In the query phase, the user can indicate cells or multicellular niches of interest, specify one or more relevant molecular panels, upon which mViSE retrieves similar communities of cells or niches, and displays their molecular profiles. This is made possible by a hierarchical strategy (**Figure 2D**) for combining the learned models for multiple panels. This versatile query-driven image analysis strategy can be used for a variety of investigational goals, including tissue exploration, identifying cells and niches of interest, comparative profiling of cell populations and niches across brain regions, delineating brain regions and cell layers, regional profiling of molecular markers, and others.

The design of mViSE for multiplex brain tissue images is not straightforward. Content-Based Medical Image Retrieval (CBMIR) systems that have been developed to identify medical images within large databases[16-22] cannot handle high-order multiplex data. For example, Yottixel[14], a digital pathology search engine, represents whole-slide images as mosaics of patch barcodes, enabling real-time, large-scale retrieval with high accuracy and reduced storage requirements. UNI[23], a large-scale self-supervised pathology foundation model pretrained on a large WSI dataset across 20 tissue types, allows resolution-agnostic and data-efficient representation learning and outperforms prior models on several computational pathology tasks. *Yet, existing applications work primarily for grayscale images or RGB color H&E images; multiplex images remain largely unexplored, and the application of search engines to multiplex fluorescence imaging is still an open challenge*.

Overcoming this challenge requires the design of multiplex encoders capable of capturing the complex and region-specific morphological and spatio-molecular patterns and their variations at the level of individual cells, tissue patches, and brain regions convincingly. This is challenging for multiple reasons.

First, deep encoders with sufficient representational capacity are needed, but they generate high-dimensional encodings. We expect the high-dimensional encodings for visually similar patches to be proximal in the high-dimensional feature space, but this property is not a given. We found that a generic encoder, e.g., the variational autoencoder (VAE)[24,25], the U-Net[26] with channel attention[27], or even the recently introduced transformer based multi-task foundational model (BEiTv3) [28] do not meet this expectation[29] (**Figure 3B**). Achieving this property requires that the encoder is specifically trained for handling the visual search query-driven retrieval task. For this reason, we built upon techniques from the field of person re-identification[30] that enable task-specific encoder training. *Another challenge emanating from the high-dimensional nature of neural network encodings is their inherently "black box" nature. To overcome this, we developed methods that provide an explicit visual confirmation of the trained encoder's ability to capture the known brain cytoarchitecture* (**Figures 2B, 2C, 2D**).

For modeling the complexity of brain tissue, an encoder must have sufficient representational capacity to capture long-range spatio-molecular relationships and integrate information across multiple imaging channels in a meaningful manner. For this, we first adopted a Vision Transformer (ViT)-based encoder to provide sufficient representational capacity for multiplex imaging data across extended spatial regions, better than conventional CNN-based models such as U-Net or VAE[25]. We also found that VAEs have an unavoidable built-in Gaussian smoothness prior that limits the model's ability to detect subtle differences across region boundaries in feature space[31]. Vision Transformer models have the capacity to model details, complex long-range relationships, and subtle differences[32]. For complex brain tissues, we found that the multivariate distribution of encodings is inherently unbalanced since some brain regions are much smaller and statistically under-represented compared to much larger regions, leading to under-fitting of the rare samples. The VAEs suffer from this limitation, in addition to bias[33,34]. To overcome this limitation, we developed an iterative strategy combining k-nearest neighbor graphs and an information-theoretic community detection algorithm (InfoMap)[35] – both scalable algorithms that treat all samples equally, and ensuring that small communities are preserved (**Figure 2A**). Furthermore, our training process integrates a contrastive loss and a triplet loss function to explicitly target discriminative representation learning. It prioritizes maximizing inter-class separability while minimizing intra-class variability, thereby enhancing the handling of imbalanced data, mitigating noise from irrelevant channels, and maintaining scalability to large datasets. Importantly, our process does not require parameter settings or tuning.

Each of the MP-WSI image channels represent specific proteins that have specific biological functions individually, and in combination with other proteins that may be present or absent. Unfortunately, current encoders are designed for 3-channel color RGB images and cannot handle MP-WSI images containing dozens of image channels requiring combinatoric interpretation. Merely extending current encoders (e.g., VAE) with additional channels becomes ineffective for 5 or more channels due to noise accumulation and instability in reconstructing multiplex image data. A more meaningful and a scalable strategy is clearly needed for handling multiplex images with more channels. In this work, we developed a 'divide and conquer' strategy in which the image channels are organized into a set of panels, each containing a list of meaningfully related protein markers, and mViSE is provided with clear guidance on the combination of panels that are relevant for a given query, and a simple logical way to combine information across panels.

Below, we describe our multifaceted strategy to overcome the above-mentioned limitations, and a usable open-source software implementation in the form of a plug-in tool incorporated into the Qu-Path software that is widely used to analyze multiplex WSI images.

Methods

The detailed specimen preparation and imaging protocols, and the MP-IHC whole slide images (WSI) used here are disseminated in our earlier paper[1] so not repeated here.

Multiplex Visual Search Engine Design

mViSE is trained on tissue patches over the set of images to be analyzed, or a representative subset. We used two patch sizes. Small patches (75×75 pixels, 330nm/pixel) centered on cell nuclei include individual cell bodies, perhaps a few basal processes, or a neighbor. Larger patches (175×175 pixels) include multicellular neighborhoods with multiple cell types, local myeloarchitecture, vascular cells or capillary fragments, etc. Other patch sizes can be used, provided the same size is used for training and retrieval. Although not essential, we find it advantageous to center the larger patches on individual cells and we refer to them as cell-centered patches (CCP). This is helpful when we are interested in the neighborhoods of specific cell types of interest. CCPs require prior cell detection, but the methods are now well established[36].

Selection of Molecular Marker Panels

Since multiplex imaging data contains a mix of channels, some of which are relevant, while other are irrelevant to an intended query, we found it best to allow the user to specify the panel of relevant molecular markers. For example, glial status markers would be irrelevant for a user interested in exploring neuronal arrangements. When the multiplexing is of high order (>10), we found it advantageous to organize the data into a set of N panels of molecular markers $\{P_1, P_2, \dots P_N\}$. These panels are not set in stone – a user can define multiple customized panels and use them in combination as described below. We found panels consisting of $5 - 8$ markers to work best. Larger panels suffer from imaging noise accumulation in the multiplex encoding steps described next. For the examples presented here, we defined six panels of molecular markers to capture various facets of brain tissue, including a cell-type panel $P_1 = \{$NeuN, S100β, IBA1, Olig2, RECA1$\}$, a glial panel $P_2 = \{$S100 β, Sox2, GFAP, GLAST, Olig2, CNPase, MBP, IBA1$\}$, a myeloarchitectonic panel $P_3 = \{$NeuN, NF-H, NF-M, MAP2, Synaptophysin (SYP), CNPase, MBP$\}$, a neuronal subtypes panel $P_4 = \{$TH, Calretinin, ChAT, GLUT, PARVALB, and GAD67$\}$, cell phenotypic status panel $P_5 = \{$DAPI, Histones, PCNA, CC-3, Nestin$\}$, and a vascular panel $P_6 = \{$RECA1, Tomato Lectin, GFP, Blood Brain Barrier, Smooth Muscle Actin$\}$.

Multiplex Vision Transformer Panel Encoders

We trained an array of multiplex tissue patch encoders $f_{\theta_P}(p)$ with weight vectors θ_P for each of the marker panels P defined above, that accepts a tissue patch p as input and computes an encoding vector $v_p = f_{\theta_P}(p)$, with the expectation that a pair of visually similar patches (p_i, p_j) have a shorter distance $d\left(f_{\theta_P}(p_i), f_{\theta_P}(p_j)\right)$, compared to a pair of dissimilar patches[27], without the need for human labeling. For retrieving patches that are cytologically and/or cytoarchitecturally similar, and organizing the retrievals based on their similarity to p_q, the encoder must extract usable semantic and positional features of the relevant tissue constituents across the image channels.

