ISCI, Volume *21*

Supplemental Information

EcSeg: Semantic Segmentation of Metaphase Images Containing Extrachromosomal DNA

Utkrisht Rajkumar, Kristen Turner, Jens Luebeck, Viraj Deshpande, Manmohan Chandraker, Paul Mischel, and Vineet Bafna

Supplemental Table Titles and Legends

Supplemental Table 1: **Performance on different neural architectures.** The table reports (1) mIOU scores of ecDNA, chromosomes, nuclei, and cytoplasm, and (2) precision and recall scores for ecDNA for each variant of neural architecture tested. Related to Figure 2a, b, c.

Supplemental Table 2: **Precision and recall scores for ecSeg and ecDetect on entire data set.** The table reports (1) the precision and recall scores for ecSeg and ecDetect, (2) ground truth ecDNA counts per image, and (3) the predicted numbed of ecDNA from ecSeg for each of the 483 images in the data set. Related to Figure 2f.

Supplemental Table 3: **Performance on test data set.** Precision and recall scores from ecSeg and ecDetect for the 7 cell lines in the test set. Related to Figure 2d,e.

Supplemental Table 4: **Entropy.** The entropy and entropy efficiency for all cell lines present across the entire data set (training, validation, and test). Related to Figure 2g.

Supplemental Table 5: **Drug treatment ecDNA counts** Sheet 1 has raw ecDNA counts for both control and case for week 0,2, and 4. Sheet 2 has the entropy values for the cases in week 0, 2, and 4. Related to Figure 3d, e, f, g.

Transparent Methods

Data set. We started with a data set from (Turner et al. 2017). To capture relevant spatial information, cells were cultured according to standard protocol, and Karyomax was added to enrich for cells in metaphase. Cells were collected and treated with a 0.075 M KCl hypotonic solution for 10 minutes, followed by fixation in 3 : 1 methanol/glacial acetic acid solution. Interphase and mitotic cells were dropped onto humidified glass slides, and mounting medium with DAPI was applied to the slides. Cells in metaphase were imaged with an Olympus BX43 microscope equipped with a QiClick CCD camera. No 3D imaging was performed. Our dataset contains 483 images of dimensions 1392×1040 sampled from 27 different tumor cell lines. All images were stained with 4,6-diamidino-2-phenylindole (DAPI). DAPI is a blue-fluorescent stain that binds to any DNA structure represent in the sample. Thus, in our data set, it defines ecDNA, chromosomal, and nucleic regions. Some components in the image are also stained with fluorescence in situ hybridization (FISH) for specific probes on the ecDNA. However, we ignored the FISH signals when constructing our ground truth as (a) some ecDNA may not carry the probe target due to heterogeneity, and (b) not all targets are bound by the probe. Thus, extrachromosomal FISH signals validate ecDNA, but absence of FISH signals is not indicative of a lack of ecDNA.

We cropped these 483 images into 9,660 patches of 256×256. Some patches were purely background and we only included patches with at least 1% of the total area being covered in DAPI. We were left with 5949 usable patches. We split this data set such that 60% was used for training (3570 patches), 20% for validation (1190 patches), and the final 20% for reporting test results (1189 patches).

Ground Truth Labeling. Manual identification of ecDNA can be laborious as a single image can easily contain more than 200 ecDNA elements, sometimes up to 500. Thus, we built a software, using off-the-shelf morphological operations, to toggle a region as being ecDNA or not. The ground truth was then obtained through a manual annotation process using that software. To reduce the annotator's work, we seeded the process by providing ecDetect annotations which the annotator could then toggle on or off.

We used Otsu's thresholding to binarize the gray-scale image (**Otsu1979**). The adaptive method demarcated the nuclei and chromosomes, but the smaller and lower intensity ecDNA were marked as background. We smoothed the edges of the chromosomes and nuclei by performing an *open* operation, which is an erosion of the connected components followed by a dilation. We next used Bradley local thresholding (**Bradley2007**), an adaptive thresholding algorithm, to perform ecDNA annotations. Bradley local thresholding uses a sliding average filter and checks if the brightness of the center pixel is $T\%$ lower than the mean intensity of the pixels in the window. If it is lower, then the pixel is set to black or otherwise set to white. We used a window size of 3×3 pixels and a threshold value of $T = 3\%$. This allowed us to segment the image to a finer resolution with ecDNA predictions. We post-processed ecDNA segmentation by removing stray components that were less than 15 pixels in size, filling in any holes, removing spurs, and performing an *open*

operation on each of the connected components. Notably, the process missed many true ecDNA, but the coarse segmentation was useful for training the U-net.

