

# UnMICST: Deep learning with real augmentation for robust segmentation of highly multiplexed images of human tissues

**Clarence Yapp**

Harvard Medical School <https://orcid.org/0000-0003-1144-5710>

**Edward Novikov**

Harvard Medical School

**Won-Dong Jang**

Harvard University

**Yu-An Chen**

Harvard Medical School

**Marcelo Cicconet**

Harvard Medical School

**Zoltan Maliga**

Harvard Medical School

**Connor Jacobson**

Harvard Medical School <https://orcid.org/0000-0002-0299-3451>

**Donglai Wei**

Harvard University

**Sandro Santagata**

Brigham and Women's Hospital/ Harvard Medical School <https://orcid.org/0000-0002-7528-9668>

**Hanspeter Pfister**

Harvard

**Peter Sorger** (✉ [Peter\\_Sorger@hms.harvard.edu](mailto:Peter_Sorger@hms.harvard.edu))

Harvard Medical School <https://orcid.org/0000-0002-3364-1838>

---

## Article

**Keywords:** UnMICST, image processing, highly multiplexed images, microscopy

**Posted Date:** June 21st, 2021

**DOI:** <https://doi.org/10.21203/rs.3.rs-501324/v1>

**License:** © ⓘ This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

**Additional Declarations:** **Yes** there is potential Competing Interest. PKS is a member of the SAB or BOD member of Applied Biomath, RareCyte Inc., and Glencoe Software, which distributes a commercial version of the OMERO database; PKS is also a member of the NanoString SAB. In the last five years the Sorger lab has received research funding from Novartis and Merck. Sorger declares that none of these relationships have influenced the content of this manuscript. SS is a consultant for RareCyte Inc. The other authors declare no outside interests.

---

**Version of Record:** A version of this preprint was published at Communications Biology on November 18th, 2022. See the published version at <https://doi.org/10.1038/s42003-022-04076-3>.

1 **UnMICST: Deep learning with real augmentation for robust segmentation of highly**  
2 **multiplexed images of human tissues**

3

4 Clarence Yapp<sup>1,2</sup>, Edward Novikov<sup>1,3</sup>, Won-Dong Jang<sup>1,3</sup>, Yu-An Chen<sup>1</sup>, Marcelo Cicconet<sup>2</sup>, Zoltan  
5 Maliga<sup>1</sup>, Connor A. Jacobson<sup>1</sup>, Donglai Wei<sup>3</sup>, Sandro Santagata<sup>1,4,5</sup>, Hanspeter Pfister<sup>3</sup>, and Peter K.  
6 Sorger<sup>1,4,6</sup>

7

8 <sup>1</sup>Laboratory of Systems Pharmacology, Harvard Medical School, Boston, MA 02115

9 <sup>2</sup>Image and Data Analysis Core, Harvard Medical School, Boston, MA 02115

10 Human Tumor Atlas Network

11 <sup>3</sup>School of Engineering and Applied Sciences, Harvard University, Cambridge, MA, 02138

12 <sup>4</sup>Ludwig Center for Cancer Research at Harvard, Harvard Medical School, Boston, MA, 02115

13 <sup>5</sup>Department of Pathology, Brigham and Women's Hospital, Harvard Medical School, Boston, MA,  
14 02115

15 <sup>6</sup>Department of Systems Biology, Harvard Medical School, Boston, MA, 02115

16

17 **ORCID IDS**

18 Clarence Yapp [clarence@hms.harvard.edu](mailto:clarence@hms.harvard.edu) 0000-0003-1144-5710

19 Edward Novikov [edward\\_novikov@seas.harvard.edu](mailto:edward_novikov@seas.harvard.edu) 0000-0003-1476-5111

20 Won-Dong Jang [wjiang@seas.harvard.edu](mailto:wjiang@seas.harvard.edu) 0000-0002-5205-1024

21 Yu-An Chen [yu-an\\_chen@hms.harvard.edu](mailto:yu-an_chen@hms.harvard.edu) 0000-0001-7228-4696

22 Marcelo Cicconet [marcelo\\_cicconet@hms.harvard.edu](mailto:marcelo_cicconet@hms.harvard.edu) 0000-0003-2649-1509

23 Zoltan Maliga [zoltan\\_maliga@hms.harvard.edu](mailto:zoltan_maliga@hms.harvard.edu) 0000-0003-4209-7253

24 Connor A. Jacobson [connor\\_jacobson@hms.harvard.edu](mailto:connor_jacobson@hms.harvard.edu) 0000-0002-0299-3451

25 Donglai Wei [donglai@seas.harvard.edu](mailto:donglai@seas.harvard.edu) 0000-0002-2329-5484

26 Sandro Santagata [ssantagata@bics.bwh.harvard.edu](mailto:ssantagata@bics.bwh.harvard.edu) 0000-0002-7528-9668

27 Hanspeter Pfister [pfister@seas.harvard.edu](mailto:pfister@seas.harvard.edu) 0000-0002-3620-2582

28 Peter K. Sorger [peter\\_sorger@hms.harvard.edu](mailto:peter_sorger@hms.harvard.edu), 0000-0002-3364-1838

29

30

31

32 **ABSTRACT**

33 Newly developed technologies have made it feasible to routinely collect highly multiplexed  
34 (20-60 channel) images at subcellular resolution from human tissues for research and diagnostic  
35 purposes. Extracting single cell data from such images requires efficient and accurate image  
36 segmentation. This starts with identification of nuclei, a challenging problem in tissue imaging that  
37 has recently benefited from the use of deep learning. In this paper, we demonstrate two generally  
38 applicable approaches to improving segmentation accuracy for multiple human tissues. The first  
39 involves the use of “real augmentations” during training. These augmentations comprise defocused  
40 and saturated image data and improve model accuracy whereas computational augmentation  
41 (Gaussian blurring) does not. The second involves collection of nuclear envelope data to better  
42 identify nuclear outlines. The two approaches cumulatively and substantially improve segmentation  
43 with three different deep learning frameworks, yielding a set of highly accurate segmentation  
44 models. We speculate that the use of real augmentations may have applications in image  
45 processing outside of microscopy.

46

47 **INTRODUCTION**

48 Optical microscopy is a method of long-standing importance in biology that is changing  
49 rapidly due to the introduction of analytical methods based on machine learning (deep learning).  
50 Machine learning (ML) models are particularly important in the case of tissue images, whose  
51 complexity challenges traditional computer vision methods. Tissues are assemblies of multiple cell  
52 types that interact with each other and with structures such as basement membranes and  
53 connective tissue in defined geometric arrangements on length scales ranging from subcellular  
54 organelles to entire tissue (0.1 to  $10^4$   $\mu\text{m}$ ). Microscopy using Hematoxylin and Eosin (H&E)  
55 complemented by immunohistochemistry<sup>1</sup> has long played a primary role in the study of tissue  
56 architecture<sup>2,3</sup> and clinical histopathology remains the primary means by which diseases such as  
57 cancer are staged and managed clinically<sup>4</sup>. However, classical histology provides insufficient  
58 molecular information to precisely identify cell subtypes, study mechanisms of development, and  
59 characterize disease genes. In the past few years, several different methods for highly multiplexed  
60 tissue imaging have been developed (**Supplementary Table 1**). These yield subcellular resolution  
61 data on the abundance of 20-60 antigens, which is sufficient to subtype cells, measure cell states  
62 (quiescent, proliferating, dying, etc.) and interrogate cell signaling pathways. These methods differ  
63 in resolution, field of view, and multiplicity (“plex”), but all generate 2D images of tissue sections 5 –  
64 10  $\mu\text{m}$  thick. Developing effective image processing algorithms for these complex data is now the  
65 primary challenge in both research and clinical settings.

66 When multiplexed images are segmented and quantified, the resulting single cell data are a  
67 natural complement to single cell RNA Sequencing (scRNASeq) data, which have had a dramatic  
68 impact on our understanding of normal and diseased tissues<sup>5,6</sup>. Unlike dissociated RNASeq,

69 however, multiplex tissue imaging preserves spatial information. Single cell analysis of data  
70 acquired via imaging requires segmentation, a computer vision technique that assigns class labels  
71 to an image in a pixel-wise manner to subdivide it; this is followed by marker quantification on a per-  
72 cell or per-organelle basis. Extensive work has gone into the development of methods for  
73 segmenting metazoan cells grown in culture, but segmentation of tissue images is a substantially  
74 more complex challenge. Recently, machine learning methods have become standard, paralleling  
75 the widespread use of convolutional neural networks (CNNs) in image recognition, object detection,  
76 and synthetic image generation<sup>7</sup>. Architectures such as ResNet, VGG16, and more recently, UNet  
77 and Mask R-CNN<sup>8,9</sup> have gained attention for their ability to learn millions of parameters and  
78 generalize across datasets, as evidenced by excellent performance in a wide range of segmentation  
79 competitions, as well as in hackathon challenges<sup>10</sup> using publicly available image datasets<sup>11,12</sup>. One  
80 limitation of machine learning on tissue images is a lack of sufficient freely-available data with  
81 ground truth labelling. Experience with natural scene images<sup>8</sup> has proven that the acquisition of  
82 labels can be time consuming and rate limiting<sup>13</sup>.

83 In both cultured cells and tissues, the localization of nuclei is an optimal starting point for  
84 segmenting cells since most cell types have one nucleus (cells undergoing mitosis, muscle and liver  
85 cells and osteoclasts are important exceptions), and nuclear stains with high signal-to-background  
86 ratios are widely available. The nucleus is generally large relative to the resolution of wide-field  
87 fluorescence microscopes, making it easy to detect at multiple magnifications, and it often lies at the  
88 approximate center of a cell. There are possible advantages to using additional markers during  
89 image acquisition; for example, Schüffler *et al.*, (2015) used multiplexed IMC data and watershed  
90 methods for multi-channel segmentation. Methods based on random forests such as Ilastik and  
91 Weka<sup>15,16</sup> also exploit multiple channels for class-wise pixel classification via an ensemble of  
92 decision trees to assign pixel-wise class probabilities in an image. However, random forest models  
93 have significantly less capacity for learning than CNNs, which is a substantial disadvantage. The  
94 possibility of using CNNs with multi-channel data to enhance nuclei segmentation has not been  
95 widely explored.

96 The accuracy of segmentation algorithms is crucially dependent on the quality of the original  
97 images and the absence of artefacts. In practice however, images have unavoidable blur artefacts  
98 and saturated channels. Segmentation methods must compensate for these aberrations. The  
99 standard way to accomplish this is via image augmentation<sup>17</sup> which expands training sets by pre-  
100 processed images via random rotation, shearing, flipping, etc. This is designed to prevent  
101 algorithms from learning irrelevant aspects of an image, such as orientation. To date, focus artifacts  
102 have been tackled by using computed Gaussian blur to augment training data<sup>18-20</sup>.