The design of the encoder is crucial since it defines the effectiveness of mViSE. As noted above, a generic encoder, e.g., the variational autoencoder[24] does not guarantee the distance requirement, fails to capture long-range relationships, and lacks representational capacity for multiplex image data. To overcome these limitations, we selected and iteratively trained the array of encoders with a cluster-

contrastive learning loss function[37] in the context of a retrieval task of specific interest to us, inspired by prior work in unsupervised person re-identification [38-41]. In these systems, clustering algorithms are used to generate pseudo-labels that provide a robust supervisory signal for optimizing the encoder's parameters via backpropagation. Unfortunately, a direct application of re-identification methods is not effective for tissue similarity search in multiplex images since they are designed for Red Green Blue (RGB) images, the query patch p_q simply does not recur, and the natural biological variability across the tissue is greater than the inter-camera variations encountered in person re-identification systems. On the other hand, brain tissue exhibits characteristic patterns of cell morphology, cellular arrangements, and characteristic expression patterns of molecular markers over brain regions, suggesting that tissue patches may be amenable to clustering (grouping), although the clustering criteria can vary greatly.

We initialized the iterative encoder training with the Bidirectional Encoder representation from Image Transformers (BEiTv3) [28], a large foundational model with a strong ability to capture morphology and positional patterns, detect long-range dependencies across an image, capture complex spatial correlations with low inductive bias, better multi-task versatility and adaptability and robustness compared to CNN-based encoders[42]. A single ViT[43] encoder with shared weights is applied independently to each channel, enabling consistent feature extraction while reducing model complexity. (**Figure 2A**). However, as BEiTv3 is pre-trained on RGB images, rather than $5 - 8$ channel multiplex data, we used an adaptive Efficient Channel Attention (ECA) [44] module to dynamically emphasize the most informative channels.

Even the powerful BEITv3 encoder with the ECA is only weakly successful in capturing the brain cytoarchitecture as evident in the initial community map shown in **Figure 2B** that weakly reflects the brain tissue architecture. As noted earlier, this is because the encoder lacks a task-specific training. Our iterative training process illustrated in **Figure 2A** that is inspired by a re-identification task fills this gap by minimizing a custom two-part loss function $L = L_{con} + L_{tri}$ composed of a contrastive loss term L_{con} and the triplet loss term L_{tri} as described below. For L_{con}, we used the Cluster Noise Contrastive Estimation (ClusterNCE) loss proposed by Dai *et al*.[30] We model the known repetitive brain tissue architecture by organizing the patch data into I clusters whose centroids in feature space are $\{\varphi_1, \varphi_2, \dots \varphi_i\}$. We refer to the list of cluster centroids as the *memory dictionary* (**Figure 2A**), and cluster indices as the *pseudo labels*[30]. With this notation, L_{con} for a point q in feature space is given by:

$$L_{con} = -\log \left[\frac{exp\left(q \cdot \frac{\varphi_+}{\tau}\right)}{\sum_{i=0}^{l} exp\left(q \cdot \frac{\varphi_i}{\tau}\right)} \right], \tag{1}$$

where φ_+ is the center of the positive cluster that q belongs to, φ_i is the center of the ith cluster, "·" indicates a dot product, and τ is a temperature term that controls the sensitivity of the loss function (we used $\tau = 0.05$). We found that the cluster-contrastive loss alone fails to discriminate well in brain regions with low variability. To address this, we incorporated the triplet loss[45] L_{tri}, that fosters the learning of a discriminative feature embedding space by minimizing the distances between similar patches and maximizing the distance between dissimilar patches. It is written in terms of the hardest positive (similar) sample p^* and the hardest negative (dissimilar) sample n^*, by at least a margin m (we used $m = 0.2$) as follows:

$$L_{tri} = \max \left((d(q, p^*) + m - d(q, n^*)), 0 \right), \tag{2}$$

where $d(,)$ is the Euclidean distance. We found that when training with L_{tri} alone, the network consistently converged on two main clusters: unmyelinated and myelinated tissue, leaving it a suboptimal choice for learning similarities between other tissue types. Training with L_{con} alone produced weak associations across all tissues, and the delineation from one tissue type to another, specifically amongst cortical layers, was not sufficiently clear. However, the combination of these loss terms proved effective for mViSE, especially in discriminating regions with subtly different cytoarchitectures, e.g., the brain cortical layers.

The pseudo labels noted above are needed to compute the loss function L, and they provide the supervisory signal for optimizing the encoders within our self-supervised training framework. At each iteration, we use the current draft encoder to generate encodings for each patch. From these encodings, we first construct a weighted undirected k-nearest neighbor tree [46] (we use $k = 24$) in which the nodes are patches and the links are established for the k most proximal nodes as defined by the following proximity measure between patches p_i and p_j:

$$S(p_i, p_j) = \lambda * \text{cosSim} \left(f_{\theta_P}(p_i), f_{\theta_P}(p_j) \right) + (1 - \lambda) * d(c_i, c_j), \tag{3}$$

where cosSim is the cosine similarity measure, $d(c_i, c_j)$ is the spatial distance between the patch

centers, and $\lambda = 0.98$ is a fixed weight parameter designed to add a modest spatial proximity bias. Each

link is weighted by the corresponding similarity. The k-nearest tree forms the input to the two-level

InfoMap algorithm[35]. This generates a map of clusters (communities) of patches such that the

information theoretic distance, measured as a code length is minimal for patches within the same

cluster, and these are the pseudo labels. From the pseudo labels, we update the cluster centers φ_k using

the following momentum update strategy needed to ensure the consistency of pseudo labels from one

iteration to the next:

$$\varphi_k \leftarrow \eta\varphi_k + (1 - \eta)q, \qquad\qquad (4)$$

where $\eta = 0.2$ is the momentum coefficient. The updated cluster centers are used to compute the loss

L for performing a backpropagation update of the encoders. Pseudo labels can be displayed as color

coded images (**Figure 2**), providing a visual confirmation of the multiplex encoder's progress during

training as iterations (epochs) progress (**Figure 2B**). A successfully trained encoder reflects the brain

tissue architecture implied by the specified panel of molecular markers when we visualize the pseudo

labels (**Figure 2D**). An encoder $f_{\theta_P}(p)$ is thus trained for each panel P of chosen molecular markers.

Query Processing: Single Cell Retrieval

Once trained, the multiplex encoders $f_{\theta_P}(p)$ transform our patches into 4,608-dimensional vectors that

are used for processing queries, and we developed two strategies for this. To enable fast retrievals, we

use principal components analysis to reduce the dimensionality of these vectors to 128. When a user

indicates a query patch p_q, the first strategy simply computes the cosine similarity with every other

patch, and retrieves the N most similar patches, where N is a number set by the user, e.g., 500 (Figure

2). We found this strategy to be useful for processing small patches ($25\mu m^2$) that typically contain a

single cell. The most appropriate panel P of molecular markers for the encoder for single-cell retrieval

consists mostly of cell-type markers.

Query Processing: Similar Muli-cellular Patch Retrieval

For handling queries p_q representing larger multi-cellular patches ($58\mu m^2$), we found it best to retrieve and display the community of patches of which p_q is a member. The user can run this query using any of the trained panel-specific encoders (**Figure 2D**). This strategy is effective when working with panels of 5 – 8 molecular markers at a time.

A more sophisticated strategy is appropriate when a larger number of molecular markers, and therefore, multiple panels of molecular panels may be relevant to the query goal. The process for handling this convergence of information is illustrated in **Figure 2C** in the context of 3 panels (Cell type, Glial, and Myelo-architectonic panels). The encodings for each panel are used to form a k-nearest neighbor graph of all patches. These panel-specific graphs are combined into a single graph by including only the graph links that are common to all the graphs, i.e., a set intersection operation. The weights of the graph links are average of the weights. We then run the InfoMap algorithm to identify the resulting communities of patches. As evident in Figure 1C, the resulting community map intelligently combines information across the panels. In this case, when a user indicates a query patch p_q, we simply retrieve the community of patches which p_q is a member from the combined community map.

Experimental Results

We evaluated mViSE on the MP-IHC whole slide images (WSI) of healthy rat brains disseminated in our earlier paper[1]. We focused on healthy brains since their tissue architecture is known and mapped extensively, allowing us to visually confirm and quantify the quality of mVISE retrievals. We trained an encoder for each of the above-described panels. **Figure 2D** provides a visual confirmation of our pseudo-label maps capturing the underlying brain tissue architecture over the six panels. Note that some of the molecular markers can be present in more than one panel, and this does not pose a limitation for mViSE.

