Morphology-Aware Profiling of Highly Multiplexed Tissue Images using Variational Autoencoders

Gregory J. Baker[1,2,3,&,*,#], Edward Novikov[1,4,*], Shannon Coy[1,2,5], Yu-An Chen[1,2], Clemens B. Hug[1], Zergham Ahmed[1,4], Sebastián A. Cajas Ordóñez[4], Siyu Huang[4,%], Clarence Yapp[1], Gaurav N. Joshi[6], Fumiki Yanagawa[6], Artem Sokolov[1], Hanspeter Pfister[4], Sandro Santagata[1,2,3,5], and Peter K. Sorger[1,2,3,#]

[1]Laboratory of Systems Pharmacology, Harvard Medical School, Boston, MA

[2]Ludwig Center for Cancer Research at Harvard, Harvard Medical School, Boston, MA

[3]Department of Systems Biology, Harvard Medical School, Boston, MA

[4]Harvard John A. Paulson School of Engineering and Applied Sciences, Harvard University, Cambridge, MA

[5]Department of Pathology, Brigham and Women's Hospital, Harvard Medical School, Boston, MA

[6]Nikon Instruments, Lexington, MA

[&]Current affiliation: Division of Oncological Sciences, Knight Cancer Institute, Oregon Health & Science University, Portland, OR

[%]Current affiliation: Visual Computing Division, School of Computing, Clemson University, Clemson, SC

*Co-first author; [#]Corresponding author

Key Words: spatial proteomics, variational autoencoder (VAE), artificial intelligence, CyCIF, MxIF, cancer

Corresponding Authors:

Gregory J. Baker, Ph.D., Pharm.D.

Laboratory of Systems Pharmacology

Harvard Medical School

Warren Alpert Building, Room 444

200 Longwood Avenue, Boston, MA 02115

Tel: 617-432-6909

Email: gregory_baker2@hms.harvard.edu

Peter K. Sorger, Ph.D.

Laboratory of Systems Pharmacology

Harvard Medical School

Armenise Building, Room 137F

200 Longwood Avenue, Boston, MA 02115

Tel: (617) 432-6081

Email: peter_sorger@hms.harvard.edu

ORCiDs:

Gregory J. Baker - 0000-0002-5196-3961

Edward Novikov - 0000-0003-1476-5111

Shannon Coy - 0000-0003-0033-9031

Yu-An Chen - 0000-0001-7228-4696

Clemens B. Hug - 0000-0002-8299-3274

Zergham Ahmed - 0000-0001-8368-0408

Sebastián A. Cajas Ordóñez - 0000-0003-0579-6178

Siyu Huang - 0000-0002-2929-0115

Clarence Yapp - 0000-0003-1144-5710

Artem Sokolov - 0000-0002-8056-0504

Hanspeter Pfister - 0000-0002-3620-2582

Sandro Santagata - 0000-0002-7528-9668

Peter K. Sorger - 0000-0002-3364-1838

Prefix definitions:

S: denotes a segmentation-based cluster

V_9: denotes a 9x9µm image patch cluster (single-tissue analysis)

V_{9M}: denotes a 9x9µm image patch cluster (multi-tissue analysis)

V_{30}: denoted a 30x30µm image patch cluster (single-tissue analysis)

Abstract Word Count: 148/150

Introduction, Results, and Discussion Word Count: 5,000/5,000

ABSTRACT

Spatial proteomics (highly multiplexed tissue imaging) provides unprecedented insight into the types, states, and spatial organization of cells within preserved tissue environments. To enable single-cell analysis, high-plex images are typically segmented using algorithms that assign marker signals to individual cells. However, conventional segmentation is often imprecise and susceptible to signal spillover between adjacent cells, interfering with accurate cell type identification. Segmentation-based methods also fail to capture the morphological detail that histopathologists rely on for disease diagnosis and staging. Here, we present a method that combines unsupervised, pixel-level machine learning using autoencoders with traditional segmentation to generate single-cell data that captures information on protein abundance, morphology, and local neighborhood in a manner analogous to human experts while overcoming signal spillover. We demonstrate the generality of this approach by applying it to CyCIF, Lunaphore COMET, and CODEX data, and show that it can learn histological features across multiple spatial scales.

INTRODUCTION

High-plex tissue imaging (spatial proteomics) has emerged to study the biology of intact tissues and tumors by measuring the abundances and spatial distributions of tens to hundreds of biomolecules using fluorescently or metal-tagged antibodies[1]. Single-cell data are most commonly derived from these images by segmentation. This involves computing masks centered on nuclei that divide an image into individual cells. Staining intensities are then computed for each antibody to generate a *cell feature table*[2] describing single-cell protein expression levels. A strength of this approach is that it facilitates multi-omic data integration by generating tabular data similar to flow cytometry data or count tables in single-cell transcriptional profiling[2] .

However, segmentation of high-plex tissue images faces three fundamental limitations: spatial crosstalk (or lateral spillover), under and over segmentation, and insensitivity to morphology. The former arises when signals from neighboring cells cross-contaminate each other because they cross segmentation boundaries[3], which confounds cell type identification and other analyses.[3,4] Recent high-plex, high-resolution, 3D imaging of tissues[5] has shown that cells have membrane-membrane distances <1 μm, their processes are often intertwined[6], and two or more cells commonly overlap along the imaging (Z) axis. These observations explain the inability of even state-of-the-art algorithms such as CellPose[7] (and its recent updates)[8] to fully overcome spatial crosstalk or related problems such as over-segmentation (subdividing cells into multiple objects) or under-segmentation (combining different cells into a single object).

Differences in cell morphology play a major role in classical histopathology, most commonly based on human inspection of hematoxylin and eosin (H&E) stained slides, but this is another area in which single cell analysis based on tissue segmentation falls short. Differences in cell morphology can be detected using supervised and unsupervised machine learning (ML) from pixel-level image data, for example using computational pathology (CPath) AI/ML approaches.[9] CPath methods use convolution neural networks, transformers, and other deep learning approaches to classify histology images by tissue type, disease state, or treatment response.[10] Pixel-level learning has also been applied to multiplexed tissue images using convolutional and transformer-based neural networks, CNN-based classifiers[11–14], generative models[15,16], and vision transformers[17,18], and unsupervised clustering of image features[19]. Supervised learning on multiplexed image data shows promise[18] but it is unclear how model outputs compare to single cell data generated using traditional segmentation methods.

In this paper we describe an approach to spatial profiling that combines image segmentation and pixel-level ML using light-weight generative models known as variational autoencoders (VAEs)[20,21]; we then compare the results to what can be obtained using optimized segmentation approaches alone. We compile these complementary methods in the MORPHAEUS toolbox to facilitate reuse and further development. We find

that VAEs can be trained to analyze high-plex images in a fully unsupervised manner using single-cell or multicellular image patches. Clustering of VAE latent space manifolds then enables the identification and cataloging of phenotypes at both the single-cell and tissue levels, allowing the model to learn histological features across spatial scales, from single cells to recurrent cell assemblies.

RESULTS

Spatial crosstalk limits accuracy of segmentation-based cell phenotyping

Image segmentation and pixel-level learning are agnostic to how images are acquired and both methods have been applied to a wide variety of fluorescence-based spatial profiling technologies including CODEX[23], COMET (Lunaphore)[24], cyclic immunofluorescence (CyCIF)[25] and other spatial proteomic methods[1]. In this paper we focus on CyCIF datasets collected as reference data[26] for the Human Tumor Atlas Network (HTAN)[27] and demonstrate that the approach generalizes to COMET and CODEX datasets.

Analysis of CyCIF images started with a 22-plex whole-slide CRC dataset previously released by HTAN (HTAN ID: HTA13_1_101) and referred to here as CyCIF-1A (**Fig. 1a** and **Supplementary Table 1**). Algorithms in MCMICRO[28] were used to perform image assembly and segmentation. An H&E image from a tissue section adjacent to CyCIF-1A (HTAN ID: HTA13_1_100) was annotated by two board-certified pathologists and found to comprise 11 distinct histological regions (HRs) of diagnostic or biological significance such as normal mucosa (the crypt-containing epithelial layer of the colon), smooth muscle and stromal tissue (**Fig. 1b,c**) as well as multiple adenocarcinoma (AC) morphologies including glandular, mucinous, and solid.[29]

To generate single-cell data from these images, per-cell mean fluorescence intensities were computed on segmented cells using an object mask appropriate for each protein and corresponding to the cell nucleus, cytoplasm, or plasma membrane (e.g. a plasma membrane mask for CD4 and CD8, and nuclear mask for PCNA; **Supplementary Table 2**). The resulting single-cell feature table underwent quality control (QC) with CyLinter[30] to remove visual and image processing artifacts (affecting ~21% of total cells; **Extended Data Fig. 1a**). For an extended discussion of segmentation-based tissue analysis see (**Supplementary Note 1** and **Supplementary Fig. 1**).

Clustering of post-QC segmented data using Leiden community detection[31] yielded 18 clusters (S0-17, with "S" denoting segmentation-based). Clusters ranged in size from 0.8% to 12.6% of total cells (**Supplementary Table 3**). Horn's parallel analysis[32] showed that the first three principal components (PCs) captured the majority (68%) of dataset variance; single cell data were therefore visualized in 3D UMAP[33] embeddings (**Fig. 1d**, **Extended Data Fig. 1b** and **Supplementary Video 1**). Hierarchical clustering (using mean, z-score-normalized, signal intensities) revealed four major cell types occupying characteristic regions of tissue: keratin$^+$ epithelium, αSMA$^+$ smooth muscle, vimentin$^+$ stroma, and CD45$^+$ immune cells (**Fig. 1e,f**, **Extended Data Fig. 1c-e** and **Supplementary Fig. 2**). Clusters S0, 1, 2, 8, and 10 represented four tumor cell subsets that differed in PCNA expression (**Extended Data Fig. 1f**) but did not correspond directly to glandular, mucinous, and solid AC histologies (**Fig. 1b,c and f**). Normal epithelial cells comprised the tumor-adjacent

139 colonic mucosa (S14, HR1) and benign reactive mesothelium in the serosa (S17, R3), the latter likely arising
140 from tumor-associated inflammation. Smooth muscle occupied the muscularis propria (S4, HR2), while
141 fibroblasts (S6) were identified in the subserosa (HR4), peritumoral stroma (HR9), and submucosa (HR10).
142 Endothelial cells (S7) were present throughout the tissue, as were immune cells, including populations of
143 macrophages (S3), $CD8^+$ T cells (S5), and both conventional (S13) and regulatory (S15) $CD4^+$ T cells
144 (**Extended Data Fig. 2** and **Supplementary Fig. 3**).

145 Despite the biological plausibility of these cell phenotypes, closer examination revealed multiple
146 clusters with unlikely immunomarker combinations. For example, cells in S12 appeared to express markers of
147 proliferating epithelial cells (keratin, ECAD, PCNA) as well as $CD8^+$ T cells (CD45, CD3ε, CD8α, CD45RO;
148 **Fig. 1e**). This was not a consequence of suboptimal clustering, since manual gating on keratin and CD8α
149 channels showed that >80% of segmented cells in S12 were positive for both markers (**Fig. 1g,h**) and
150 inspection of the underlying image showed that they corresponded to $CD8^+$ T cells immediately adjacent to
151 $keratin^+$ tumor cells (**Fig. 1i**). We conclude that S12 cells represent $CD8^+$ T cells affected by spatial crosstalk
152 from tumor cells (**Supplementary Fig. 4**).

153 Clusters S11 and S16 were also affected by spatial crosstalk. Both clusters represented tightly packed B
154 and T cells in lymphoid aggregates (HR*), with S16 cells found exclusively in the serosa (HR3) and subserosa
155 (HR4). Non-tumor infiltrating $CD8^+$ T cells in S5 were impacted by spatial crosstalk as well, but this cell type
156 was distributed across multiple additional subclusters in UMAP space. A detailed investigation of the
157 relationship between spatial crosstalk and dimensionally reduced embeddings (UMAP, tSNE) showed that
158 spatial crosstalk: (i) was pervasive, affecting up to 30% of the ~1M cells in CyCIF-1A; (ii) promoted formation
159 of non-biological clusters due to distinctive (yet artifactual) marker combinations and (iii) ii) represented a
160 significant barrier to accurate cell phenotyping (**Supplementary Note 2** and **Supplementary Figs. 5 & 6**)
161

162 **VAE classification of human CRC at single-cell resolution**

163 To analyze the same CyCIF data at the pixel level we developed a method for training VAE models on
164 image patches derived from whole-slide data. This approach used Cellcutter[34] to generate 9x9µm (14x14 pixel)
165 multi-channel patches centered on individual nuclei (as determined by image segmentation) and store them in a
166 memory-efficient Zarr file (**Fig. 2a,b** and **Methods**). After subsampling, the final dataset for CyCIF-1A
167 consisted of 479,909 patches, 86% with one and 13% with two nuclei (**Fig. 2c**). Image patch pre-processing
168 involved: (i) log transformation, (ii) clipping to remove bright and dim outlier pixels, (iii) per-channel
169 normalization (0-1) based on fluorescence intensities from the whole-slide image, and (iv) rotational data
170 augmentation to reduce sensitivity to cell orientation (**Extended Data Fig. 3a,b**). Patches were divided into

171 training, validation, and test sets at a ratio of 80:10:10 and the first set used to train a standard VAE network
172 consisting of four convolutional encoding layers followed by a dense layer (**Fig. 2d**, **Extended Data Fig. 3c**,
173 and **Methods**). To focus the attention of the network on the cell at the center of the patch, each channel was
174 convolved with a 2D Gaussian kernel having a standard deviation of 7 pixels. This vignetting technique
175 reduced the intensity of features towards patch borders, causing the network to perceive them with less acuity
176 (42% of maximum intensity at patch corners; **Fig. 2e**).

177 A VAE treats individual image pixels as independent features (or dimensions). Thus, 9x9µm image
178 patches from our 21-plex CRC image comprised 4,116 total dimensions (14 pixels x 14 pixels x 21 channels).
179 The VAE latent space dimension (an adjustable parameter influencing model accuracy) was set to 10% of total
180 image patch dimensionality, resulting in a 412-dimensional latent space. Visual examination of learned VAE
181 reconstructions confirmed accurate capture of marker location and intensity and exclusion of idiosyncratic
182 differences among cells of the same type (**Fig. 2 f, g**). At convergence, the model achieved a validation loss of
183 8.3×10^{-4}.

184 Leiden clustering of VAE image patch encodings generated 24 clusters ($V_9$0-23, with "V_9" denoting a
185 VAE trained on 9x9µm patches (**Fig. 2h**, **Supplementary Table 4**, **Supplementary Video 2**, and
186 **Supplementary Image Gallery 1**) and these corresponded to the primary lineages identified by segmentation-
187 based analysis: keratin$^+$ epithelium, αSMA$^+$ smooth muscle, vimentin$^+$ stroma, and CD45$^+$ immune cells
188 (**Extended Data Fig. 4a,b** and **Supplementary Fig. 7**). Primary clusters could be grouped into 10 meta-
189 clusters (A-J; **Fig. 2i**) by hierarchical clustering based on median, z-score normalized, pixel-level marker
190 intensities. To evaluate the agreement between VAE and segmentation-based cluster pairs, we used the Jaccard
191 index (JI). Six clusters (29% of total cells) exhibited substantial overlap (JI ≥ 0.44). These comprised Tregs
192 ($V_9$4/S15), smooth muscle cells ($V_9$8/S4), normal crypt-forming epithelial cells ($V_9$10/S14), lymphoid
193 aggregates ($V_9$11/S11), CD8$^+$ T cells ($V_9$13/S5), and mesothelial cells ($V_9$19/S17; **Extended Data Fig. 4c,d**,
194 **Extended Data Fig. 5** and **Supplementary Fig. 8**). Since segmentation-based clustering is based on marker
195 intensity, we surmised that formation of these VAE clusters was also driven primarily by differences in marker
196 expression. To demonstrate that VAE-based cell state inference generalizes across spatial proteomics platforms
197 despite differences in image bit depth, resolution, and tissue type, we performed similar analysis of images of
198 non-small cell lung cancer collected using COMET and head & neck squamous cell carcinoma collected using
199 CODEX (**Supplementary Note 3**, **Extended Data Fig. 6**, and **Supplementary Figs. 9 & 10**).
200

201 **VAE image patch classification is morphology-aware and overcomes spatial crosstalk**

202 VAE-based learning identified seven tumor cell clusters in CyCIF CRC data (meta-cluster H; **Fig. 2i**),
203 compared to five with segmentation. While marker expression profiles for tumor-associated segmentation and
204 VAE clusters were similar (**Fig. 3a**), only VAE-based clusters mapped to distinct physical regions of the tumor
205 (**Fig. 3b,c**, **Extended Data Fig. 7a** and **Fig. 1b,c**). To quantify this correspondence, we calculated normalized
206 mean posterior probabilities for VAE clusters relative to four previously developed kNN classifiers trained to
207 predict normal, glandular, mucinous, or solid AC histologies from CyCIF data[35]. Cells in clusters $V_9$0, 3, 5, 21,
208 and 22 showed high mean probabilities of glandular AC (≥71%), while those in clusters $V_9$6 and $V_9$1 had the
209 highest probabilities of mucinous (54%) and solid AC (47%), respectively. These results suggested that the
210 VAE learned to distinguish tumor cells with different morphologies, and that mucinous and solid AC subtypes
211 exhibit greater morphological variability than glandular AC (**Fig. 3d** and **Extended Data Fig. 7b**). To quantify
212 the degree of uncertainty in AC subtype assignment, we calculated the information entropy of VAE clusters
213 relative to their kNN class probabilities (**Fig. 3d**, red dotted line). Coloring clusters by entropy scores and
214 mapping them back onto CyCIF-1A, we observed a distinctive pattern, with lower entropy glandular AC
215 appearing at the left of the tissue and higher entropy mucinous and solid AC at the right (**Fig. 3e**). This result
216 revealed a spatial gradient of tumor dedifferentiation across the specimen. Inspection by pathologists confirmed
217 that the VAE had correctly learned differences in tumor phenotypes in an unsupervised manner and was
218 superior to kNN classifiers generated by supervised training.

219 To determine the contribution of morphology to tumor-associated VAE clusters, we isolated the keratin
220 channel of patches from meta-cluster H to and binarized them to remove intensity variation; we then trained a
221 new VAE model. We found that cells with mucinous ($V_9$6) and glandular ($V_9$0, 3, and 5) morphologies
222 occupied distinct regions in feature space and exhibited differences in keratin organization (**Fig. 3f,g**), although
223 this was not true for solid AC ($V_9$1). Thus, the shape of the keratin signal alone was sufficient to distinguish
224 glandular from mucinous cells. This finding, combined with the marker intensity map for V_9 clusters (**Fig. 3a**,
225 right panel) as well as spatial gradients observed in cluster marker intensity (**Fig. 3h**) and tissue morphology
226 (**Fig. 3i**), strongly suggest that the VAE network simultaneously captures intensity and morphology
227 information when distinguishing tumor cell subtypes.

228

229 **Identifying factors driving VAE cluster membership**

230 $V_9$22 cancer cells were found throughout the tumor and were consistently surrounded by $CD8^+$ T cells
231 from cluster $V_9$18 (**Fig. 3j**). The Jaccard index plot (**Extended Data Fig. 4c**) showed that these VAE clusters
232 were combined in segmentation cluster S12. Thus, VAE-based learning correctly resolved $CD8^+$ T cells and
233 tumor cells affected by spatial crosstalk. To investigate how this occurred, $V_9$18 and $V_9$22 image patches were

compared at a pixel level using mean squared error (MSE) analysis (**Fig. 3k**). This showed that variance between $V_9$18 and $V_9$22 patches was significantly greater than variance within $V_9$18 patches for channels CD3ε, CD45, CD8α, and PCNA. Visualizing average MSE representations for each cluster showed that CD3ε, CD45, and CD8α signals were concentrated at the center of $V_9$18 patches but the periphery of $V_9$22 patches (**Fig. 3l**), while PCNA signals exhibited the opposite pattern. This result was confirmed by inspection of image patch galleries for $V_9$18 and $V_9$22 (**Fig. 3m**). $CD8^+$ T cells were also found in $V_9$13, but these cells were adjacent to different stromal and immune cell types in the lamina propria of normal epithelium (HR1), muscularis propria (HR2), and peritumoral stroma (HR9; **Fig. 3n**). We conclude that a VAE can distinguish adjacent cell types affected by spatial crosstalk in segmentation-based analysis by correctly learning different spatial patterns of protein expression and can further differentiate similar cell types (e.g., $CD8^+$ T cells) based on tissue context.

To further study how patch features influenced cluster membership, we performed concept saliency analysis[22]—a method in interpretable AI that seeks to explain the features driving differences in ML classes. Archetypal image patch encodings or "concept vectors" (denoted, *Zc*) were first selected for each cluster based on maximum cosine similarity to the average encoding for all patches bearing the same cluster label. Individual patches were then scored for similarity to cluster archetypes by computing the dot product between the VAE encoding for a particular image patch (*Zi*) and the previously identified concept vector ($Sc = Zi \cdot Zc$). This showed that image patches with and without a specific cluster label formed discrete concept score distributions (**Extended Data Fig. 7c**) and that the VAE identified groups of image patches with unique spatial features. We then generated concept saliency maps using the rectified gradient method[36], which uses backpropagation to calculate pixel-wise contributions to a given concept score while minimizing noise (**Fig. 3o**, **Extended Data Fig. 7d** and **Supplementary Image Gallery 2**). In these maps, pixels with high attribution values corresponded to signals consistently identified with specific clusters. For example, in the case of $CD68^+$ $CD163^-$ $PD\text{-}L1^+$ macrophages (cluster $V_9$14), the CD68 and PD-L1 channels showed positive attribution values in regions of the patch with high fluorescence signals, whereas biologically unrelated signals from CD163 and CD8α had negative values (**Fig. 3o**). These results provide further evidence that intensity signals and spatial distributions learned in an unsupervised manner by a VAE are interpretable and biologically meaningful.

Patch-based learning achieves greater precision relative to image segmentation

The Jaccard index plot showed that segmentation-based clusters were often subdivided by the VAE into multiple subclusters, suggesting higher precision in distinguishing cell types (**Extended Data Fig. 4c**). For example, within segmentation cluster S3 (annotated as macrophages) the VAE identified three phenotypically

and spatially distinct cell types: (i) CD163$^+$ CD68$^+$ tumor-reactive macrophages in the peritumoral stroma ($V_9$7, HR9), (ii) CD163$^+$ tissue-resident macrophages in the muscularis propria ($V_9$12, HR2), and (iii) CD68$^+$ PD-L1$^+$ tumor-associated macrophages (TAMs) infiltrating the tumor mass ($V_9$14, HR5-7; **Fig. 4a**). S3 also contained multiple non-macrophage cells populations, including CD31$^+$ vascular endothelial cells ($V_9$2, 23.1% of S3 cells) in the serosa and subserosa (HR3 & 4) and vimentin$^+$ fibroblasts ($V_9$16, 11.7% of S3 cells) in the lamina propria of healthy mucosa (HR1). Co-clustering of such dissimilar cell types confounds methods for cell-type identification. A second example of cell types incorrectly grouped together in segmentation-based analysis but separated by the VAE was S13 (conventional CD4$^+$ T cells, Tcons). Twenty-seven (27) percent of S13 cells were identified as CD163$^+$ CD4$^+$ tissue-resident macrophages ($V_9$15)[37] in HR1 and HR2 **(Fig. 4b-d**) with the remaining cells distributed across fibroblast and macrophage clusters ($V_9$7, 14, 16). We reasoned that shared expression of CD4 between Tcons and this macrophage population caused them to co-cluster in S13. Thus, pixel-level learning enabled by a VAE was able to discriminate cell populations occupying distinct spatial contexts, lacking definitive immunomarker profiles (e.g., fibroblasts), or expressing atypical combinations of markers (CD163$^+$ CD4$^+$ macrophages). The identification of these CD4$^+$ macrophages in CyCIF-1A highlights how unsupervised learning can "discover" unexpected cell types.

$V_9$23 presented another example of how unsupervised learning can identify recurrent spatial patterns not part of an *a priori* analysis plan. $V_9$23 represented a rare population of αSMA$^+$ smooth muscle cells lining the basement membrane of an artery within the muscularis propria (0.1% of total cells, HR1; **Fig. 4e**). $V_9$23 image patches were distinguished by the external elastic lamina—an acellular structure that provides mechanical support to the blood vessel wall and can be detected based on its strong and photostable autofluorescence[38] (**Fig. 4f**). Acellular structures such as this are visible in images but lack nuclei and are therefore absent from segmentation-based spatial feature tables.

Limitations of patch-based learning relative to image segmentation

In a few cases, VAE-based learning led to co-clustering of cell types that segmentation-based analysis correctly separated. For example, $V_9$4 contained both Tcons (S13) and regulatory T cells (Tregs, S15; **Extended Data Fig. 8a,b** and **Supplementary Video 3**). In biplots of segmented cells, CD4 and FOXP3 formed discrete distributions corresponding to Tcons and Tregs, but this was not true in patch-based representations (**Extended Data Fig. 8c-e**). We speculate that this occurred because nuclear FOXP3 signals represent only a small fraction of total patch dimensionality. The localization of Tcons and Tregs to similar stromal niches failed to provide additional discriminatory information (**Extended Data Fig. 8f,g**). This

297 highlights a limitation of unsupervised learning that could be addressed by increasing the weight of signals for
298 specific lineage markers or cell compartments (e.g., the nucleus) during image patch encoding.
299 In both segmentation and VAE-based analysis, clusters corresponding to lymphoid aggregates (S11,
300 $V_9$11 and $V_9$20, HR*; **Fig. 4g**) were heterogeneous with respect to cell type, containing both B and T cells.
301 With segmentation, this arose from spatial crosstalk, but in VAE analysis it was due to image patches
302 containing multiple nuclei due to close cell packing; more than 22% of patches from $V_9$11 and $V_9$20 contained
303 >1 nucleus (**Fig. 4h**). Thus, while unsupervised VAE-based learning has greater sensitivity to morphological
304 and contextual features it cannot separate whole cells that consistently co-appear in an image patch. A lack of
305 prior knowledge about what is important but subtle differences (e.g., nuclear FOXP3) imposes additional
306 limitations on unsupervised learning and underscores the value of performing pixel-level inference and
307 traditional segmentation in parallel.
308
309 **VAE analysis identifies matched cell states within and across specimens**
310 To explore the consistency of VAE-based learning across tissue sections and tumors, we trained a new
311 network on 9x9µm image patches from three CyCIF images of CRC labeled with the same 21 antibodies
312 (**Supplementary Table 2**). Tissues included (i) CyCIF-1A, to test technical reproducibility of the method; (ii)
313 a serial section from the same tissue block located ~20 µm along the Z-axis, to evaluate concordance across
314 biological replicates (CyCIF-1B; **Fig. 5a**); and (iii) a tissue section from a different patient, to compare clusters
315 across tumors (CyCIF-2; **Fig. 5b**). Leiden clustering of combined VAE image patch encodings from all three
316 samples yielded 33 clusters (V_{9M}0-32, with "V_{9M}" denoting a VAE trained on 9x9µm patches from multiple
317 tissues; **Fig. 5c**, **Supplementary Table 5** and **Supplementary Video 4**). Three clusters (V_{9M}29, 31, and 32)
318 were identified as microscopy artifacts (<0.3% of cells) that underwent no further analysis. The remaining
319 clusters corresponded to the expected four primary cell lineages: keratin$^+$ epithelial, αSMA$^+$ smooth muscle,
320 vimentin$^+$ stromal, and CD45$^+$ immune cells (**Extended Data Fig. 9a,b** and **Supplementary Fig. 11**). V_{9M}
321 clusters mapping to CyCIF-1A and 1B revealed strong concordance, with matched clusters occupying similar
322 regions of tissue (**Fig. 5d** and **Supplementary Fig. 12**). Hierarchical clustering of V_{9M} clusters by protein
323 expression profiles showed that 60% (18 of 30) of two-way dendrogram merges involved matched clusters
324 from CyCIF-1A and 1B, and an additional 37% (11 clusters) involved three-way dendrogram merges between
325 matched clusters from all three samples (**Fig. 5e**). Thus, 97% (60% + 37%) of V_{9M} clusters were matched
326 between serial sections of the same tumor.
327 Clusters common to all three images included: endothelial cells (V_{9M}4), smooth muscle cells (V_{9M}6),
328 crypt-forming epithelial cells (V_{9M}14), CD8$^+$ T cells (V_{9M}19), CD68$^+$ macrophages (V_{9M}21) and mitotic tumor

329 cells ($V_{9M}30$; **Fig. 5f**). The latter were identified based on their condensed nuclear chromatin (a feature of pro-
330 metaphase and metaphase cells), which was visible in the DNA channel and as a silhouette in the PCNA
331 channel (**Extended Data Fig. 9c**). Due to their scarcity (<0.08% of cells in each specimen), these cells were
332 missed in single-tissue analysis but were detected in multi-tissue analysis simply because there was more data.
333 Four clusters ($V_{9M}0$, 8, 13, and 23) were unique to CyCIF-2, together accounting for 85.5% of cells (**Fig. 5g,h**
334 and **Extended Data Fig. 9d** and **Supplementary Video 5**). The largest clusters, $V_{9M}0$ (41.2%) and $V_{9M}13$
335 (17.6%), corresponded to $PCNA^{high}$ and $PCNA^{low/neg}$ tumor cells, respectively (**Extended Data Fig. 9e**). Both
336 exhibited abnormal CD3ε labeling, likely due to expression of cross-reactive epitopes, and explaining why
337 CyCIF-2 tumor cells did not co-cluster with those in CyCIF-1 (**Extended Data Fig. 9f,g**). While $V_{9M}8$ cells
338 (fibroblasts) were found in all three samples, they were 7-fold more abundant in CyCIF-2 due to a higher
339 stroma-to-tumor ratio (**Fig. 5g** and **Extended Data Fig. 9h**). $V_{9M}23$ represented $CD45^+$ cells specific to CyCIF-
340 2 that lacked expression of other immune lineage markers (**Extended Data Fig. 9i**). However, the same tumor-
341 infiltrating $CD8^+$ T cells ($V_9 18$) and surrounding $keratin^+$ tumor cells ($V_9 22$) in CyCIF-1A were also identified
342 in the other two images ($V_{9M}24$ and $V_{9M}27$; **Extended Data Fig. 9j**), emphasizing the ability of a VAE to
343 reproducibly overcome spatial crosstalk. These results show that VAE-based learning reliably identified similar
344 cell types across tissues, including rare populations.
345

346 **VAE analysis of multicellular image patches identifies recurrent tissue motifs**

347 VAE networks can be trained on image patches of arbitrary dimensions. To classify CRC according to
348 multi-cellular architecture, we trained a VAE model on 30x30µm image patches containing an average of 5
349 nuclei per patch (**Fig. 6a**). Although patches were centered on reference cells for indexing purposes, Gaussian
350 vignetting was omitted to ensure that all pixels were visible to the network at full intensity. At convergence, the
351 model achieved a validation loss of 1.4×10^{-3}, with VAE-generated representations closely resembling input
352 image patches (**Fig. 6b**). As in V_9 analysis, the distribution of concept scores for each V_{30} cluster were bimodal,
353 indicating that clusters with distinct features had been identified (**Extended Data Fig. 10a**). Leiden clustering
354 applied to 30x30µm image patch encodings generated 12 clusters (V_{30}0-11) including one microscopy artifact
355 ($V_{30}11$, 0.1% of cells; **Fig. 6c**, **Supplementary Table 6**, **Supplementary Video 6**, **Supplementary Image**
356 **Gallery 3**). The spatial distributions of V_{30} clusters corresponded to histologic regions in the H&E image,
357 including glandular ($V_{30}2/8$, HR5), mucinous ($V_{30}4$, HR6), and solid AC ($V_{30}0$, HR7); the muscularis propria
358 ($V_{30}5/6$; HR2); normal mucosa ($V_{30}7$, HR1); and serosal mesothelium ($V_{30}10$, HR3). In addition, $V_{30}9$
359 correctly captured all 17 lymphoid aggregates annotated by our histopathologists in the H&E image plus
360 several smaller ones ($V_{30}9$, HR*; **Fig. 6d**).