103 In this paper, we investigate ways to maximize the accuracy of image segmentation in  
104 multiplexed tissue data by including different types of information in the images and by augmenting  
105 the training data. We generate a set of training and test data with ground-truth annotations via

106 human curation of multiple normal tissues and tumors, and use these data to score segmentation  
107 accuracy achieved on three deep learning algorithms: UNet, Mask R-CNN, and Pyramid Scene  
108 Parsing Network (PSPNet); we use this to generate a family of *Universal Models for Identifying Cells*  
109 *and Segmenting Tissue* (UnMICST). We identify two generalizable and easily implemented ways to  
110 improve segmentation accuracy. The first involves adding images of nuclear envelope staining  
111 (NES) to images of nuclear chromatin acquired using DNA-intercalating dyes. The second involves  
112 adding real augmentations, defined here as intentionally defocused and over-saturated images  
113 (collected from the same specimens), to the training data to make models more robust to the types  
114 of artefacts encountered in real tissue images. We find that augmentation with real data  
115 significantly outperforms conventional Gaussian blur augmentation, offering a statistically significant  
116 improvement in model robustness.

117

## 118 **RESULTS**

119 Semantic segmentation is a coarse-grained ML approach that assigns objects to distinct  
120 trained classes, while instance segmentation is fine grained and identifies individual instances of  
121 objects. We implemented and evaluated two semantic and one instance segmentation algorithms  
122 (UNet, PSPNet, and Mask R-CNN respectively). We trained each of these models (UnMICST-U,  
123 UnMICST-P, and UnMICST-M, respectively) on manually curated and labelled data from seven  
124 distinct tissue types. The accuracy of each model was assessed based on the number of correctly  
125 segmented cells relative to the total number using a variable intersection over union (IoU) threshold  
126 ranging from the least stringent, 0.55, to most stringent, 0.8. The IoU (the Jaccard Index) is  
127 calculated by measuring the overlap between the ground truth annotation and the prediction via a  
128 ratio of the intersection to the union of pixels in both masks. The greater the IoU, the higher the  
129 accuracy, with an ideal value of 1 (although this is very rarely achieved). Unlike the standard pixel  
130 accuracy metric (the fraction of pixels in an image that were correctly classified), IoU is not sensitive  
131 to class-imbalance.

132

### 133 **Data sets and ground truth annotation of nuclear boundaries**

134 It is well established that various cell types in tissues have nuclear morphologies that are  
135 different from the spherical and ellipsoidal shape observed in cultured cells<sup>21</sup>. Nuclear pleomorphism  
136 (variation in nuclear size and shape) is widely used in histopathology to grade cancers and has  
137 been automated for H&E images using classical machine vision methods<sup>22</sup>. We generated training,  
138 validation, and test datasets with manually annotated ground-truth of nuclei contours, centers, and  
139 background in images of a 120-core tissue microarray representing 30 distinct normal and diseased  
140 tissue types. Human inspection identified seven distinct nuclear morphologies ranging from mixtures  
141 of cells that were large vs. small, round cells vs. narrow, and densely and irregularly packed vs.  
142 organized in clusters. Because ground-truth annotation is a laborious process, we focused on

143 distinctive nuclei from a subset of morphologies in seven tissue types (lung adenocarcinoma, non-  
144 neoplastic small intestine, normal prostate, colon adenocarcinoma, glioblastoma, non-neoplastic  
145 ovary, and tonsil) from 12 cores representing a total of ~10,400 nuclei.

146

### 147 **Real augmentations increase model robustness to focus artefacts**

148 To study the impact of real and computed augmentations on the performance of  
149 segmentation methods, we trained models with different sets of data, involving both real and  
150 computed augmentations and then tested the data on images that were acquired in focus, out of  
151 focus or blurred using a Gaussian kernel. We assessed segmentation accuracy quantitatively based  
152 on IoU and qualitatively by visual inspection of predicted masks overlaid on image data. Real  
153 augmentation involved adding additional empirical, rather than computed, training data having the  
154 types of imperfections most commonly encountered in tissue. This was accomplished by positioning  
155 the focal plane 3  $\mu\text{m}$  above and below the specimen, resulting in de-focused images. A second set  
156 of images was collected at long exposure times, thereby saturating 70-80% of pixels. Because  
157 blurred and saturated images were collected sequentially without changing stage positions, it was  
158 possible to use the same set of ground truth annotations. For computed augmentations, we  
159 convolved a Gaussian kernel to the in-focus images with a range of standard deviations chosen to  
160 cover a broad spectrum of experimental cases (**Figure 1a**). In both scenarios, the resulting models  
161 were evaluated on a test set prepared in the same way as the training set.

162 In an initial set of studies, we found that models trained with data having Gaussian blur  
163 augmentation performed well on Gaussian blurred test data but when IoU was quantified for  
164 defocused and saturated test data, we found that the use of training data augmented with Gaussian  
165 blur improved accuracy only slightly relative to baseline models lacking augmentations (**Figure 1b**).  
166 In contrast, the use of training data supplemented with real augmentations increased the fraction of  
167 cells retained at an IoU threshold of 0.6 by 40-60%; statistically significant improvement was  
168 observed up to an IoU cutoff of 0.8 with all three learning frameworks (UnMICST-U, UnMICST-M,  
169 and UnMICST-P models). To perform a balanced comparison, we created two sets of training data  
170 having equal numbers of images. The first set contained the original data plus computed 90- and  
171 180-degree rotations, and the second set contained original data plus defocused data collected from  
172 above and below the specimen. Again, we found that models trained with real augmentations  
173 substantially outperformed rotationally augmented models when tested on defocused test data  
174 (**Figure. 1c**). Repeating this experiment with NES yielded the same outcome. Thus, training deep  
175 learning architectures with real augmentation generates models that outperform models with  
176 computed augmentation when data contain commonly encountered artefacts.

177

178

179

## 180 **Addition of NES improves segmentation accuracy**

181           When we stained our TMA panel (the Exemplar Microscopy Images of Tissues and Tumors  
182 (EMIT) TMA) we found that antibodies against lamin A and C (**Figure 2a**) (which are different splice  
183 forms of *LMNA* gene) stained approximately only half as many nuclei as antibodies against lamin B1  
184 (**Figure 2b**) or lamin B2 (**Figure 2c**) (products of the *LMNB1* and *LMNB2* genes). Staining for the  
185 lamin B receptor (**Figure 2e**) exhibited poor image contrast. A pan-tissue survey showed that a  
186 mixture of antibodies for nucleoporin NUP98 and lamin B2 conjugated to the same fluorophore  
187 (Alexafluor-647) generated nuclear envelope staining (NES) for nearly all nuclei across multiple  
188 tissues (**Figure 2f-h**). We judged this to be the optimal antibody cocktail. However, only some cell  
189 types, epithelia in colorectal adenocarcinoma for example, exhibited the ring-like structure that is  
190 characteristic of nuclear lamina in cultured epithelial cells. The nuclear envelope in immune and  
191 other cells has folds and invaginations<sup>23</sup> and in our data, NES staining could be irregular and  
192 diffuse, further emphasizing the difficulty of finding a broadly useful NES stain in tissue.

193           The value of NES images for model performance was also assessed quantitatively and  
194 qualitatively. In images of colon adenocarcinoma, non-neoplastic small intestine, and tonsil tissue,  
195 we found that the addition of NES images resulted in significant improvements in segmentation  
196 accuracy based on IoU with all three learning frameworks; improvements in other tissues, such as  
197 lung adenocarcinoma, were more modest and sporadic (**Figure 3a**, Lung). For nuclear  
198 segmentation of fibroblasts in prostate cancer tissue, UnMICST-U and UnMICST-M models with  
199 NES data were no better than models trained on DNA staining alone. Most striking were cases in  
200 which NES data slightly decreased performance (UnMICST-P segmentation on prostate fibroblasts  
201 and UnMICST-U segmentation of glioblastoma). Inspection of the UnMICST-P masks suggested  
202 that the segmentation of well-separated fibroblast nuclei was already optimal with DNA images  
203 alone (~60% of nuclei retained at IoU of 0.6), implying the addition of NES images afforded little  
204 improvement. With UnMICST-U masks in glioblastoma, the problem appeared to involve atypical  
205 NES morphology, which is consistent with a high level of nuclear pleomorphism and the presence of  
206 “giant cells,” both of which are well-established features of high-grade glioblastoma<sup>24,25</sup>. We also  
207 note that NES data alone was marginally inferior to DNA staining as a sole source of training data  
208 and should therefore be used in combination with images of DNA (**Supplementary Figure 1**). Thus,  
209 adding NES to training data broadly but not universally improves segmentation accuracy.

210

## 211 **Combining NES images and real augmentation has a cumulative effect**

212           To determine whether real augmentation and NES would combine during model training to  
213 achieve superior segmentation precision relative to the use of either type of data alone, we trained  
214 and tested models under four different scenarios (using all three learning frameworks; **Figure 4**).  
215 We used images from the small intestine, a tissue containing nuclei having a wide variety of  
216 morphologies, and then extended the analysis to other tissue types (see below). Models were

217 evaluated on defocused DNA test data to increase the sensitivity of the experiment. In the first  
218 scenario, we trained baseline models using in-focus DNA image data and tested models on unseen  
219 in-focus DNA images. With tissues such as the small intestine, which are challenging to segment  
220 because they contain densely-packed nuclei, scenario A resulted in slightly undersegmented  
221 predictions. In Scenario B and for all subsequent scenarios, defocused DNA images were included  
222 in the test set, giving rise to contours that were substantially misaligned with ground truth  
223 annotations and resulting in higher undersegmentation. False-positive predictions and imprecise  
224 localizations of the nuclei membrane were observed in areas devoid of nuclei and with very low  
225 contrast (**Figure 4a**). When NES images were included in the training set (Scenario C), nuclear  
226 boundaries were more consistent with ground truth annotations, although false-positive predicted  
227 nuclei still remained. The best performance was observed when NES images and real augmentation  
228 were combined: accurate nuclear boundaries were well aligned with ground truth annotations in  
229 both shape and in size. Observable differences in the placement of segmentation masks were  
230 reflected in improvements in IoU: for all three deep learning frameworks, including NES data and  
231 real augmentations increased the fraction of nuclei retained by 50% at an IoU threshold of 0.6  
232 (**Figure 4b**). The accuracy of UnMICST-P (blue curve) trained on in-focus DNA data alone was  
233 higher than the other two baseline models at all IoU thresholds, suggesting that UnMICST-P has a  
234 greater capacity to learn. UnMICST-P may have an advantage in experiments in which staining the  
235 nuclear envelope proves difficult or impossible.

236

### 237 **Combining NES and real augmentation is advantageous across multiple tissue types**

238 To determine if improvements in segmentation would extend to multiple tissue types we  
239 repeated the analysis described above using three scenarios for training and testing with both in-  
240 focus (**Figure 5a**) and defocused images (**Figure 5b**). Scenario 1 used in-focus DNA images for  
241 training (blue bars), scenario 2 used in-focus DNA and NES images (red bars), and scenario 3 used  
242 in-focus DNA and NES images plus real augmentation (green bars). While the magnitude of the  
243 improvement varied with tissue type and test set (panel a vs b), the results as a whole support the  
244 conclusion that including both NES and real augmentations during model training confers  
245 statistically significant improvement in segmentation accuracy with multiple tissue types and models.  
246 The accuracy boost was greatest when models performed poorly (e.g., in scenario 1 where models  
247 were tested on defocused colon image data; **Figure 5b**, blue bars), so that segmentation accuracy  
248 became relatively uniform across tissue and cell types.