For single-cell retrieval experiments with small $25\mu m^2$ image patches, we used the cell-type panel P_1, and the neuronal subtype panel P_4. **Figure 3** illustrates the retrieval of $25\mu m^2$ image patches starting from a single query (indicated in red) and retrieving the 500 most similar patches/cells whose spatial distribution is indicated as green dots overlayed on the fitted Swansson atlas[4] to provide a spatial reference for judging the retrievals of similar cortical neurons, hippocampal neurons, RECA1+ vascular cells, IBA1+ microglia, S100β+ astrocytes, Olig2+ oligodendrocytes, vessel-proximal oligodendrocytes, vessel-proximal astrocytes, patches containing astrocytes and oligodendrocytes in close proximity,

GABAergic neurons, Glutamatergic interneurons, and cholinergic neurons. For each query, the top five retrievals are presented as close-ups for visual confirmation. The full retrieval is provided in the electronic **Supplement**.

To validate the single-cell retrieval performance, we computed the top-k retrieval scores using the results from our previous study[30] as ground truth (**Figure 5D**). Our method achieved an average Top-1 accuracy of 0.90 and a Top-5 accuracy of 0.96 across the five major cell types. Notably, the RECA score, which marks endothelial cells, was comparatively lower due to vessel-proximal structures (**Figure 3C**). Beyond cell type classification, our approach also captures multiplex morphological features, enabling accurate retrieval in challenging scenarios where small, densely packed cells are overlapping.
Next, we evaluated mViSE for its ability to delineate small brain regions using a combination of three panels $\{P_1, P_2, P_3\}$. Combining the panels resulted in a higher quality community structure as measured by the lower average code lengths, as summarized in **Figure 5E**.

The code length [47], is an information-theoretic metric used to evaluate the quality of a community structure. It quantifies the efficiency of a random walk constrained within the same community in the feature space. A lower code length indicates a better-defined community structure. Each panel's community codelength reflects the quality of its corresponding structure. Panels 1, 2 and 3 community organization exhibits to capture the global brain structure, delineate the functional regions, while Panels 4, 5 and 6 tend to capture the single cell level information. (**Figure 2D and Figure 6**). Importantly, the combination of panels yields a lower code length, and the corresponding community structure effectively captures higher-order relationships, enabling more discriminative and robust community representations, which can also be observed visually in **Figure 2C**.

Figure 4 illustrates the retrieval of similar $58\mu m^2$ multi-cellular image patches, also starting from a single query, in the RSGc, mt, IGP, and LHB regions of the rat brain. The retrieved image patches are visually similar and delineate these brain regions effectively. The full retrieval is provided in the electronic **Supplement**.

Using the same three panels, we evaluated the ability of mViSE to effectively delineate the cortical cell layers, a challenging task, generate molecular profiles of the layers, and further, to reveal the inter- and intra-layer variations. **Figure 5** shows the results for Layers L1 - L6. Each cortical layer consists of a union

of multiple communities in the combined community map. The first two rows of **Figure 5** show the individual communities for each layer in different colors. The third row shows the expression maps for each of the layers for each panel. The 4th and 5th rows provide additional details for L1. The fourth row of **Figure 5** shows the representative patches for L1, under each of the three panels. The 5th row of **Figure 5** shows the corresponding expression maps.

To assess the performance of mViSE, we manually fit the Swanson Atlas[4] to the images and calculated the confusion matrix and Intersection over Union (IoU) scores (**Figure 5D)**. Our method achieved true positive rates of 0.8 – 0.9 across distinct cortical layers, with an average IoU of 0.70. These results indicate a high level of accuracy in capturing cortical organization in a fully unsupervised manner from the multiplex data, without customized programming, driven only by visual queries. This highlights the versatility and practical usefulness of mViSE. **Figures 6** provides additional close-up illustrations of microenvironment communities for four selected tissue regions {A, B, C, D}. The bottom row of **Figure 6** shows the combined community maps for panels $\{P_1, P_2, P_3\}$ illustrating the ability of the model to capture subtle differences in tissue architecture, including small regions that are statistically under-represented. **Figure 6** shows the ability of the model to sense subtle differences in tissue architecture within brain regions (RSGC, Hippocampus).

We compared the Intersection over Union (IoU) scores of our method with state-of-the-art general vision foundation models (Beit V3 , DINO V3[48]) and pathology-specific foundation models (Prov-GigaPath[49], UNI2, Titan[50]). Because these models are designed for three-channel RGB inputs, the seven-channel images in Panel 3 were merged into RGB images using the ImageJ channel-merging strategy. The merged images were directly used for inference, and the resulting feature maps extracted from the foundation model encoders were subsequently subjected to community detection using the same settings as our method. As shown in **Table 1**, our model achieves the highest IoU scores across all cortical layers compared with the other foundation models. **Figure S-1** in the Supplement presents the full set of community maps, providing visual confirmation of the superior performance of our approach. The results demonstrate that our model captures more fine-grained information, produces clearer layer boundaries, and identifies subregions that are missed by other models. This advantage arises because our model encodes each channel independently, preserving channel-wise information that may be lost during RGB channel merging. In addition, our method incorporates explicit locational information,

further enhancing performance. Finally, the use of a self-supervised training framework contributes to the improved generalization and overall performance of our model relative to existing approaches.

Table 1: Comparison of mVISE against contemporary Foundation Models. We measured the Intersection over Union (IoU) scores of mapping accuracy over Brain Cortical Layers (L1, L2+L3, L4, L5, L6a, L6b).

Models	IoU Scores (higher is better)					
	L1	L2+L3	L4	L5	L6a	L6b
DINOV3	0.69	0.45	0.32	0.30	0.39	0.55
BEITV3	0.67	0.42	0.25	0.29	0.35	0.51
Prov-GigaPath	0.69	0.47	0.31	0.29	0.40	0.44
UNI2	0.64	0.52	0.31	0.31	0.39	0.39
Titan	0.65	0.49	0.32	0.29	0.37	0.39
mVISE(Ours)	**0.78**	**0.67**	**0.49**	**0.40**	**0.50**	**0.68**

Ablation study

Because several components of our model are specifically designed to handle multiplex images (e.g., the ECA layer), we focus our analysis on the contributions of the ViT encoder and the incorporation of locational information to overall performance. **Figure S-2** in the Supplement provides a close-up view of the community maps, demonstrating that the ViT-based encoder is more effective at capturing patch-level information compared to U-Net, a commonly used CNN-based model. In the community maps, blank regions indicate patches that are assigned as isolated nodes that do not form a community (we require a minimum of five similar patches to form a community). While the U-Net encoder performs reasonably well in certain regions, it fails to capture structures in some of the cortical layers, as evidenced by the prominent blank areas, indicating difficulty in modeling these regions. In contrast, the ViT-based encoder shows more consistent and complete community formation across cortical layers. Furthermore, locational information was incorporated during training, as described in Eq. (3). The inclusion of spatial information encourages the model to converge toward more homogeneous regions with clearly defined cortical layer boundaries.

Experimental Setup

Dataset preparation for a 16-bit, 30-channel image of 43,054 × 29,398 pixels containing 219,616 detected cells took about 1.5 hours. Half of the datasets were used for training due to a dense cell-centered sample generation strategy. The encoder was initialized from the pretrained BEiTv3-base vision encoder, containing 86M parameters. To speed up training, the first nine layers of the encoder were frozen, leaving the final three layers trainable, combined with a linear projection head and the channel attention. Model training was parallelized on two NVIDIA RTX A6000 GPUs (48 GB) with a batch size of 256 for each panel. For a 5-channel panel with 109,808 samples and a patch size of 175 × 175 × 5 pixels, training converged in 6 hours, and computation of the community structure took 7 minutes. In practice, it is best to perform the model training on a GPU cluster overnight. The results of the training are saved as a community index table (in the .csv file format) containing the community index for each cell and for each panel, and the intensities of each of the protein marker signals. This table is used to handle the user's visual queries as described below.

We developed an extension for the widely used QuPath[51] viewer to display the multiplex data, accept visual queries from the user, and display the retrievals, along with expression maps for the retrieved cells or patches. A query is handled simply by displaying the members of the community that the queried cell belongs to. This simple strategy is surprisingly effective since the retrieved items are optimal in the sense of being most proximal as measured by the information-theoretic codelength. It is also very fast (sub-second). As our results show, most retrievals also delineate a brain region or a sub-region thereof, and this is a re-assuring visual confirmation of correctness.