361 Because V_{30} clusters contained more pixels and were not vignetted, concept saliency maps had higher
362 resolution, with high-attribution pixels forming clear multi-cellular spatial patterns (**Fig. 6e**, **Extended Data**
363 **Fig. 10b**, and **Supplementary Image Gallery 4**). For example, keratin saliency maps reproduced the
364 characteristic morphologies of epithelial cells in normal mucosal crypts ($V_{30}7$) and glandular AC ($V_{30}8$), while
365 maps for lymphoid aggregates ($V_{30}9$) highlighted strong contributions from $CD20^+$ B cells and $CD4^+$ helper T
366 cells, suggesting their importance in defining this cluster concept. In contrast, $CD8^+$ T cells contributed
367 negatively to the lymphoid aggregate cluster concept, consistent with the observation that $CD8^+$ T cells were
368 less abundant in these structures than B and helper T cells.

369 To better understand the cell type composition of V_{30} clusters, we used information entropy computed
370 against V_9 cluster labels. Entropy scores (which had possible values between 0 to 4.6 bits based on 24 possible
371 V_9 labels) ranged from 0.38 bits for $V_{30}10$ to 3.18 bits for $V_{30}1$ (**Fig. 6f**). At least 75% of cells in low-entropy
372 clusters $V_{30}7$, 8 and 10 corresponded to a single V_9 cluster label, showing that these structures were primarily
373 composed of a single cell type. For example, 92% of cells in $V_{30}7$ were normal crypt-forming epithelial cells
374 ($V_9$10; **Fig. 6g**). However, cells in $V_{30}1$ (a high-entropy cluster) comprised ten different V_9 clusters, including
375 $CD4^+$ T cells ($V_9$4, 19.0%), $CD68^+$ $CD163^+$ macrophages ($V_9$7, 17.3%), $CD31^+$ endothelial cells ($V_9$2, 15.8%),
376 $CD68^+$ PD-$L1^+$ macrophages ($V_9$14, 10.9%), vimentin$^+$ fibroblasts ($V_9$16, 10.7%), $CD8^+$ T cells ($V_9$13, 9.5%),
377 and $CD163^+$ $CD4^+$ macrophages ($V_9$15, 6.3%; **Fig. 6h**). This diversity of cell types reflects the role of the
378 colonic stroma in structural support, nutrient exchange, and wound healing mediated by intermixed fibroblasts,
379 immune, and other cell types. Notably, both the lamina propria of healthy mucosal crypts (HR1) and the
380 peritumoral stroma surrounding cancer cells (HR9) were both part of $V_{30}1$, suggesting that these compartments
381 share a common set of constituent cell types. We conclude that a VAE trained on multicellular image patches
382 can correctly identify morphologically distinct arrangements of single cell types as well as complex cellular
383 communities comprising multiple lineages in normal and desmoplastic stromal tissue.

DISCUSSION

We have described a framework (MORPHAEUS, Morphology Oriented Research in Pathology Harnessing Autoencoder Unsupervised Strategies) that combines the strengths of unsupervised pixel-level ML and traditional image segmentation to support tissue classification across spatial scales, from single cells to tissue neighborhoods. The resulting data fit into current single-cell analysis pipelines where adjacent cells that cannot be separated are treated as a community. We show that VAEs can extract information from images corresponding to marker intensity, cell morphology, and local neighborhood in a manner analogous to human experts, providing deeper insight into tissue composition and organization than segmentation-based analysis alone while addressing spatial crosstalk.

Generative models such as VAEs, adversarial autoencoders, deep convolutional generative adversarial networks, and combinations thereof require no ground truth annotations, making them well-suited to discovery of unanticipated features. In our work VAEs detected (i) an unexpected population of $CD163^+$ $CD4^+$ tissue-resident macrophages[37] ($V_9$15) missed by segmentation-based clustering, (ii) mitotic tumor cells (V_{9M}30) with distinctive DNA morphology, and (iii) cells adjacent to acellular structures such as the arterial elastic lamina ($V_9$23). Because VAEs exploit spatial patterns in marker levels to learn neighborhood features, they represent a novel approach to overcoming spatial crosstalk in segmentation-based methods. VAEs accomplish this by identifying multi-cellular cellular niches such as $CD8^+$ T cells in stromal ($V_9$13) versus tumor tissue ($V_9$18). Moreover, whereas segmentation assumes all cells are separable and spatial relationships are inferred using spatial statistics, training VAEs on patches of different sizes enables learning of both cell-level and neighborhood features simultaneously. This represents a conceptually different approach to spatial profiling in which we accept that it is not possible, or even desirable, to singulate every cell and then attempt to reassemble neighborhoods from single cell data. Instead, VAE-based approaches leave neighborhoods intact depending on patch size: 9μm patches captured single cells or closely associated pairs (e.g. T cells and tumor cells), while 30μm patches captured multi-cellular neighborhoods.

Despite the strengths of unsupervised VAE-based learning, it can miss subtle nuclear features such as the difference between Tcons and Tregs based on expression of the nuclear transcription factor FOXP3.[39] Segmentation-based methods may be superior in this case because they implicitly increase the saliency of nuclear signals due to the use of nuclear segmentation masks. However, the same effect may be achievable via feature engineering during VAE image patch encoding.[40] Both segmentation and VAE models also fail to separate B and T cells in dense lymphoid aggregates. A potential solution is a multi-pass learning in which a smaller patch size is used for cell-dense regions of tissue such as lymphoid aggregates with larger patches used to infer cells in the remainder of tissue. However, such approaches will likely be most effective using higher

resolution data[41] that can more effectively discern between cell types. In some cases it may be reasonable to view T and B cell pairs in densely-packed lymphoid aggregates as a recurrent tissue motif rather than sets of fully discriminable cells. Thus, we suggest performing traditional segmentation and patch-based learning in parallel, and using the comparison to optimize patch size, adjust vignetting, and tune clustering parameters.

SUPPLEMENTARY INFORMATION

Supplementary Note 1 – Additional information on segmentation methods and cell type calling

Highly multiplex tissue imaging is a rapidly evolving technology whose analysis often involves defining boundaries around individual cells through the process of image segmentation. In the case of whole-slide tissue imaging, this involves the acquisition of tens to hundreds of image tiles stitched and registered together using algorithms such as ASHLAR[42] that form multi-channel mosaic images stored in metadata-rich open file formats like OME-TIFF[43] (level 2 data according to MITI guidelines[2], **Supplementary Fig. 1a**). Semantic segmentation algorithms, such as the popular U-Net[44,45] architecture and its derivatives such as UnMICST[46], TissueNet[47], and Cellpose[7] are then applied to images of cell nuclei labeled with DNA intercalating dyes (e.g., Hoechst or DAPI) to identify boundaries between cells and generate cell segmentation masks—reference images in which pixels are labeled according to the cell they belong to (MITI level 3 data, **Supplementary Fig. 1b**).

Mean immunomarker signals for individual cells are then programmatically computed and organized into a tabular data format like CSV (level 4 data according to MITI guidelines, **Supplementary Fig. 1c**). The derived tables are often augmented by measurements of the segmentation instances themselves, which serve as proxies for nuclear morphology (e.g., area, eccentricity, etc.). However, the accuracy of these measurements is conditional on segmentation quality. While standard pipelines such as MCMICRO[28] have been developed to automate the aforementioned image processing tasks, no consensus exists on how best to classify cell phenotypes in tissue-derived, single-cell data.

A common method for segmentation-based cell state inference involves manual thresholding per-cell signal intensities (i.e., hard gating) performed using univariate histograms or bivariate scatter plots (**Supplementary Fig. 1d**). However, as positive and negative cell populations often overlap, such methods can be sensitive to even small changes in gate placement, leading to ambiguous annotation and potentially changing biological interpretation. Methods based on Gaussian mixture models (GMMs) and naïve Bayes classifiers such as those implemented by the marker-scoring functions in CELESTA[48] and MCMICRO[28], respectively, represent "soft" gating methods. Such methods have the advantage of assigning class-conditional probability weights to individual cells as an alternative to strict thresholding (**Supplementary Fig. 1e**). Nevertheless, these methods also require manual parameter tuning and are sensitive to the same limitations imposed by segmentation-based analysis.

Unsupervised clustering methods (e.g., Louvain[49] or Leiden[31] community detection, k-means[50], SOMs[51] (self-organizing maps), and HDBSCAN[52] (hierarchical density-based spatial clustering of applications with noise)) are an alternative to gate-based approaches for cell state inference. In these methods, high-dimensional

452 single-cell data are projected onto lower dimensional embeddings (typically 2D or 3D) using data
453 dimensionality reduction algorithms such as t-SNE (t-distributed stochastic neighbor embedding)[53] or UMAP
454 (uniform manifold approximation and projection)[33] before or after clustering (**Supplementary Fig. 1f**).
455 Regardless of the method used for cell phenotyping, protein expression profiles for inferred cell populations are
456 often visualized as clustered heatmaps that can be interpreted by domain experts according to prior knowledge
457 of cellular protein expression profiles (**Supplementary Fig. 1g**).
458

459 **Supplementary Note 2 – Analysis of Errors in Segmentation Cluster S5**

460 Based on a heatmap of marker expression, segmentation cluster S5 appeared to express markers
461 associated with multiple cell types, including smooth muscle cells (αSMA, collagen IV), macrophages (CD68,
462 CD163), endothelial cells (CD31), mesenchymal cells (vimentin), and $CD8^+$ T cells (CD45, CD3ε, CD8α,
463 CD45RO; **Fig. 1e**). Unexpectedly, cells in this cluster gave rise to multiple subclusters in UMAP space, which
464 we termed S5.1, S5.2, and S5.3 (**Supplementary Fig. 5a**). Viewing these cells in CyCIF-1A revealed that they
465 corresponded to phenotypically similar $CD8^+$ T cells occupying distinct histological regions that also differed
466 with respect to their protein expression profiles (**Supplementary Fig. 5d,c**). Closer inspection showed cells in
467 S5.1 were influenced by spatial crosstalk with $CD68^+$ PD-$L1^+$ macrophages located in the peritumoral stroma
468 (HR9) and stromal compartments throughout the muscularis propria (HR2; **Supplementary Fig. 5d**), while
469 cells in S5.2 were affected by crosstalk with $Desmin^+$ smooth muscle cells in the muscularis (**Supplementary**
470 **Fig. 5e**). S5.3 cells, which were located in the serosa (HR3) and subserosa (HR4), did not appear to be
471 impacted by spatial crosstalk. Instead, they were distinguished by high CD45 expression, a feature we had
472 previously observed in immune cells occupying these tissue compartments (**Supplementary Fig. 5f**).

473 Although all S5 cells were grouped under the same label by the Leiden clustering algorithm, they
474 formed three distinct subpopulations in 3D UMAP space. This observation led us to hypothesize that the effects
475 of spatial crosstalk are more apparent in low-dimensional embeddings of high-plex data due to data
476 compression. This idea aligns with previous findings that distances and relationships between data points are
477 not reliably preserved during dimensionality reduction[54]. To test this hypothesis, we performed density-based
478 clustering using HDBSCAN[52] on both 2D and 3D UMAP embeddings of cells from CyCIF-1A. Supporting our
479 hypothesis, we found that cells grouped into a single cluster in 3D UMAP space were often redistributed into
480 multiple clusters in a 2D embedding. For example, a subset of tumor cells originally grouped within cluster 17
481 in a 3D embedding (**Supplementary Fig. 6a-c**) were delocalized in a 2D embedding of the same data
482 (**Supplementary Fig. 6d**). These delocalized cells were found to reside at the tumor-stromal interface and were
483 influenced by spatial crosstalk from collagen IV signals originating in neighboring stromal cells

484 (**Supplementary Fig. 6e,f**). Examples like this help explain the delocalization of three $CD8^+$ T cell subsets
485 comprising S5 (S5.1, S5.2, and S5.3) in UMAP feature space. They also underscore the creation of artifactual
486 clusters by spatial crosstalk.
487
488 **Supplementary Note 3 – Testing VAE-based analysis on different types of high-plex tissue images**
489 To assess the ability of the method to generalize across different imaging platforms and tissue types,
490 independent VAE models were trained on Lunaphore COMET and Akoya PhenoCycler (CODEX) data. The
491 Lunaphore image (referred to as Lunaphore-1) comprised a section of human non-small cell lung cancer
492 (NSCLC) stained for 27 markers and acquired at a nominal resolution of 0.28 µm/pixel (compared to 0.65
493 µm/pixel for CyCIF-1A). We found that image patches in this dataset cropped at 5x5µm (compared to 9x9µm
494 for CyCIF-1A) best approximated single cells. Leiden community detection identified 19 VAE image patch
495 clusters ($V_5$0-$V_5$18) well separated in UMAP space (**Extended Data Fig. 6a**). Clustering protein expression
496 profiles revealed biologically recognizable populations, including IDO-1^+ and IDO-1^- lung cancer cells ($V_5$0
497 and $V_5$11, respectively), αSMA^{lo} and αSMA^{hi} vasculature ($V_5$1 and $V_5$2, respectively), $FOXP3^+$ Tregs ($V_5$12),
498 $CD66b^+$ granulocytes ($V_5$16), and podoplanin$^+$ lymphatic vessels ($V_5$17; **Extended Data Fig. 6b**). Mapping
499 clusters onto the original image confirmed their identities based on characteristic morphologies and spatial
500 distributions (**Extended Data Fig. 6c,d** and **Supplementary Fig. 9**). For example, clusters $V_5$1 and $V_5$2
501 corresponded to well-structured, pre-existing ($V_5$1) and disorganized, neo-angiogeneic ($V_5$2) blood vessels.
502 The PhenoCycler image (referred to as CODEX-1) comprised a tissue microarray (TMA) core of
503 human tongue squamous cell carcinoma stained for 29 markers and imaged at a nominal resolution of 0.5
504 µm/pixel. Unlike Lunaphore and CyCIF images, which were stored at 16-bit depth, these data were stored at 8-
505 bit depth. Single cells in this image were approximated using 11x11µm image patches. Despite the lower bit
506 depth and relatively small number of cells in this image (~10K), the Leiden algorithm identified 19 VAE image
507 patch clusters (V_{11}0-V_{11}18) that were also well separated in UMAP space (**Extended Data Fig. 6e**). Multiple
508 tumor cell phenotypes were identified in this sample based on differences in marker intensity and morphology,
509 including $CD44^+$ stem-like (V_{11}0) and $CK17^+$ basal-like (V_{11}5) carcinoma cells. Other identified cell types
510 included cancer-associated fibroblasts (CAFs, V_{11}4), normoxic and hypoxic macrophages (V_{11}11 and V_{11}18,
511 respectively), and memory $CD4^+$ T cells (V_{11}16; **Extended Data Fig. 6f**). These clusters were also validated by
512 referencing the primary image (**Extended Data Fig. 6g,h** and **Supplementary Fig. 10**). Taken together, these
513 additional analyses demonstrated that, despite differences in image bit depth, nominal resolution, and cell
514 number, VAE-based cell state inference generalizes effectively across diverse spatial proteomics platforms.

FIGURES and LEGENDS

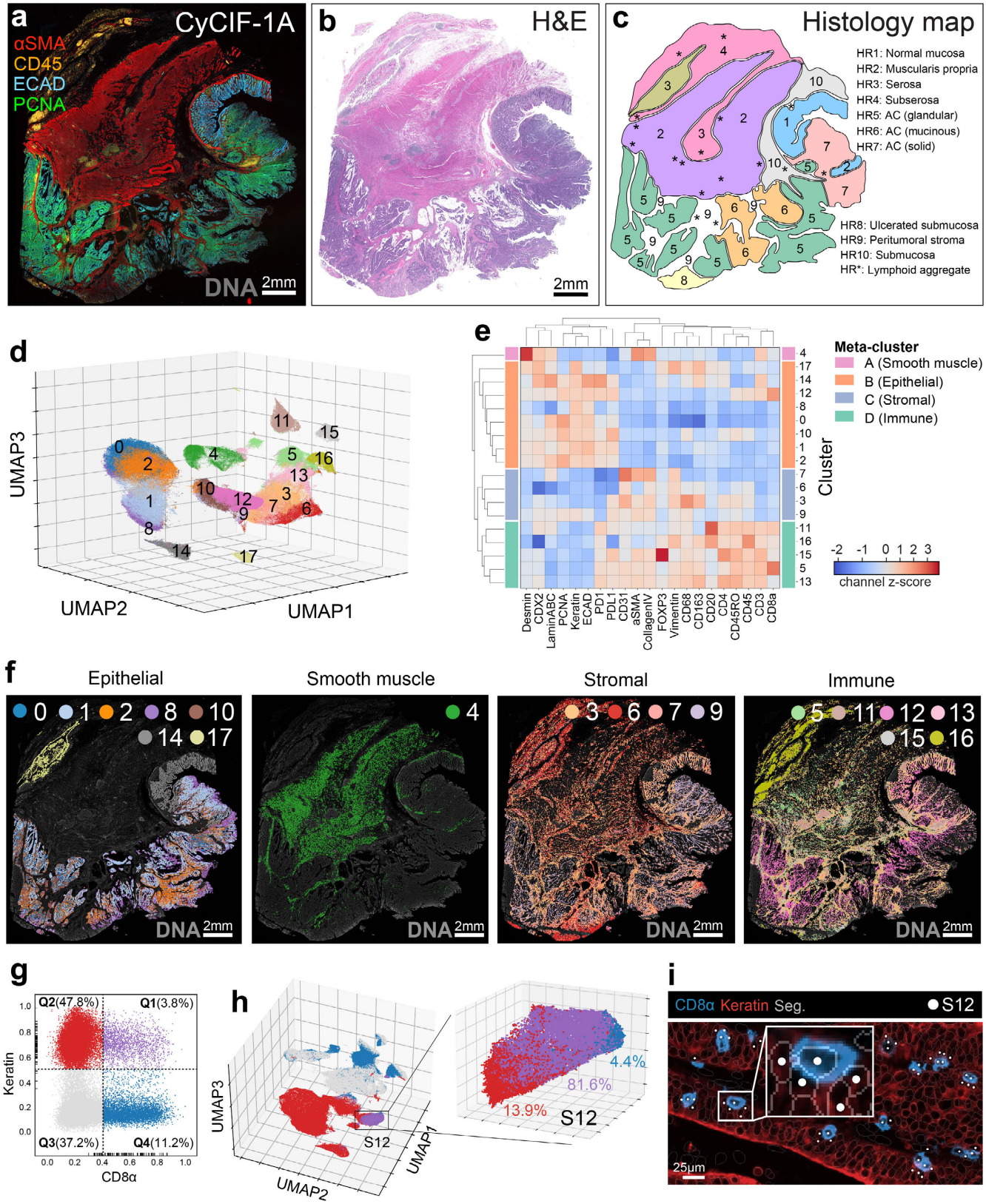


Fig. 1 | Spatial crosstalk limits accuracy of segmentation-based cell phenotyping. a, CyCIF-1A: a 22-plex, whole-slide CyCIF image of primary human CRC (HTAN ID: HTA13_1_101) showing five channels: αSMA

(smooth muscle, red), CD45 (leukocytes, orange), ECAD (epithelium, blue), PCNA (proliferating cells, green), and DNA (Hoechst-stained nuclei, gray). See **Supplementary Table 2** for a complete list of antibodies used to label this tissue. **b**, Adjacent tissue section to CyCIF-1A (~4μm away in Z-axis) stained with routine histology dyes Hematoxylin & Eosin (H&E). **c**, Histology map showing the locations of 11 histological regions (HRs) observed by two board-certified pathologists in the H&E image shown in panel b. **d**, UMAP embedding of segmented cells from CyCIF-1A colored by Leiden cluster; see **Supplementary Video 1** for 360° plot rotation. **e**, Hierarchical clustering of Leiden cluster z-score-normalized mean marker intensities revealing four primary meta-clusters (A-D). **f**, Location of Leiden clusters in CyCIF-1A grouped by major cell lineages: Epithelial, Smooth muscle, Stromal, and Immune. **g**, Gated bivariate plot showing the percentage of CD8α$^+$ keratin$^+$ (Q1, purple), keratin$^+$ (Q2, red), CD8α$^-$ keratin$^-$ (Q3, gray), and CD8α$^+$ cells (Q4, blue) in CyCIF-1A. Rugplots (small black ticks along x- and y-axes) indicate marker intensity values for 100 cells drawn at random from S12 showing that the majority of cells in this cluster are positive for both markers. **h**, UMAP embedding of segmented cells from CyCIF-1A colored by the gating strategy shown in panel g; cluster S12 is highlighted by the black rectangle. Inset shows S12 in further detail highlighting its percentages of keratin$^+$ (red), CD8α$^+$ (blue), and CD8α$^+$ keratin$^+$ cells (purple). **i**, Field-of-view from CyCIF-1A showing that cells in S12 (white scatter points) consist of both CD8α$^+$ T cells (blue) and neighboring keratin$^+$ epithelial cells (red). Inset shows spillover of marker signals across segmentation boundaries (translucent perimeters) between the two different cell types, which we refer to as “spatial crosstalk”.

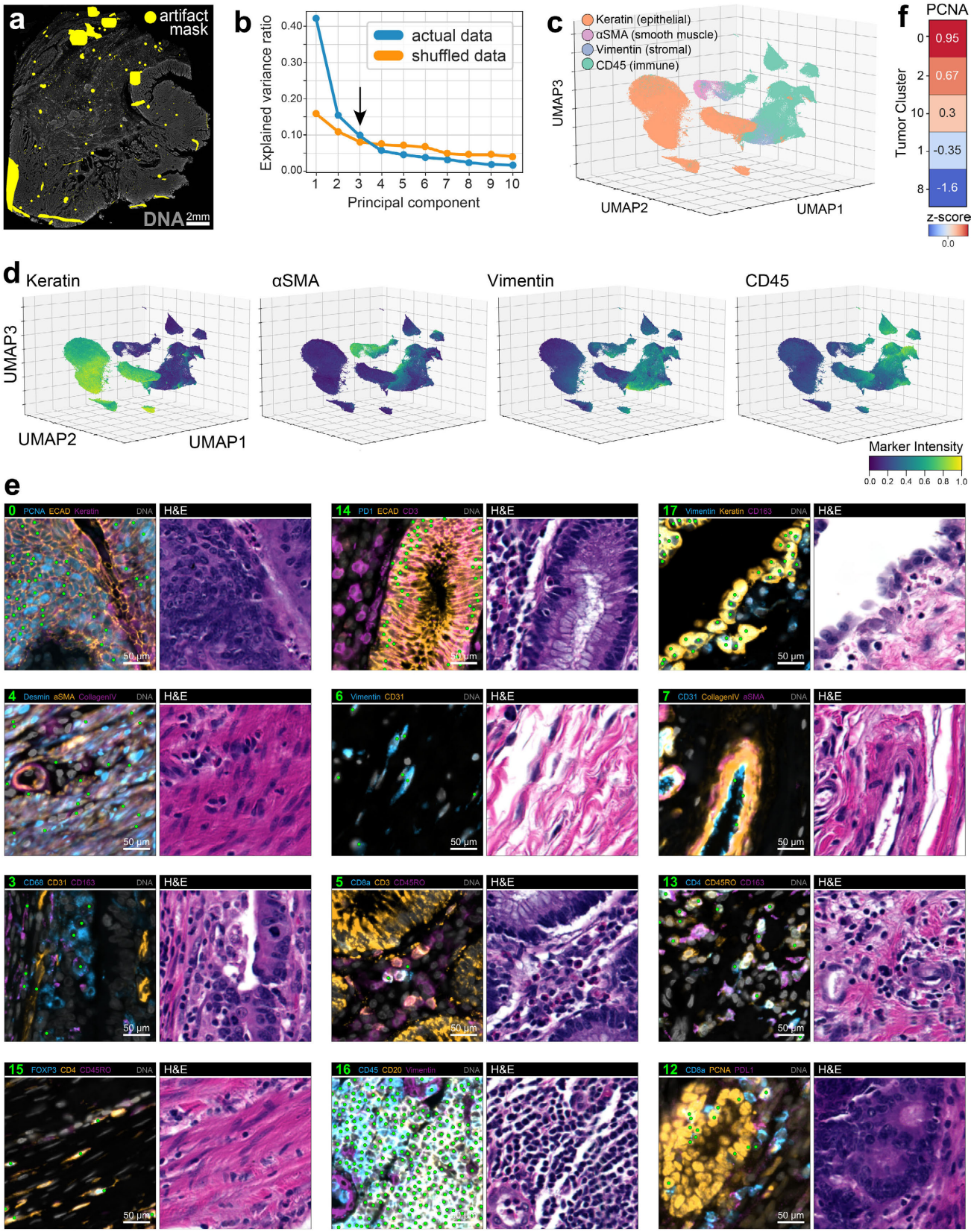
537

Image	Data Type	Plex	Image dimensions (X/Y pixels)	Nominal resolution (µm/pixel)	Tissue Area (mm^2)	HTAN ID	Description
CyCIF-1A	CyCIF	21 (+Hoechst)	(26139, 27120)	0.65	~172	HTA13_1_101	Whole-slide section of human colorectal adenocarcinoma from a 69-year-old Caucasian male
CyCIF-1B	CyCIF	21 (+Hoechst)	(24897, 27121)	0.65	~165	HTA13_1_106	Whole-slide section of human colorectal adenocarcinoma from a 69-year-old Caucasian male (~20µm away from CyCIF-1A in the Z-axis)
H&E	H&E	N/A	(26139, 27120)	0.65	~164	HTA13_1_100	Whole-slide section of human colorectal adenocarcinoma from a 69-year-old Caucasian male (~4µm away from CyCIF-1A in the Z-axis)
CyCIF-2	CyCIF	21 (+Hoechst)	(27411, 22981)	0.65	~161	HTA7_922_2	Whole-slide section of human colorectal adenocarcinoma from a 62-year-old Caucasian male
Lunaphore-1	Lunaphore COMET	27 (+DAPI)	(44643, 44643)	0.28	~41	N/A	Whole-slide section of human non-small cell carcinoma from a 64-year-old Caucasian male
CODEX-1	PhenoCycler (CODEX)	29 (+DAPI)	(5695, 5695)	0.5	~12	N/A	1mm diameter tissue microarray core of human tonugue squamous cell carcinoma

538 **Supplementary Table 1 | Metadata for biospecimens used in this study.** Primary imaging data for the first
539 four samples can be found at the HTAN Data Portal (https://data.humantumoratlas.org/explore) by referencing
540 their corresponding HTAN IDs. Lunaphore and CODEX images are specific to this study and can be accessed
541 upon request from the authors. Additional information on CyCIF antibodies used to label samples CyCIF-1A,
542 CyCIF-1B, and CyCIF-2 can be found in the STAR Methods section of Lin et al. *Cell* 2023 (DOI:
543 10.1016/j.cell.2022.12.028).

