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], Artem Sokolov[1], Hanspeter Pfister[4], Sandro Santagata[1,2,3,5], 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
[&]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 (AI), 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: gbak7696@gmail.com

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

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 the problem of signal spillover. The result is a more accurate and nuanced characterization of cell types and states than segmentation-based analysis alone.

INTRODUCTION

High-plex tissue imaging (spatial proteomics) has emerged as a method for measuring the abundance and spatial distributions of tens to hundreds of biomolecules in intact tissue, most commonly using fluorescently or metal-tagged antibodies[1]. Single-cell data are obtained from these images by computing segmentation masks centered on individual nuclei and quantifying integrated per antibody staining intensities within whole-cell or subcellular segmentation boundaries. This generates cell feature tables analogous to count tables in transcriptional profiling.[2] These tables can be used to identify cell types and states and combined with other forms of single-cell data to develop molecular and mechanistic descriptions of tissue structures and biological processes. However, close packing and overlap of cells in tissues makes segmentation imprecise.[3,4] Despite many years of effort, both over-segmentation (subdividing cells into multiple objects) and under-segmentation (combining different cells into a single object) are common and confound cell type identification and other analyses.[3,5] Moreover, current approaches to segmentation provide little information about cell morphology or intracellular protein distributions.

An alternative approach to image analysis widely used in computational pathology involves deep learning on pixel-level data, most commonly, images of tissue stained with hematoxylin and eosin (H&E).[6] Computational pathology builds on advances in the use of neural networks and artificial intelligence (AI) for facial recognition[7] and other applications, to automatically classify images by tissue type, disease state, or treatment response.[8] Pixel-level learning has also been applied to multiplexed tissue images using convolutional and transformer-based neural networks, CNN-based classifiers[9–12], generative models[13,14], vision transformers[15,16], and unsupervised clustering of image features[17]. Methods based on supervised learning can perform better with complex image data than those based on unsupervised learning, but the former are limited to discovering pre-specified features or "classes", and training can be labor intensive. Regardless of approach, studies to date do not directly compare pixel-level learning with segmentation nor do they explore how machine learning compares to human annotation.

In this paper, we compare and then combine traditional segmentation-based single-cell analysis with pixel-level machine learning (ML) using light-weight generative models based on variational autoencoders (VAEs).[18,19] We show that VAEs can be trained to analyze high-plex images in a fully unsupervised manner using single-cell image patches centered on segmented nuclei. By changing the size of these patches to include multiple cells, we show that it is possible to learn histological features at different length scales. We find that clustering of VAE latent space manifolds makes it possible to enumerate both single-cell phenotypes and local cellular communities in a manner that overcomes "spatial crosstalk" (lateral spillover)[3] arising from errors in segmentation. Using concept saliency analysis[20]—a method in interpretable AI—we show that a combination

of VAE-based classification and traditional segmentation is able to draw on information related to marker intensity, cell morphology, and local neighborhoods in a manner analogous to human experts, providing superior insight into tissue composition and organization than segmentation-based analysis alone.

RESULTS

Spatial crosstalk limits accuracy of segmentation-based cell phenotyping

Algorithms in MCMICRO[21] were used to segment cells in a 22-plex whole-slide cyclic immunofluorescence (CyCIF)[22] image of human colorectal cancer (CRC) previously released by the Human Tumor Atlas Network and referred to here as CyCIF-1A (**Fig. 1a** and **Supplementary Table 1**; HTAN ID: HTA13_1_101). This image is one of a series of interleaved CyCIF and H&E images previously used for 2D and 3D analysis of 16 CRC cases.[23] 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 (**Fig. 1b,c**). Regions included normal mucosa (the crypt-containing epithelial layer of the colon), smooth muscle and stromal tissue, and regions of adenocarcinoma (AC) exhibiting glandular, mucinous, or solid morphologies. Per-cell mean fluorescence intensities were extracted from CyCIF-1A for three sets of object masks corresponding to the cell nucleus, cytoplasm, and plasma membrane. The integrated intensity of each marker was then determined using the mask appropriate for that protein (**Supplementary Table 2**). For example, staining intensities for CD4 and CD8, which localize to the plasma membranes of lymphocytes, were computed using plasma membrane masks, while those for PCNA, a cell proliferation marker localized to the nucleus, were extracted using nuclei masks. The resulting single-cell feature table underwent quality control (QC) with CyLinter[24] to remove visual and image processing artifacts (affecting ~21% of total cells; **Extended Data Fig. 1a**). For more information on segmentation-based tissue analysis see **Supplementary Note 1** and **Supplementary Fig. 1**.

Clustering of post-QC segmented data using Leiden community detection[25] 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**). Data were visualized in a 3D UMAP[26] embedding since Horn's parallel analysis[27] showed that the first three principal components (PCs) captured the majority (68%) of dataset variance (**Fig. 1d**, **Extended Data Fig. 1b** and **Supplementary Video 1**; supplementary videos and image galleries are available on Synapse.org with a free account: https://www.synapse.org/Synapse:syn53216852/files/). Hierarchical clustering by mean z-score-normalized signal intensities revealed the presence of 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, differing primarily in PCNA expression levels (**Extended Data Fig. 1f**). However, these clusters did not correspond directly to the glandular, mucinous, and solid AC histologies observed in the H&E image (**Fig. 1b,c and f**). Normal epithelial cells comprised the tumor-adjacent colonic mucosa (S14, HR1) and benign reactive mesothelium in the serosa (S17, R3), the latter likely arising

134 from tumor-associated inflammation. Smooth muscle occupied the muscularis propria (S4, HR2), while
135 fibroblasts (S6) were identified in the subserosa (HR4), peritumoral stroma (HR9), and submucosa (HR10).
136 Endothelial cells (S7) were present throughout the tissue, as were immune cells, including populations of
137 macrophages (S3), $CD8^+$ T cells (S5), and both conventional (S13) and regulatory (S15) $CD4^+$ T cells
138 (**Extended Data Fig. 2** and **Supplementary Fig. 3**).

139 Despite the biological plausibility of these cell phenotypes, closer examination revealed multiple
140 clusters with unlikely immunomarker combinations. For example, cells in S12 appeared to express markers of
141 proliferating epithelial cells (keratin, ECAD, PCNA) and $CD8^+$ T cells (CD45, CD3ε, CD8α, CD45RO; **Fig.**
142 **1e**). This was not a consequence of suboptimal clustering, since manual gating on keratin and CD8α channels
143 showed that >80% of cells in this cluster were positive for both markers (**Fig. 1g,h**). Inspection of the image
144 showed that S12 comprised $CD8^+$ T cells adjacent to $keratin^+$ tumor cells (**Fig. 1i**). Since neighboring cells can
145 have overlapping membrane extensions (e.g. lamellipodia) or overlapping cell bodies along the Z-axis[28],
146 segmentation masks often include portions of adjacent cells of different types, resulting in unavoidable error
147 referred to as "spatial crosstalk" (or lateral spillover[3]—a phenomenon analogous to multiplets in scRNA-seq[29]).
148 We conclude that S12 cells represent $CD8^+$ T cells affected by spatial crosstalk from adjacent tumor cells.

149 Clusters S11 and S16 were also affected by spatial crosstalk. Both clusters represented tightly packed B
150 and T cells comprising lymphoid aggregates (HR*), with S16 cells found exclusively in the serosa (HR3) and
151 subserosa (HR4). Non-tumor infiltrating $CD8^+$ T cells captured by S5 were also impacted by spatial crosstalk
152 and were distributed across multiple subclusters in UMAP space. A broader investigation into the relationship
153 between spatial crosstalk and reduced dimensionality representations of high-plex imaging data showed that
154 spatial crosstalk: (i) is pervasive in segmented high-plex image data, affecting up to 30% of the ~1M cells in
155 CyCIF-1A; (ii) represents a significant barrier to accurate cell phenotyping; and (iii) promotes formation of
156 discrete clusters in low dimensionality embeddings, likely due to the presence of distinctive (yet artifactual)
157 marker combinations (**Supplementary Note 2** and **Supplementary Figs. 4 & 5**).
158

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

160 To analyze cell types in high-plex tissue images at the pixel level, we developed a method for training
161 VAE models on square image patches extracted from high-plex tissue images. This approach used Cellcutter[30]
162 to generate 9x9μm (14x14pixel) multi-channel patches centered on individual nuclei and store them in a
163 memory-efficient Zarr file (**Fig. 2a,b** and **Methods**). After subsampling, the final dataset for CyCIF-1A
164 consisted of 479,909 patches, 86% with one and 13% with two nuclei (**Fig. 2c**). Image patch pre-processing
165 involved: (i) log transformation, (ii) clipping to remove bright and dim outlier pixels, (iii) per-channel

normalization (0-1) based on fluorescence intensities from the whole-slide image, and (iv) rotational data augmentation to reduce sensitivity to cell orientation (**Extended Data Fig. 3a,b**). Patches were divided into training, validation, and test sets at a ratio of 80:10:10 and used to train a standard VAE network consisting of four convolutional encoding layers followed by a dense layer (**Fig. 2d**, **Extended Data Fig. 3c**, and **Methods**). To focus the attention of the network on the cell at the center of the patch, each channel was convolved with a 2D Gaussian kernel with a standard deviation of 7 pixels. This vignetting technique reduced the intensity of features towards patch borders, causing the network to perceive them with less acuity (42% of maximum intensity at patch corners; **Fig. 2e**).

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

Leiden clustering of VAE image patch encodings generated 24 clusters ($V_9$0-23, with "V_9" denoting a VAE trained on 9x9µm patches; **Fig. 2h**, **Supplementary Table 4**, **Supplementary Video 2**, and **Supplementary Image Gallery 1**). As expected, V_9 clusters corresponded to the same primary lineages identified in segmentation-based analysis: keratin$^+$ epithelium, αSMA$^+$ smooth muscle, vimentin$^+$ stroma, and CD45$^+$ immune cells (**Extended Data Fig. 4a,b** and **Supplementary Fig. 6**). Hierarchical clustering of V_9 clusters (according to median, z-score normalized, pixel-level marker intensities) generated 10 meta-clusters (A-J; **Fig. 2i**). To evaluate the agreement between VAE and segmentation-based cluster pairs, we used the Jaccard index (JI). Six clusters (29% of total cells) exhibited substantial overlap (JI ≥ 0.44). These comprised Tregs ($V_9$4/S15), smooth muscle cells ($V_9$8/S4), normal crypt-forming epithelial cells ($V_9$10/S14), lymphoid aggregates ($V_9$11/S11; primarily tertiary lymphoid structures, TLSs), CD8$^+$ T cells ($V_9$13/S5), and mesothelial cells ($V_9$19/S17; **Extended Data Fig. 4c,d**, **Extended Data Fig. 5** and **Supplementary Fig. 7**). Since segmentation-based clustering is based on marker intensity, we surmised that formation of these VAE clusters was also driven primarily by differences in marker expression.

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

VAE-based learning identified seven tumor cell clusters (meta-cluster H; **Fig. 2i**), compared to five with segmentation. While marker expression profiles of tumor-associated VAE clusters were highly similar

(**Fig. 3a**), some differences were observed; for example, $V_{9}21$ was higher in PCNA and lower in CDX2 compared to $V_{9}0$. However, whereas segmentation-based clusters were intermixed in location when mapped back onto CyCIF-1A, VAE-based clusters corresponded to distinct tumor regions observed in the H&E image (**Fig. 3b,c**, **Extended Data Fig. 6a** and **Fig. 1b,c**). To quantify this correspondence, we calculated normalized mean posterior probabilities for VAE clusters using four previously developed kNN classifiers trained to predict normal, glandular, mucinous, or solid AC histologies based on protein expression[23]. Cells in clusters $V_{9}0$, 3, 5, 21, and 22 showed high mean probabilities of glandular AC (≥71%), while those in clusters $V_{9}6$ and $V_{9}1$ had the highest probabilities of mucinous (54%) and solid AC (47%), respectively. These results suggested that the VAE learned to distinguish cells with different morphologies learned by the network, and that mucinous and solid AC subtypes exhibit greater morphological variability than glandular AC (**Fig. 3d** and **Extended Data Fig. 6b**). To quantify the degree of uncertainty in AC subtype assignment, we calculated the information entropy of VAE clusters relative to their kNN class probabilities (**Fig. 3d**, red dotted line). Coloring clusters by entropy scores and mapping them back onto CyCIF-1A, we observed a distinctive pattern, with lower entropy glandular AC appearing at the left of the image and higher entropy mucinous and solid AC at the right (**Fig. 3e**). This result revealed a spatial gradient of tumor dedifferentiation that was confirmed examination of the primary imaging data (**Fig. 3f,g**).

To determine the contribution of morphology in distinguishing tumor-associated VAE clusters, we isolated and binarized the keratin channel of tumor cell image patches from meta-cluster H to remove intensity variation, and then trained a new VAE model. We found that cells with mucinous ($V_{9}6$) and glandular ($V_{9}0$, 3, and 5) morphologies occupied distinct regions in feature space and exhibited differences in keratin organization (**Fig. 3h,i**), although this was not true for solid AC ($V_{9}1$). Thus, the shape of the keratin signal alone was sufficient to distinguish glandular from mucinous cells but was unreliable for identifying solid AC. This finding, combined with the marker intensity map for V_{9} clusters (**Fig. 3a**, right panel), strongly suggest that the VAE network simultaneously captures intensity and morphology information when distinguishing tumor cell subtypes.

Identifying factors driving cluster membership using concept saliency analysis

$V_{9}22$ cancer cells were found throughout the tumor and were consistently surrounded by $CD8^{+}$ T cells from cluster $V_{9}18$ (**Fig. 3j**). The Jaccard index plot (**Extended Data Fig. 4c**) showed that these two VAE clusters were combined in segmentation cluster S12. Thus, VAE-based learning had correctly resolved $CD8^{+}$ T cells and 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

230 variance between $V_9$18 and $V_9$22 patches was significantly greater than variance within $V_9$18 patches for
231 channels CD3ε, CD45, CD8α, and PCNA. Visualizing average MSE representations for each cluster showed
232 that CD3ε, CD45, and CD8α signals were concentrated at the center of $V_9$18 patches but the periphery of $V_9$22
233 patches (**Fig. 3l**), while PCNA signals exhibited the opposite pattern. This result was confirmed by inspection
234 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
235 were adjacent to different stromal and immune cell types in the lamina propria of normal epithelium (HR1),
236 muscularis propria (HR2), and peritumoral stroma (HR9; **Fig. 3n**). We conclude that a VAE can distinguish
237 adjacent cell types affected by spatial crosstalk in segmentation by learning patterns of protein expression, and
238 can further differentiate similar cell types (e.g., $CD8^+$ T cells) based on tissue context.

239 To study how patch features influenced cluster membership, we performed concept saliency
240 analysis[20]—a method in interpretable AI that seeks to explain the features driving different ML classes.
241 Archetypal image patch encodings or "concept vectors" (denoted, *Zc*) were first selected for each cluster based
242 on maximum cosine similarity to the average encoding for all patches bearing the same cluster label. Individual
243 patches were then scored for similarity to cluster archetypes by computing the dot product between the VAE
244 encoding for a particular image patch (*Zi*) and the previously identified concept vector ($Sc = Zi \cdot Zc$). This
245 showed that image patches with and without a specific cluster label formed discrete concept score distributions
246 (**Extended Data Fig. 7a**), showing that the VAE identified groups of image patches with unique features. We
247 then generated concept saliency maps using the rectified gradient method[31], which uses backpropagation to
248 calculate pixel-wise contributions to a given concept score while minimizing noise (**Fig. 3o**, **Extended Data**
249 **Fig. 7b** and **Supplementary Image Gallery 2**). In these maps, pixels with high attribution values corresponded
250 to signals consistently identified with specific clusters. For example, in the case of $CD68^+$ $CD163^-$ PD-$L1^+$
251 macrophages in cluster $V_9$14, the CD68 and PD-L1 channels showed positive attribution values in regions of
252 the patch with high fluorescence signals, whereas biologically unrelated signals from CD163 and CD8α had
253 negative values (**Fig. 3o**). These results provide further evidence that intensity signals and spatial distributions
254 learned in an unsupervised manner by a VAE are interpretable and biologically meaningful.
255

256 **Patch-based learning achieves greater precision relative to image segmentation**

257 The Jaccard index plot showed that segmentation-based clusters were often subdivided by the VAE into
258 multiple subclusters, suggesting higher precision in distinguishing cell types (**Extended Data Fig. 4c**). For
259 example, within segmentation cluster S3 (annotated as macrophages) the VAE identified three phenotypically
260 and spatially distinct cell types: (i) $CD163^+$ $CD68^+$ tumor-reactive macrophages in the peritumoral stroma
261 ($V_9$7, HR9), (ii) $CD163^+$ tissue-resident macrophages in the muscularis propria ($V_9$12, HR2), and (iii) $CD68^+$

262 PD-L1$^+$ tumor-associated macrophages (TAMs) infiltrating the tumor mass ($V_9$14, HR5-7; **Fig. 4a**). S3 also
263 contained non-macrophage populations, including CD31$^+$ vascular endothelial cells ($V_9$2, 23.1% of S3 cells) in
264 the serosa and subserosa (HR3 & 4) and vimentin$^+$ fibroblasts ($V_9$16, 11.7% of S3 cells) in the lamina propria
265 of healthy mucosa (HR1). A second example of cell types incorrectly grouped together in segmentation-based
266 analysis but separated by the VAE was S13 (conventional CD4$^+$ T cells, Tcons). Twenty-seven (27) percent of
267 S13 cells were identified as CD163$^+$ CD4$^+$ tissue-resident macrophages ($V_9$15)[32] in HR1 and HR2 **(Fig. 4b-d**)
268 with the remaining cells distributed across fibroblast and macrophage clusters ($V_9$7, 14, 16). We reasoned that
269 shared expression of CD4 between Tcons and this atypical macrophage population caused them to co-cluster in
270 S13. Thus, pixel-level learning enabled by a VAE was able to discriminate cell populations occupying distinct
271 spatial contexts, lacking definitive immunomarker profiles (e.g., fibroblasts), or expressing atypical
272 combinations of markers (CD163$^+$ CD4$^+$ macrophages). The identification of CD4$^+$ macrophages in CyCIF-1A
273 highlighted how unsupervised learning can "discover" unexpected cell types.
274 $V_9$23 presented another example of how unsupervised learning identified recurrent spatial patterns not
275 part of an *a priori* analysis plan. $V_9$23 represented a rare population of αSMA$^+$ smooth muscle cells lining the
276 basement membrane of an artery within the muscularis propria (0.1% of total cells, HR1; **Fig. 4e**). $V_9$23 image
277 patches were distinguished by the presence of a portion of the external elastic lamina—an acellular structure
278 that provides mechanical support to the blood vessel wall and can be detected based on its strong and
279 photostable autofluorescence[33] (**Fig. 4f**). Acellular structures such as this are visible in images but lack nuclei
280 and are therefore absent from segmentation-based spatial feature tables.
281
282 **Limitations of patch-based learning relative to image segmentation**
283 In some cases, VAE-based learning led to co-clustering of cell types that segmentation-based analysis
284 correctly separated. For example, $V_9$4 contained both Tcons (S13) and regulatory T cells (Tregs, S15;
285 **Extended Data Fig. 8a,b** and **Supplementary Video 3**). In biplots of segmented cells, CD4 and FOXP3
286 formed discrete distributions corresponding to Tcons and Tregs, but this was not true in patch-based
287 representations (**Extended Data Fig. 8c-e**). We speculate that this occurred because nuclear FOXP3 signals
288 represent only a small fraction of total patch dimensionality (4,116 dimensions) and the localization of Tcons
289 and Tregs to similar stromal niches which provides no additional discriminatory information (**Extended Data
290 Fig. 8f,g**). This highlights a limitation of unsupervised learning that may be addressed by increasing the weight
291 of signals for specific lineage markers or cell compartments (e.g., the nucleus) during image patch encoding.
292 In both segmentation and VAE-based analysis, clusters corresponding to lymphoid aggregates (S11,
293 $V_9$11 and $V_9$20, HR*; **Fig. 4g**) were heterogeneous with respect to cell type, containing both B and T cells.

With segmentation, this arose from spatial crosstalk, but in VAE analysis it was due to image patches containing multiple closely packed cells; more than 22% of patches from $V_9$11 and $V_9$20 contained >1 nucleus (**Fig. 4h**). Thus, while unsupervised VAE-based learning has greater sensitivity to morphological and contextual features it cannot separate whole cells that consistently co-appear in image patches. Moreover, a lack of prior knowledge about what is important (e.g., nuclear FOXP3) imposes additional limitations on unsupervised learning and underscores the value of performing pixel-level inference and traditional segmentation in parallel.

VAE analysis identifies matched cell states within and across specimens

To explore the consistency of VAE-based learning across tissue sections and tumors, we trained a new network on 9x9µm image patches from three CyCIF images of CRC labeled with the same 21 antibodies (**Supplementary Table 2**). Tissues included (i) CyCIF-1A, to test technical reproducibility of the method; (ii) a serial section from the same tissue block located ~20 µm along the Z-axis, to evaluate concordance across biological replicates (CyCIF-1B; **Fig. 5a**); and (iii) a tissue section from a different patient, to compare clusters across tumors (CyCIF-2; **Fig. 5b**). Leiden clustering of VAE image patch encodings from all three samples yielded 33 clusters (V_{9M}0-32, with "V_{9M}" denoting a VAE trained on 9x9µm patches from multiple tissues; **Fig. 5c**, **Supplementary Table 5** and **Supplementary Video 4**). Three clusters (V_{9M}29, 31, and 32) were identified as microscopy artifacts (<0.3% of cells) that underwent no further analysis. The remaining clusters corresponded to the expected four primary cell lineages: keratin$^+$ epithelial, αSMA$^+$ smooth muscle, vimentin$^+$ stromal, and CD45$^+$ immune cells (**Extended Data Fig. 9a,b** and **Supplementary Fig. 8**). V_{9M} clusters mapped to CyCIF-1A and 1B revealed strong concordance, with matched clusters occupying similar regions of tissue (**Fig. 5d** and **Supplementary Fig. 9**). Hierarchical clustering of V_{9M} clusters by their protein expression profiles showed that 60% (18 of 30) of two-way dendrogram merges involved CyCIF-1A and 1B, and an additional 37% (11 clusters) involved three-way dendrogram merges between all three samples (**Fig. 5e**). Thus, 97% (60% + 37%) of V_{9M} clusters were matched between serial sections of the same tumor.

Clusters common to all three images included: endothelial cells (V_{9M}4), smooth muscle cells (V_{9M}6), crypt-forming epithelial cells (V_{9M}14), CD8$^+$ T cells (V_{9M}19), CD68$^+$ macrophages (V_{9M}21) and mitotic tumor cells (V_{9M}30; **Fig. 5f**). The latter were likely identified based on the presence of condensed nuclear chromatin (a feature of pro-metaphase and metaphase cells), which was visible in the DNA channel and as a silhouette in the PCNA channel (**Extended Data Fig. 9c**). Due to their scarcity (<0.08% of cells in each specimen), these cells were missed in single-tissue analysis, but detected in multi-tissue analysis due to the inclusion of more data. Four clusters (V_{9M}0, 8, 13, and 23) were unique to CyCIF-2, together accounting for 85.5% of cells (**Fig.**

5g,h and **Extended Data Fig. 9d** and **Supplementary Video 5**). The largest clusters, $V_{9M}0$ (41.2%) and $V_{9M}13$ (17.6%), corresponded to PCNA^{high} and PCNA$^{low/neg}$ tumor cells, respectively (**Extended Data Fig. 9e**). Both exhibited abnormal CD3ε labeling, likely due to expression of cross-reactive epitopes, and explaining why CyCIF-2 tumor cells did not co-cluster with those in CyCIF-1 (**Extended Data Fig. 9f,g**). While $V_{9M}8$ cells (fibroblasts) were found in all three samples, they were 7-fold more abundant in CyCIF-2 due to a higher stroma-to-tumor ratio (**Fig. 5g** and **Extended Data Fig. 9h**). $V_{9M}23$ represented CD45$^+$ cells specific to CyCIF-2 that lacked expression of other immune lineage markers (**Extended Data Fig. 9i**). However, the same tumor-infiltrating CD8$^+$ T cells ($V_9 18$) and surrounding keratin$^+$ tumor cells ($V_9 22$) in CyCIF-1A were also identified in the two other images ($V_{9M}24$ and $V_{9M}27$; **Extended Data Fig. 9j**), emphasizing the ability of a VAE to reproducibly overcome spatial crosstalk. These results show that VAE-based learning reliably identified similar cell types across tissues, including rare populations. However, differences in tumor cells across patients led to patient-specific clusters.

