

Supplementary Information of

Stain-free identification of cell nuclei using tomographic phase microscopy in flow cytometry

Daniele Pirone^{1,2}, Joowon Lim³, Francesco Merola¹, Lisa Miccio¹, Martina Mugnano¹, Vittorio Bianco¹, Flora Cimmino⁴, Feliciano Visconte⁴, Annalaura Montella^{4,5}, Mario Capasso^{4,5}, Achille Iolascon^{4,5}, Pasquale Memmolo¹, Demetri Psaltis^{3,**}, Pietro Ferraro^{1,*}

¹ CNR-ISASI, Institute of Applied Sciences and Intelligent Systems “E. Caianiello”, Via Campi Flegrei 34, 80078 Pozzuoli, Napoli, Italy

² DIETI, Department of Electrical Engineering and Information Technologies, University of Naples “Federico II”, Via Claudio 21, 80125 Napoli, Italy

³ EPFL, Ecole Polytechnique Fédérale de Lausanne, Optics Laboratory, CH-1015 Lausanne, Switzerland

⁴ CEINGE - Advanced Biotechnologies, Via Gaetano Salvatore 486, 80131 Napoli, Italy

⁵ DMMBM, Department of Molecular Medicine and Medical Biotechnology, University of Naples “Federico II”, Via Pansini 5, 80131 Napoli, Italy

* pietro.ferraro@cnr.it

** demetri.psaltis@epfl.ch

S1. CSSI algorithm

In order to describe the steps of the proposed CSSI algorithm, sketched in Fig. 3c, we exploit the 3D numerical cell phantom shown in Figs. 3a,b.

1. The 3D RI tomogram of the analyzed cell is centered in its $L_x \times L_y \times L_z$ array, that is then divided into distinct cubes, each of which has an edge measuring ε pixel, as shown in the central xz -slice in Fig. S1a.

The ε parameter is the resolution factor at which the 3D array is firstly analyzed. It must be an even number and, after dividing each side of the 3D array by ε , an odd number must be obtained. Therefore, each distinct cube contains ε^3 voxels (i.e. RI values). The cubes completely contained within the cell shell are the investigated cubes C_I (yellow cubes within the blue cell shell in Fig. S1a). The central cube is not an investigated cube, since it is taken as a reference cube C_R (green cube in Fig. S1a), which vertices have x -, y -, and z -coordinates taken from pairs (V_{1x}, V_{2x}) , (V_{1y}, V_{2y}) , and (V_{1z}, V_{2z}) , respectively. In Fig. S1b, we also display the 3D array from which the central xz -slice of Fig. S1a has been selected.

As discussed in the Main Text, the CSSI algorithm is based on the WMW test^{35,36}. It is a rank-based non-parametric statistical test, thus distributions do not have to be normal. With a certain significance level γ , it allows rejecting or not the null hypothesis H_0 for which two sets of values have been drawn from the same distribution. The significance level γ is the probability of making an error of 1st species, i.e. of rejecting the null hypothesis H_0 when it is true. The confidence level is defined as $1-\gamma$, i.e. it is the probability of not rejecting the null hypothesis H_0 when it is true. An important parameter in a statistical test is the p-value, which ranges from 0 to 1. The p-value is the observed significance level, i.e. the smallest significance level at which H_0 is rejected. It can be also defined as the probability of obtaining results at least as extreme as the results actually observed, when the null hypothesis H_0 is true. Therefore, a low p-value leads to reject the null hypothesis H_0 , because it means that such an extreme observed result is very unlikely when the null hypothesis H_0 is true. In fact, if the $p\text{-value} \geq \gamma$, H_0 is not rejected with significance level γ , while if $p\text{-value} < \gamma$, H_0 is rejected with significance level γ . Therefore, the greater the p-value the greater the confidence level with which two sets of values have been extracted from the same population. Hence, our algorithm performs multiple comparisons between the investigated cubes C_I and the reference one C_R through the WMW test, because we are assuming that the C_R voxels belong to the subcellular structure of interest, thus if a certain C_I has been drawn from its same distribution, then also the C_I voxels belong to the subcellular structure of interest. Without loss of generality, here we are describing and testing the method in the case of the nucleus segmentation. As discussed in the Main Text, for many kinds of

suspended cells (e.g., cancer cell lines) the central voxels of the cell belong to the nucleus, therefore we associate the reference cube C_R to the central cube of the 3D RI tomogram.