However, for the 96 test set images (1189 patches), where we needed a more precise accounting of false negative and false positives, we used additional annotators who refined the predictions by manually examining each image and correcting any ecDNA that were falsely classified during the coarse annotation.

Segmentation. Inspired by the U-Net, we used a modified fully convolutional neural network presented in Fig. 1f. We optimized the architecture by performing grid search over the network's hyper-parameters. We varied the number of filters in the first layer ($\{16, 32, 64\}$), input patch sizes ($\{128^2, 256^2, 512^2\}$) and L2 regularization ($\{1, 0.1, 0.01, 0.001, 0.0001\}$). We applied multi-scale context aggregation using dilated convolution (Yu and Koltun 2015). We found that although the chromosomal IoU increased, the ecDNA precision and recall remained the same. We also experimented with pre-trained weights from VGG16 trained on ImageNet. However, because ImageNet contains images of everyday objects, our model had a more difficult time generalizing to the microscopy images. In each case, we minimized loss on the network variants using the Adam optimizer on 8 GeForce GTX 1080 Ti GPUs. We trained the network on the training set and used the validation set to evaluate loss and mIoU. The training was halted if the loss on the validation set did not change for 7 epochs (the 'patience' time). The test data was a "holdout" set that was only used for final quantification of the model and had no direct effect on the training itself. The performance was optimized on a network with 32 filters in the first layer and doubling the number of filters in each layer, input image sizes of 256×256, and a L2 regularization parameter of 0.0001.

We decided not to perform any data augmentation through warping and stretching. The relative size and shapes of ecDNA are very critical, and often times, certain ecDNA are almost the size of chromosomes, such as in the COLO205 cell line Supplementary Fig. 1. Any warping and stretching could cause the ecDNA and chromosomes to be indistinguishable even for the human eye. Rotations were not used either as our images have no rotational significance. All the images were taken from a top-down view with no bias towards orientation. Finally, as we collected data from a large number of cell lines, we had sufficient variation in our dataset.

We denoted each ground truth image as a collection of pixels $\mathcal{P}$ with the goal of classifying the pixels into one class from $\mathcal{C} = \{b, n, h, e\}$, representing background (b), nucleus (n), chromosome (h), and ecDNA (e). The ground truth was described by a binary function $y_c(x) \in \{0, 1\}$ for all $x \in \mathcal{P}$, $c \in \mathcal{C}$. Additionally, $\sum_c y_c(x) = 1$ for all pixels, enforcing a single class assignment. For each $x \in \mathcal{P}$, $c \in \mathcal{C}$, the network outputs a class score, $P_c(x) \in [0, 1]$. We trained the network to minimize a custom loss function defined below.

Loss function. We defined loss L as a weighted binary cross entropy (BCE) minus the Sørensen-Dice coefficient (Dice). Specifically, the BCE loss for class c was computed using:

$$\mathrm{BCE}[x] = -\frac{1}{|\mathcal{C}|} \sum_{c \in \mathcal{C}} \left[y_c(x) \ln \left(\frac{1}{1 + e^{-P_c(x)}} \right) + (1 - y_c(x)) \ln \left(1 - \frac{1}{1 + e^{-P_c(x)}} \right) \right].$$

Similarly, we compute Dice loss as:

$$\text{Dice} = \left[1 - \frac{2\sum_c \mathbf{P_c} \cdot \mathbf{y_c}}{\sum_c (\|\mathbf{P_c}\|_1 + \|\mathbf{y_c}\|_1)}\right]$$

We used weights to boost the under-represented classes. Let n_b, n_n, n_h, n_e denote the total number of pixels belonging to each class in background, nuclei, chromosome, and ecDNA, respectively, for the entire training and validation dataset. As $n_b >> n_n > n_h > n_e$, we assigned weight w_c to each class $c \in \{b, n, h, e\}$ as follows:

$$w_c = \max\left\{1, \frac{n_n}{n_c}\right\}$$

Correspondingly, the weight of a pixel was given by:

$$w_x = \sum_c y_c(x) w_c \tag{3}$$

and the net loss was computed using

$$L = \frac{1}{|\mathcal{P}|} \sum_x w_x(\text{BCE}[x] + \text{Dice})$$

To prevent over-fitting, we trained for 45 epochs with an early stopping "patience" of 7 which stopped training if the loss on the Validation set did not improve for 7 epochs.