249

### 250 **Applying UnMICST to highly multiplex tissue images**

251 To investigate the overall improvement achievable with UnMICST models, we tested them  
252 with and without real or computed augmentations and NES data on all six tissues as a set, including

253 in-focus, saturated and out-focus images (balancing the total amount of training data in each case).  
254 A 1.7-fold improvement in accuracy was observed at an IoU of 0.6 for the fully trained model (i.e.  
255 with NES data and real augmentations; **Figure 6a**). Inspection of segmentation masks also  
256 demonstrated more accurate contours for nuclei across a wide range of shapes. The overall  
257 improvement in accuracy was substantially greater than any difference observed between semantic  
258 and instance segmentation frameworks.

259 We then tested a fully trained UnMICST-U model on a 64-plex CyCIF image of non-  
260 neoplastic small intestine tissue from the EMIT TMA (**Figure 6b**). Staining intensities were  
261 quantified on a per-cell basis, and the results visualized using Uniform Manifold Approximation and  
262 Projection (UMAP; **Figure 6c**). Segmentation masks were found to be well-located with little  
263 evidence of under or over-segmentation (**Figure 6d**). Moreover, whereas 21% of cells with  
264 segmented nuclei stained positive (as determined by using Gaussian-mixture model) for the  
265 immune cell marker CD45, and 53% stained positive for the epithelial cell marker E-cadherin, less  
266 than 3% were positive for both. No known cell type is actually positive for both CD45 and E-  
267 cadherin, and the very low abundance of these double-positive “cells” is evidence of accurate  
268 segmentation. When we examined some of the 830 double positive cells (blue dashed circle in  
269 **Figure 6c**) we found multiple examples of a CD3<sup>+</sup> T cell (yellow arrowheads; light yellow dots in  
270 **Figure 6e**) tightly associated with or between the epithelial cells of intestinal villi (green “kiwi”  
271 structure visible in **Figure 6e**). This is consistent with the known role of the intestinal epithelium in  
272 immune homeostasis<sup>26</sup>. In these cases, the ability of humans to distinguish immune and epithelial  
273 cells relies on prior knowledge, multi-dimensional intensity features and subtle differences in shape  
274 and texture – none of which were aspects of model training. Thus, future improvements in tissue  
275 segmentation are likely to require the development of CNNs able to classify rare but biologically  
276 significant spatial arrangements, rather than simple extensions of the general purpose segmentation  
277 algorithms described here.

278

### 279 **Some tissues still pose a challenge for nuclei segmentation**

280 Of all the tissue types annotated and tested in this paper, non-neoplastic ovary was the most  
281 difficult to segment (**Supplementary Figure 2a**) and addition of ovarian training data to models  
282 trained on data from other tissues decreased their accuracy (**Supplementary Figure 2b**). We have  
283 previously imaged ovarian cancers at even higher resolution (60x/1.42NA sampled at 108 nm pixel  
284 size; (Färkkilä et al., 2020) using optical sectioning and deconvolution microscopy; inspection of  
285 these images reveals nuclei with highly irregular morphology, poor image contrast, and dense  
286 packing (**Supplementary Figure 2c**) unlike colon adenocarcinoma (**Supplementary Figure 2d**).  
287 Thus, additional research, possibly involving different NES antibodies, will be required to improve  
288 performance with ovarian and other difficult to segment tissues. Until then, caution is warranted  
289 when combining training data from tissues with very different nuclear morphologies.

## 290 **DISCUSSION**

291 This paper makes three primary contributions to the growing literature on the segmentation  
292 of tissue images, which is an essential step in single-cell data analysis. First, we show that addition  
293 of real augmentations comprising defocused and saturated images to model training data improves  
294 segmentation accuracy to a significant extent whereas Gaussian blurring does not. These results  
295 extend to deep learning frameworks based on instance segmentation (UnMICST-M) and semantic  
296 segmentation (UnMICST-U and UnMICST-P). Second, we show that it is often possible to increase  
297 segmentation accuracy by including additional data (NES) on nuclear envelop morphology, although  
298 identifying suitable antibody cocktails is not trivial. Finally, using newly generated labeled training  
299 data for multiple tissue types, we show that real augmentation and NES combine to improve the  
300 robustness and accuracy of segmentation across many tissues; these improvements are directly  
301 applicable to the real-world task of segmenting high dimensional tissue and tumor images.

302 UnMICST models are therefore a good starting point for performing image segmentation on rapidly  
303 growing tissue data repositories. Errors remaining when multiplexed images are segmented using  
304 optimized UnMICST models appear to have a subtle biological basis. Development of additional  
305 “physiology aware” machine-learning models may be necessary to reduce these apparent errors.

306 From a machine learning perspective, the value of adding images to training data is self-  
307 evident. Experimental feasibility is less clear. A key tradeoff is that the greater the number of  
308 fluorescence channels used for segmentation, the fewer channels available for the collection of data  
309 on other markers. Fortunately, the development of highly multiplexed imaging has made this less  
310 relevant because collection of 20-40 or more image channels (each corresponding to a different  
311 fluorescent antibody) has become routine. This makes it straightforward to reserve two channels for  
312 segmentation. The cost-benefit ratio of adding extra segmentation data will be different in high  
313 content screening of cells in multi-well plates, however, for which inexpensive reagents are  
314 generally essential. In tissues, lamins change in expression and morphology with disease state<sup>27</sup>  
315 and imaging lamins is also likely to provide valuable biological information in addition to improving  
316 segmentation<sup>28</sup>. To allow others to build on the current work, we are releasing all training and test  
317 images, their segmentation masks and annotations, and real augmentations for multiple types of  
318 tissue (tonsil, ovary, small intestine and cancers of the colon, brain, lung, prostate) via the EMIT  
319 resource; models are released as components of the UnMICST model resource (see data and code  
320 availability information).

321 The most immediately generalizable finding from this work is that real augmentation  
322 outperforms computed augmentation generated using Gaussian kernels. Blurring is an inevitable  
323 consequence of the limited bandwidth of optical systems (their point spread functions) as well as  
324 light scattering and diffraction, and in tissue imaging, the use of non-immersion objective lenses and  
325 consequent refractive index mismatches. Real out-of-focus blur also differs when the focal plane is  
326 above and below the specimen. Areas for future application of real augmentations could include

327 inhomogeneous light sources and stage jitter. While it will be useful to determine kernels for more  
328 effective computed augmentation, collecting real augmentation data imposes a minimal burden in a  
329 real-world setting. Our observation that real augmentation outperforms computed augmentation may  
330 also have general significance outside of the field of microscopy: with any high-performance camera  
331 system, real out-of-focus data will inevitably be more complicated than Gaussian blur.

332

### 333 **CODE AND DATA AVAILABILITY**

334 To allow others to build on the current work we are releasing all training, validation and test images,  
335 their annotations, and real augmentations for multiple types of tissue (tonsil, ovary, small intestine  
336 and cancers of the colon, brain, lung, prostate) via the EMIT resource; models and scripts for  
337 training and inference are released as components of the UnMICST model resource (see data and  
338 code availability information). <https://labsyspharm.github.io/UnMICST-info/>

339

### 340 **METHODS**

#### 341 ***Sample preparation for imaging***

342 To generate images for model training and testing, human tissue specimens (from 42  
343 patients) were used to construct a multi-tissue microarray (HTMA427) under an excess (discarded)  
344 tissue protocol approved by the Institutional Review Board (IRB) at Brigham and Women's Hospital  
345 (BWH IRB 2018P001627). Two 1.5 mm diameter cores were acquired from each of 60 tissue  
346 regions with the goal of acquiring one or two examples of as many tumors as possible (with  
347 matched normal tissue from the same resection when feasible), as well as several non-neoplastic  
348 medical diseases and secondary lymphoid tissues such as tonsil, spleen and lymph nodes. Overall,  
349 the tissue microarray (TMA) contained 120 cores plus three additional “marker cores,” which are  
350 cores added to the TMA in a manner that makes it possible to orient the image. Slides were stained  
351 with the following reagents from Cell Signaling Technologies (Beverly MA, USA) and Abcam  
352 (Cambridge UK).

353

Target	Fluorochrome	Species	Clone	Vendor	Cat. No.	RRID
DNA	Hoechst 33342	NA	NA	CST	4082	AB_10626776
Lamin B2	Alexafluor 647	Rabbit	EPR9701(B)	Abcam	ab200427	AB_2889288
NUP98	Alexafluor 647	Rabbit	C39A3	CST	13393	AB_2728831

354

355 Before imaging, slides were mounted with 90% glycerol and a #1.5 coverslip. Prior to  
356 algorithmic evaluation, the images were split into three mutually disjoint subsets and used for  
357 training, validation, and testing.

358

359

## 360 **Acquisition of image data and real augmentations**

361 The stained TMA was imaged on a INCell 6000 (General Electric Life Sciences) microscope  
362 equipped with a 20x/0.75 objective lens (370 nm nominal lateral resolution at 550 nm wavelength)  
363 and a pixel size of 0.325  $\mu\text{m}$  per pixel. Hoechst and lamin-A647 were excited with a 405 nm and  
364 642 nm laser, respectively. Emission was collected with the “DAPI” (455/50 nm) and “Cy5” (682/60  
365 nm) filter sets with exposure times of 60 ms and 100 ms, respectively. Whole-slide imaging involved  
366 acquisition of 1,215 tiles with an 8% overlap, which is recommended for stitching in ASHLAR, a next  
367 generation stitching and registration algorithm for large images  
368 (<https://github.com/labsyspharm/ashlar>). To generate defocused data, we acquired images from  
369 above and below the focal plane by varying the Z-axis by 3  $\mu\text{m}$  in both directions. To generate  
370 saturated images of DNA staining, a 150ms exposure time was used. These two types of  
371 “suboptimal” data were then used for “real augmentation” during model training, as described below.

372 Representative cores for lung adenocarcinoma, non-neoplastic small intestine, normal  
373 prostate, colon adenocarcinoma, glioblastoma, non-neoplastic ovary, and tonsil were extracted from  
374 image mosaics and down-sampled by a factor of 2 to match the pixel size of images routinely  
375 acquired and analyzed in MCMICRO<sup>29</sup>. Images were then cropped to 256 x 256-pixel tiles, and in-  
376 focus DNA and NES were imported into Adobe Photoshop to facilitate human annotation of nuclear  
377 boundaries. Annotations for contours and background classes were labelled on separate layers  
378 while swapping between DNA and NES as necessary. To save time, we drew complete contours of  
379 nuclei and filled these in using the Matlab *imfill* operation to generate nuclei centers. For nuclei at  
380 the image borders where contours would be incomplete, we manually annotated nuclei centers. As  
381 described by Ronneberger et al. (2015), a fourth layer was used to mark areas between clumped  
382 cells. These additional annotations made it possible to specifically penalize models that incorrectly  
383 classified these pixels.