Recognizing that many users may not have access to a GPU cluster, we also developed a "quick search" tool that does not utilize the neural network model and can be run fully within QuPath. This tool can also be used along-side the deep model based (comprehensive) search engine. Specifically, the user can use QuPath tools to identify all cells, compute cell morphometrics, establish cell phenotypes, and compute the distribution of protein markers within an image patch centered on each cell. As illustrated in **Figure 1**, the user can select a query cell and flexibly choose among molecular markers, morphological features, and neighborhood cell markers. The system then returns the top N most similar matches (as specified by the user, not automatic) along with the corresponding similarity matrix.

Data availability

The datasets including the original acquired images from the microscope and the results of each step are publicly hosted at https://doi.org/10.7910/DVN/9H9LV9. We provide a comprehensive tutorial covering the machine learning model, and the QuPath extension on our GitHub repository: https://github.com/moxx799/mViSE. A complete dataset that can be loaded onto QuPath is posted at https://zenodo.org/records/18601179. This dataset includes: (i) A 36-channel TIF file that serves as the primary raw data source; (ii) A JavaScript extension for mViSE; (iii) Output files containing the results of the required experiments, required for mViSE QuPath extension; (iv) Annotation files manually curated to provide reference labels and validation for the experimental results; and (v) A step-by-step guide on how to use the files, detailing workflows for running the JavaScript extension, and accessing the experimental results.

Conclusions and Discussion

Viewed as a software method and a practical tool, a multiplex visual search engine can be a useful addition to the brain spatial proteomics investigators toolkit. Ongoing advances in instrumentation are fueling the generation of ever larger multiplex images capturing the intricate multiscale and multifaceted complexity of brain tissue in ever increasing spatial and molecular detail. They contain far more information than current pipeline-based image analysis tools can uncover and require computer programming to extend the capabilities. In this context, a multiplex visual search engine offers a direct visual query-driven alternative strategy for analyzing high-order multiplexed images without requiring complex computer programming. For a start, mVISE can complement current analysis methods based on delineating and identifying the phenotype of each cell and aggregating the resulting data over progressively larger spatial regions. It also opens direct image analysis strategies that are driven by visual queries that select groups of cells and multi-cellular image patches representing microenvironments of interest for quantitative profiling. Its interactive nature allows the user to quickly reject irrelevant or non-useful query retrievals and move on to more useful queries.

Visual confirmation has been an important aspect of our method design at each step. Even before a query is issued, the encoders for each panel provide visual confirmation of successfully capturing the facet of tissue architecture described by each panel. Once a visual query is issued, we get immediate visual confirmation of the retrievals on the screen. Many queries illustrated here effectively highlight

known brain regions offering a clear confirmation of correctness of the encoders and their chosen combinations. This, in turn provides confidence in the accompanying expression maps correctly reflecting the protein distributions for the retrieved community of cells or cellular neighborhoods. This also made it possible for us to quantify the quality of retrievals by fitting a brain atlas to the images and measuring how well the relevant brain region / sub-region was delineated by the retrievals. Finally, our strategy for combining two or more panels is conceptually simple, and this process too provides a visual confirmation in the form of a combined patch community map.

The vision transformer based deep neural network models that form the core of mVISE seamlessly integrate morphological and molecular information across multiple image channels, and the iterative model training process is designed to ensure that encodings of visually similar cells/patches are proximal in feature space. From a practical standpoint, it is important that the model training process is unsupervised and does not require any time-consuming and expensive human annotation efforts or tuning. Importantly, it culminates with a clear visual confirmation of successfully capturing the facet of tissue architecture described by each panel. At the next level, our divide-and-conquer strategy allows the user to specify the set of biologically meaningful molecular panels for processing a query, and the way panels are combined. This provides a very flexible, scalable, and principled control over the query-driven analysis. The user can choose to analyze as few as one facet of brain tissue architecture or a chosen combination of facets at a time. As spatial proteomics technologies continue to advance, future datasets are expected to include tens to hundreds of multiplexed channels, posing a significant challenge for how to effectively integrate and interpret the resulting high-dimensional biological information. Notwithstanding this growth, we expect our divide-and-conquer strategy to remain relevant due to its roots in the underlying tissue biology. Additionally, our feature-space community detection strategy for combining panels has proven to be robust and predictable notwithstanding its high dimensionality. We find the accompanying code length estimates to be a valuable qualitative indicator of the quality of the community structure, and we expect these data to become even more valuable in the foreseeable future.

Going forward, we foresee a growing role for query-driven image analysis strategies in spatial proteomics, and there are opportunities to build upon mVISE recognizing its limitations. First, we note that our encoders are specifically trained for the class of images of interest and for the molecular panels of interest by customizing a foundational model, rather than being foundational encoders that are pre-

trained on a comprehensive collection of datasets. In this paper, we focused on healthy brain tissue in our examples given the ease of visual confirmation, but the proposed method is equally applicable to diseased or injured brains. This will require manual ground truth data to quantify the quality of retrievals. The training of the multiplex encoders is currently not feasible at interactive speeds. In the future, we foresee the development of pre-trained foundational models that directly offer the all-important guarantee that encodings of visually similar cells/patches are proximal in feature space. Such a development will obviate our training process, and the time that it entails. In the interim, rapid advances in GPU computing and model quantization will reduce the training time to cope with hyper-plex imaging data including transcriptomic data on a large scale. Finally, we foresee opportunities to extend the search engine for handling a combination of text and imagery using vision language models, and direct *comparative* profiling of brain regions across tissue samples. In this spirit, all our methods are provided in open-source form to enable the next round of advancements.

Funding Declaration

References

1   Maric, D. *et al.* Whole-brain tissue mapping toolkit using large-scale highly multiplexed immunofluorescence imaging and deep neural networks. *Nat Commun* **12**, 1550 (2021). https://doi.org:10.1038/s41467-021-21735-x

2   Langlieb, J. *et al.* The molecular cytoarchitecture of the adult mouse brain. *Nature* **624**, 333-342 (2023). https://doi.org:10.1038/s41586-023-06818-7

3   Hawrylycz, M. J. *et al.* An anatomically comprehensive atlas of the adult human brain transcriptome. *Nature* **489**, 391-399 (2012). https://doi.org:10.1038/nature11405

4   Swanson, L. W. Brain maps 4.0-Structure of the rat brain: An open access atlas with global nervous system nomenclature ontology and flatmaps. *J Comp Neurol* **526**, 935-943 (2018). https://doi.org:10.1002/cne.24381

5   Schleicher, A., Morosan, P., Amunts, K. & Zilles, K. Quantitative architectural analysis: a new approach to cortical mapping. *J Autism Dev Disord* **39**, 1568-1581 (2009). https://doi.org:10.1007/s10803-009-0790-8

6   Paxinos, G., Furlong, T. & Watson, C. *Human brainstem: cytoarchitecture, chemoarchitecture, myeloarchitecture*. (Academic Press, 2020).

7   Lake, B. B. *et al.* Neuronal subtypes and diversity revealed by single-nucleus RNA sequencing of the human brain. *Science* **352**, 1586-1590 (2016). https://doi.org:10.1126/science.aaf1204

8   Osanai, Y., Yamazaki, R., Shinohara, Y. & Ohno, N. Heterogeneity and regulation of oligodendrocyte morphology. *Front Cell Dev Biol* **10**, 1030486 (2022). https://doi.org:10.3389/fcell.2022.1030486

9   Hennigs, J. K., Matuszcak, C., Trepel, M. & Korbelin, J. Vascular Endothelial Cells: Heterogeneity and Targeting Approaches. *Cells* **10** (2021). https://doi.org:10.3390/cells10102712

10   Costantini, I. *et al.* Large-scale, cell-resolution volumetric mapping allows layer-specific investigation of human brain cytoarchitecture. *Biomed Opt Express* **12**, 3684-3699 (2021). https://doi.org:10.1364/BOE.415555

11   Todorov, M. I. *et al.* Machine learning analysis of whole mouse brain vasculature. *Nat Methods* **17**, 442-449 (2020). https://doi.org:10.1038/s41592-020-0792-1

12   Maric, D. *et al.* Whole-brain tissue mapping toolkit using large-scale highly multiplexed immunofluorescence imaging and deep neural networks. *Nature communications* **12**, 1550 (2021).

13   Paxinos, G., Watson, C., Calabrese, E., Badea, A. & Johnson, G. A. *MRI/DTI atlas of the rat Brain*. (Academic Press, 2015).