Accuracy. For each class c, and threshold $\tau \in \mathcal{T}$, where $\mathcal{T} = \{0.05, 0.1, 0.5\}$, define an indicator $\theta_{c,\tau}(x) = \{1 \text{ if } P_c(x) \geq \tau; 0 \text{ otherwise}\}$. Define the mean Intersection over Union (mIoU) score across all classes as:

$$M = \frac{1}{|\mathcal{T}|} \sum_\tau \frac{1}{|\mathcal{C}|} \sum_c \frac{\boldsymbol{\theta}_{c,\tau} \cdot \mathbf{y_c}}{\|\boldsymbol{\theta}_{c,\tau}\|_1 + \|\mathbf{y_c}\|_1}$$

Post-processing of segmentation. Post-training, the network outputs a 256×256 matrix O, with

$$O[x] = \arg\max_c P_c(x)$$

To filter noise, we computed connected components for each class. Connected components are regions of adjacent pixels with the same class value. We filled all holes in each of the connected components such that the hole is assigned the same class as the surrounding pixels. We performed secondary size thresholding on the ecDNA elements such that all ecDNA components less than 15 pixels are marked as background and those greater than 125 pixels are marked as chromosomes. We also removed any ecDNA that were attached to the edges of chromosomes or nuclei as these regions are often just spurs of the larger class.

Accuracy Metrics. To compute component level accuracy, we computed true positive, false positive, and false negative rates. If the centroid of a predicted ecDNA component was within a 5 pixel euclidean distance of the centroid of a ground truth ecDNA component, we marked this as

a true positive (TP). If there are no ground truth ecDNAs within that distance, we classified the component as a false positive (FP). We found that the average area of ecDNA across our entire dataset was 75 pixels and thus a distance threshold of $\sqrt{75/\pi} \simeq 5$ pixels ensures that ecDNAs detected on the periphery of the boundary from the annotated center pixel is still considered a true positive. Inversely, if there were no predicted ecDNAs within a 5 pixel distance of a ground truth annotation, we classified it as a false negative (FN). We compute our precision and recall for each image as:

$$\text{precision} = \frac{\text{TP}}{\text{TP} + \text{FP}}$$

$$\text{recall} = \frac{\text{TP}}{\text{TP} + \text{FN}}$$

We also measured accuracy using the F1 score, a harmonic average of precision and recall.

$$\text{F1} = 2 \times \frac{\text{precision} \times \text{recall}}{\text{precision} + \text{recall}}$$

Entropy and Entropy efficiency. Consider a sample with n cells. Let n_i (respectively $p_i = \frac{n_i}{n}$ denote the number (respectively, fraction) of cells with i copies. We defined heterogeneity of copy number using Shannon entropy:

$$\mathcal{H}_n = -\sum_i p_i \log_2 p_i,$$

The entropy efficiency, defined by $\frac{\mathcal{H}_n}{\log_2 n}$ normalizes the value between 0 (no heterogeneity) and 1 (maximum heterogeneity).

Drug Treatment Quantification. We cultured GBM39 cells as neurospheres under serum-free conditions (DMEM/F12 basal media with 1X Glutamax, EGF, FGF, and heparin). Cells were cultured in 5 uM Erlotinib. The EGFR-containing ecDNA was quantified via ecSeg at 0, 2, and 4 weeks.