384 Because original, defocused, and saturated images of DNA were all acquired in the same  
385 image stack, it was possible to use a single registered set of DNA annotations across all augmented  
386 image channels. To produce the training set, each image was cropped into 64 x 64 patches,  
387 normalized to use the full dynamic range, and further augmented using 90-degree rotations,  
388 reflections, and 20% upscaling. Consistent with the training set, the validation and test sets also  
389 include defocused and saturated examples but were not augmented with standard transformations.  
390 The ratio of data examples present in the training, validation, and test set split was 0.36:0.24:0.4.  
391 For a fair comparison across models, the same dataset and split were used for the three deep  
392 learning frameworks described in this manuscript (**Supplementary Table 2**).

393

## 394 **Model implementation**

395 To facilitate model training, three distinct state-of-the-art architectures were implemented  
396 and evaluated. They are, in no particular order, UNet, Mask R-CNN, and PSPNet. UNet was

397 selected for its prior success in the biomedical domain, Mask R-CNN was selected for its ability to  
398 perform both object detection and mask generation, and PSPNet was selected for its capacity to  
399 integrate image features from multiple spatial scales. Training, validation, and test data were  
400 derived from 12 cores in 7 tissues and a total of 10,359 nuclei in the composition of colon – 1,142;  
401 glioblastoma (GBM) – 675; lung – 1735; ovarian – 956; fibroblast – 922; small intestine – 1677;  
402 tonsil – 3252. To maintain consistency of evaluation across segmentation algorithms, segmentation  
403 accuracy was calculated by counting the fraction of cells in a held out test set that passed a  
404 sweeping Intersection over Union (IoU) threshold.

405

## 406 **Model Training**

### 407 ***UnMICST-U models***

408 A three-class UNet model<sup>8</sup> was trained based on annotation of nuclei centers, nuclei  
409 contours, and background. The neural network is comprised of 4 layers and 80 input features.  
410 Training was performed using a batch size of 32 with the Adam Optimizer and a learning rate of  
411 0.00005 with a decay rate of 0.98 every 5,000 steps until there was no improvement in accuracy or  
412 ~100 epochs had been reached. Batch normalization was used to improve training speed. During  
413 training, the bottom layer had a dropout rate of 0.35, and L1 regularization was implemented to  
414 minimize overfitting<sup>30,31</sup> and early stopping. Training was performed on workstations equipped with  
415 NVidia GTX 1080 or NVidia TitanX GPUs.

416

### 417 ***UnMICST-M models***

418 Many segmentation models are based on the Mask R-CNN architecture<sup>9</sup>, Mask R-CNN has  
419 previously exhibited excellent performance on a variety of segmentation tasks. Mask R-CNN begins  
420 by detecting bounding boxes of nuclei and subsequently performs segmentation within each box.  
421 This approach eliminates the need for an intermediate watershed, or equivalent, segmentation step.  
422 Thus, Mask R-CNN directly calculates a segmentation mask, significantly reducing the overhead in  
423 traditional segmentation pipelines. We adopted a ResNet50<sup>32</sup> backbone model in the UnMICST-M  
424 implementation and initialized the weights using pretrained values from the COCO object instance  
425 segmentation challenge<sup>33</sup> to improve convergence properties. For efficient training, we upsampled  
426 the original input images to 800 x 800-pixels and trained a model for 24 epochs using a batch size  
427 of 8. The Adam optimizer, with a weight decay of 0.0001 to prevent overfitting, was exploited with a  
428 variable learning rate, initially set to 0.01 and decreased by a factor of 0.1 at epochs 16 and 22.  
429 Training was performed on a compute node cluster using 4 NVidia TitanX or NVidia Tesla V100  
430 GPUs. For evaluation and comparison, we used the model with the highest performance on the  
431 validation set, following standard practice.

432

433

#### 434 **UnMICST-P models**

435 We trained a three class PSPNet model<sup>34</sup> to extract cell nuclei centers, nuclei contours, and  
436 background from a wide variety of tissue types. PSPNet is one of the most widely used  
437 convolutional neural networks for the semantic segmentation of natural scene images in the  
438 computer vision field. The network employs a so-called pyramid pooling module whose purpose is  
439 to learn global as well as local features. The additional contextual information used by PSPNet  
440 allowed the segmentation algorithm to produce realistic probability maps with greater confidence.  
441 We used ResNet101 as a backbone. Training of the network was performed using a batch size of 8  
442 with an image size of 256 x 256-pixels for 15,000 iterations or until the minimum loss model was  
443 obtained. A standard cross entropy loss function was used during training. Gradient descent was  
444 performed using the Adam optimizer with a learning rate of 0.0001 and a weight decay parameter of  
445 0.005 via L2 regularization. Batch normalization was employed for faster convergence, and a  
446 dropout probability of 0.5 was used in the final network layer to mitigate overfitting. The model  
447 training was performed on a compute cluster node equipped with NVidia Tesla V100 GPUs.

448

#### 449 **Analysis of multi-dimensional data**

450 For the analysis shown in Figure 6, a 64-plex CyCIF image of non-neoplastic small intestine  
451 tissue from the EMIT TMA (<https://www.synapse.org/#!Synapse:syn22345748/>) was stained with a  
452 total of 45 antibodies as described in protocols [https://www.protocols.io/view/ffpe-tissue-pre-](https://www.protocols.io/view/ffpe-tissue-pre-treatment-before-t-cycif-on-leica-bji2kkqge)  
453 [treatment-before-t-cycif-on-leica-bji2kkqge](https://www.protocols.io/view/ffpe-tissue-pre-treatment-before-t-cycif-on-leica-bji2kkqge) and [dx.doi.org/10.17504/protocols.io.bjiukkew](https://doi.org/10.17504/protocols.io.bjiukkew). Images  
454 were segmented using the UnMICST-U model trained on DNA with NES data and real  
455 augmentations. Mean fluorescence intensities across 45 markers for 27,847 segmented nuclei were  
456 quantified as described in<sup>29</sup>. E-cadherin positive and CD45 positive cells were identified using  
457 Gaussian-mixture models on log-transformed data. For multivariate clustering, log-transformed  
458 mean intensities of all single cells of 14 selected protein markers (E-cadherin, pan-cytokeratin,  
459 CD45 CD4, CD3D, CD8, RF3, PML, GLUT1, GAPDH TDP43, OGT, COLL4, an EPCAM) were pre-  
460 processed using Uniform Manifold Approximation and Projection (UMAP)<sup>35</sup> and clustered using  
461 Hierarchical Density-Based Spatial Clustering of Applications with Noise (HDBSCAN)<sup>36</sup>. Clusters  
462 expressing a high level of both E-cadherin and CD45 were identified and overlaid on a false-colored  
463 image showing the staining of DNA, E-cadherin, and CD45.

464

#### 465 **ACKNOWLEDGEMENTS**

466 We thank Alyce Chen and Madison Tyler for their help with this manuscript. The work was funded  
467 by NIH grants U54-CA225088 and U2C-CA233262 to P.K.S. and S.S and by the Ludwig Cancer  
468 Center at Harvard. Z.M. is supported by NCI grant R50-CA252138. We thank Dana-Farber/Harvard  
469 Cancer Center (P30- CA06516) for the use of its Specialized Histopathology Core.

470

471 **AUTHOR CONTRIBUTIONS**

472 The study design was conceived by CY, WDJ, EN, and PKS. Image acquisition and annotation was  
473 done by CY. SS provided the EMIT TMA sample and validated the tissue types. TMA staining was  
474 performed by ZM and CAJ. Data analysis was performed by CY, WDJ, EN, and YAC. YAC and CY  
475 performed the single cell quantitative analysis and analysis found in figure 6. Additional coding was  
476 done by MC. Additional experiments were conducted by DW. PKS, SS, HP supervised the study. All  
477 authors contributed to the writing and editing of the manuscript.

478

479 **OUTSIDE INTERESTS**

480 PKS is a member of the SAB or BOD member of Applied Biomath, RareCyte Inc., and Glencoe  
481 Software, which distributes a commercial version of the OMERO database; PKS is also a member  
482 of the NanoString SAB. In the last five years the Sorger lab has received research funding from  
483 Novartis and Merck. Sorger declares that none of these relationships have influenced the content of  
484 this manuscript. SS is a consultant for RareCyte Inc. The other authors declare no outside interests.

485

486

487

488

489

490

491

492

493

494

495

496

497

498

499

500

501

502

503

504

505

506

507

508 **REFERENCES**

- 509 1. Immunologists, A. A. of. The Demonstration of Pneumococcal Antigen in Tissues by the Use of  
510 Fluorescent Antibody. *J. Immunol.* **45**, 159–170 (1942).
- 511 2. Albertson, D. G. Gene amplification in cancer. *Trends Genet.* **22**, 447–455 (2006).
- 512 3. Shlien, A. & Malkin, D. Copy number variations and cancer. *Genome Med.* **1**, 62 (2009).
- 513 4. Amin, M. B. *et al.* The Eighth Edition AJCC Cancer Staging Manual: Continuing to build a bridge  
514 from a population-based to a more ‘personalized’ approach to cancer staging. *CA. Cancer J.*  
515 *Clin.* **67**, 93–99 (2017).
- 516 5. Achim, K. *et al.* High-throughput spatial mapping of single-cell RNA-seq data to tissue of origin.  
517 *Nat. Biotechnol.* **33**, 503–509 (2015).
- 518 6. Slyper, M. *et al.* A single-cell and single-nucleus RNA-Seq toolbox for fresh and frozen human  
519 tumors. *Nat. Med.* **26**, 792–802 (2020).
- 520 7. LeCun, Y., Bengio, Y. & Hinton, G. Deep learning. *Nature* **521**, 436–444 (2015).
- 521 8. Ronneberger, O., Fischer, P. & Brox, T. U-Net: Convolutional Networks for Biomedical Image  
522 Segmentation. *ArXiv150504597 Cs* (2015).
- 523 9. He, K., Gkioxari, G., Dollár, P. & Girshick, R. Mask R-CNN. *ArXiv170306870 Cs* (2018).
- 524 10. Caicedo, J. C. *et al.* Nucleus segmentation across imaging experiments: the 2018 Data Science  
525 Bowl. *Nat. Methods* **16**, 1247–1253 (2019).
- 526 11. Kromp, F. *et al.* An annotated fluorescence image dataset for training nuclear segmentation  
527 methods. *Sci. Data* **7**, 262 (2020).
- 528 12. Schwendy, M., Unger, R. E. & Parekh, S. H. EVICAN—a balanced dataset for algorithm  
529 development in cell and nucleus segmentation. *Bioinformatics* **36**, 3863–3870 (2020).
- 530 13. Gurari, D. *et al.* How to Collect Segmentations for Biomedical Images? A Benchmark Evaluating  
531 the Performance of Experts, Crowdsourced Non-experts, and Algorithms. *2015 IEEE Winter*  
532 *Conf. Appl. Comput. Vis.* (2015) doi:10.1109/WACV.2015.160.
- 533 14. Schüffler, P. J. *et al.* Automatic single cell segmentation on highly multiplexed tissue images.  
534 *Cytometry A* **87**, 936–942 (2015).