14   Kalra, S. *et al.* Yottixel–an image search engine for large archives of histopathology whole slide images. *Medical Image Analysis* **65**, 101757 (2020).

15   Tizhoosh, H. R. & Pantanowitz, L. On image search in histopathology. *Journal of Pathology Informatics*, 100375 (2024).

16   Kaur, P. & Singh, R. K. in *2020 international conference on intelligent engineering and management (ICIEM).* 187-192 (IEEE).

17   Qayyum, A., Anwar, S. M., Awais, M. & Majid, M. Medical image retrieval using deep convolutional neural network. *Neurocomputing* **266**, 8-20 (2017).

18   Song, A. H. *et al.* Artificial intelligence for digital and computational pathology. *Nature Reviews Bioengineering* **1**, 930-949 (2023).

19   Lu, M. Y. *et al.* A visual-language foundation model for computational pathology. *Nature medicine* **30**, 863-874 (2024).

20   Adnan, M., Kalra, S., Cresswell, J. C., Taylor, G. W. & Tizhoosh, H. R. Federated learning and differential privacy for medical image analysis. *Scientific reports* **12**, 1953 (2022).

21   Wang, X. *et al.* RetCCL: Clustering-guided contrastive learning for whole-slide image retrieval. *Medical image analysis* **83**, 102645 (2023).

22   Chen, C. *et al.* Fast and scalable search of whole-slide images via self-supervised deep learning. *Nature Biomedical Engineering* **6**, 1420-1434 (2022).

23   Chen, R. J. *et al.* Towards a general-purpose foundation model for computational pathology. *Nature medicine* **30**, 850-862 (2024).

24   Kim, G. & Chun, H. Similarity-assisted variational autoencoder for nonlinear dimension reduction with application to single-cell RNA sequencing data. *BMC Bioinformatics* **24**, 432 (2023). https://doi.org:10.1186/s12859-023-05552-1

25   Pinheiro Cinelli, L., Araújo Marins, M., Barros da Silva, E. A. & Lima Netto, S. in *Variational methods for machine learning with applications to deep networks*   111-149 (Springer, 2021).

26   Aung, T. M. M. & Khan, A. A. Enhanced U-Net with Attention Mechanisms for Improved Feature Representation in Lung Nodule Segmentation. *Curr Med Imaging* (2025). https://doi.org:10.2174/0115734056386382250902064757

27   Ronneberger, O., Fischer, P. & Brox, T. in *Medical image computing and computer-assisted intervention–MICCAI 2015: 18th international conference, Munich, Germany, October 5-9, 2015, proceedings, part III 18.* 234-241 (Springer).

28   Wang, W. *et al.* in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition.* 19175-19186.

29   Wang, Y., Blei, D. & Cunningham, J. P. Posterior collapse and latent variable non-identifiability. *Advances in neural information processing systems* **34**, 5443-5455 (2021).

30   Dai, Z., Wang, G., Yuan, W., Zhu, S. & Tan, P. in *Proceedings of the Asian conference on computer vision.* 1142-1160.

31   Casale, F. P., Dalca, A., Saglietti, L., Listgarten, J. & Fusi, N. Gaussian process prior variational autoencoders. *Advances in neural information processing systems* **31** (2018).

32   Takahashi, S. *et al.* Comparison of vision transformers and convolutional neural networks in medical image analysis: A systematic review. *Journal of Medical Systems* **48**, 84 (2024).

33   Damm, S. *et al.* in *International Conference on Artificial Intelligence and Statistics.* 3931-3960 (PMLR).

34   Bredell, G., Flouris, K., Chaitanya, K., Erdil, E. & Konukoglu, E. Explicitly minimizing the blur error of variational autoencoders. *arXiv preprint arXiv:2304.05939* (2023).

35   Held, P., Krause, B. & Kruse, R. in *2016 Third European Network Intelligence Conference (ENIC).* 61-68 (IEEE).

36   Bamwenda, J., Ozerdem, M. S., Ayyildiz, O. & Akpolat, V. A Hybrid Deep Learning Framework for Accurate Cell Segmentation in Whole Slide Images Using YOLOv11, StarDist, and SAM2. *Bioengineering (Basel)* **12** (2025). https://doi.org:10.3390/bioengineering12060674

37   Wang, F. & Liu, H. in *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition.* 2495-2504.

38   Zhong, Z., Zheng, L., Luo, Z., Li, S. & Yang, Y. in *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition.* 598-607.

39   Ge, Y., Zhu, F., Chen, D. & Zhao, R. Self-paced contrastive learning with hybrid memory for domain adaptive object re-id. *Advances in neural information processing systems* **33**, 11309-11321 (2020).

40   Wang, D. & Zhang, S. in *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition.* 10981-10990.

41   Zhong, Z., Zheng, L., Cao, D. & Li, S. in *Proceedings of the IEEE conference on computer vision and pattern recognition.* 1318-1327.

42   Bourriez, N. *et al.* ChAda-ViT : Channel Adaptive Attention for Joint Representation Learning of Heterogeneous Microscopy Images. *Proc Cvpr Ieee*, 11556-11565 (2024). https://doi.org:10.1109/Cvpr52733.2024.01098

43   Dosovitskiy, A. *et al.* An image is worth 16x16 words: Transformers for image recognition at scale. *arXiv preprint arXiv:2010.11929* (2020).

44   Wang, Q. *et al.* in *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition.* 11534-11542.

45   Hermans, A., Beyer, L. & Leibe, B. In defense of the triplet loss for person re-identification. *arXiv preprint arXiv:1703.07737* (2017).

46   Johnson, J., Douze, M. & Jégou, H. Billion-scale similarity search with GPUs. *IEEE Transactions on Big Data* **7**, 535-547 (2019).

47   Rosvall, M., Axelsson, D. & Bergstrom, C. T. The map equation. *The European Physical Journal Special Topics* **178**, 13-23 (2009).

48   Siméoni, O. *et al.* Dinov3. *arXiv preprint arXiv:2508.10104* (2025).

49   Xu, H. *et al.* A whole-slide foundation model for digital pathology from real-world data. *Nature* **630**, 181-188 (2024).