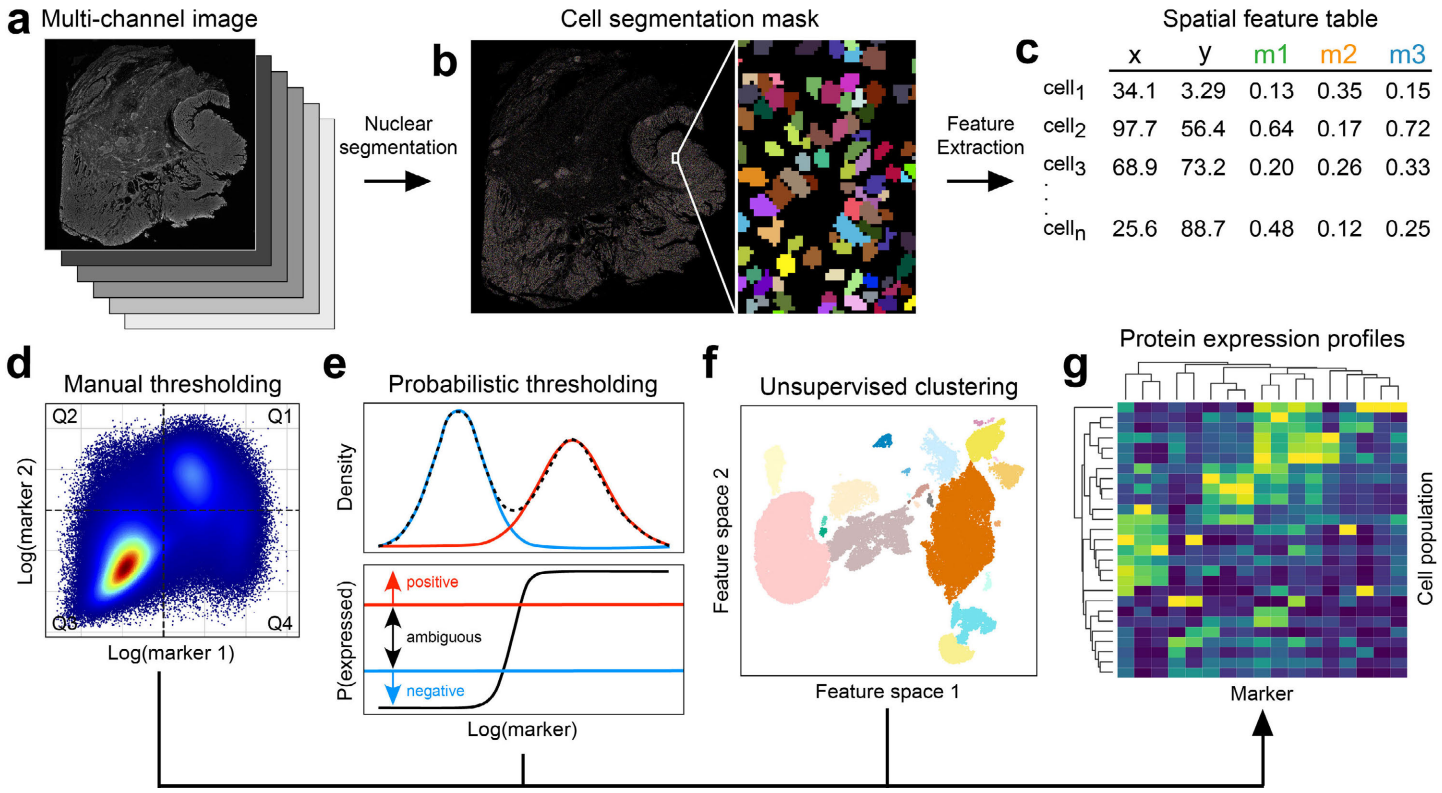
Channel	Marker	Segmentation Object	Analyzed
1	Hoechst-cycle1	nucleus	**yes**
2	AF488	N/A	no
3	AF555	N/A	no
4	AF647	N/A	no
5	Hoechst-cycle2	nucleus	no
6	A488	N/A	no
7	A555	N/A	no
8	A647	N/A	no
9	Hoechst-cycle3	nucleus	no
10	CD3	membrane	**yes**
11	NaKATPase	membrane	no
12	CD45RO	membrane	**yes**
13	Hoechst-cycle4	nucleus	no
14	Ki67	nucleus	no
15	Keratin	cytoplasm	**yes**
16	αSMA	cytoplasm	**yes**
17	Hoechst-cycle5	nucleus	no
18	CD4	membrane	**yes**
19	CD45	membrane	**yes**
20	PD1	membrane	**yes**
21	Hoechst-cycle6	nucleus	no
22	CD20	membrane	**yes**
23	CD68	cytoplasm	**yes**
24	CD8a	membrane	**yes**
25	Hoechst-cycle7	nucleus	no
26	CD163	cytoplasm	**yes**
27	FOXP3	nucleus	**yes**
28	PD-L1	membrane	**yes**
29	Hoechst-cycle8	nucleus	no
30	ECAD	cytoplasm	**yes**
31	Vimentin	cytoplasm	**yes**
32	CDX2	cytoplasm	**yes**
33	Hoechst-cycle9	nucleus	no
34	LaminABC	nucleus	**yes**
35	Desmin	cytoplasm	**yes**
36	CD31	nucleus	**yes**
37	Hoechst-cycle10	nucleus	no
38	PCNA	nucleus	**yes**
39	Ki67	nucleus	no
40	CollagenIV	cytoplasm	**yes**

Supplementary Table 2 | Metadata for CyCIF antibodies. The "Segmentation Object" column indicates the cellular compartment used to extract marker signals for that channel. The column labeled "Analyzed" indicates whether a channel was evaluated in the current study. Unused channels represent secondary antibodies (e.g., channel 2, AF488) used to block non-specific antibody binding in downstream CyCIF cycles, repeated DNA images used for cross-cycle image registration, or markers that failed initial quality control.



Extended Data Fig. 1 | Segmentation-based cell phenotyping of human CRC. a, Artifact masks (yellow ROIs) highlighting regions in CyCIF-1A affected by microscopy artifacts (e.g., tissue folds, antibody

aggregates, illumination aberrations, etc.) censored from further analysis. **b**, Horn's parallel analysis showing the fraction of variance in post-QC single-cell data from CyCIF-1A explained by the first 10 principal components (blue curve) relative to the same data after randomization (orange curve). Black arrow highlights the point of intersection between the two curves, indicating that the first three principal components capture non-random variation in the dataset. **c**, UMAP embedding of segmented cells from CyCIF-1A colored by major cell lineage markers: keratin (epithelial, orange), αSMA (smooth muscle, pink), vimentin (stromal, blue), and CD45 (immune, green). **d**, UMAP embedding of segmented cells from CyCIF-1A colored by mean marker intensity for major cell lineage markers keratin, αSMA, vimentin, and CD45. See **Supplementary Fig. 2** for colormaps of all 21 immunomarkers. **e**, Paired CyCIF (left) and H&E (right) fields-of-view from CyCIF-1A showing the location of random examples of cells from various clusters (green scatter points in CyCIF images). See **Supplementary Fig. 3** for views of all 18 clusters. **f**, Heatmap of z-score-normalized mean PCNA intensities among tumor-associated clusters.

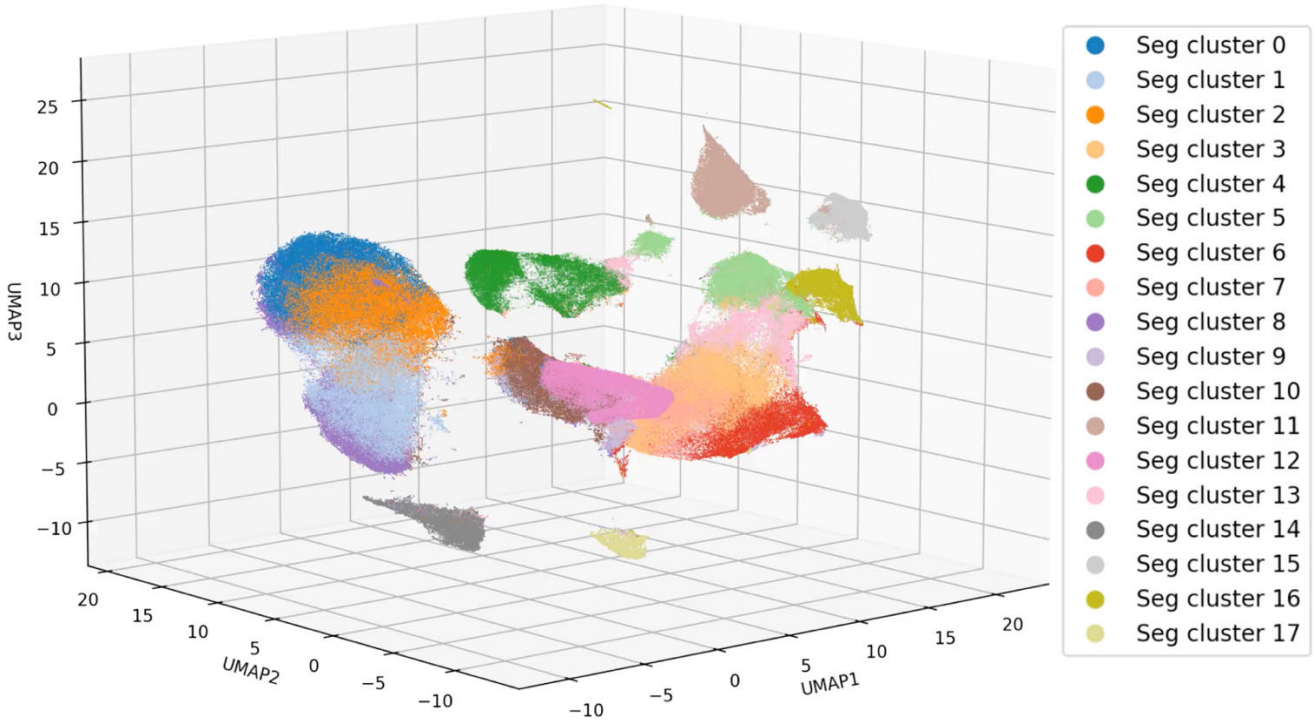


Supplementary Fig. 1 | Workflow for segmentation-based cell phenotyping of multiplex tissue images. **a**, Multi-channel OME-TIFF/TIFF file. **b**, Segmentation mask generated by semantic segmentation of counterstained cell nuclei. **c**, Spatial feature table specifying X/Y spatial coordinates and mean marker intensities (e.g., m1-m3) for segmented cells (rows). **d**, Quadrant gating of single-cell data according to two markers colored by cell density. **e**, Two-mode Gaussian mixture model for assigning continuous probability weights for single marker expression values of single-cell data. Probability of immunoreactivity is computed as the Bayesian posterior probability (y-axis, bottom plot) at a given channel intensity (x-axis). Cells with posterior probabilities above the horizontal red line are positive for the marker, those beneath the horizontal blue line are negative. Those with posterior probabilities between the red and blue lines are deemed ambiguous for the marker in this example. **f**, Unsupervised clustering of high-dimensional single-cell data embedded in a 2D feature space (e.g., UMAP or t-SNE). **g**, Clustered heatmap of mean marker intensities (columns) for different cell populations (rows) identified by any of the methods shown in panels d-f.

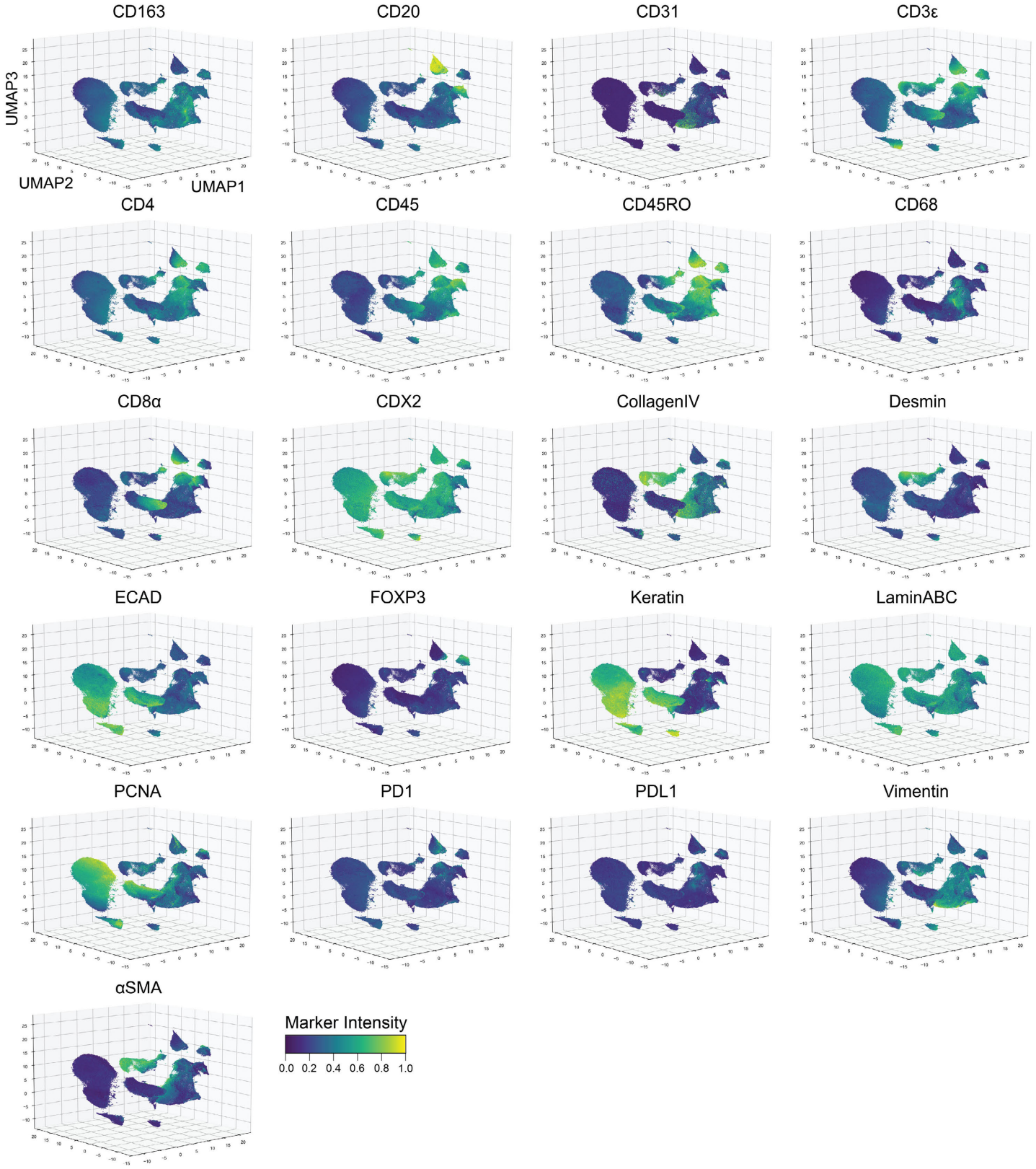
578

Leiden cluster	Annotation	% total cells	Spatial crosstalk
0	PCNA^{5+} adenocarcinoma (AC)	12.6	False
1	PCNA^{2+} AC	9.0	False
2	PCNA^{4+} AC	7.9	False
3	Macrophages	7.7	False
4	Smooth muscle cells	6.8	False
5	CD8^{+} T cells	6.5	**True**
6	Fibroblasts	6.1	False
7	Endothelial cells	5.8	False
8	PCNA^{1+} AC	5.8	False
9	Mixed AC and stromal cells (tumor-stromal boundry)	5.3	**True**
10	PCNA^{3+} AC	5.2	**True**
11	Mixed B and T lymphocytes (lymphoid aggregates)	5.1	**True**
12	Mixed CD8^{+} T and AC cells	4.2	**True**
13	Conventional CD4^{+} T cells (Tcons)	4.1	False
14	Crypt-forming epithelial cells	2.6	False
15	Regulatory CD4^{+} T cells (Tregs)	2.4	False
16	Mixed lymphocytes (CD45^{2+})	2.2	**True**
17	Benign reactive mesothelial cells	0.8	False

579 **Supplementary Table 3 | Segmentation-based Leiden cluster annotations.** The column labeled "% total
580 cells" indicates the percentage of total cells in CyCIF-1A accounted for by that cluster. Column labeled
581 "Spatial crosstalk" indicates whether signal spillover across segmentation boundaries of neighboring cells was
582 observed in CyCIF-1A.



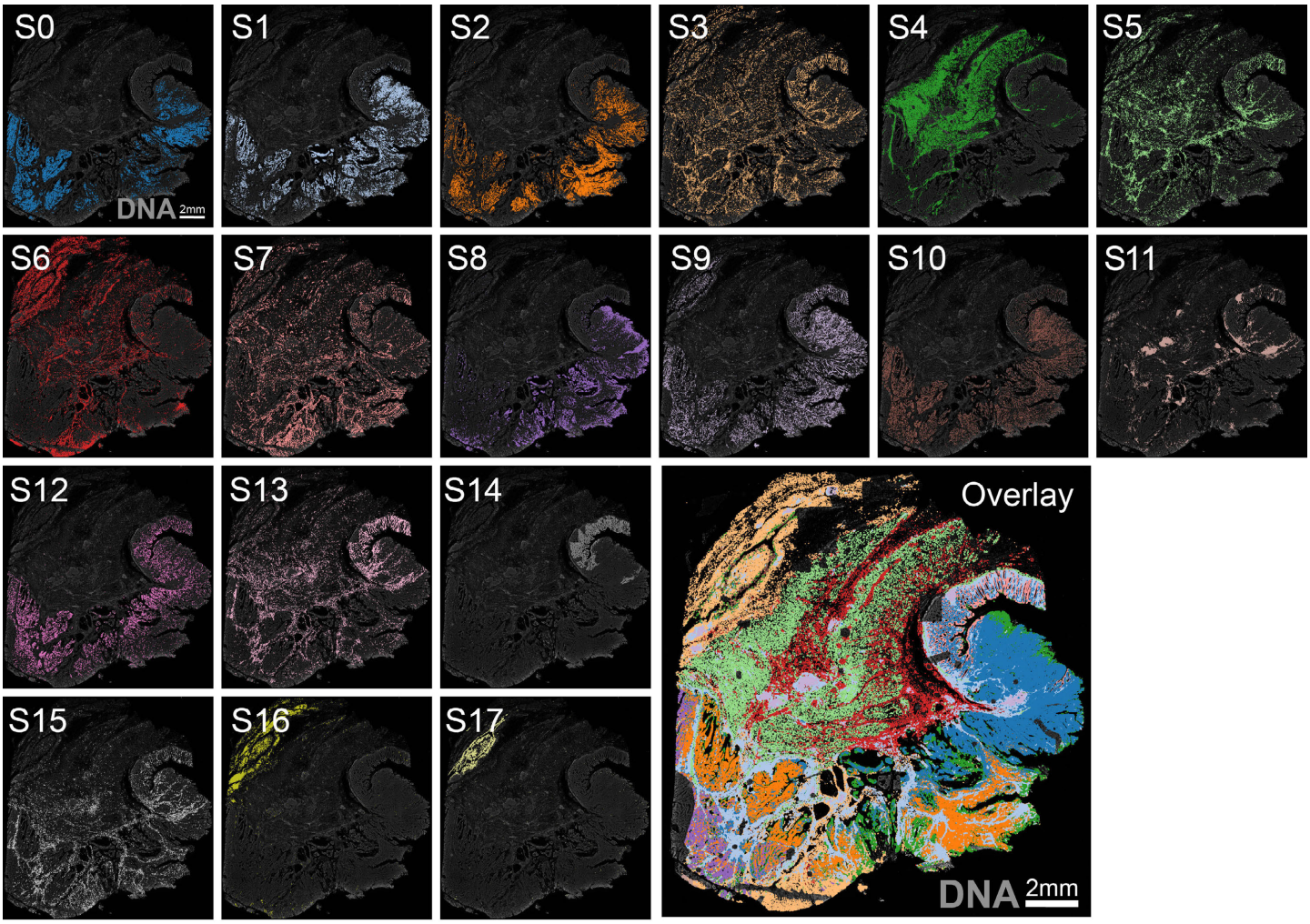
Supplementary Video 1 | 360° UMAP plot rotation showing segmented cells colored by Leiden cluster label.



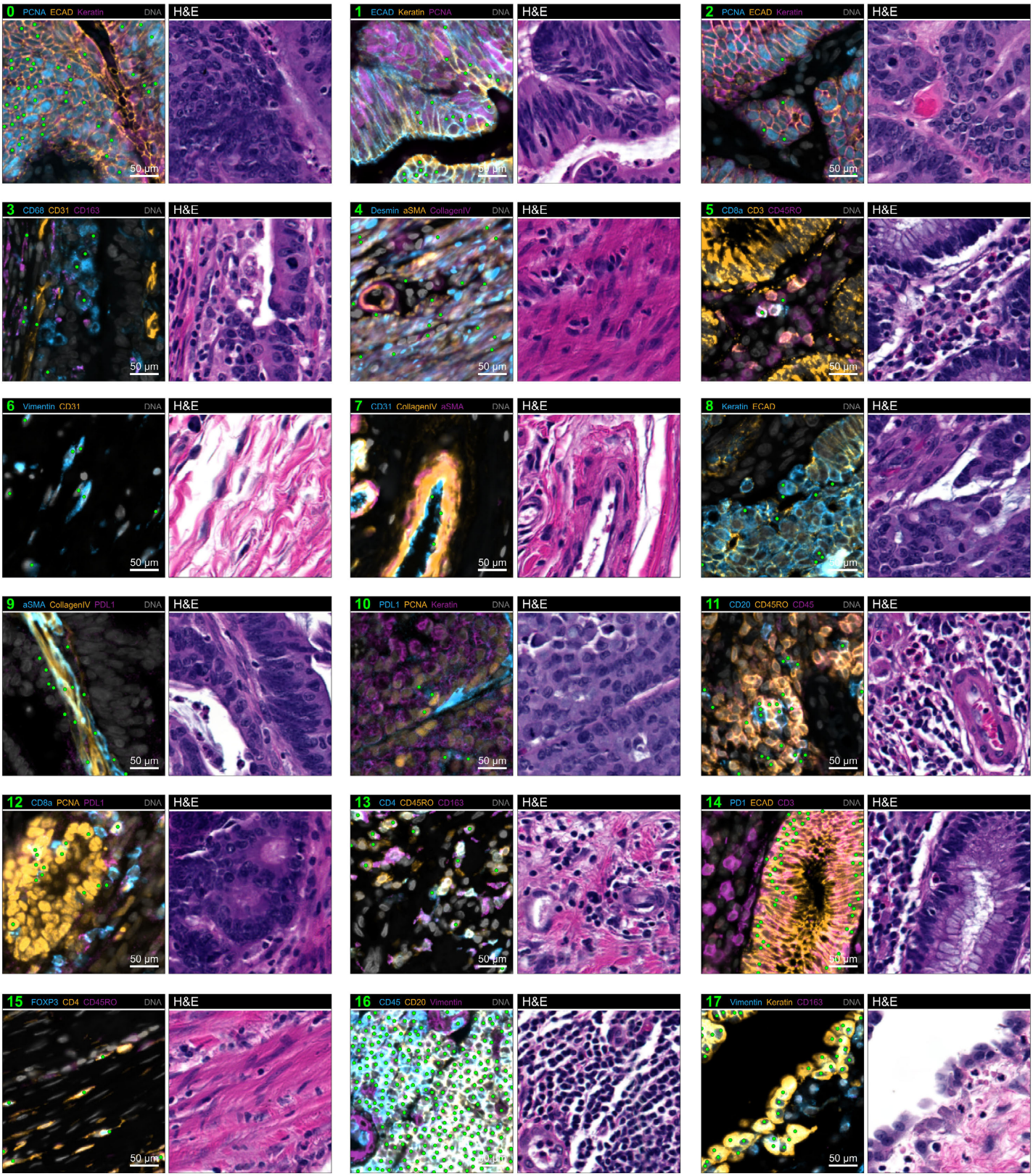
585

586 **Supplementary Fig. 2 | Channel colormaps for segmented cells in CyCIF-1A.** UMAP embeddings of

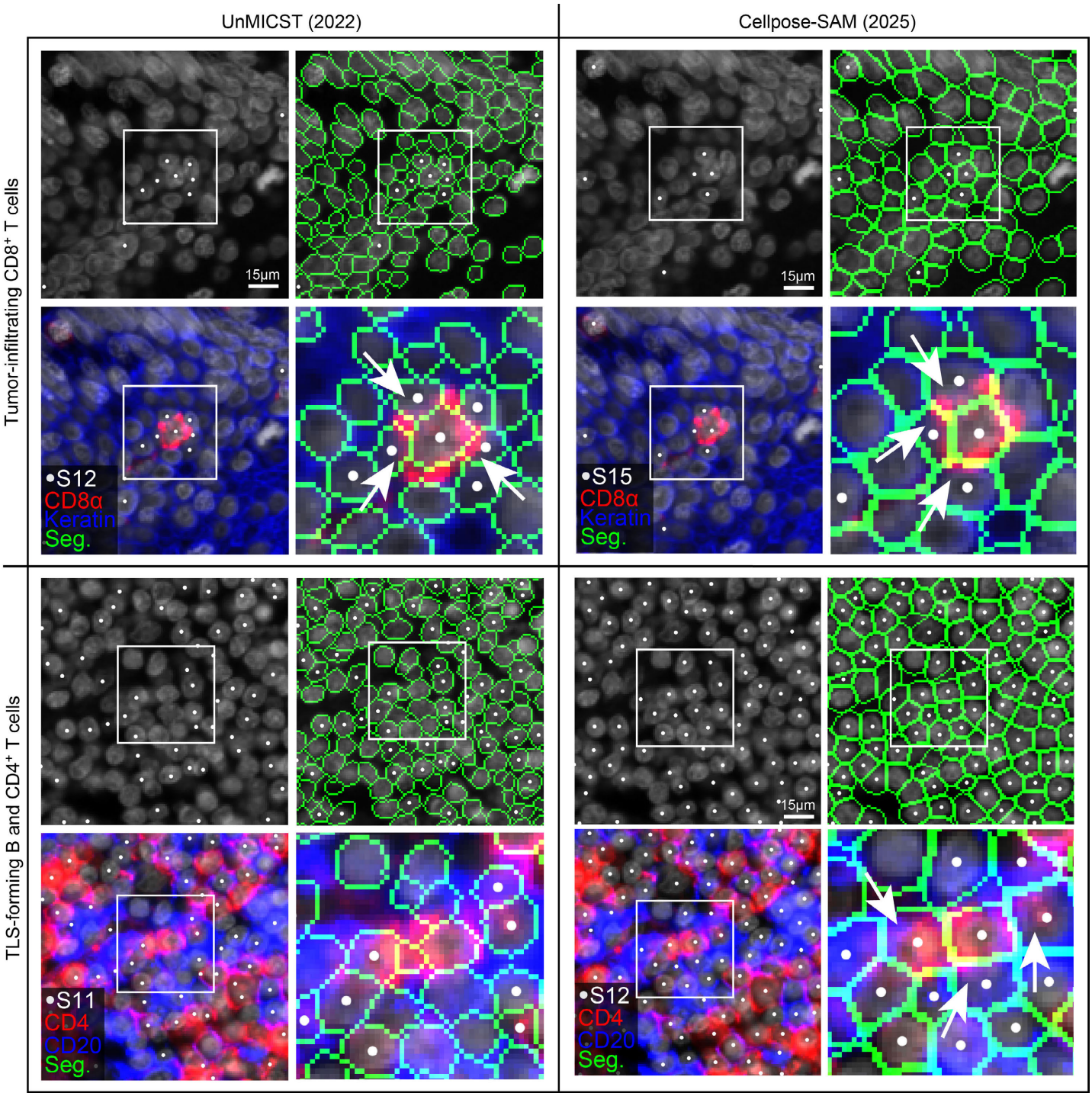
587 segmented cells colored by mean marker intensity for 21 immunomarkers.



Extended Data Fig. 2 | Location of segmentation-based Leiden clusters in CyCIF-1A. DNA (gray background) shown for reference.