- 535 15. Arganda-Carreras, I. *et al.* Trainable Weka Segmentation: a machine learning tool for  
536 microscopy pixel classification. *Bioinformatics* **33**, 2424–2426 (2017).
- 537 16. Berg, S. *et al.* ilastik: interactive machine learning for (bio)image analysis. *Nat. Methods* **16**,  
538 1226–1232 (2019).
- 539 17. Krizhevsky, A., Sutskever, I. & Hinton, G. E. ImageNet Classification with Deep Convolutional  
540 Neural Networks. in *Advances in Neural Information Processing Systems 25* (eds. Pereira, F.,  
541 Burges, C. J. C., Bottou, L. & Weinberger, K. Q.) 1097–1105 (Curran Associates, Inc., 2012).
- 542 18. Ahmed Raza, S. E. *et al.* MIMO-Net: A multi-input multi-output convolutional neural network for  
543 cell segmentation in fluorescence microscopy images. in *2017 IEEE 14th International  
544 Symposium on Biomedical Imaging (ISBI 2017)* 337–340 (IEEE, 2017).  
545 doi:10.1109/ISBI.2017.7950532.
- 546 19. Shorten, C. & Khoshgoftaar, T. M. A survey on Image Data Augmentation for Deep Learning. *J.*  
547 *Big Data* **6**, 60 (2019).
- 548 20. Horwath, J. P., Zakharov, D. N., Mégrét, R. & Stach, E. A. Understanding important features of  
549 deep learning models for segmentation of high-resolution transmission electron microscopy  
550 images. *Npj Comput. Mater.* **6**, 1–9 (2020).
- 551 21. Skinner, B. M. & Johnson, E. E. P. Nuclear morphologies: their diversity and functional  
552 relevance. *Chromosoma* **126**, 195–212 (2017).
- 553 22. Dalle, J.-R. *et al.* Nuclear pleomorphism scoring by selective cell nuclei detection. in *IEEE  
554 Workshop on Applications of Computer Vision (WACV 2009), 7-8 December, 2009, Snowbird,  
555 UT, USA* (IEEE Computer Society, 2009).
- 556 23. Fischer, E. G. Nuclear Morphology and the Biology of Cancer Cells. *Acta Cytol.* **64**, 511–519  
557 (2020).
- 558 24. Kros, J. M. Grading of Gliomas: The Road From Eminence to Evidence. *J. Neuropathol. Exp.*  
559 *Neurol.* **70**, 101–109 (2011).
- 560 25. Louis, D., Ohgaki, H., Wiestler, O. & Cavenee, W. *WHO Classification of Tumours of the Central  
561 Nervous System.* (2016).

- 562 26. Allaire, J. M. *et al.* The Intestinal Epithelium: Central Coordinator of Mucosal Immunity. *Trends*  
563 *Immunol.* **39**, 677–696 (2018).
- 564 27. Sakthivel, K. M. & Sehgal, P. A Novel Role of Lamins from Genetic Disease to Cancer  
565 Biomarkers. *Oncol. Rev.* **10**, 309 (2016).
- 566 28. Bell, E. S. & Lammerding, J. Causes and consequences of nuclear envelope alterations in  
567 tumour progression. *Eur. J. Cell Biol.* **95**, 449–464 (2016).
- 568 29. Schapiro, D. *et al.* MCMICRO: A scalable, modular image-processing pipeline for multiplexed  
569 tissue imaging. *bioRxiv* 2021.03.15.435473 (2021) doi:10.1101/2021.03.15.435473.
- 570 30. Ng, A. Y. Feature selection, L1 vs. L2 regularization, and rotational invariance. in *Proceedings*  
571 *of the twenty-first international conference on Machine learning* 78 (Association for Computing  
572 Machinery, 2004). doi:10.1145/1015330.1015435.
- 573 31. Srivastava, N., Hinton, G., Krizhevsky, A., Sutskever, I. & Salakhutdinov, R. Dropout: a simple  
574 way to prevent neural networks from overfitting. *J. Mach. Learn. Res.* **15**, 1929–1958 (2014).
- 575 32. He, K., Zhang, X., Ren, S. & Sun, J. Deep Residual Learning for Image Recognition. in *2016*  
576 *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)* 770–778 (2016).  
577 doi:10.1109/CVPR.2016.90.
- 578 33. Lin, T.-Y. *et al.* Microsoft COCO: Common Objects in Context. in *Computer Vision – ECCV*  
579 *2014* (eds. Fleet, D., Pajdla, T., Schiele, B. & Tuytelaars, T.) 740–755 (Springer International  
580 Publishing, 2014). doi:10.1007/978-3-319-10602-1\_48.
- 581 34. Zhao, H., Shi, J., Qi, X., Wang, X. & Jia, J. Pyramid Scene Parsing Network. *ArXiv161201105*  
582 *Cs* (2017).
- 583 35. Becht, E. *et al.* Dimensionality reduction for visualizing single-cell data using UMAP. *Nat.*  
584 *Biotechnol.* **37**, 38–44 (2019).
- 585 36. Campello, R. J. G. B., Moulavi, D. & Sander, J. Density-Based Clustering Based on Hierarchical  
586 Density Estimates. in *Advances in Knowledge Discovery and Data Mining* (eds. Pei, J., Tseng,  
587 V. S., Cao, L., Motoda, H. & Xu, G.) 160–172 (Springer, 2013). doi:10.1007/978-3-642-37456-  
588 2\_14.
- 589

590 **The following references are related to supplementary materials only**

- 591 37. Otsu, N. A Threshold Selection Method from Gray-Level Histograms. (1979)  
592 doi:10.1109/TSMC.1979.4310076.
- 593 38. Byun, J. *et al.* Automated tool for the detection of cell nuclei in digital microscopic images:  
594 application to retinal images. *Mol. Vis.* **12**, 949–960 (2006).
- 595 39. Xu, H., Lu, C. & Mandal, M. An efficient technique for nuclei segmentation based on ellipse  
596 descriptor analysis and improved seed detection algorithm. *IEEE J. Biomed. Health Inform.* **18**,  
597 1729–1741 (2014).
- 598 40. Xu, H., Lu, C., Berendt, R., Jha, N. & Mandal, M. Automatic Nuclei Detection Based on  
599 Generalized Laplacian of Gaussian Filters. *IEEE J. Biomed. Health Inform.* **21**, 826–837 (2017).
- 600 41. Qi, X., Xing, F., Foran, D. J. & Yang, L. Robust Segmentation of Overlapping Cells in  
601 Histopathology Specimens Using Parallel Seed Detection and Repulsive Level Set. *IEEE Trans.*  
602 *Biomed. Eng.* **59**, 754–765 (2012).
- 603 42. Liu, C., Shang, F., Ozolek, J. A. & Rohde, G. K. Detecting and segmenting cell nuclei in two-  
604 dimensional microscopy images. *J. Pathol. Inform.* **7**, 42 (2016).
- 605 43. Lux, F. & Matula, P. Cell Segmentation by Combining Marker-Controlled Watershed and Deep  
606 Learning. *ArXiv200401607 Cs Eess* (2020).
- 607 44. Wang, C. *et al.* Deep learning pipeline for cell edge segmentation of time-lapse live cell images.  
608 *bioRxiv* 191858 (2019) doi:10.1101/191858.
- 609 45. Shahzad M *et al.* Robust Method for Semantic Segmentation of Whole-Slide Blood Cell  
610 Microscopic Images. *Comput. Math. Methods Med.* **2020**, 4015323–4015323 (2020).
- 611 46. Lee, H. & Jeong, W.-K. Scribble2Label: Scribble-Supervised Cell Segmentation via Self-  
612 Generating Pseudo-Labels with Consistency. *ArXiv200612890 Cs* (2020).
- 613 47. Al-Kofahi, Y., Zaltsman, A., Graves, R., Marshall, W. & Rusu, M. A deep learning-based  
614 algorithm for 2-D cell segmentation in microscopy images. *BMC Bioinformatics* **19**, 365 (2018).
- 615 48. McQuin, C. *et al.* CellProfiler 3.0: Next-generation image processing for biology. *PLOS Biol.* **16**,  
616 e2005970 (2018).

- 617 49. Schmidt, U., Weigert, M., Broaddus, C. & Myers, G. Cell Detection with Star-convex Polygons.  
618 *ArXiv180603535 Cs* **11071**, 265–273 (2018).
- 619 50. Wen, C. *et al.* Deep-learning-based flexible pipeline for segmenting and tracking cells in 3D  
620 image time series for whole brain imaging. *bioRxiv* 385567 (2018) doi:10.1101/385567.
- 621 51. Vu, Q. D. *et al.* Methods for Segmentation and Classification of Digital Microscopy Tissue  
622 Images. *Front. Bioeng. Biotechnol.* **7**, (2019).
- 623 52. Lugagne, J.-B., Lin, H. & Dunlop, M. J. DeLTA: Automated cell segmentation, tracking, and  
624 lineage reconstruction using deep learning. *PLOS Comput. Biol.* **16**, e1007673 (2020).
- 625 53. Kromp, F. *et al.* Deep Learning architectures for generalized immunofluorescence based  
626 nuclear image segmentation. (2019).
- 627 54. Vuola, A. O., Akram, S. U. & Kannala, J. Mask-RCNN and U-net Ensembled for Nuclei  
628 Segmentation. *ArXiv190110170 Cs* (2019).
- 629 55. Korfhage, N. *et al.* Detection and segmentation of morphologically complex eukaryotic cells in  
630 fluorescence microscopy images via feature pyramid fusion. *PLOS Comput. Biol.* **16**, e1008179  
631 (2020).
- 632 56. Liu, D. *et al.* Unsupervised Instance Segmentation in Microscopy Images via Panoptic Domain  
633 Adaptation and Task Re-Weighting. in *2020 IEEE/CVF Conference on Computer Vision and*  
634 *Pattern Recognition (CVPR)* 4242–4251 (IEEE, 2020). doi:10.1109/CVPR42600.2020.00430.
- 635 57. Masubuchi, S. *et al.* Deep-learning-based image segmentation integrated with optical  
636 microscopy for automatically searching for two-dimensional materials. *Npj 2D Mater. Appl.* **4**, 1–  
637 9 (2020).
- 638 58. Deng, J. *et al.* ImageNet: A large-scale hierarchical image database. in *2009 IEEE Conference*  
639 *on Computer Vision and Pattern Recognition* 248–255 (2009).  
640 doi:10.1109/CVPR.2009.5206848.
- 641 59. Torr, A., Basaran, D., Sero, J., Rittscher, J. & Sailem, H. DeepSplit: Segmentation of  
642 Microscopy Images Using Multi-task Convolutional Networks. in *Medical Image Understanding*  
643 *and Analysis* (eds. Papież, B. W., Namburete, A. I. L., Yaqub, M. & Noble, J. A.) vol. 1248 155–  
644 167 (Springer International Publishing, 2020).

