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

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

[1]Laboratory of Systems Pharmacology, Harvard Medical School, Boston, MA 02115
[2]Image and Data Analysis Core, Harvard Medical School, Boston, MA 02115
Human Tumor Atlas Network
[3]School of Engineering and Applied Sciences, Harvard University, Cambridge, MA, 02138
[4]Ludwig Center for Cancer Research at Harvard, Harvard Medical School, Boston, MA, 02115
[5]Department of Pathology, Brigham and Women's Hospital, Harvard Medical School, Boston, MA, 02115
[6]Department of Systems Biology, Harvard Medical School, Boston, MA, 02115

ORCID IDS

Clarence Yapp clarence@hms.harvard.edu 0000-0003-1144-5710
Edward Novikov edward_novikov@seas.harvard.edu 0000-0003-1476-5111
Won-Dong Jang wdjang@seas.harvard.edu 0000-0002-5205-1024
Yu-An Chen yu-an_chen@hms.harvard.edu 0000-0001-7228-4696
Marcelo Cicconet marcelo_cicconet@hms.harvard.edu 0000-0003-2649-1509
Zoltan Maliga zoltan_maliga@hms.harvard.edu 0000-0003-4209-7253
Connor A. Jacobson connor_jacobson@hms.harvard.edu 0000-0002-0299-3451
Donglai Wei donglai@seas.harvard.edu 0000-0002-2329-5484
Sandro Santagata ssantagata@bics.bwh.harvard.edu 0000-0002-7528-9668
Hanspeter Pfister pfister@seas.harvard.edu 0000-0002-3620-2582
Peter K. Sorger peter_sorger@hms.harvard.edu, 0000-0002-3364-1838

ABSTRACT

Newly developed technologies have made it feasible to routinely collect highly multiplexed (20-60 channel) images at subcellular resolution from human tissues for research and diagnostic purposes. Extracting single cell data from such images requires efficient and accurate image segmentation. This starts with identification of nuclei, a challenging problem in tissue imaging that has recently benefited from deep learning. In this paper we demonstrate two generally applicable approaches to improving segmentation accuracy as scored using new human-labelled segmentation masks spanning multiple human tissues. The first approach involves the use of "real augmentations" during training. These comprise defocused and saturated image data and improve model accuracy when computational augmentation (Gaussian blurring) does not. The second involves collection of nuclear envelope data. The two approaches cumulatively and substantially improve segmentation with three different deep learning frameworks, yielding a set of high accuracy segmentation models. Moreover, the use of real augmentation may have applications outside of microscopy.

1. INTRODUCTION

Tissues are complex assemblies of multiple cell types that interact with each other and with structures such as basement membranes and connective tissue in defined geometric arrangements. In diseases such as cancer, disruption of cell-intrinsic regulatory mechanisms and the tissue environment cause morphological and structural changes over spatial scales ranging from subcellular organelles to entire tissue (0.1 to 10^4 µm). Microscopy using Hematoxylin and Eosin (H&E), Romanowsky–Giemsa stains, and other colorimetric dyes, complemented by immunohistochemistry (Immunologists, 1942) has long played the primary role in the study of tissue architecture in humans and other organisms (Albertson, 2006; Shlien and Malkin, 2009). In a clinical setting, histopathology remains the primary means by which diseases such as cancer are staged and managed clinically (Amin *et al.*, 2017). However, classical histology provides insufficient molecular information to precisely identify cell subtypes, study mechanisms of development, and characterize disease genes. In the past few years, several different methods for highly multiplexed tissue imaging have been developed. These yield subcellular resolution data on the abundance of 20-60 antigens, which is sufficient to subtype cells, measure cell states (quiescent, proliferating, dying, etc.) and interrogate cell signaling pathways (Rozenblatt-Rosen *et al.*, 2020a). Methods based on immunofluorescence imaging include cyclic immunofluorescence (CyCIF) (Lin *et al.*, 2018), Multiplexed Immunofluorescence (MxIF) (Gerdes *et al.*, 2013), CO-Detection by indEXing (CODEX) (Goltsev *et al.*, 2018), and Signal Amplification By Exchange Reaction (immuno-SABER) (Saka *et al.*, 2019). Mass spectrometry-based methods include Imaging Mass Cytometry (IMC) (Giesen *et al.*, 2014) and Multiplexed Ion Beam Imaging (MIBI) (Angelo *et al.*, 2014). These methods differ in resolution, field of view, and multiplicity ("plex") but all generate 2D images of tissue sections that are typically 5 – 10 µm thick.

When multiplexed images are segmented and quantified, the resulting single cell data are a natural complement to single cell RNA Sequencing (scRNASeq) data, which have had a dramatic

impact on our understanding of normal and diseased tissues (Achim *et al.*, 2015; Slyper *et al.*, 2020). Unlike dissociated RNASeq, however, multiplex tissue imaging preserves spatial information. Single cell analysis of data acquired via imaging requires segmentation, a computer vision technique that assigns class labels to an image in a pixel-wise manner to subdivide it; this is followed by marker quantification on a per-cell or per organelle basis. Extensive work has gone into the development of methods for segmenting metazoan cells grown in culture, but segmentation of tissue images is a substantially more complex challenge. Cell types, and thus cell sizes and shapes, are more diverse in tissues; cells are often closely packed and may be only partly captured by any individual section. Segmentation of tissue culture cells has long relied on classical approaches such as the watershed algorithm, whose parameters are manually tuned for different cell types. More recently, machine learning methods have become standard, paralleling the widespread use of convolutional neural networks (CNNs) in image recognition, object detection, and synthetic image generation (LeCun, Bengio and Hinton, 2015).

CNNs are well-suited to image analysis since they automatically extract image features at multiple scales at each layer of the neural network. Architectures such as ResNet, VGG16, and more recently, UNet and Mask R-CNN (Ronneberger, Fischer and Brox, 2015; He *et al.*, 2018) have gained considerable attention for their ability to learn millions of parameters and generalize across datasets, as evidenced by excellent performance in a wide range of segmentation competitions, as well as in hackathon challenges (Caicedo *et al.*, 2019) using publicly available image sets (Kromp *et al.*, 2020; Schwendy, Unger and Parekh, 2020). UNet, in particular, has gained popularity in the biomedical imaging domain due to its ease of deployment on Graphical Processing Units (GPUs) and its superior performance on previously challenging cell segmentation and tracking datasets. One limitation of machine learning on tissue images is a lack of sufficient freely-available data with ground truth labelling. Experience with natural scene images (Ronneberger, Fischer and Brox, 2015) has proven that the acquisition of labels can be time consuming and rate limiting (Gurari *et al.*, 2015).

In both cultured cells and tissues, the localization of nuclei is an optimal starting point for segmenting cells since most cell types have one nucleus (cells undergoing mitosis, muscle and liver cells and osteoclasts are important exceptions), and nuclear stains with high signal-to-background ratios are widely available. The nucleus is generally quite large (relative to the resolution of wide-field fluorescence microscopes), making it easy to detect at multiple magnifications, and it often lies at the approximate center of a cell. Nuclei can be identified in the absence of staining (Chalfoun *et al.*, 2014; Stylianidou *et al.*, 2016; Lux and Matula, 2020) but for fluorescence imaging, labeling with DAPI (4′,6-diamidino-2-phenylindole) or Hoechst 33342 is common; in transfected cells or transgenic animals, fluorescently tagged proteins such as histones can also be used (Al-Kofahi *et al.*, 2018; Schmidt *et al.*, 2018; Vuola, Akram and Kannala, 2019; Wang *et al.*, 2019; Kromp *et al.*, 2020). In classical histology, nuclei are labelled along with other structures using chromogenic dyes such as H&E (Al-Kofahi *et al.*, 2010; Qi *et al.*, 2012; Xu, Lu and Mandal, 2014; Chen *et al.*, 2016; Xu *et al.*, 2017; Lee and Jeong, 2020; Liu *et al.*, 2020; Shahzad M *et al.*, 2020).

Although segmentation is commonly performed using nuclei alone, there are possible advantages to using other markers in combination with nuclear stains; for example, Schüffler *et al.*, (2015) used multiplexed IMC data and watershed methods for multi-channel segmentation. Methods based on random forests such as Ilastik and Weka (Arganda-Carreras *et al.*, 2017; Berg *et al.*, 2019) also exploit multiple channels for class-wise pixel classification via an ensemble of decision trees to assign pixel-wise class probabilities in an image. However, random forest models have significantly less capacity for learning than CNNs, which is a substantial disadvantage. The possibility of using CNNs with multi-channel data to enhance nuclei segmentation has not been widely explored.

The accuracy of segmentation algorithms is crucially dependent on the quality of the original images and the absence of artefacts. In practice however, images commonly have at least some unavoidable focus or blur artefacts and saturated channels. This is particularly true in the case of whole slide imaging in which up to 1,000 sequentially acquired image tiles are used to create mosaic images of specimen as large as several square centimeters. Segmentation methods must compensate for aberrations commonly occurring in such large images. One way to accomplish this is via image augmentation. Image augmentation (Krizhevsky, Sutskever and Hinton, 2012) is a method commonly used in the expansion of training sets for machine learning tasks in which a set of original images are pre-processed to include random rotation, shearing, flipping, etc. This is designed to prevent algorithms from learning irrelevant aspects of an image, such as orientation. Image augmentation is also a method for generating training data for images with specific types of artefacts including focus problems. To date this has been accomplished by applying various degrees of Gaussian blur to the training data (Ahmed Raza *et al.*, 2017; Shorten and Khoshgoftaar, 2019; Horwath *et al.*, 2020).

1.1 Related work

One of the earliest methods to extract cells from the background involves specifying a global threshold based on the Otsu method (Otsu, 1979) in which a threshold is selected that maximizes intraclass variance between two classes. For cell objects that remain clumped and unseparated after simple thresholding, the Otsu method is usually replaced with marker-controlled watershed segmentation. This entails a distance transform operation on the clumped objects to identify a single central ‘seed’ point per object. The presence of multiple seed points per object (cell) results in oversegmentation and this can usually be resolved with blurring followed by a regional maxima operation. Other methods to identify markers include, but are not limited to, graph cuts or filtering with a blob detector (usually a Laplacian of Gaussian filter) (Byun *et al.*, 2006; Xu, Lu and Mandal, 2014; Xu *et al.*, 2017), level sets followed by mean-shift clustering (Qi *et al.*, 2012), or a filter bank of rings (Liu *et al.*, 2016). Many of these methods have been applied to both fluorescence and H&E images of tissues and while ground truth labeled datasets are not required, fairly extensive parameter tuning and empirical testing is necessary.

A variety of neural net and other architectures have been explored to improve detection and classification accuracy and expand generalizability to new types of images. Several deep learning architectures developed for natural images have been adapted for marker detection in images of cells including Fully Convolutional Networks (FCNs) (Lux and Matula, 2020), Visual Geometry Group (VGG16) (Wang *et al.*, 2019; Shahzad M *et al.*, 2020), Residual Networks (ResNets) (Lee and Jeong, 2020), UNet (Al-Kofahi *et al.*, 2018; McQuin *et al.*, 2018; Schmidt *et al.*, 2018; Wen *et al.*, 2018; Vu *et al.*, 2019; Horwath *et al.*, 2020; Lugagne, Lin and Dunlop, 2020), and Mask R-CNN (Kromp *et al.*, 2019; Vuola, Akram and Kannala, 2019, 2019; Korfhage *et al.*, 2020; Liu *et al.*, 2020; Masubuchi *et al.*, 2020). In classical image analysis, advances in methodology commonly involve the development of new algorithms; any changes in parameter settings needed to accommodate new data are regarded as project-specific details. In contrast, in machine learning approaches, advances in algorithms, learned models and labelled data are all significant; this is particularly true as differentiable architectures for deep learning become increasingly standardized and the quality of trained models is a key point of comparison. While machine learning algorithms outperform classical methods in many image processing applications, model training and evaluation remain substantial challenges in the biomedical domain.

