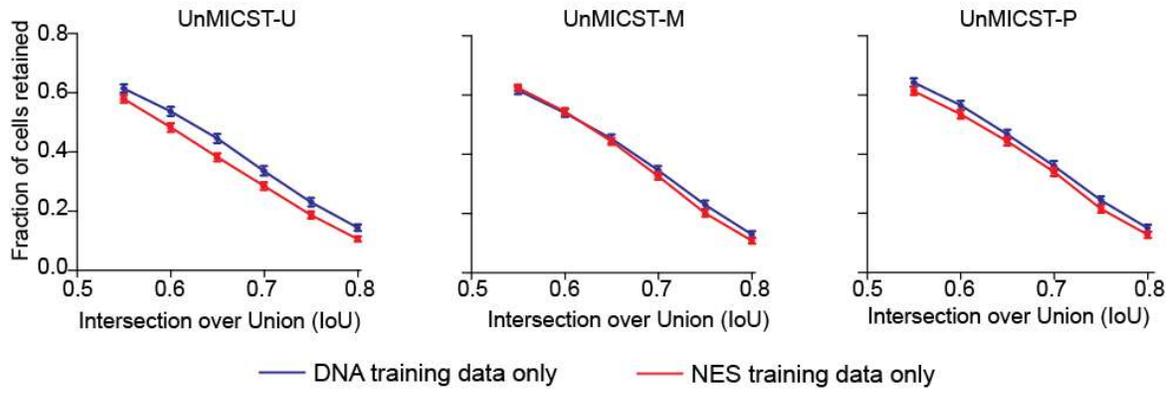


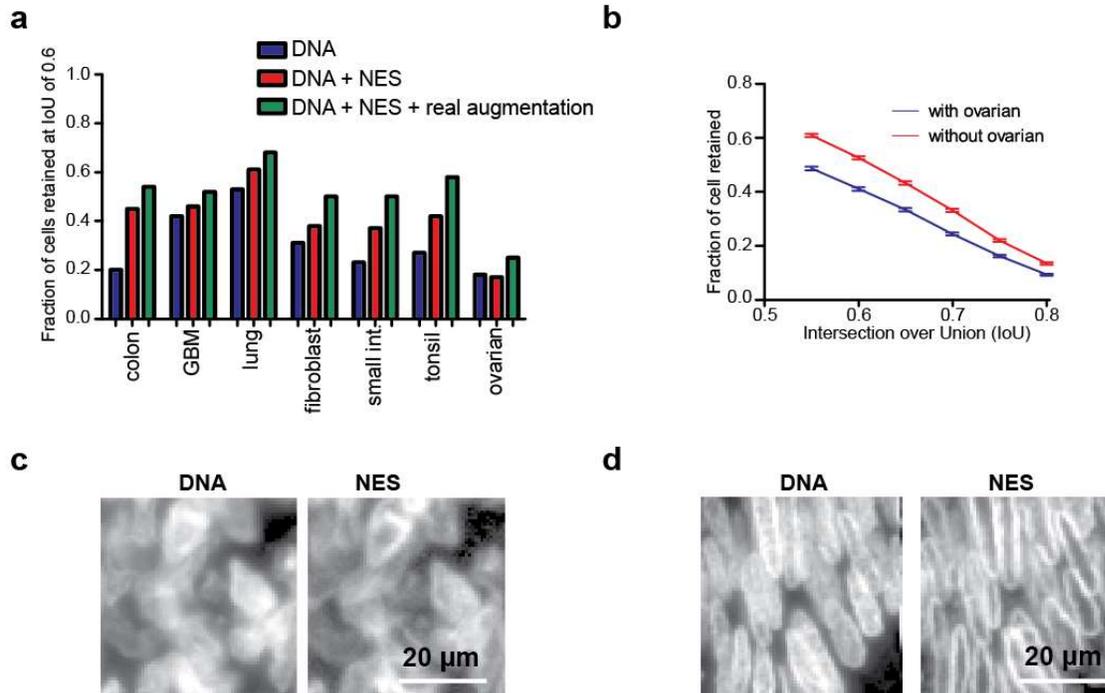
1 **SUPPLEMENTARY MATERIALS**  
**Supplementary Figure 1**



2 — DNA training data only — NES training data only  
3 **Supplementary Figure 1: NES data does not substitute for DNA in model training.** Test  
4 results when training was performed on DNA (blue curve) and lamin (red curve) individually as  
5 single channels for a) UnMICST-U, b) UnMICST-M, and c) UnMICST-P.