- 645 60. Stringer, C., Wang, T., Michaelos, M. & Pachitariu, M. Cellpose: a generalist algorithm for  
646 cellular segmentation. *Nat. Methods* 1–7 (2020) doi:10.1038/s41592-020-01018-x.
- 647 61. Ljosa, V., Sokolnicki, K. L. & Carpenter, A. E. Annotated high-throughput microscopy image sets  
648 for validation. *Nat. Methods* **9**, 637 (2012).
- 649 62. Xing, F. *et al.* Towards pixel-to-pixel deep nucleus detection in microscopy images. *BMC*  
650 *Bioinformatics* **20**, 472 (2019).
- 651 63. Yang, L. *et al.* NuSeT: A deep learning tool for reliably separating and analyzing crowded cells.  
652 *PLOS Comput. Biol.* **16**, e1008193 (2020).
- 653 64. Al-Kofahi, Y., Lassoued, W., Lee, W. & Roysam, B. Improved Automatic Detection and  
654 Segmentation of Cell Nuclei in Histopathology Images. *IEEE Trans. Biomed. Eng.* **57**, 841–852  
655 (2010).
- 656 65. Roth, H. R. *et al.* DeepOrgan: Multi-level Deep Convolutional Networks for Automated Pancreas  
657 Segmentation. in *Medical Image Computing and Computer-Assisted Intervention -- MICCAI*  
658 *2015* (eds. Navab, N., Hornegger, J., Wells, W. M. & Frangi, A.) 556–564 (Springer International  
659 Publishing, 2015). doi:10.1007/978-3-319-24553-9\_68.
- 660 66. Chen, H., Qi, X., Yu, L. & Heng, P.-A. DCAN: Deep Contour-Aware Networks for Accurate  
661 Gland Segmentation. *ArXiv160402677 Cs* (2016).
- 662 67. Höfener, H. *et al.* Deep learning nuclei detection: A simple approach can deliver state-of-the-art  
663 results. *Comput. Med. Imaging Graph.* **70**, 43–52 (2018).
- 664 68. Khadangi, A., Boudier, T. & Rajagopal, V. EM-net: Deep learning for electron microscopy image  
665 segmentation. *bioRxiv* 2020.02.03.933127 (2020) doi:10.1101/2020.02.03.933127.
- 666 69. Raza, S. E. A. *et al.* Micro-Net: A unified model for segmentation of various objects in  
667 microscopy images. *Med. Image Anal.* **52**, 160–173 (2019).
- 668 70. Berryman, S., Matthews, K., Lee, J. H., Duffy, S. P. & Ma, H. Image-based Cell Phenotyping  
669 Using Deep Learning. *bioRxiv* 817544 (2019) doi:10.1101/817544.
- 670 71. Challen, G. A. & Goodell, M. A. Promiscuous Expression of H2B-GFP Transgene in  
671 Hematopoietic Stem Cells. *PLOS ONE* **3**, e2357 (2008).
- 672 72. Tumber, T. *et al.* Defining the epithelial stem cell niche in skin. *Science* **303**, 359–363 (2004).

673 73. Chalfoun, J. *et al.* FogBank: a single cell segmentation across multiple cell lines and image  
674 modalities. *BMC Bioinformatics* **15**, 431 (2014).

675 74. Stylianidou, S., Brennan, C., Nissen, S. B., Kuwada, N. J. & Wiggins, P. A. SuperSegger: robust  
676 image segmentation, analysis and lineage tracking of bacterial cells. *Mol. Microbiol.* **102**, 690–  
677 700 (2016).

678

679

680

681

682

683

684

685

686

687

688

689

690

691

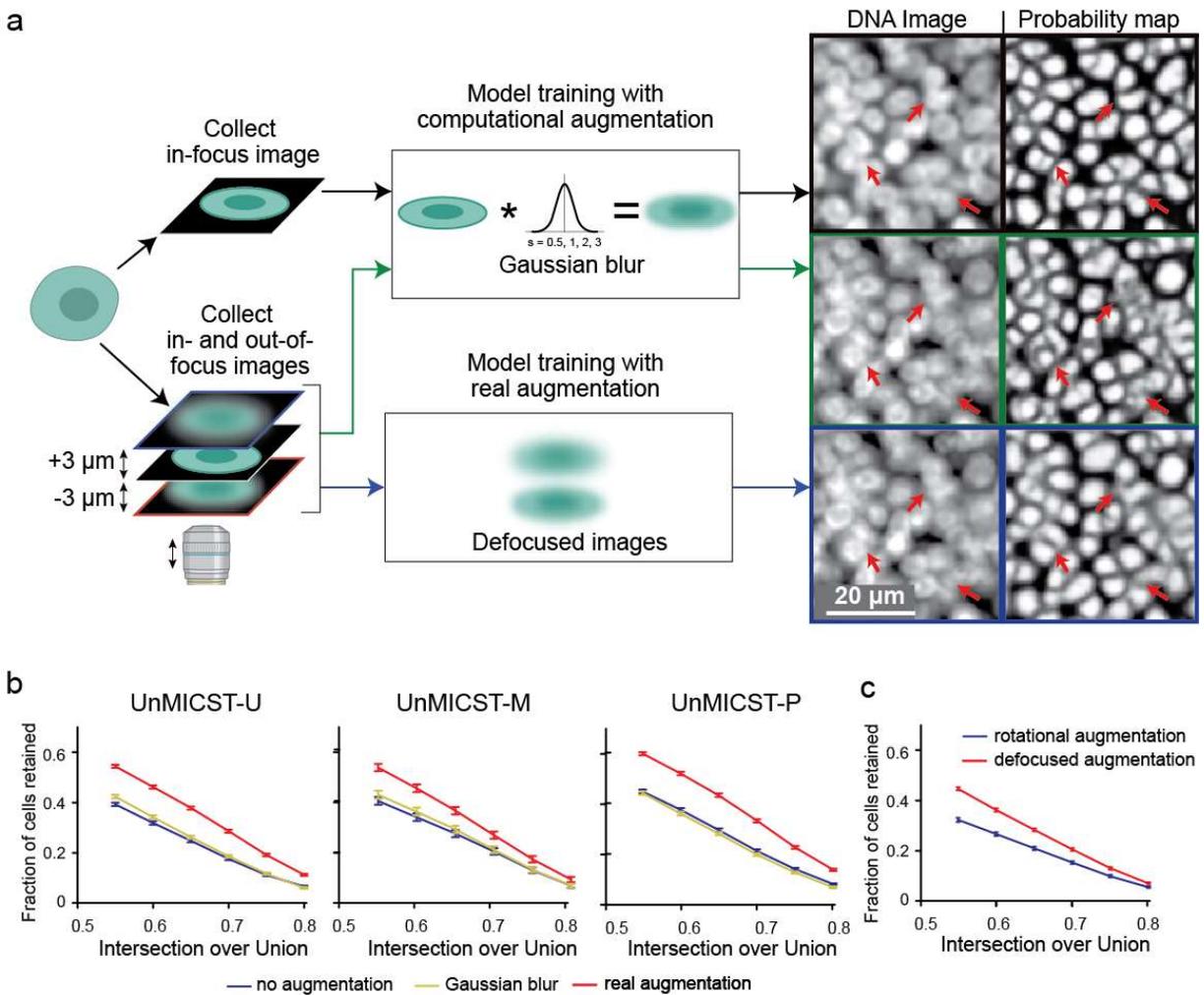
692

693

694

695

**Figure 1**



696

697

698 **Figure 1: Comparing the use of real augmentations (defocused and overexposed images)**

699 **and Gaussian blur. a)** Schematic diagram showing the approach comparing test images on models

700 trained with Gaussian-blurred or defocused image data. Higher contrast probability maps signify

701 more confidence – areas of interest are highlighted with red arrows. Corresponding probability maps

702 indicate a model trained with defocused images performs better on defocused test images than a

703 Gaussian-blurred model. Scale bar denotes 20 micrometers. **b)** Plots show that incorporating real

704 augmentations (red curve) into the training set is statistically significantly superior to training sets

705 with Gaussian blur (yellow curve) and without real augmentations (blue curve) for UnMICST-U,

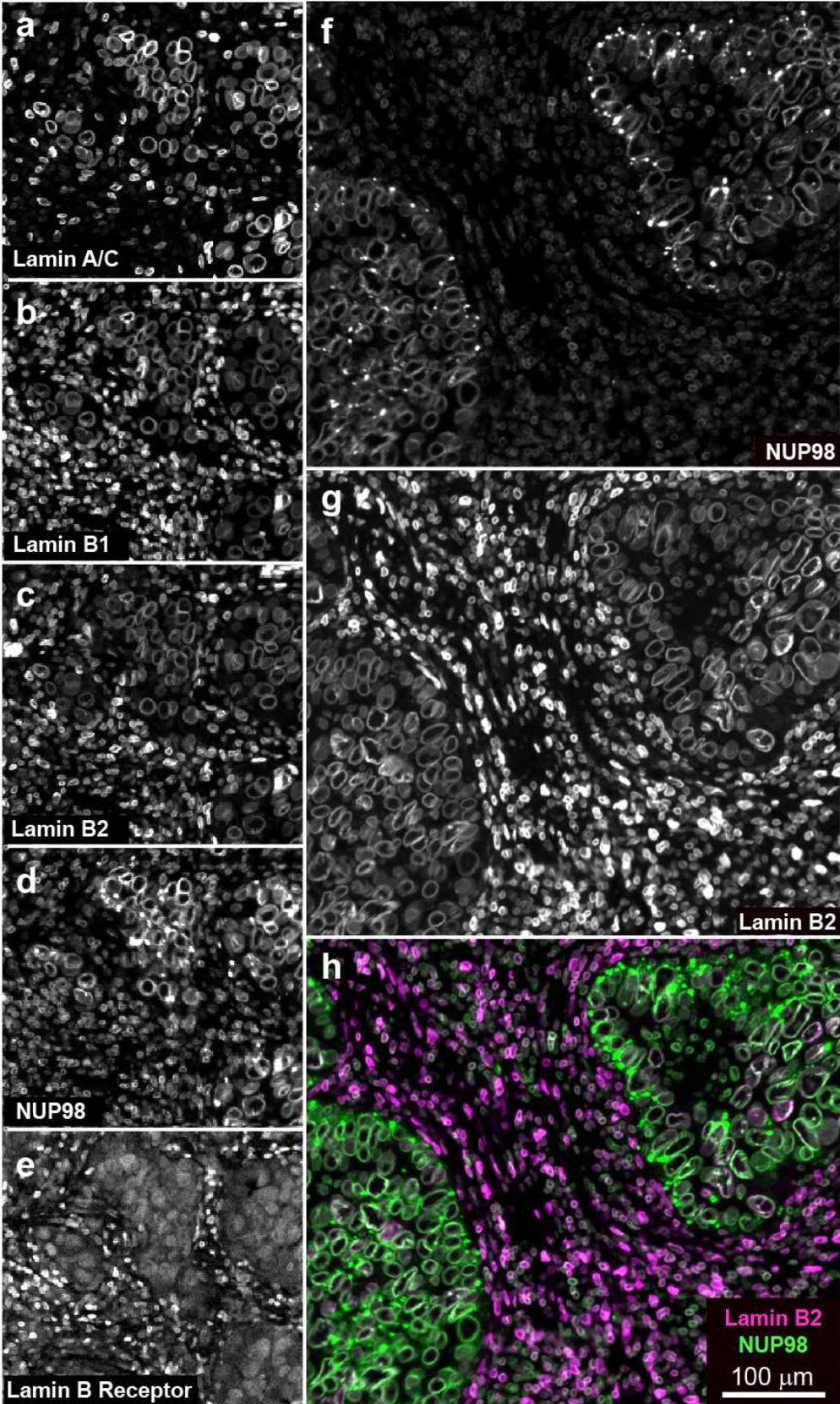
706 UnMICST-M, and UnMICST-P. Simulating defocused images with Gaussian blur is only marginally