1.1.1 Biomedical Image datasets

Image recognition models are commonly trained on diverse sets of natural images, and transfer-style learning is used to improve performance on specific domains. The introduction of the ImageNet dataset (Deng *et al.*, 2009) catalyzed the application of deep learning to image analysis by providing a set of real-world images along with ground truth annotations. Subsequent applications of ImageNet make clear that machine learning models are only as strong as the underlying training and validation data. In biomedical research, segmentation methods for cell lines (Caicedo *et al.*, 2019; Kromp *et al.*, 2020; Liu *et al.*, 2020; Torr *et al.*, 2020) have benefited from large public datasets such as Expert Visual Cell ANnotation (EVICAN) (Schwendy, Unger and Parekh, 2020), Cellpose (Stringer *et al.*, 2020), the Broad Bioimage Benchmark Collection (BBBC) (Ljosa, Sokolnicki and Carpenter, 2012), The Cancer Genome Atlas (TCGA) (Xing *et al.*, 2019), and Kaggle datasets (Yang *et al.*, 2020).

Few human-labelled datasets are currently available for highly multiplexed tissue images and existing learned segmentation models generally apply only to specific tissue types (Byun *et al.*, 2006; Al-Kofahi *et al.*, 2010; Qi *et al.*, 2012; Xu, Lu and Mandal, 2014; Roth *et al.*, 2015; Chen *et al.*, 2016; Liu *et al.*, 2016; Xu *et al.*, 2017; Höfener *et al.*, 2018; Kromp *et al.*, 2020). Moreover, in most existing datasets, images comprise H&E and not fluorescence images. One exception is a broad 23-tissue study that is part of TCGA, however, these annotations are not publicly available. Moreover, because only the center of the nuclei (as opposed to the nuclear boundary) was marked, TCGA annotations are more suited to cell counting as opposed to segmentation. To address this issue, this current manuscript includes a set of tissue images and densely labelled ground-truth annotations of whole nuclei.

1.1.2 Image augmentation to improve model training

Deep learning models have a high capacity to learn large numbers of features. This results in high accuracy but with a substantial danger of overfitting. The use of augmentation methods helps to remedy overfitting by increasing the diversity of the training data (Krizhevsky, Sutskever and Hinton, 2012). The most common method employs a combination of translational shifts, 90 degree rotations, reflections, and affine transformations (Vu *et al.*, 2019; Khadangi, Boudier and Rajagopal, 2020; Liu *et al.*, 2020; Lux and Matula, 2020). A few studies have also added elastic deformations using B-splines (Ronneberger, Fischer and Brox, 2015; Raza *et al.*, 2019; Torr *et al.*, 2020). These methods are not unique to microscopy images, however, and only a few studies have used augmentation to address variation in the brightness and contrast of otherwise identical images (Lugagne, Lin and Dunlop, 2020) or added synthetically generated camera noise and non-cellular debris to make model training less sensitive to artefacts (Schmidt *et al.*, 2018; Yang *et al.*, 2020). A particularly interesting form of augmentation used by Kromp *et al.* (2019) involved manually separating cells from the background and arranging nuclei in grids with random positions and orientations, effectively generating new training examples. The authors found that imposing a grid structure did not substantially improve accuracy because cut-out nuclei had hard edges that are atypical of fluorescence images. The authors, therefore, used a generative model trained to relate images of nuclei to their artificial counterparts – nuclei that had been cut out and placed back on their original positions.

To date, common artefacts such as image saturation and defocus have been addressed computationally using image histogram modification and Gaussian blurring (Shorten and Khoshgoftaar, 2019; Yang *et al.*, 2020) as well as deliberate defocusing of an actual microscope (Ljosa, Sokolnicki and Carpenter, 2012). The BBBC described by Ljosa et al. is primarily derived from tissue culture cells or transmission and differential interference contrast (DIC) microscopy of model organisms as opposed to human tissues. Ground truth information in Ljosa et al. was also generated automatically as opposed to by human annotation, where the latter is likely to be more accurate.

1.1.3 Use of stains to aid in segmentation

In fluorescence microscopy, nuclei are most commonly stained using intercalating dyes such as DAPI or Hoechst 33342 (Al-Kofahi *et al.*, 2010; Schmidt *et al.*, 2018; Berryman *et al.*, 2019; Vuola, Akram and Kannala, 2019), SiR-DNA (Yang *et al.*, 2020), TO-PRO (Byun *et al.*, 2006), and hematoxylin (Al-Kofahi *et al.*, 2010; Qi *et al.*, 2012; Xu, Lu and Mandal, 2014; Chen *et al.*, 2016; Xu *et al.*, 2017; Vu *et al.*, 2019; Lee and Jeong, 2020; Shahzad M *et al.*, 2020). When expression of recombinant proteins is feasible (e.g., in cell lines), cells expressing fluorescent protein fusions to histones (Challen and Goodell, 2008) or spindle components is an effective means to label nuclei (Wen *et al.*, 2018; Wang *et al.*, 2019); this has also been done in genetically engineered mouse models (Tumbar *et al.*, 2004) but is not relevant to analysis of human tissues. In the case of

brightfield or phase contrast imaging, nuclear labels are not utilized (Chalfoun *et al.*, 2014; Stylianidou *et al.*, 2016; Schwendy, Unger and Parekh, 2020), but nuclei can often be identified based on brightness. Almost all of these studies have used data from a single imaging channel for nuclei localization, which can be problematic when nuclei are diffuse or close together, both of which are common in cancer specimens. In general, the results of segmentation are superior using fluorescence as opposed to brightfield data (Gurari *et al.*, 2015), but the use of additional channels to more precisely define nuclear boundaries by staining for nuclear lamins or nucleoporins, for example, has not been broadly explored.

1.2 Contributions

In this paper, we investigate ways to maximize the accuracy of image segmentation on multiplexed tissue data by including different types of information in the images and by augmenting the training data. We generate a set of training and test data with ground-truth annotations via human inspection of lung adenocarcinoma, prostate-derived fibroblasts, normal tonsil, non-neoplastic small intestine, colon adenocarcinoma, glioblastoma, and non-neoplastic ovary, and use these data to score segmentation accuracy achieved by using three deep learning algorithms: UNet, Mask R-CNN, and Pyramid Scene Parsing Network (PSPNet). We introduce a family of models termed Universal Models for Identifying Cells and Segmenting Tissue (UnMICST). The three frameworks presented here are referred to based on the algorithms they use as UnMICST-U, UnMICST-M, and UnMICST-P, respectively. We identify two generalizable and easily implemented ways of improving segmentation accuracy. The first involves collecting images of nuclear envelope staining (NES) acquired using a cocktail of antibodies against lamins and the nuclear envelope protein NUP98 as a complement to standard images of chromatin acquired using DNA-intercalating dyes. The second involves adding real augmentations, defined here as intentionally defocused and over-saturated images, to the training data to make models more robust to the types of artefacts encountered in real tissue images.

We find that augmentation with real data significantly outperforms conventional Gaussian blur augmentation, offering a statistically significant improvement in model robustness to defocused artefacts. In general, the use of additional nuclear markers and real augmentation improves segmentation across multiple tissue types and generalizes to each of the three tested frameworks. Thus, nuclei segmentation models should use staining data for DNA (e.g., Hoechst 33342) and NES as well as data with real artefacts. All of the data and models in this paper are made freely available for further research (www.synapse.org/#!Synapse:syn24192218/files/). A list of key contributions is as follows:

- Generate and release a range of multiplexed tissue and human-generated ground truth annotations for nuclear segmentation
- Demonstrate that combined use of two markers for identifying nuclei involving a DNA stain to visualize chromatin and antibodies to stain the nuclear envelope improves segmentation accuracy in multiple settings

- Show that segmentation models trained on data containing real artefacts (defocused and saturated pixels) outperform models solely based on computational augmentation
- Demonstrate that these findings generalize to three different deep learning architectures (UNet, Mask R-CNN, and PSPNet)

2. Method

2.1 Sample preparation for imaging

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

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

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

2.2 Acquisition of image data and real augmentations

The stained TMA was imaged on a INCell 6000 (General Electric Life Sciences) microscope equipped with a 20x/0.75 objective lens (370 nm nominal lateral resolution at 550 nm wavelength) and a pixel size of 0.325 µm per pixel. Hoechst and lamin-A647 were excited with a 405 nm and 642 nm laser, respectively. Emission was collected with the “DAPI” (455/50 nm) and “Cy5” (682/60 nm) filter sets with exposure times of 60 ms and 100 ms, respectively. Whole-slide imaging involved acquisition of 1,215 tiles with an 8% overlap, which is recommended for stitching in ASHLAR, a next generation stitching and registration algorithm for large images (https://github.com/labsyspharm/ashlar). To generate defocused data, we acquired images from above and below the focal plane by varying the Z-axis by 3 µm in both directions. To generate

saturated images of DNA staining, a 150ms exposure time was used. These two types of "suboptimal" data were then used for "real augmentation" during model training, as described below.

Representative cores for lung adenocarcinoma, non-neoplastic small intestine, normal prostate, colon adenocarcinoma, glioblastoma, non-neoplastic ovary, and tonsil were extracted from image mosaics and down-sampled by a factor of 2 to match the pixel size of images routinely acquired and analyzed in MCMICRO (Schapiro *et al.*, 2021). Images were then cropped to 256 x 256-pixel tiles, and in-focus DNA and NES were imported into Adobe Photoshop to facilitate human annotation of nuclear boundaries. Annotations for contours and background classes were labelled on separate layers while swapping between DNA and NES as necessary. To save time, we drew complete contours of nuclei and filled these in using the Matlab *imfill* operation to generate nuclei centers. For nuclei at the image borders where contours would be incomplete, we manually annotated nuclei centers. As described by Ronneberger et al. (2015), a fourth layer was used to mark areas between clumped cells. These additional annotations made it possible to specifically penalize models that incorrectly classified these pixels.

Because original, defocused, and saturated images of DNA were all acquired in the same image stack, it was possible to use a single registered set of DNA annotations across all augmented image channels. To produce the training set, each image was cropped into 64 x 64 patches, normalized to use the full dynamic range, and further augmented using 90-degree rotations, reflections, and 20% upscaling. Consistent with the training set, the validation and test sets also include defocused and saturated examples but were not augmented with standard transformations. The ratio of data examples present in the training, validation, and test set split was 0.36:0.24:0.4. For a fair comparison across models, the same dataset and split were used for the three deep learning frameworks described in this manuscript (**Supplementary Table 1**).

2.3 Model implementation

To facilitate model training, three distinct state-of-the-art architectures were implemented and evaluated. They are, in no particular order, UNet, Mask R-CNN, and PSPNet. UNet was selected for its prior success in the biomedical domain, Mask R-CNN was selected for its ability to perform both object detection and mask generation, and PSPNet was selected for its capacity to integrate image features from multiple spatial scales. Training, validation, and test data were derived from 12 cores in 7 tissues and a total of 10,359 nuclei in the composition of colon – 1,142; glioblastoma (GBM) – 675; lung – 1735; ovarian – 956; fibroblast – 922; small intestine – 1677; tonsil – 3252. To maintain consistency of evaluation across segmentation algorithms, segmentation accuracy was calculated by counting the fraction of cells in a held out test set that passed a sweeping Intersection over Union (IoU) threshold.

2.3.1 UnMICST-U model training

A three-class UNet model (Ronneberger, Fischer and Brox, 2015) was trained based on annotation of nuclei centers, nuclei contours, and background. The neural network is comprised of 4 layers and 80 input features. Training was performed using a batch size of 32 with the Adam Optimizer and a learning rate of 0.00005 with a decay rate of 0.98 every 5,000 steps until there was no improvement in accuracy or ~100 epochs had been reached. Batch normalization was used to improve training speed. During training, the bottom layer had a dropout rate of 0.35, and L1 regularization was implemented to minimize overfitting (Ng, 2004; Srivastava *et al.*, 2014) and early stopping. Training was performed on workstations equipped with NVidia GTX 1080 or NVidia TitanX GPUs.