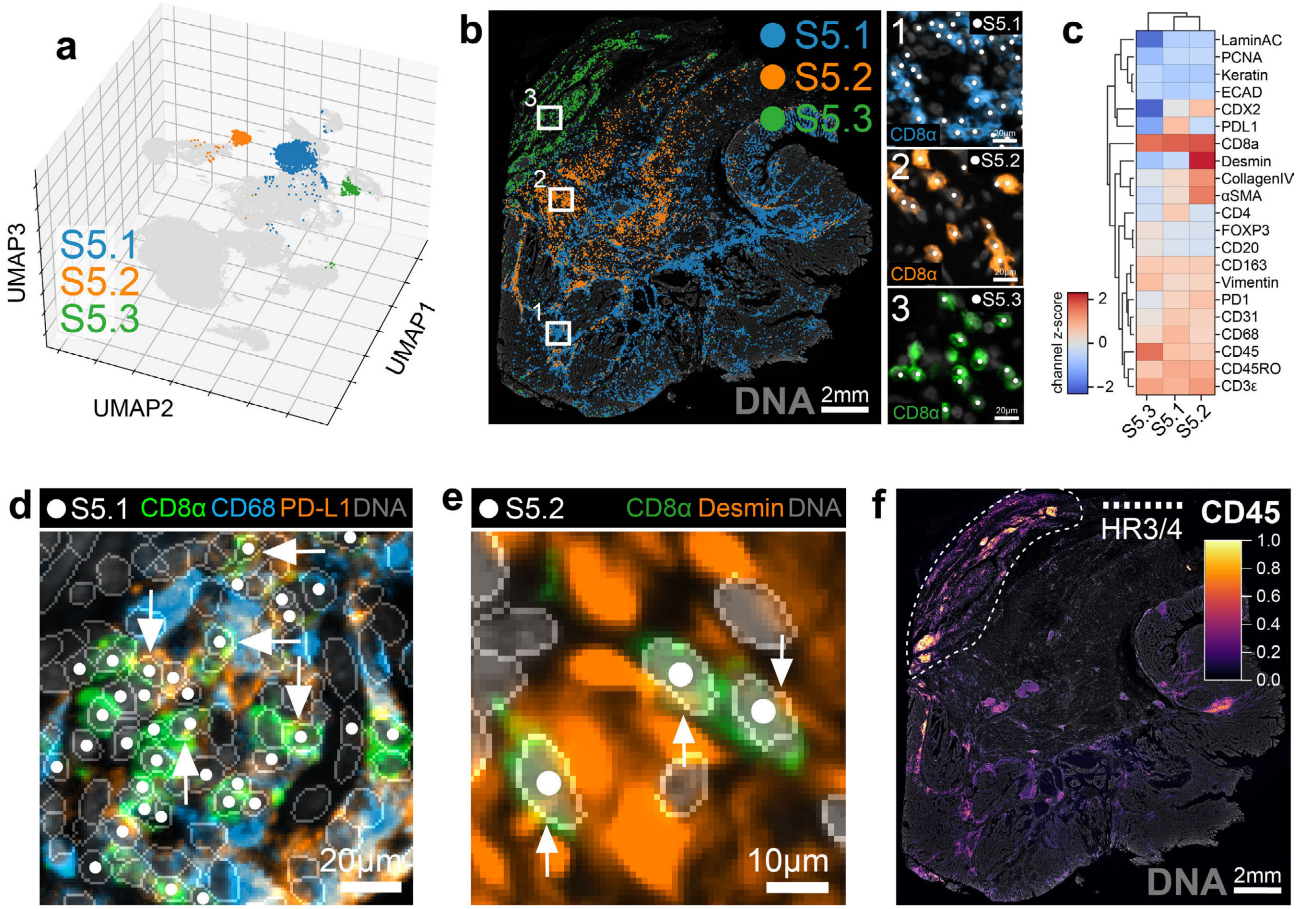


Supplementary Fig. 3 | Fields-of-view showing the location of segmentation-based Leiden clusters in CyCIF-1A. Paired CyCIF (left) and H&E (right) fields-of-view in CyCIF-1A showing the location of random examples of cells from all 18 Leiden clusters (green scatter points in CyCIF images).

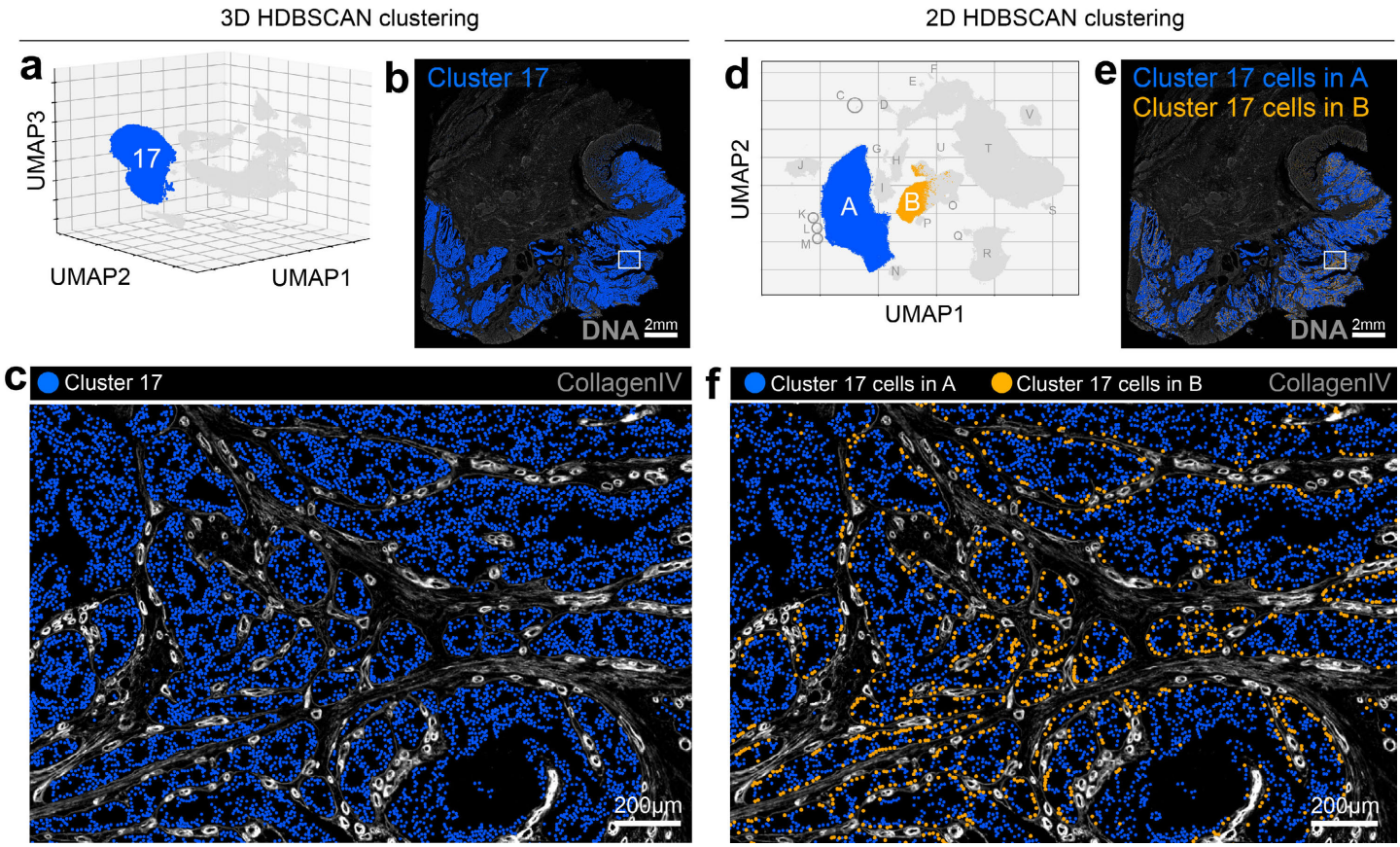


Supplementary Fig. 4 | CNN-based (UnMICST) and transformer-based (Cellpose-SAM) segmentation algorithms are affected by spatial crosstalk, independent of model accuracy. Since neighboring cells can have overlapping membrane extensions (e.g. lamellipodia) or overlapping cell bodies along the Z-axis, segmentation masks often include portions of adjacent cells of different types referred to as "spatial crosstalk". As a result, even state-of-the-art transformer-based segmentation methods like Cellpose-SAM[55] introduce unavoidable crosstalk errors when cell boundaries are imposed on an image. Top two rows: UnMICST (left

column) and Cellpose-SAM (right column) segmentation outlines in the same field-of-view in CyCIF-1A showing that, in both cases, tumor-infiltrating $CD8^+$ T cells (red) and neighboring $keratin^+$ tumor cells (blue) fall into the same cluster due to the sharing of signals across segmentation boundaries (white arrows) independent of segmentation quality. Bottom two rows: UnMICST (left column) and Cellpose-SAM (right column) segmentation outlines in the same field-of-view in CyCIF-1A showing that, in both cases, lymphoid aggregates consisting of B cells (blue) and $CD4^+$ T cells (red) also co-cluster due to sharing of signals across segmentation boundaries (white arrows); an unavoidable consequence of spatial crosstalk.



Supplementary Fig. 5 | CD8$^+$ T cells in segmentation cluster S5 are delocalized in UMAP space and affected by spatial crosstalk with cells in different tissue niches. a, 3D UMAP embedding of segmented cells from CyCIF-1A highlighting three subclusters of S5 cells: S5.1 (blue), S5.2 (orange), and S5.3 (green). **b**, Location of S5 subclusters in CyCIF-1A showing that they localize to different regions of tissue: S5.1 (peritumoral stroma, HR9), S5.2 (muscularis propria, HR2), S5.3 (sersosa/subserosa, HR3/4). Insets at right show that each subcluster represents a population of CD8$^+$ T cells. **c**, Clustered heatmap of z-score-normalized mean marker intensities for S5 subclusters computed relative to all 18 segmentation-based Leiden clusters (not only S5 subclusters) revealing differences in their protein expression profiles. **d**, Enlarged view of the region highlighted by box 1 in panel b revealing spatial crosstalk between CD8α$^+$ T cells from S5.1 (white scatter points) and neighboring CD68$^+$ PD-L1$^+$ macrophages. White arrows point to instances of signal spillover across segmentation boundaries. **e**, Enlarged view of the region highlighted by box 2 in panel b showing spatial crosstalk between CD8α$^+$ T cells from S5.2 cells (white scatter points) and neighboring Desmin$^+$ smooth muscle cells. White arrows point to instances of signal spillover across segmentation boundaries. **f**, CyCIF-1A as seen in the CD45 channel (colormap) showing that immune cells in the serosa/subserosa (HR3/4), in which CD8$^+$ T cells in S5.3 reside, exhibit higher CD45 intensity compared to other areas of the tissue, helping to explain why these cells form their own cluster.



Supplementary Fig. 6 | Cells affected by spatial crosstalk form discrete clusters on embedding in low-dimensional feature space. a, 3D UMAP embedding of segmented cells from CyCIF-1A with HDBSCAN cluster 17 highlighted in blue. (HDBSCAN is a density-based clustering algorithm that was applied directly to the embedded data points.) **b**, Location of cluster 17 cells in CyCIF-1A. **c**, Enlarged view of the region highlighted by the white box in panel b showing that cluster 17 cells do not include collagen IV$^+$ (gray) stromal cells at the tumor-stromal interface. **d**, 2D UMAP embedding of the same segmented cells from CyCIF-1A showing that cluster 17 cells from panel now form two discrete clusters (A, blue and B, orange) on embedding in 2D feature space. **e**, Location of cells from clusters A and B in CyCIF-1A. **f**, Enlarged view of the region highlighted by the white box in panel e showing that cluster B represents tumor cells at the tumor-stromal boundary affected by spatial crosstalk with collagen IV$^+$ (gray) stromal cells.

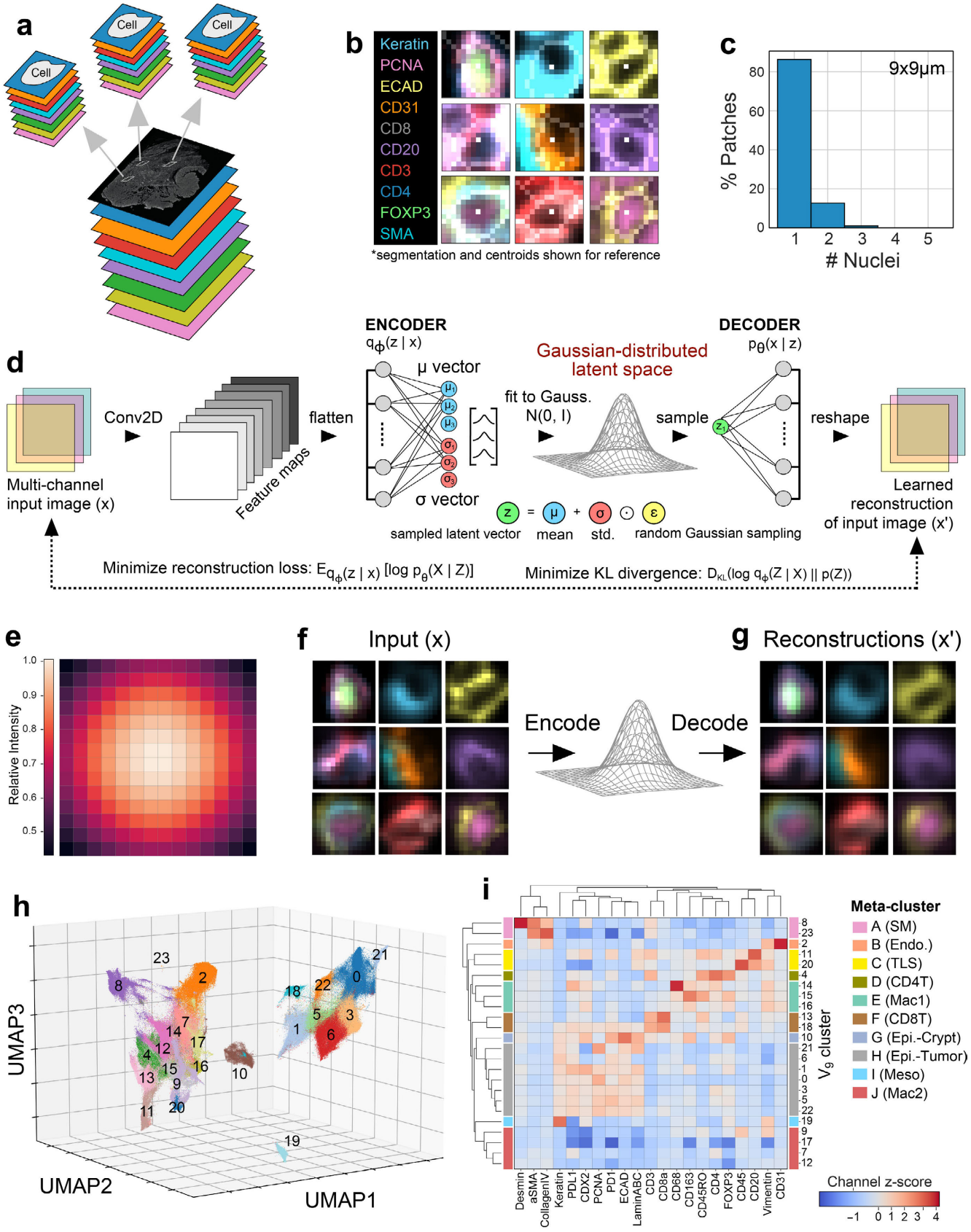
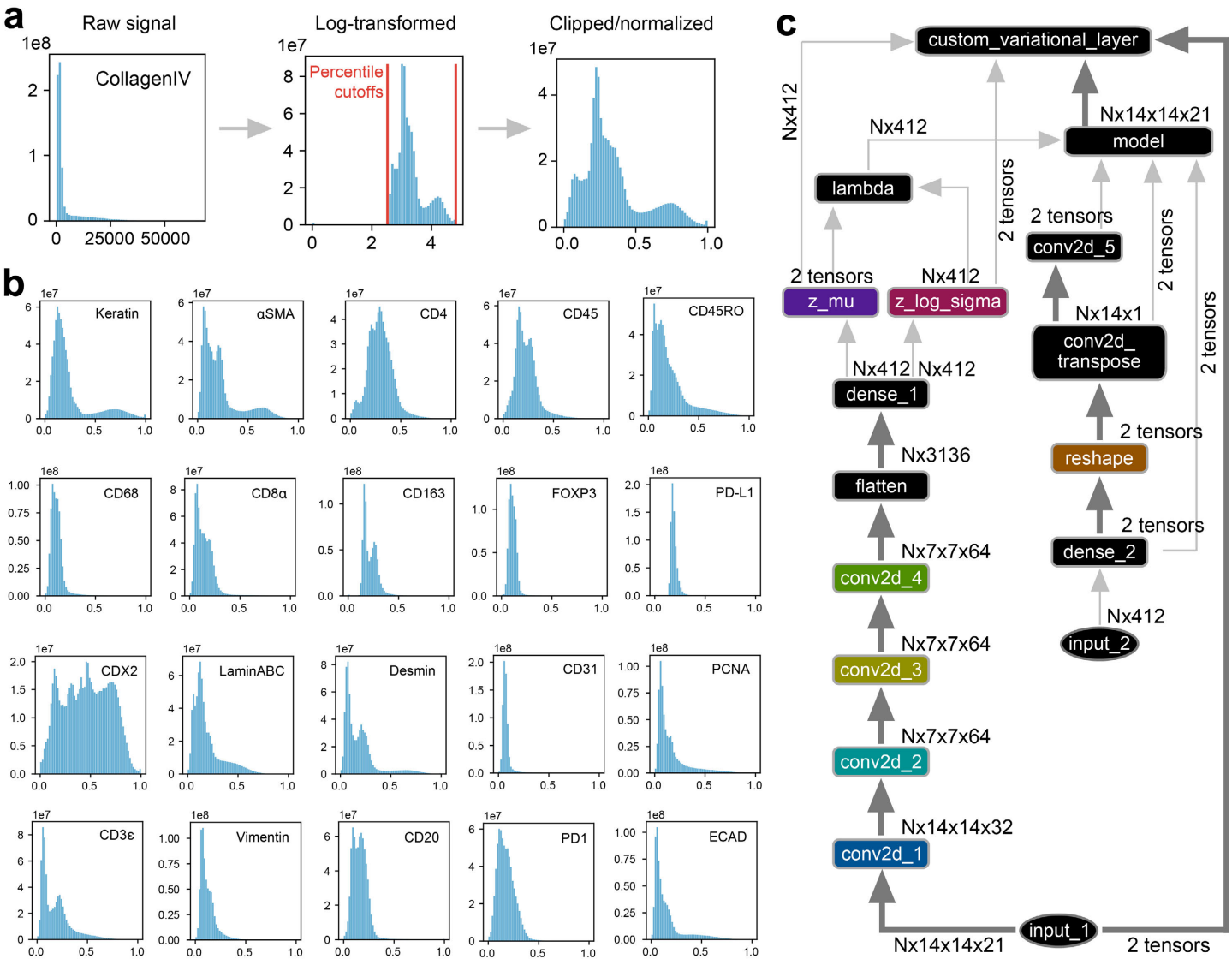


Fig. 2 | VAE-based classification of human CRC at single-cell resolution. a, Schematic showing the process of extracting multi-channel, single-cell image patches from whole-slide multiplex tissue images using

Cellcutter. **b**, Example 9x9µm multi-channel image patches extracted from CyCIF-1A. Nuclear centroids of reference cells at the center of each patch (white pixels) and segmentation boundaries (translucent perimeters) shown for reference only and not used during VAE model training or inference. **c**, Histogram showing the percentage of 9x9µm image patches from CyCIF-1A containing different numbers of cell nuclei (median = 1, Q1 = 1, Q3 = 1, IQR = 0). **d**, Schematic of a standard VAE network trained on multi-channel image patches. The encoding network is trained to map vector representations of input images onto a lower-dimensional, Gaussian-distributed latent space manifold. Image patch encodings are passed through the decoding network to visualize learned reconstructions whose similarity to corresponding input images is assessed using the average of two metrics: the mean squared error (MSE) and Kullback–Leibler (KL) divergence from a multi-dimensional Gaussian distribution. **e**, Heatmap showing relative pixel intensities of a 2D Gaussian-distributed vignette mask with a standard deviation of 7-pixels applied per-channel and used to focus the attention of the VAE network on the reference cell at the center of each 9x9µm image patch. **f**, Image patches shown in panel b after vignette masking, which are input into the VAE encoding network. **g**, Learned reconstructions of vignetted input images shown in panel f showing that information on protein expression and morphology have been learned by the network. **h**, 3D UMAP embedding of 9x9µm VAE image patch encodings from CyCIF-1A colored by Leiden cluster; see **Supplementary Video 2** for 360° plot rotation. **i**, Hierarchical clustering of z-score-normalized mean pixel-level marker intensities for 9x9µm VAE image patch Leiden clusters revealing 10 meta-clusters (A-J).

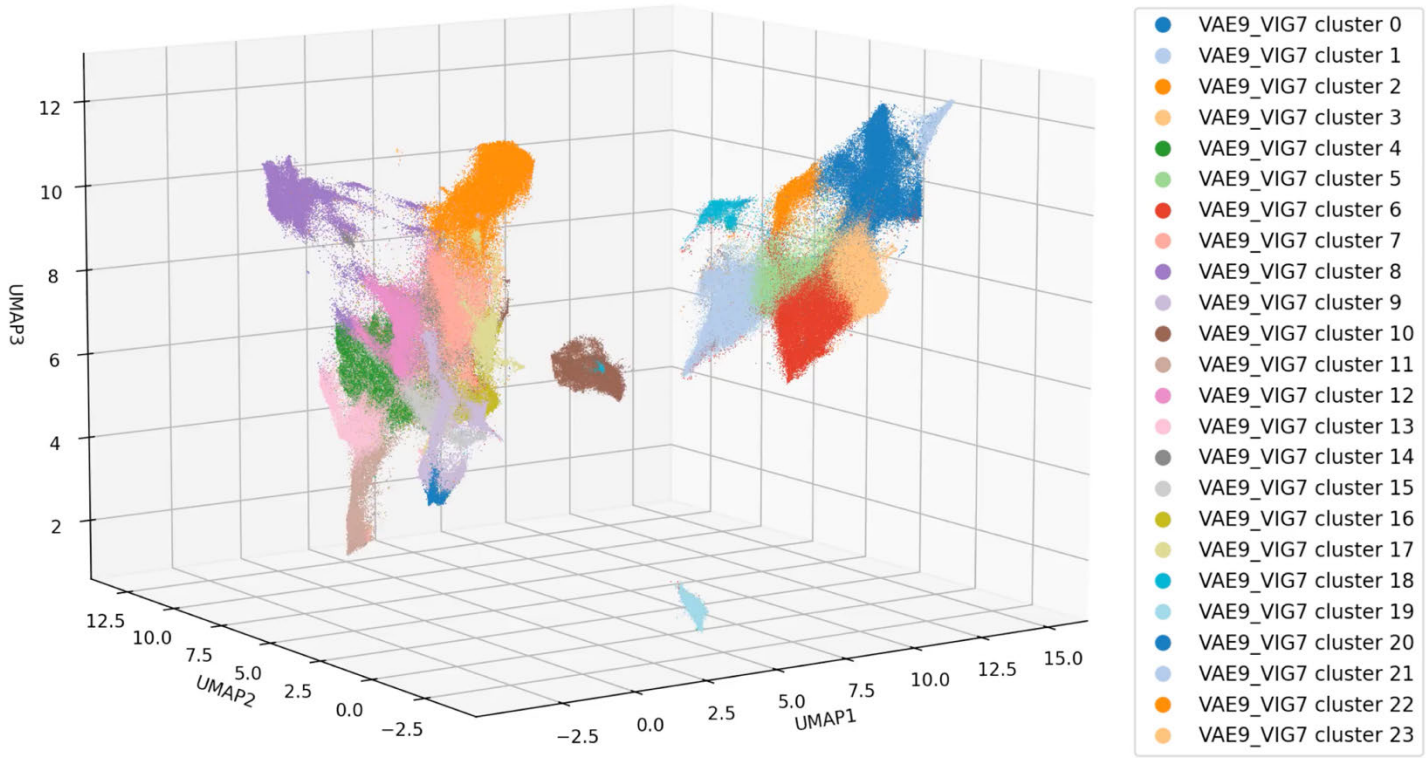


Extended Data Fig. 3 | VAE network architecture and image patch preprocessing. a, Preprocessing steps undergone by image patches prior to entering the VAE encoding network using the collagen IV channel as an example showing the raw pixel intensity distribution for the full CyCIF-1A image channel (right histogram), the same data after log-transformation (center histogram) and log-transformed data after clipping and 0-1 normalization (right histogram). Lower (0.17th percentile) and upper (99.99th percentile) clipping cutoffs are highlighted by vertical red bars in the center histogram. **b**, Clipped and normalized channel intensity distributions for the other 20 immunomarkers analyzed in this study. **c**, Schematic of the VAE network architecture used in this study. Conv2d = 2D convolutional layer, N = image patch batch size.

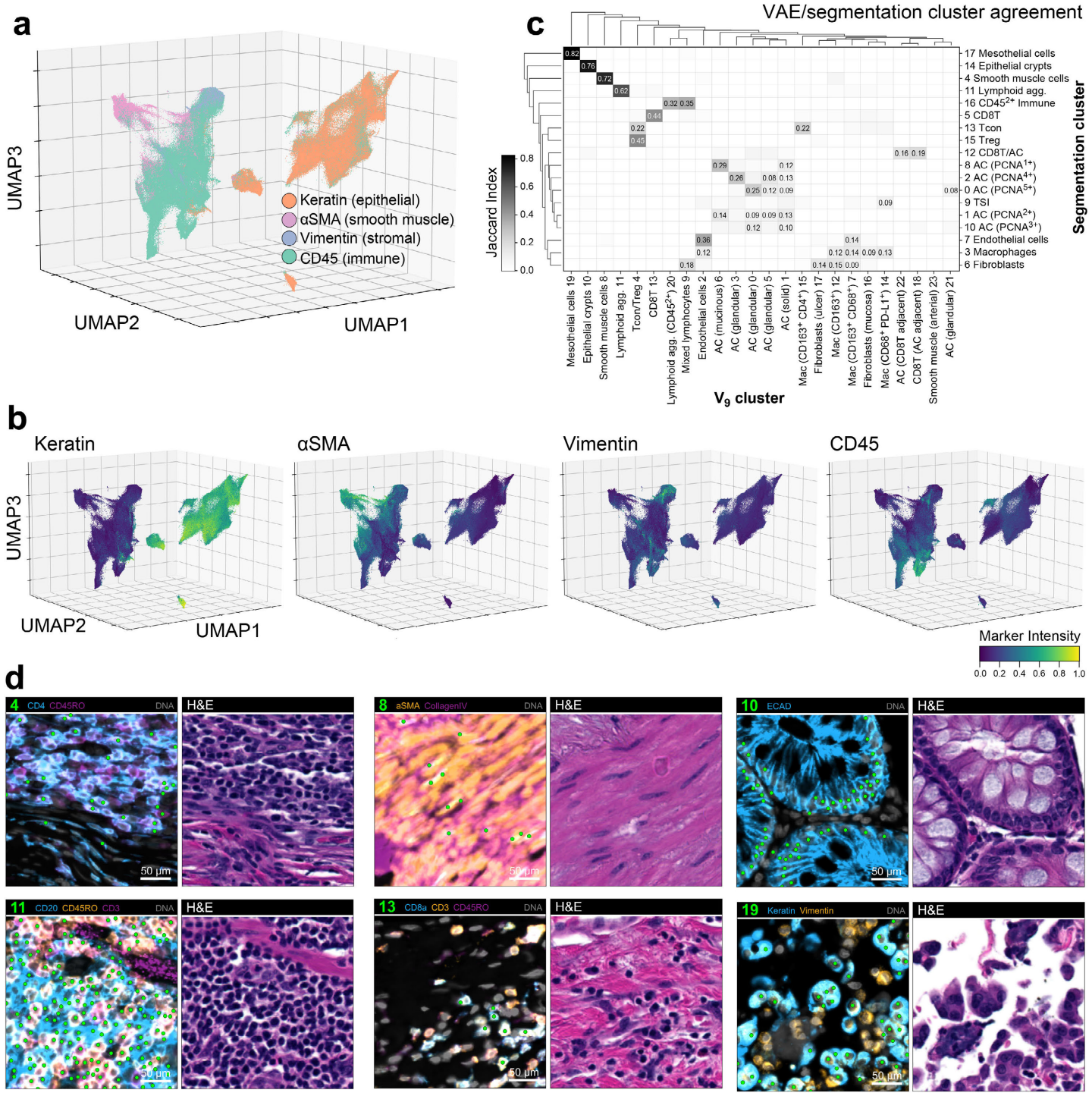
667

Leiden cluster	Annotation	% total cells	Histological region (HR)
0	Glandular adenocarcinoma (AC)	9.3	HR5 (glandular AC, predominently submucosal)
1	Solid AC	8.7	HR7 (solid AC)
2	Endothelial cells	7.3	Diffuse
3	Glandular AC	6.6	HR5 (glandular AC, predominently luminal/superficial)
4	$CD4^+$ T cells (Tcon/Treg)	6.4	Diffuse
5	Glandular AC	6.0	HR5 (glandular AC)
6	Mucinous AC	6.0	HR6 (mucinous AC)
7	Macrophages ($CD68^+$ $CD163^+$)	6.0	HR9 (peritumoral stroma)
8	Smooth muscle cells	5.3	HR2 (muscularis propria)
9	Mixed lymphocytes ($CD20^-$/$CD4^-$/$CD8^-$)	5.0	HR4 (subserosa)
10	Crypt-forming epithelial cells	5.0	HR1 (normal mucosa)
11	Mixed B and T lymphocytes	4.1	HR* (lymphoid aggregates)
12	Macrophages ($CD163^+$)	4.0	HR2 (muscularis propria)
13	$CD8^+$ T cells	3.7	Diffuse
14	Macrophages ($CD68^+$ PD-$L1^+$)	3.3	Tumoral/peritumoral
15	Macrophages ($CD163^+$ $CD4^+$)	2.7	Diffuse
16	Fibroblasts	2.7	HR2 (lamina propria of normal mucosa)
17	Fibroblasts	1.9	HR8 (ulcerated submucosa)
18	$CD8^+$ T cells (AC adjacent)	1.4	HR5,6,7 (glandular, mucinous, and solid AC)
19	Benign reactive mesothelial cells	1.3	HR3 (serosa)
20	Mixed lymphocytes	1.3	HR4 (subserosa)
21	Glandular AC	1.2	HR5 (glandular AC, predominently submucosal)
22	Mixed AC ($CD8^+$ T cell adjacent)	1.1	HR5,6,7 (glandular, mucinous, and solid AC)
23	Smooth muscle cells	0.1	HR2 (artery, muscularis propria)

668 **Supplementary Table 4 | 9x9µm VAE image patch Leiden cluster annotations.**

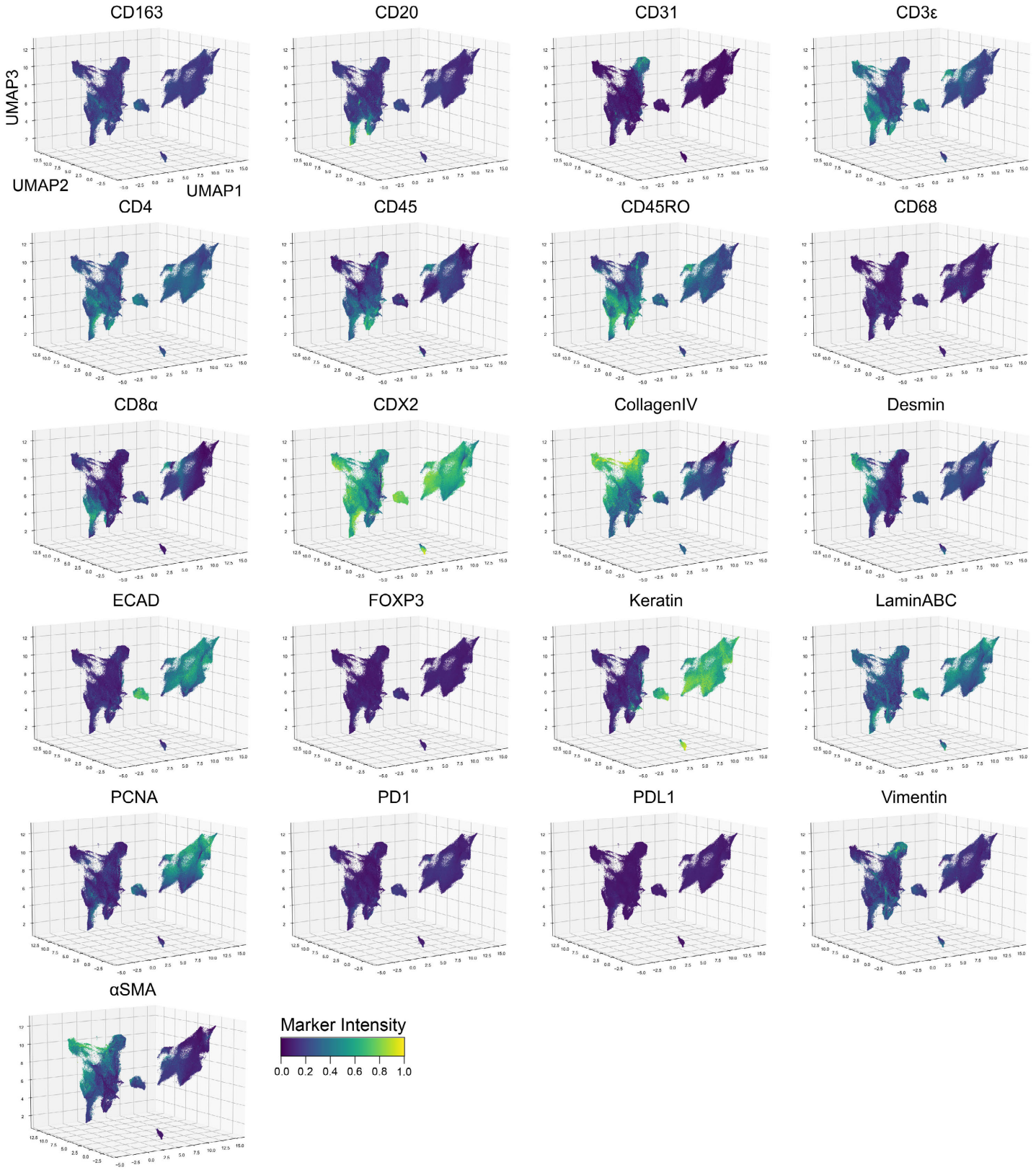


Supplementary Video 2 | 360° UMAP plot rotation showing 9x9µm VAE image patch encodings colored by Leiden cluster label.

