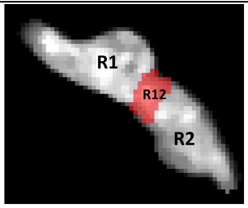
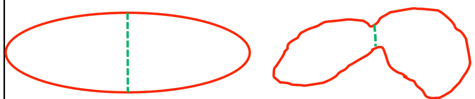
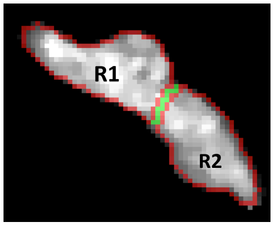
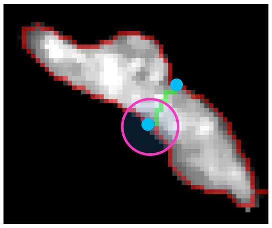
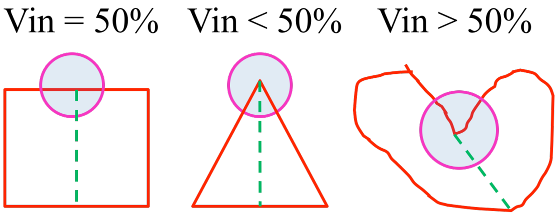
Supplementary Table 1: Features used by the learning-based region merging model of our nuclei segmentation algorithm

Appearance Similarity Features	Brightness similarity	We first compute the median intensities of the two regions and then compute the ratio between the minimum of the two median intensity values with the maximum of the two values.
	Similarity of Intensity Distributions	We compute the 64-bin intensity histogram for each of the two regions and measure the similarity between the two histograms using the well-known bhattacharya-coefficient.
	Texture Similarity	We compute nine haralick texture features (Contrast, Homogeneity, Energy, Entropy, Correlation, AutoCorrelation, ClusterShade, ClusterProminence, MaximumProbability) for each of the two regions at 3 different scales (GLCM neighborhood offsets of 1, 2, and 4 voxels with the offsets along z-axis adjusted depending on the anisotropy in the voxel size) resulting in a 27-dimensional vector that quantifies the texture of the region. The similarity in texture between the two regions is then computed using the cosine similarity measure.
Boundary Saliency Features	Watershed saliency (quantifies the amount of edge information in the watershed region)	 Let R1 and R2 be the given pair of adjacent regions and let R12 be a dilated version of the common boundary shared between them which we henceforth refer to as the watershed region (see 2D illustration on right). We quantify the saliency or edgeness of the watershed region using two types of measures: • Ratio of the median intensity (M_R12) of the watershed region with the median intensities (M_R1 and M_R2) of each of the given pair of adjacent regions. Specifically, we use as features the ratio of M_R12 with (i) minimum of M_R1 and M_R2, and (ii) maximum of M_R1 and M_R2. • Ratio of inter-quartile range (IQR_R12) of the watershed region with the inter-quartile ranges (IQR_R1 and IQR_R2) of the given pair of adjacent regions. Specifically, we use as features the ratio of IQR_R12 with (i) minimum of IQR_R1 and IQR_R2, and (ii) maximum of IQR_R1 and IQR_R2.
Boundary Continuity Features These features quantify the likelihood of merging the two given regions from a geometric point of view	Amount of common boundary	We first compute an estimate of areas of the bounding surfaces (A1 and A2) of the two given regions and the area of the common boundary (A12) shared by both of them. We then use as features the ratio of A12 with (i) minimum of A1 and A2, and (ii) maximum of A1 and A2. Shown below is a 2D schematic illustrating how this feature could help in determining whether or not the merge a given pair of adjacent regions  OK To Merge NOT OK To Merge
	Convexity Measures	We often observed that when a given pair of adjacent regions belong to two distinct nuclei then merging them results in a highly non-convex region. To capture this insight, we use as features (i) convexity of the merged region, and (ii) the ratio of the convexity of the merged region with the convexity values of the individual regions.

	Gap between the bounding surfaces of the two regions	Here, we intend to measure the gap created by the watershed algorithm between the given pair of adjacent regions (see green region in 2D illustration on right). Because of the anisotropic voxel spacing in our data, this gap varies depending on the way two adjacent regions touch each other geometrically in 3D.  To quantify the gap, we first consider one of the two given regions as the reference region and determine the part of its boundary that touches/interacts with the other region, which we denote here as B. For each voxel in B, we compute the distance, in physical space (microns), to the closest voxel on the boundary of the other region. We then compute the minimum, maximum, mean, and standard deviation of these distance values as features.
	Junction/Neck Curvature	We compute a coarse estimate of the curvature at each voxel on the neck (see blue pixels in 2D illustration on right). Specifically, for each voxel p on the neck, we compute the fraction of foreground (cell nuclei) voxels inside a sphere of a specified size centered at the neck voxel, which we denote as Vin(p).  When the continuity between the bounding surfaces of the two given regions at a neck voxel is smooth (see 2D schematic in Fig. (a) below) approximately 50% of the voxels inside the sphere will belong to the foreground and when the continuity between the two bounding surfaces at a neck voxel is non-smooth (see 2D schematics in Figs. (b) and (c) below). Based on this insight, we compute the curvature at a neck voxel p as C(p) =|Vin(p) - 0.5| where |.| denotes the absolute value. Observe that C(p) will be close to zero whenever the continuity is smooth and increases as the continuity is non-smooth.  Low Curvature (a) High Curvature (b) High Curvature (c) We compute the aforementioned curvature measure for all junction voxels and then compute the min, max, mean, and standard deviation of these values as features.