2. An adaptive threshold T_P is set according to the p-values computed through the WMW test between the investigated cubes C_I and the reference cube C_R . It is chosen as the maximum value less than or equal to τ , such that for at least one C_I it happens that $\text{p-value} \geq T_P$.
3. A first rough clustering is performed through repeated M -iterations loops, to create a preliminary nucleus set \mathcal{N}^P . For each of them
 - a. A temporary set \mathcal{N}^T is created with the RIs of the sole reference cube C_R .
 - b. At each of M iterations
 - i. A reference set \mathcal{R} is created by randomly drawing ε^3 values from the temporary set \mathcal{N}^T .
 - ii. For each investigated cube C_I , the corresponding p-value is computed with respect to the reference set \mathcal{R} through the WMW test.
 - iii. The investigated cubes C_I such that their $\text{p-value} \geq T_P$ are added to the temporary set \mathcal{N}^T .
 - c. After an M -iterations loop, all the investigated cubes C_I added to the temporary set \mathcal{N}^T are moved to the preliminary nucleus set \mathcal{N}^P , and then the temporary set \mathcal{N}^T is reset.
 - d. Steps a-c are repeated until at least n investigated cubes C_I have been stored within the preliminary nucleus set \mathcal{N}^P , which is shown in Fig. S1c.
4. A filtering operation is performed to delete outlier cubes from the preliminary nucleus set \mathcal{N}^P , thus creating a filtered nucleus set \mathcal{N}^F . Let $C_{\mathcal{N}^P, i}$ be a cube within \mathcal{N}^P , with $i = 1, 2, \dots, n$.
 - a. The reduced nucleus set \mathcal{N}_i^{P-} is created after removing the cube $C_{\mathcal{N}^P, i}$ from the preliminary nucleus set \mathcal{N}^P , with $i = 1, 2, \dots, n$.
 - b. A p-value vector \bar{p} of length n is created, which i -th element is the p-value computed through the WMW test between the cube $C_{\mathcal{N}^P, i}$ and the reduced nucleus set \mathcal{N}_i^{P-} .
 - c. A distance vector \bar{d} of length n is created, which i -th element is the Euclidean distance between the centre of cube $C_{\mathcal{N}^P, i}$ and point B , i.e., the centroid of the preliminary nucleus set \mathcal{N}^P .
 - d. The p-value vector \bar{p} is sorted in ascending order, thus obtaining the sorted p-value vector \bar{p}^S , shown in Fig. S1d.
 - e. The distance vector \bar{d} is sorted in ascending order and normalized to its maximum, thus obtaining the sorted distance vector \bar{d}^S , shown in Fig. S1e by blue dots.

Both vectors \bar{p}^S and \bar{d}^S are used to remove outlier cubes within the preliminary nucleus set \mathcal{N}^P . In fact, a cube $C_{\mathcal{N}^P,i}$ is considered an outlier if it is far from the centroid B and has a low p-value with respect to the other cubes in \mathcal{N}^P .

- f. A fourth-degree polynomial is fitted to the sorted distance vector \bar{d}^S , thus obtaining the vector \bar{d}^{SF} and its first difference \bar{D}^{SF} , that are reported in red in Figs. S1e,f, respectively.
- g. Let m be the index of the lowest value with null slope in vector \bar{D}^{SF} , as highlighted by the black dot in Fig. S1f. If in the vector \bar{D}^{SF} there is no point with null slope, m is chosen as the index of the global minimum.
- h. After computing thresholds T_1, T_2, T_3, T_4 , and T_5 , the filtered nucleus set \mathcal{N}^F reported in Fig. S1g is formed by cubes $C_{\mathcal{N}^F,i}$ that satisfy one of the following conditions

$$\begin{aligned} 1) \quad & \frac{d_i^{SF}}{d_{max}^{SF}} \leq T_1 \\ 2) \quad & T_1 < \frac{d_i^{SF}}{d_{max}^{SF}} \leq T_2 \ \& \ p_i^S > T_4, \\ 3) \quad & T_2 < \frac{d_i^{SF}}{d_{max}^{SF}} \leq T_3 \ \& \ p_i^S > T_5 \end{aligned} \tag{S1}$$

where $\&$ is the logical *and* operator, d_i^{SF} and p_i^S are elements of vectors \bar{d}^{SF} and \bar{p}^S , respectively, with $i = 1, 2, \dots, n$, and d_{max}^{SF} is the maximum value of vector \bar{d}^{SF} .