50   Ding, T. *et al.* A multimodal whole-slide foundation model for pathology. *Nature medicine*, 1-13 (2025).

51   Bankhead, P. *et al.* QuPath: Open source software for digital pathology image analysis. *Scientific reports* **7**, 1-7 (2017).
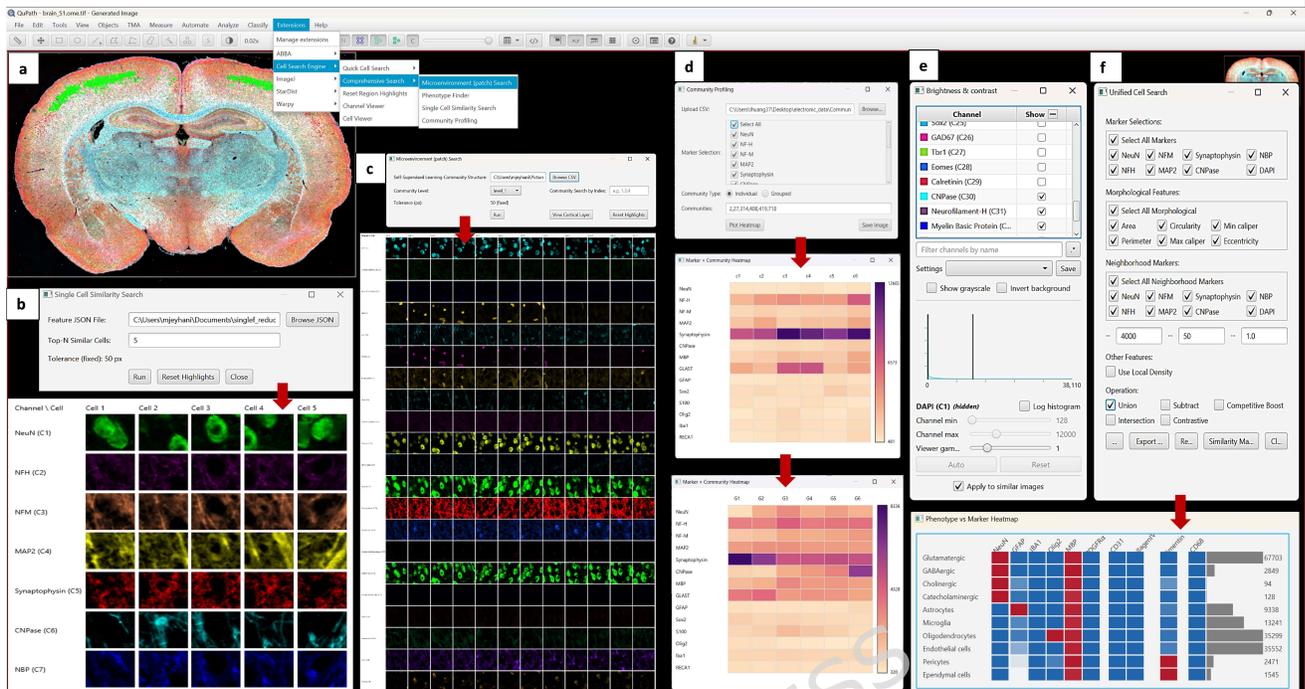
Fig. 1. mViSE is a visual search engine for profiling multiplex whole brain slice images interactively, without programming. (a) Search queries are submitted by clicking on a chosen cell in the viewing window, upon which similar other cells and/or similar cell neighborhoods are retrieved and displayed as green dots overlaid on the multiplex WSI image. (b) The single-cell similarity search retrieves the top N matches with a 75×75 pixel window size. (c) The multicellular (175×175 pixel patch) search retrieves similar patches. (d) Community profiling displays mean marker expression data for a chosen set of communities (individually or grouped) and markers. (e) Channel selection panel for viewer. (f) Unified Cell Search identifies cells similar to a query cell by combining selectable molecular markers, morphological descriptors, and neighborhood profiles, and exports phenotype and molecular marker expression heatmap.
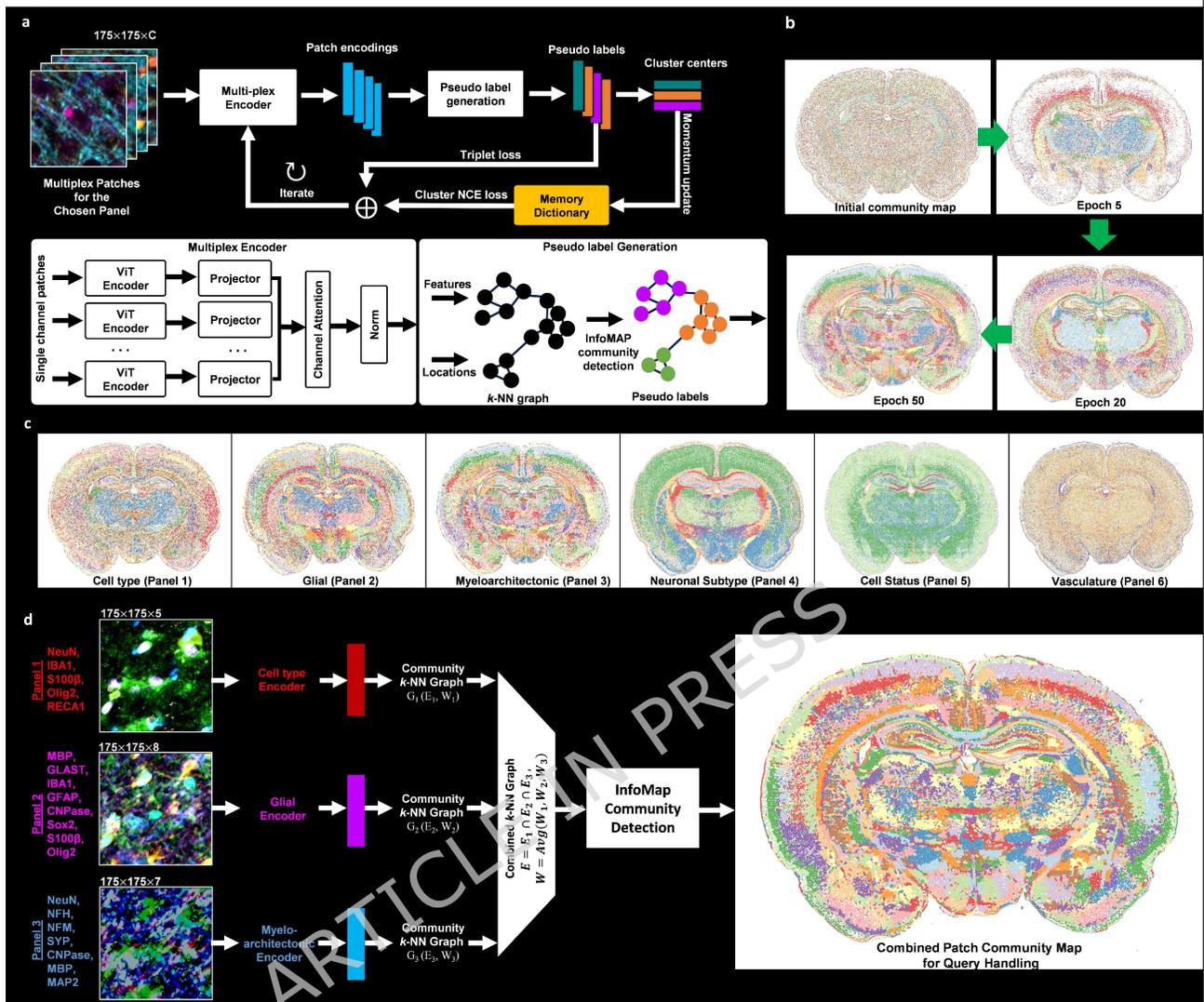
Fig. 2

Fig. 2. **Illustrating the self-supervised learning and query handling methods**. (a) The self-supervised model architecture. The encoder is designed to process a multiplex image patch of size 175×175 pixels with C channels. (b) Pseudo labels provide a visual confirmation of training progress. The initial community map generated by the BEITv3 foundation model encoder for Panel 3 barely captures the brain tissue structure. During 50 epochs of task-specific training, the community map clearly and gradually reflects the brain tissue structure and is ready for handling queries. (c) The learned community structure for 6 different panels. (d) The method for logically combining three different panels with 15 channels altogether to generate a combined community map for handling sophisticated queries.
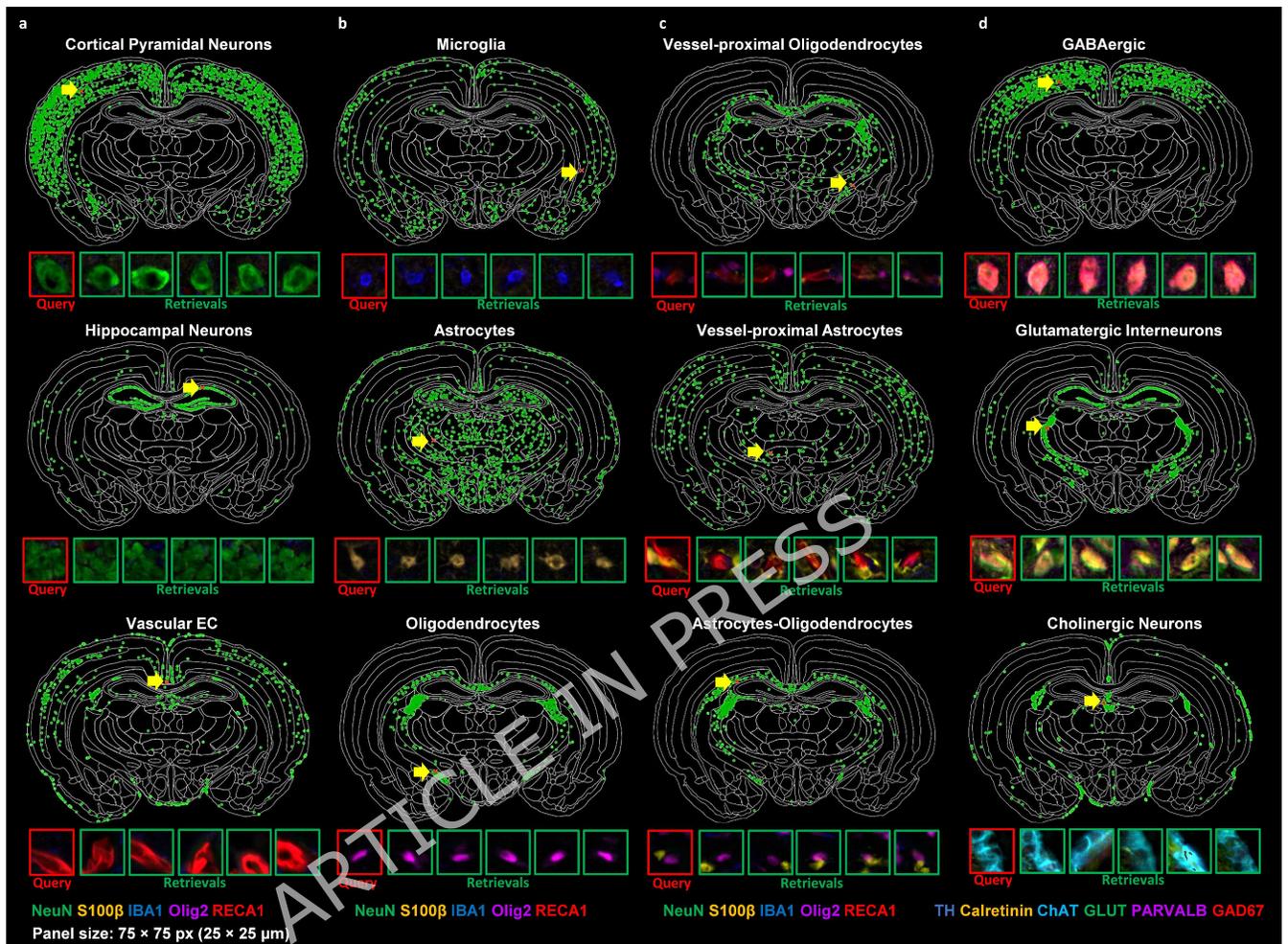
Fig. 3