2.3.2 UnMICST-M model training

Many segmentation models are based on the Mask R-CNN architecture (He *et al.*, 2018); Mask R-CNN has previously exhibited excellent performance on a variety of segmentation tasks. Mask R-CNN begins by detecting bounding boxes of nuclei and subsequently performs segmentation within each box. This approach eliminates the need for an intermediate watershed, or equivalent, segmentation step. Thus, Mask R-CNN directly calculates a segmentation mask, significantly reducing the overhead in traditional segmentation pipelines. We adopted a ResNet50 (He *et al.*, 2016) backbone model in the UnMICST-M implementation and initialized the weights using pretrained values from the COCO object instance segmentation challenge (Lin *et al.*, 2014) to improve convergence properties. For efficient training, we upsampled the original input images to 800 x 800-pixels and trained a model for 24 epochs using a batch size of 8. The Adam optimizer, with a weight decay of 0.0001 to prevent overfitting, was exploited with a variable learning rate, initially set to 0.01 and decreased by a factor of 0.1 at epochs 16 and 22. Training was performed on a compute node cluster using 4 NVidia TitanX or NVidia Tesla V100 GPUs. For evaluation and comparison, we used the model with the highest performance on the validation set, following standard practice.

2.3.3 UnMICST-P model training

We trained a three class PSPNet model (Zhao *et al.*, 2017) to extract cell nuclei centers, nuclei contours, and background from a wide variety of tissue types. PSPNet is one of the most widely used convolutional neural networks for the semantic segmentation of natural scene images in the computer vision field. The network employs a so-called pyramid pooling module whose purpose is to learn global as well as local features. The additional contextual information used by PSPNet allowed the segmentation algorithm to produce realistic probability maps with greater confidence. We used ResNet101 as a backbone. Training of the network was performed using a batch size of 8 with an image size of 256 x 256-pixels for 15,000 iterations or until the minimum loss model was obtained. A standard cross entropy loss function was used during training. Gradient descent was performed using the Adam optimizer with a learning rate of 0.0001 and a weight decay parameter of

0.005 via L2 regularization. Batch normalization was employed for faster convergence, and a dropout probability of 0.5 was used in the final network layer to mitigate overfitting. The model training was performed on a compute cluster node equipped with NVidia Tesla V100 GPUs.

2.4 Analysis of multi-dimensional data

For the analysis shown in Figure 6, a 64-plex CyCIF image of non-neoplastic small intestine tissue from the EMIT TMA (https://www.synapse.org/#!Synapse:syn22345748/) was stained with a total of 45 antibodies as described in protocols https://www.protocols.io/view/ffpe-tissue-pre-treatment-before-t-cycif-on-leica-bji2kkge and dx.doi.org/10.17504/protocols.io.bjiukkew. Images were segmented using the UnMICST-U model trained on DNA with NES data and real augmentations. Mean fluorescence intensities across 45 markers for 27,847 segmented nuclei were quantified as described in (Schapiro *et al.*, 2021). E-cadherin positive and CD45 positive cells were identified using Gaussian-mixture models on log-transformed data. For multivariate clustering, log-transformed mean intensities of all single cells of 14 selected protein markers (E cadherin, pan-cytokeratin, CD45 CD4, CD3D, CD8', RF3, PML, GLUT1, GAPDH TDP43, OGT, COLL4, an EPCAM) were pre-processed using Uniform Manifold Approximation and Projection (UMAP) (Becht *et al.*, 2019) and clustered using Hierarchical Density-Based Spatial Clustering of Applications with Noise (HDBSCAN) (Campello, Moulavi and Sander, 2013). Clusters expressing a high level of both E-cadherin and CD45 were identified and overlaid on a false-colored image showing the staining of DNA, E-cadherin, and CD45.

3 Results

Semantic segmentation is a coarse-grained approach to segmentation that assigns objects to distinct trained classes, while instance segmentation is fine grained and identifies individual instances of objects. We implemented and evaluated two semantic and one instance segmentation algorithms (UNet, Mask R-CNN, and PSPNet, respectively) to investigate the relative performance of deep learning architectures with different types of training data. We trained each of these models (UnMICST-U, UnMICST-M, and UnMICST-P, respectively) on manually curated and labelled data from seven distinct tissue types (discussed in section 3.1). The accuracy of each model was assessed based on the number of correctly segmented cells relative to the total number using a variable intersection over union (IoU) threshold ranging from the least stringent, 0.55, to most stringent, 0.8. The IoU (the Jaccard Index) is calculated by measuring the overlap between the ground truth annotation and the prediction via a ratio of the intersection to the union of pixels in both masks. The greater the IoU, the higher the accuracy, with an ideal value of 1 (although this is very rarely achieved). Unlike the standard pixel accuracy metric (the fraction of pixels in an image that were correctly classified), IoU is not sensitive to class-imbalance.

3.1 – Data sets and ground truth annotation of nuclear boundaries

It is well established that various cell types in tissues have nuclear morphologies that are different from the spherical and ellipsoidal shape observed in cultured cells (Skinner and Johnson, 2017). Nuclear pleomorphism (variation in nuclear size and shape) is widely used in histopathology to grade cancers and has been automated for H&E images using classical machine vision methods (Dalle et al., 2009). However, the impact on segmentation accuracy of variation in nuclear morphology by cell type and tissue or tumor of origin has not been widely studied for tissue imaging. We, therefore, generated training, validation, and test datasets with manually annotated ground-truth of nuclei contours, centers, and background. These images were acquired from a 120-core tissue microarray representing 30 distinct normal and diseased tissue types. After inspecting all cores, we found that nuclei morphologies could be placed in at least seven groups by eye. These ranged from mixtures of cells that were large vs. small, round cells vs. narrow, and densely and irregularly packed vs. organized in clusters. Additionally, there were multinucleated cells (i.e., glioblastoma) and weakly stained nuclei (i.e., non-neoplastic ovary). Because ground-truth annotation is a laborious process, we focused on distinctive nuclei from a subset of morphologies in seven tissue types (lung adenocarcinoma, non-neoplastic small intestine, normal prostate, colon adenocarcinoma, glioblastoma, non-neoplastic ovary, and tonsil) from 12 cores representing a total of ~10,400 nuclei.

3.2 – Real augmentations increase model robustness to focus artefacts

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

In an initial set of studies, we found that models trained on training data having Gaussian blur augmentation performed well on Gaussian blurred test data but poorly on the same data when it included actual defocused and saturated images. When IoU was quantified for defocused and saturated test data, we found that the use of training data having Gaussian blur augmentations

improved accuracy only slightly relative to baseline models lacking augmentations (**Figure 1B**). In contrast, the use of training data with real augmentation increased the fraction of cells retained by 40-60% at an IoU threshold of 0.6; statistically, significant improvement was observed up to an IoU cutoff of 0.8 with all three learning frameworks (UnMICST-U, UnMICST-M, and UnMICST-P models). To perform a balanced comparison, we created two sets of training data having equal numbers of images. The first set contained the original data plus computed 90- and 180-degree rotations, and the second set contained original data plus defocused data collected from above and below the specimen. We again found that models trained with real augmentations substantially outperform rotationally augmented models when tested on defocused test data (**Figure. 1C**). Repeating this experiment with NES yielded the same outcome. Thus, training deep learning architectures with real augmentation will generate models capable of outperforming models with computed augmentation when data contain commonly encountered artefacts.

3.3 – Addition of NES (lamin B2 and NUP98) improves segmentation accuracy

When we stained our TMA panel (the Exemplar Microscopy Images of Tissues and Tumors (EMIT) TMA) we found that antibodies against lamin A and C (**Figure 2A)** (which are different splice forms of LMNA gene) stained approximately only half as many nuclei as antibodies against lamin B1 (**Figure 2B**) or lamin B2 (**Figure 2C**) (products of the LMNB1 and LMNB2 genes). Staining for the lamin B receptor (**Figure 2E**) exhibited poor image contrast. A pan-tissue survey showed that a mixture of antibodies for nucleoporin NUP98 and lamin B2 conjugated to the same fluorophore (Alexafluor-647) resulted in nuclear envelope staining (NES) for nearly all nuclei across multiple tissues (**Figure 2F- H**). However, using a mixture of NUP98 and lamin B2 antibodies only some cell types, epithelia in colorectal adenocarcinoma for example, exhibited the ring-like structure that is characteristic of nuclear lamina in cultured epithelial cells. The nuclear envelope in immune and other cells has folds and invaginations (Fischer, 2020) and in our data, NES staining could be irregular and diffuse (**Fig. 3A & B**), further emphasizing the difficulty of finding a broadly useful NES stain in tissue. Future studies might therefore explore a wider range of nuclear lamins and nucleoporins in diverse tissue and tumor types to further improve cell type coverage and NES quality.

The value of NES images for model performance was assessed quantitatively and qualitatively. In images of colon adenocarcinoma, non-neoplastic small intestine, and tonsil tissue, we found that the addition of NES images resulted in significant improvements in segmentation accuracy based on IoU with all three learning frameworks; improvements in other tissues, such as lung adenocarcinoma, were more modest and sporadic (**Figure 3A,** Lung). For example, for nuclear segmentation of fibroblasts in prostate cancer tissue, UnMICST-U and UnMICST-M models with NES data were no better than models trained on DNA staining alone. Most striking were cases in which NES data decreased performance (UnMICST-P segmentation on prostrate fibroblasts and UnMICST-U segmentation of glioblastoma). Inspection of the UnMICST-P masks suggested that the segmentation of well-separated fibroblast nuclei was already optimal with DNA images alone (~60%

of nuclei retained at IoU of 0.6), implying the addition of NES images afforded little improvement. With UnMICST-U masks in glioblastoma, the problem appeared to involve atypical NES morphology, which is consistent with a high level of nuclear pleomorphism and the presence of "giant cells," both of which are well-established features of high-grade glioblastoma (Kros, 2011; Louis *et al.*, 2016). We also note that NES data alone was marginally inferior to DNA staining as a sole source of training data and should therefore be used in combination with images of DNA (**Supplementary Figure 1**). In sum we conclude that training on image data that also includes NES, a morphological and molecular feature of nuclei distinct from DNA staining, broadly but not universally results in improvements in segmentation accuracy

3.3b – Combining NES images and real augmentation has a cumulative effect on segmentation accuracy

To determine whether real augmentation and NES combine during model training to achieve superior segmentation precision relative to the use of either type of data alone, we trained and tested models under four different scenarios (using all three learning frameworks; **Figure 4**). We used images from the small intestine, a tissue containing nuclei having a wide variety of morphologies, and then extended the analysis to other tissue types (see below). Models were evaluated on defocused DNA test data to increase the sensitivity of the experiment. In the first scenario, we trained baseline models using in-focus DNA image data and tested models on unseen in-focus DNA images. With tissues such as the small intestine, which are challenging to segment because they contain densely-packed nuclei, scenario A resulted in slightly undersegmented predictions. In Scenario B and for all subsequent scenarios, defocused DNA images were included in the test set, giving rise to contours that were substantially misaligned with ground truth annotations and resulting in higher undersegmentation. False-positive predictions and imprecise localizations of the nuclei membrane were observed in areas devoid of nuclei and with very low contrast (**Figure 4A**). When NES images were included in the training set (Scenario C), nuclear boundaries were more consistent with ground truth annotations, although false-positive predicted nuclei still remained. The best performance was observed when NES images and real augmentation were combined: accurate nuclear boundaries were well aligned with ground truth annotations in both shape and in size. Observable differences in the placement of segmentation masks were reflected in improvements in IoU: for all three deep learning frameworks, including NES data and real augmentations increased the fraction of nuclei retained by 50% at an IoU threshold of 0.6 (**Figure 4B**). The accuracy of UnMICST-P (blue curve) trained on in-focus DNA data alone was higher than the other two baseline models at all IoU thresholds, suggesting that UnMICST-P has a greater capacity to learn. UnMICST-P may have an advantage in experiments in which staining the nuclear envelope proves difficult or impossible.

3.4 – A combination of NES and real augmentations results in improved models across multiple tissue types

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

3.5 – Applying UnMICST to highly multiplex tissue images

To evaluate the overall impact of the innovations described above, UnMICST models with and without real augmentations, NES data, and computed augmentations were used to segment six tissues as a set, including in-focus, saturated and out-focus images (balancing the total amount of training data in each case). A 1.7-fold improvement in accuracy was observed at an IoU of 0.6 for the fully trained model (i.e. with NES data and real augmentations; **Figure 6A**). Inspection of segmentation masks also revealed more accurate contours for nuclei having a wide range of shapes. The overall improvement in accuracy was substantially greater than any difference observed between semantic and instance segmentation frameworks.