Evolutionary model for ecDNA driven copy number. Consider an initial population of cells, with each cell carrying $k \geq 0$ copies of an oncogene on ecDNA. We modeled the population using a discrete generation Galton-Watson branching process (Bozic et al. 2010). In this simplified model, each cell in the current generation containing k amplicons (amplifying an oncogene) either dies with probability d_k, or replicates with probability b_k to create the next generation. We set the selective advantage

$$\frac{b_k}{d_k} = \begin{cases} 1 + f_{m,\alpha}(k), & 0 \leq k < M_a \\ 0 & \text{otherwise} \end{cases} \tag{4}$$

$$d_k = 1 - b_k \tag{5}$$

In other words, cells with k copies of the amplicon stop dividing after reaching a limit of M_a amplicons. Otherwise, they have a selective advantage for $0 < k \leq M_a$, where the strength of selection ($b_k - d_k \propto f_{m,\alpha}(k)$) is governed by parameters m, α. Initially, the selective advantage

increases with increasing copies, but later diminishes due to increasing metabolic load. We modeled this by defining

$$f_{m,\alpha}(k) = \begin{cases} \frac{k}{M_s} & (0 \leq k \leq M_s)\ , \\ \frac{1}{1+e^{\alpha(k-m)}} & (M_s < k < M_a). \end{cases} \tag{6}$$

Here, parameters m and α are the 'mid-point', and 'steepness' parameters of the logistic function, respectively. Initially, $f_{m,\alpha}(k)$ grows linearly, reaching a peak value of $f_{m,\alpha}(k) = 1$ for $k = M_s$. As the viability of cells with large number of amplicons is limited by available metabolites (**Pavlova2016**), $f_{m,\alpha}(k)$ decreases logistically in value for $k > M_s$ reaching $f_{m,\alpha}(k) \to 0$ for $k \geq M_a$. We model the decrease by a sigmoid function with a single mid-point parameter m s.t. $f_{m,\alpha}(m) = \frac{1}{2}$. The 'steepness' parameter α is automatically adjusted to ensure that $\max\{1 - f_{m,\alpha}(M_s), f_{m,\alpha}(M_a)\} \to 0$. We empirically chose $M_a = 20, m = 100, \alpha = 0.1$ to match a mean copy number of 50 ecDNA per cell observed prior to drug treatment.

The addition of a drug targeting the oncogene provides a disadvantage (negative fitness) to cells carrying extra copies of the oncogene. Therefore, after drug treatment, we used the selective function

$$f_{r,\alpha}(k) = -\frac{e^{\alpha(k-r)}}{1 + e^{\alpha(k-r)}}. \tag{7}$$

$f_{r,\alpha}(k)$ provides negative selection pressure causing a steep decline in the average number of ecDNA per cell. We simulated the effect of the drug using $r \in \{5, 20, 50, 100\}$, $\alpha \in \{0.07, 0.04, 0.03\}$. Supplementary Figure 5 shows the values for $\alpha = 0.04$. We observed that $r = 20, \alpha = 0.04$ best matched the empirical observations with Eb treatment (Figure 3d).

FISH analysis. ecSeg also incorporates FISH analysis. It allows the user to specify the color of the FISH signal used to illuminate the gene of interest and the intensity threshold T ($T = 120$ by default). It then extracts binary images highlighting only the pixels that have the minimum intensity in the appropriate color channel and additionally marks the pixels as either ecDNA or chromosomes. ecSeg outputs a table containing the total number of FISH pixels, the fraction of FISH pixels that are also marked as ecDNA, and the fraction marked as chromosomal for each image in the user-specified file path.

DATA AND SOFTWARE AVAILABILITY

ecSeg is available at https://github.com/ucrajkumar/ecSeg. The accession number for the data reported in this paper is 10.17632 : m7n3zvg539.3.