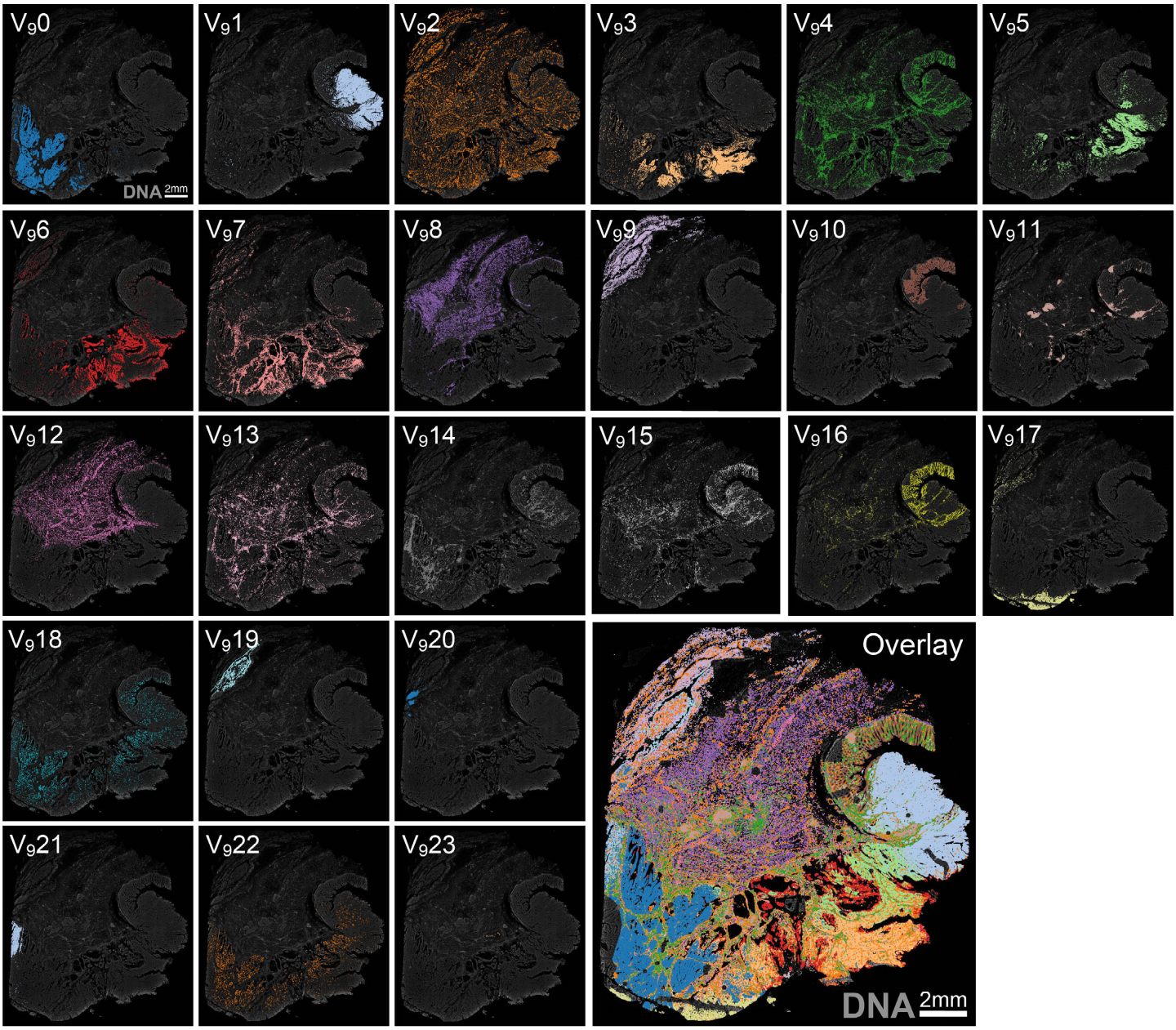


Extended Data Fig. 4 | VAE-based classification of human CRC at single-cell resolution. a, 3D UMAP embedding of 9x9µm VAE image patch encodings from CyCIF-1A colored by major cell lineage markers: keratin (epithelial, orange), αSMA (smooth muscle, pink), vimentin (stromal, blue), and CD45 (immune, green). **b**. 3D UMAP embedding of VAE image patch encodings from CyCIF-1A colored by pixel-level median marker intensity of major cell lineage markers keratin, αSMA, vimentin, and CD45. See **Supplementary Fig. 7** for colormaps of all 21 immunomarkers. **c**, Jaccard similarity indices (defined as the

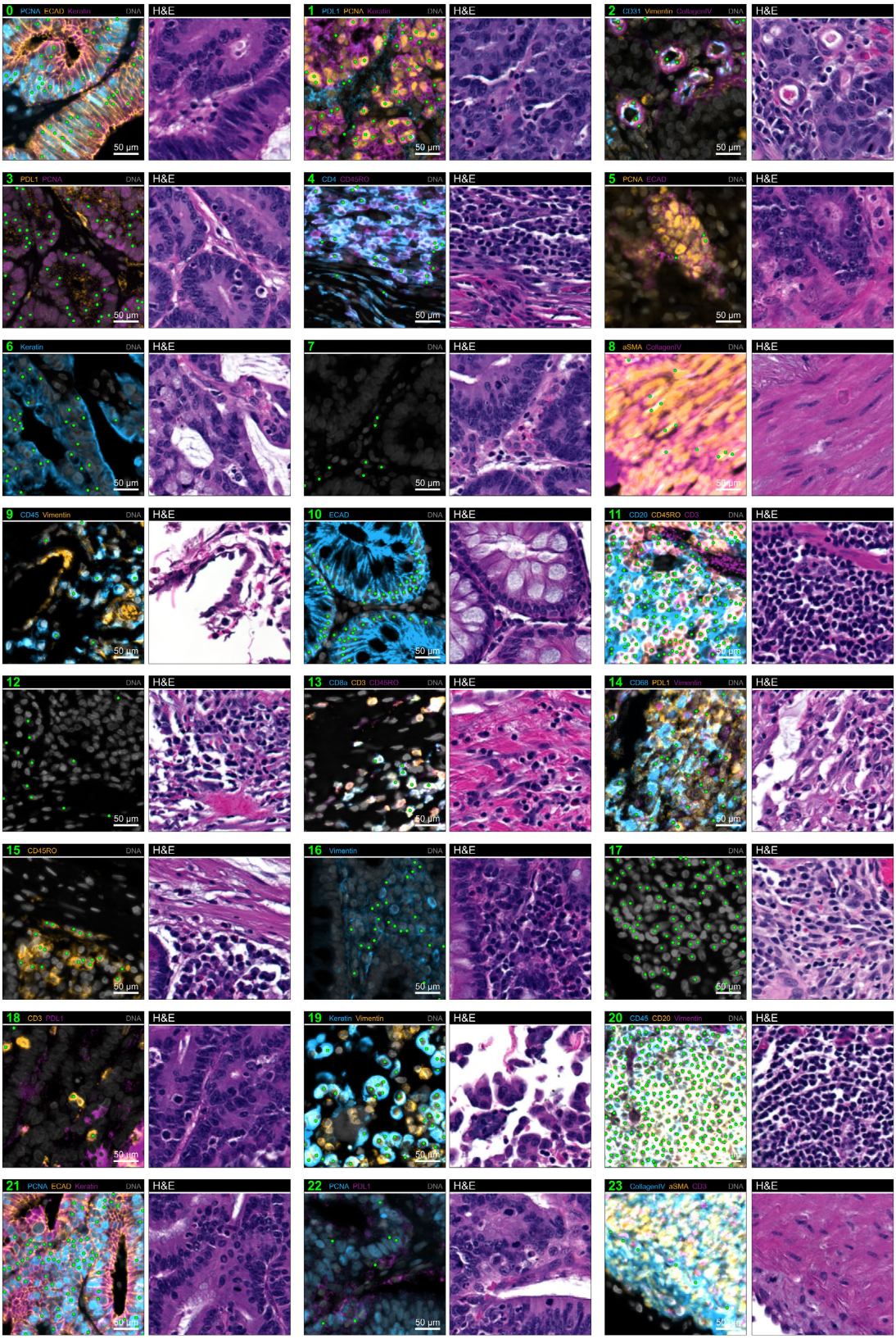
679 ratio of the intersection over the union of two sets) for cells in pairs of VAE and segmentation-based clusters.
680 Possible values range from 0 (no overlap) to 1 (perfect overlap). **d**, Paired CyCIF (left) and H&E (right) fields-
681 of-view from CyCIF-1A showing the location of random examples of cells from various Leiden clusters (green
682 scatter points in CyCIF images). See **Supplementary Fig. 8** for views of all 24 clusters.



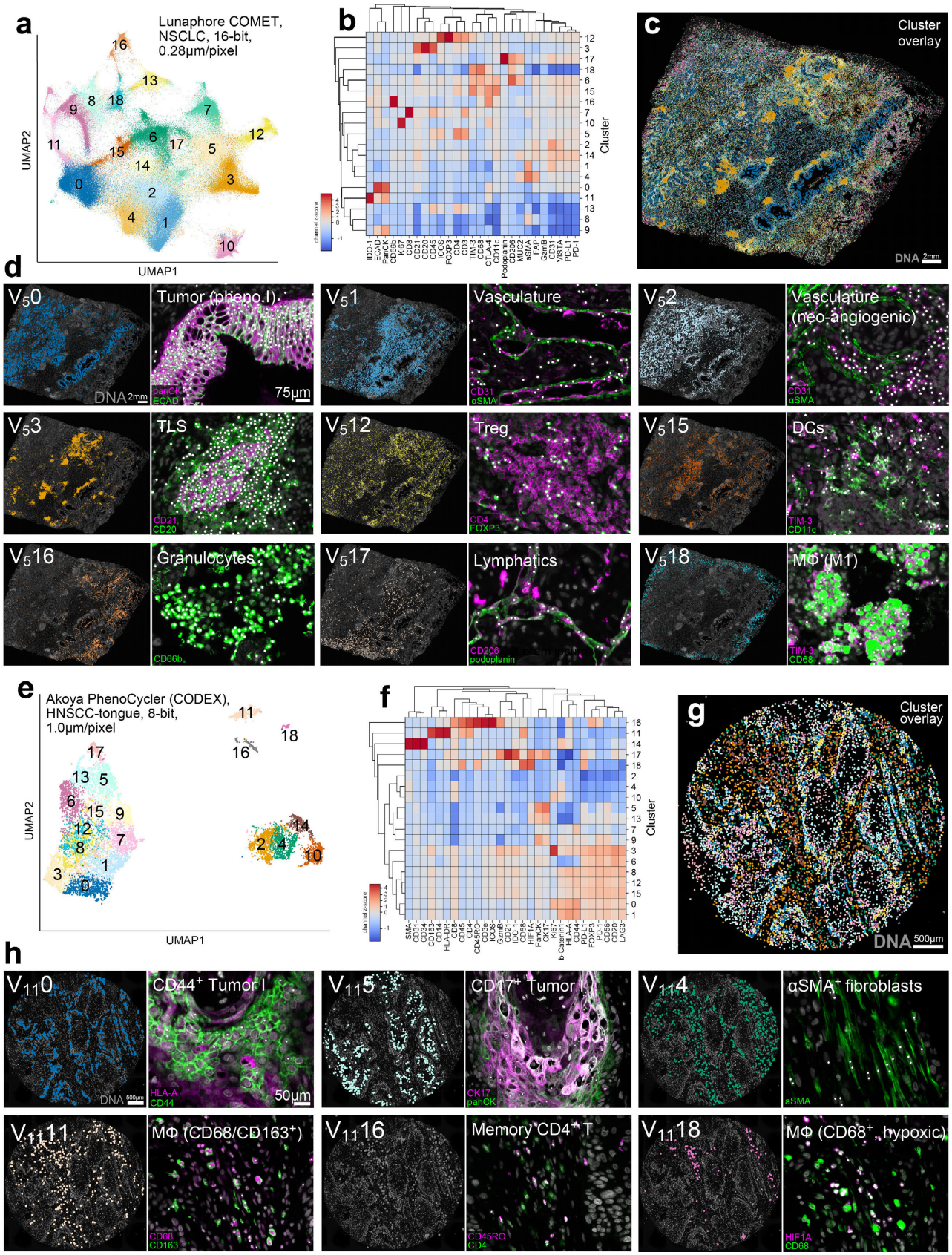
Supplementary Fig. 7 | Channel colormaps for 9x9µm VAE image patch encodings from CyCIF-1A. UMAP embeddings of VAE image patch encodings colored by pixel-level median marker intensity for 21 immunomarkers.



Extended Data Fig. 5 | Location of 9x9µm VAE image patch Leiden clusters in CyCIF-1A. DNA (gray background) shown for reference.

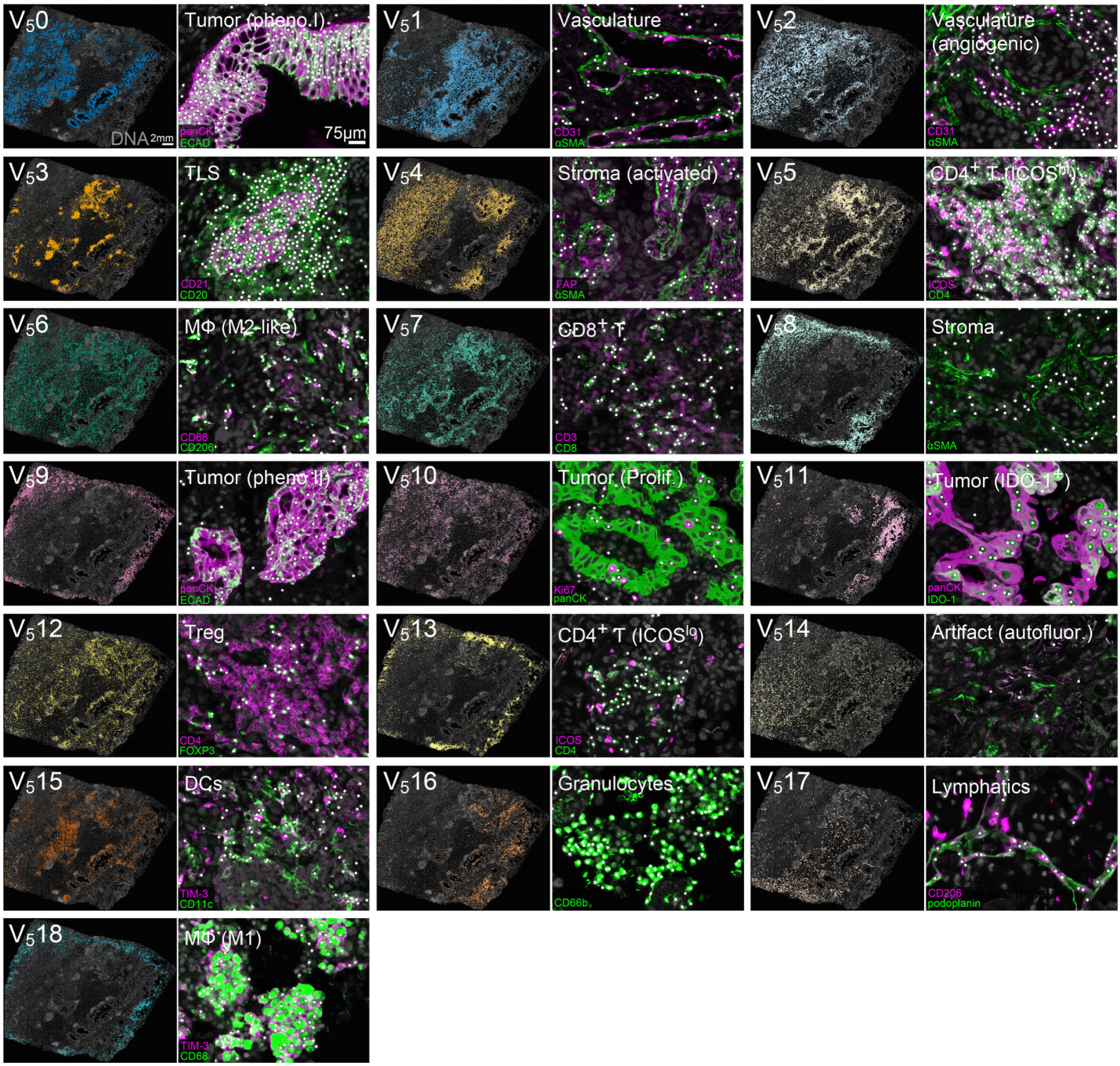


Supplementary Fig. 8 | Fields-of-view showing the location of 9x9µm VAE image patch Leiden clusters in CyCIF-1A. Paired CyCIF (left) and H&E (right) fields-of-view in CyCIF-1A showing the location of random examples of cells from all 24 Leiden clusters (green scatter points in CyCIF images).

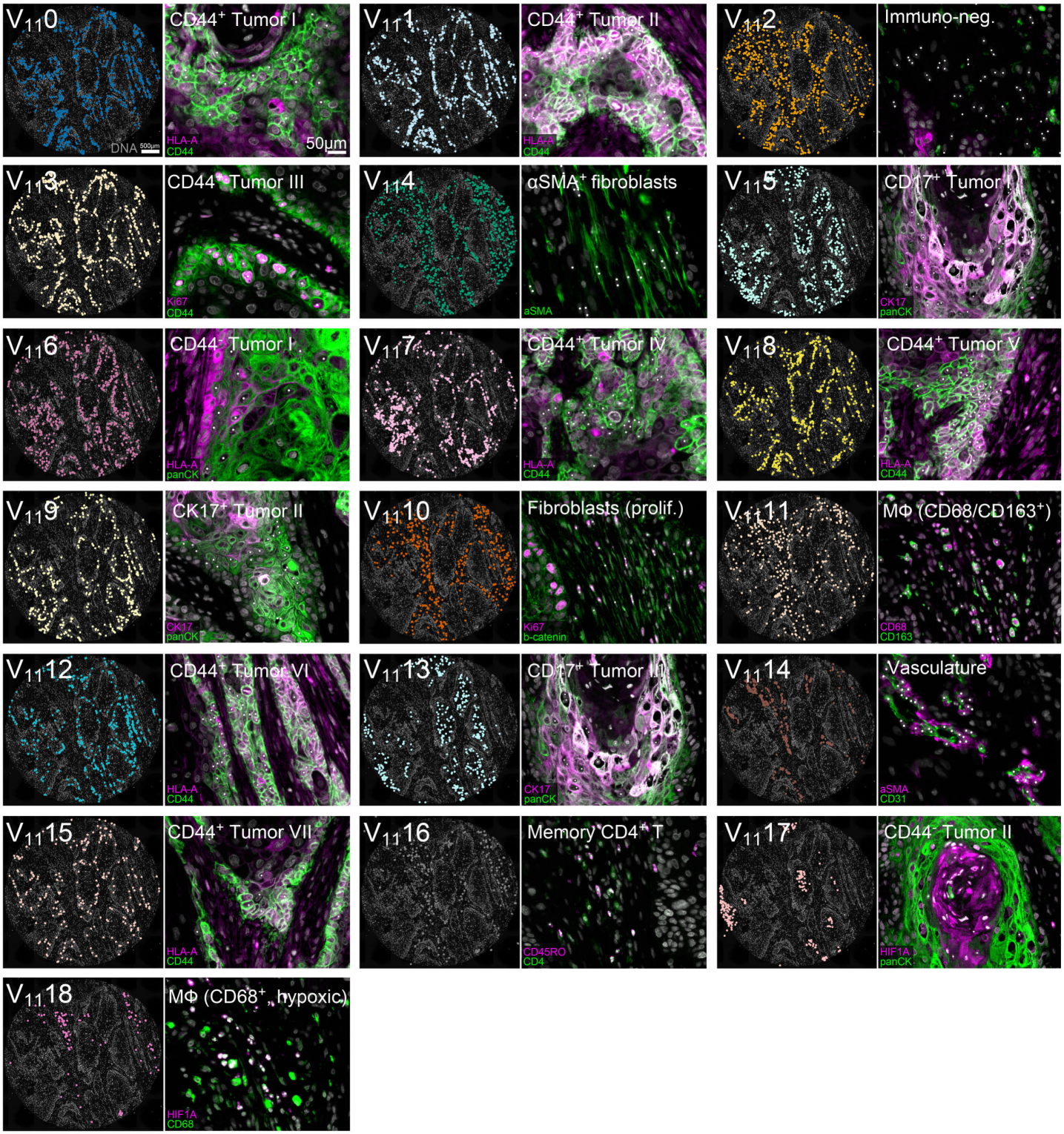


Extended Data Fig. 6 | VAE-based classification of Lunaphore and PhenoCycler data. a-d, 27-plex Lunaphore COMET imaging of human NSCLC: **a**, 2D UMAP embedding of 5x5µm VAE image patch

encodings colored by Leiden cluster. **b**, Hierarchical clustering of z-score-normalized mean pixel-level marker intensities for Leiden clusters shown in panel a. **c**, Location of Leiden clusters in Lunaphore-1; DNA (gray) shown for reference. **d**, Examples of individual clusters showing their locations (left panels) and phenotypes (right panels) as seen in characteristic image channels. See **Supplementary Fig. 9** for views of all 19 clusters. **e-h**, 29-plex Akoya PhenoCycler imaging of human tongue squamous cell carcinoma: **e**, 2D UMAP embedding of 11x11µm VAE image patch encodings colored by Leiden cluster. **f**, Hierarchical clustering of z-score-normalized mean pixel-level marker intensities for Leiden clusters shown in panel e. **g**, Location of Leiden clusters in CODEX-1; DNA (gray) shown for reference. **h**, Examples of individual clusters showing their locations (left panels) and phenotypes (right panels) as seen in characteristic image channels. See **Supplementary Fig. 10** for views of all 19 clusters.



Supplementary Fig. 9 | Locations and phenotypes of 5x5µm VAE image patch Leiden clusters in Lunaphore-1. Left panels show the overall spatial distribution of clustering cells with DNA (gray) for reference. Right panels show examples of clustering cells (white points) as seen in their characteristic marker channels.



Supplementary Fig. 10 | Locations and phenotypes of 11x11µm VAE image patch Leiden clusters in CODEX-1. Left panels show the overall spatial distribution of clustering cells with DNA (gray) for reference. Right panels show examples of clustering cells (white points) as seen in their characteristic marker channels.