We then tested a fully trained UnMICST-U model on a 64-plex CyCIF image of non-neoplastic small itestine tissue from the EMIT TMA (**Figure 6B**). Staining intensities were quantified and on a per-cell basis, and the results visualized using Uniform Manifold Approximation and Projection (UMAP; **Figure 6C**). Segmentation masks were found to be well-located with little evidence of under or over-segmentation (**Figure 6D**). Moreover, whereas 21% of cells with segmented nuclei stained positive (as determined by using Gaussian-mixture model) for the immune cell marker CD45, and 53% stained positive for the epithelial cell marker E cadherin, less than 3% were positive for both. No known cell type is actually positive for both CD45 and E cadherin, and the very low abundance of these double-positive “cells” is evidence of accurate segmentation. When we examined some of the 830 double positive cells (blue dashed circle in **Figure 6C**) we found multiple examples of a $CD3^+$ T cell (yellow arrowheads; light yellow dots in **Figure 6E**) tightly associated with or between the epithelial cells of intestinal villi (green “kiwi” structure visible in **Figure 6E**). This is consistent with the know role of the intestinal epithelium in immune homeostasis (Allaire *et al.*, 2018). In these cases, the ability of humans to distinguish immune and epithelial cells relies on prior knowledge, multi-dimensional intensity features and

subtle differences in shape and texture – none of which were elements of model training. Thus, future improvements in tissue segmentation are likely to require the development of CNNs able to classify rare but biologically significant spatial arrangements, rather than simple extension of the general purpose segmentation algorithms described here.

3.6 – Some tissue still pose a challenge for nuclei segmentation

Of all the tissue types annotated and tested in this paper, non-neoplastic ovary was the most difficult to segment (**Supplementary Figure 2A**). We also observed that the addition of ovarian training data to any of the three models also trained on data from other tissues caused a decrease in model accuracy when tested against other images from those tissues (**Supplementary Figure 2B**). We speculate that this reflects the highly irregular morphology, poor image contrast, and dense packing of many non-neoplastic ovarian cell nuclei as compared to colon adenocarcinoma (**Supplementary Figure 2 C,D**). We found that increasing the resolution by oversampling the image (at 0.325 microns/pixel) only marginally improved our ability to discern nuclei. We have previously imaged ovarian cancers at even higher resolution (60x/1.42NA sampled at 108 nm pixel size; (Färkkilä et al., 2020) using optical sectioning and deconvolution and inspection of these images showed that discernibility between nuclei remains poor and that multiple cells are often found stacked on top of each other in a single 5 µm section. Additional research, possibly involving different NES antibodies will be required to improve performance with ovarian and other difficult to segment tissues. Until then, caution is warranted when combining training data from tissues with highly irregular nuclear morphologies.

4 Discussion

In this paper, we show that the use of real image augmentation and data on the morphology of the nuclear envelope (NES) improves the accuracy of nuclear image segmentation in a range of tissue and tumor types. We show that this is true with three different deep learning frameworks based on instance segmentation (UnMICST-M) or semantic segmentation (UnMICST-U and UnMICST-P). The generality and utility of our findings are further enhanced by the use of whole-slide fluorescence microscopy technologies being developed in multiple laboratories for large-scale atlasing projects such as Human Tumor Atlas Network (HTAN) and Human BioMolecular Atlas Program (HubMAP) (HuBMAP Consortium, 2019; Rozenblatt-Rosen *et al.*, 2020b; Lin *et al.*, 2021). Augmenting training data with intentionally defocused and saturated images significantly improves segmentation accuracy across multiple tissue types, whereas conventional computational augmentation (inclusion of a Gaussian blur during training) does not. The use of NES images in training also results in a statistically significant improvement for most tissue types. Based on segmentation performance, NES staining is not a substitute for DNA staining – it should instead be used in conjunction with DNA. Combined use of real augmentation and NES data during model training results in a substantial and cumulative improvement in performance, making segmentation robust to differences in modeling framework and data quality in all but one of the tissues tested.

Moreover, we show that these improvements pertain not only to test data generated for this study, but also the real-world task of segmenting high dimensional images of resected human tissues.

We find that the three deep learning models tested are broadly similar in performance when judged by IoU and direct inspection of segmentation masks, with UnMICST-M somewhat better than UnMICST-U overall. Thus, instance segmentation may have advantages over semantic segmentation. However, the slightly better performance of UnMICST-P with DNA-only training data suggests that deeper semantic segmentation may be helpful when training data are limited. These differences were all smaller in magnitude and more scenario-specific than improvements observed when training involved real augmentation and NES data. From a machine learning perspective, the addition of images to the training data is expected to improve model performance. Experimental feasibility is less clear. A key tradeoff is that the greater the number of fluorescence channels used for segmentation, the fewer channels available for the collection of data on other functional markers. Fortunately, the development of highly multiplexed imaging has made this less relevant because collection of 20-40 or more image channels (each corresponding to a different fluorescent antibody) has become routine using techniques such as CyCIF. This makes it straightforward to reserve two channels for segmentation. Moreover, adding labelled NES antibodies to highly multiplexed tissue imaging studies has a negligible impact on cost. The cost-benefit ratio of extra segmentation data is different in high content screening of cells in multi-well plates, for which inexpensive reagents are generally essential.

As a means other than DNA staining of identifying nuclei, components of the nuclear lamina and nuclear envelope are obvious choices. These proteins comprise a dense network of intermediate filaments that forms a prominent ring just inside the nuclear envelope. However, identifying antibodies that stain multiple cell types and tissues is not trivial because lamin isoforms are differentially expressed. The best combination for NES that we have identified is a mixture of antibodies against lamin B2 and the nuclear pore protein NUP98, but further study of other markers is warranted, particularly in ovarian tissue. Lamins have a wide range of possible functions in cancer and have been documented to change in expression levels and morphology with disease state (Sakthivel and Sehgal, 2016). Thus, imaging lamin is likely to provide valuable biological information in addition to improving segmentation (Bell and Lammerding, 2016). Abnormal nuclear morphology is an important feature of cancer cells with established diagnostic utility (Uhler and Shivashankar, 2018).

Our discovery that real augmentation outperforms computed augmentation may have general significance outside of the field of microscopy. With any high-performance camera system, real out-of-focus blur is more complicated than Gaussian blur. In microscopy, this arises from light scattering and diffraction, and in tissue imaging, the use of non-immersion objective lenses results in a large refractive index mismatch between the air surrounding the lens and the glass slide. Thus, the appearance of out-of-focus blur differs when the microscope focal plane is above and below the specimen. In other settings, optical aberrations might have different causes but the overall concept would be the same. The collection of images for real augmentation does not impose a major burden

on data collection for model training since only one set of ground truth annotations needs to be curated. Areas of future investigation of real augmentation in microscopy include different settings for correction collars, inhomogeneous light sources, and stage jitter. We note, however, that blur is the most common problem encountered with well-aligned instruments and can never be fully eliminated: some cells will always be out-of-focus when imaged by widefield epifluorescence microscopy. It will also be useful to determine whether a kernel can be derived from real augmentation to make them computable, but our preliminary studies suggest that this will not be trivial.

As part of the current work, we are releasing all training and test images, their segmentation masks and annotations, and real augmentations for multiple types of tissue (tonsil, ovary, small intestine and cancers of the colon, brain, lung, prostate) via the EMIT resource; models are released as components of the UnMICST model resource (www.synapse.org/#!Synapse:syn24192218/files/). Both resources are maintained by the HTAN consortium and lay the groundwork for future improvements in models and training/test data. Crowdsourcing may help to overcome a limitation in the current study, which does not extend to all 34 tissue types in the EMIT dataset. This derives primarily from the high burden imposed by ground-truth labeling.

4.1 Conclusion

This paper makes three primary contributions to the growing literature on the segmentation of multiplexed tissue images, which is an essential step in single-cell data analysis but remains a substantial technical challenge. First, we show that it is often possible to increase segmentation accuracy by including additional channels on nuclear morphology. Second, we show that real augmentations involving the use of defocused and saturated images during model training improve segmentation accuracy to a significant extent whereas Gaussian blurring does not. Third, we create labeled training data for multiple tissue types and three different deep learning models. These models are a good starting point for performing image segmentation on tissue data collected in ongoing projects and can be further enhanced by the addition of new training data.

With the rapid growth of multiplexed tissue imaging and digital pathology, the generation of high-performance segmentation models for real pathology data has immediate utility. Our data suggest that the current generation of models work well with lung adenocarcinoma, in which nuclei are generally circular and well separated, and colorectal cancers, in which residual glandular morphology results in elliptical and generally well-aligned nuclei. However, segmentation remains challenging in ovarian tissue and glioblastoma, and additional experimental and computational work is required to improve model accuracy. This might include the validation of additional morphological markers and the generation of improved training data through the annotation of higher-resolution 3D images. Our analysis of errors remaining when multiplexed images of tissues such as small intestine tissue are segmented using optimized UnMICST models suggests that many errors have a subtle biological basis. Development of additional “physiology aware” machine-learning models may be necessary to reduce these apparent errors.

ACKNOWLEDGEMENTS

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

OUTSIDE INTERESTS

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.

Figure 1: Comparing the use of real augmentations (defocused and overexposed images) and Gaussian blur. A) Schematic diagram showing the approach comparing test images on models trained with Gaussian-blurred or defocused image data. Higher contrast probability maps signify more confidence – areas of interest are highlighted with red arrows. Corresponding probability maps indicate a model trained with defocused images performs better on defocused test images than a Gaussian-blurred model. Scale bar denotes 20 micrometers. **B**) Plots show that incorporating real augmentations (red curve) into the training set is statistically significantly superior to training sets with Gaussian blur (yellow curve) and without real augmentations (blue curve) for UnMICST-U, UnMICST-M, and UnMICST-P. Simulating defocused images with Gaussian blur is only marginally better than not augmenting the training data at all. **C**) Comparing UnMICST-U model accuracy when the training dataset size was held constant by replacing defocused augmentations (red curve) with 90 and 180 degree rotations (blue curve).

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

Figure 3: NES with DNA improves nuclear segmentation. NES – nuclear envelope staining. Assessing the addition of NES as a 2nd marker to DNA on segmentation accuracy on a per tissue and per model basis. **A**) Variable IoU plots comparing the DNA-only model (blue curve) and the DNA + NES model (red curve) across frameworks. Adding NES increased accuracy for densely packed nuclei such as colon, small intestine, tonsil, and to some extent, lung tissue. Error bars are standard errors of mean. **B**) Representative grayscale images of tissues stained with DNA and NES

comparing their variable morphologies, followed by UnMICST-U mask predictions (green) overlaid onto ground truth annotations (purple). In tissue with sparse nuclei, such as fibroblasts from prostate tissue, NES did not add an additional benefit to DNA alone. In tissues where NES does not exhibit the characteristic nuclear ring, as in glioblastoma, the accuracy was similarly not improved. Scale bar denotes 20 micrometers.

Figure 4: Combination of NES and real image augmentations on segmentation performance. NES - nuclear envelope staining. **A**) Models trained with in-focus DNA data alone produced probability maps that were undersegmented, especially in densely-packed tissue such as small intestine (Scenario A). When tested on defocused data, nuclei borders were largely incorrect (Scenario B). Adding NES restored nuclei border shapes (Scenario C). Combining NES and real augmentations reduced false positive detections and produced nuclei masks better resembling the ground truth labels (Scenario D). Scalebar denotes 20 micrometers. Table legend shows conditions used for each scenario A-D. **B**) Graphs compare the accuracy represented as the number of cells retained across varying IoU thresholds with all models from UnMICST-U (top), UnMICST-M (center), and UnMICST-P (bottom). In all models, more nuclei were retained when NES and real augmentations were used together during training (yellow curves) compared to using NES without real augmentations (red curves) or DNA alone (blue curves).