However, to build a filtered nucleus set \mathcal{N}^F , a strong spatial and statistical filtering has been made, in order to store only cubes that belong to the nucleus with high probability, thus leading to a strong underestimation of the nucleus region. Moreover, to increase the statistical power of the WMW test, the resolution factor ε should not be too small. As a consequence, the ε -cubic structuring element leads to a low spatial resolution.

5. A refinement step is performed, in order to transform the filtered nucleus set \mathcal{N}^F into a refined nucleus set \mathcal{N}^R , shown in Fig. S1h. For each cube $C_{\mathcal{N}^F,i}$ within the filtered nucleus set \mathcal{N}^F ,
 - a. Let the augmented cube $C_{\mathcal{N}^F,i}^A$ be the smallest cube centered in $C_{\mathcal{N}^F,i}$ with an edge multiple of ε px, such that the p-value computed through the WMW test between its RIs and all the \mathcal{N}^F values is greater than or equal to $\beta\mu(\bar{p})$, where $\mu(\cdot)$ is the average operator.
 - b. To enhance the resolution, the augmented cube $C_{\mathcal{N}^F,i}^A$ is in turn divided into distinct sub-cubes with edges measuring $\varepsilon/2$ px.
 - c. For each of these sub-cubes
 - i. Its $\varepsilon^3/8$ values are compared with $\varepsilon^3/8$ RIs randomly drawn from the filtered nucleus set \mathcal{N}^F .
 - ii. If the computed p-value $\geq \alpha T_p$, the examined sub-cube is inserted into the refined nucleus set \mathcal{N}^R .
6. All the possible pairs of sub-cubes in the refined nucleus set \mathcal{N}^R are linked through a line segment.

7. A morphological closing is performed to smooth the corners of the resulting 3D polygonal and fill its holes, thus finally obtaining the partial nucleus set \mathcal{N}_j , displayed in Fig. S1i.
8. Steps 1-7 are repeated K times on the same cell, thus obtaining K partial nucleus sets \mathcal{N}_j , with $j = 1, 2, \dots, K$, that are slightly different from each other, because in some of the performed WMW tests, the reference set is randomly drawn from a greater one.
9. The sum of all the K partial nucleus sets \mathcal{N}_j provides a tomogram of occurrences, in which each voxel can take integer values $k \in [0, K]$ since each voxel may have been classified nucleus k times. In Fig. S2a, the central slice of this tomogram of occurrences is reported.
10. An adaptive threshold k^* is set to segment the tomogram of occurrences. Let V_k be the number of voxels that have been classified nucleus at least k times, with $k = 1, 2, \dots, K$. Therefore, V_1 is the number of voxels of logical *or* among all the K partial nucleus sets \mathcal{N}_j , while V_K is the number of voxels of logical *and* among all the K partial nucleus sets \mathcal{N}_j .

- a. A vector \bar{V}^P of percentage volumes is created, which elements are computed as

$$V_k^P = \frac{V_k}{V_1}, \quad (\text{S2})$$

with $k = 1, 2, \dots, K$, as reported in Fig. S2b by blue dots.

The 3D segmented nuclear OCH should be computed as the set of voxels that have occurred at least k^{opt} times. The parameter k^{opt} should maximize simultaneously the accuracy, sensitivity, and specificity of the proposed CSSI method. In Fig. S2c, these performances are reported for each 3D segmented region composed by voxels that have occurred at least k times, with $k = 1, 2, \dots, K$, along with the k^{opt} value, highlighted by the vertical green line. However, in a real experiment these trends are unknown, thus the intersection point cannot be computed. Therefore, a criterion is requested to find k^* threshold, i.e., a suitable estimate of the k^{opt} threshold.

- b. The k^* threshold (red vertical line in Fig. S2b) is found as the k index at which the percentage volume vector \bar{V}^P is nearest to a threshold T_V (orange horizontal line in Fig. S2b).

In Fig. S2c, where the k^* threshold is highlighted by the red vertical line, it is clear that, despite $k^* \neq k^{opt}$, k^* is located in the same quasi-constant region of k^{opt} , hence this estimated threshold leads to very little differences in terms of clustering performances with respect to the optimal one.