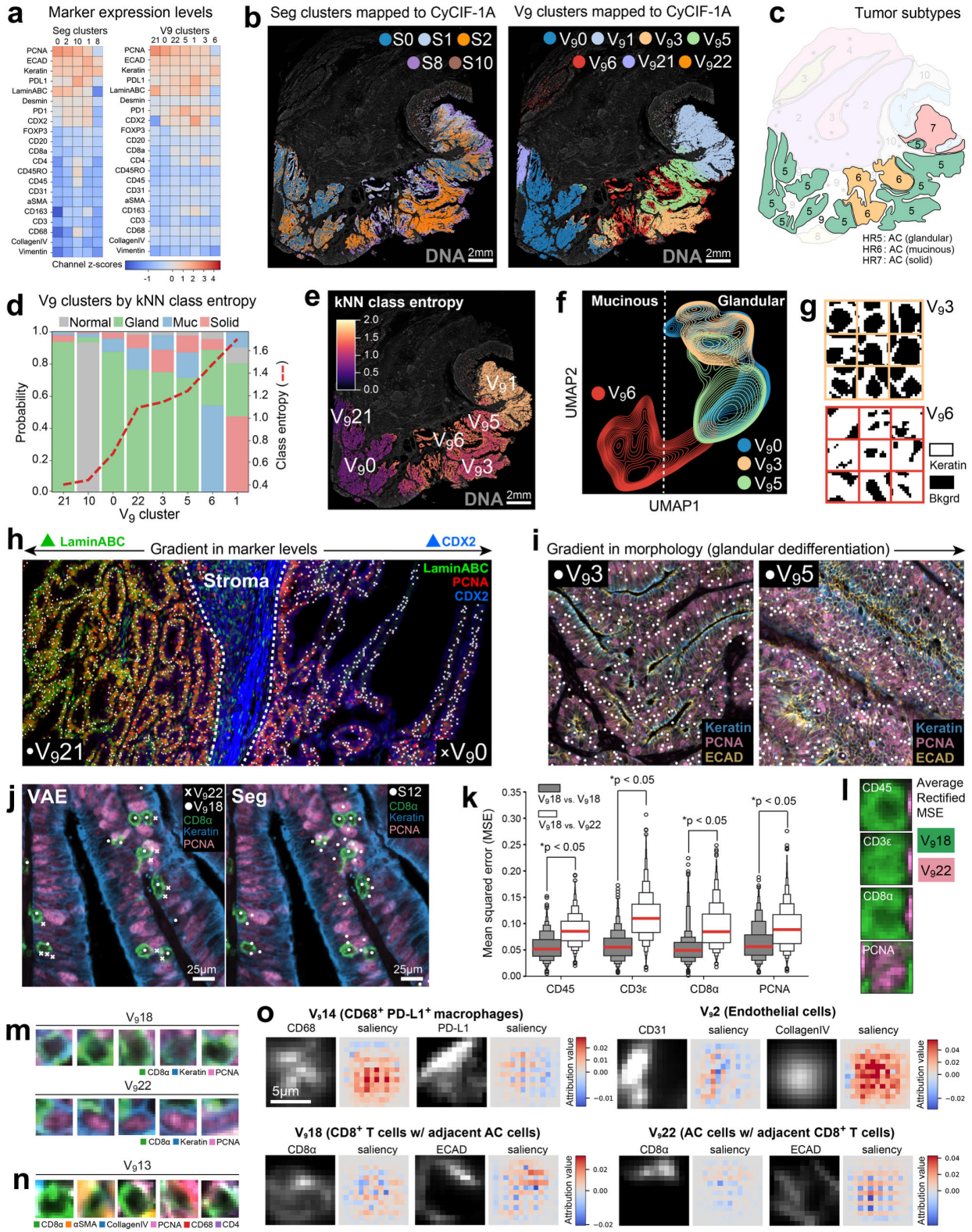
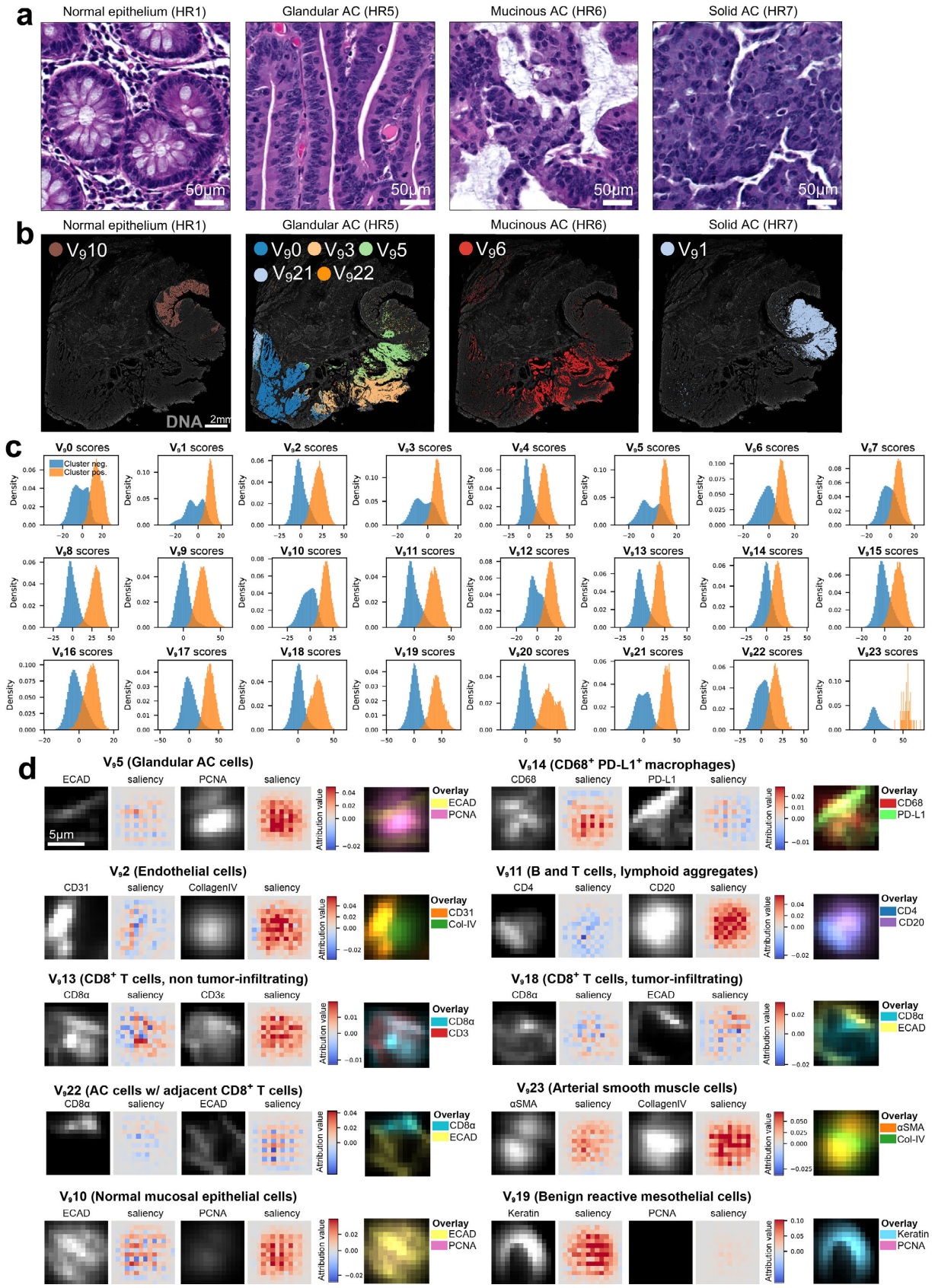


Fig. 3 | VAE image patch classification is morphology-aware and overcomes spatial crosstalk. a, z-score normalized protein expression profiles for tumor-associated segmentation (left) and VAE (right) Leiden clusters. **b**, Location of tumor-associated segmentation (left) and VAE (right) Leiden clusters in CyCIF-1A showing that V_9 clusters are organized into spatially defined regions throughout the tumor domain. **c**, Histology

map highlighting the locations of different AC subtypes observed in the H&E image: glandular (green), mucinous (orange), and solid (red; **Fig. 1b**). **d**, Stacked bar chart showing normalized mean posterior probability scores (left y-axis) for kNN classes corresponding to normal epithelium (gray) and glandular (green), mucinous (blue), and solid AC (red) among epithelial V_9 clusters sorted from left to right in ascending rank order based on class entropy (right y-axis). **e**, Location of cells from tumor-associated VAE clusters from meta-cluster H in CyCIF-1A colored by kNN class entropy revealing a spatial gradient from lower entropy glandular AC at the left of the image to higher entropy mucinous and solid AC at the right. **f**, UMAP embedding showing kernel density estimates for binarized VAE image patch encodings of the keratin channel from tumor-associated V_9 clusters 0, 3, 5, and 6. **g**, Example binary keratin image patches from V_9 clusters annotated as glandular ($V_9$3) and mucinous AC ($V_9$6) showing their morphological differences. **h**, Field-of-view in CyCIF-1A at the interface between $V_9$21 (left of stroma region) and $V_9$0 (right of stroma region) showing that these two clusters differ in their expression of LaminABC (green), PCNA (red), and CDX2 (blue). **i**, Fields-of-view in CyCIF-1A of glandular AC clusters $V_9$3 (left) and $V_9$5 (right) as seen in channels keratin (blue), PCNA (pink), and ECAD (orange) revealing morphological difference in the density and organization of the glandular structures formed by these cells. **j**, Matched fields-of-view in CyCIF-1A as seen in the CD8α (green), keratin (blue), and PCNA (pink) channels showing that VAE classification correctly distinguishes between tumor-infiltrating $CD8^+$ T cells ($V_9$18) and neighboring $keratin^+$ epithelial cells ($V_9$22; left image) that co-cluster in segmentation cluster S12 due to spatial crosstalk (right image). Note: due to random data subsampling, not all data points match between left and right images. **k**, Boxen plots showing mean squared error (MSE) distributions among pairs of $V_9$18 (gray) image patches and between $V_9$18 and $V_9$22 (white) image patches for channels CD3ε, CD45, CD8α, and PCNA. Horizontal red lines indicate the median MSE at the center quantile of each distribution ($p<0.05$, two-sided Mann-Whitney U-test). **l**, Average rectified pixel-wise MSE for channels CD3ε, CD45, CD8α, and PCNA in $V_9$18 (green) and $V_9$22 (pink) image patches revealing CD45, CD3ε, and CD8α signals appear at the center of $V_9$18 patches and towards the periphery of $V_9$22 patches; the opposite is true for PCNA. This indicates that reference cells in $V_9$18 patches represent $CD8^+$ T cells and those in $V_9$22 patches represent $PCNA^+$ tumor cells. **m**, Image patch galleries confirming that cluster $V_9$18 cells represent $CD8\alpha^+$ (green) T cells (top row) with surrounding $PCNA^+$ (pink) $keratin^+$ (blue) tumor cells while $V_9$22 represents the opposite scenario (bottom row). **n**, Image patch gallery showing that cells in $V_9$13 represent non-tumor infiltrating $CD8^+$ T cells neighbored by diverse cell types as evidenced by variable immunomarker expression at patch borders. **o**, Grayscale channels (left images) and corresponding concept saliency maps (right images) for image patches within the 95th percentile of positive concept score distributions from four V_9 clusters showing characteristic channels for each.



Extended Data Fig. 7 | Morphology and concept saliency analysis for 9x9µm VAE image patch Leiden clusters. | a, Fields-of-view from the H&E image (**Fig. 1b**) showing examples of various normal and cancerous

epithelial histologies. **b**, Location of epithelial V_9 clusters in CyCIF-1A grouped by histological region. See (**Fig. 1c**) for histological annotations. **c**, Distributions of concept scores (Sc), defined as the dot product between the ith VAE image patch encoding (*Zi*) and the corresponding cluster concept vector (*Zc*) such that $Sc = Zi \cdot Zc$ for patches with (orange) and without (blue) a given cluster label. **d**, Channel intensities (left images) and corresponding concept saliency maps (right images) for 10 different V_9 clusters. Overlay color images shown for reference; see **Supplementary Image Gallery 2** for fifteen examples of saliency maps across all 21 channels for all 24 V_9 clusters.

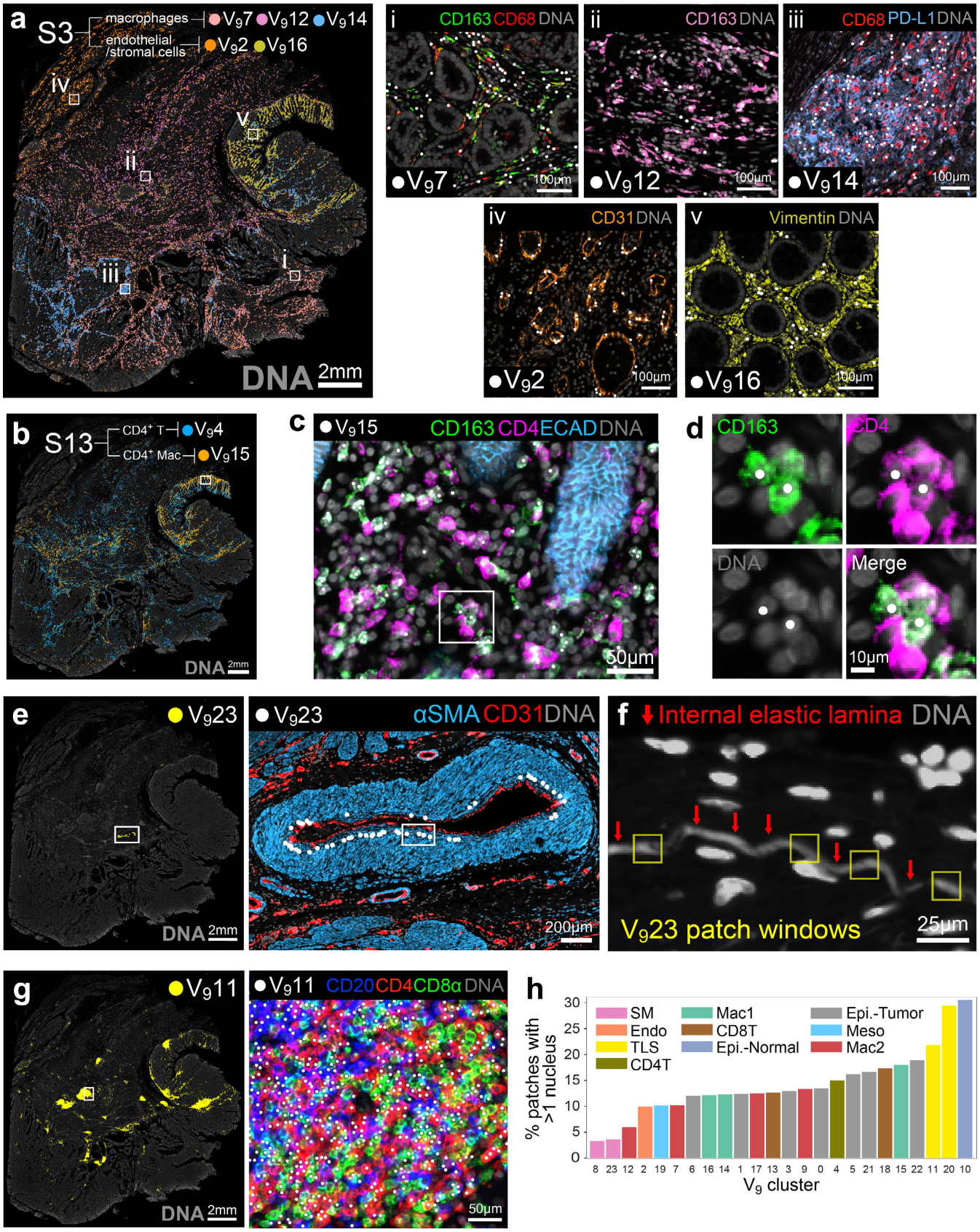
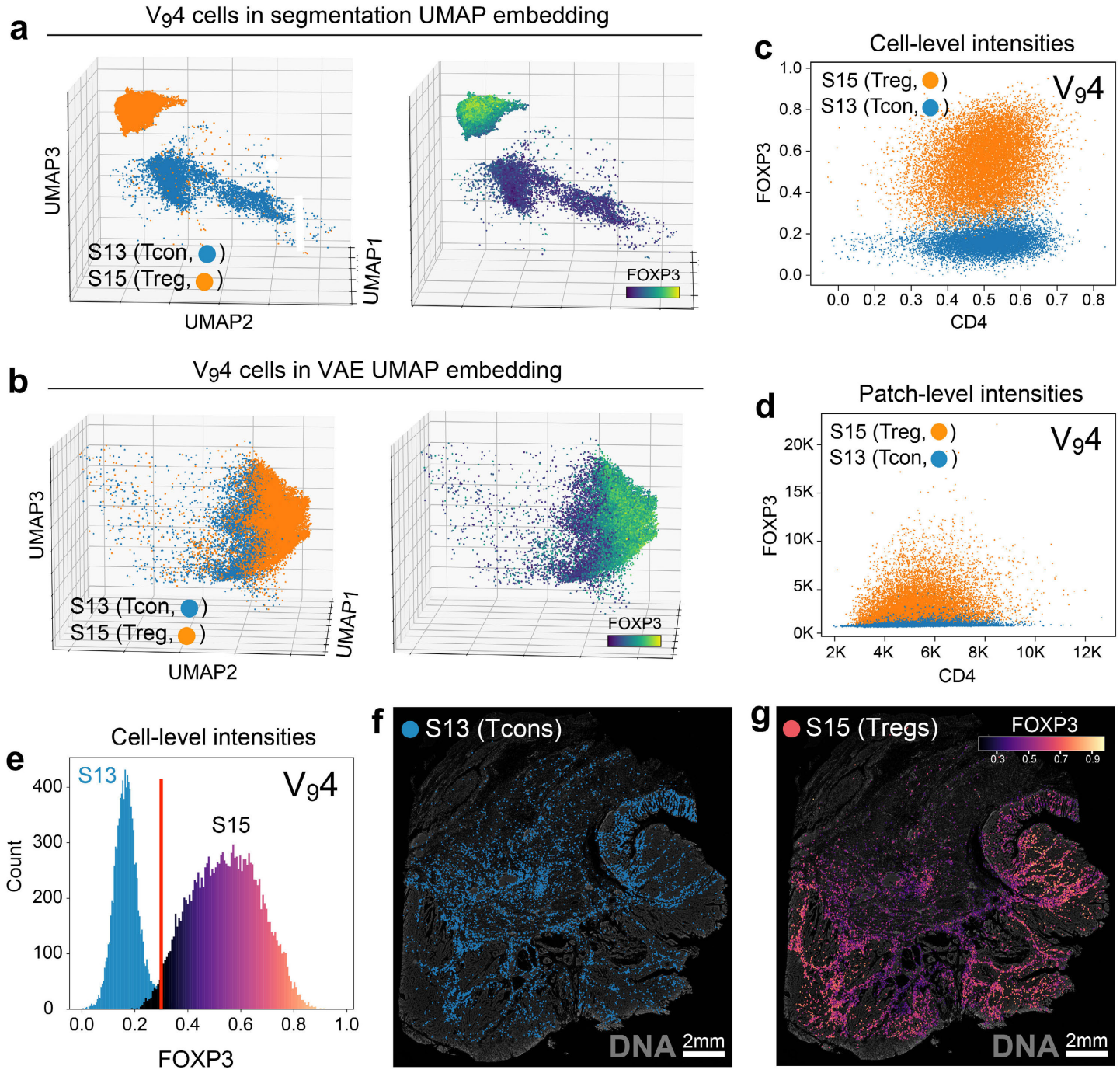


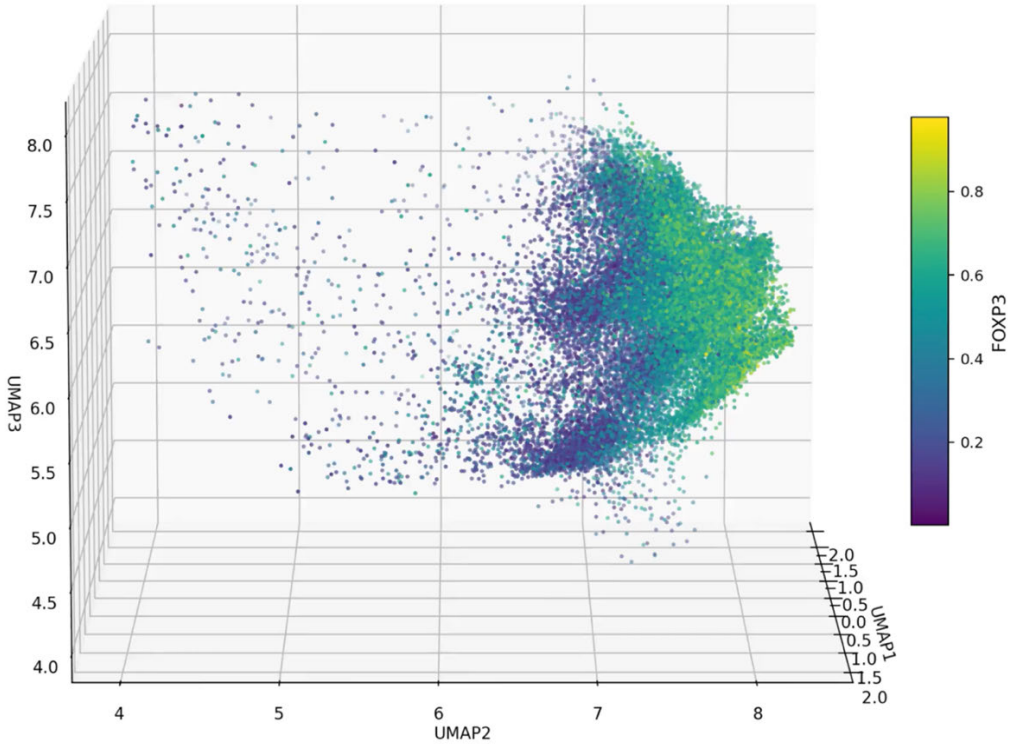
Fig. 4 | Strengths and limitations of patch-based learning for single-cell phenotyping. a, Left panel: Location of segmentation cluster S3 in CyCIF-1A colored by V_9 cluster labels showing multiple subclusters

identified by the VAE. Right panels: Fields-of-view corresponding to white boxes in left image showing example cells (white scatter points) identified by the VAE as macrophages (i-iii), fibroblasts (iv), and endothelial cells (v). **b**, Location of cells in segmentation cluster S13 in CyCIF-1A corresponding to $V_9$4 ($CD4^+$ T cells) and $V_9$15 ($CD4^+$ macrophages) in VAE analysis. **c**, Enlarged view of the region highlighted by the white box in panel b as seen in the CD163 (green), CD4 (magenta), and ECAD (blue) channels confirming that $V_9$15 represents $CD163^+$ $CD4^+$ macrophages in the lamina propria of normal mucosal crypts (HR1); DNA (gray) channel shown for reference. **d**, Enlarged view of the region highlighted by the white box in panel c showing $CD163^+$ $CD4^+$ macrophages interacting with $CD163^-$ $CD4^+$ helper T cells; channels shown separately for clarity. **e**, Left panel: Location of $V_9$23 cells in the muscularis propria (HR2) of CyCIF-1A. Right panel: Enlarged view of the region highlighted by the white box in left image as seen in the αSMA (blue) and CD31 (red) channels showing that $V_9$23 represents smooth muscle cells proximal to the abluminal wall of an artery. **f**, Enlarged view of the region highlighted by the white box in the image at left as seen in the DNA (gray) channel highlighting the internal elastic lamina (red arrows), an autofluorescent acellular structure present in $V_9$23 image patches (yellow boxes). **g**, Left image: Location of $V_9$11 cells comprising lymphoid aggregates in CyCIF-1A. Right image: Enlarged view of the region highlighted by the white box in the image at left as seen in the CD20 (blue), CD4 (green), and CD8α (red) channels showing that $V_9$11 comprises a mixed of $CD20^+$ B cells, $CD4^+$ T cells, and $CD8^+$ T cells. **h**, Bar plot showing the percentage of V_9 cluster cells with greater than one nucleus colored by meta-cluster labels from (**Fig. 2i**). Note that image patch clusters associated with lymphoid aggregates ($V_9$11 and $V_9$20, yellow bars) show a high proportion of cells with multiple nuclei, helping to explain why B and T cells comprising these structures do not form discrete clusters



Extended Data Fig. 8 | Tregs and Tcons exhibit greater similarity at the image patch level compared to the single-cell level. a, Distribution of $V_9$4 cells in the segmentation-based UMAP embedding colored by S13 (Tcons, blue) and S15 (Tregs, orange) in the left plot and FOXP3 intensity in the right plot. **b**, Distribution of $V_9$4 cells in the VAE-based UMAP embedding colored by S13 (Tcons, blue) and S15 (Treg, orange) in the left plot and FOXP3 intensity in the right plot. **c**, Scatter plot showing the distribution of $V_9$4 cells according to CD4 (x-axis) and FOXP3 (y-axis) intensity at the level of segmented cells colored by S13 (Tcons, blue) and S15 (Tregs, orange) showing that the two cell types form discrete populations along the FOXP3 axis. **d**, Scatter

794 plot showing the distribution of $V_9$4 cells according to CD4 (x-axis) and FOXP3 (y-axis) intensity in 9x9µm
795 image patches colored by S13 (Tcons, blue) and S15 (Tregs, orange) showing that the FOXP3 axis no longer
796 exhibits a bimodal distribution. **e**, Distribution of $V_9$4 cells according to FOXP3 intensity colored by S13
797 (Tcons, blue) and S15 (Tregs, colormap) again revealing a bimodal distribution for the two cell populations
798 according to FOXP3 at the level of segmented cells. Red vertical line indicates a manually defined gate
799 distinguishing positive (right of gate) and negative (left of gate) cell populations. **f**, Location of gated S13 cells
800 (Tcons, FOXP3$^-$ cells in panel e) in CyCIF-1A. **g**, Location of gated S15 cells (Tregs, FOXP3$^+$ cells in panel e)
801 in CyCIF-1A revealing that Tregs occupy similar spatial niches as Tcons in CyCIF-1A and tend to express
802 higher FOXP3 levels in tumor regions (HR5, 6, and 7).



Supplementary Video 3 | 360° UMAP plot rotation showing 9x9µm VAE image patch encodings for VAE cluster 4 ($V_9$4) colored by FOXP3 expression.

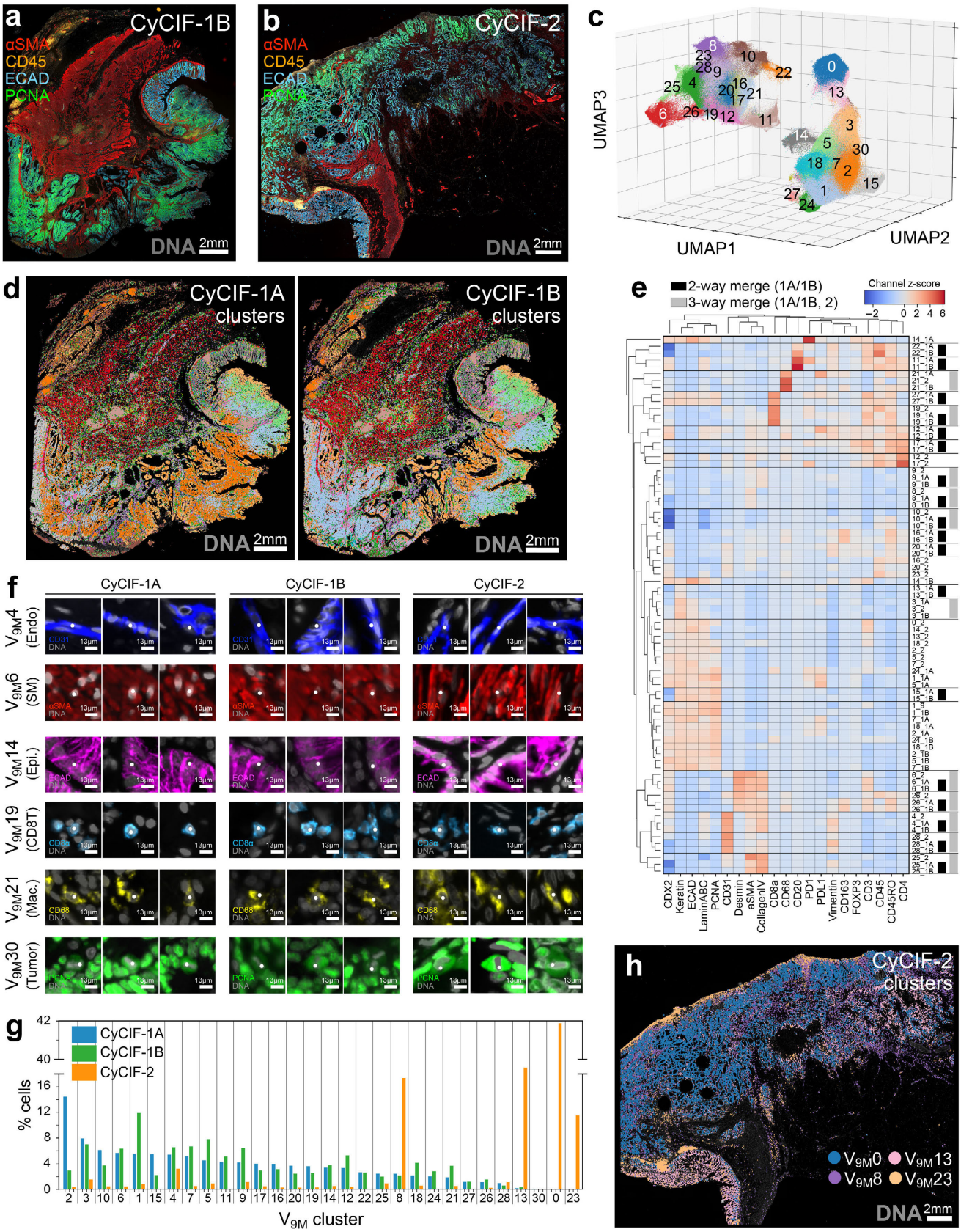
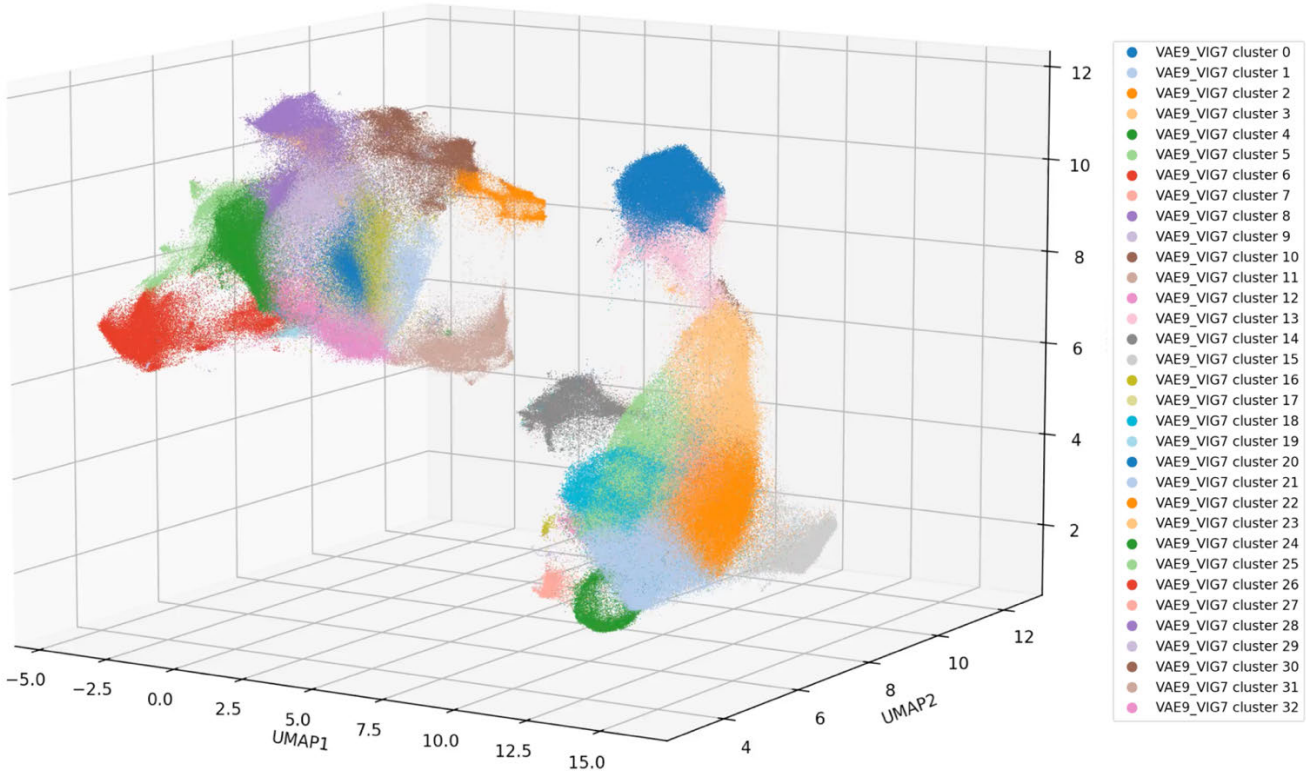