VAE analysis of multicellular image patches identifies recurrent tissue motifs

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

Because V_{30} clusters contained more pixels and were not vignetted, concept saliency maps were higher resolution, with high-attribution pixels forming clear multi-cellular spatial patterns (**Fig. 6e**, **Extended Data Fig. 10b**, and **Supplementary Image Gallery 4**). For example, keratin saliency maps reproduced the characteristic morphologies of epithelial cells in normal mucosal crypts ($V_{30}7$) and glandular AC ($V_{30}8$), while

358 those for lymphoid aggregates ($V_{30}9$) highlighted strong contributions from $CD20^+$ B cells and $CD4^+$ helper T
359 cells, suggesting their importance in defining this cluster concept. In contrast, $CD8^+$ T cells contributed
360 negatively to the lymphoid aggregate cluster concept, consistent with the observation that $CD8^+$ T cells were
361 less abundant in these structures than B and helper T cells.

362 To better understand the cell type composition of V_{30} clusters, we used information entropy computed
363 against V_9 cluster labels. Entropy scores (which had possible values between 0 to 4.585 bits based on 24
364 possible 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-
365 entropy clusters $V_{30}7$, 8 and 10 corresponded to a single V_9 cluster label, showing that these structures were
366 primarily composed of a single cell type. For example, 92% of cells in $V_{30}7$ were normal crypt-forming
367 epithelial cells ($V_9 10$; **Fig. 6g**). However, cells in $V_{30}1$ (a high-entropy cluster) comprised ten different V_9
368 clusters, including $CD4^+$ T cells ($V_9 4$, 19.0%), $CD68^+$ $CD163^+$ macrophages ($V_9 7$, 17.3%), $CD31^+$ endothelial
369 cells ($V_9 2$, 15.8%), $CD68^+$ PD-$L1^+$ macrophages ($V_9 14$, 10.9%), $vimentin^+$ fibroblasts ($V_9 16$, 10.7%), $CD8^+$ T
370 cells ($V_9 13$, 9.5%), and $CD163^+$ $CD4^+$ macrophages ($V_9 15$, 6.3%; **Fig. 6h**). This diversity reflects the role of
371 the colonic stroma in structural support, nutrient exchange, and wound healing mediated by fibroblasts,
372 immune, and other cell types. Notably, both the lamina propria of healthy mucosal crypts (HR1) and the
373 peritumoral stroma surrounding cancer cells (HR9) were both part of $V_{30}1$, suggesting that these compartments
374 share a common set of constituent cell types (at least at this level of detail). We conclude that a VAE trained on
375 multicellular image patches can correctly identify morphologically distinct arrangements of single cell types as
376 well as complex cellular communities comprising multiple lineages in normal and desmoplastic stromal tissue.

DISCUSSION

In this paper, we describe a framework (Morphology Oriented Research in Pathology Harnessing Autoencoder Unsupervised Strategies, MORPHAEUS) that combines the strengths of unsupervised pixel-level ML and traditional image segmentation. MORPHAEUS supports tissue classification across spatial scales, from single cells to tissue neighborhoods and has tools for image patch pre-processing, VAE model training, unsupervised clustering, and saliency analysis. VAEs are advantageous because their learned representations are interpretable. By interrogating VAE clustering results using a combination of learned representations, concept saliency maps, and expert histopathologist review, we find that VAEs simultaneously learn information on protein abundance, cell morphology, and local neighborhood. For example VAE-based learning on 9x9μm image patches successfully distinguished glandular, mucinous, and solid tumor subtypes[6] as well as morphological variation associated with tumor dedifferentiation. VAEs also exploit spatial patterns in marker levels to learn neighborhood features directly, thereby overcoming problems with spatial crosstalk that plague segmentation-based methods. This enabled the VAE to directly identify biologically significant cellular niches such as $CD8^+$ T cells in stromal ($V_9$13) versus tumor tissue ($V_9$18). Moreover, whereas segmentation isolates single cells whose spatial relationships must be inferred through explicit distance calculations, training VAEs on patches of different sizes makes it possible to learn both cell-level and contextual features simultaneously.

Generative models such as VAEs and their variants (e.g., adversarial autoencoders (AAEs), deep convolutional generative adversarial networks (DCGANs), and combinations thereof such as VAE-GANs) require no ground truth annotations, making them well-suited to analysis of complex multiplex images. For example, in this study VAEs detected an unexpected population of $CD163^+$ $CD4^+$ tissue-resident macrophages[32] ($V_9$15) missed by segmentation-based clustering, mitotic tumor cells (V_{9M}30) based on their distinctive DNA morphology, and cells adjacent to acellular structures such as the arterial elastic lamina ($V_9$23). Acellular features are readily observable by humans and important in understanding tissue biology, but are missed by segmentation-based workflows.

Despite the strengths of unsupervised VAE-based learning, we show that it can miss subtle nuclear features such as the difference between Tcons and Tregs based on the presence of the nuclear transcription factor, FOXP3.[34] Segmentation-based methods are superior in this case because they implicitly increase the saliency of nuclear signals due to the use of nuclear segmentation masks. The same effect might be achievable during VAE image patch encoding via feature engineering.[35] 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 approach in which a smaller patch size is used for cell-dense regions of tissue such as TLSs with larger patches used to infer cells in the remainder of tissue. However, such approaches will likely be most effective using higher

resolution or 3D imaging[4] that can more effectively discern between cell types. In some cases it may also be reasonable to simply view T and B cell pairs in densely-packed lymphoid aggregates as a recurrent tissue motif rather than sets of fully discriminable cells. Thus, as a general approach for generating reliable single cell data from high plex images, 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.

414 **SUPPLEMENTARY INFORMATION**

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

416 Highly multiplex tissue imaging is a rapidly evolving technology whose analysis often involves
417 defining boundaries around individual cells through the process of image segmentation. In the case of whole-
418 slide tissue imaging, this involves the acquisition of tens to hundreds of image tiles stitched and registered
419 together using algorithms such as ASHLAR[36] that form multi-channel mosaic images stored in metadata-rich
420 open file formats like OME-TIFF[37] (level 2 data according to MITI guidelines[2], **Supplementary Fig. 1a**).
421 Semantic segmentation algorithms, such as the popular U-Net[38,39] architecture and its derivatives such as
422 UnMICST[28], TissueNet[40], and Cellpose[41] are then applied to images of cell nuclei labeled with DNA
423 intercalating dyes (e.g., Hoechst or DAPI) to identify boundaries between cells and generate cell segmentation
424 masks—reference images in which pixels are labeled according to the cell they belong to (MITI level 3 data,
425 **Supplementary Fig. 1b**).

426 Mean immunomarker signals for individual cells are then programmatically computed and organized
427 into a tabular data format like CSV (level 4 data according to MITI guidelines, **Supplementary Fig. 1c**). The
428 derived tables are often augmented by measurements of the segmentation instances themselves, which serve as
429 proxies for nuclear morphology (e.g., area, eccentricity, etc.). However, the accuracy of these measurements is
430 conditional on segmentation quality. While standard pipelines such as MCMICRO[21] have been developed to
431 automate the aforementioned image processing tasks, no consensus exists on how best to classify cell
432 phenotypes in tissue-derived, single-cell data.

433 A common method for segmentation-based cell state inference involves manual thresholding per-cell
434 signal intensities (i.e., hard gating) performed using univariate histograms or bivariate scatter plots
435 (**Supplementary Fig. 1d**). However, as positive and negative cell populations often overlap, such methods can
436 be sensitive to even small changes in gate placement, leading to ambiguous annotation and potentially
437 changing biological interpretation. Methods based on Gaussian mixture models (GMMs) and naïve Bayes
438 classifiers such as those implemented by the marker-scoring functions in CELESTA[42] and MCMICRO[21],
439 respectively, represent "soft" gating methods. Such methods have the advantage of assigning class-conditional
440 probability weights to individual cells as an alternative to strict thresholding (**Supplementary Fig. 1e**).
441 Nevertheless, these methods also require manual parameter tuning and are sensitive to the same limitations
442 imposed by segmentation-based analysis.

443 Unsupervised clustering methods (e.g., Louvain[43] or Leiden[25] community detection, k-means[44], SOMs[45]
444 (self-organizing maps), and HDBSCAN[46] (hierarchical density-based spatial clustering of applications with
445 noise)) are an alternative to gate-based approaches for cell state inference. In these methods, high-dimensional

single-cell data are projected onto lower dimensional embeddings (typically 2D or 3D) using data dimensionality reduction algorithms such as t-SNE (t-distributed stochastic neighbor embedding)[47] or UMAP (uniform manifold approximation and projection)[26] before or after clustering (**Supplementary Fig. 1f**). Regardless of the method used for cell phenotyping, protein expression profiles for inferred cell populations are often visualized as clustered heatmaps that can be interpreted by domain experts according to prior knowledge of cellular protein expression profiles (**Supplementary Fig. 1g**).

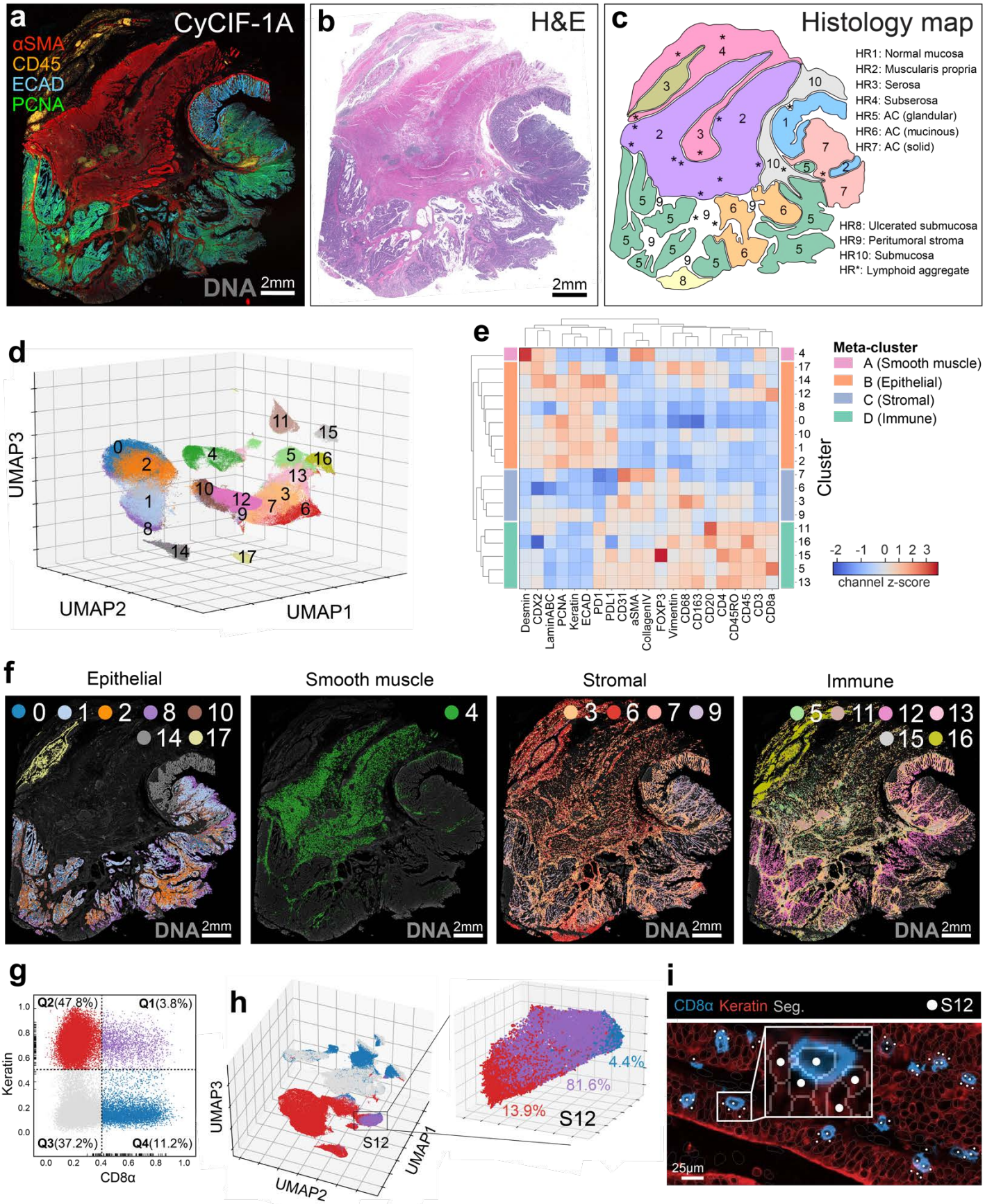
Supplementary Note 2 – Analysis of Errors in Segmentation Cluster S5

Based on a heatmap of marker expression, segmentation cluster S5 appeared to express markers associated with multiple cell types, including smooth muscle cells (αSMA, collagen IV), macrophages (CD68, CD163), endothelial cells (CD31), mesenchymal cells (vimentin), and $CD8^+$ T cells (CD45, CD3ε, CD8α, CD45RO; **Fig. 1e**). Unexpectedly, cells in this cluster gave rise to multiple subclusters in UMAP space, which we termed S5.1, S5.2, and S5.3 (**Supplementary Fig. 4a**). Viewing these cells in CyCIF-1A revealed that they corresponded to phenotypically similar $CD8^+$ T cells occupying distinct histological regions that also differed with respect to their protein expression profiles (**Supplementary Fig. 4d,c**). Closer inspection showed cells in S5.1 were influenced by spatial crosstalk with $CD68^+$ PD-$L1^+$ macrophages located in the peritumoral stroma (HR9) and stromal compartments throughout the muscularis propria (HR2; **Supplementary Fig. 4d**), while cells in S5.2 were affected by crosstalk with $Desmin^+$ smooth muscle cells in the muscularis (**Supplementary Fig. 4e**). S5.3 cells, which were located in the serosa (HR3) and subserosa (HR4), did not appear to be impacted by spatial crosstalk. Instead, they were distinguished by high CD45 expression, a feature we had previously observed in immune cells occupying these tissue compartments (**Supplementary Fig. 4f**).

Although all S5 cells were grouped under the same label by the Leiden clustering algorithm, they formed three distinct subpopulations in 3D UMAP space. This observation led us to hypothesize that the effects of spatial crosstalk are more apparent in low-dimensional embeddings of high-plex data due to data compression. This idea aligns with previous findings that distances and relationships between data points are not reliably preserved during dimensionality reduction[48]. To test this hypothesis, we performed density-based clustering using HDBSCAN[46] on both 2D and 3D UMAP embeddings of cells from CyCIF-1A. Supporting our hypothesis, we found that cells grouped into a single cluster in 3D UMAP space were often redistributed into multiple clusters in a 2D embedding. For example, a subset of tumor cells originally grouped within cluster 17 in a 3D embedding (**Supplementary Fig. 5a-c**) were delocalized in a 2D embedding of the same data (**Supplementary Fig. 5d**). These delocalized cells were found to reside at the tumor-stromal interface and were influenced by spatial crosstalk from collagenIV signals originating in neighboring stromal cells

(**Supplementary Fig. 5e,f**). Examples like this help explain the delocalization of three $CD8^+$ T cell subsets comprising S5 (S5.1, S5.2, and S5.3) in UMAP feature space. They also underscore the creation of artifactual clusters by spatial crosstalk.

481 **FIGURES and LEGENDS**



482

483 **Fig. 1 | Spatial crosstalk limits accuracy of segmentation-based cell phenotyping. a**, CyCIF-1A: a 22-plex,
484 whole-slide CyCIF image of primary human CRC (HTAN ID: HTA13_1_101) showing five channels: αSMA
485 (smooth muscle, red), CD45 (leukocytes, orange), ECAD (epithelium, blue), PCNA (proliferating cells, green),
486 and DNA (Hoechst-stained nuclei, gray). See **Supplementary Table 2** for a complete list of antibodies used to
487 label this tissue. **b**, Adjacent tissue section to CyCIF-1A (~4μm away in Z-axis) stained with routine histology
488 dyes Hematoxylin & Eosin (H&E). **c**, Histology map showing the locations of 11 histological regions (HRs)
489 observed by two board-certified pathologists in the H&E image shown in panel b. **d**, UMAP embedding of
490 segmented cells from CyCIF-1A colored by Leiden cluster; see **Supplementary Video 1** for 360° plot rotation.
491 **e**, Hierarchical clustering of Leiden cluster z-score-normalized mean marker intensities revealing four primary
492 meta-clusters (A-D). **f**, Location of Leiden clusters in CyCIF-1A grouped by major cell lineages: Epithelial,
493 Smooth muscle, Stromal, and Immune. **g**, Gated bivariate plot showing the percentage of CD8α$^+$ keratin$^+$ (Q1,
494 purple), keratin$^+$ (Q2, red), CD8α$^-$ keratin$^-$ (Q3, gray), and CD8α$^+$ cells (Q4, blue) in CyCIF-1A. Rugplots
495 (small black ticks along x- and y-axes) indicate marker intensity values for 100 cells drawn at random from
496 S12 showing that the majority of cells in this cluster are positive for both markers. **h**, UMAP embedding of
497 segmented cells from CyCIF-1A colored by the gating strategy shown in panel g; cluster S12 is highlighted by
498 the black rectangle. Inset shows S12 in further detail highlighting its percentages of keratin$^+$ (red), CD8α$^+$
499 (blue), and CD8α$^+$ keratin$^+$ cells (purple). **i**, Field-of-view from CyCIF-1A showing that cells in S12 (white
500 scatter points) consist of both CD8α$^+$ T cells (blue) and neighboring keratin$^+$ epithelial cells (red). Inset shows
501 spillover of marker signals across segmentation boundaries (translucent perimeters) between the two different
502 cell types, which we refer to as "spatial crosstalk".

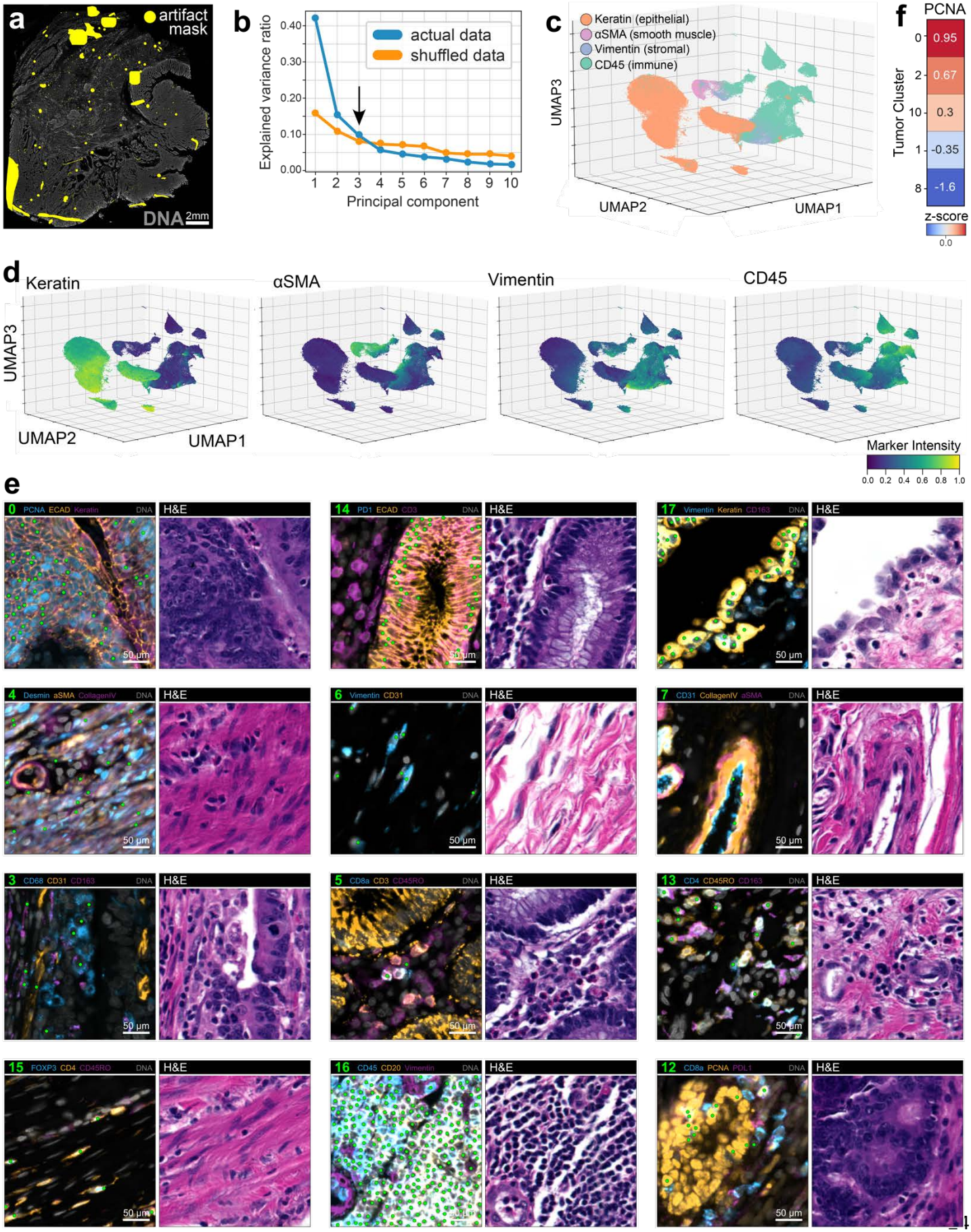
503

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

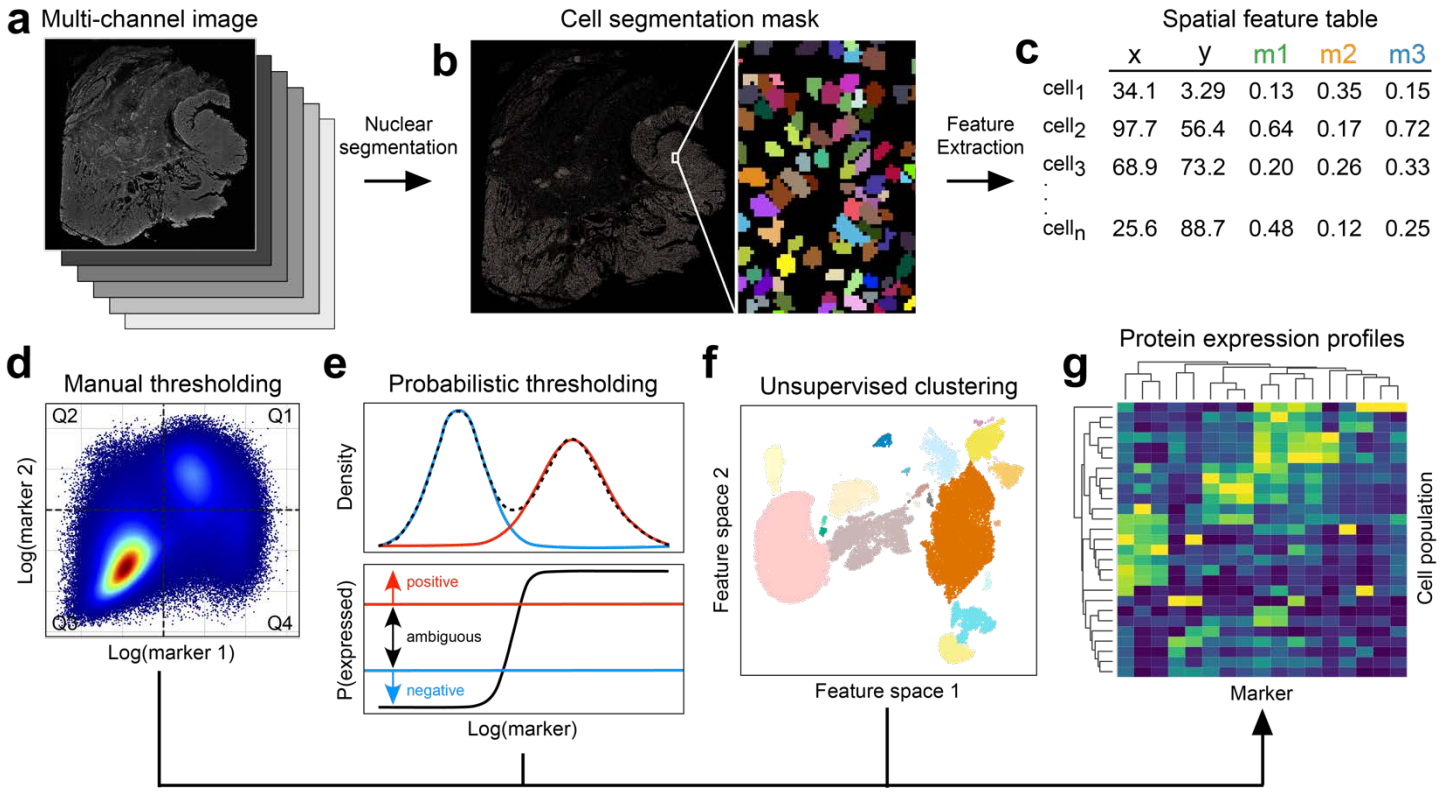
504 **Supplementary Table 1 | Metadata for biospecimens used in this study.** Primary imaging data for all four
505 samples can be found at the HTAN Data Portal (https://data.humantumoratlas.org/explore) by referencing their
506 corresponding HTAN IDs. Additional information on the CyCIF antibodies used to label the tissue can be
507 found in the STAR Methods section of Lin et al. *Cell* 2023 (DOI: 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.