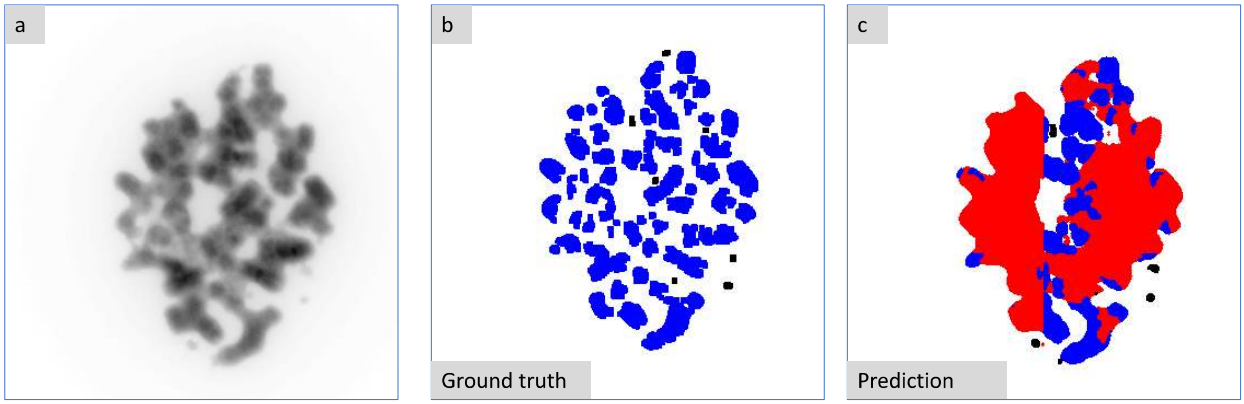
Supplemental References

1. Bradley, D., Roth, G., 2007. Adaptive thresholding using the integral image. J. Graph. Tools 12, 1321.

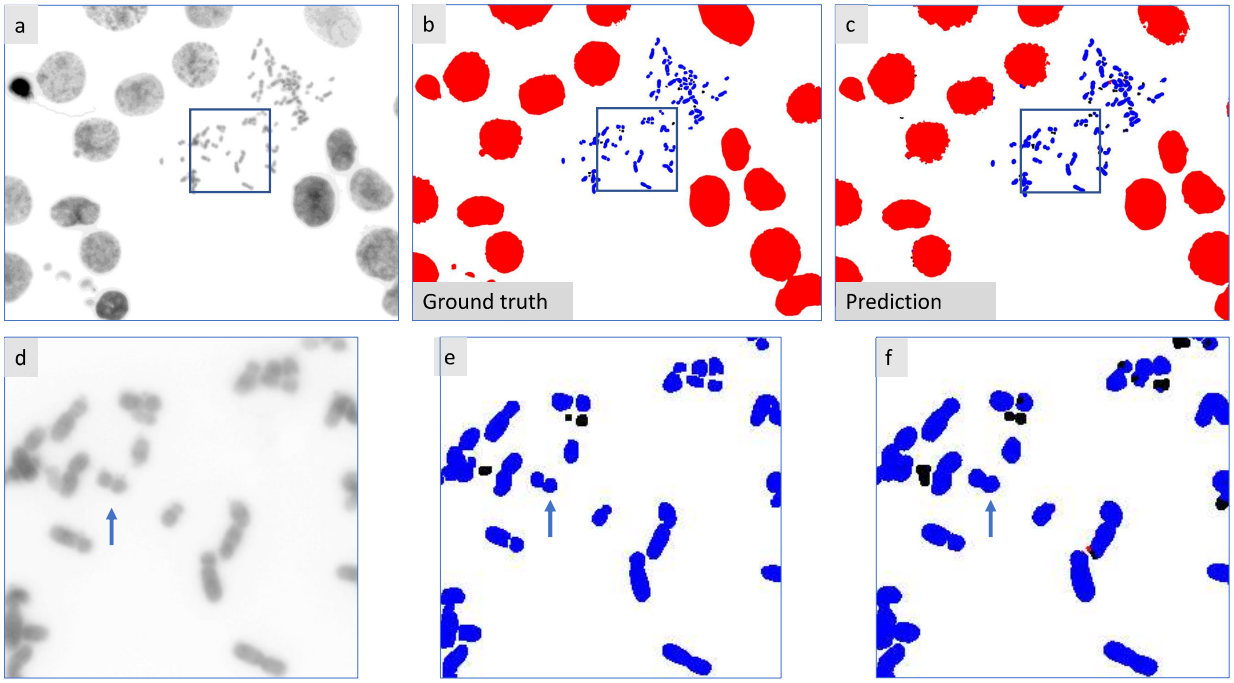
2. Otsu N., 1979. A threshold selection method from gray-level histograms. IEEE Trans. Sys. Man. Cyber. 9 (1): 6266.