6  
7  
8  
9  
10  
11  
12  
13  
14  
15  
16  
17  
18  
19  
20  
21  
22  
23

## Supplementary Figure 2



24

25 **Supplementary Figure 2: Assessing the use of normal ovarian data in model training. a)**

26 Normal ovarian tissue performed the worst with respect to segmentation accuracy among all

27 annotated tissues. Using NES and real augmentations conferred a modest benefit. **b)** The

28 segmentation accuracy across all tissues was lower when ovarian data was included in the

29 training set (blue curve) as opposed to excluded (red curve). **c)** Images showing that annotation

30 of normal ovarian nuclei from DNA is a challenging task, even with NES staining at four times

31 the nominal resolution (0.325 microns per pixel and with a 20x/0.75 objective lens). **d)** In

32 contrast, colon was straightforward to annotate, particularly with NES staining, which forms a

33 distinct halo around the nuclear periphery.

34

35

36

37

38

39

40

41

42

43 **Supplementary Table 1: High-plex tissue imaging methods**

44

Optical	Fluorophore-based	Single staining cycle but with sequential data acquisition	CO-Detection by indEXing (CODEX) <sup>2</sup> Immunostaining with Signal Amplification By Exchange Reaction (Immuno-SABER) <sup>6</sup>
		Cyclic staining and sequential data detection	Cyclic Immunofluorescence (CyCIF) <sup>7</sup> Multiplex immunofluorescence (MxIF) <sup>3</sup> Iterative indirect immunofluorescence imaging (4i) <sup>9</sup>
	Enzyme-based		Multiplex Immunohistochemistry (mIHC) <sup>4</sup> Multiplexed immunohistochemical consecutive staining on single slide (MICSSS) <sup>8</sup>
Non-optical	Detection based on atomic mass spectrometry of metal-labelled antibodies		Imaging Mass Cytometry (IMC) <sup>1</sup> Multiplexed Ion Beam Imaging (MIBI) <sup>5</sup>

45

46

47

48

49

50

51

52

53

54

55

56

57

58

59 **Supplementary Table 2:** Data set size and composition across disjoint training and testing data  
60 splits. The model used for training is indicated in the left-hand column and was identical across  
61 all three segmentation architectures.

62

<b>Model</b>	<b>Training Set Size</b>	<b>Test Set Size</b>
In-focus DNA	3636	217
In-focus DNA + NES	3636	217
In-focus DNA + Real Augmentations	21,816	1,302
In-focus DNA + NES + Real Augmentations	21,816	1,302
In-focus DNA + Gaussian blur	14,544	868
DNA + Real Augmentations	5,100	645
DNA + 90°/180° rotations	5,100	645
DNA + NES + Real Augmentations	5,100	645
DNA + NES + 90°/180° rotations	5,100	645

63

## 64 **Supplementary Note 1: Historical information on segmentation of microscopy images**

65 One of the earliest methods to extract cells from the background involves specifying a  
66 global threshold based on the Otsu method<sup>37</sup> in which a threshold is selected that maximizes  
67 intraclass variance between two classes. For cell objects that remain clumped and unseparated  
68 after simple thresholding, the Otsu method is usually replaced with marker-controlled watershed  
69 segmentation. This entails a distance transform operation on the clumped objects to identify a  
70 single central 'seed' point per object. The presence of multiple seed points per object (cell)  
71 results in oversegmentation and this can usually be resolved with blurring followed by a regional  
72 maxima operation. Other methods to identify markers include, but are not limited to, graph cuts  
73 or filtering with a blob detector (usually a Laplacian of Gaussian filter)<sup>38-40</sup>, level sets followed by  
74 mean-shift clustering<sup>41</sup>, or a filter bank of rings<sup>42</sup>. Many of these methods have been applied to  
75 both fluorescence and H&E images of tissues and while ground truth labeled datasets are not  
76 required, fairly extensive parameter tuning and empirical testing is necessary.

77 A variety of neural network, and other sophisticated, architectures have been explored to  
78 improve detection and classification accuracy and expand generalizability to new types of  
79 images. Several deep learning architectures developed for natural images have been adapted  
80 for marker detection in images of cells including Fully Convolutional Networks (FCNs)<sup>43</sup>, Visual  
81 Geometry Group (VGG16)<sup>44,45</sup>, Residual Networks (ResNets)<sup>46</sup>, UNet<sup>20,47-52</sup>, and Mask R-  
82 CNN<sup>53,54,54-57</sup>. In classical image analysis, advances in methodology commonly involve the  
83 development of new algorithms; any changes in parameter settings needed to accommodate  
84 new data are regarded as project-specific details. In contrast, in machine learning approaches,  
85 advances in algorithms, learned models and labelled data are all significant; this is particularly  
86 true as differentiable architectures for deep learning become increasingly standardized and the  
87 quality of trained models is a key point of comparison. While machine learning algorithms  
88 outperform classical methods in many image processing applications, model training and  
89 evaluation remain substantial challenges in the biomedical domain.