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

Figure 6: Applying UnMICST models to highly multiplexed image data. A) Accuracy improvement of UnMICST models trained with and without NES (nuclear envelope staining) as compared to DNA alone, and real augmentations as compared to computed blur (GN; Gaussian noise). To balance training dataset size, GN was substituted for NES data and computed 90/180 degree rotations were substituted for real augmentations. **B)** A 64-plex CyCIF image of a non-neoplastic small intestine TMA core from the EMIT dataset. Dashed box indicates region of interest for panels **D** and **E**. **C)** UMAP projection using single cell staining intensities for 14 marker proteins (see methods). The color of the data points represents the intensity of E-cadherin (top left) or CD45 (bottom left) across all segmented nuclei. Density-based clustering using HDBSCAN identified distinct clusters (each denoted by a different color) that were positive for either E-cadherin or CD45 as well as a small number of double-positive cells (blue dashed circle). **D)** Enlarged region of yellow dashed box from **B** showing segmentation mask outlines (magenta) overlayed onto DNA channel

(green). **E)** Composite image of DNA, E-cadherin, and CD45 of the same region. Nuclei centroids from segmentation denoted by brown dots. Cells positive for both E-cadherin and CD45 (from blue dashed circle in panel **C** are marked with yellow arrows and yellow dots. Inset: enlarged view of boxed region showing overlapping immune and epithelial cells.

SUPPLEMENTARY FIGURES

Supplementary Figure 1: NES data does not substitute for DNA in model training. Test results when training was performed on DNA (blue curve) and lamin (red curve) individually as single channels for **A**) UnMICST-U, **B**) UnMICST-M, and **C**) UnMICST-P.

Supplementary Figure 2: Assessing the use of normal ovarian data in model training. **A**) Normal ovarian tissue performed the worst with respect to segmentation accuracy among all annotated tissues. Using NES and real augmentations conferred a modest benefit. **B**) The segmentation accuracy across all tissues was lower when ovarian data was included in the training set (blue curve) as opposed to excluded (red curve). **C**) Images showing that annotation of normal ovarian nuclei from DNA is a challenging task, even with NES staining at four times the nominal resolution (0.325 microns per pixel and with a 20x/0.75 objective lens). **D**) In contrast, colon was straightforward to annotate, particularly with NES staining, which forms a distinct halo around the nuclear periphery.

Supplementary Table 1: Data set size and composition across disjoint training and testing data splits. The model used for training is indicated in the left-hand column and was identical across all three segmentation architectures.

Model	**Training Set Size**	**Test Set Size**
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 + $90^o/180^o$ rotations	5,100	645
DNA + NES + Real Augmentations	5,100	645
DNA + NES + $90^o/180^o$ rotations	5,100	645

REFERENCES

Achim, K. *et al.* (2015) ‘High-throughput spatial mapping of single-cell RNA-seq data to tissue of origin’, *Nature Biotechnology*, 33(5), pp. 503–509. doi: 10.1038/nbt.3209.

Ahmed Raza, S. E. *et al.* (2017) ‘MIMO-Net: A multi-input multi-output convolutional neural network for cell segmentation in fluorescence microscopy images’, in *2017 IEEE 14th International Symposium on Biomedical Imaging (ISBI 2017)*. *2017 IEEE 14th International Symposium on Biomedical Imaging (ISBI)*, Melbourne, VIC: IEEE, pp. 337–340. doi: 10.1109/ISBI.2017.7950532.

Albertson, D. G. (2006) ‘Gene amplification in cancer’, *Trends in Genetics*, 22(8), pp. 447–455. doi: 10.1016/j.tig.2006.06.007.

Al-Kofahi, Y. *et al.* (2010) ‘Improved Automatic Detection and Segmentation of Cell Nuclei in Histopathology Images’, *IEEE Transactions on Biomedical Engineering*, 57(4), pp. 841–852. doi: 10.1109/TBME.2009.2035102.

Al-Kofahi, Y. *et al.* (2018) ‘A deep learning-based algorithm for 2-D cell segmentation in microscopy images’, *BMC Bioinformatics*, 19(1), p. 365. doi: 10.1186/s12859-018-2375-z.

Allaire, J. M. *et al.* (2018) ‘The Intestinal Epithelium: Central Coordinator of Mucosal Immunity’, *Trends in Immunology*, 39(9), pp. 677–696. doi: 10.1016/j.it.2018.04.002.

Amin, M. B. *et al.* (2017) ‘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: a cancer journal for clinicians*, 67(2), pp. 93–99. doi: 10.3322/caac.21388.

Angelo, M. *et al.* (2014) ‘Multiplexed ion beam imaging (MIBI) of human breast tumors’, *Nature medicine*, 20(4), pp. 436–442. doi: 10.1038/nm.3488.

Arganda-Carreras, I. *et al.* (2017) ‘Trainable Weka Segmentation: a machine learning tool for microscopy pixel classification’, *Bioinformatics*, 33(15), pp. 2424–2426. doi: 10.1093/bioinformatics/btx180.

Becht, E. *et al.* (2019) ‘Dimensionality reduction for visualizing single-cell data using UMAP’, *Nature Biotechnology*, 37(1), pp. 38–44. doi: 10.1038/nbt.4314.

Bell, E. S. and Lammerding, J. (2016) ‘Causes and consequences of nuclear envelope alterations in tumour progression’, *European Journal of Cell Biology*, 95(11), pp. 449–464. doi: 10.1016/j.ejcb.2016.06.007.

Berg, S. *et al.* (2019) ‘ilastik: interactive machine learning for (bio)image analysis’, *Nature Methods*, 16(12), pp. 1226–1232. doi: 10.1038/s41592-019-0582-9.

Berryman, S. *et al.* (2019) ‘Image-based Cell Phenotyping Using Deep Learning’, *bioRxiv*, p. 817544. doi: 10.1101/817544.

Byun, J. *et al.* (2006) ‘Automated tool for the detection of cell nuclei in digital microscopic images: application to retinal images’, *Molecular Vision*, 12, pp. 949–960.

Caicedo, J. C. *et al.* (2019) ‘Nucleus segmentation across imaging experiments: the 2018 Data Science Bowl’, *Nature Methods*, 16(12), pp. 1247–1253. doi: 10.1038/s41592-019-0612-7.

Campello, R. J. G. B., Moulavi, D. and Sander, J. (2013) ‘Density-Based Clustering Based on Hierarchical Density Estimates’, in Pei, J. et al. (eds) *Advances in Knowledge Discovery and Data*

Mining. Berlin, Heidelberg: Springer (Lecture Notes in Computer Science), pp. 160–172. doi: 10.1007/978-3-642-37456-2_14.

Chalfoun, J. *et al.* (2014) 'FogBank: a single cell segmentation across multiple cell lines and image modalities', *BMC Bioinformatics*, 15(1), p. 431. doi: 10.1186/s12859-014-0431-x.

Challen, G. A. and Goodell, M. A. (2008) 'Promiscuous Expression of H2B-GFP Transgene in Hematopoietic Stem Cells', *PLOS ONE*, 3(6), p. e2357. doi: 10.1371/journal.pone.0002357.

Chen, H. *et al.* (2016) 'DCAN: Deep Contour-Aware Networks for Accurate Gland Segmentation', *arXiv:1604.02677 [cs]*. Available at: http://arxiv.org/abs/1604.02677 (Accessed: 5 October 2020).

Deng, J. *et al.* (2009) 'ImageNet: A large-scale hierarchical image database', in *2009 IEEE Conference on Computer Vision and Pattern Recognition*. *2009 IEEE Conference on Computer Vision and Pattern Recognition*, pp. 248–255. doi: 10.1109/CVPR.2009.5206848.

Fischer, E. G. (2020) 'Nuclear Morphology and the Biology of Cancer Cells', *Acta Cytologica*, 64(6), pp. 511–519. doi: 10.1159/000508780.

Gerdes, M. J. *et al.* (2013) 'Highly multiplexed single-cell analysis of formalin-fixed, paraffin-embedded cancer tissue', *Proceedings of the National Academy of Sciences*, 110(29), pp. 11982–11987. doi: 10.1073/pnas.1300136110.

Giesen, C. *et al.* (2014) 'Highly multiplexed imaging of tumor tissues with subcellular resolution by mass cytometry', *Nature Methods*, 11(4), pp. 417–422. doi: 10.1038/nmeth.2869.

Goltsev, Y. *et al.* (2018) 'Deep Profiling of Mouse Splenic Architecture with CODEX Multiplexed Imaging', *Cell*, 174(4), pp. 968-981.e15. doi: 10.1016/j.cell.2018.07.010.

Gurari, D. *et al.* (2015) 'How to Collect Segmentations for Biomedical Images? A Benchmark Evaluating the Performance of Experts, Crowdsourced Non-experts, and Algorithms', *2015 IEEE Winter Conference on Applications of Computer Vision*. doi: 10.1109/WACV.2015.160.

He, K. *et al.* (2016) 'Deep Residual Learning for Image Recognition', in *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 770–778. doi: 10.1109/CVPR.2016.90.

He, K. *et al.* (2018) 'Mask R-CNN', *arXiv:1703.06870 [cs]*. Available at: http://arxiv.org/abs/1703.06870 (Accessed: 15 March 2021).

Höfener, H. *et al.* (2018) 'Deep learning nuclei detection: A simple approach can deliver state-of-the-art results', *Computerized Medical Imaging and Graphics*, 70, pp. 43–52. doi: 10.1016/j.compmedimag.2018.08.010.

Horwath, J. P. *et al.* (2020) 'Understanding important features of deep learning models for segmentation of high-resolution transmission electron microscopy images', *npj Computational Materials*, 6(1), pp. 1–9. doi: 10.1038/s41524-020-00363-x.

HuBMAP Consortium (2019) 'The human body at cellular resolution: the NIH Human Biomolecular Atlas Program', *Nature*, 574(7777), pp. 187–192. doi: 10.1038/s41586-019-1629-x.

Immunologists, A. A. of (1942) 'The Demonstration of Pneumococcal Antigen in Tissues by the Use of Fluorescent Antibody', *The Journal of Immunology*, 45(3), pp. 159–170.

Khadangi, A., Boudier, T. and Rajagopal, V. (2020) 'EM-net: Deep learning for electron microscopy image segmentation', *bioRxiv*, p. 2020.02.03.933127. doi: 10.1101/2020.02.03.933127.

Korfhage, N. *et al.* (2020) ‘Detection and segmentation of morphologically complex eukaryotic cells in fluorescence microscopy images via feature pyramid fusion’, *PLOS Computational Biology*, 16(9), p. e1008179. doi: 10.1371/journal.pcbi.1008179.

Krizhevsky, A., Sutskever, I. and Hinton, G. E. (2012) ‘ImageNet Classification with Deep Convolutional Neural Networks’, in Pereira, F. et al. (eds) *Advances in Neural Information Processing Systems 25*. Curran Associates, Inc., pp. 1097–1105. Available at: http://papers.nips.cc/paper/4824-imagenet-classification-with-deep-convolutional-neural-networks.pdf (Accessed: 5 October 2020).

Kromp, F. *et al.* (2019) ‘Deep Learning architectures for generalized immunofluorescence based nuclear image segmentation’. Available at: https://arxiv.org/abs/1907.12975v1 (Accessed: 10 October 2020).

Kromp, F. *et al.* (2020) ‘An annotated fluorescence image dataset for training nuclear segmentation methods’, *Scientific Data*, 7(1), p. 262. doi: 10.1038/s41597-020-00608-w.

Kros, J. M. (2011) ‘Grading of Gliomas: The Road From Eminence to Evidence’, *Journal of Neuropathology & Experimental Neurology*, 70(2), pp. 101–109. doi: 10.1097/NEN.0b013e31820681aa.

LeCun, Y., Bengio, Y. and Hinton, G. (2015) ‘Deep learning’, *Nature*, 521(7553), pp. 436–444. doi: 10.1038/nature14539.

Lee, H. and Jeong, W.-K. (2020) ‘Scribble2Label: Scribble-Supervised Cell Segmentation via Self-Generating Pseudo-Labels with Consistency’, *arXiv:2006.12890 [cs]*. Available at: http://arxiv.org/abs/2006.12890 (Accessed: 5 October 2020).

Lin, J.-R. *et al.* (2018) ‘Highly multiplexed immunofluorescence imaging of human tissues and tumors using t-CyCIF and conventional optical microscopes’, *eLife*, 7. doi: 10.7554/eLife.31657.