a
artifact mask
DNA
2mm
b
Explained variance ratio
0.40
0.30
0.20
0.10
0.00
actual data
shuffled data
1 2 3 4 5 6 7 8 9 10
Principal component
c
Keratin (epithelial)
αSMA (smooth muscle)
Vimentin (stromal)
CD45 (immune)
UMAP3
UMAP2
UMAP1
f
PCNA
Tumor Cluster
0 2 10 1 8
0.95
0.67
0.3
-0.35
-1.6
z-score
0.0
d
Keratin
αSMA
Vimentin
CD45
UMAP3
UMAP2
UMAP1
Marker Intensity
0.0 0.2 0.4 0.6 0.8 1.0
e
0 PCNA ECAD Keratin DNA
H&E
14 PD1 ECAD CD3 DNA
H&E
17 Vimentin Keratin CD163 DNA
H&E
4 Desmin αSMA CollagenIV DNA
H&E
6 Vimentin CD31 DNA
H&E
7 CD31 CollagenIV αSMA DNA
H&E
3 CD68 CD31 CD163 DNA
H&E
5 CD8a CD3 CD45RO DNA
H&E
13 CD4 CD45RO CD163 DNA
H&E
15 FOXP3 CD4 CD45RO DNA
H&E
16 CD45 CD20 Vimentin DNA
H&E
12 CD8a PCNA PDL1 DNA
H&E
50 µm

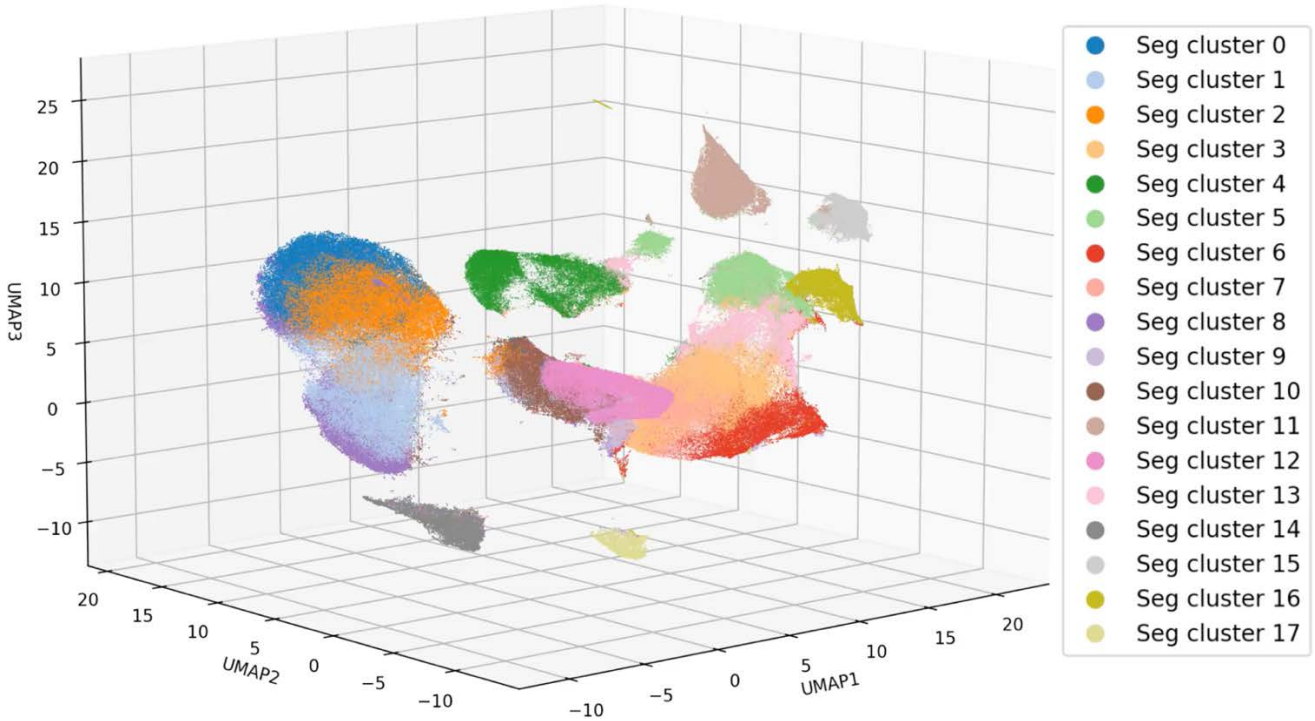
Extended Data Fig. 1 | Segmentation-based cell phenotyping of human CRC. a, Artifact masks (yellow ROIs, n=460) highlighting regions in CyCIF-1A affected by microscopy artifacts (e.g., tissue folds, antibody aggregates, illumination aberrations, etc.) that were censored from further analysis. **b**, 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.



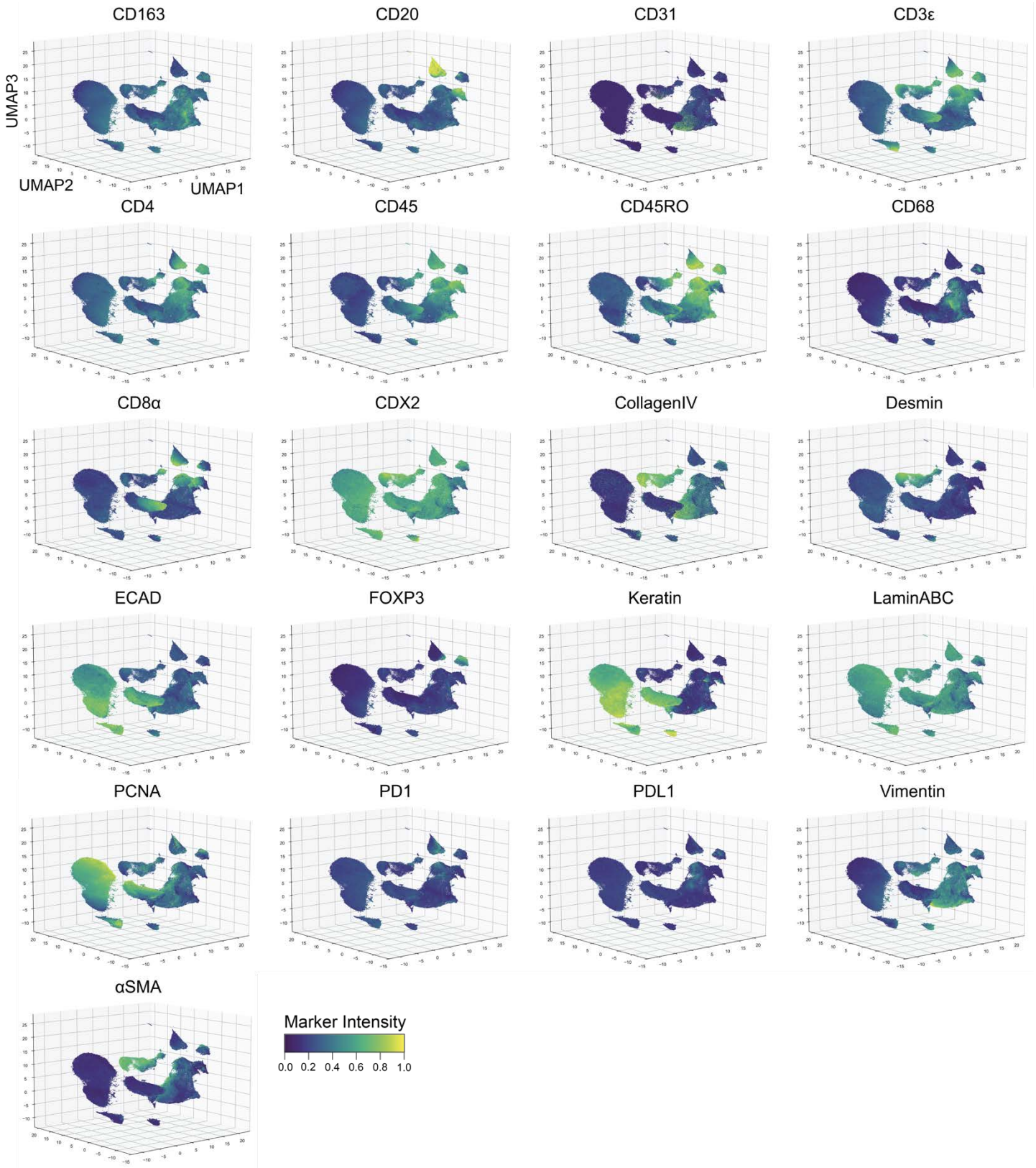
Supplementary Fig. 1 | Workflow for segmentation-based cell phenotyping of multiplex tissue images. **a**, Multi-channel OME-TIFF/TIFF file. **b**, Segmentation mask generated by semantic segmentation of counterstained cell nuclei. **c**, Spatial feature table specifying X/Y spatial coordinates and mean 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 Baysian 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.

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

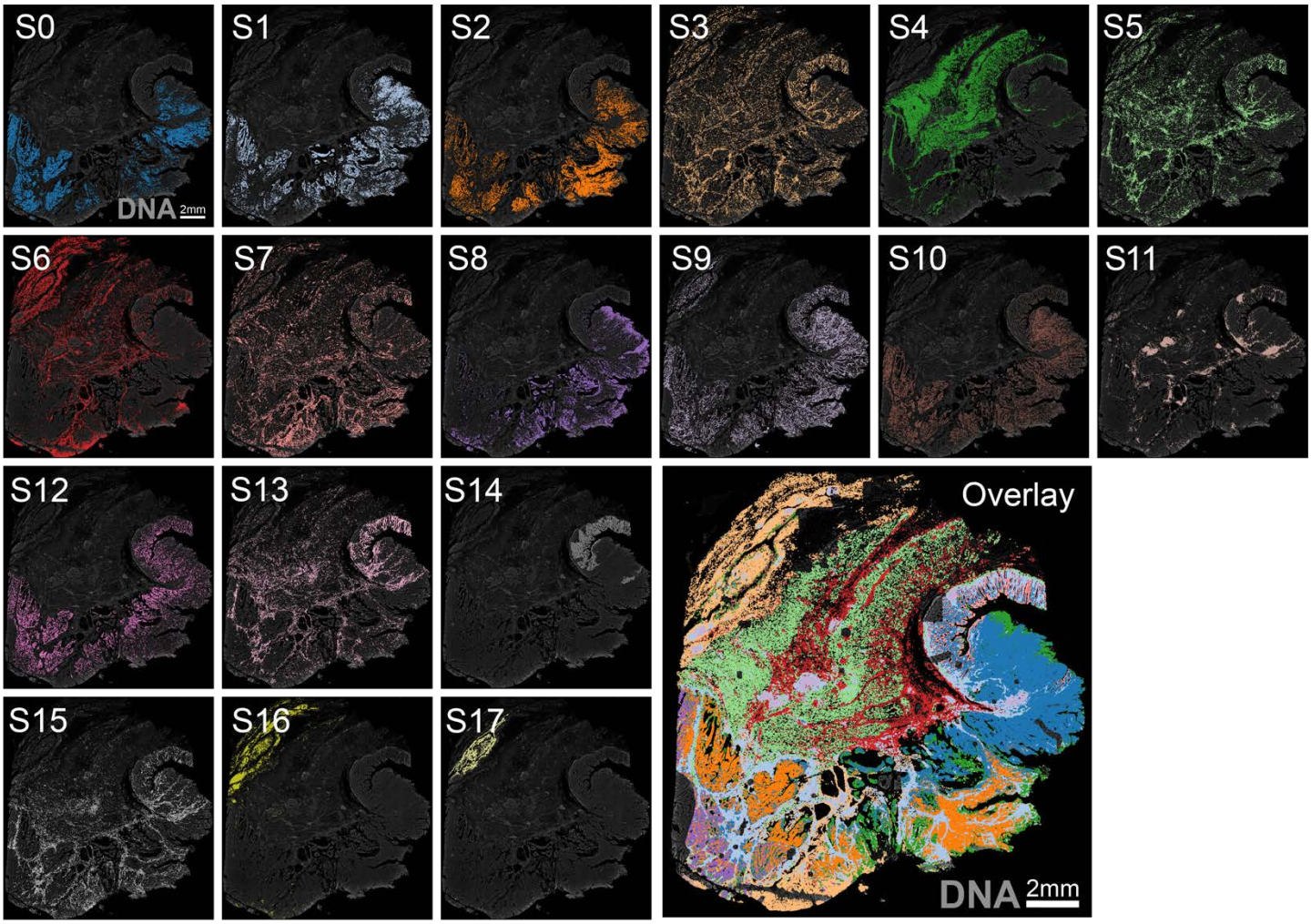
Supplementary Table 3 | Segmentation-based Leiden cluster annotations. The column labeled "% total cells" indicates the percentage of total cells in CyCIF-1A accounted for by that cluster. Column labeled "Spatial crosstalk" indicates whether signal spillover across segmentation boundaries of neighboring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



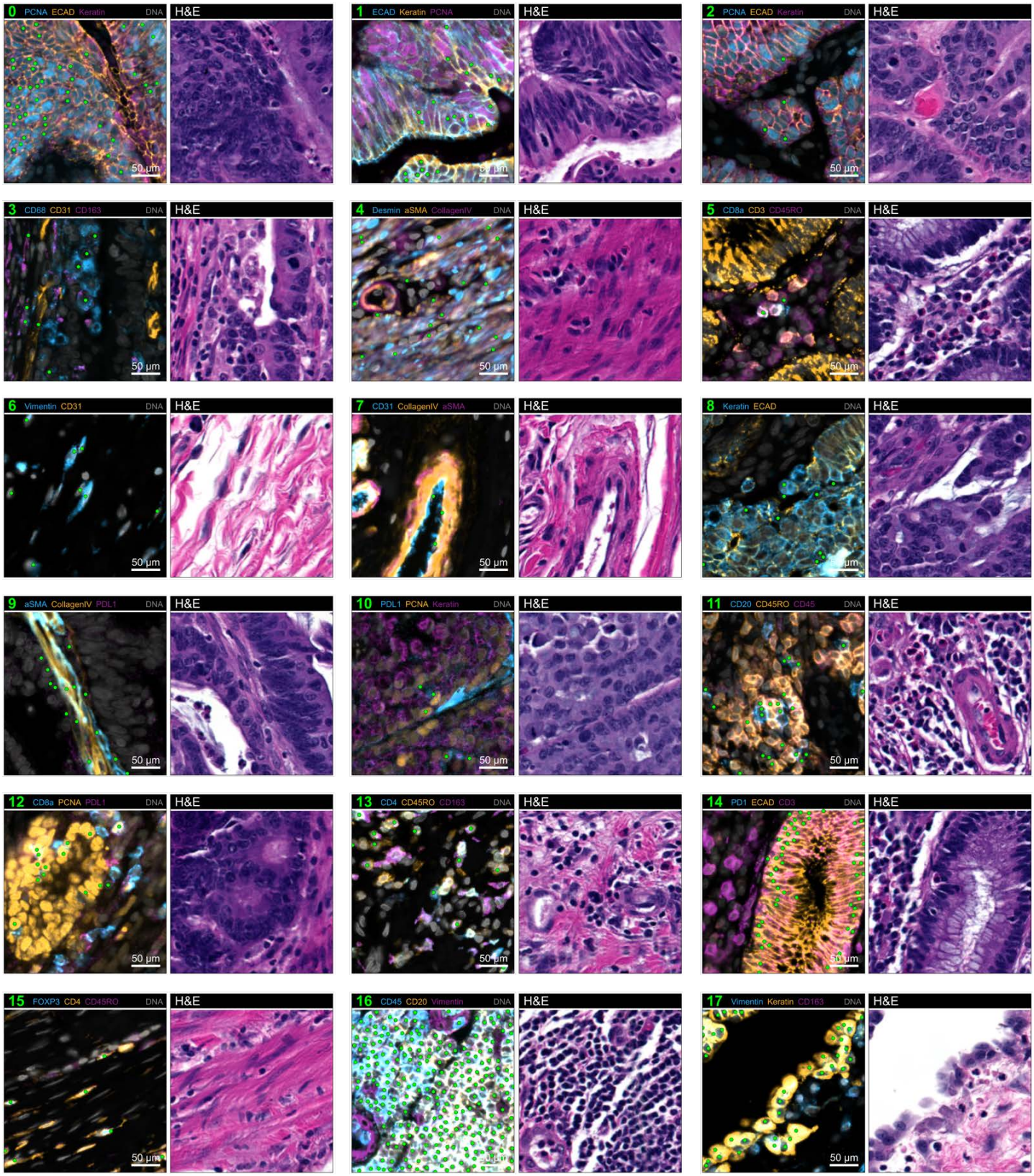
Supplementary Video 1 | 360° UMAP plot rotation showing segmented cells colored by Leiden cluster label.



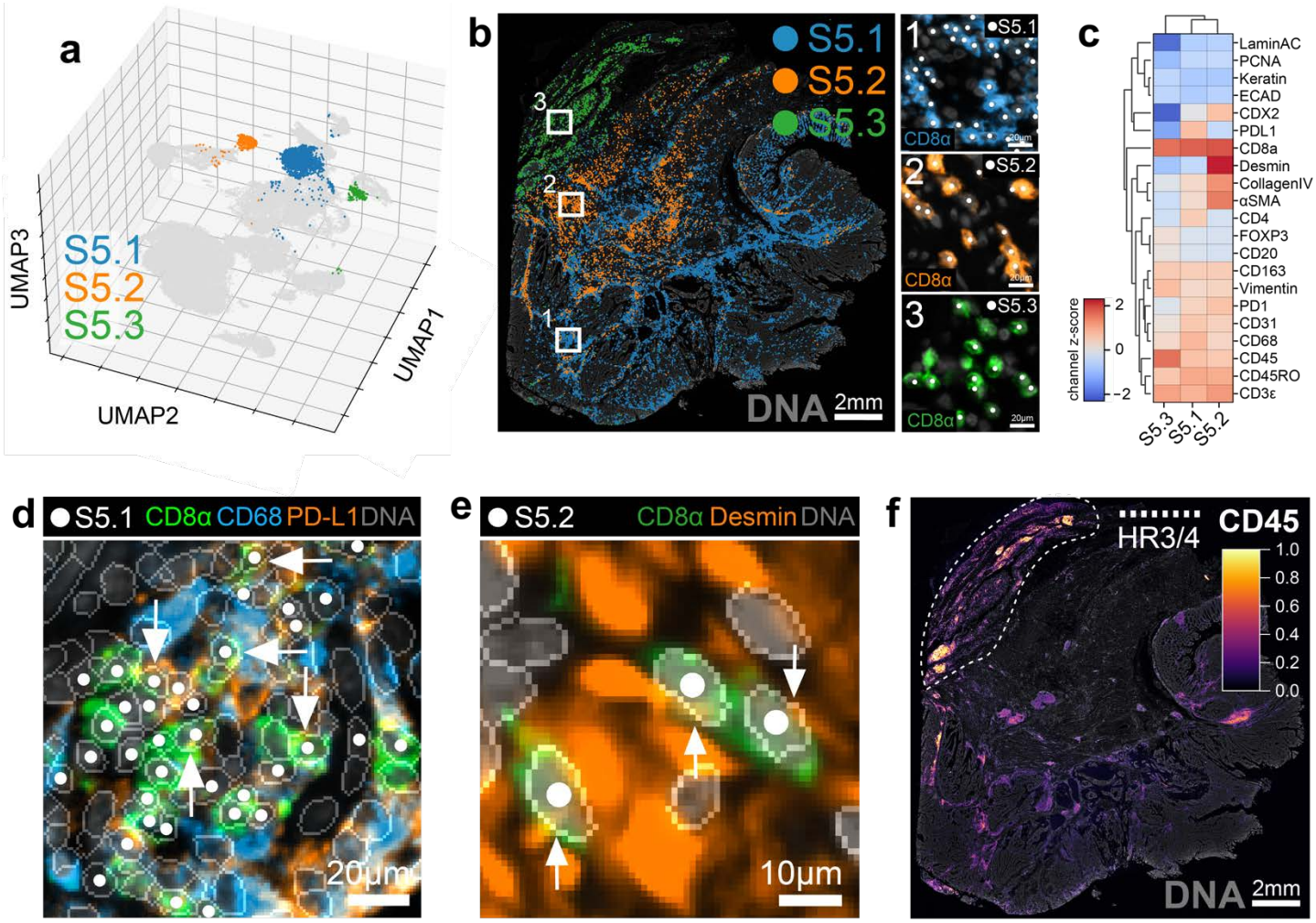
Supplementary Fig. 2 | Channel colormaps for segmented cells. UMAP embeddings of segmented cells from CyCIF-1A colored by mean marker intensity for 21 immunomarkers.