### 90 *Biomedical Image datasets*

91 Image recognition models are commonly trained on diverse sets of natural images, and  
92 transfer-style learning is used to improve performance on specific domains. The introduction of  
93 the ImageNet dataset<sup>58</sup> catalyzed the application of deep learning to image analysis by  
94 providing a set of real-world images along with ground truth annotations. Subsequent  
95 applications of ImageNet make clear that machine learning models are only as strong as the  
96 underlying training and validation data. In biomedical research, segmentation methods for cell  
97

98 lines<sup>10,11,56,59</sup> have benefited from large public datasets such as Expert Visual Cell ANnotation  
99 (EVICAN)<sup>12</sup>, Cellpose<sup>60</sup>, the Broad Bioimage Benchmark Collection (BBBC)<sup>61</sup>, The Cancer  
100 Genome Atlas (TCGA)<sup>62</sup>, and Kaggle datasets<sup>63</sup>.

101 Few human-labelled datasets are currently available for highly multiplexed tissue images  
102 and existing learned segmentation models generally apply only to specific tissue types<sup>11,38–42,64–</sup>  
103 <sup>67</sup>. Moreover, in most existing datasets, images comprise H&E and not fluorescence  
104 images. One exception is a broad 23-tissue study that is part of TCGA, however, these  
105 annotations are not publicly available. Moreover, because only the center of the nuclei (as  
106 opposed to the nuclear boundary) was marked, TCGA annotations are more suited to cell  
107 counting as opposed to segmentation. To address this issue, this current manuscript includes a  
108 set of tissue images and densely labelled ground-truth annotations of whole nuclei.

109

### 110 *Image augmentation to improve model training*

111 Deep learning models have a high capacity to learn large numbers of features. This  
112 results in high accuracy but with a substantial danger of overfitting. The use of augmentation  
113 methods helps to remedy overfitting by increasing the diversity of the training data<sup>17</sup>. The most  
114 common method employs a combination of translational shifts, 90 degree rotations, reflections,  
115 and affine transformations<sup>43,51,56,68</sup>. A few studies have also added elastic deformations using B-  
116 splines<sup>8,59,69</sup>. These methods are not unique to microscopy images, however, and only a few  
117 studies have used augmentation to address variation in the brightness and contrast of otherwise  
118 identical images<sup>52</sup> or added synthetically generated camera noise and non-cellular debris to  
119 make model training less sensitive to artefacts<sup>49,63</sup>. A particularly interesting form of  
120 augmentation used by Kromp *et al.* (2019) involved manually separating cells from the  
121 background and arranging nuclei in grids with random positions and orientations, effectively  
122 generating new training examples. The authors found that imposing a grid structure did not  
123 substantially improve accuracy because cut-out nuclei had hard edges that are atypical of  
124 fluorescence images. The authors, therefore, used a generative model trained to relate images  
125 of nuclei to their artificial counterparts – nuclei that had been cut out and placed back on their  
126 original positions.

127 To date, common artefacts such as image saturation and defocus have been addressed  
128 computationally using image histogram modification and Gaussian blurring<sup>19,63</sup> as well as  
129 deliberate defocusing of an actual microscope<sup>61</sup>. The BBBC described by Ljosa *et al.* is primarily  
130 derived from tissue culture cells or transmission and differential interference contrast (DIC)  
131 microscopy of model organisms as opposed to human tissues. Ground truth information in Ljosa

132 et al. was also generated automatically as opposed to by human annotation, where the latter is  
133 likely to be more accurate.

#### 134 *Use of stains to aid in segmentation*

135 In fluorescence microscopy, nuclei are most commonly stained using intercalating dyes  
136 such as DAPI or Hoechst 33342<sup>49,54,64,70</sup>, SiR-DNA<sup>63</sup>, TO-PRO<sup>38</sup>, and hematoxylin<sup>39-41,45,46,51,64,66</sup>.  
137 When expression of recombinant proteins is feasible (e.g., in cell lines), cells expressing  
138 fluorescent protein fusions to histones<sup>71</sup> or spindle components is an effective means to label  
139 nuclei<sup>44,50</sup>; this has also been done in genetically engineered mouse models<sup>72</sup>, but is not  
140 relevant to the analysis of human tissues. In the case of brightfield or phase contrast imaging,  
141 nuclear labels are not utilized<sup>12,73,74</sup>, but nuclei can often be identified based on brightness.  
142 Almost all of these studies have used data from a single imaging channel for nuclei localization,  
143 which can be problematic when nuclei are diffuse or close together, both of which are common  
144 in cancer specimens. In general, the results of segmentation are superior using fluorescence as  
145 opposed to brightfield data<sup>13</sup>, but the use of additional channels to more precisely define nuclear  
146 boundaries by staining for nuclear lamins or nucleoporins, for example, has not been broadly  
147 explored.