11. The final 3D nuclear OCH \mathcal{N} is made of voxels that have been classified nucleus at least k^* times, as shown in Fig. 3d.

In Fig. 3d, it is evident that the proposed CSSI algorithm allows segmenting a 3D nuclear OCH very close to the original one, as underlined by accuracy, sensitivity, and specificity reported below the tomograms. Moreover, these values are very close to the optimal ones that could be obtained by using the optimal threshold

k^{opt} instead of the estimated one k^* , i.e. $ACC^{opt} = 98.22 \%$, $SENS^{opt} = 98.57 \%$, and $SPEC^{opt} = 98.10 \%$.

All the parameters involved in the proposed CSSI algorithm are described in Table S1. It is worth underlining that, in our experiments, a resolution factor $\varepsilon=10$ px has been set to analyse arrays made of at least $190 \times 190 \times 190$ voxels, since it was an optimum compromise between the need of having both high resolution in nucleus segmentation and high statistical power in WMW test (see the analysis about the imaging spatial resolution in Section S3).

Our methodology requires as first step the processing of tens of images for computing the 3D tomographic reconstruction and as second step the implementation of the CSSI algorithm for recovering the 3D nuclear OCH for each cell. The overall computational processing takes tens of minutes on a conventional desktop computer. Instead, the numerical processing to recover nuclear organelles from the recorded images takes few minutes in the 2D case of Amnis ImageStreamX due to the use of a commercial software. However, our computational processing can be easily speeded up using Graphic Processing Units, parallel processing strategies, and faster programming languages (in our implementation we used MATLAB[®] R2020a).