Lin, J.-R. *et al.* (2021) ‘Multiplexed 3D atlas of state transitions and immune interactions in colorectal cancer’, *bioRxiv*, p. 2021.03.31.437984. doi: 10.1101/2021.03.31.437984.

Lin, T.-Y. *et al.* (2014) ‘Microsoft COCO: Common Objects in Context’, in Fleet, D. et al. (eds) *Computer Vision – ECCV 2014*. Cham: Springer International Publishing (Lecture Notes in Computer Science), pp. 740–755. doi: 10.1007/978-3-319-10602-1_48.

Liu, C. *et al.* (2016) ‘Detecting and segmenting cell nuclei in two-dimensional microscopy images’, *Journal of Pathology Informatics*, 7(1), p. 42. doi: 10.4103/2153-3539.192810.

Liu, D. *et al.* (2020) ‘Unsupervised Instance Segmentation in Microscopy Images via Panoptic Domain Adaptation and Task Re-Weighting’, in *2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*. *2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, Seattle, WA, USA: IEEE, pp. 4242–4251. doi: 10.1109/CVPR42600.2020.00430.

Ljosa, V., Sokolnicki, K. L. and Carpenter, A. E. (2012) ‘Annotated high-throughput microscopy image sets for validation’, *Nature methods*, 9(7), p. 637. doi: 10.1038/nmeth.2083.

Louis, D. *et al.* (2016) *WHO Classification of Tumours of the Central Nervous System*. Available at: https://publications.iarc.fr/Book-And-Report-Series/Who-Classification-Of-Tumours/WHO-Classification-Of-Tumours-Of-The-Central-Nervous-System-2016 (Accessed: 15 March 2021).

Lugagne, J.-B., Lin, H. and Dunlop, M. J. (2020) ‘DeLTA: Automated cell segmentation, tracking, and lineage reconstruction using deep learning’, *PLOS Computational Biology*, 16(4), p. e1007673. doi: 10.1371/journal.pcbi.1007673.

Lux, F. and Matula, P. (2020) ‘Cell Segmentation by Combining Marker-Controlled Watershed and Deep Learning’, *arXiv:2004.01607 [cs, eess]*. Available at: http://arxiv.org/abs/2004.01607 (Accessed: 5 October 2020).

Masubuchi, S. *et al.* (2020) ‘Deep-learning-based image segmentation integrated with optical microscopy for automatically searching for two-dimensional materials’, *npj 2D Materials and Applications*, 4(1), pp. 1–9. doi: 10.1038/s41699-020-0137-z.

McQuin, C. *et al.* (2018) ‘CellProfiler 3.0: Next-generation image processing for biology’, *PLOS Biology*, 16(7), p. e2005970. doi: 10.1371/journal.pbio.2005970.

Ng, A. Y. (2004) ‘Feature selection, L1 vs. L2 regularization, and rotational invariance’, in *Proceedings of the twenty-first international conference on Machine learning*. New York, NY, USA: Association for Computing Machinery (ICML ’04), p. 78. doi: 10.1145/1015330.1015435.

Otsu, N. (1979) ‘A Threshold Selection Method from Gray-Level Histograms’. doi: 10.1109/TSMC.1979.4310076.

Qi, X. *et al.* (2012) ‘Robust Segmentation of Overlapping Cells in Histopathology Specimens Using Parallel Seed Detection and Repulsive Level Set’, *IEEE Transactions on Biomedical Engineering*, 59(3), pp. 754–765. doi: 10.1109/TBME.2011.2179298.

Raza, S. E. A. *et al.* (2019) ‘Micro-Net: A unified model for segmentation of various objects in microscopy images’, *Medical Image Analysis*, 52, pp. 160–173. doi: 10.1016/j.media.2018.12.003.

Ronneberger, O., Fischer, P. and Brox, T. (2015) ‘U-Net: Convolutional Networks for Biomedical Image Segmentation’, *arXiv:1505.04597 [cs]*. Available at: http://arxiv.org/abs/1505.04597 (Accessed: 5 October 2020).

Roth, H. R. *et al.* (2015) ‘DeepOrgan: Multi-level Deep Convolutional Networks for Automated Pancreas Segmentation’, in Navab, N. et al. (eds) *Medical Image Computing and Computer-Assisted Intervention -- MICCAI 2015*. Cham: Springer International Publishing (Lecture Notes in Computer Science), pp. 556–564. doi: 10.1007/978-3-319-24553-9_68.

Rozenblatt-Rosen, O. *et al.* (2020a) ‘The Human Tumor Atlas Network: Charting Tumor Transitions across Space and Time at Single-Cell Resolution’, *Cell*, 181(2), pp. 236–249. doi: 10.1016/j.cell.2020.03.053.

Rozenblatt-Rosen, O. *et al.* (2020b) ‘The Human Tumor Atlas Network: Charting Tumor Transitions across Space and Time at Single-Cell Resolution’, *Cell*, 181(2), pp. 236–249. doi: 10.1016/j.cell.2020.03.053.

Saka, S. K. *et al.* (2019) ‘Immuno-SABER enables highly multiplexed and amplified protein imaging in tissues’, *Nature Biotechnology*, 37(9), pp. 1080–1090. doi: 10.1038/s41587-019-0207-y.

Sakthivel, K. M. and Sehgal, P. (2016) ‘A Novel Role of Lamins from Genetic Disease to Cancer Biomarkers’, *Oncology Reviews*, 10(2), p. 309. doi: 10.4081/oncol.2016.309.

Schapiro, D. *et al.* (2021) ‘MCMICRO: A scalable, modular image-processing pipeline for multiplexed tissue imaging’, *bioRxiv*, p. 2021.03.15.435473. doi: 10.1101/2021.03.15.435473.

Schmidt, U. *et al.* (2018) ‘Cell Detection with Star-convex Polygons’, *arXiv:1806.03535 [cs]*, 11071, pp. 265–273. doi: 10.1007/978-3-030-00934-2_30.

Schüffler, P. J. *et al.* (2015) ‘Automatic single cell segmentation on highly multiplexed tissue images’, *Cytometry Part A*, 87(10), pp. 936–942. doi: https://doi.org/10.1002/cyto.a.22702.

Schwendy, M., Unger, R. E. and Parekh, S. H. (2020) ‘EVICAN—a balanced dataset for algorithm development in cell and nucleus segmentation’, *Bioinformatics*, 36(12), pp. 3863–3870. doi: 10.1093/bioinformatics/btaa225.

Shahzad M *et al.* (2020) ‘Robust Method for Semantic Segmentation of Whole-Slide Blood Cell Microscopic Images.’, *Computational and Mathematical Methods in Medicine*, 2020, pp. 4015323–4015323. doi: 10.1155/2020/4015323.

Shlien, A. and Malkin, D. (2009) ‘Copy number variations and cancer’, *Genome Medicine*, 1(6), p. 62. doi: 10.1186/gm62.

Shorten, C. and Khoshgoftaar, T. M. (2019) ‘A survey on Image Data Augmentation for Deep Learning’, *Journal of Big Data*, 6(1), p. 60. doi: 10.1186/s40537-019-0197-0.

Slyper, M. *et al.* (2020) ‘A single-cell and single-nucleus RNA-Seq toolbox for fresh and frozen human tumors’, *Nature Medicine*, 26(5), pp. 792–802. doi: 10.1038/s41591-020-0844-1.

Srivastava, N. *et al.* (2014) ‘Dropout: a simple way to prevent neural networks from overfitting’, *The Journal of Machine Learning Research*, 15(1), pp. 1929–1958.

Stringer, C. *et al.* (2020) ‘Cellpose: a generalist algorithm for cellular segmentation’, *Nature Methods*, pp. 1–7. doi: 10.1038/s41592-020-01018-x.

Stylianidou, S. *et al.* (2016) ‘SuperSegger: robust image segmentation, analysis and lineage tracking of bacterial cells’, *Molecular Microbiology*, 102(4), pp. 690–700. doi: 10.1111/mmi.13486.

Torr, A. *et al.* (2020) ‘DeepSplit: Segmentation of Microscopy Images Using Multi-task Convolutional Networks’, in Papież, B. W. et al. (eds) *Medical Image Understanding and Analysis*. Cham: Springer International Publishing (Communications in Computer and Information Science), pp. 155–167. doi: 10.1007/978-3-030-52791-4_13.

Tumbar, T. *et al.* (2004) ‘Defining the epithelial stem cell niche in skin’, *Science (New York, N.Y.)*, 303(5656), pp. 359–363. doi: 10.1126/science.1092436.

Uhler, C. and Shivashankar, G. V. (2018) ‘Nuclear Mechanopathology and Cancer Diagnosis’, *Trends in Cancer*, 4(4), pp. 320–331. doi: 10.1016/j.trecan.2018.02.009.

Vu, Q. D. *et al.* (2019) ‘Methods for Segmentation and Classification of Digital Microscopy Tissue Images’, *Frontiers in Bioengineering and Biotechnology*, 7. doi: 10.3389/fbioe.2019.00053.

Vuola, A. O., Akram, S. U. and Kannala, J. (2019) ‘Mask-RCNN and U-net Ensembled for Nuclei Segmentation’, *arXiv:1901.10170 [cs]*. Available at: http://arxiv.org/abs/1901.10170 (Accessed: 5 October 2020).

Wang, C. *et al.* (2019) ‘Deep learning pipeline for cell edge segmentation of time-lapse live cell images’, *bioRxiv*, p. 191858. doi: 10.1101/191858.

Wen, C. *et al.* (2018) ‘Deep-learning-based flexible pipeline for segmenting and tracking cells in 3D image time series for whole brain imaging’, *bioRxiv*, p. 385567. doi: 10.1101/385567.

Xing, F. *et al.* (2019) ‘Towards pixel-to-pixel deep nucleus detection in microscopy images’, *BMC Bioinformatics*, 20(1), p. 472. doi: 10.1186/s12859-019-3037-5.

Xu, H. *et al.* (2017) ‘Automatic Nuclei Detection Based on Generalized Laplacian of Gaussian Filters’, *IEEE Journal of Biomedical and Health Informatics*, 21(3), pp. 826–837. doi: 10.1109/JBHI.2016.2544245.

Xu, H., Lu, C. and Mandal, M. (2014) ‘An efficient technique for nuclei segmentation based on ellipse descriptor analysis and improved seed detection algorithm’, *IEEE journal of biomedical and health informatics*, 18(5), pp. 1729–1741. doi: 10.1109/JBHI.2013.2297030.

Yang, L. *et al.* (2020) ‘NuSeT: A deep learning tool for reliably separating and analyzing crowded cells’, *PLOS Computational Biology*, 16(9), p. e1008193. doi: 10.1371/journal.pcbi.1008193.

Zhao, H. *et al.* (2017) ‘Pyramid Scene Parsing Network’, *arXiv:1612.01105 [cs]*. Available at: http://arxiv.org/abs/1612.01105 (Accessed: 15 March 2021).