707 better than not augmenting the training data at all. **c)** Comparing UnMICST-U model accuracy when

708 the training dataset size was held constant by replacing defocused augmentations (red curve) with

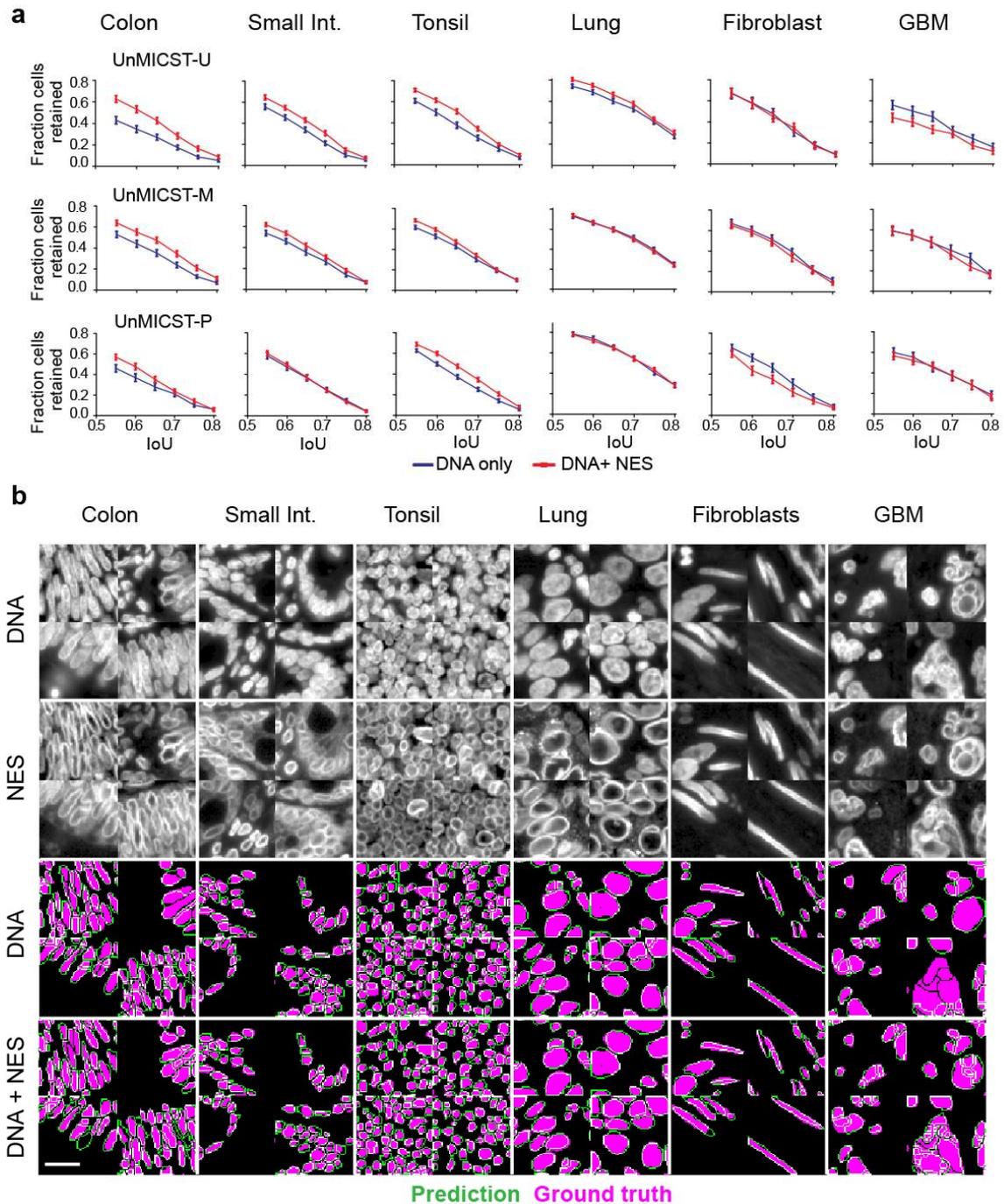
709 90 and 180 degree rotations (blue curve).

Figure 2



711 **Figure 2: Comparing different nuclear envelope stains in colon adenocarcinoma. a-e)**  
712 Showcasing lamin A/C, lamin B1, lamin B2, NUP98, and the lamin B receptor in the same field of  
713 view. Lamin B1 and B2 appear to stain similar proportions of nuclei while lamin A/C stains fewer  
714 nuclei. The stain against the lamin B receptor was comparatively weaker. Lamin B2 (f) and NUP98  
715 (g) are complementary and, when used in combination, maximize the number of cells stained. h)  
716 Composite of lamin B2 (purple) and NUP98 (green). Scale bar denotes 100 micrometers.

**Figure 3**



717

718 **Figure 3: NES with DNA improves nuclear segmentation.** NES – nuclear envelope staining.

719 Assessing the addition of NES as a 2<sup>nd</sup> marker to DNA on segmentation accuracy on a per tissue

720 and per model basis. **a)** Variable IoU plots comparing the DNA-only model (blue curve) and the

721 DNA + NES model (red curve) across frameworks. Adding NES increased accuracy for densely

722 packed nuclei such as colon, small intestine, tonsil, and to some extent, lung tissue. Error bars are

723 standard errors of mean. **b)** Representative grayscale images of tissues stained with DNA and NES  
724 comparing their variable morphologies, followed by UnMICST-U mask predictions (green) overlaid  
725 onto ground truth annotations (purple). In tissue with sparse nuclei, such as fibroblasts from prostate  
726 tissue, NES did not add an additional benefit to DNA alone. In tissues where NES does not exhibit  
727 the characteristic nuclear ring, as in glioblastoma, the accuracy was similarly not improved. Scale  
728 bar denotes 20 micrometers.

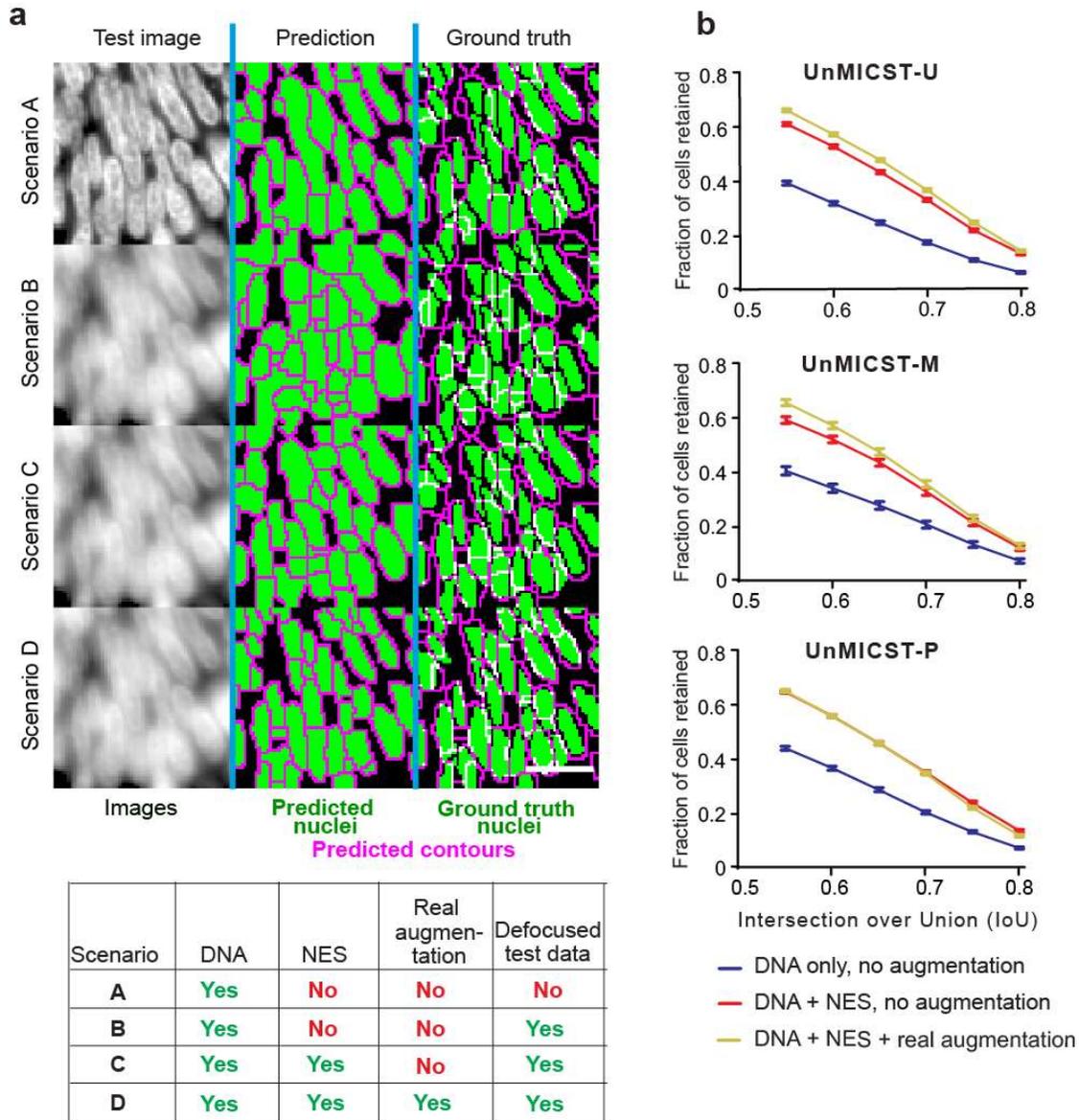
729

730

731

732

**Figure 4**



733

734

735 **Figure 4: Combination of NES and real image augmentations on segmentation performance.**

736 NES - nuclear envelope staining. **a)** Models trained with in-focus DNA data alone produced  
 737 probability maps that were undersegmented, especially in densely-packed tissue such as small  
 738 intestine (Scenario A). When tested on defocused data, nuclei borders were largely incorrect  
 739 (Scenario B). Adding NES restored nuclei border shapes (Scenario C). Combining NES and real  
 740 augmentations reduced false positive detections and produced nuclei masks better resembling the  
 741 ground truth labels (Scenario D). Scalebar denotes 20 micrometers. Table legend shows conditions  
 742 used for each scenario A-D. **b)** Graphs compare the accuracy represented as the number of cells

743 retained across varying IoU thresholds with all models from UnMICST-U (top), UnMICST-M  
744 (center), and UnMICST-P (bottom). In all models, more nuclei were retained when NES and real  
745 augmentations were used together during training (yellow curves) compared to using NES without  
746 real augmentations (red curves) or DNA alone (blue curves).