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

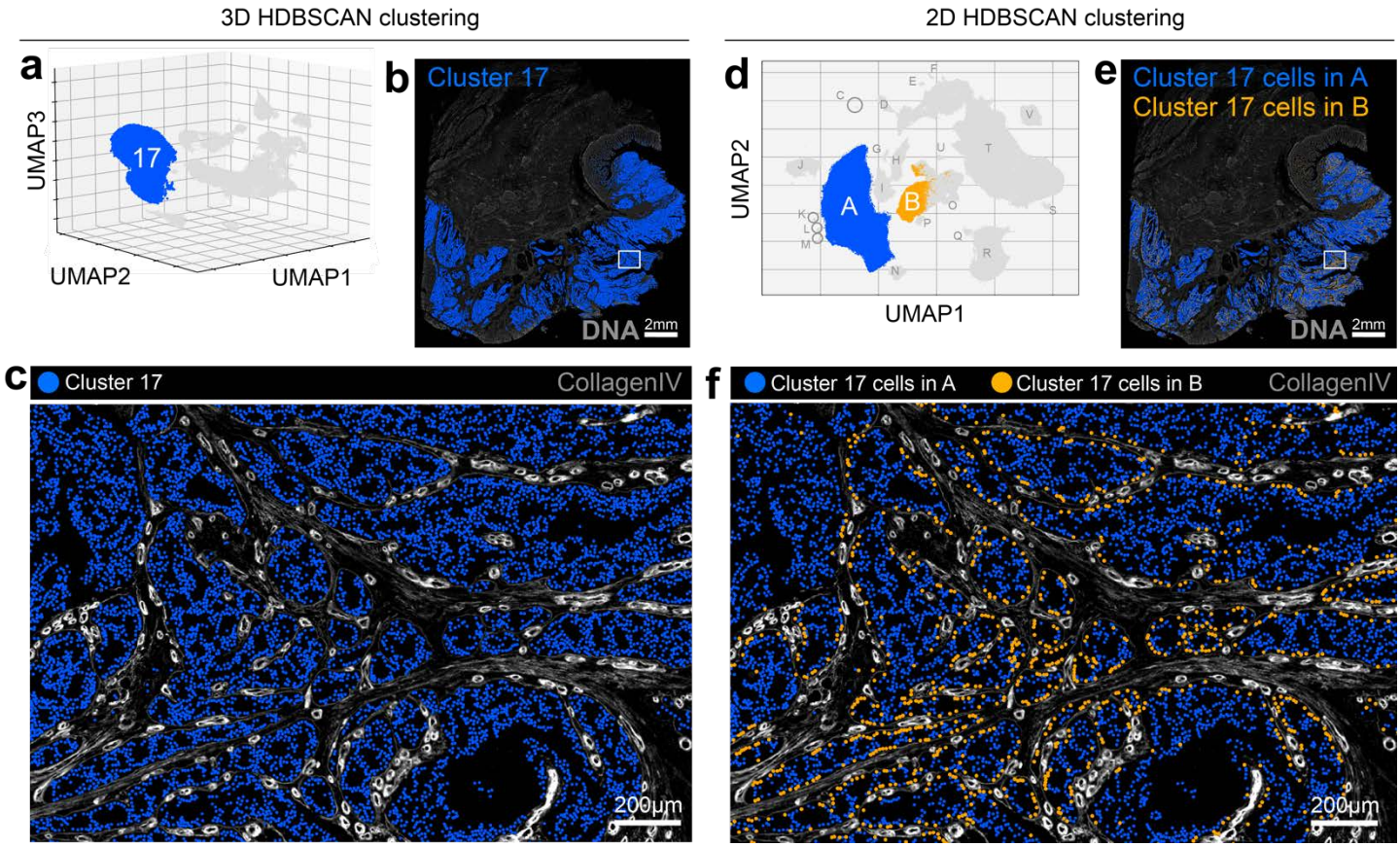


Supplementary Fig. 3 | Fields-of-view showing the location of segmentation-based Leiden clusters in CyCIF-1A. Paired CyCIF (left) and H&E (right) fields-of-view in CyCIF-1A showing the location of random examples of cells from all 18 Leiden clusters (green scatter points in CyCIF images).

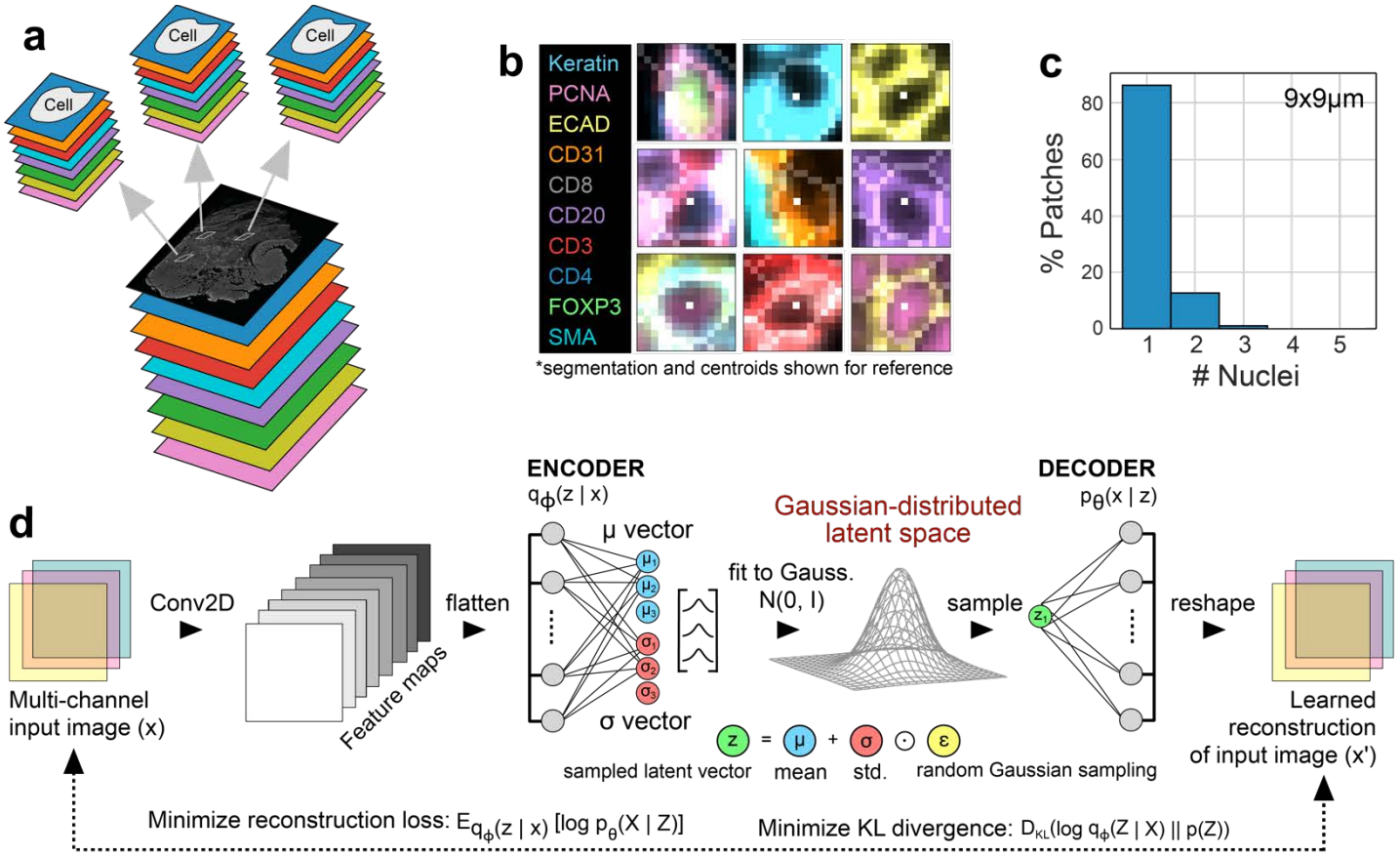


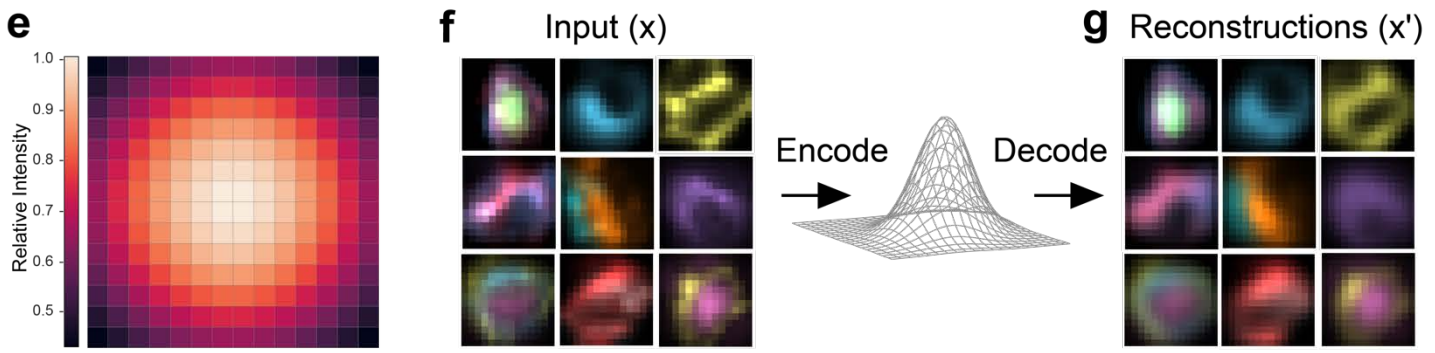
Supplementary Fig. 4 | CD8$^+$ T cells in segmentation cluster S5 are delocalized in UMAP space and affected by spatial crosstalk with cells in different tissue niches. a, 3D UMAP embedding of segmented cells from CyCIF-1A highlighting three subclusters of S5 cells: S5.1 (blue), S5.2 (orange), and S5.3 (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

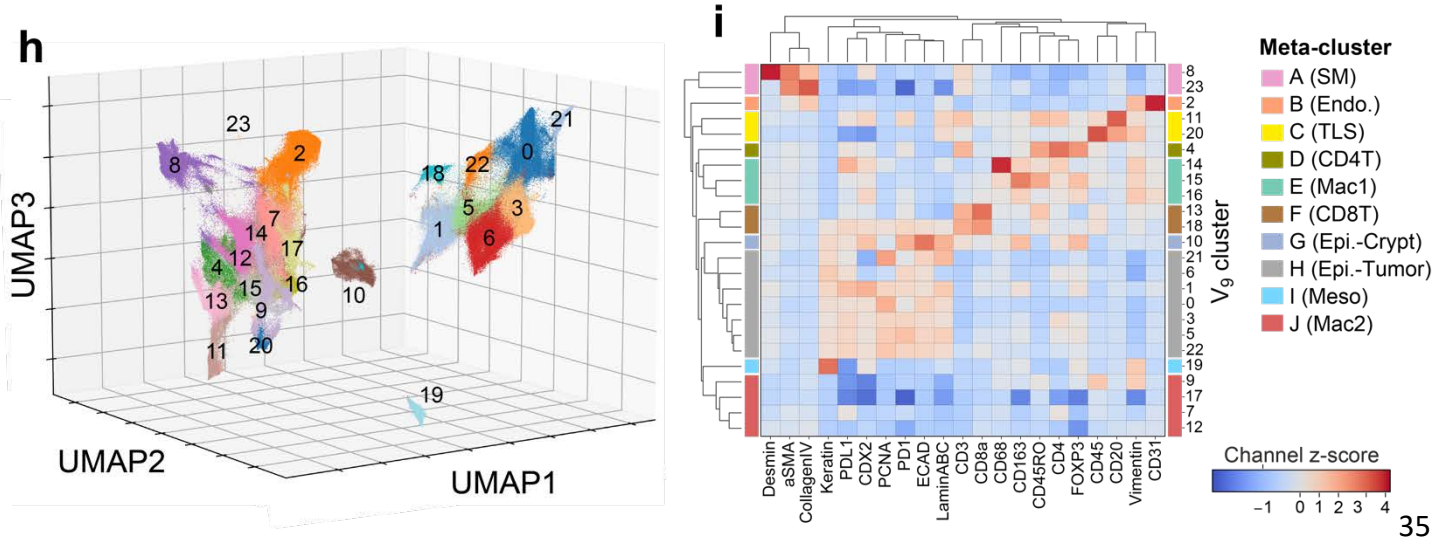
$CD8^+$ T cells in S5.3 reside, exhibit higher CD45 intensity compared to other areas of the tissue, helping to explain why these cells form their own cluster.



Supplementary Fig. 5 | 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.

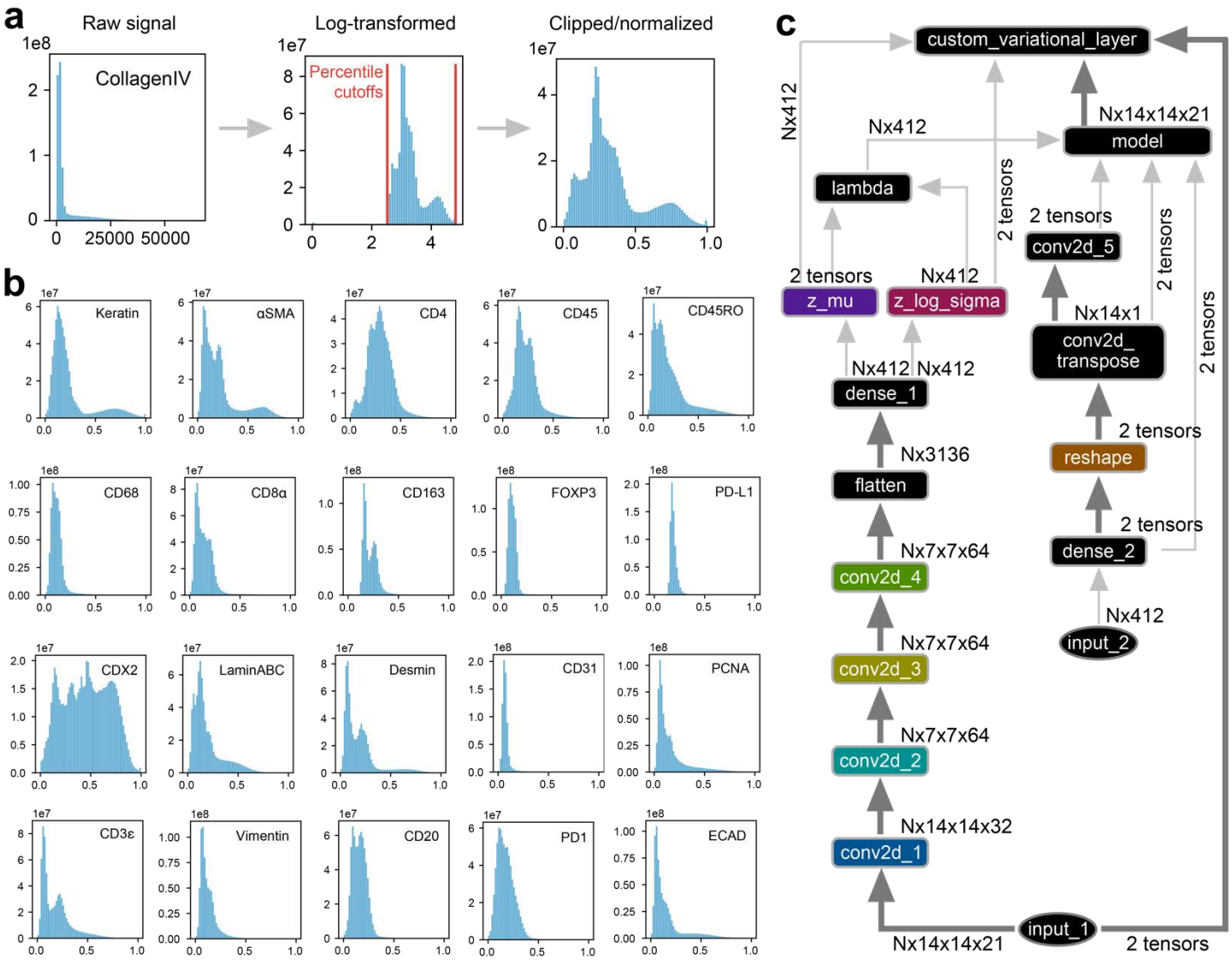

a
Cell
Cell
Cell
b
Keratin
PCNA
ECAD
CD31
CD8
CD20
CD3
CD4
FOXP3
SMA
*segmentation and centroids shown for reference
c
9x9µm
% Patches
# Nuclei
d
Multi-channel input image (x)
Conv2D
Feature maps
flatten
ENCODER
$q_\phi(z \mid x)$
µ vector
σ vector
Gaussian-distributed latent space
fit to Gauss. N(0, I)
sample
DECODER
$p_\theta(x \mid z)$
reshape
Learned reconstruction of input image (x')
sampled latent vector
mean
std.
random Gaussian sampling
Minimize reconstruction loss: $E_{q_\phi(z \mid x)}[\log p_\theta(X \mid Z)]$
Minimize KL divergence: $D_{KL}(\log q_\phi(Z \mid X) \mid\mid p(Z))$


e
Relative Intensity
f
Input (x)
Encode
Decode
g
Reconstructions (x')


h
UMAP3
UMAP2
UMAP1
i
Meta-cluster
A (SM)
B (Endo.)
C (TLS)
D (CD4T)
E (Mac1)
F (CD8T)
G (Epi.-Crypt)
H (Epi.-Tumor)
I (Meso)
J (Mac2)
V_9 cluster
Desmin
aSMA
CollagenIV
Keratin
PDL1
CDX2
PCNA
PD1
ECAD
LaminABC
CD3
CD8a
CD68
CD163
CD45RO
CD4
FOXP3
CD45
CD20
Vimentin
CD31
Channel z-score

587

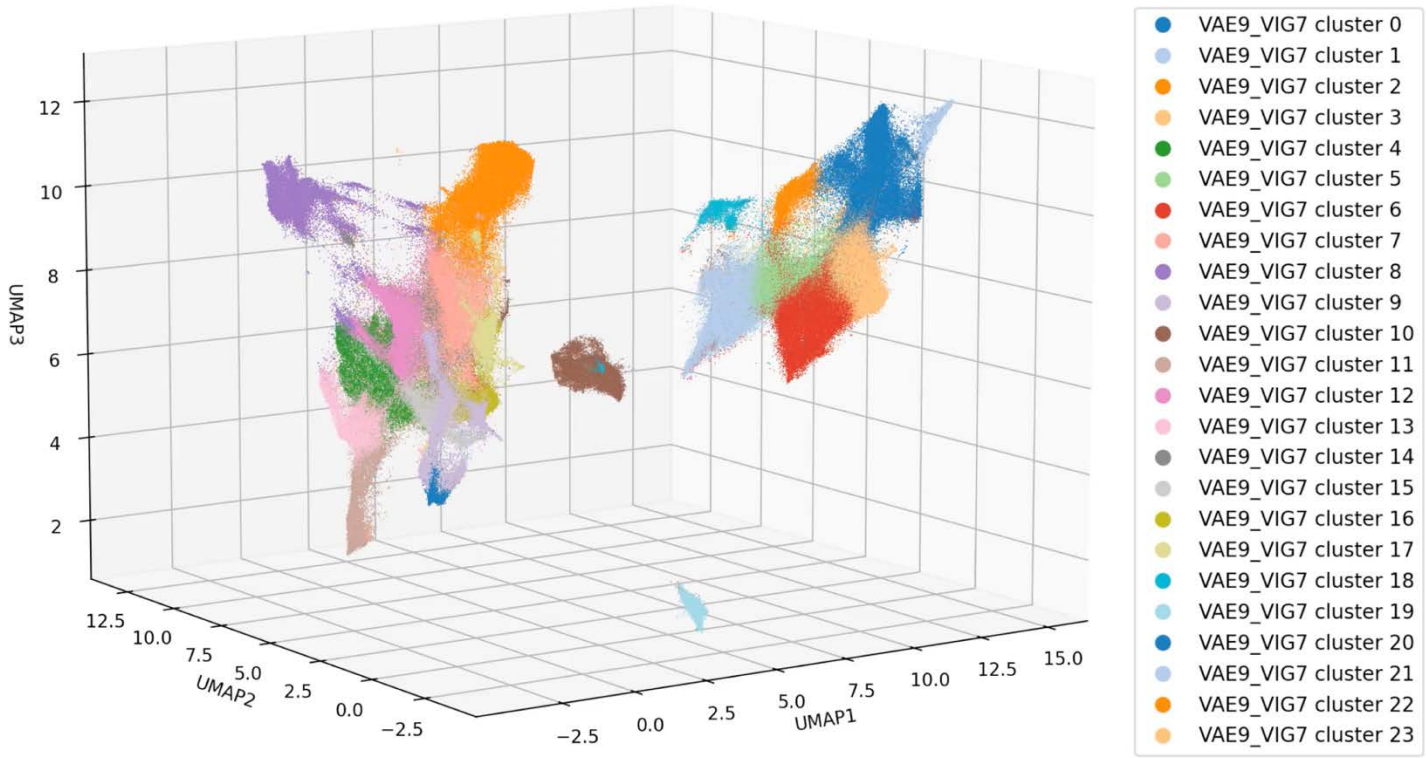
Fig. 2 | VAE-based classification of human CRC at single-cell resolution. a, Schematic showing the process of extracting multi-channel, single-cell image patches from whole-slide multiplex tissue images using Cellcutter. **b**, Example 9x9µm multi-channel image patches extracted from CyCIF-1A. Nuclear 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).



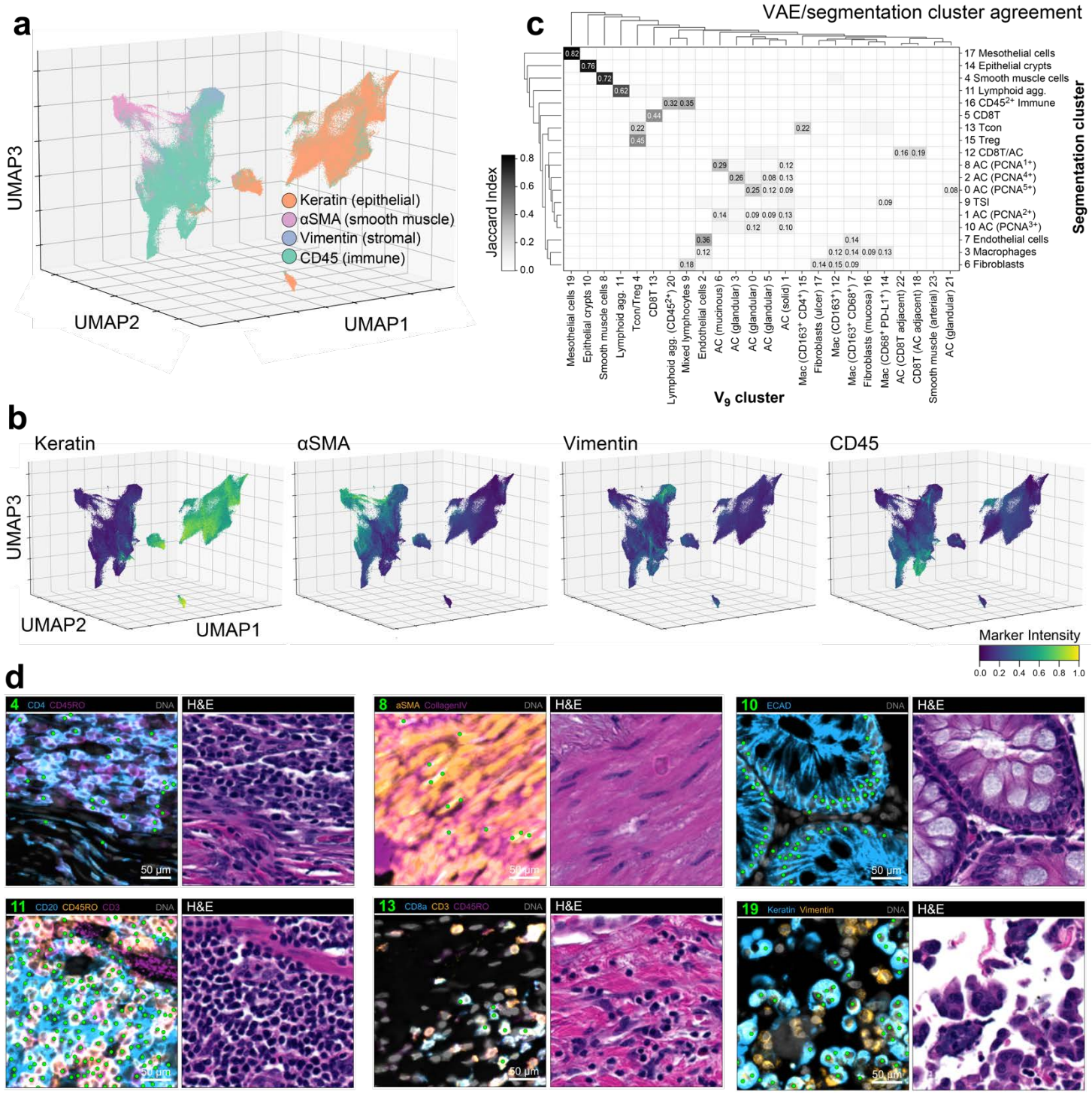
Extended Data Fig. 3 | VAE network architecture and image patch preprocessing. a, Preprocessing steps undergone by image patches prior to entering the VAE encoding network using the collagen IV channel as an example showing the raw pixel intensity distribution for the full CyCIF-1A image channel (right 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.