3. Pavlova, N.N., Thompson, C.B., 2016. The emerging hallmarks of cancer metabolism. Cell Metab. 23, 2747.

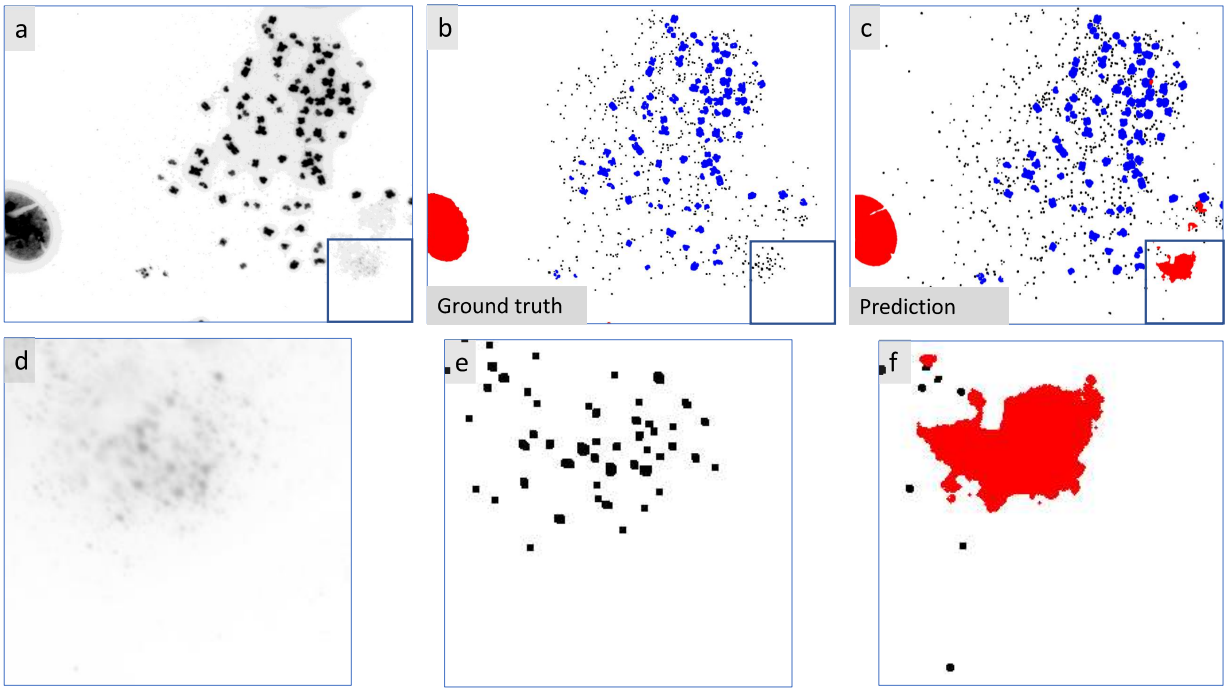
Supplemental Figures



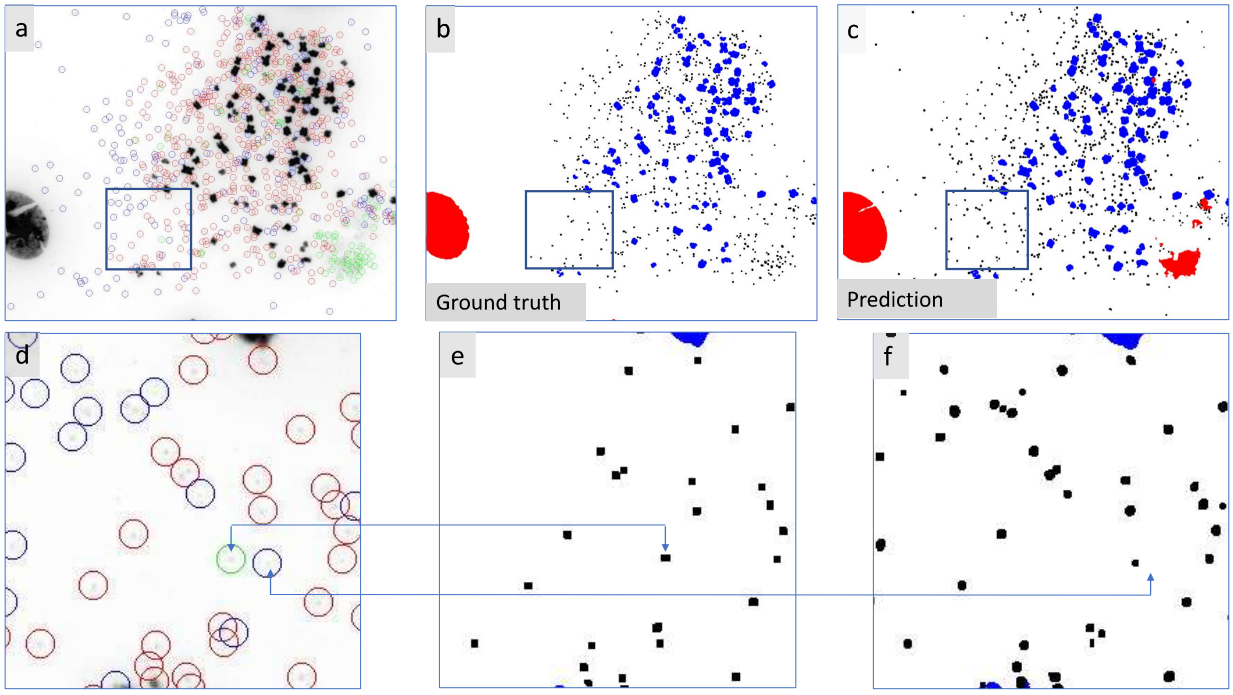
Supplemental Figure 1: **Incorrect classification of chromosomes as nuclei in COLO205.** (a) DAPI of original image from cell line COLO205. (b) Ground truth annotation with intact nuclei, chromosomes, and ecDNA being represented by red, blue, and black, respectively. (c) Segmentation map. COLO205 tumor cell remain tightly clumped even after the nucleic membrane has disintegrated. The network mis-classifies these chromosomes as nuclei due to the tight clustering. Related to main Fig. 2b.