Fig. 5 | VAE analysis identifies matched cell states within and across patient specimens. a, CyCIF-1B: a second section from the same tissue block as CyCIF-1A (~20µm away in the Z-axis) imaged for the same

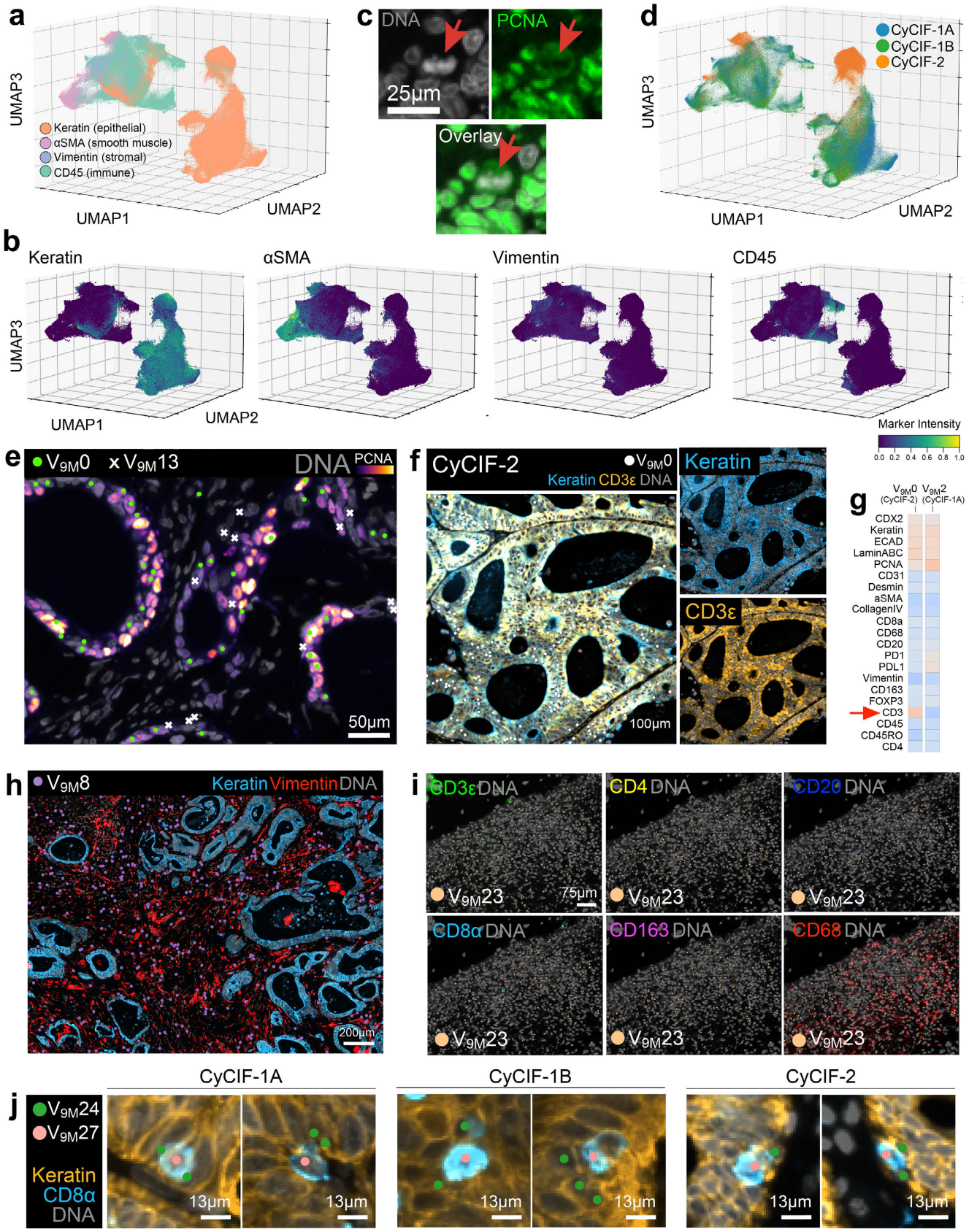
809 CyCIF antibodies listed in (**Supplementary Table 2**). **b**, CyCIF-2: a tissue section from a second CRC patient
810 imaged for the same CyCIF antibodies as CyCIF-1A and 1B. **c**, 3D UMAP embedding of 9x9µm VAE image
811 patch encodings from CyCIF-1A, CyCIF-1B, and CyCIF-2 colored by Leiden cluster. Clusters 29, 31, and 32
812 were identified as microscopy artifacts and are not labeled; see **Supplementary Video 4** for 360° plot rotation.
813 **d**, Location of V_{9M} clusters in CyCIF-1A (left) and 1B (right) revealing color-matched clusters in overlapping
814 locations in serial sections of this tissue; DNA channel (gray) shown for reference. **e**, Hierarchical clustering of
815 z-score-normalized median marker intensities of tissue-specific V_{9M} clusters showing that many share similar
816 protein expression profiles. Black vertical bars indicate two-way dendrogram merges between CyCIF-1A and
817 1B. Gray vertical bars indicate three-way merges between all three specimens. Data are filtered to remove
818 clusters associated with microscopy artifacts (V_{9M}29, 31, and 32) and clusters representing <0.1% of their
819 respective tissue. **f**, V_{9M} clusters identified in all three specimens as seen in their characteristic marker channels
820 with reference cells highlighted by white scatter points; DNA (gray) shown for reference. **g**, Grouped bar chart
821 showing the percentage of cells in each specimen accounted for by each V_{9M} cluster sorted in descending rank-
822 order according to cluster frequency in CyCIF-1A. Clusters identified as microscopy artifacts (V_{9M}29, 31, and
823 32) are not shown. **h**, Location of the four most abundant V_{9M} clusters in CyCIF-2; DNA channel (gray) shown
824 for reference.

Leiden cluster	Annotation	Histological region (HR)	CyCIF-2 enriched
0	PCNA^{high} glandular adenocarcinoma (AC)	HR5 (glandular AC)	**True**
1	Glandular AC	HR5 (glandular AC)	False
2	Glandular AC	HR5 (glandular AC)	False
3	Mucinous AC	HR6 (mucinous AC)	False
4	Endothelial cells	Diffuse	False
5	Solid AC	HR7 (solid AC)	False
6	Smooth muscle cells	HR2 (muscularis propria)	False
7	Glandular AC	HR5 (glandular AC)	False
8	Fibroblasts	αSMA^{+} stroma	**True**
9	Fibroblasts	CollagenIV^{+} stroma	False
10	Mixed lymphocytes (CD20^{-}/CD4^{-}/CD8^{-})	HR4 (subserosa)	False
11	Mixed B and T lymphocytes	HR* (lymphoid aggregates)	False
12	Macrophages (CD163^{+} CD4^{+})	HR9 (peritumor stroma)	False
13	PCNA$^{neg/low}$ Glandular AC	HR5 (glandular AC)	**True**
14	Crypt-forming epithelial cells	HR1 (normal mucosa)	False
15	Glandular AC	HR5 (glandular AC, predominently submucosal)	False
16	Macrophages (CD163^{+})	Diffuse	False
17	CD4^{+} T cells (Tcon/Treg)	Diffuse	False
18	Mixed AC and stromal cells (tumor-stromal boundry)	Tumor-stromal boundry	False
19	CD8^{+} T cells	Diffuse	False
20	Fibroblasts	Lamina propria	False
21	Macrophages (CD68^{+} PD-L1^{+})	Tumoral/peritumoral	False
22	Mixed lymphocytes (CD45^{2+})	HR4 (subserosa), HR* (lymphoid aggregates)	False
23	Mixed B and T lymphocytes	lymphoid aggregates	**True**
24	Mixed AC (CD8^{+} T cell adjacent)	HR5,6,7 (glandular, mucinous, and solid AC)	False
25	Perivascular smooth muscle cells	Diffuse	False
26	Macrophages (CD68^{+} CD163^{+})	HR2 (muscularis propria)	False
27	CD8^{+} T cells	HR5,6,7 (glandular, mucinous, and solid AC)	False
28	Endothelial cells (proximal to CollagenIV)	HR2, 3, 4, 9	False
29	Artifact	N/A	False
30	Mixed AC (dividing)	HR5, 6, 7 (glandular, mucinous, and solid AC)	False
31	Artifact	N/A	False
32	Artifact	N/A	False

Supplementary Table 5 | 9x9µm VAE image patch Leiden cluster annotations (multi-tissue analysis). Column labeled "CyCIF-2 enriched" indicates whether the given cluster is predominantly found in that specimen.

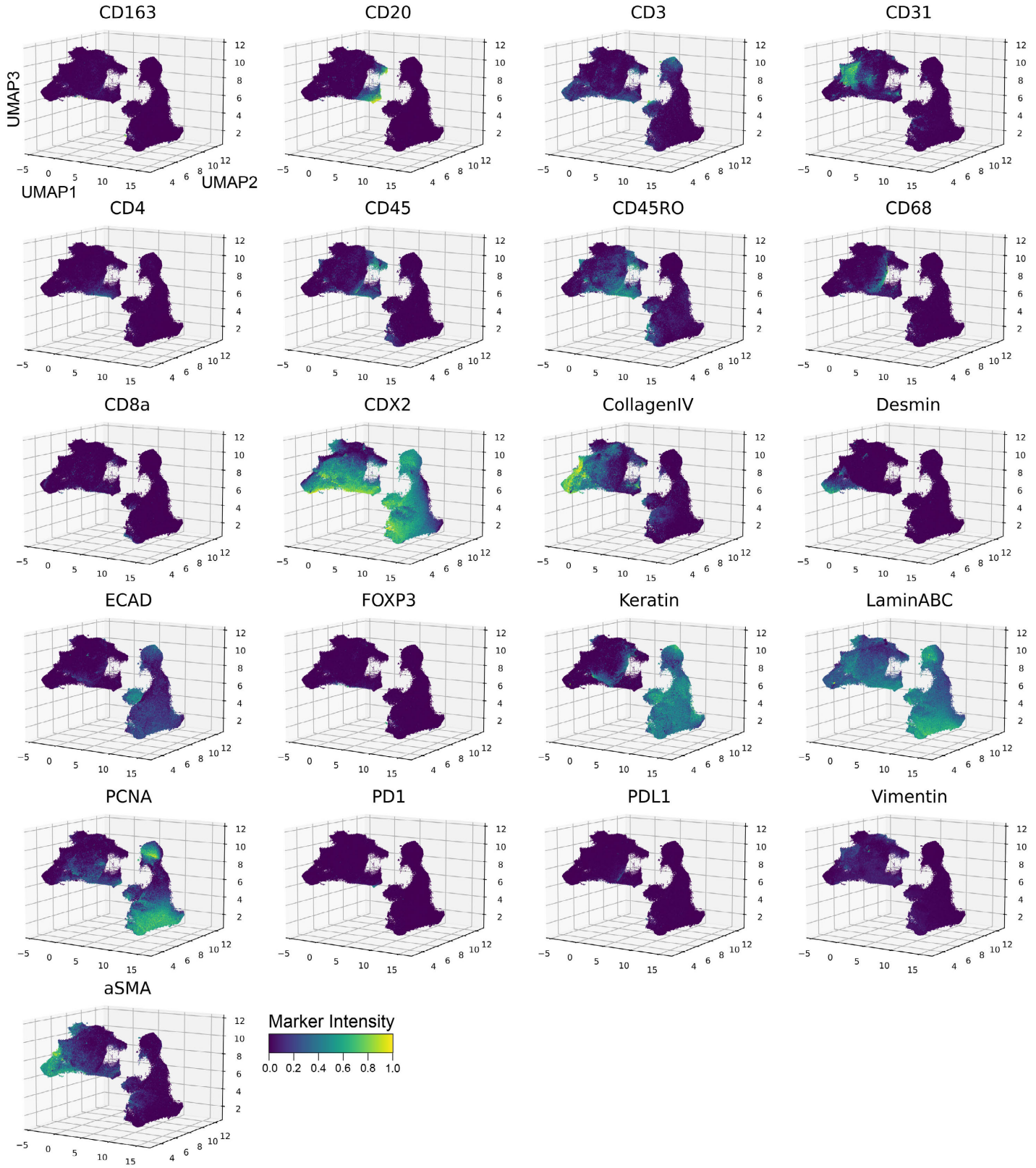


Supplementary Video 4 | 360° UMAP plot rotation showing 9x9µm VAE image patch encodings from CyCIF-1A, CyCIF-1B, and CyCIF-2 colored by Leiden cluster label.

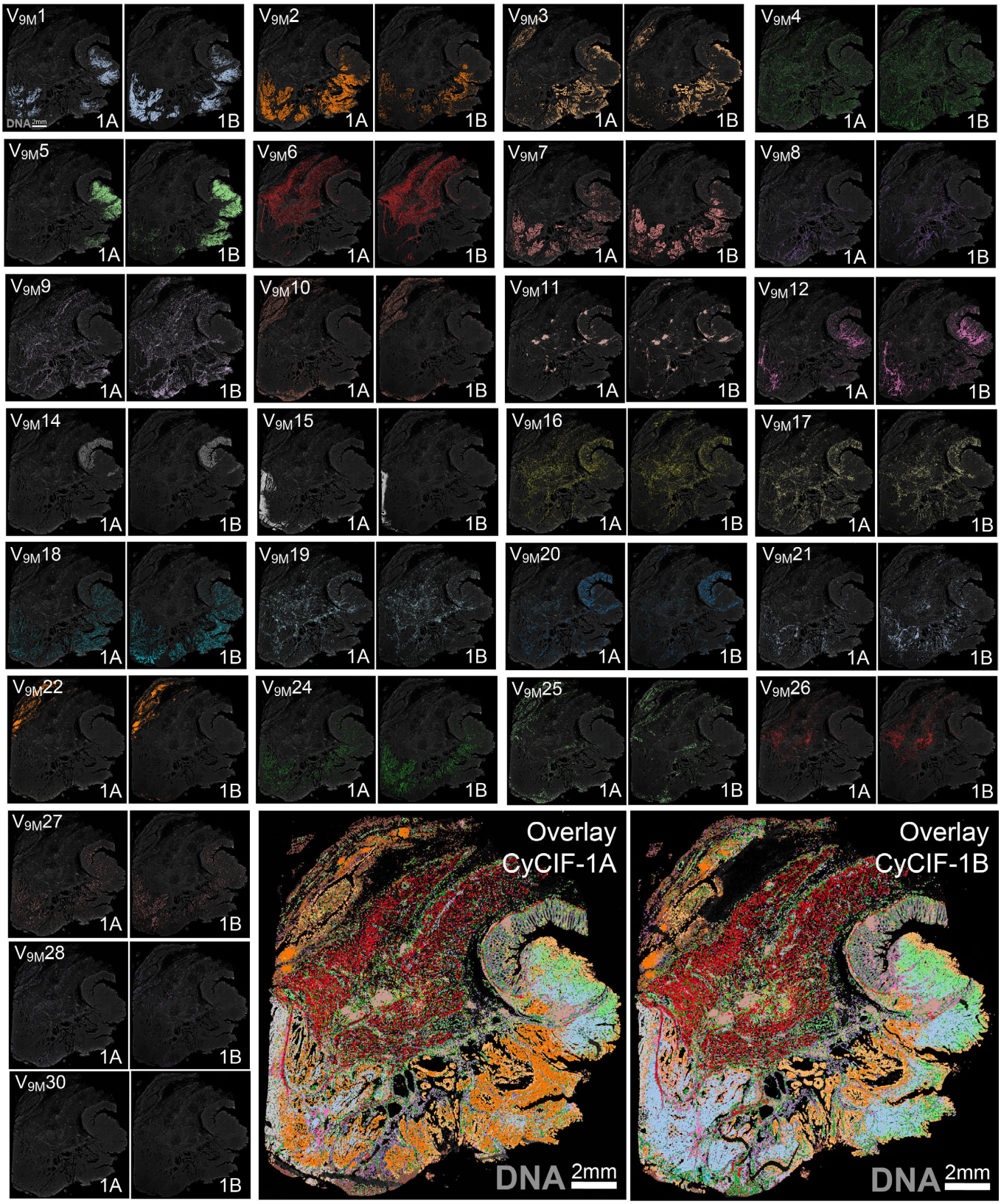


Extended Data Fig. 9 | VAE analysis identifies matched cell states within and across patient specimens. a, 3D UMAP embedding of 9x9μm VAE image patch encodings from CyCIF-1A, CyCIF-1B, and CyCIF-2

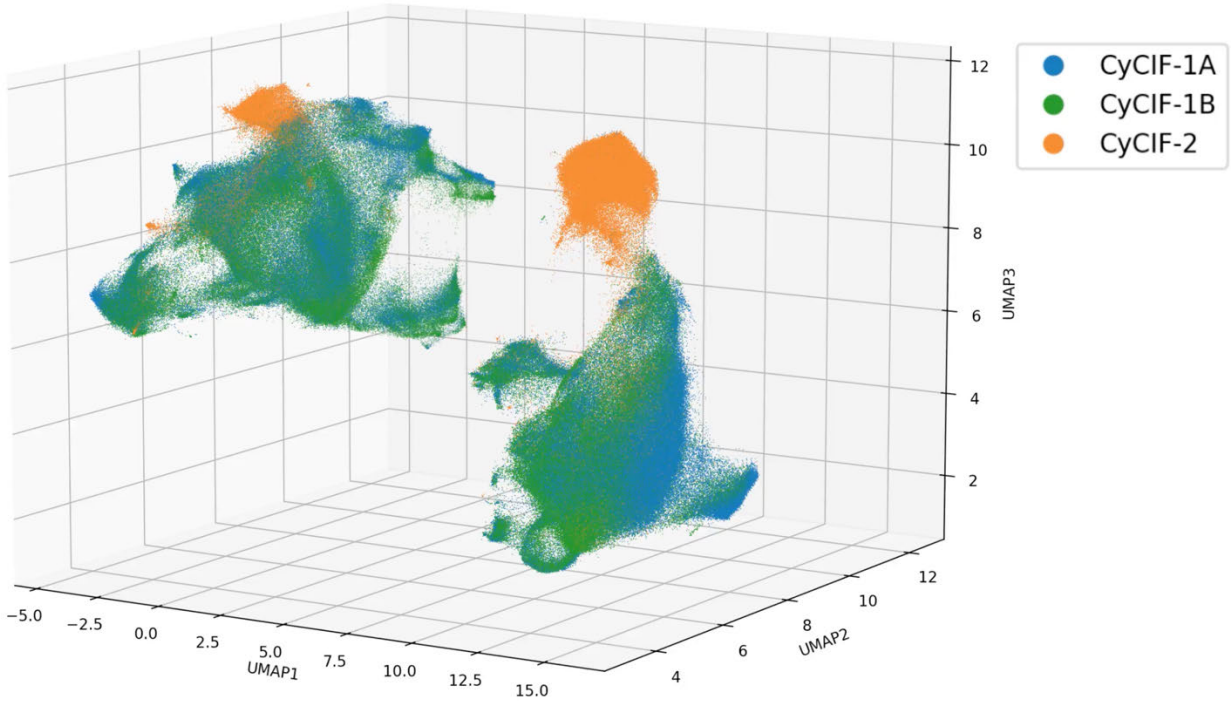
835 colored by major cell lineage markers: keratin (epithelial, orange), αSMA (smooth muscle, pink), vimentin
836 (stromal, blue), and CD45 (immune, green). **b**. 3D UMAP embedding of VAE image patch encodings from
837 CyCIF-1A, CyCIF-1B, and CyCIF-2 colored by pixel-level median marker intensity for major cell lineage
838 markers keratin, αSMA, vimentin, and CD45. See **Supplementary Fig. 11** for colormaps of all 21
839 immunomarkers. **c**, Field-of-view in CyCIF-1A showing a tumor cell (red arrow) undergoing cell division as
840 evidenced by its condensed chromatin seen in the DNA channel (gray) and in silhouette in the PCNA channel
841 (green). **d**, 3D UMAP embedding of 9x9μm VAE image patch encodings from CyCIF-1A, CyCIF-1B, and
842 CyCIF-2 colored by specimen; see **Supplementary Video 5** for 360° plot rotation. **e**, Field-of-view in CyCIF-2
843 showing that $V_{9M}0$ (green circles) and $V_{9M}13$ (white X symbols) both represent glandular AC cells but $V_{9M}0$
844 cells are brighter for the cell proliferation marker PCNA. **f**, Field-of-view in CyCIF-2 showing glandular AC
845 cells in $V_{9M}0$ (white scatter points). Note that these tumor cells are unexpectedly reactive to anti-CD3ε
846 antibodies. **g**, Channel z-score-normalized protein expression signatures for glandular AC cells in $V_{9M}0$ in
847 CyCIF-2 and analogous cells in $V_{9M}2$ in CyCIF-1A showing that tumor cells in CyCIF-2 are brighter for CD3ε
848 (red arrow) due to non-specific anti-CD3ε antibody binding. **h**, Field-of-view in CyCIF-2 showing fibroblasts
849 in $V_{9M}8$ (purple scatter points) and revealing a large ratio of stromal (vimentin, red)-to-tumor (keratin, blue)
850 tissue. **i**, Field-of-view in CyCIF-2 showing a region of tissue in which $CD45^+$ cells are negative for other,
851 more specific, immune lineage markers. **j**, Examples of $CD8^+$ T cells and adjacent $keratin^+$ tumor cells in
852 CyCIF-1A (left), CyCIF-1B (center), and CyCIF-2 (right) showing that a VAE correctly and reproducibly
853 identifies these neighboring cell types that co-cluster in segmentation-based analysis due to spatial crosstalk.



Supplementary Fig. 11 | Channel colormaps for 9x9µm VAE image patch encodings from multi-tissue analysis. UMAP embeddings of VAE image patch encodings from CyCIF-1A, CyCIF-1B, and CyCIF-2 colored by pixel-level median marker intensity for 21 immunomarkers.



Supplementary Fig. 12 | 9x9µm VAE image patch Leiden clusters exhibit concordance between serial sections CyCIF-1A and CyCIF-1B. Location of V_{9M} clusters in CyCIF-1A and CyCIF-1B revealing overlap in their histological regions; DNA (gray background) shown for reference.



Supplementary Video 5 | 360° UMAP plot rotation showing 9x9µm VAE image patch encodings from CyCIF-1A, CyCIF-1B, and CyCIF-2 colored by tissue specimen.

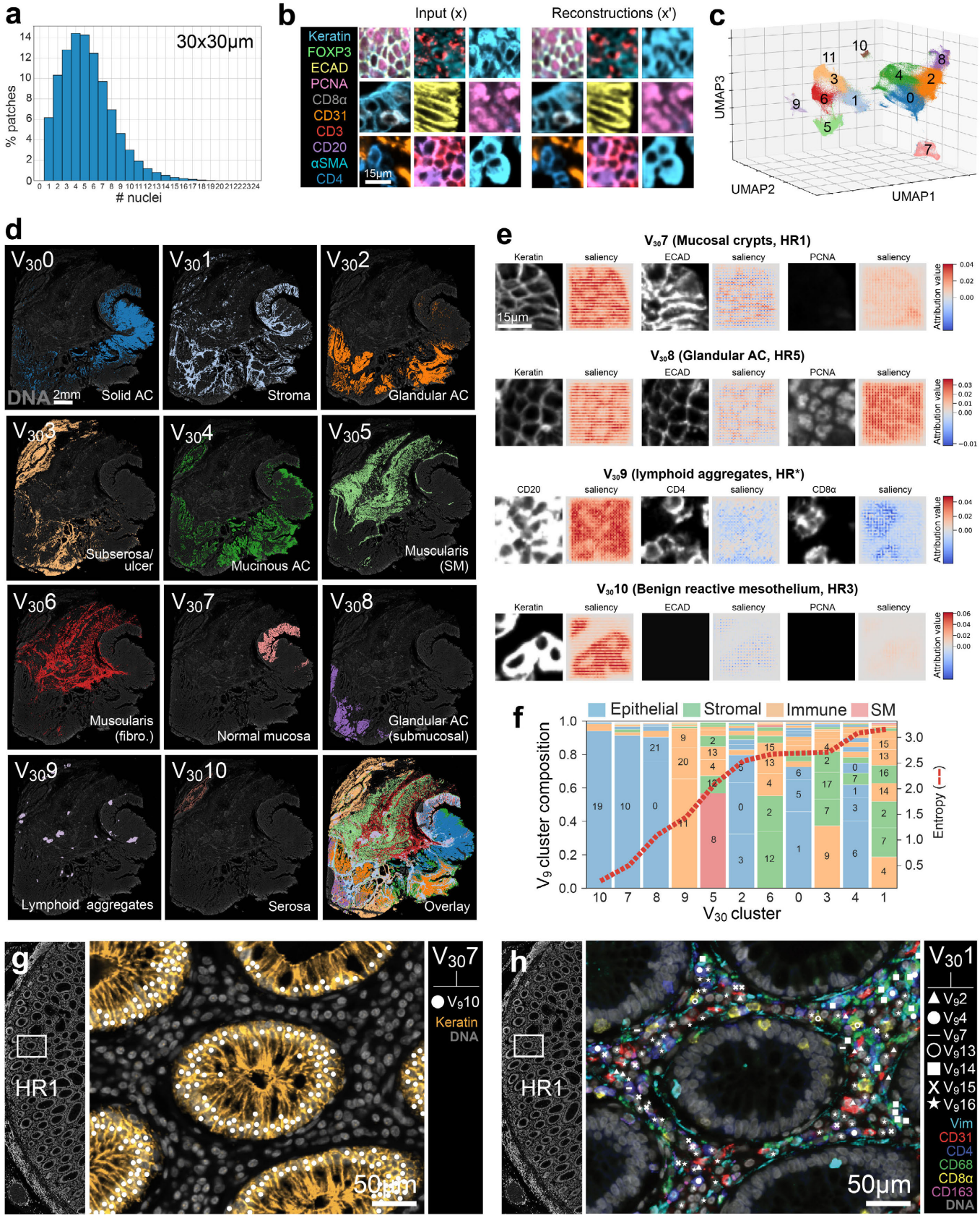
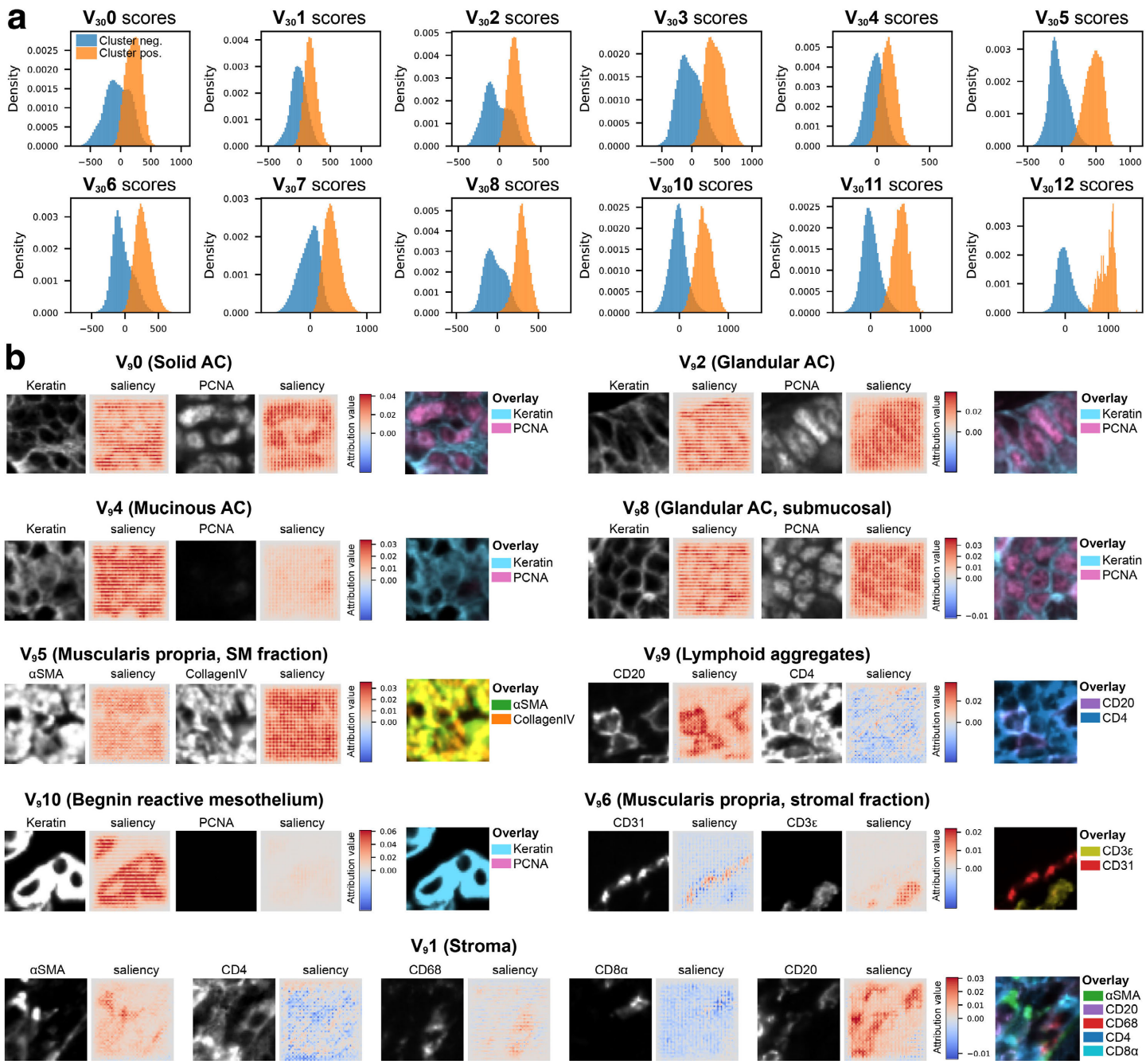