Figure 1

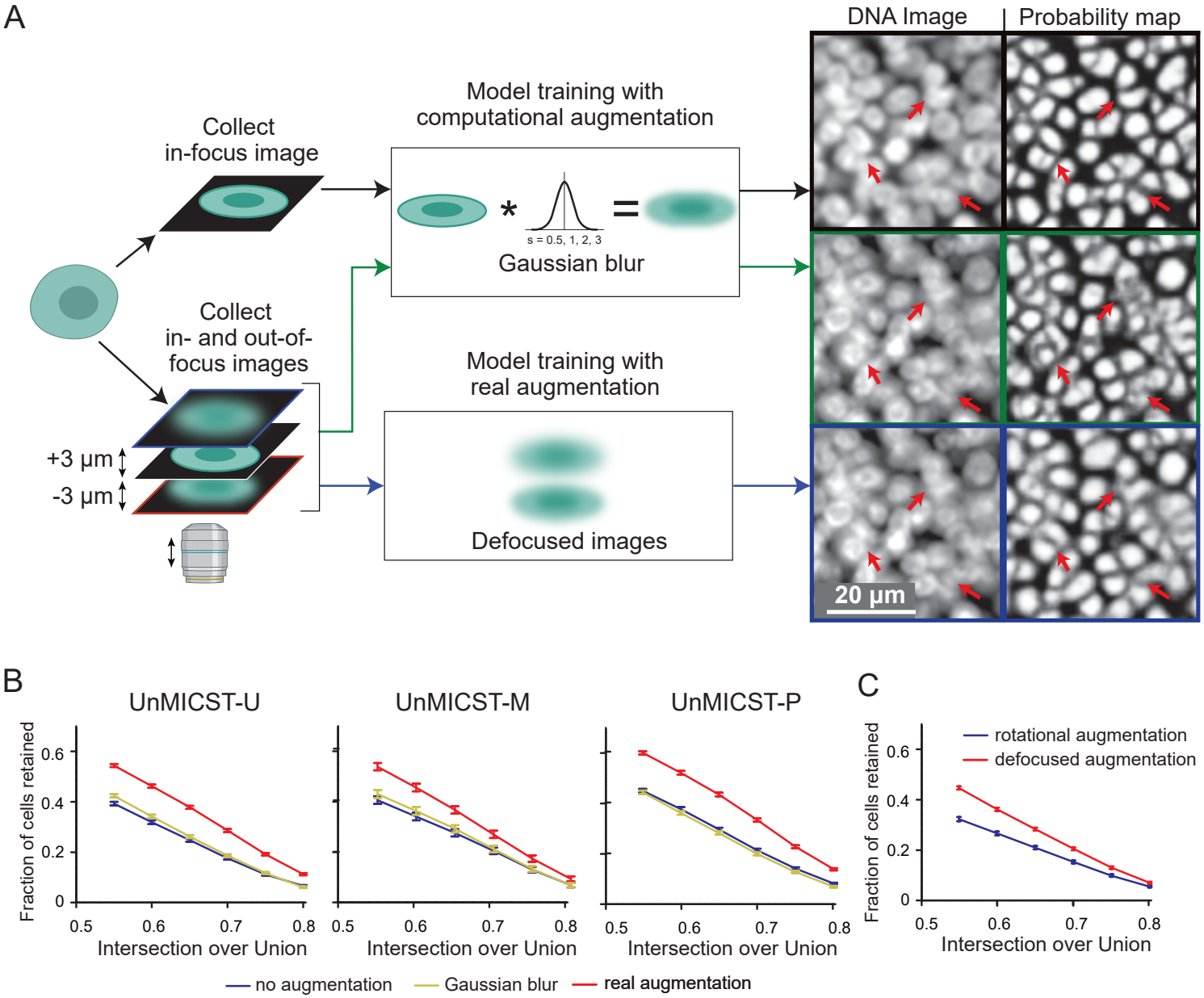

A
DNA Image
Probability map
Model training with computational augmentation
Collect in-focus image
s = 0.5, 1, 2, 3
Gaussian blur
Collect in- and out-of-focus images
Model training with real augmentation
+3 µm
-3 µm
Defocused images
20 µm
B
UnMICST-U
UnMICST-M
UnMICST-P
Fraction of cells retained
Intersection over Union
no augmentation
Gaussian blur
real augmentation
C
rotational augmentation
defocused augmentation

Figure 2

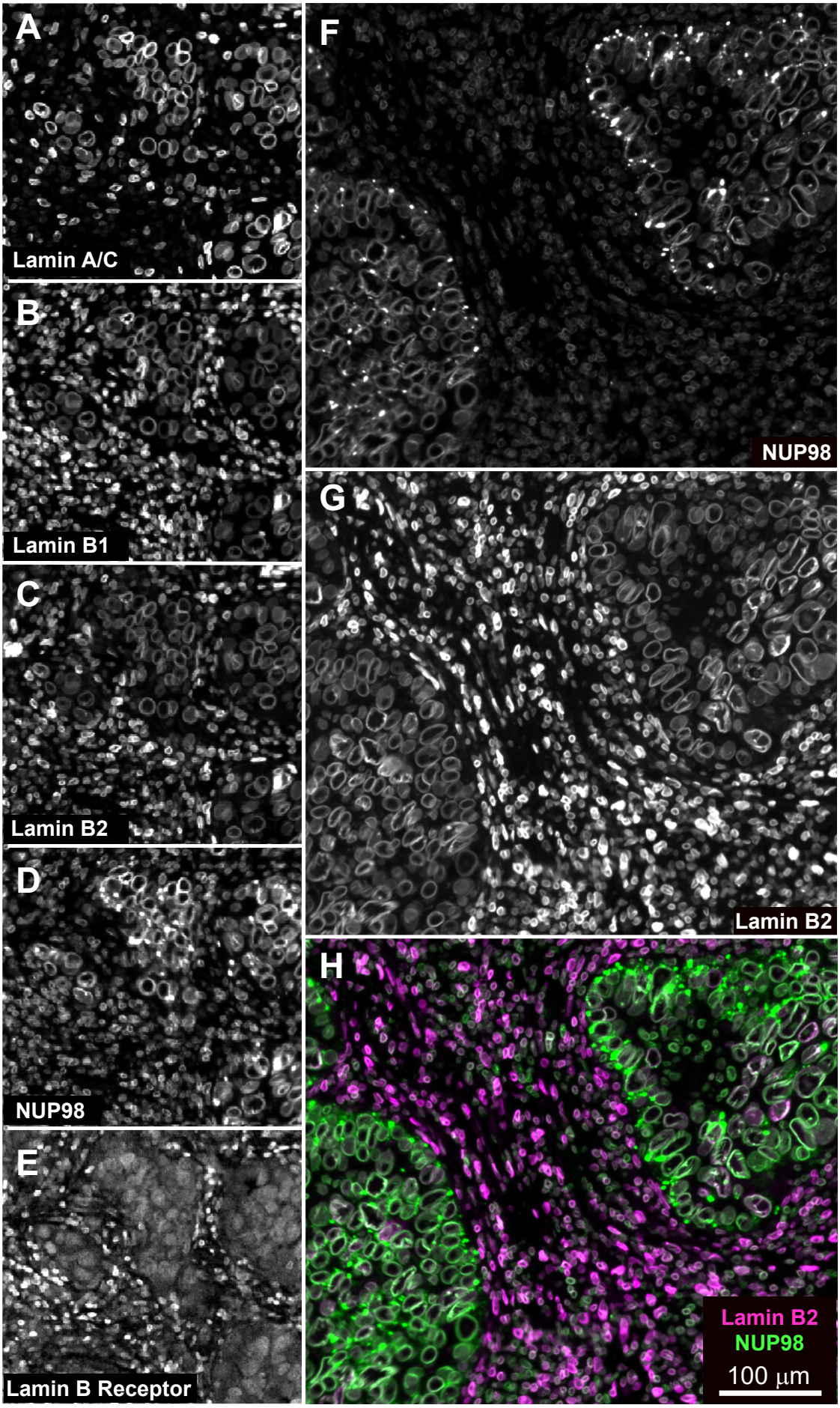

A
Lamin A/C
B
Lamin B1
C
Lamin B2
D
NUP98
E
Lamin B Receptor
F
NUP98
G
Lamin B2
H
Lamin B2
NUP98
100 μm

Figure 3

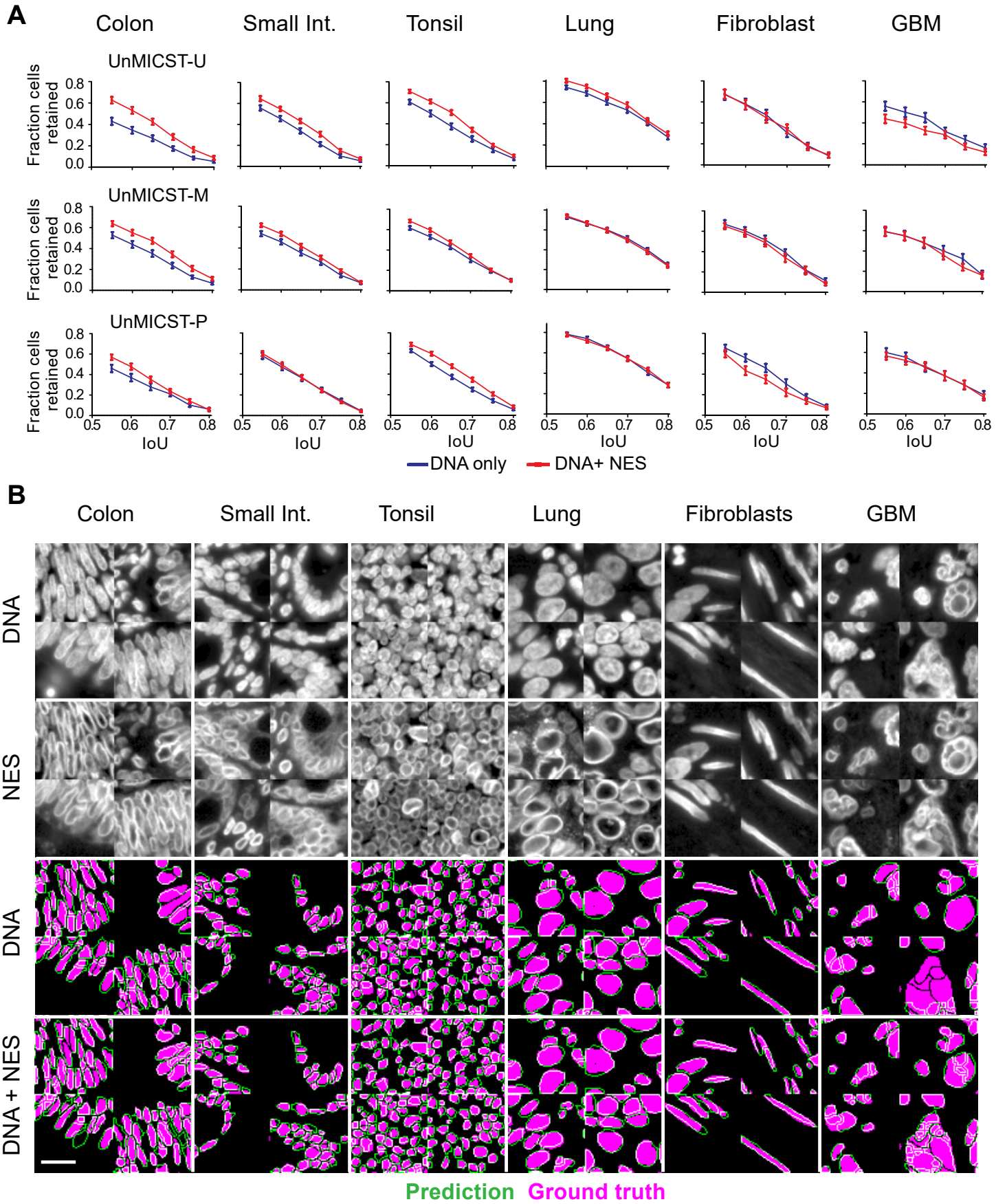
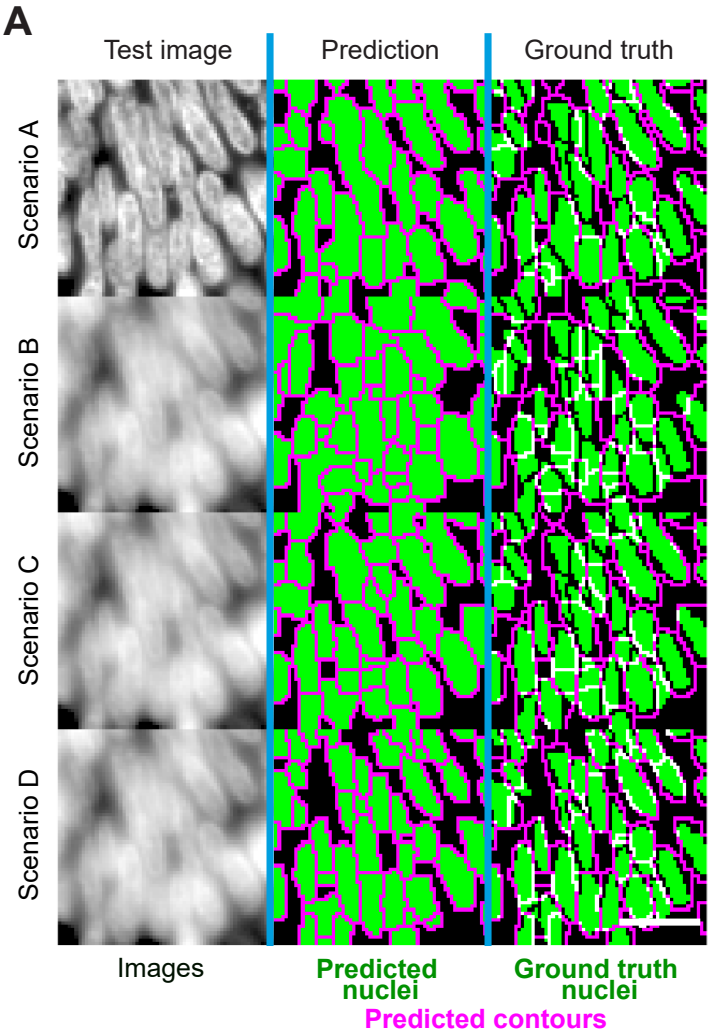