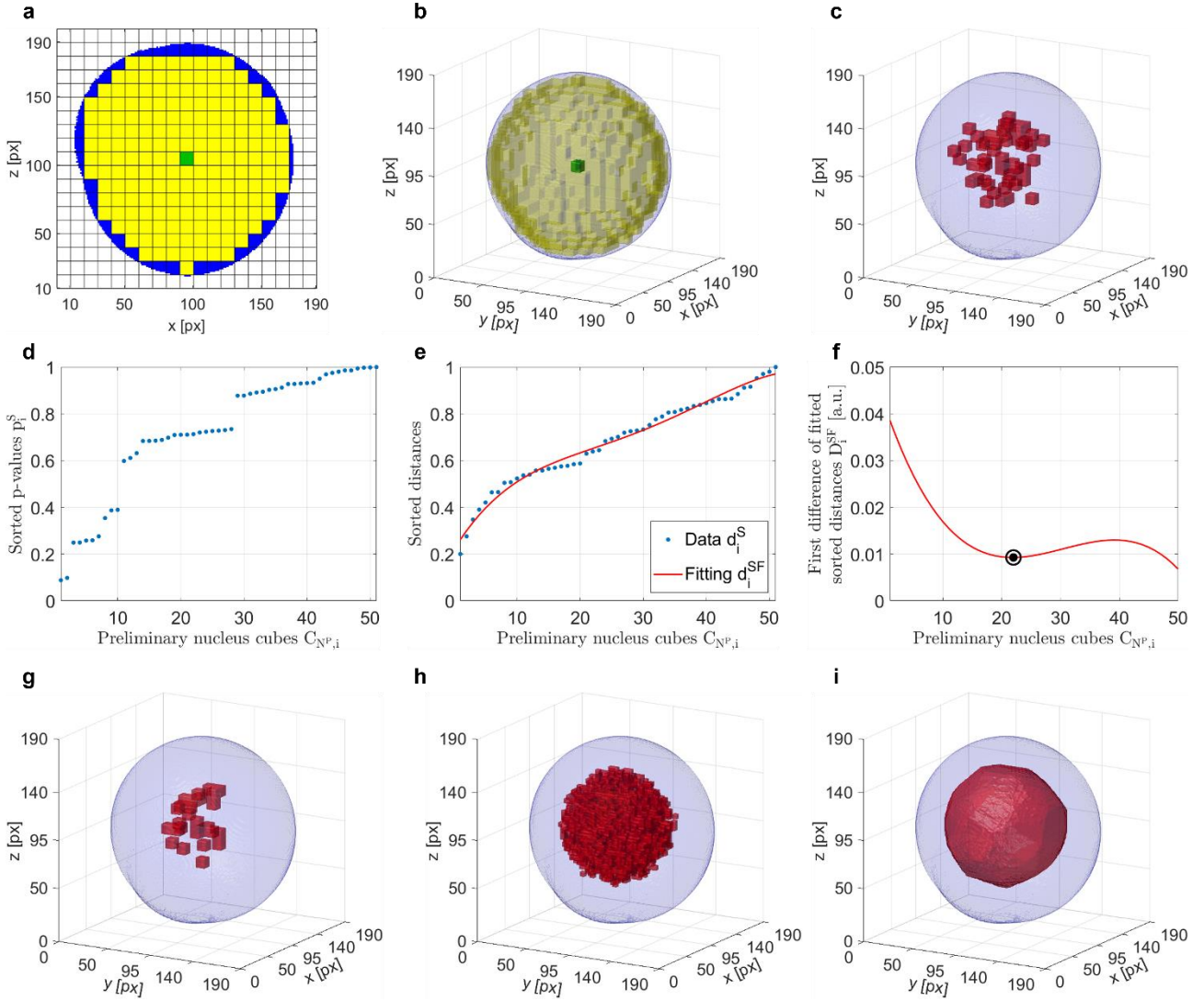


Fig. S1. CSSI of the stain-free nuclear OCH in a 3D numerical cell phantom (Supplementary Video 1). **a** Central xz -slice of the 3D array divided into distinct cubes with edge $\varepsilon = 10$ px. The central reference cube C_R is highlighted in green while the investigated cubes C_I are highlighted in yellow within the cell shell in blue. **b** 3D array from which the central xz -slice in (a) has been selected. **c** Preliminary nucleus set \mathcal{N}^P made of ε -cubes classified nucleus after a first rough clustering. **d** Vector of sorted p-values \bar{p}^S computed through the WMW test between each cube $C_{\mathcal{N}^P,i}$ in the preliminary nucleus set \mathcal{N}^P and the reduced nucleus set \mathcal{N}_i^{P-} , with $i = 1, 2, \dots, n$. **e** Vector of sorted and normalized Euclidean distances \bar{d}^S (blue dots) between each cube $C_{\mathcal{N}^P,i}$ in the preliminary nucleus set \mathcal{N}^P and the centroid of all cubes in \mathcal{N}^P , along with the fourth-degree polynomial fitting \bar{d}^{SF} (red line), with $i = 1, 2, \dots, n$. **f** First difference \bar{D}^{SF} (red line) of vector of fitted sorted distances \bar{d}^{SF} in (b), with highlighted in black the lowest value with null slope. **g** Filtered nucleus set \mathcal{N}^F made of ε -cubes classified nucleus after a spatial and statistical filtering of the preliminary nucleus set \mathcal{N}^P in (c). **h** Refined nucleus set \mathcal{N}^R made of $\varepsilon/2$ -cubes classified nucleus after increasing resolution in the filtered nucleus set \mathcal{N}^F in (d) through sub-cubes of size $\varepsilon/2$. **i** Partial nucleus set \mathcal{N}_j obtained by linking sub-cubes in (e) through segment lines and by using morphological closing. In (b,c,g-i), the blue region is the cell shell and the red region is the segmented nucleus at different steps of the CSSI algorithm.