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)

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

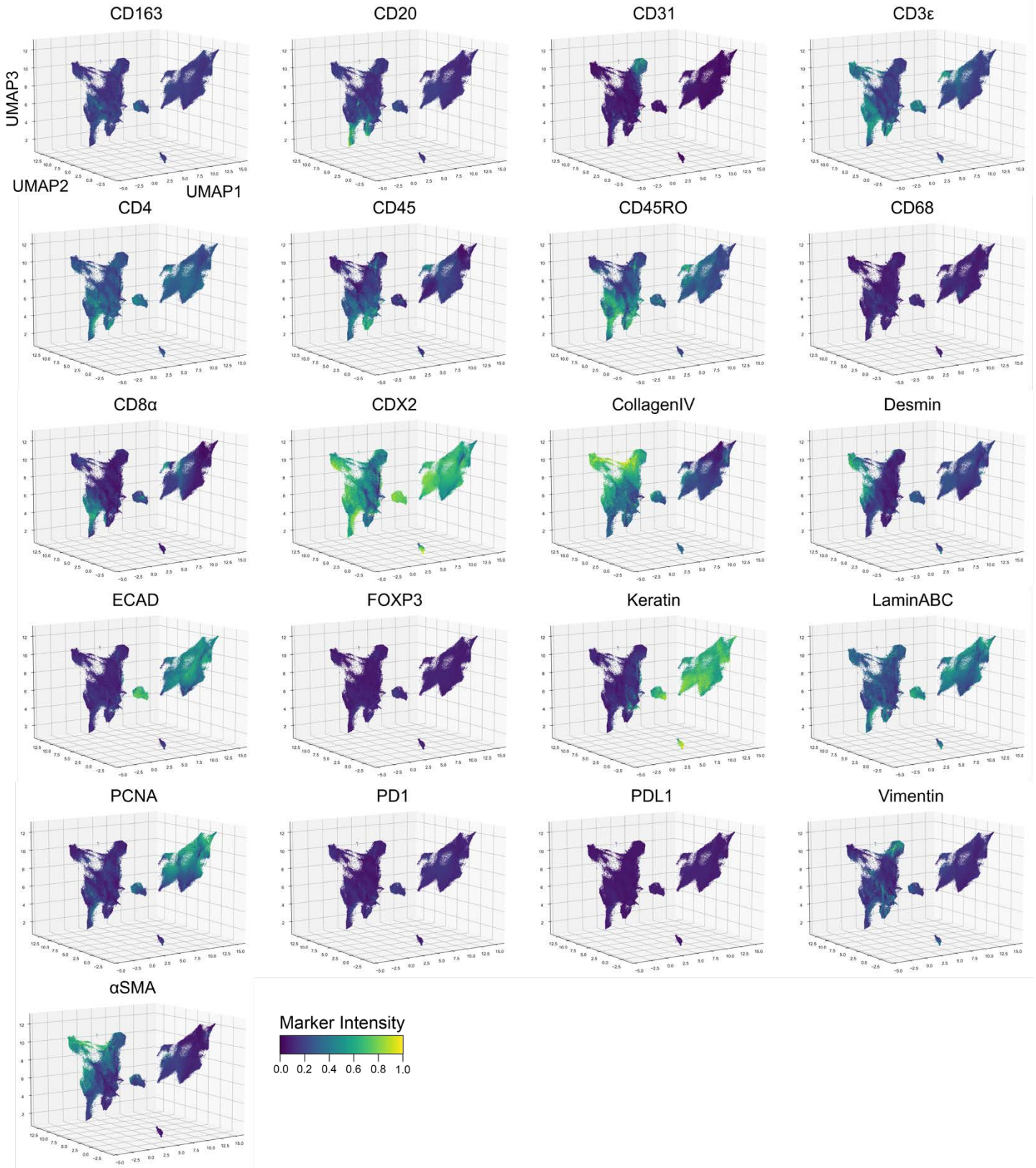


Supplementary Video 2 | 360° UMAP plot rotation showing 9x9µm VAE image patch encodings colored by Leiden cluster label.

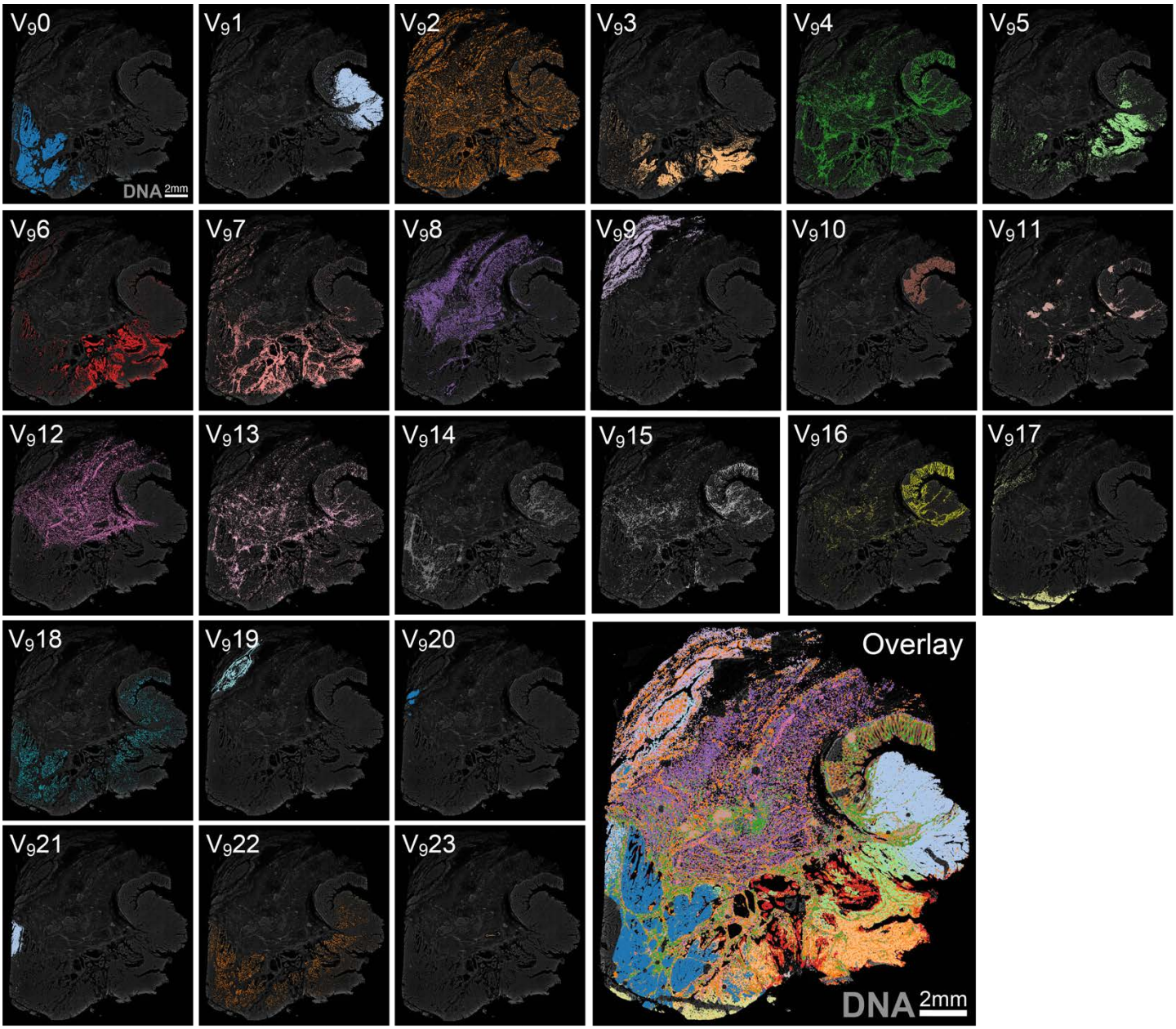


Extended Data Fig. 4 | VAE-based classification of human CRC at single-cell resolution. a, 3D UMAP embedding of 9x9µm VAE image patch encodings from CyCIF-1A colored by major cell lineage markers: keratin (epithelial, orange), αSMA (smooth muscle, pink), vimentin (stromal, blue), and CD45 (immune, green). **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. 6** for colormaps of all 21 immunomarkers. **c**, Jaccard similarity indices (defined as the

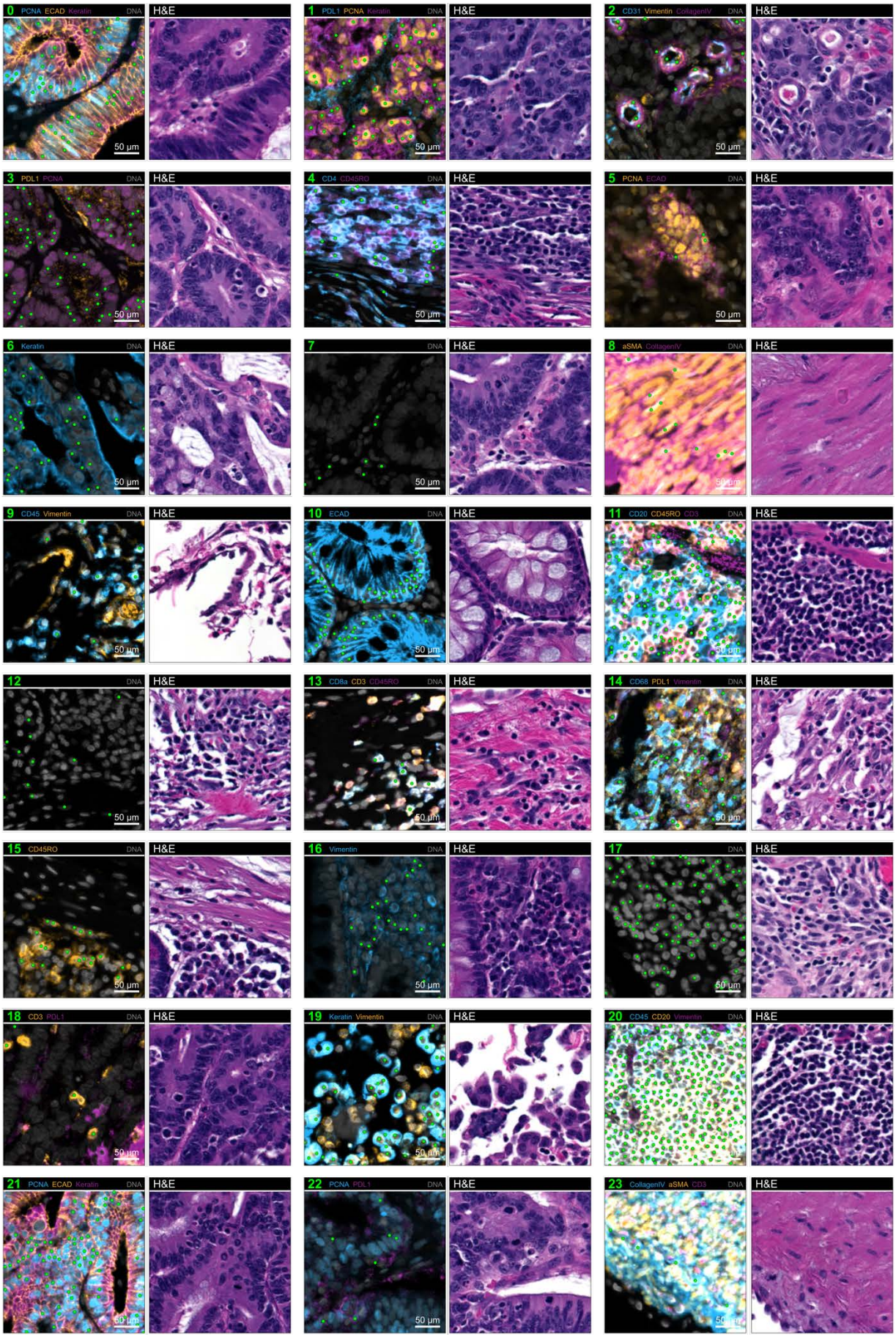
629 ratio of the intersection over the union of two sets) for cells in pairs of VAE and segmentation-based clusters.
630 Possible values range from 0 (no overlap) to 1 (perfect overlap). **d**, Paired CyCIF (left) and H&E (right) fields-
631 of-view from CyCIF-1A showing the location of random examples of cells from various Leiden clusters (green
632 scatter points in CyCIF images). See **Supplementary Fig. 7** for views of all 24 clusters.



Supplementary Fig. 6 | Channel colormaps for 9x9µm VAE image patch encodings. UMAP embeddings of VAE image patch encodings from CyCIF-1A 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) shown for reference.


0 PCNA ECAD Keratin DNA
H&E
1 PDL1 PCNA Keratin DNA
H&E
2 CD31 Vimentin CollagenIV DNA
H&E
3 PDL1 PCNA DNA
H&E
4 CD4 CD45RO DNA
H&E
5 PCNA ECAD DNA
H&E
6 Keratin DNA
H&E
7 DNA
H&E
8 aSMA CollagenIV DNA
H&E
9 CD45 Vimentin DNA
H&E
10 ECAD DNA
H&E
11 CD20 CD45RO CD3 DNA
H&E
12 DNA
H&E
13 CD8a CD3 CD45RO DNA
H&E
14 CD68 PDL1 Vimentin DNA
H&E
15 CD45RO DNA
H&E
16 Vimentin DNA
H&E
17 DNA
H&E
18 CD3 PDL1 DNA
H&E
19 Keratin Vimentin DNA
H&E
20 CD45 CD20 Vimentin DNA
H&E
21 PCNA ECAD Keratin DNA
H&E
22 PCNA PDL1 DNA
H&E
23 CollagenIV aSMA CD3 DNA
H&E
50 µm

641 **Supplementary Fig. 7 | Fields-of-view showing the location of 9x9µm VAE image patch Leiden clusters in**
642 **CyCIF-1A.** Paired CyCIF (left) and H&E (right) fields-of-view in CyCIF-1A showing the location of random
643 examples of cells from all 24 Leiden clusters (green scatter points in CyCIF images).

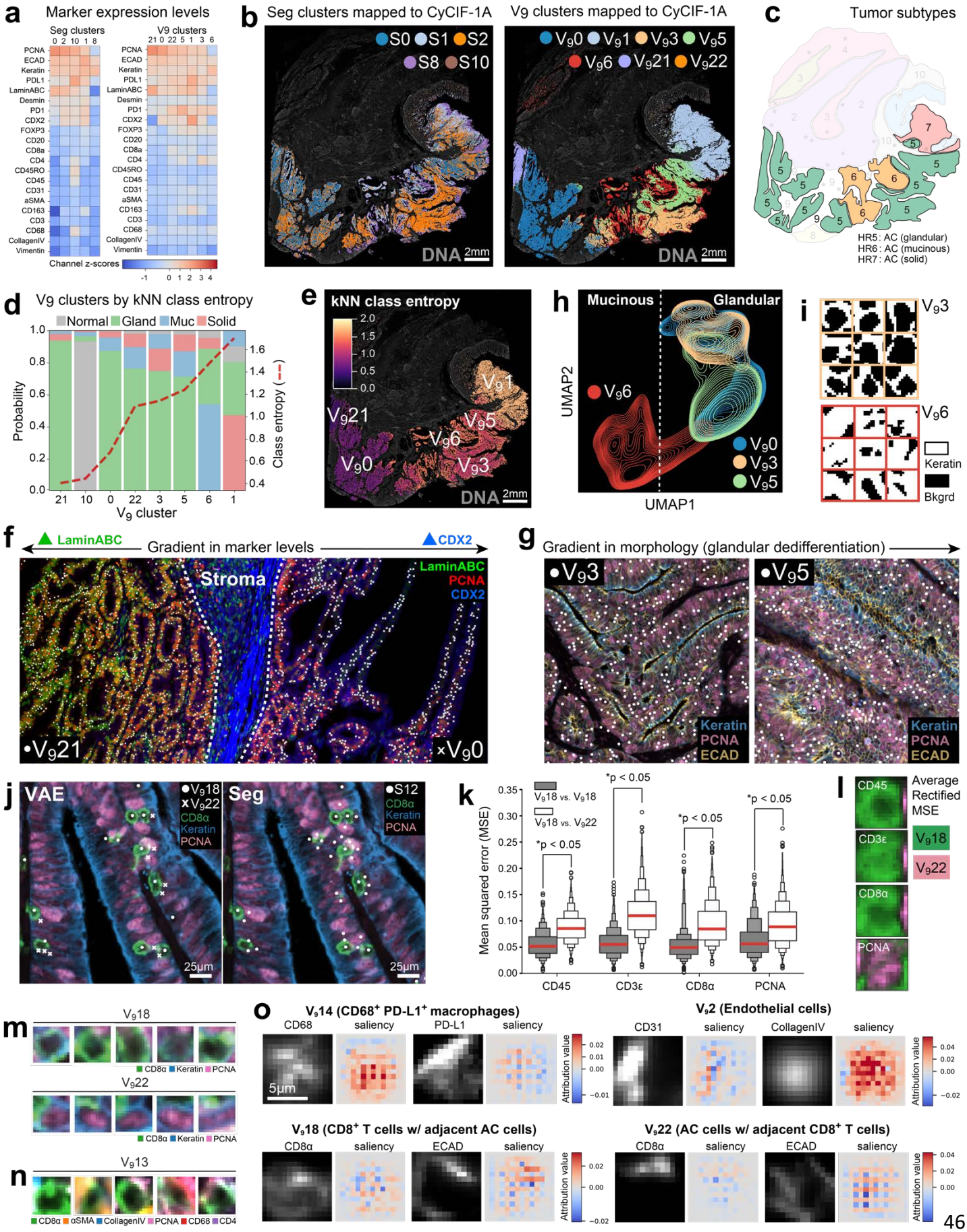
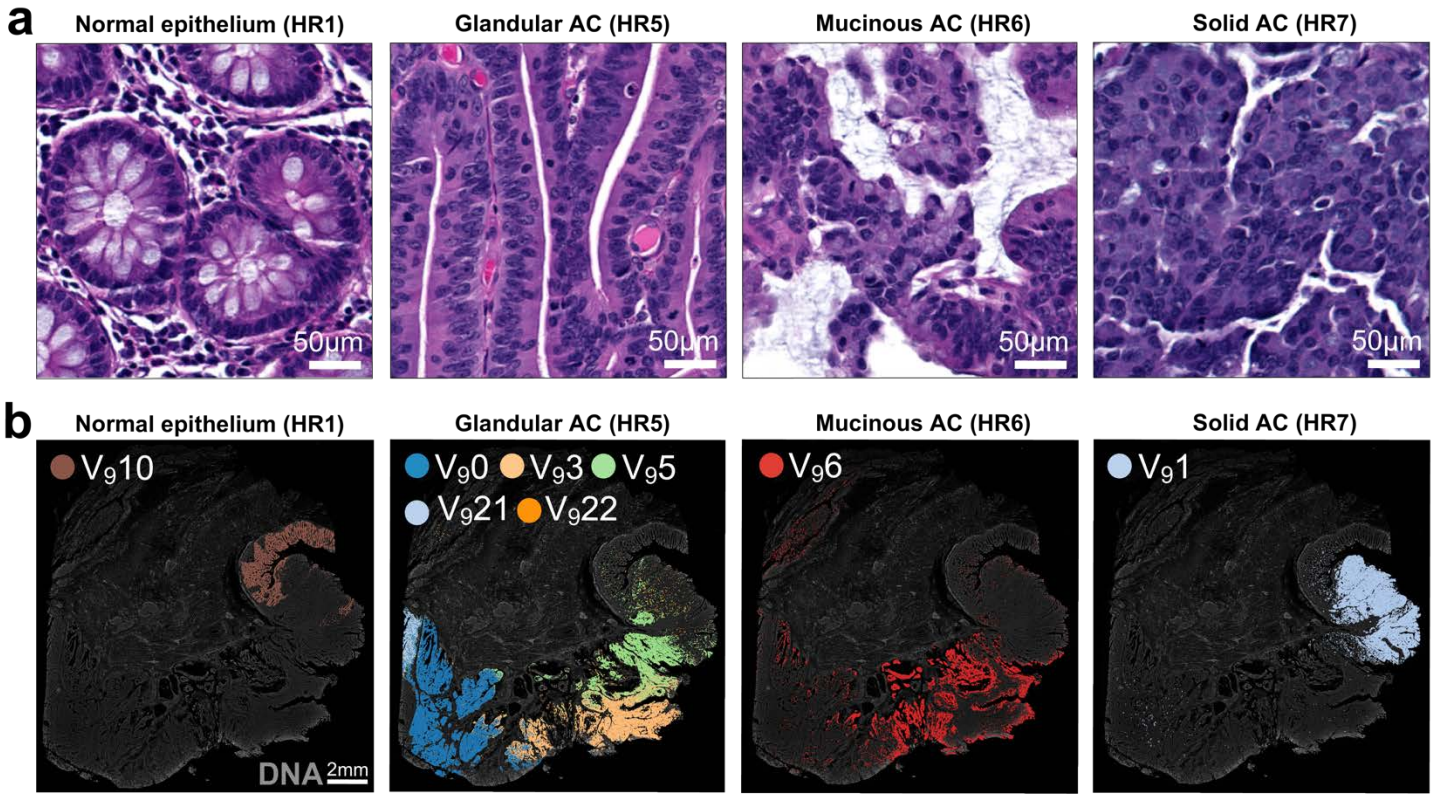
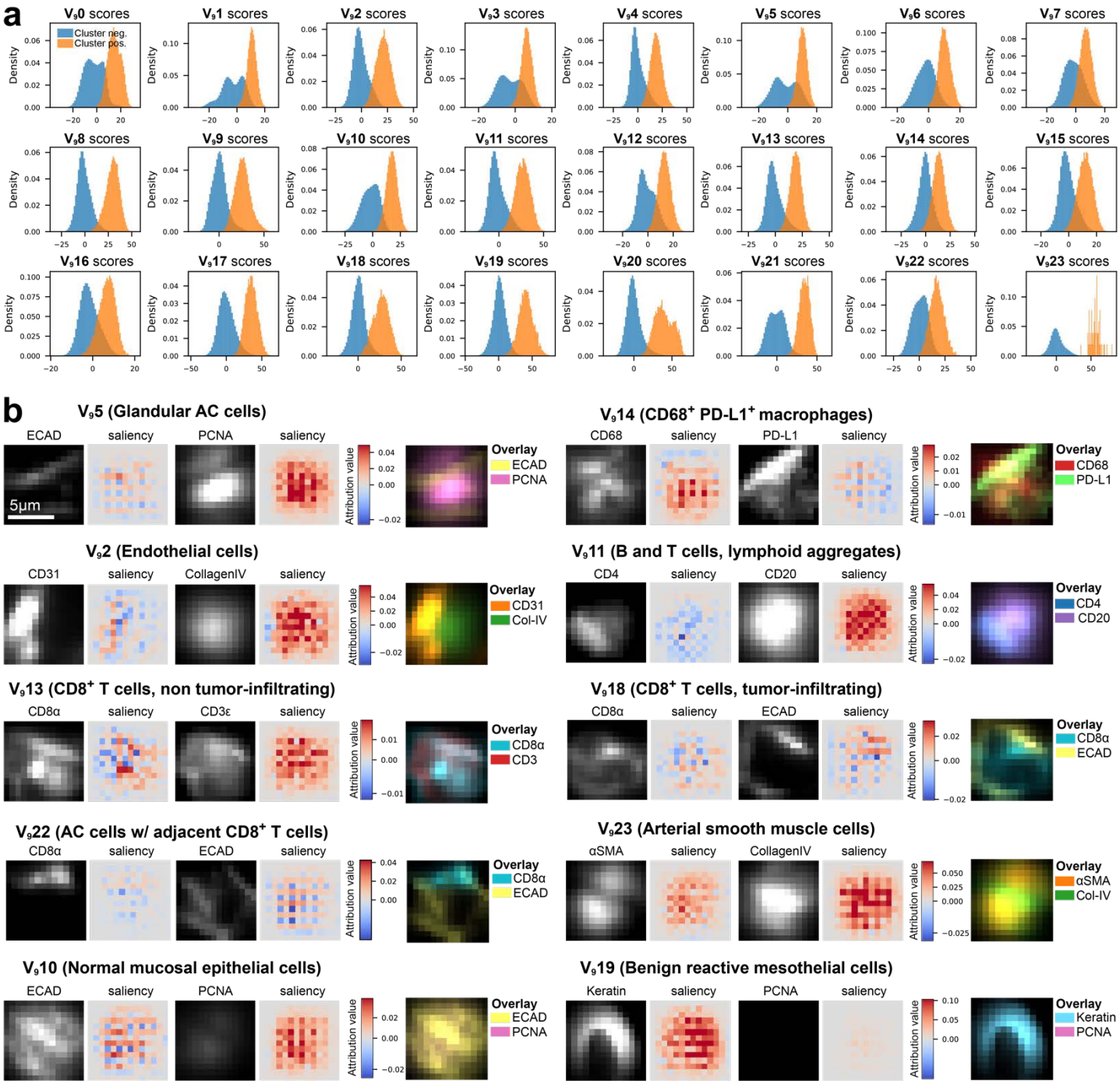