Fig. 3. **Retrievals of single cell types, neuronal subtypes, and proximal cell patterns.** The queried cell is indicated with a red cross at the tip of a yellow arrow. The green dots show the spatial locations of the top 500 retrieved cells overlaid on the Swanson Atlas. The applicable panel of markers is displayed on the bottom. For each example, a close-up of the queried cell is displayed within a red box, and the top five retrievals are displayed within green boxes, providing a visual confirmation. (a, b) Retrievals for major cell types, including cortical pyramidal neurons, hippocampal neurons, vascular endothelial cells, microglia, astrocytes, and oligodendrocytes, and their distribution. (c) Showing our search engine's ability to detect vessel-proximal oligodendrocytes and astrocytes, and proximal cell pairs comprising astrocytes and oligodendrocytes. (d) Retrievals of neuronal subtypes, highlighting the search engine's capacity to retrieve GABAergic and glutamatergic interneurons, and cholinergic neurons.

Fig. 4

Panel 1: NeuN S100β IBA1 Olig2 RECA1
Panel 2: S100β Sox2 GFAP GLAST Olig2 CNPase MBP IBA1
Panel 3: NeuN NF-H NF-M MAP2 Synaptophysin CNPase MBP
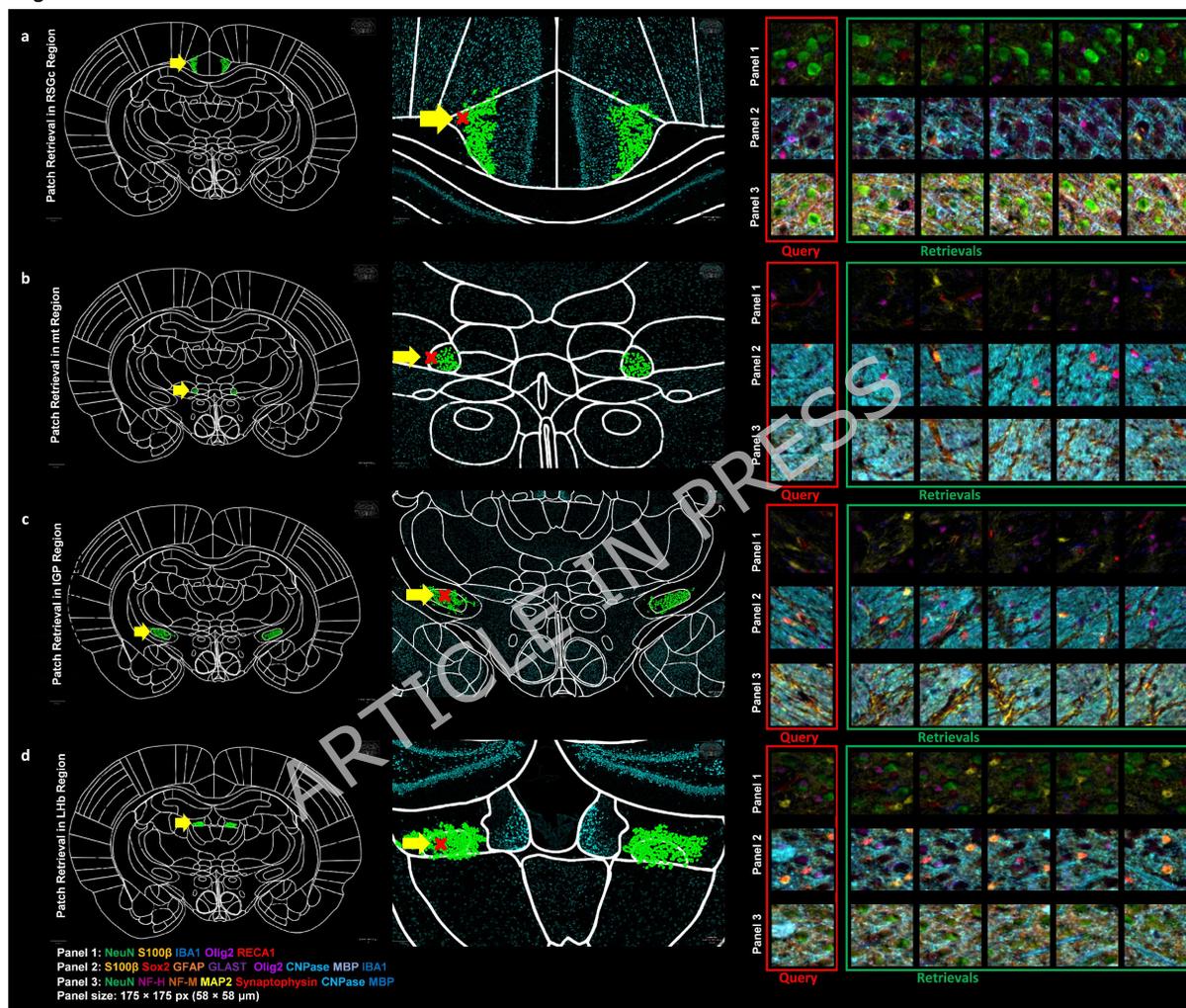Panel size: 175 × 175 px (58 × 58 μm)

Fig. 4. Sample retrievals of similar cellular neighborhoods (patches) across four small brain regions.

The red crosses pointed by the yellow arrows indicates the query patches, while the green dots retrieved patches within the same community overlaid on a fitted Paxinos Atlas. The community searches are self-limiting, so the user does not need to provide the number of expected retrievals. The query patches and a set of randomly selected five retrievals are displayed to the right, comprising Panel 1 (cell types with 5 markers), Panel 2 (glial components with 8 channels), and Panel 3 (myeloarchitecture with 7 markers). The retrieved patches are highlighted in green boxes alongside random samples from