Supplementary Table 2: Performance statistics of the nuclei segmentation algorithm

	Mean +/- stddev (%)	**Range (%)**
Segmentation Accuracy	91.04 +/- 4.87	[77.61, 97.89]
Over-segmentation Errors	3.82 +/- 2.65	[0.70 14.00]
Under-segmentation Errors	4.91 +/- 3.42	[0.00 12.68]

Supplementary Table 3: Comparison of segmentation accuracy obtained using our nuclei segmentation algorithm with and without region merging on a small 5-volume test dataset

	Mean +/- stddev (%)	**Range (%)**
Proposed method with region merging	94.76 +/- 2.14	[92.59, 97.88]
Proposed method without region merging	87.70 +/- 2.11	[84.23, 89.87]

Supplementary Table 4: Features used for cell cycle state identification

Histone Features	Intensity statistics (computed from the voxels inside cell mask in the histone channel)	Entropy of 8-bin histogram Median intensity Median absolute deviation (MAD) Inter-quartile range (IQR) Robust measure of skewness* Robust measure of kurtosis* Median of Local standard deviation in a 3x3 neighborhood
	Haralick texture features at 4 scales (Intensities of the cell voxels were normalized using the minimum and maximum intensities of the cell; Gray-level Co-occurence Matrix was computed in a rotation invariant fashion using 64 gray-levels and neighborhood offsets of 1, 2, 4, and 8 voxels wherein the offset in Z-dimension was adjusted based on the anisotropy in the voxel spacing)	Contrast Homogeneity Energy Entropy Correlation AutoCorrelation ClusterShade ClusterProminence MaximumProbability
	Shape features (computed both from the cell mask in 3D and its maximum intensity projection (MIP) along the z-axis)	Convexity of MIP Eccentricity of MIP Circularity of MIP 3D convexity
FUCCI Features	Intensity statistics (computed separately from the voxels inside the cell mask in FUCCI-GFP and FUCCI-RFP channels)	Entropy of 8-bin histogram Median intensity Median absolute deviation Inter-quartile range Robust measure of skewness Robust measure of kurtosis Median of Local standard deviation in a 3x3 neighborhood
	Cross-channel ratios of a selected set of intensity statistics	Ratio of median intensity in FUCCI-GFP and Histone Ratio of MAD in FUCCI-GFP and Histone Ratio of IQR in FUCCI-GFP and Histone Ratio of median intensity in FUCCI-RFP and Histone Ratio of MAD in FUCCI-RFP and Histone Ratio of IQR in FUCCI-RFP and Histone Ratio of median intensity in FUCCI-RFP and FUCCI-GFP Ratio of MAD in FUCCI-RFP and FUCCI-GFP Ratio of IQR in FUCCI-RFP and FUCCI-GFP
	Cross-channel colocalization measures using the pearson correlation coefficient	Pearson correlation between FUCCI-GFP and Histone Pearson correlation between FUCCI-RFP and Histone Pearson correlation between FUCCI-GFP and FUCCI-RFP

*Kim, T.-H. & White, H. On more robust estimation of skewness and kurtosis. Finance Research Letters 1, 56-73 (2004).

Supplementary Table 5: Confusion matrix for the 4-class cell cycle state identification model

Human\Computer	**G1**	**Late G1/Early S**	**S/G2**	**mitotic**
G1	2801	173	30	8
Late G1/Early S	40	550	8	3
G2	12	100	630	45
M	1	5	19	382

Supplementary Table 6: Per-class performance of the 4-class cell cycle state identification model

	Precision	**Recall**	**F-Measure**
G1	0.98	0.93	0.95
Late G1/Early S	0.66	0.92	0.77
G2	0.92	0.80	0.85
M	0.87	0.94	0.90

Supplementary Table 7: Overall performance of the 4-class cell cycle state identification model

	Accuracy	**Average Precision**	**Average Recall**	**Average F-Measure**
10-fold CV on Training dataset	91.72%	0.88	0.91	0.89
Test dataset	90.76%	0.86	0.90	0.87

Supplementary Table 8: Confusion matrix for the Interphase vs mitotic classification model

Human\Computer	**Interphase**	**Mitotic**
Interphase	4287	113
Mitotic	11	396

Supplementary Table 9: Per-class performance of the Interphase vs Mitotic classification model

	Precision	**Recall**	**F-Measure**
Interphase	0.99	0.97	0.99
Mitotic	0.78	0.97	0.86

Supplementary Table 10: Overall performance of the Interphase vs Mitotic classification model

	Accuracy	**Average Precision**	**Average Recall**	**Average F-Measure**
10-fold CV on Training dataset	96.22%	0.89	0.97	0.92
Test dataset	97.42%	0.89	0.97	0.93

Supplementary Table 11: Confusion matrix for the G1 vs S/G2/M classification model

Human\Computer	**G1**	**S/G2/M**
G1	2901	111
S/G2/M	99	1696

Supplementary Table 12: Per-class performance of the G1 vs S/G2/M classification model

	Precision	**Recall**	**F-Measure**
G1	0.97	0.96	0.97
S/G2/M	0.94	0.94	0.94

Supplementary Table 13: Overall performance of the G1 vs S/G2/M classification model

	Accuracy	**Average Precision**	**Average Recall**	**Average F-Measure**
10-fold CV on Training dataset	97.04%	0.97	0.97	0.97
Test dataset	95.63%	0.95	0.95	0.95