a
Marker expression levels
Seg clusters
V9 clusters
PCNA
ECAD
Keratin
PDL1
LaminABC
Desmin
PD1
CDX2
FOXP3
CD20
CD8a
CD4
CD45RO
CD45
CD31
aSMA
CD163
CD3
CD68
CollagenIV
Vimentin
Channel z-scores
b
Seg clusters mapped to CyCIF-1A
S0 S1 S2 S8 S10
V9 clusters mapped to CyCIF-1A
V90 V91 V93 V95 V96 V921 V922
DNA 2mm
c
Tumor subtypes
HR5 : AC (glandular)
HR6 : AC (mucinous)
HR7 : AC (solid)
d
V9 clusters by kNN class entropy
Normal Gland Muc Solid
Probability
Class entropy
V9 cluster
e
kNN class entropy
V91 V921 V95 V96 V90 V93
DNA 2mm
h
Mucinous
Glandular
UMAP2
UMAP1
V96
V90 V93 V95
i
V93
V96
Keratin
Bkgrd
f
LaminABC
Gradient in marker levels
CDX2
Stroma
LaminABC
PCNA
CDX2
V921
V90
g
Gradient in morphology (glandular dedifferentiation)
V93
V95
Keratin
PCNA
ECAD
j
VAE
V918
V922
CD8α
Keratin
PCNA
Seg
S12
CD8α
Keratin
PCNA
25μm
k
V918 vs. V918
V918 vs. V922
*p < 0.05
Mean squared error (MSE)
CD45
CD3ε
CD8α
PCNA
l
Average Rectified MSE
CD45
CD3ε
CD8α
PCNA
V918
V922
m
V918
CD8α Keratin PCNA
V922
CD8α Keratin PCNA
n
V913
CD8α αSMA CollagenIV PCNA CD68 CD4
o
V914 (CD68+ PD-L1+ macrophages)
CD68
saliency
PD-L1
saliency
5μm
Attribution value
V92 (Endothelial cells)
CD31
saliency
CollagenIV
saliency
V918 (CD8+ T cells w/ adjacent AC cells)
CD8α
saliency
ECAD
saliency
V922 (AC cells w/ adjacent CD8+ T cells)
CD8α
saliency
ECAD
saliency

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**, 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). **g,** Fields-of-view in CyCIF-1A of glandular AC clusters $V_9$3 (left) and $V_9$5 (right) as seen in channels keratin (blue), PCNA (pink), and ECAD (orange) revealing morphological difference in the density and organization of the glandular structures formed by these cells. **h**, UMAP embedding showing kernel density estimates for binarized VAE image patch encodings of the keratin channel from tumor-associated V_9 clusters 0, 3, 5, and 6. **i**, Example binary keratin image patches from V_9 clusters annotated as glandular ($V_9$3) and mucinous AC ($V_9$6) showing their morphological differences. **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. 6 | VAE-based learning identifies AC cells with distinct morphologies. 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.



Extended Data Fig. 7 | Concept saliency analysis for 9x9µm VAE image patch Leiden 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 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.

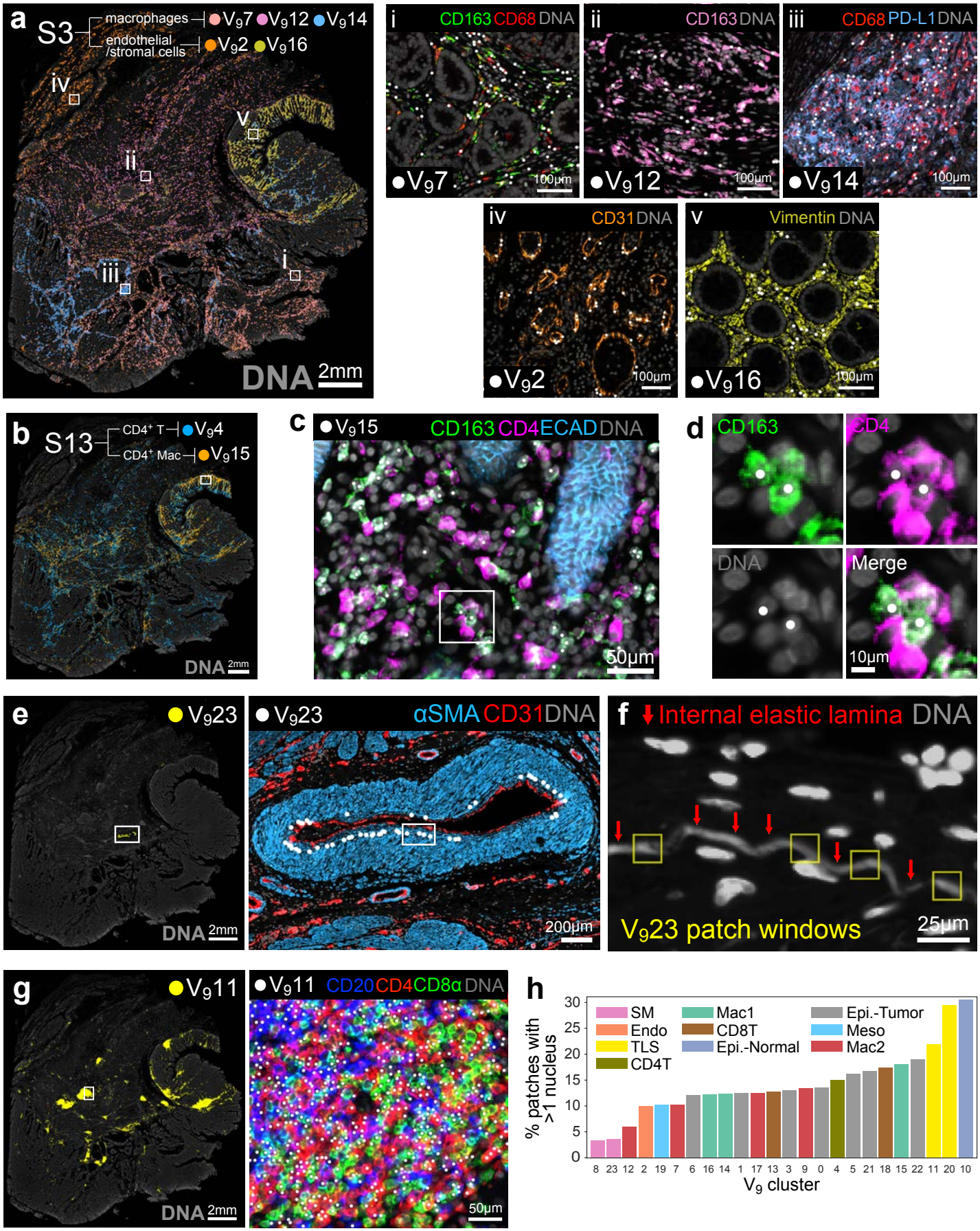
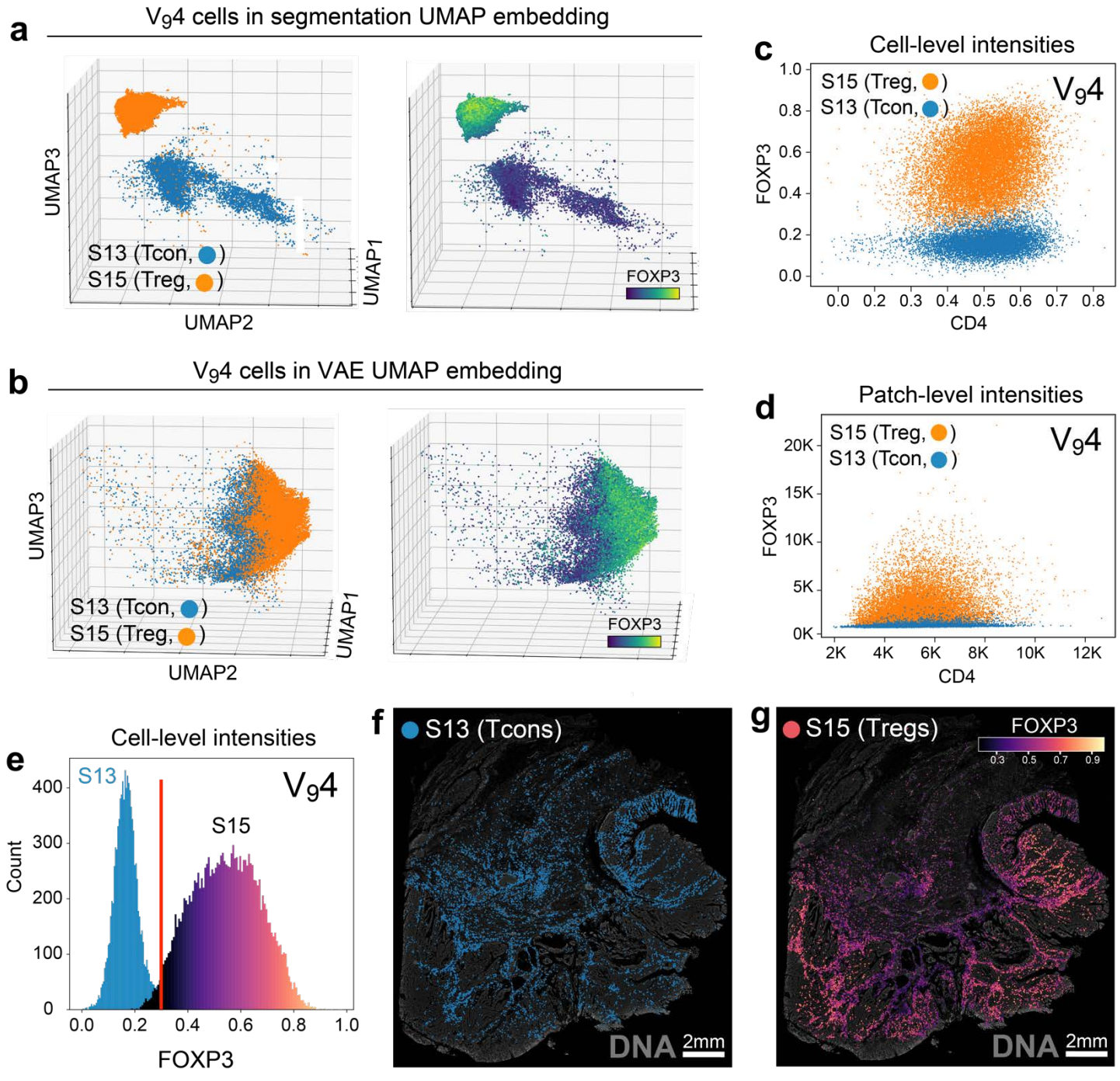
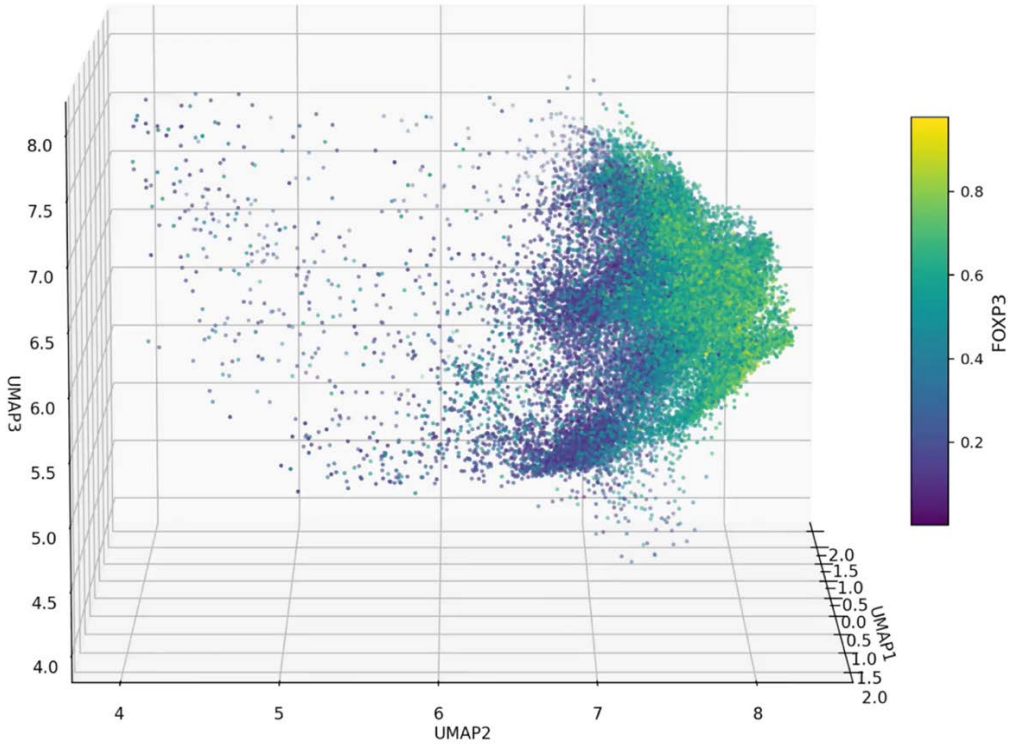

a
S3
macrophages
$V_9$7 $V_9$12 $V_9$14
endothelial /stromal cells
$V_9$2 $V_9$16
iv
v
ii
iii
i
DNA 2mm
i CD163 CD68 DNA
$V_9$7
100µm
ii CD163 DNA
$V_9$12
100µm
iii CD68 PD-L1 DNA
$V_9$14
100µm
iv CD31 DNA
$V_9$2
100µm
v Vimentin DNA
$V_9$16
100µm
b
S13
CD4+ T
$V_9$4
CD4+ Mac
$V_9$15
DNA 2mm
c
$V_9$15 CD163 CD4 ECAD DNA
50µm
d
CD163
CD4
DNA
Merge
10µm
e
$V_9$23
DNA 2mm
$V_9$23 αSMA CD31 DNA
200µm
f
Internal elastic lamina DNA
$V_9$23 patch windows
25µm
g
$V_9$11
DNA 2mm
$V_9$11 CD20 CD4 CD8α DNA
50µm
h
% patches with >1 nucleus
SM
Endo
TLS
CD4T
Mac1
CD8T
Epi.-Normal
Epi.-Tumor
Meso
Mac2
0 5 10 15 20 25 30
8 23 12 2 19 7 6 16 14 1 17 13 3 9 0 4 5 21 18 15 22 11 20 10
V_9 cluster

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

724 plot showing the distribution of $V_9$4 cells according to CD4 (x-axis) and FOXP3 (y-axis) intensity in 9x9µm
725 image patches colored by S13 (Tcons, blue) and S15 (Tregs, orange) showing that the FOXP3 axis no longer
726 exhibits a bimodal distribution. **e**, Distribution of $V_9$4 cells according to FOXP3 intensity colored by S13
727 (Tcons, blue) and S15 (Tregs, colormap) again revealing a bimodal distribution for the two cell populations
728 according to FOXP3 at the level of segmented cells. Red vertical line indicates a manually defined gate
729 distinguishing positive (right of gate) and negative (left of gate) cell populations. **f**, Location of gated S13 cells
730 (Tcons, FOXP3$^-$ cells in panel e) in CyCIF-1A. **g**, Location of gated S15 cells (Tregs, FOXP3$^+$ cells in panel e)
731 in CyCIF-1A revealing that Tregs occupy similar spatial niches as Tcons in CyCIF-1A and tend to express
732 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.

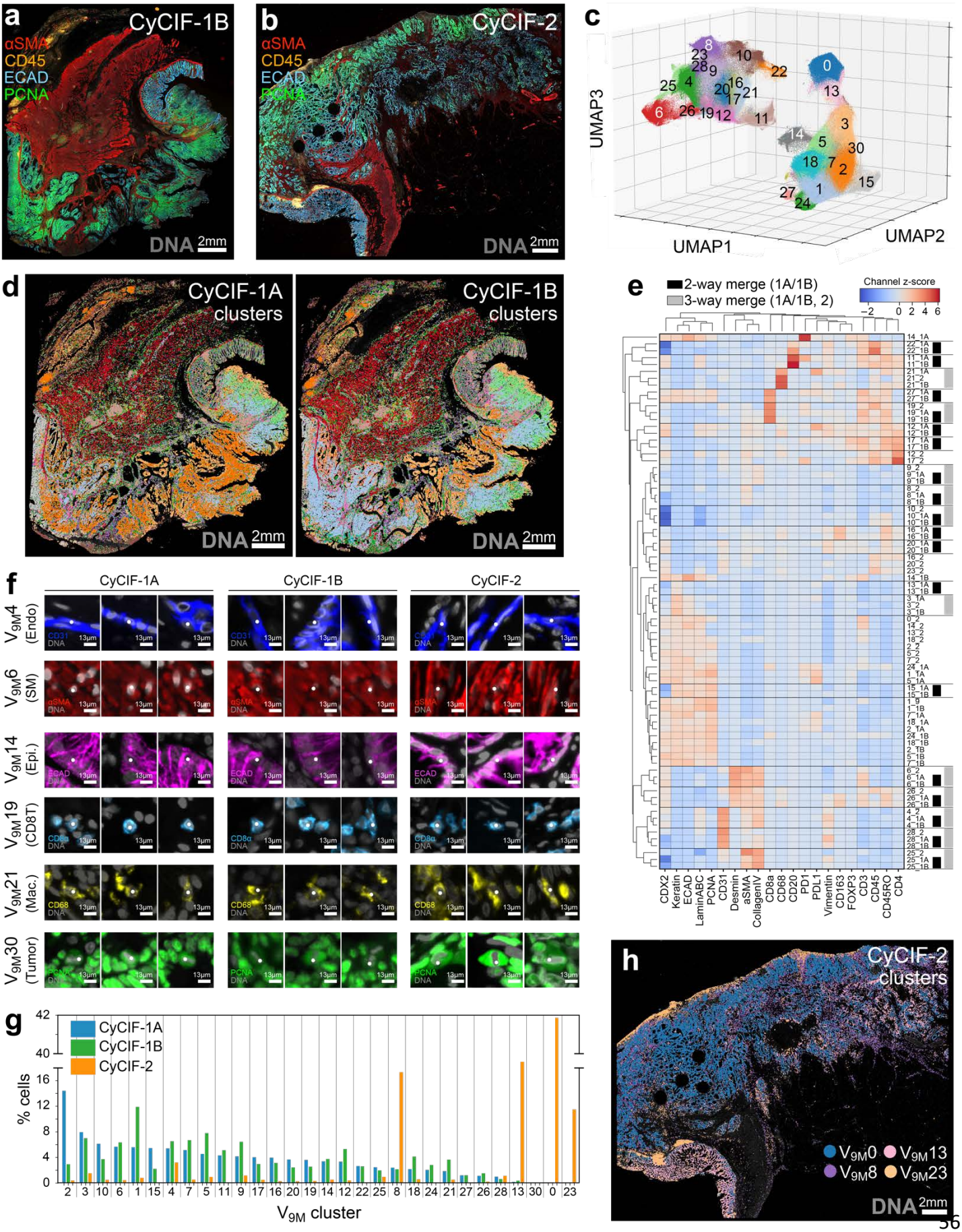
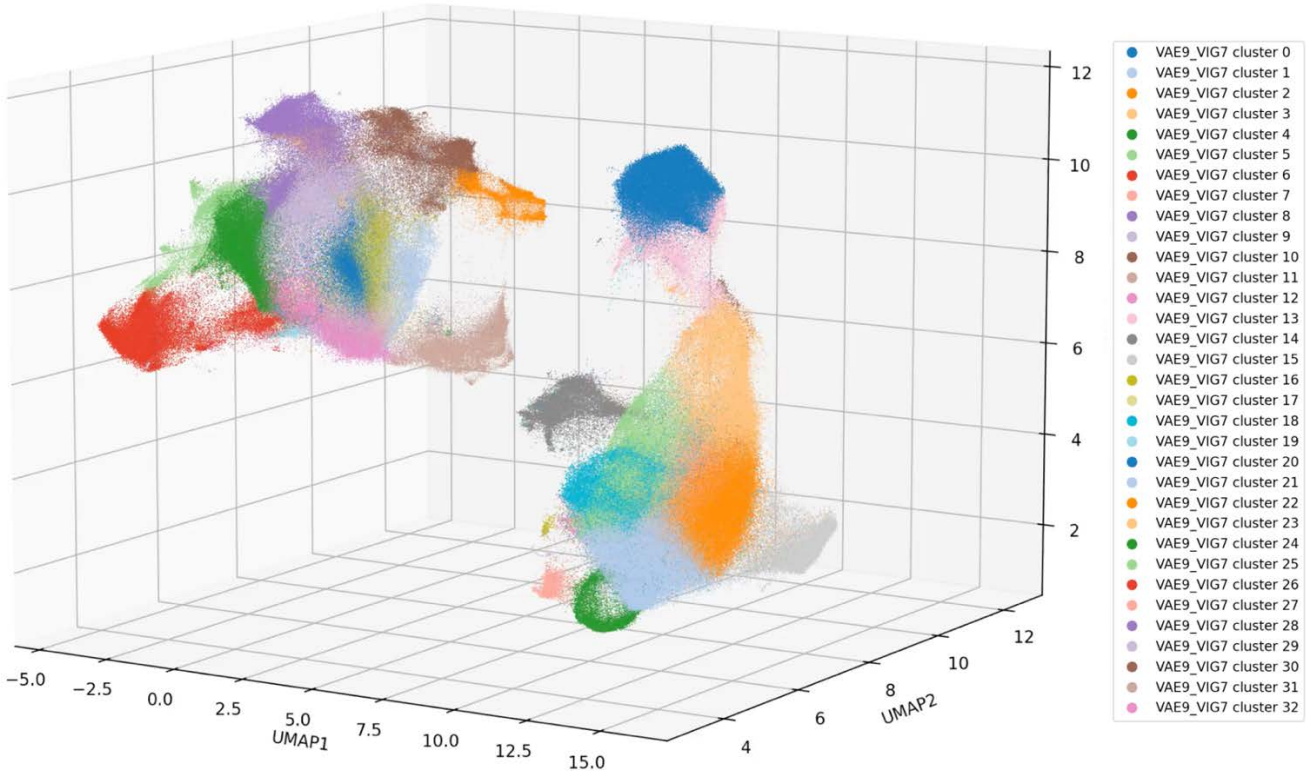

a
CyCIF-1B
αSMA
CD45
ECAD
PCNA
DNA 2mm
b
CyCIF-2
αSMA
CD45
ECAD
PCNA
DNA 2mm
c
UMAP3
UMAP1
UMAP2
d
CyCIF-1A clusters
CyCIF-1B clusters
DNA 2mm
e
2-way merge (1A/1B)
3-way merge (1A/1B, 2)
Channel z-score
CDX2
Keratin
ECAD
LaminABC
PCNA
CD31
Desmin
aSMA
CollagenIV
CD8a
CD68
CD20
PD1
PDL1
Vimentin
CD163
FOXP3
CD3
CD45
CD45RO
CD4
f
CyCIF-1A
CyCIF-1B
CyCIF-2
V9M4 (Endo)
V9M6 (SM)
V9M14 (Epi.)
V9M19 (CD8T)
V9M21 (Mac.)
V9M30 (Tumor)
g
CyCIF-1A
CyCIF-1B
CyCIF-2
% cells
V9M cluster
h
CyCIF-2 clusters
V9M0
V9M13
V9M8
V9M23
DNA 2mm