Fig. 6 | VAE classification of multi-cellular image patches identifies recurrent tissue motifs. a, Histogram showing the percentage of 30x30µm image patches containing different numbers of cell nuclei (median = 5, Q1

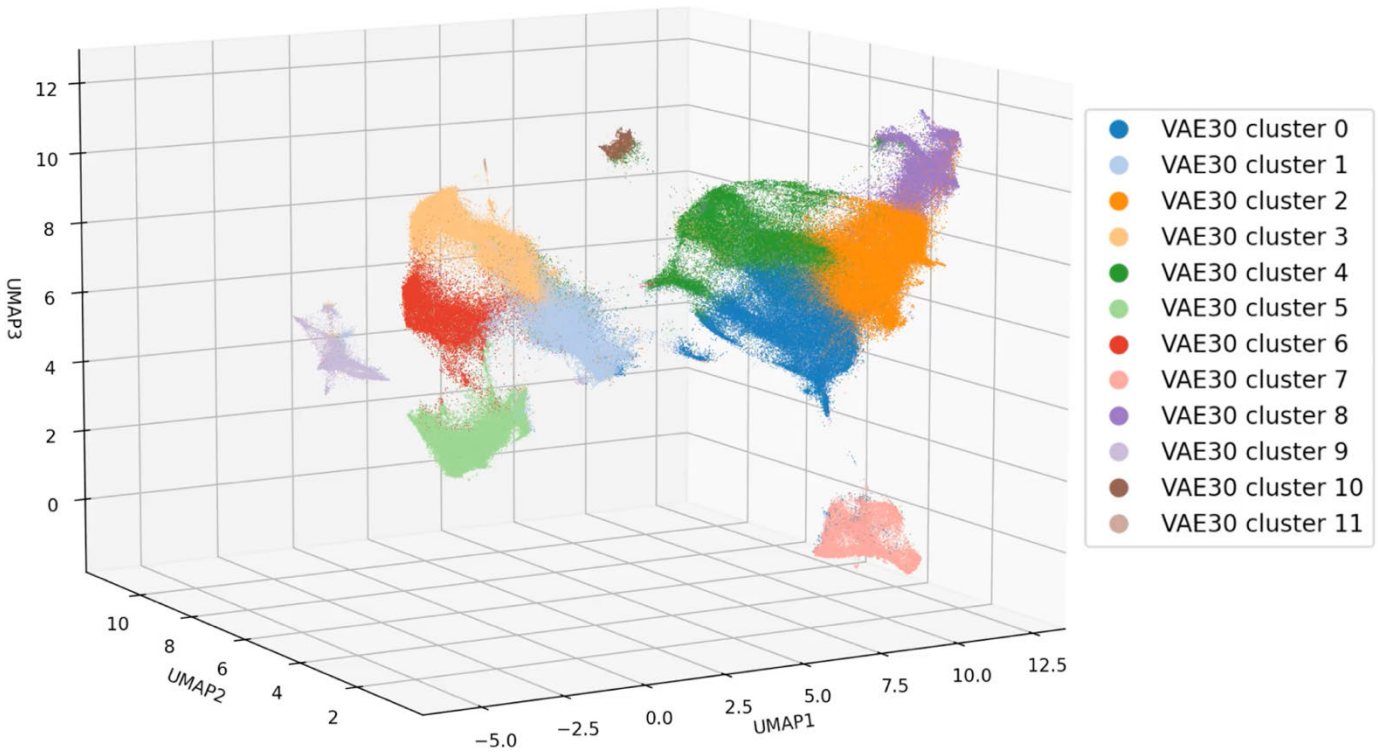
869 = 3, Q3 = 7, IQR = 4). **b**, Left: Example 30x30µm multi-channel image patches extracted from CyCIF-1A and
870 used as input into the VAE encoding network. Right: Learned reconstructions of input images shown at left. **c**,
871 3D UMAP embedding of 30x30µm VAE image patch encodings from CyCIF-1A colored by Leiden cluster;
872 see **Supplementary Video 6** for 360° plot rotation. **d**, Locations of V_{30} clusters in CyCIF-1A corresponding to
873 histological regions observed in the H&E image (**Fig. 1b,c**); DNA channel (gray) shown for reference. **e**,
874 Grayscale channels (left images) and corresponding concept saliency maps (right images) for four V_{30} clusters
875 showing characteristic channels. **f**, Stacked bar chart showing the composition of reference cells at the center of
876 V_{30} cluster image patches (x-axis) according to their V_9 cluster labels (left y-axis) and their cell type
877 heterogeneity as measured by information entropy (right y-axis, dashed red curve). **g**, Location of reference
878 cells in $V_{30}7$ (a low-entropy cluster) from CyCIF-1A composed of $V_9 10$ cells as seen in the keratin (orange)
879 channel, indicating that this V_{30} cluster represents healthy mucosal crypts consisting of normal epithelial cells
880 ($V_9 10$). **h**, Location of reference cells in $V_{30}1$ (a high-entropy cluster) from CyCIF-1A composed of multiple
881 V_9 cell types, including fibroblasts (vimentin, cyan), endothelial cells (CD31, red), helper T cells (CD4, blue),
882 M1-like macrophages (CD68, green), cytotoxic T cells (CD8α, yellow), and M2-like macrophages (CD163,
883 magenta) indicating that this cluster captures cellularly heterogeneous stromal tissue as a recurrent tissue motif.



Extended Data Fig. 10 | Concept saliency analysis for 30x30µm VAE clusters. a, Distributions of concept scores (Sc), defined as the dot product between the ith VAE image patch encoding (*Zi*) and the corresponding cluster concept vector (*Zc*) such that $Sc = Zi \cdot Zc$, for patches with (orange) and without (blue) a given cluster label. **b**, Channel intensities (left images) and corresponding concept saliency maps (right images) for nine different V_{30} clusters. Overlay color images at far right shown for reference; see **Supplementary Image Gallery 4** for fifteen examples of saliency maps across all 21 channels for each of the twelve V_{30} clusters.

Leiden cluster	% total cells	Histological region (HR)
0	16.8	HR7 (solid AC)
1	15.1	HR9 (peritumoral stroma)
2	14.7	HR5 (glandular AC)
3	10.7	HR4 (subserosa), HR8 (ulcerated submucosa)
4	10.5	HR6 (mucinous AC)
5	8.8	HR2 (muscularis propria, smooth muscle)
6	7.8	HR2 (muscularis propria, stroma)
7	5.1	HR1 (normal mucosa)
8	5.0	HR 5 (glandular AC, predominently submucosal)
9	4.5	HR* (lymphoid aggregates)
10	1.0	HR3 (serosa)
11	0.1	Microscopy artifact

Supplementary Table 6 | 30x30µm VAE image patch Leiden cluster annotations.



Supplementary Video 6 | 360° UMAP plot rotation showing 30x30µm VAE image patch encodings from CyCIF-1A colored by Leiden cluster labels.

896 **MATERIALS AND METHODS**

897 **Provenance and Characteristics of Tissue used in this Study**

898 The colorectal cancer (CRC) specimens referred to in this paper as "CyCIF-1A" (HTAN biospecimen ID:
899 HTA13_1_101), "CyCIF-1B" (HTAN biospecimen ID: HTA13_1_106), and "H&E" (HTAN biospecimen ID:
900 HTA13_1_100) consist of 5µm-thick, formalin-fixed paraffin-embedded (FFPE) tissue sections, derived from
901 the same tumor block (referred to as CRC1). These specimens were acquired through the Cooperative Human
902 Tissue Network (CHTN) and were first described in a prior study supported by the Human Tumor Atlas
903 Network (HTAN)[35]. The tumor originated from a right-sided, poorly differentiated stage IIIB colorectal
904 adenocarcinoma (pT3N1bM0) of the cecum, taken from a 69-year-old Caucasian male with microsatellite
905 instability (MSI-H) and a BRAFV600E (c.1799T>A) mutation. The CRC specimen referred to as "CyCIF-2"
906 (HTAN ID: HTA7_922_2) was obtained from Brigham and Women's Hospital (BWH) in compliance with all
907 relevant ethical regulations established by the BWH Institutional Review Board (IRB). This sample, also
908 analyzed as part of the same HTAN study[35], represents a stage IVA adenocarcinoma (pT4aN1bcM1a) from a
909 62-year-old Caucasian male with p53 and APC mutations. See (**Supplementary Table 1**) for an overview of
910 biospecimen metadata. The Lunaphore COMET and Akoya PhenoCycler images are specific to this study and
911 can be accessed upon request from the authors.

912

913 **Image Processing**

914 CRC tissue used in this study underwent 10 rounds of four-color multiplex immunofluorescence imaging by
915 CyCIF[56]. DNA plus 3 immunomarkers were imaged at 20x magnification (0.65µm/pixel) with 2x2 pixel
916 binning at each cycle, for a total of 40 channels (**Supplementary Table 2**). Nineteen (19) channels were
917 withheld from analysis in the current study, as they represented either cycle-specific Hoechst DNA stain
918 (channels 1, 5, 9, 13, 17, 21, 25, 29, 33, 37), secondary antibodies alone (channels 2, 3, 4, 6, 7, 8), or channels
919 determined to be of insufficient quality. Raw microscopy tiles (in .rcpnl format) were stitched, registered,
920 segmented, and quantified using the MCMICRO image processing pipeline[28]. Cells in the image were
921 quantified according to three different cell segmentation mask objects: nucleus, cytoplasm, and plasma
922 membrane. To ensure the most accurate segmentation-based signal quantification, the choice of mask selection
923 was determined on a per immunomarker basis according to prior knowledge of the immunolabeling pattern for
924 that marker. The raw single-cell spatial feature table data consisted of 1,218,752 total segmentation instances,
925 which were reduced to 962,343 after QC with CyLinter software[30] to remove cells affected by visual and
926 image-processing artifacts (e.g., tissue folds, slide debris, illumination aberrations, mis-segmentation, etc.).

927

928 **Detailed Experimental Protocols**

929 The following two protocols were used in the processing of tissue for this study: 1) FFPE Tissue Pre-treatment
930 before t-CyCIF on Leica Bond RX V.2 (dx.doi.org/10.17504/protocols.io.bji2kkge); 2) Tissue Cyclic
931 Immunofluorescence (t-CyCIF) V.2 (dx.doi.org/10.17504/protocols.io.bjiukkew)

932

933 **Cellcutter**

934 Cellcutter (https://github.com/labsyspharm/cellcutter) is a Python module developed in our lab for extracting
935 square image patches of arbitrary pixel size centered on individual cells from larger multi-channel TIFF/OME-
936 TIFF images by referring to their corresponding cell ID masks. Patches are stored in the efficient Zarr file
937 format. In this study, Cellcutter was instantiated as a subprocess in the MORPHAEUS data analysis pipeline
938 (https://github.com/labsyspharm/vae) by passing the following inputs to the tool's *cut_cells* command (see
939 Cellcutter's README.md file at https://github.com/labsyspharm/cellcutter#readme for further implementation
940 details):

941 --window-size 14 (for 9x9µm patches), 46 (for 30x30µm patches)
942 --cells-per-chunk 1
943 --cache-size 57711
944 <path to CRC OME-TIFF>
945 <path to CRC cell ID mask>
946 <path to CRC spatial feature table>
947 <path to Zarr output file>
948 --channels ["anti_CD3", "anti_CD45RO", "Keratin_570", "aSMA_660", "CD4_488", "CD45_PE",
949 "PD1_647", "CD20_488", "CD68_555", "CD8a_660", "CD163_488", "FOXP3_570",
950 "PDL1_647", "Ecad_488", "Vimentin_555", "CDX2_647", "LaminABC_488",
951 "Desmin_555", "CD31_647", "PCNA_488", "CollagenIV_647"]

952

953 **Image Patch Preprocessing**

954 Cellcutter image patches were centered on the nucleus of individual cells identified during image processing
955 with MCMICRO[28] and stored as Zarr arrays (https://zarr.readthedocs.io/en/stable/index.html) with dimensions
956 [channels, cells, x_dim, y_dim]. Image patches contained the following channels (in order): CD3ε, CD45RO,
957 keratin, αSMA, CD4, CD45, PD1, CD20, CD68, CD8α, CD163, FOXP3, PD-L1, ECAD, vimentin, CDX2,
958 laminABC, desmin, CD31, PCNA, collagenIV. Pixel intensity values for each channel were $\log_{10}$-transformed.
959 Pixels dimmer than the 0.17th and brighter than the 99.99th percentiles of signal intensity for full CRC image

channels were clipped and normalized to a range between 0-1 prior to model training to mitigate the impact of excessively dim and bright outlier pixels.

VAE Network Implementation

Model training was performed using Tensorflow 2.15.0. Batches of 64 image patches were converted from 16-bit unsigned integer to 32-bit floating point format prior to entering to the VAE network for both the 9x9µm and 30x30µm analyses. The VAE architecture was composed of four 2D convolutional layers. The filter size for the first layer was 32 pixels and the subsequent three were 64 pixels. All filters used the ReLU activation function with a kernel size of 3 pixels. Convolutional layers were followed by a 412-dimensional dense layer (in 9x9µm analysis) from which latent means and (log) variances were computed per image patch. A custom variational layer was used to calculate the loss between input image patches and their corresponding learned reconstructions defined as the mean squared error (MSE) between multi-channel input image patches and their learned reconstructions, plus the latent space structure loss (i.e., deviation from a multivariate Gaussian distribution) computed as the Kullback-Leibler (KL) divergence, the coefficient of which was -5.0×10^{-4}. The root mean squared propagation (RMSprop) algorithm was used for model optimization and the model learning rate was set to 1.0×10^{-6}. Training image patches were shuffled prior to each epoch of model training. Model weights were saved only when validation loss decreased relative to the preceding epoch. To make the VAE encoding and decoding networks end-to-end differentiable, the "reparameterization trick" was employed to enable the standard Gaussian (with a mean of zero and a standard deviation of one) to be the reference distribution in the KL divergence calculation.

Reporting Summary

Further information on research design is available in the Nature Portfolio Reporting Summary linked to this article.

Ethics and IRB Statement

HTAN tissue specimens and metadata used in this study complied with all relevant ethical regulations. The research described in this study is considered Non-Human Subjects Research.

Data Availability Statement

Imaging data analyzed in this paper are stored in a public Amazon S3 bucket (s3://lsp-public-data/baker-2025-vae/) and can be accessed by following instructions under "Downloading input data files" at the dedicated

GitHub repository for this paper (https://github.com/labsyspharm/vae-paper). Primary imaging data for all four tissues is also available through the HTAN Data Portal (https://data.humantumoratlas.org/explore) and can be accessed by searching for their respective HTAN biospecimen IDs: CyCIF-1A (HTA13_1_101), CyCIF-1B (HTA13_1_106), H&E (HTA13_1_100), CyCIF-2 (HTA7_922_2). With a free account, access to supplementary videos and image galleries is provided at the dedicated Sage Synapse Wiki page for this paper: https://www.synapse.org/Synapse:syn53216852/files/.

Code Availability Statement

Source code for the MORPHAEUS data analysis pipeline is available for academic re-use under the MIT open-source license agreement at Github (https://github.com/labsyspharm/vae) and is archived on Zenodo at XXX (to be indexed post-review). Python code used to generate the figures shown in this paper is available at (https://github.com/labsyspharm/vae-paper) and archived on Zenodo at XXX (to be indexed post-review).

Author Contribution Statements

G.J.B., A.S., and E.N. conceived and designed the study. H.P., and P.K.S. supervised and funded the work. G.J.B. developed and implemented the MORPHAEUS data analysis pipeline. E.N., Y.-A.C., Z.A., S.A.C., S.H., and C.Y. assisted with model development, data analysis, and visualization. G.J.B. and C.H. conceived of and developed the Cellcutter software. G.N.J and FY provided the Lunaphore COMET dataset. S.C. and S.S. provided histopathological annotations. G.J.B. performed data interpretation and generated the figures. G.J.B., E.N., S.S. and P.K.S. wrote the manuscript with input from all authors.

Acknowledgements

This work was supported by NCI grant U01-CA284207, the Harvard Ludwig Center (P.K.S., S.S.), an ASPIRE Award from The Mark Foundation for Cancer Research, and the David Liposarcoma Research Initiative, and was initiated as part of the computational toolbox for the Human Tissue Atlas Network (HTAN). Histopathology support was provided by P30-CA06516. S.S. is supported by the BWH President's Scholars Award. We thank J. Tefft for superb editorial support. We also thank Paul Lambert and Taja Lozar from the McArdle Laboratory for Cancer Research at the University of Wisconsin-Madison for kindly providing the Akoya PhenoCycler dataset. Results shown in this study are in part based upon data generated by HTAN (https://humantumoratlas.org/).

Competing Interests

P.K.S. is a cofounder and member of the Board of Directors of Glencoe Software and a member of the Scientific Advisory Board for RareCyte and Montai Health; he holds equity in Glencoe and RareCyte. P.K.S. is a consultant for Merck. PKS declares that none of these relationships have influenced the content of this manuscript. A.S in currently an employee of Flagship Pioneering. The other authors declare no competing interests.

1029 **REFERENCES**

1030 1. Method of the Year 2024: spatial proteomics. *Nat Methods* **21**, 2195–2196 (2024).

1031 2. Schapiro, D. *et al.* MITI minimum information guidelines for highly multiplexed tissue images. *Nat*
1032 *Methods* **19**, 262–267 (2022).

1033 3. Bai, Y. *et al.* Adjacent Cell Marker Lateral Spillover Compensation and Reinforcement for Multiplexed
1034 Images. *Front Immunol* **12**, 652631 (2021).

1035 4. Chevrier, S. *et al.* Compensation of Signal Spillover in Suspension and Imaging Mass Cytometry. *Cell Syst* **6**,
1036 612-620.e5 (2018).

1037 5. Baker, G. J. *et al.* Quality control for single-cell analysis of high-plex tissue profiles using CyLinter. *Nat*
1038 *Methods* **21**, 2248–2259 (2024).

1039 6. Yapp, C. *et al.* Highly multiplexed 3D profiling of cell states and immune niches in human tumors. *Nat*
1040 *Methods* 1–14 (2025) doi:10.1038/s41592-025-02824-x.

1041 7. Stringer, C., Wang, T., Michaelos, M. & Pachitariu, M. Cellpose: a generalist algorithm for cellular
1042 segmentation. *Nat Methods* **18**, 100–106 (2021).

1043 8. Pachitariu, M., Rariden, M. & Stringer, C. Cellpose-SAM: superhuman generalization for cellular
1044 segmentation. 2025.04.28.651001 Preprint at https://doi.org/10.1101/2025.04.28.651001 (2025).

1045 9. Amin, M. B. *et al.* The Eighth Edition AJCC Cancer Staging Manual: Continuing to build a bridge from a
1046 population-based to a more 'personalized' approach to cancer staging. *CA Cancer J Clin* **67**, 93–99 (2017).

1047 10. Cui, M. & Zhang, D. Y. Artificial intelligence and computational pathology. *Laboratory Investigation* **101**,
1048 412–422 (2021).

1049 11. Markovits, E. *et al. A Novel Deep Learning Pipeline for Cell Typing and Phenotypic Marker Quantification in*
1050 *Multiplex Imaging*. http://biorxiv.org/lookup/doi/10.1101/2022.11.09.515776 (2022)
1051 doi:10.1101/2022.11.09.515776.

1052 12. Amitay, Y. *et al.* CellSighter: a neural network to classify cells in highly multiplexed images. *Nat Commun*
1053 **14**, 4302 (2023).
1054 13. Nirmal, A. J., Yapp, C., Santagata, S. & Sorger, P. K. Cell Spotter (CSPOT): A machine-learning approach to
1055 automated cell spotting and quantification of highly multiplexed tissue images. Preprint at
1056 https://doi.org/10.1101/2023.11.15.567196 (2023).
1057 14. Zhu, B. *et al.* CellLENS enables cross-domain information fusion for enhanced cell population delineation
1058 in single-cell spatial omics data. *Nat Immunol* https://doi.org/10.1038/s41590-025-02163-1 (2025)
1059 doi:10.1038/s41590-025-02163-1.
1060 15. Nguyen, H., Vuong, H., Truong, T. & Pham, S. *IC-VAE: A Novel Deep Learning Framework for Interpreting*
1061 *Multiplexed Tissue Imaging Data*. http://biorxiv.org/lookup/doi/10.1101/2023.11.06.565771 (2023)
1062 doi:10.1101/2023.11.06.565771.
1063 16. Spitzer, H., Berry, S., Donoghoe, M., Pelkmans, L. & Theis, F. J. Learning consistent subcellular landmarks
1064 to quantify changes in multiplexed protein maps. *Nat Methods* **20**, 1058–1069 (2023).
1065 17. Tan, J. *et al.* Characterization of tumour heterogeneity through segmentation-free representation learning
1066 on multiplexed imaging data. *Nat Biomed Eng* **9**, 405–419 (2025).
1067 18. Wang, X. (Julie) *et al.* Generalized cell phenotyping for spatial proteomics with language-informed vision
1068 models. 2024.11.02.621624 Preprint at https://doi.org/10.1101/2024.11.02.621624 (2025).
1069 19. Liu, C. C. *et al.* Robust phenotyping of highly multiplexed tissue imaging data using pixel-level clustering.
1070 *Nat Commun* **14**, 4618 (2023).
1071 20. Kingma, D. P. & Welling, M. Auto-Encoding Variational Bayes. Preprint at http://arxiv.org/abs/1312.6114
1072 (2014).
1073 21. Kingma, D. P. & Welling, M. An Introduction to Variational Autoencoders. *FNT in Machine Learning* **12**,
1074 307–392 (2019).

1075 22. Brocki, L. & Chung, N. C. Concept Saliency Maps to Visualize Relevant Features in Deep Generative
1076 Models. Preprint at https://doi.org/10.48550/arXiv.1910.13140 (2019).
1077 23. Goltsev, Y. *et al.* Deep Profiling of Mouse Splenic Architecture with CODEX Multiplexed Imaging. *Cell* **174**,
1078 968-981.e15 (2018).
1079 24. Najem, H. *et al.* Protocol to quantify immune cell distribution from the vasculature to the glioma
1080 microenvironment on sequential immunofluorescence multiplex images. *STAR Protoc* **5**, 103079 (2024).
1081 25. Lin, J.-R. *et al.* Highly multiplexed immunofluorescence imaging of human tissues and tumors using t-CyCIF
1082 and conventional optical microscopes. *eLife* **7**, e31657 (2018).
1083 26. de Bruijn, I. *et al.* Sharing data from the Human Tumor Atlas Network through standards, infrastructure
1084 and community engagement. *Nat Methods* **22**, 664–671 (2025).
1085 27. Rozenblatt-Rosen, O. *et al.* The Human Tumor Atlas Network: Charting Tumor Transitions across Space
1086 and Time at Single-Cell Resolution. *Cell* **181**, 236–249 (2020).
1087 28. Schapiro, D. *et al.* MCMICRO: a scalable, modular image-processing pipeline for multiplexed tissue
1088 imaging. *Nat Methods* **19**, 311–315 (2022).
1089 29. Lin, J.-R. *et al.* Multiplexed 3D atlas of state transitions and immune interaction in colorectal cancer. *Cell*
1090 **186**, 363-381.e19 (2023).
1091 30. Baker, G. J. *et al.* Quality control for single-cell analysis of high-plex tissue profiles using CyLinter. *Nat*
1092 *Methods* https://doi.org/10.1038/s41592-024-02328-0 (2024) doi:10.1038/s41592-024-02328-0.
1093 31. Traag, V. A., Waltman, L. & van Eck, N. J. From Louvain to Leiden: guaranteeing well-connected
1094 communities. *Sci Rep* **9**, 5233 (2019).
1095 32. Horn, J. L. A RATIONALE AND TEST FOR THE NUMBER OF FACTORS IN FACTOR ANALYSIS. *Psychometrika*
1096 **30**, 179–185 (1965).

1097 33. McInnes, L., Healy, J. & Melville, J. UMAP: Uniform Manifold Approximation and Projection for Dimension
1098 Reduction. *arXiv:1802.03426 [cs, stat]* http://arxiv.org/abs/1802.03426 (2020).
1099 34. Hug, C. B. cellcutter. (2021).
1100 35. Lin, J.-R. *et al.* Multiplexed 3D atlas of state transitions and immune interaction in colorectal cancer. *Cell*
1101 **186**, 363-381.e19 (2023).
1102 36. Kim, B. *et al.* Why are Saliency Maps Noisy? Cause of and Solution to Noisy Saliency Maps. Preprint at
1103 https://doi.org/10.48550/arXiv.1902.04893 (2019).
1104 37. Shaw, T. N. *et al.* Tissue-resident macrophages in the intestine are long lived and defined by Tim-4 and
1105 CD4 expression. *J Exp Med* **215**, 1507–1518 (2018).
1106 38. Lin, J.-R. *et al.* High-plex immunofluorescence imaging and traditional histology of the same tissue section
1107 for discovering image-based biomarkers. *Nat Cancer* **4**, 1036–1052 (2023).
1108 39. Salama, P. *et al.* Tumor-infiltrating FOXP3+ T regulatory cells show strong prognostic significance in
1109 colorectal cancer. *J Clin Oncol* **27**, 186–192 (2009).
1110 40. Geographic Data Science with Python (Chapman & Hall/CRC Texts in Statistical Science): Rey, Sergio,
1111 Arribas-Bel, Dani, Wolf, Levi John: 9780367263119: Amazon.com: Books.
1112 https://www.amazon.com/Geographic-Science-Python-Chapman-Statistical/dp/0367263114.
1113 41. Yapp, C. *et al.* Highly Multiplexed 3D Profiling of Cell States and Immune Niches in Human Tumours.
1114 2023.11.10.566670 Preprint at https://doi.org/10.1101/2023.11.10.566670 (2025).
1115 42. Muhlich, J. L. *et al.* Stitching and registering highly multiplexed whole-slide images of tissues and tumors
1116 using ASHLAR. *Bioinformatics* **38**, 4613–4621 (2022).
1117 43. Besson, S. *et al.* Bringing Open Data to Whole Slide Imaging. *Digit Pathol (2019)* **2019**, 3–10 (2019).
1118 44. Ronneberger, O., Fischer, P. & Brox, T. U-Net: Convolutional Networks for Biomedical Image
1119 Segmentation. in *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2015* (eds

Navab, N., Hornegger, J., Wells, W. M. & Frangi, A. F.) vol. 9351 234–241 (Springer International Publishing, Cham, 2015).

45. Long, J., Shelhamer, E. & Darrell, T. Fully convolutional networks for semantic segmentation. in *2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)* 3431–3440 (IEEE, Boston, MA, USA, 2015). doi:10.1109/CVPR.2015.7298965.
46. Yapp, C. *et al.* UnMICST: Deep learning with real augmentation for robust segmentation of highly multiplexed images of human tissues. *Commun Biol* **5**, 1263 (2022).
47. Greenwald, N. F. *et al.* Whole-cell segmentation of tissue images with human-level performance using large-scale data annotation and deep learning. *Nat Biotechnol* **40**, 555–565 (2022).
48. Zhang, W. *et al.* Identification of cell types in multiplexed in situ images by combining protein expression and spatial information using CELESTA. *Nat Methods* **19**, 759–769 (2022).
49. Blondel, V. D., Guillaume, J.-L., Lambiotte, R. & Lefebvre, E. Fast unfolding of communities in large networks. *J. Stat. Mech.* **2008**, P10008 (2008).
50. Ikotun, A. M., Ezugwu, A. E., Abualigah, L., Abuhaija, B. & Heming, J. K-means clustering algorithms: A comprehensive review, variants analysis, and advances in the era of big data. *Information Sciences* **622**, 178–210 (2023).
51. Kohonen, T. Self-organized formation of topologically correct feature maps. *Biol. Cybern.* **43**, 59–69 (1982).
52. McInnes, L., Healy, J. & Astels, S. hdbscan: Hierarchical density based clustering. *JOSS* **2**, 205 (2017).
53. van der Maaten et al. Visualizing high-dimensional data using t-SNE. *J Mach Learn Res* **9**, 2579–2605 (2008).
54. Wang, S., Sontag, E. D. & Lauffenburger, D. A. What cannot be seen correctly in 2D visualizations of single-cell ‘omics data? *Cell Systems* **14**, 723–731 (2023).

1143 55. Pachitariu, M., Rariden, M. & Stringer, C. Cellpose-SAM: superhuman generalization for cellular
1144 segmentation. Preprint at https://doi.org/10.1101/2025.04.28.651001 (2025).
1145 56. Lin, J.-R. *et al.* Highly multiplexed immunofluorescence imaging of human tissues and tumors using t-CyCIF
1146 and conventional optical microscopes. *Elife* **7**, (2018).