the same community. Using a single query patch encompassing 175 × 175 patch with 15 channels, the search engine identifies similar patches with results aligning closely and reassuringly to the corresponding brain regions in the fitted atlas, including the retrosplenial granular cortex (RSGc), mammillothalamic tract (mt), indusium griseum posterior (IGP), and the lateral habenula (LHb).
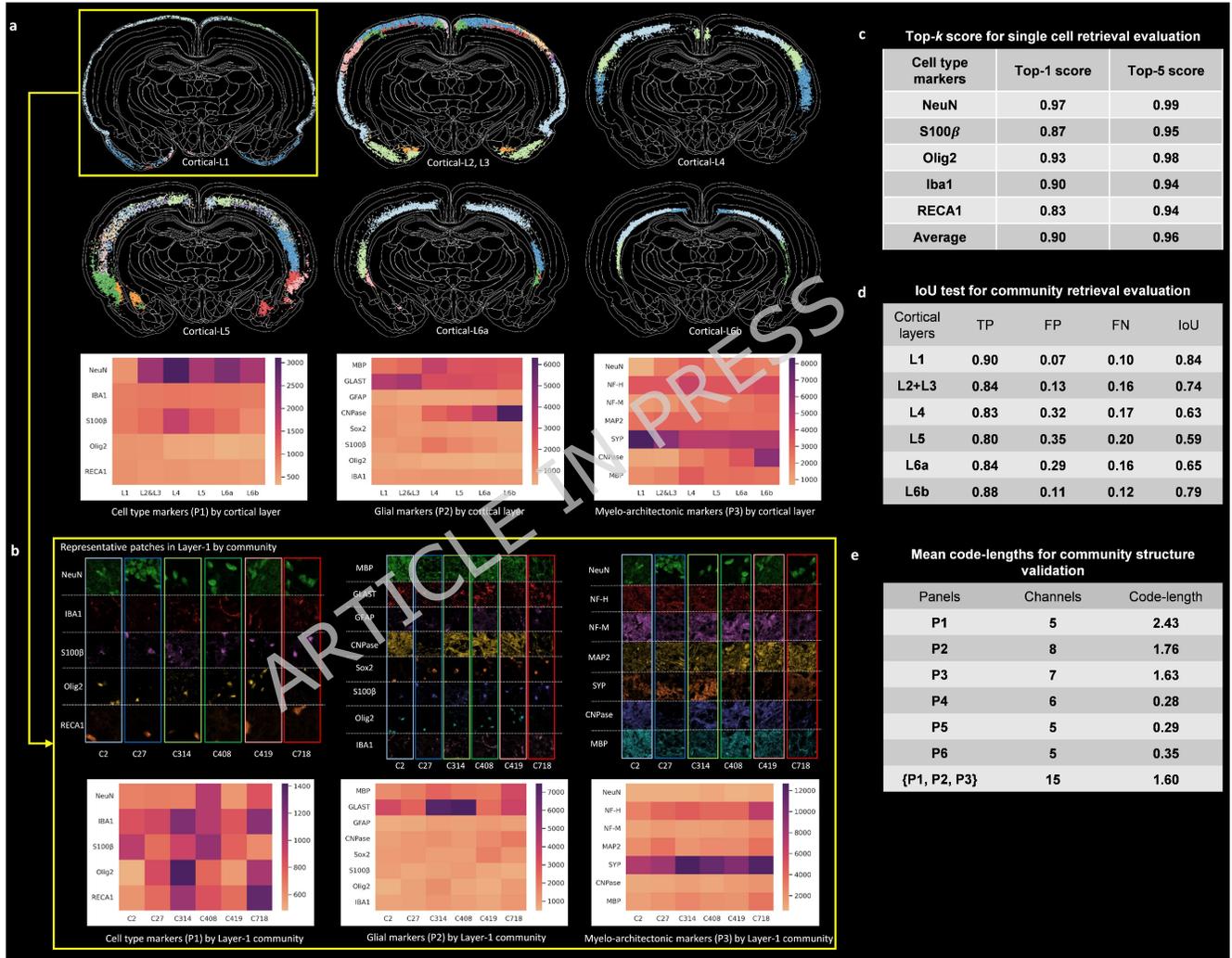
Fig. 5

Fig. 5. **Illustrating the retrieval of cortical cell layers and profiling intra-layer tissue heterogeneity, with validation**. Multiple communities are merged to form a cortical layer due to the heterogeneity of each region. (a) The combined communities delineating cortical layers 1 through 6. Each of the constituent communities is displayed in a distinct color. For each layer, we profile the markers across the cell type, glial, and myeloarchitectural panels. (b) An expanded profiling of cortical layer 1, which comprises six communities (indices 2, 27, 314, 408, 419, and 718). We present the most representative patches (175

×175 pixels) for each community across three panels, accompanied by an expression map for the corresponding markers. (c) Single cell retrieval for main cell type makers is validated by the top-1 and top-5 score respectively. (d) The IoU metric captures how well the retrievals match the atlas. (e) The mean code length is an indicator of the cohesiveness of a community's structure (smaller numbers are better). Note that combining Panels 1, 2, and 3 resulted in a more cohesive community structure compared to any of the individual panels.
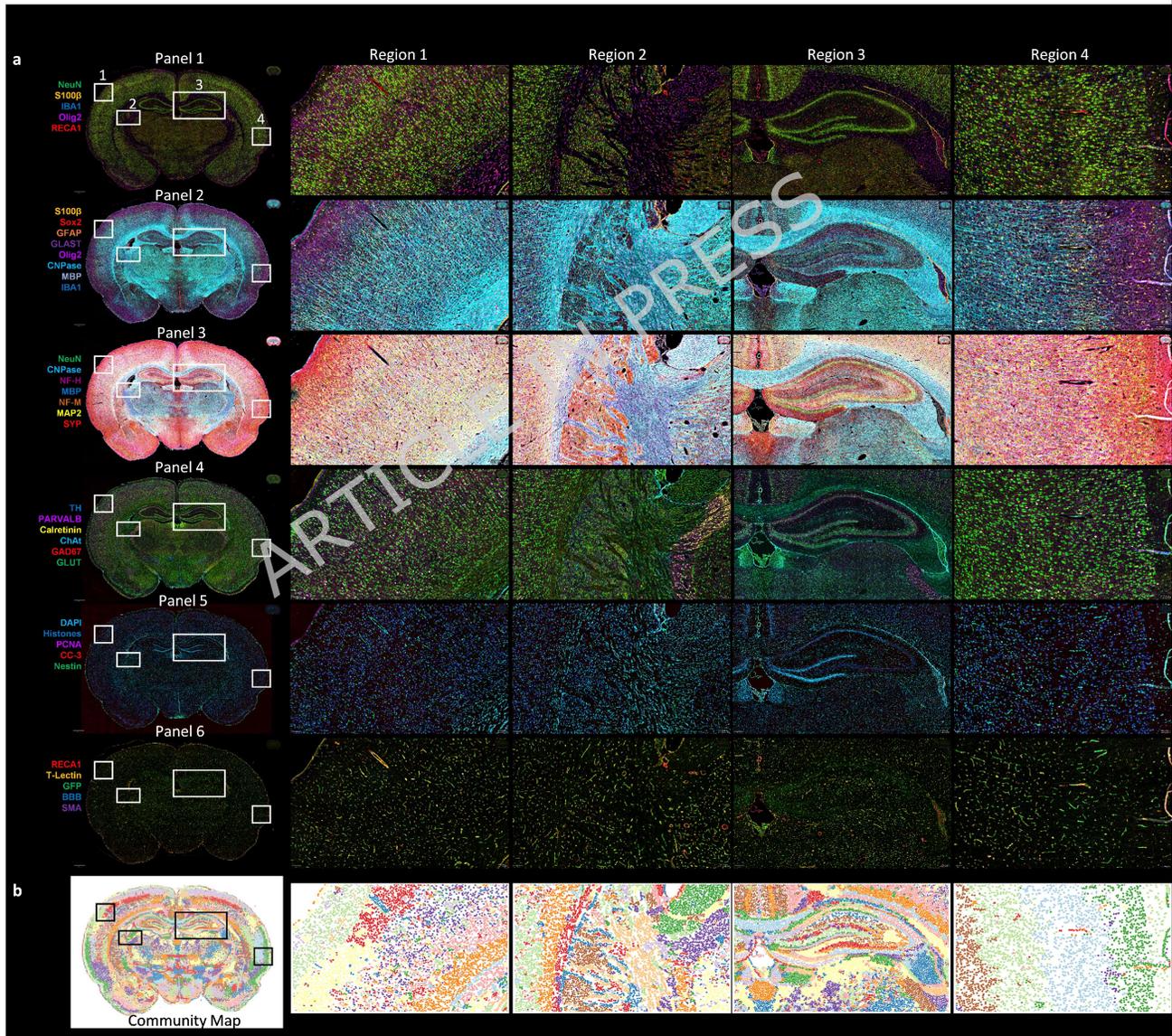
Fig. 6

Fig. 6. **Learned community map captures multiplex features and delineates tissue heterogeneity.** (a) High-plex imaging reveals similarities and differences across brain regions differently depending on the chosen panels, and show the brain tissue's complexity, including variability in size, shape, texture, and differences in protein expression. (b) The community maps provide a visual confirmation of the encoder's ability to capture complex patterns, from large-scale spatial patterns to localized regions, and to detect subtle differences such as cortical layer structure in Region 4.