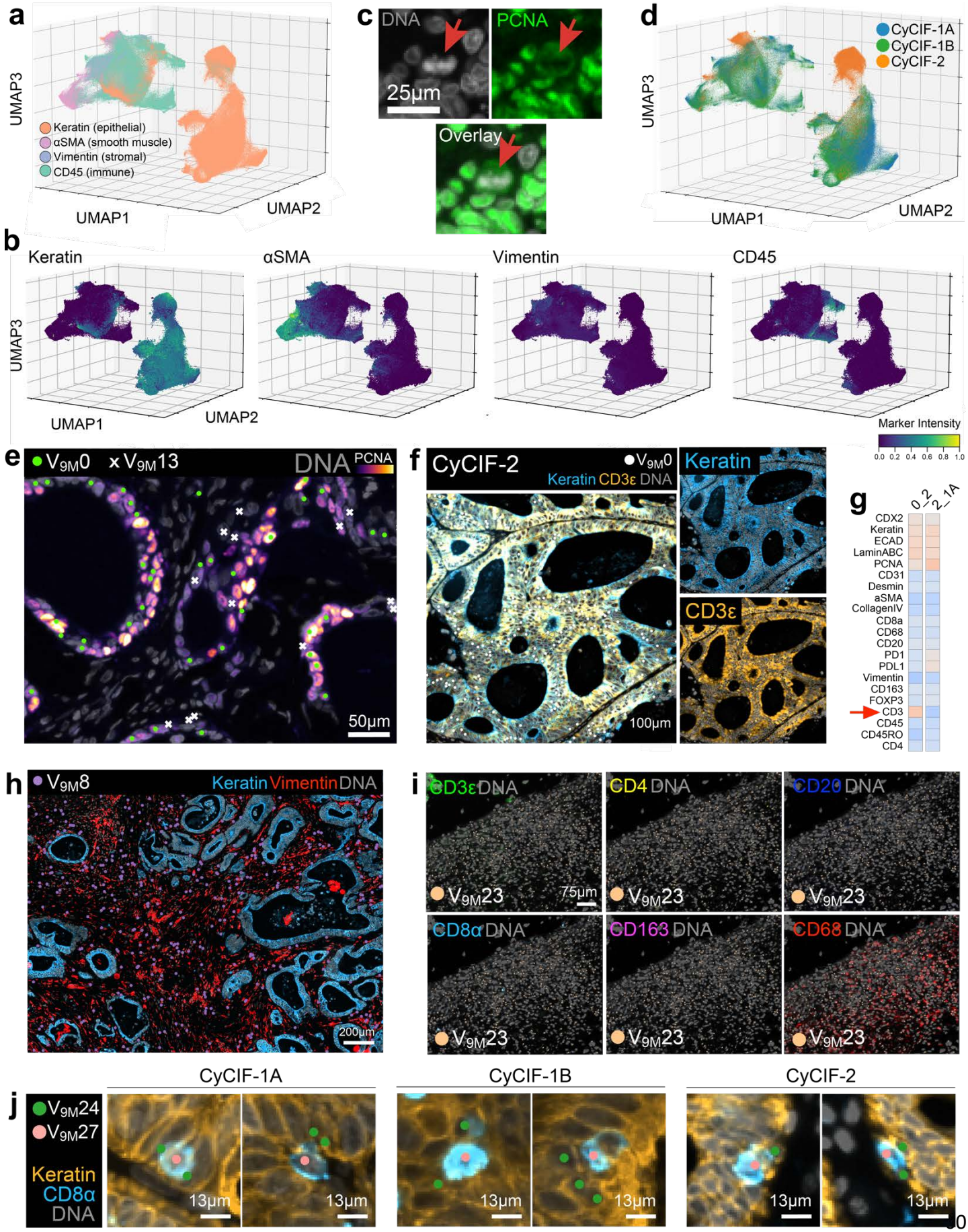
Fig. 5 | VAE analysis identifies matched cell states within and across patient specimens. a, CyCIF-1B: a second section from the same tissue block as CyCIF-1A (~20µm away in the Z-axis) imaged for the same CyCIF antibodies listed in (**Supplementary Table 2**). **b**, CyCIF-2: a tissue section from a second CRC patient imaged for the same CyCIF antibodies as CyCIF-1A and 1B. **c**, 3D UMAP embedding of 9x9µm VAE image patch encodings from CyCIF-1A, CyCIF-1B, and CyCIF-2 colored by Leiden cluster. Clusters 29, 31, and 32 were identified as microscopy artifacts and are not labeled; see **Supplementary Video 4** for 360° plot rotation. **d**, Location of V_{9M} clusters in CyCIF-1A (left) and 1B (right) revealing color-matched clusters in overlapping locations in serial sections of this tissue; DNA channel (gray) shown for reference. **e**, Hierarchical clustering of z-score-normalized median marker intensities of tissue-specific V_{9M} clusters showing that many share similar protein expression profiles. Black vertical bars indicate two-way dendrogram merges between CyCIF-1A and 1B. Gray vertical bars indicate three-way merges between all three specimens. Data are filtered to remove clusters associated with microscopy artifacts (V_{9M}29, 31, and 32) and clusters representing <0.1% of their respective tissue. **f**, V_{9M} clusters identified in all three specimens as seen in their characteristic marker channels with reference cells highlighted by white scatter points; DNA (gray) shown for reference. **g**, Grouped bar chart showing the percentage of cells in each specimen accounted for by each V_{9M} cluster sorted in descending rank-order according to cluster frequency in CyCIF-1A. Clusters identified as microscopy artifacts (V_{9M}29, 31, and 32) are not shown. **h**, Location of the four most abundant V_{9M} clusters in CyCIF-2; DNA channel (gray) shown 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

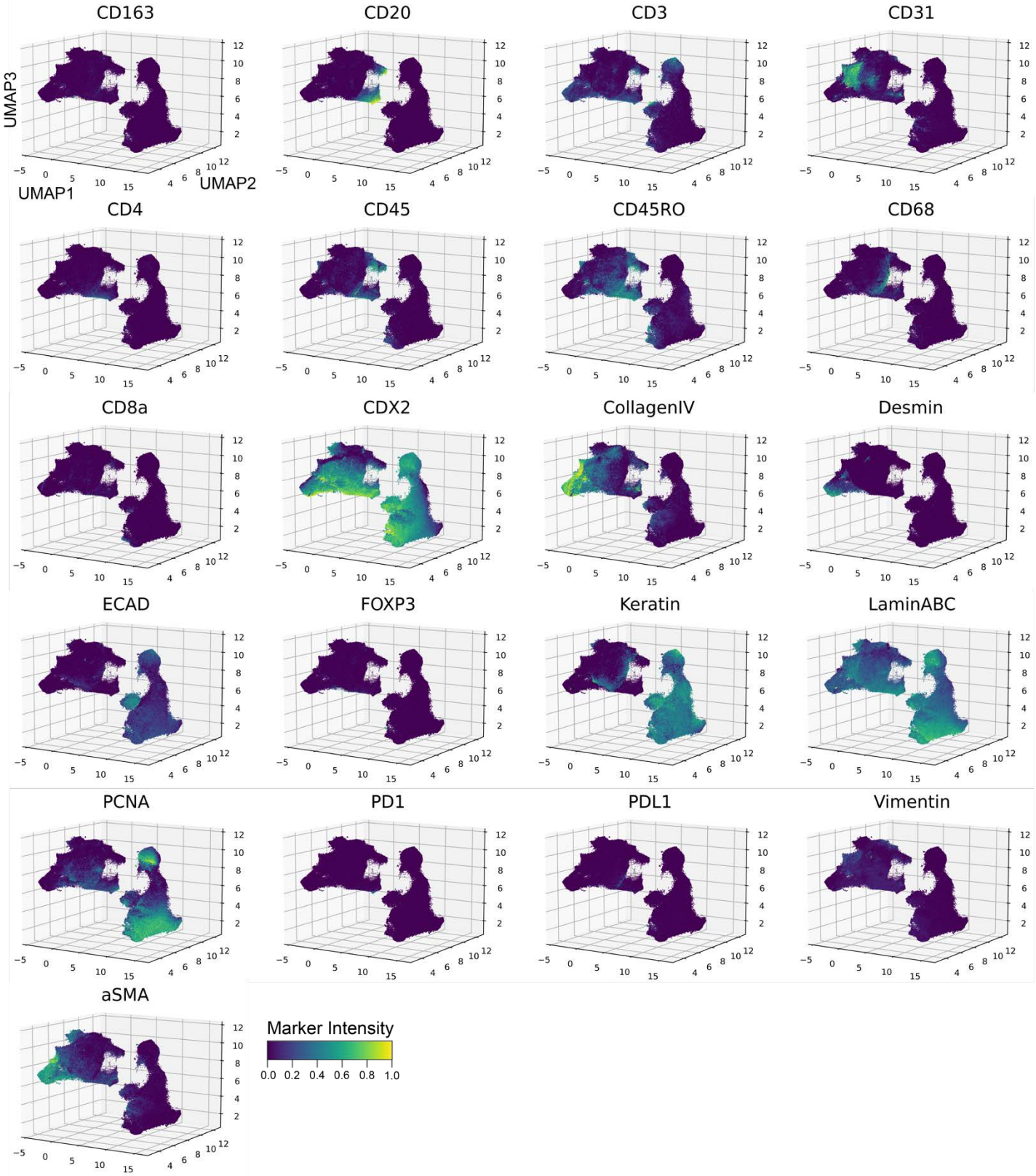
Supplementary Table 5 | 9x9µm VAE image patch Leiden cluster annotations (multi-tissue analysis). Column labeled "CyCIF-2 enriched" indicates whether the given cluster is predominantly found in that specimen.



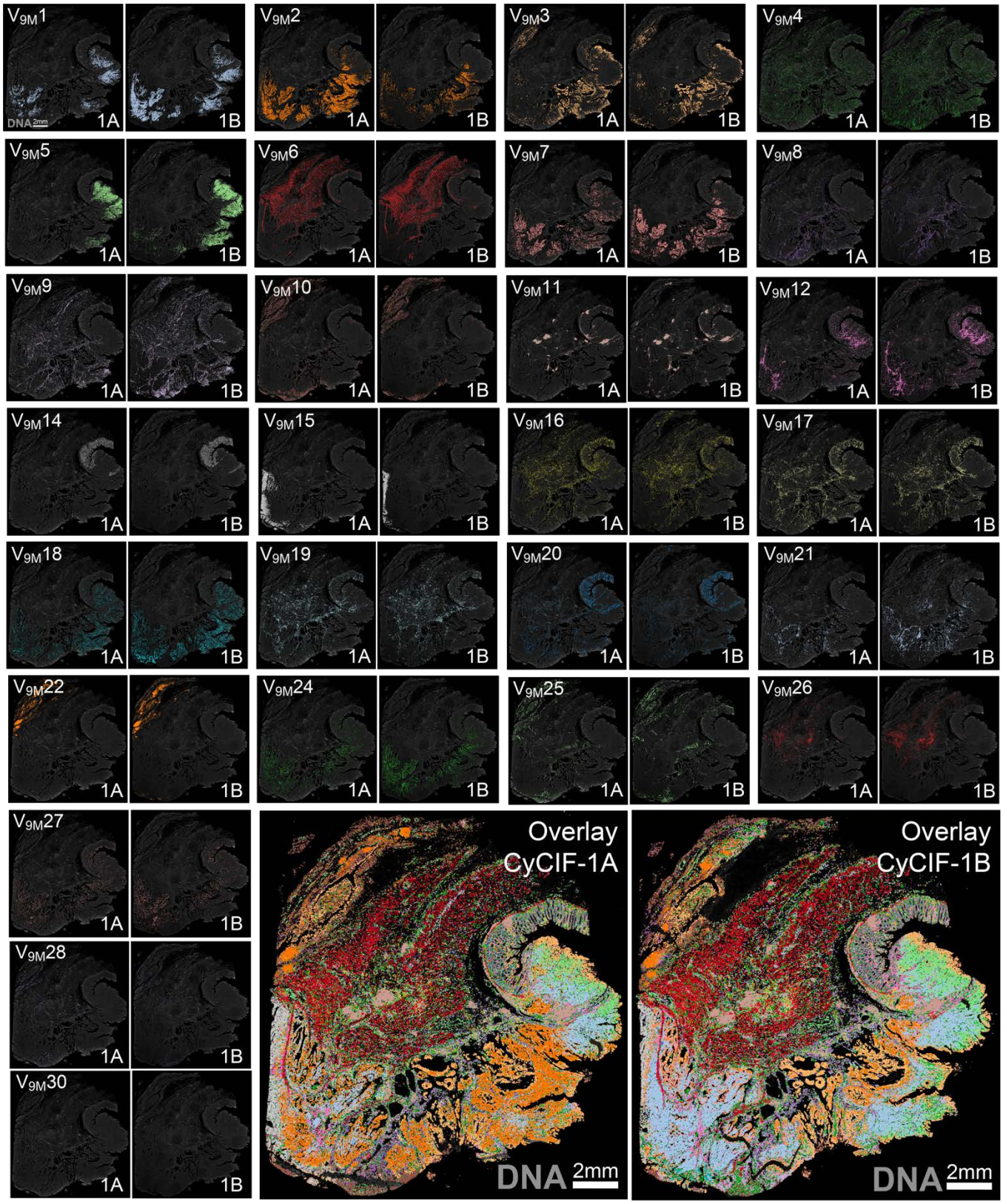
Supplementary Video 4 | 360° UMAP plot rotation showing 9x9µm VAE image patch encodings from CyCIF-1A, CyCIF-1B, and CyCIF-2 colored by Leiden cluster label.


a
UMAP3
Keratin (epithelial)
αSMA (smooth muscle)
Vimentin (stromal)
CD45 (immune)
UMAP1
UMAP2
c
DNA
PCNA
25µm
Overlay
d
CyCIF-1A
CyCIF-1B
CyCIF-2
UMAP3
UMAP1
UMAP2
b
Keratin
αSMA
Vimentin
CD45
UMAP3
UMAP1
UMAP2
Marker Intensity
0.0 0.2 0.4 0.6 0.8 1.0
e
V9M0 x V9M13
DNA PCNA
50µm
f
CyCIF-2
V9M0
Keratin CD3ε DNA
Keratin
CD3ε
100µm
g
0_2
2_1A
CDX2
Keratin
ECAD
LaminABC
PCNA
CD31
Desmin
aSMA
CollagenIV
CD8a
CD68
CD20
PD1
PDL1
Vimentin
CD163
FOXP3
CD3
CD45
CD45RO
CD4
h
V9M8
Keratin Vimentin DNA
200µm
i
CD3ε DNA
CD4 DNA
CD20 DNA
V9M23
75µm
CD8α DNA
CD163 DNA
CD68 DNA
j
CyCIF-1A
CyCIF-1B
CyCIF-2
V9M24
V9M27
Keratin
CD8α
DNA
13µm

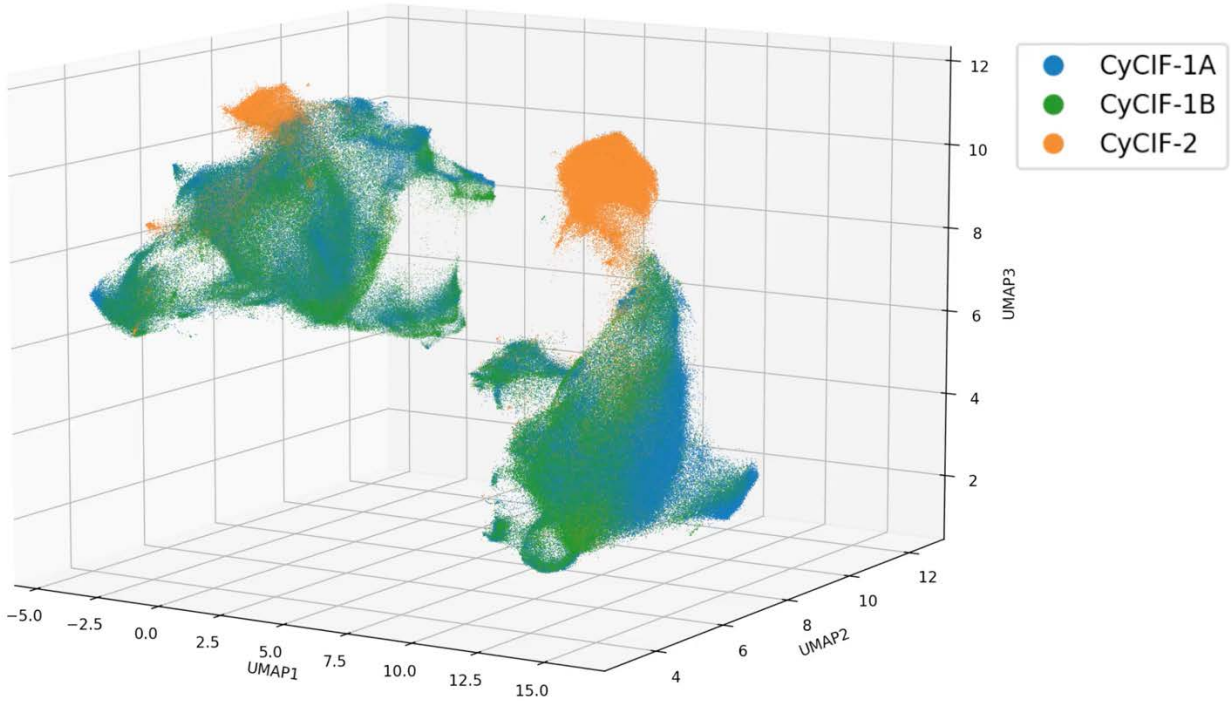
Extended Data Fig. 9 | VAE analysis identifies matched cell states within and across patient specimens. a, 3D UMAP embedding of 9x9µm VAE image patch encodings from CyCIF-1A, CyCIF-1B, and CyCIF-2 colored by major cell lineage markers: keratin (epithelial, orange), αSMA (smooth muscle, pink), vimentin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



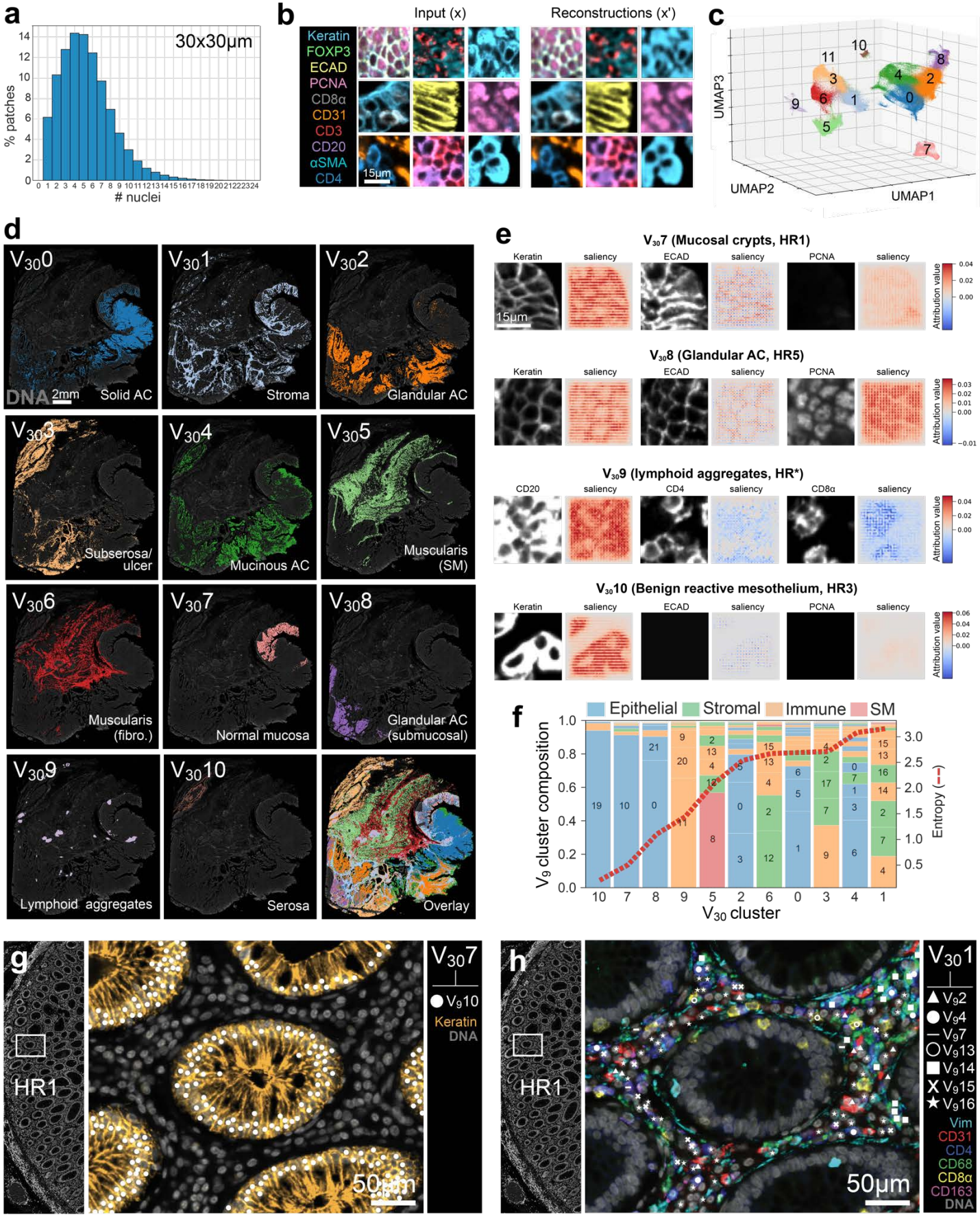
Supplementary Fig. 8 | Channel colormaps for 9x9µm VAE image patch encodings from multi-tissue analysis. UMAP embeddings of VAE image patch encodings from CyCIF-1A, CyCIF-1B, and CyCIF-2 colored by pixel-level median marker intensity for 21 immunomarkers.