A
Colon
Small Int.
Tonsil
Lung
Fibroblast
GBM
UnMICST-U
UnMICST-M
UnMICST-P
Fraction cells retained
0.0
0.2
0.4
0.6
0.8
0.5
0.6
0.7
0.8
IoU
DNA only
DNA+ NES
B
Colon
Small Int.
Tonsil
Lung
Fibroblasts
GBM
DNA
NES
DNA
DNA + NES
Prediction
Ground truth

Figure 4



Scenario	DNA	NES	Real augmen-tation	Defocused test data
A	Yes	No	No	No
B	Yes	No	No	Yes
C	Yes	Yes	No	Yes
D	Yes	Yes	Yes	Yes

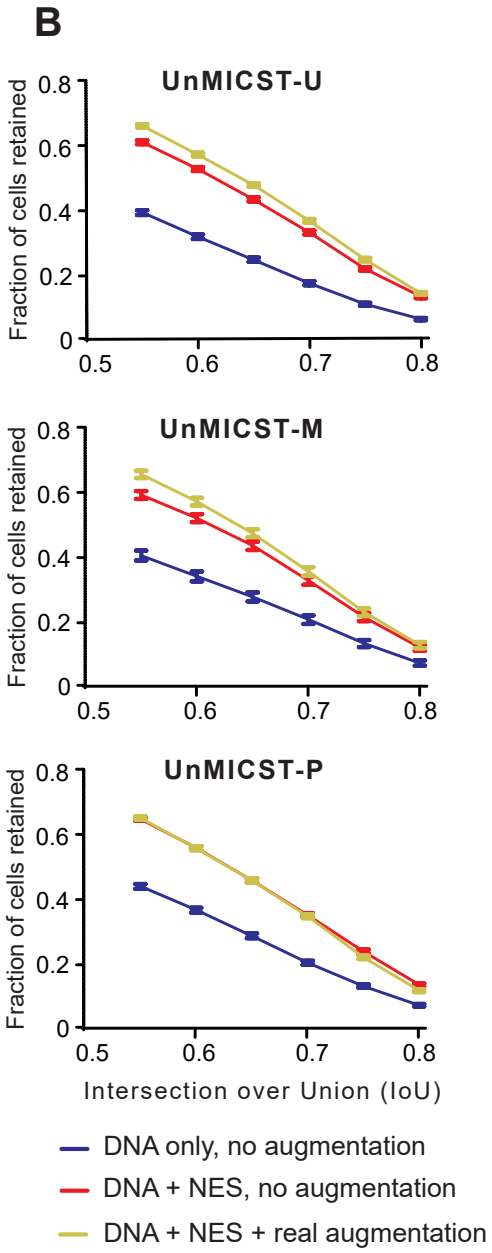


Figure 5

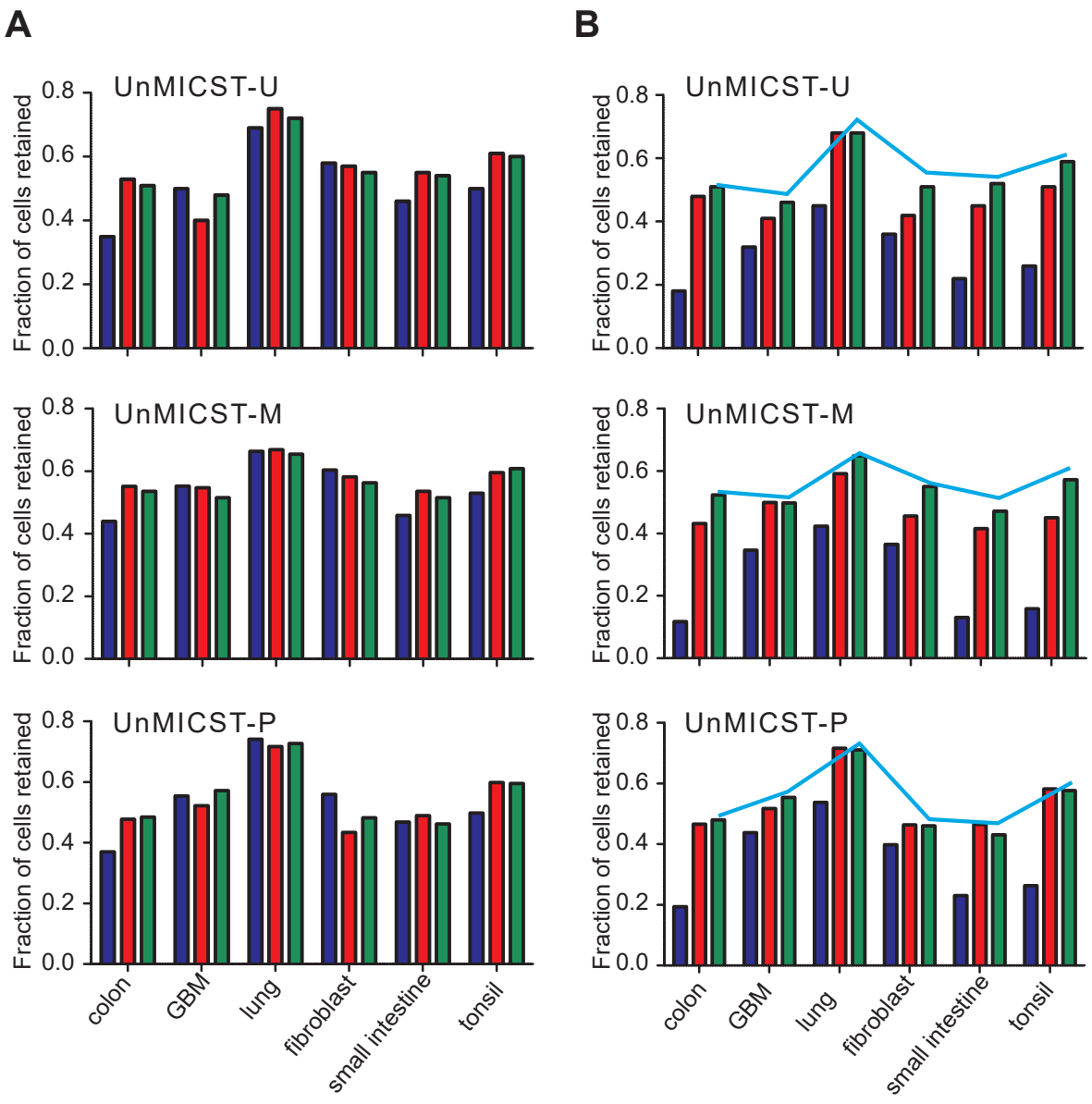
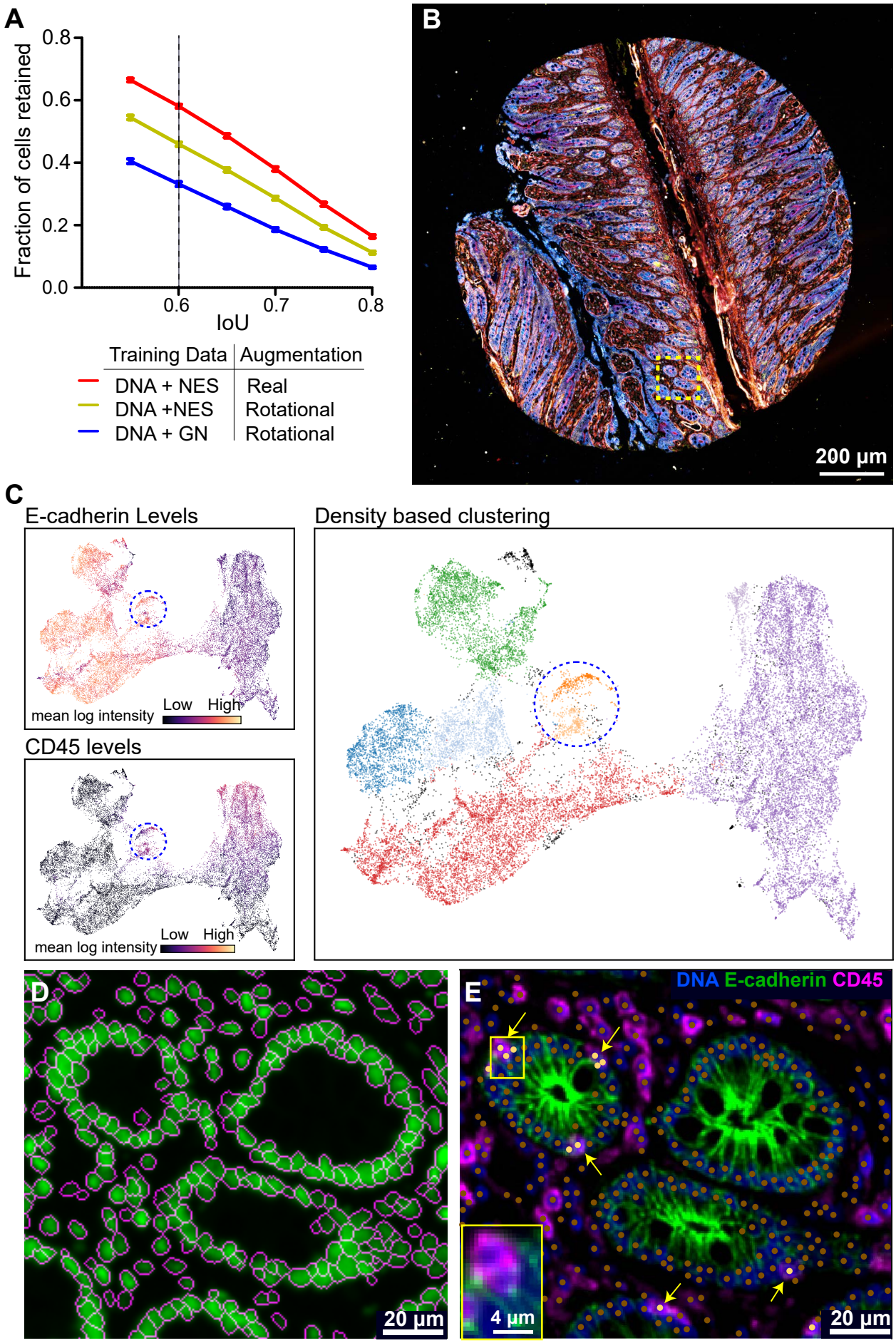

A
B
UnMICST-U
UnMICST-M
UnMICST-P
Fraction of cells retained
0.0
0.2
0.4
0.6
0.8
colon
GBM
lung
fibroblast
small intestine
tonsil

Figure 6


A
Fraction of cells retained
0.8
0.6
0.4
0.2
0.0
0.6
0.7
0.8
IoU
Training Data
Augmentation
DNA + NES
Real
DNA +NES
Rotational
DNA + GN
Rotational
B
200 µm
C
E-cadherin Levels
mean log intensity
Low
High
CD45 levels
mean log intensity
Low
High
Density based clustering
D
20 µm
E
DNA E-cadherin CD45
4 µm
20 µm