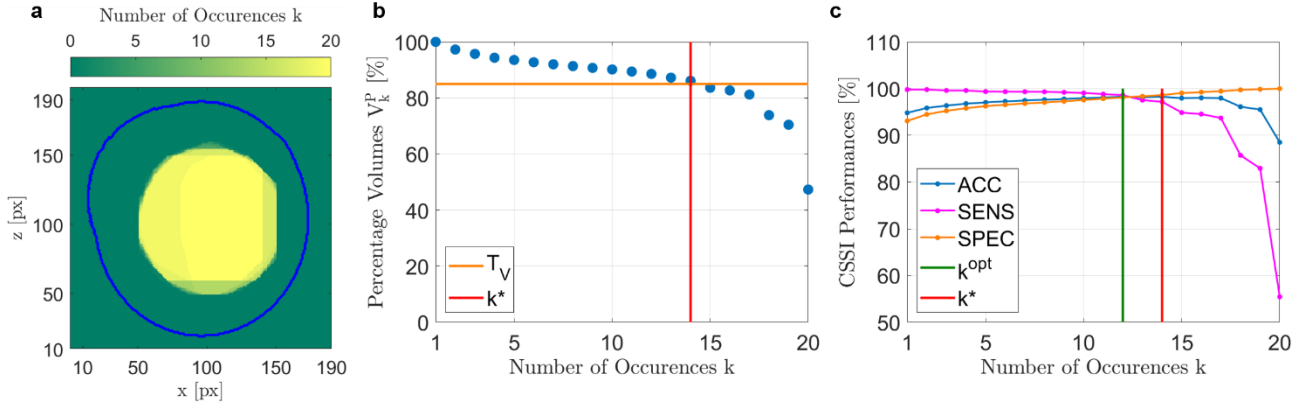


Fig. S2. Setting of the estimated threshold k^* . **a** Central xz -slice of the tomogram of occurrences, in which each voxel can take an integer value k from 0 to $K = 20$, i.e. the number of times it has been classified nucleus after repeating K times steps 1-7 of the CSSI algorithm on the same cell. The blue line is the cell contour. **b** Percentage volumes V_k^P (blue dots), i.e. number of voxels V_k classified nucleus at least k times normalized to V_{20} , along with the threshold T_V (horizontal orange line) used to find the estimated threshold k^* (vertical red line) by an intersection analysis. **c** CSSI performances associated to each possible threshold k for creating the final 3D nucleus set N , expressed in terms of accuracy (blue), sensitivity (magenta), and specificity (orange), along with the optimum threshold k^{opt} (vertical green line in which performances are simultaneously maximized) and its estimation k^* (vertical red line) computed in (b).