Supplementary Fig. 9 | 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) shown for reference.

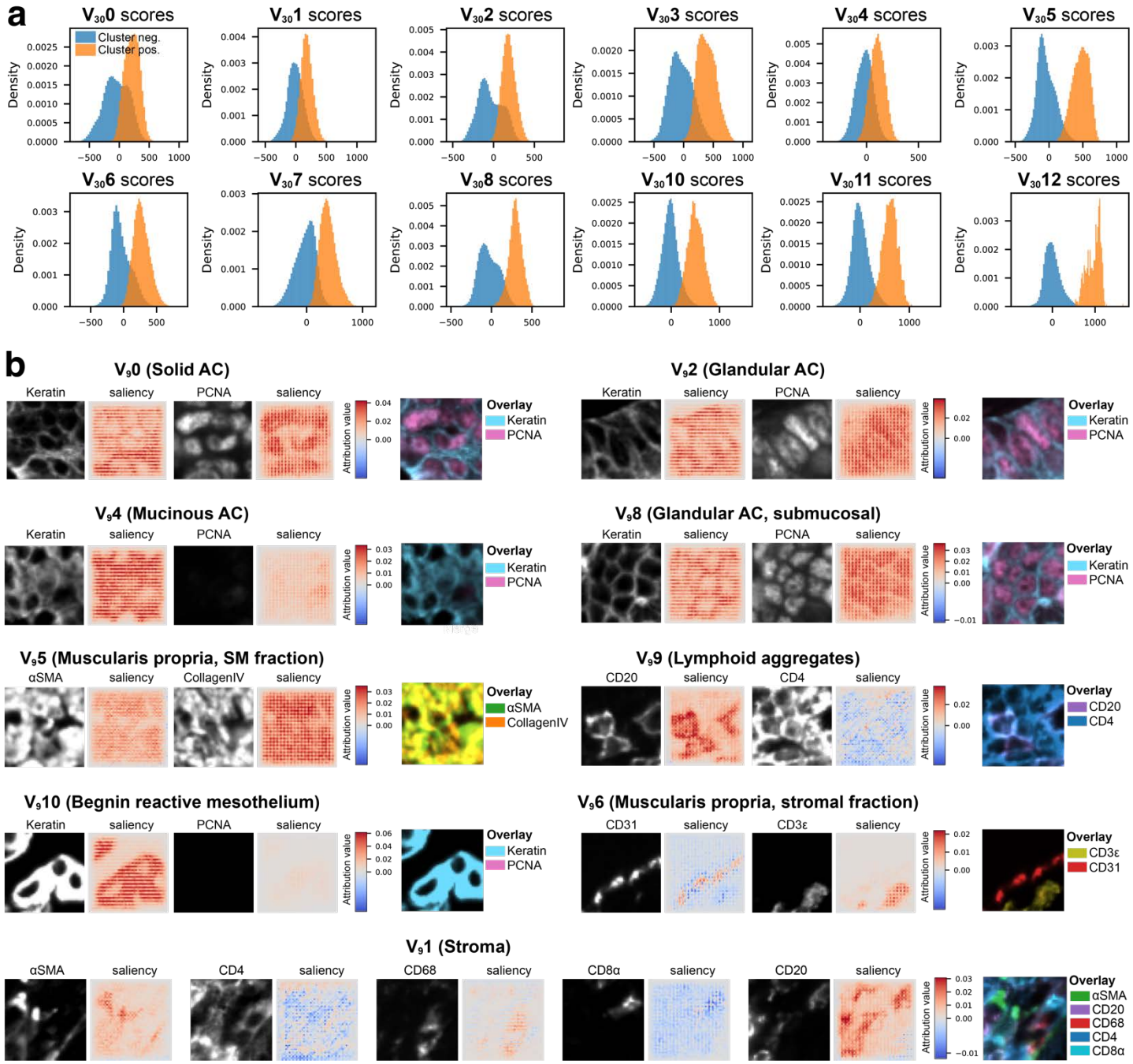


Supplementary Video 5 | 360° UMAP plot rotation showing 9x9µm VAE image patch encodings from CyCIF-1A, CyCIF-1B, and CyCIF-2 colored by tissue specimen.


a
30x30µm
% patches
# nuclei
b
Input (x)
Reconstructions (x')
Keratin
FOXP3
ECAD
PCNA
CD8α
CD31
CD3
CD20
αSMA
CD4
15µm
c
UMAP3
UMAP2
UMAP1
d
V30 0
V30 1
V30 2
V30 3
V30 4
V30 5
V30 6
V30 7
V30 8
V30 9
V30 10
DNA
2mm
Solid AC
Stroma
Glandular AC
Subserosa/ ulcer
Mucinous AC
Muscularis (SM)
Muscularis (fibro.)
Normal mucosa
Glandular AC (submucosal)
Lymphoid aggregates
Serosa
Overlay
e
V30 7 (Mucosal crypts, HR1)
V30 8 (Glandular AC, HR5)
V30 9 (lymphoid aggregates, HR*)
V30 10 (Benign reactive mesothelium, HR3)
Keratin
saliency
ECAD
PCNA
CD20
CD4
CD8α
Attribution value
15µm
f
Epithelial
Stromal
Immune
SM
V9 cluster composition
Entropy
V30 cluster
g
HR1
50µm
V30 7
V9 10
Keratin
DNA
h
HR1
50µm
V30 1
V9 2
V9 4
V9 7
V9 13
V9 14
V9 15
V9 16
Vim
CD31
CD4
CD68
CD8α
CD163
DNA

795

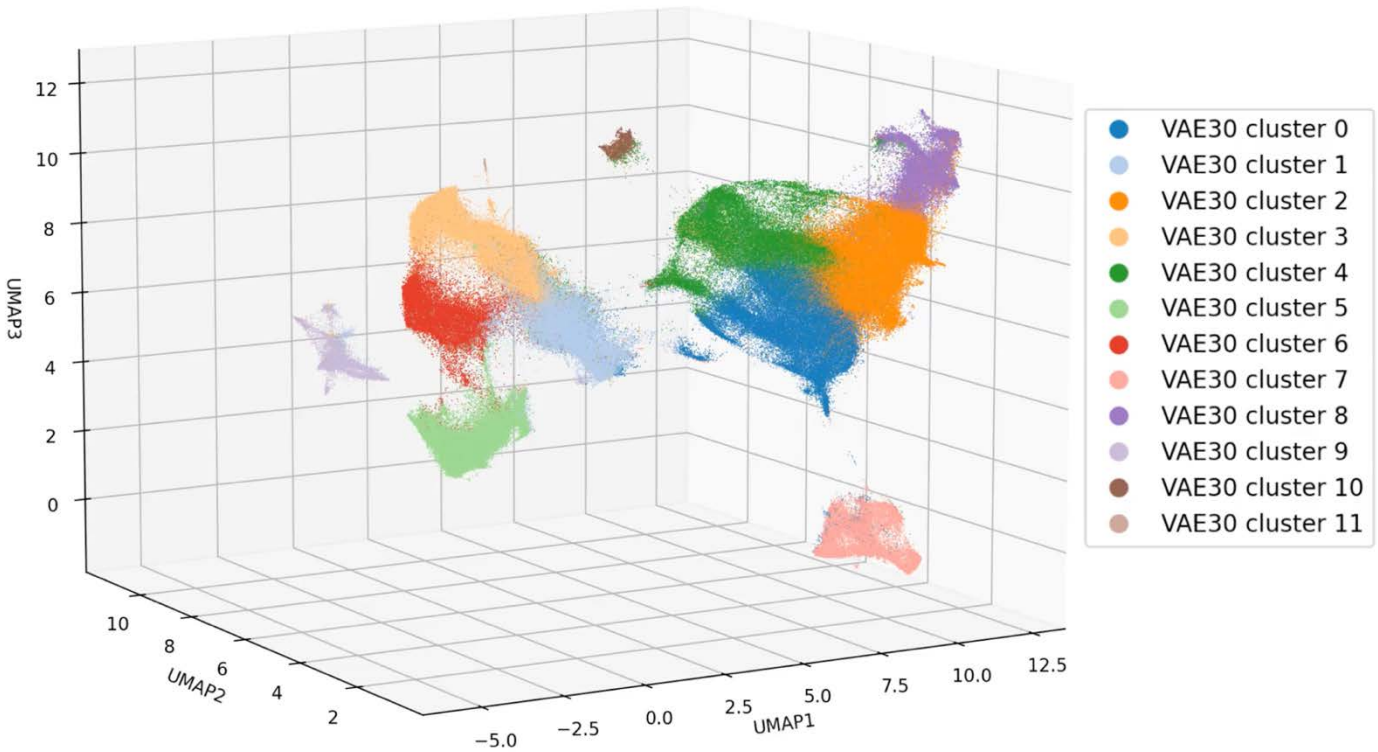
Fig. 6 | VAE classification of multi-cellular image patches identifies recurrent tissue motifs. a, Histogram showing the percentage of 30x30µm image patches containing different numbers of cell nuclei (median = 5, Q1 = 3, Q3 = 7, IQR = 4). **b**, Left: Example 30x30µm multi-channel image patches extracted from CyCIF-1A and used as input into the VAE encoding network. Right: Learned reconstructions of input images shown at left. **c**, 3D UMAP embedding of 30x30µm VAE image patch encodings from CyCIF-1A colored by Leiden cluster; see **Supplementary Video 6** for 360° plot rotation. **d**, Locations of V_{30} clusters in CyCIF-1A corresponding to histological regions observed in the H&E image (**Fig. 1b,c**); DNA channel (gray) shown for reference. **e**, Grayscale channels (left images) and corresponding concept saliency maps (right images) for four V_{30} clusters showing characteristic channels. **f**, Stacked bar chart showing the composition of reference cells at the center of V_{30} cluster image patches (x-axis) according to their V_9 cluster labels (left y-axis) and their cell type heterogeneity as measured by information entropy (right y-axis, dashed red curve). **g**, Location of reference cells in $V_{30}7$ (a low-entropy cluster) from CyCIF-1A composed of $V_9 10$ cells as seen in the keratin (orange) channel, indicating that this V_{30} cluster represents healthy mucosal crypts consisting of normal epithelial cells ($V_9 10$). **h**, Location of reference cells in $V_{30}1$ (a high-entropy cluster) from CyCIF-1A composed of multiple V_9 cell types, including fibroblasts (vimentin, cyan), endothelial cells (CD31, red), helper T cells (CD4, blue), M1-like macrophages (CD68, green), cytotoxic T cells (CD8α, yellow), and M2-like macrophages (CD163, 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.

825 **MATERIALS AND METHODS**

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

827 The colorectal cancer (CRC) specimens analyzed in this paper and referred to as "CyCIF-1A" (HTAN
828 biospecimen ID: HTA13_1_101), "CyCIF-1B" (HTAN biospecimen ID: HTA13_1_106), and "H&E" (HTAN
829 biospecimen ID: HTA13_1_100) consist of 5µm-thick, formalin-fixed paraffin-embedded (FFPE) tissue
830 sections, each derived from the same tumor block (referred to as CRC1). These specimens were acquired
831 through the Cooperative Human Tissue Network (CHTN) and were first described in a prior study supported
832 by the Human Tumor Atlas Network (HTAN)[23]. The tumor originated from a right-sided, poorly differentiated
833 stage IIIB colorectal adenocarcinoma (pT3N1bM0) of the cecum, taken from a 69-year-old Caucasian male
834 with microsatellite instability (MSI-H) and a BRAFV600E (c.1799T>A) mutation. The CRC specimen referred
835 to as "CyCIF-2" (HTAN ID: HTA7_922_2) was obtained from Brigham and Women's Hospital (BWH) in
836 compliance with all relevant ethical regulations established by the BWH Institutional Review Board (IRB).
837 This sample, also analyzed as part of the same HTAN study[23], represents a stage IVA adenocarcinoma
838 (pT4aN1bcM1a) from a 62-year-old Caucasian male with p53 and APC mutations. See (**Supplementary Table**
839 **1**) for an overview of biospecimen metadata.

840

841 **Image Processing**

842 CRC tissue used in this study underwent 10 rounds of four-color multiplex immunofluorescence imaging by
843 CyCIF[49]. DNA plus 3 immunomarkers were imaged at 20x magnification (0.65µm/pixel) with 2x2 pixel
844 binning at each cycle, for a total of 40 channels (**Supplementary Table 2**). Nineteen (19) channels were
845 withheld from analysis in the current study, as they represented either cycle-specific Hoechst DNA stain
846 (channels 1, 5, 9, 13, 17, 21, 25, 29, 33, 37), secondary antibodies alone (channels 2, 3, 4, 6, 7, 8), or channels
847 determined to be of insufficient quality. Raw microscopy tiles (in .rcpnl format) were stitched, registered,
848 segmented, and quantified using the MCMICRO image processing pipeline[21]. Cells in the image were
849 quantified according to three different cell segmentation mask objects: nucleus, cytoplasm, and plasma
850 membrane. To ensure the most accurate segmentation-based signal quantification, the choice of mask selection
851 was determined on a per immunomarker basis according to prior knowledge of the immunolabeling pattern for
852 that marker. The raw single-cell spatial feature table data consisted of 1,218,752 total segmentation instances,
853 which were reduced to 962,343 after QC with CyLinter software[24] to remove cells affected by visual and
854 image-processing artifacts (e.g., tissue folds, slide debris, illumination aberrations, mis-segmentation, etc.).

855

856 **Detailed Experimental Protocols**

The following protocols, available at Protocols.io, were used in the processing of tissue for this study:

1. FFPE Tissue Pre-treatment before t-CyCIF on Leica Bond RX V.2 (dx.doi.org/10.17504/protocols.io.bji2kkge)
2. Tissue Cyclic Immunofluorescence (t-CyCIF) V.2 (dx.doi.org/10.17504/protocols.io.bjiukkew)

Cellcutter

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

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

Image Patch Preprocessing

Cellcutter image patches were centered on the nucleus of individual cells identified during image processing with MCMICRO[21] and stored as Zarr arrays (https://zarr.readthedocs.io/en/stable/index.html) with dimensions [channels, cells, x_dim, y_dim]. Image patches contained the following channels (in order): CD3ε, CD45RO, keratin, αSMA, CD4, CD45, PD1, CD20, CD68, CD8α, CD163, FOXP3, PD-L1, ECAD, vimentin, CDX2, laminABC, desmin, CD31, PCNA, collagenIV. Pixel intensity values for each channel were $\log_{10}$-transformed. Pixels dimmer than the 0.17^{th} and brighter than the 99.99^{th} 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/). This bucket can be accessed using Cyberduck or other graphical S3 tools, or AWS Command Line tools.

Detailed instructions for accessing the data can be found here: https://zenodo.org/records/10223574. Primary imaging data for all four tissues is also freely 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). Access to supplementary videos and image galleries can be found at the dedicated Sage Synapse Wiki page for this paper: https://www.synapse.org/Synapse:syn53216852/wiki/625664.

Code Availability Statement

Source code for the VAE 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 VAE 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. 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 Ludwig Cancer Research and the Ludwig Center at Harvard (P.K.S., S.S.), the Gray Foundation, and by NIH NCI grants U01-CA284207, and U2C-CA233262. S.S. is supported by the BWH President's Scholars Award. We acknowledge J. Tefft for superb editorial support. Results shown in this study are in part based upon data generated by the Human Tumor Atlas Network (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

953 manuscript. A.S in currently an employee of Flagship Pioneering. The other authors declare no competing
954 interests.

REFERENCES

1. Method of the Year 2024: spatial proteomics. *Nat Methods* **21**, 2195–2196 (2024).
2. Schapiro, D. *et al.* MITI minimum information guidelines for highly multiplexed tissue images. *Nat Methods* **19**, 262–267 (2022).
3. Bai, Y. *et al.* Adjacent Cell Marker Lateral Spillover Compensation and Reinforcement for Multiplexed Images. *Front Immunol* **12**, 652631 (2021).
4. Yapp, C. *et al.* Highly Multiplexed 3D Profiling of Cell States and Immune Niches in Human Tumours. 2023.11.10.566670 Preprint at https://doi.org/10.1101/2023.11.10.566670 (2025).
5. Chevrier, S. *et al.* Compensation of Signal Spillover in Suspension and Imaging Mass Cytometry. *Cell Syst* **6**, 612-620.e5 (2018).
6. Amin, M. B. *et al.* The Eighth Edition AJCC Cancer Staging Manual: Continuing to build a bridge from a population-based to a more 'personalized' approach to cancer staging. *CA Cancer J Clin* **67**, 93–99 (2017).
7. Balaban, S. Deep learning and face recognition: the state of the art. in (eds. Kakadiaris, I. A., Kumar, A. & Scheirer, W. J.) 94570B (Baltimore, Maryland, United States, 2015). doi:10.1117/12.2181526.
8. Cui, M. & Zhang, D. Y. Artificial intelligence and computational pathology. *Laboratory Investigation* **101**, 412–422 (2021).
9. Markovits, E. *et al. A Novel Deep Learning Pipeline for Cell Typing and Phenotypic Marker Quantification in Multiplex Imaging*. http://biorxiv.org/lookup/doi/10.1101/2022.11.09.515776 (2022) doi:10.1101/2022.11.09.515776.
10. Amitay, Y. *et al.* CellSighter: a neural network to classify cells in highly multiplexed images. *Nat Commun* **14**, 4302 (2023).
11. Nirmal, A. J., Yapp, C., Santagata, S. & Sorger, P. K. Cell Spotter (CSPOT): A machine-learning approach to automated cell spotting and quantification of highly multiplexed tissue images. Preprint at https://doi.org/10.1101/2023.11.15.567196 (2023).
12. Zhu, B. *et al.* CellLENS enables cross-domain information fusion for enhanced cell population delineation in single-cell spatial omics data. *Nat Immunol* (2025) doi:10.1038/s41590-025-02163-1.
13. Nguyen, H., Vuong, H., Truong, T. & Pham, S. *IC-VAE: A Novel Deep Learning Framework for Interpreting Multiplexed Tissue Imaging Data*. http://biorxiv.org/lookup/doi/10.1101/2023.11.06.565771 (2023) doi:10.1101/2023.11.06.565771.
14. Spitzer, H., Berry, S., Donoghoe, M., Pelkmans, L. & Theis, F. J. Learning consistent subcellular landmarks to quantify changes in multiplexed protein maps. *Nat Methods* **20**, 1058–1069 (2023).
15. Tan, J. *et al.* Characterization of tumour heterogeneity through segmentation-free representation learning on multiplexed imaging data. *Nat Biomed Eng* **9**, 405–419 (2025).
16. Wang, X. (Julie) *et al.* Generalized cell phenotyping for spatial proteomics with language-informed vision models. Preprint at https://doi.org/10.1101/2024.11.02.621624 (2024).

17. Liu, C. C. *et al.* Robust phenotyping of highly multiplexed tissue imaging data using pixel-level clustering. *Nat Commun* **14**, 4618 (2023).

18. Kingma, D. P. & Welling, M. Auto-Encoding Variational Bayes. Preprint at http://arxiv.org/abs/1312.6114 (2014).

19. Kingma, D. P. & Welling, M. An Introduction to Variational Autoencoders. *FNT in Machine Learning* **12**, 307–392 (2019).

20. Brocki, L. & Chung, N. C. Concept Saliency Maps to Visualize Relevant Features in Deep Generative Models. Preprint at https://doi.org/10.48550/arXiv.1910.13140 (2019).

21. Schapiro, D. *et al.* MCMICRO: a scalable, modular image-processing pipeline for multiplexed tissue imaging. *Nat Methods* **19**, 311–315 (2022).

22. Lin, J.-R. *et al.* Highly multiplexed immunofluorescence imaging of human tissues and tumors using t-CyCIF and conventional optical microscopes. *eLife* **7**, e31657 (2018).

23. Lin, J.-R. *et al.* Multiplexed 3D atlas of state transitions and immune interaction in colorectal cancer. *Cell* **186**, 363-381.e19 (2023).

24. Baker, G. J. *et al.* Quality control for single-cell analysis of high-plex tissue profiles using CyLinter. *Nat Methods* (2024) doi:10.1038/s41592-024-02328-0.

25. Traag, V. A., Waltman, L. & van Eck, N. J. From Louvain to Leiden: guaranteeing well-connected communities. *Sci Rep* **9**, 5233 (2019).

26. McInnes, L., Healy, J. & Melville, J. UMAP: Uniform Manifold Approximation and Projection for Dimension Reduction. *arXiv:1802.03426 [cs, stat]* (2020).

27. Horn, J. L. A RATIONALE AND TEST FOR THE NUMBER OF FACTORS IN FACTOR ANALYSIS. *Psychometrika* **30**, 179–185 (1965).

28. 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).

29. McGinnis, C. S., Murrow, L. M. & Gartner, Z. J. DoubletFinder: Doublet Detection in Single-Cell RNA Sequencing Data Using Artificial Nearest Neighbors. *Cell Systems* **8**, 329-337.e4 (2019).

30. Hug, C. B. cellcutter. (2021).

31. Kim, B. *et al.* Why are Saliency Maps Noisy? Cause of and Solution to Noisy Saliency Maps. Preprint at https://doi.org/10.48550/arXiv.1902.04893 (2019).

32. Shaw, T. N. *et al.* Tissue-resident macrophages in the intestine are long lived and defined by Tim-4 and CD4 expression. *J Exp Med* **215**, 1507–1518 (2018).

33. Lin, J.-R. *et al.* High-plex immunofluorescence imaging and traditional histology of the same tissue section for discovering image-based biomarkers. *Nat Cancer* **4**, 1036–1052 (2023).

34. Salama, P. *et al.* Tumor-infiltrating FOXP3+ T regulatory cells show strong prognostic significance in colorectal cancer. *J Clin Oncol* **27**, 186–192 (2009).

35. Geographic Data Science with Python (Chapman & Hall/CRC Texts in Statistical Science): Rey, Sergio, Arribas-Bel, Dani, Wolf, Levi John: 9780367263119: Amazon.com: Books. https://www.amazon.com/Geographic-Science-Python-Chapman-Statistical/dp/0367263114.

36. Muhlich, J. L. *et al.* Stitching and registering highly multiplexed whole-slide images of tissues and tumors using ASHLAR. *Bioinformatics* **38**, 4613–4621 (2022).

37. Besson, S. *et al.* Bringing Open Data to Whole Slide Imaging. *Digit Pathol (2019)* **2019**, 3–10 (2019).

38. Ronneberger, O., Fischer, P. & Brox, T. U-Net: Convolutional Networks for Biomedical Image 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).

39. 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.

40. 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).

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

42. 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).

43. Blondel, V. D., Guillaume, J.-L., Lambiotte, R. & Lefebvre, E. Fast unfolding of communities in large networks. *J. Stat. Mech.* **2008**, P10008 (2008).

44. 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).

45. Kohonen, T. Self-organized formation of topologically correct feature maps. *Biol. Cybern.* **43**, 59–69 (1982).

46. McInnes, L., Healy, J. & Astels, S. hdbscan: Hierarchical density based clustering. *JOSS* **2**, 205 (2017).

47. van der Maaten et al. Visualizing high-dimensional data using t-SNE. *J Mach Learn Res* **9**, 2579–2605 (2008).

48. 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).

49. Lin, J.-R. *et al.* Highly multiplexed immunofluorescence imaging of human tissues and tumors using t-CyCIF and conventional optical microscopes. *Elife* **7**, (2018).