747

748

749

750

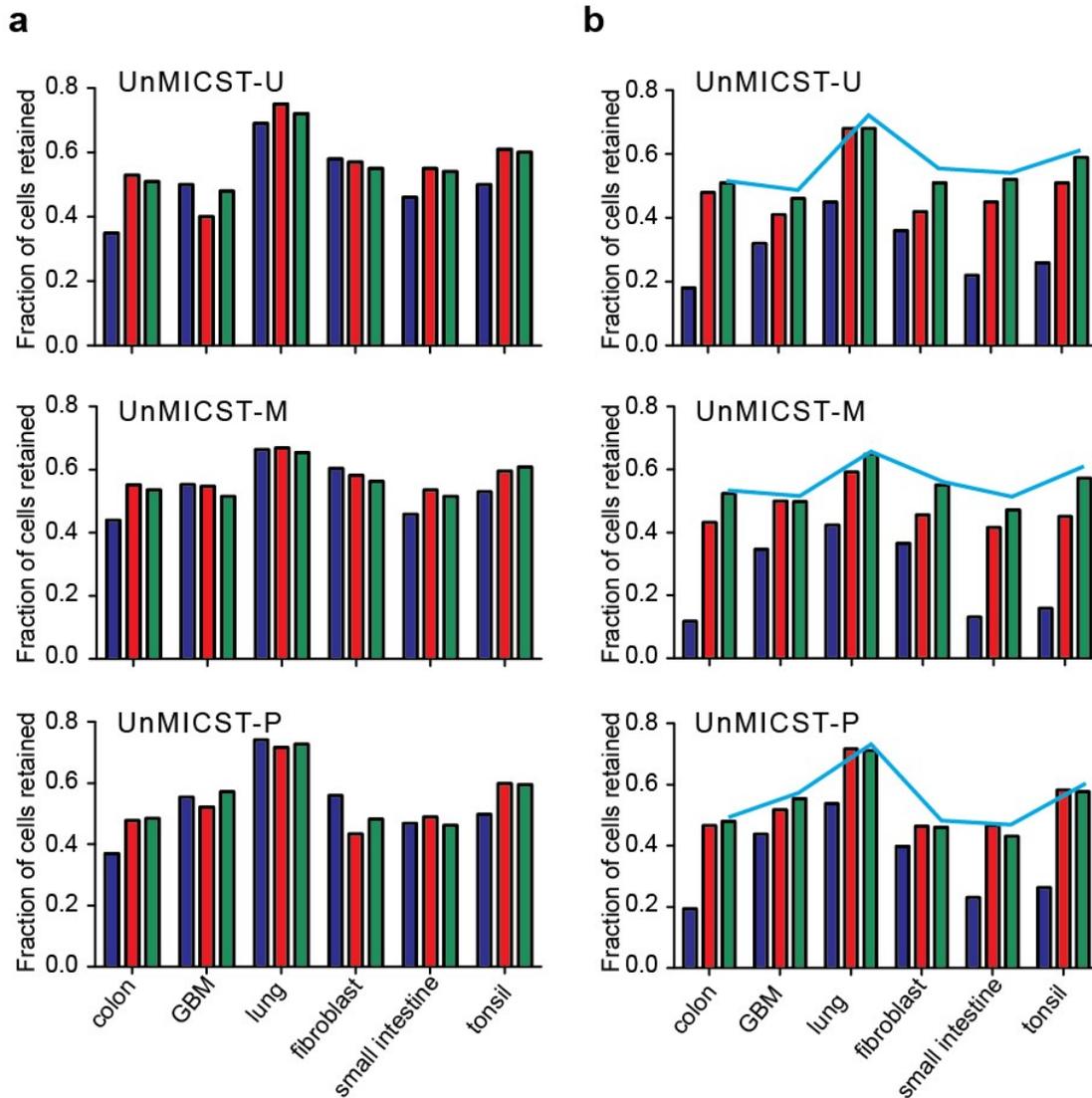
751

752

753

754

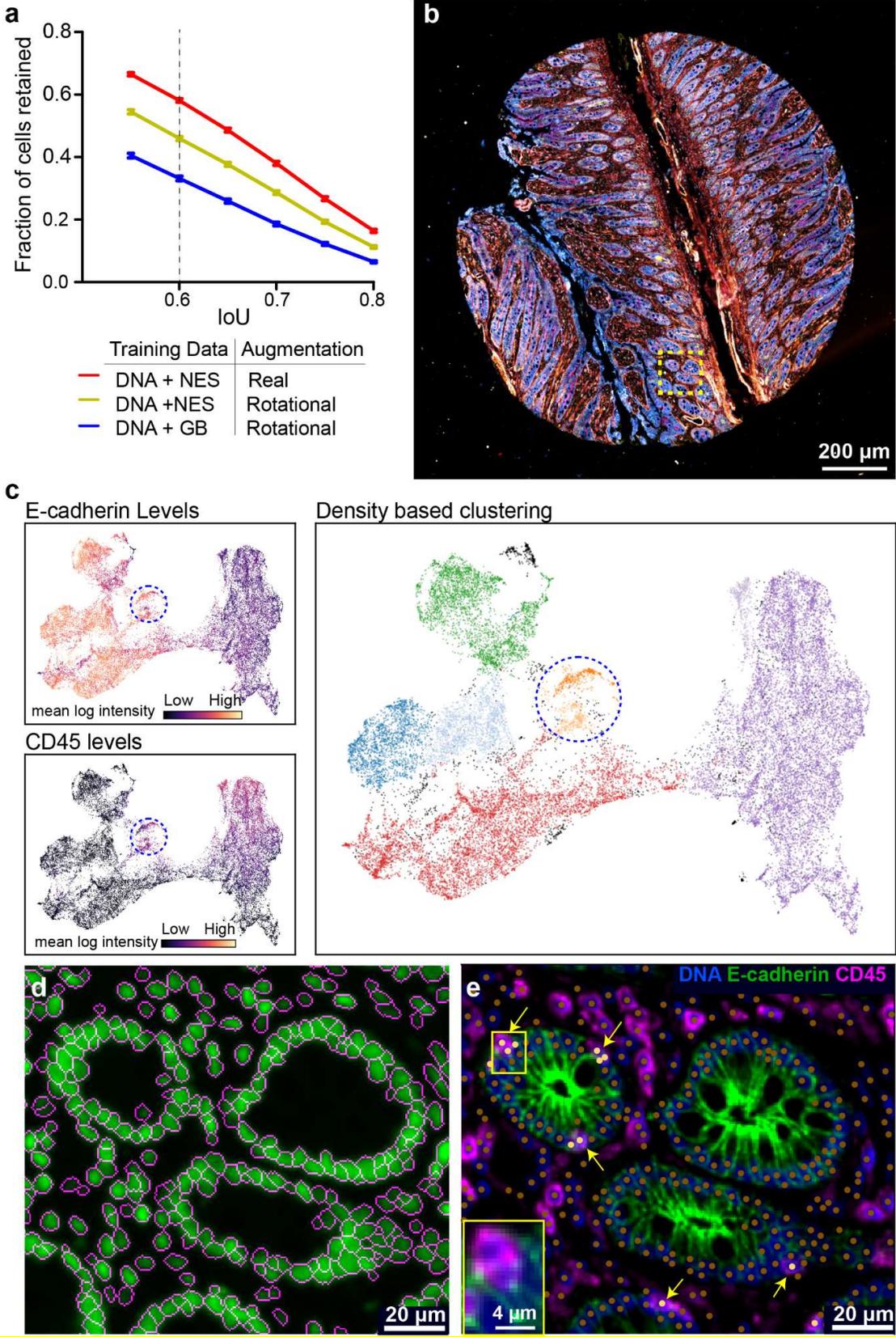
**Figure 5**



755

756 **Figure 5: Assessing different training strategies on a) in-focus and b) defocused test data for**  
 757 **different tissue types. a)** In all tissue types apart from GBM, the addition of NES (red bars) and the  
 758 use of real augmentations combined with NES (green bars) in training data offered superior  
 759 accuracy compared to using DNA alone (blue bars). **b)** When the models were tested on defocused  
 760 data, all tissues (including GBM unexpectedly) showed benefits resulting from using NES (red bars)  
 761 combined with real augmentations (green bars). The line plot indicates highest accuracy achieved  
 762 for each tissue when tested on in-focus data from panel a.

Figure 6



764 **Figure 6: Applying UnMICST models to highly multiplexed image data. a)** Accuracy  
765 improvement of UnMICST models trained with and without NES (nuclear envelope staining) as  
766 compared to DNA alone, and real augmentations as compared to computed blur (GB; Gaussian  
767 blur). To balance training dataset size, GB was substituted for NES data and computed 90/180  
768 degree rotations were substituted for real augmentations. **b)** A 64-plex CyCIF image of a non-  
769 neoplastic small intestine TMA core from the EMIT dataset. Dashed box indicates region of interest  
770 for panels **d** and **e**. **c)** UMAP projection using single cell staining intensities for 14 marker proteins  
771 (see methods). The color of the data points represents the intensity of E-cadherin (top left) or CD45  
772 (bottom left) across all segmented nuclei. Density-based clustering using HDBSCAN identified  
773 distinct clusters (each denoted by a different color) that were positive for either E-cadherin or CD45  
774 as well as a small number of double-positive cells (blue dashed circle). **d)** Enlarged region of yellow  
775 dashed box from **b** showing segmentation mask outlines (magenta) overlaid onto DNA channel  
776 (green). **e)** Composite image of DNA, E-cadherin, and CD45 of the same region. Nuclei centroids  
777 from segmentation denoted by brown dots. Cells positive for both E-cadherin and CD45 (from blue  
778 dashed circle in panel **c** are marked with yellow arrows and yellow dots. Inset: enlarged view of  
779 boxed region showing overlapping immune and epithelial cells.  
780

## Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- [SupplementaryMaterials.pdf](#)