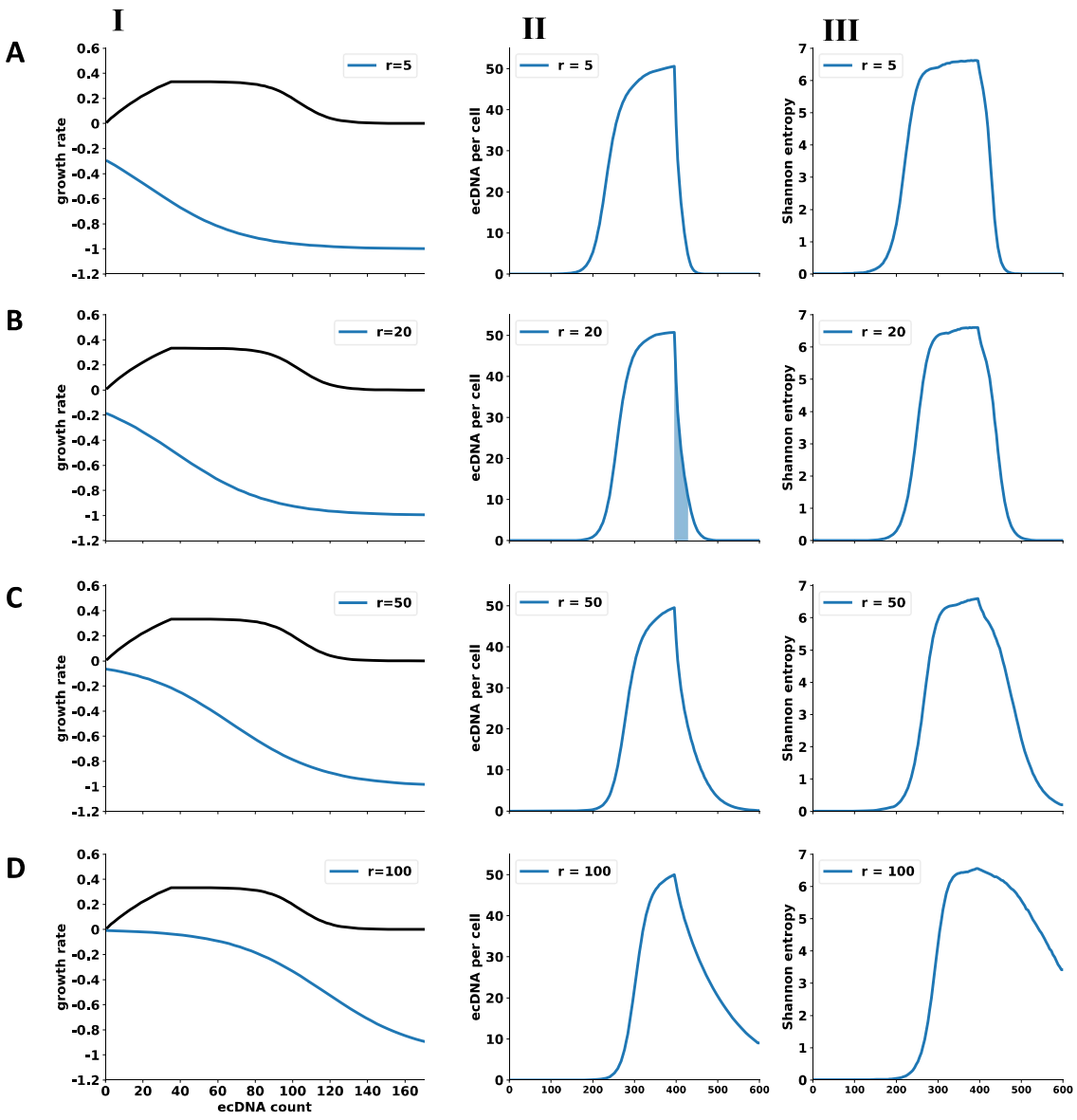
Supplemental Figure 2: **Incorrect detection of large ecDNA in COLO205.** (a) DAPI of original image from cell-line COLO205. (b) 'Ground truth' annotation with intact nuclei, chromosomes, and ecDNA being represented by red, blue, and black, respectively. (c) ecSeg Segmentation map. (d,e,f) Crops of DAPI, ground truth annotation, and ecSeg segmentation. In COLO205, replicating ecDNA structures (double minutes) often closely resemble chromosomes, making it difficult to identify. These structures are marked as chromosomes in both the ground truth and the segmentation map. Related to main Fig. 2b.



Supplemental Figure 3: **Incorrect false negative calls in cell line CA718.** (f) is burst nucleus, but appears to show as ecDNA when zoomed in, and was marked as ecDNA during human annotation. ecSeg correctly annotates it as a nucleus identifying a mistake in the human annotation. Related to main Fig. 2b.



Supplemental Figure 4: **Incorrect annotation of ecDNA in cell line CA718.** (a) DAPI of original image from cell-line CA718. (b) 'Ground truth' annotation (c) ecSeg Segmentation map (d,e,f) Crops of DAPI, ground truth annotation, and ecSeg segmentation. Blue circles denote false positives, red circles are true positives, and green circles are false negatives. As can be verified by looking at the DAPI image, many of the annotated false positives are actually true ecDNA with low-intensity DAPI signals. These ecDNA were missed during the ground truth annotation. False negatives are rare, and often indicate a problem with the ground truth annotation, as shown in Supplementary Fig. 3. Related to main Fig. 2d.



Supplemental Figure 5: **Simulating the impact of drug on ecDNA counts and heterogeneity.** Column I shows the modeled growth rates $b_k - d_k$ as a function of ecDNA count (k) for untreated (black line) and drug-treated (blue) lines, for $\alpha = 0.04$, and $r \in \{5, 20, 50, 100\}$ (rows A-D). Columns II and III show simulated changes in the mean copy number and Shannon entropy as a function of time, when the drug is applied at day 400. Upon drug application, the ecDNA counts and heterogeneity both decline in a manner dependent upon the the strength of selection modeled using α, r. Panel B.II ($r = 20, \alpha = 0.04$; shaded region) best fit the experimental data of GBM cells treated with Erltonib (related to main Figure 2h).