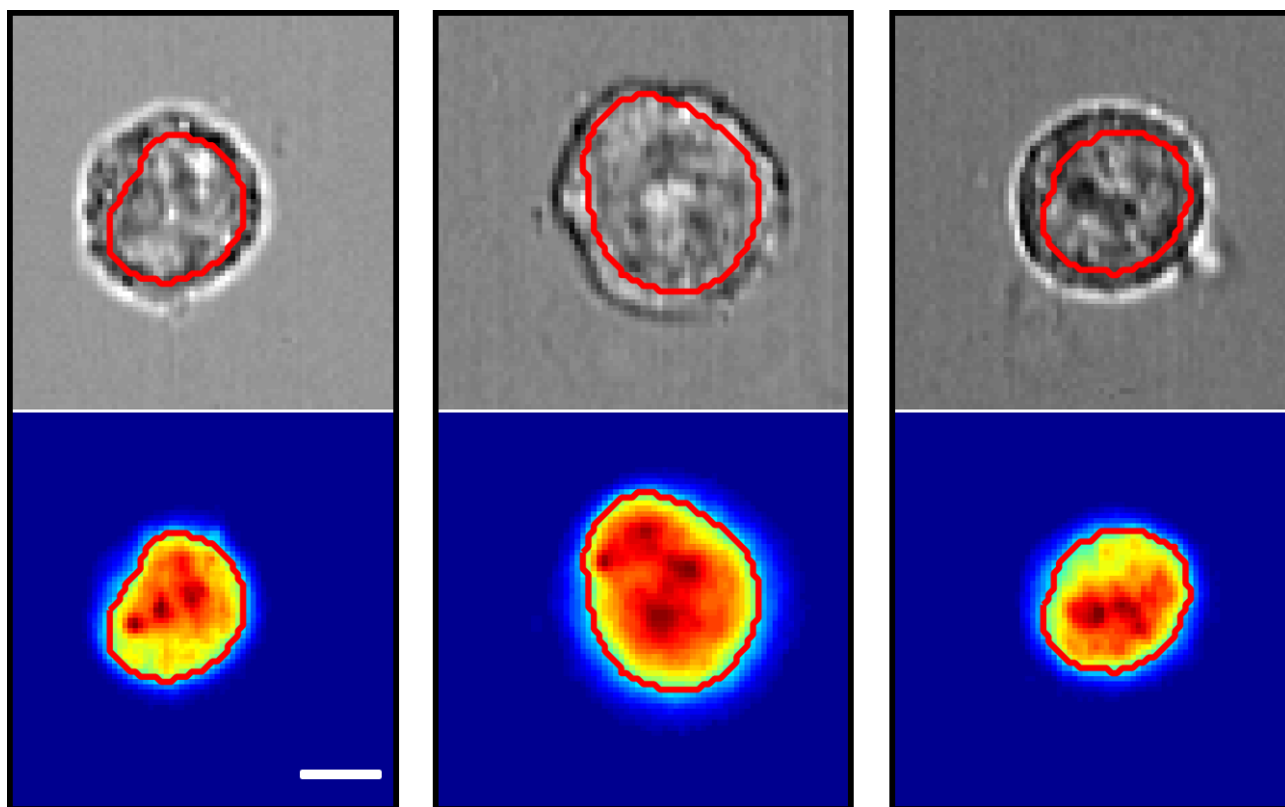


Fig. S3. 2D cyto-fluorimetric images of SK-N-SH cells recorded by Amnis ImageStreamX®. Three cells recorded simultaneously in brightfield images (top) and fluorescent images with the stained nucleus (bottom). The contour of the nucleus segmented by using the fluorescence information is overlapped in red. Scale bar is 5 μm .

S2. CSSI performances of the stain-free nucleus identification in presence of nucleoli

The nucleolus is the biggest structure inside the nucleus and it is a region particularly dense of genetic material, therefore its malfunctioning has been related to different human diseases^{S1,S2}. For example, the nuclear-nucleolar volume ratio is expected to change in a cancer cell^{S3}, as well as the number of nucleoli^{S4}. Moreover, due to its denser genetic composition, on average the nucleolus has a higher RI than the surrounding nucleus^{S2}. Therefore, we integrated the nucleolus to the same 1000 numerical 3D cell phantoms described in the Methods section in order to check whether its presence could negatively alter the correct identification of the stain-free nucleus by CSSI. Since in diploid human cells there are up to ten possible nucleoli per cell^{S4}, for each phantom we randomly drawn the number of nucleoli from the uniform distribution $U_3\{a_3, b_3\}$. The overall volume of all the nucleoli in a cell is computed from the nucleus-nucleolus volume ratio randomly drawn from the uniform distribution $U_4\{a_4, b_4\}$ ^{S3}, and then it is divided equally among all the nucleoli, each of them being simulated as a sphere. The nucleolus RIs are randomly drawn from the gaussian distribution $N_5(\mu_5, \sigma^2)$, where the mean value μ_5 , in turn randomly drawn from the uniform distribution $U_5\{a_5, b_5\}$, depends on the nucleus RIs in order to emulate the biological condition in which the nucleolus average RI is higher than the nucleus one. All the described parameters are reported in Table S2. Within the 1000 3D numerical cell phantoms, the multiple nucleoli are randomly localized inside the nucleus. In Figs. S5a,d,g,j, four numerical cell phantoms are displayed with the same 15 mitochondria and the same nucleus, and with 0, 3, 6, and 9 nucleoli, respectively. As reported in the corresponding RI histograms in Figs. S5b,e,h,k, respectively, the RI values assigned to the nucleoli are greater than the corresponding outer nuclei but still included in their RI distributions in order to consider the worst case condition for segmentation. In Fig. S5c, the nuclear OCH segmented by the CSSI method without considering the presence of nucleoli is mostly overlapped to the nuclear OCHs displayed in Figs. S5f,i,l obtained by the CSSI method in presence of 3, 6, and 9 nucleoli, respectively, as also suggested by the values of accuracy, sensitivity, and specificity reported in Figs. S5c,f,i,l. After applying the CSSI algorithm to the 1000 numerical 3D cell phantoms, we calculated the values of 9 metrics to quantify the CSSI performances in segmenting the stain-free nucleus in the presence of the nucleoli (see the second column of Table S3). The values of the 9 metrics can be directly compared with the corresponding ones in the first column of Table S3 regarding the segmentation of the stain-free nucleus without the simulation of the nucleoli. Moreover, the comparison in Fig. S4 between the histograms of these 9 metrics about these two case studies clearly show that, despite the presence of the nucleolus, even if performances of the CSSI in identifying the stain-free nucleus have slightly decreased, they still keep very high, over the 90 %. Therefore, the presence of the nucleoli inside the nucleus does not substantially worsen the performances of the CSSI nucleus segmentation algorithm.

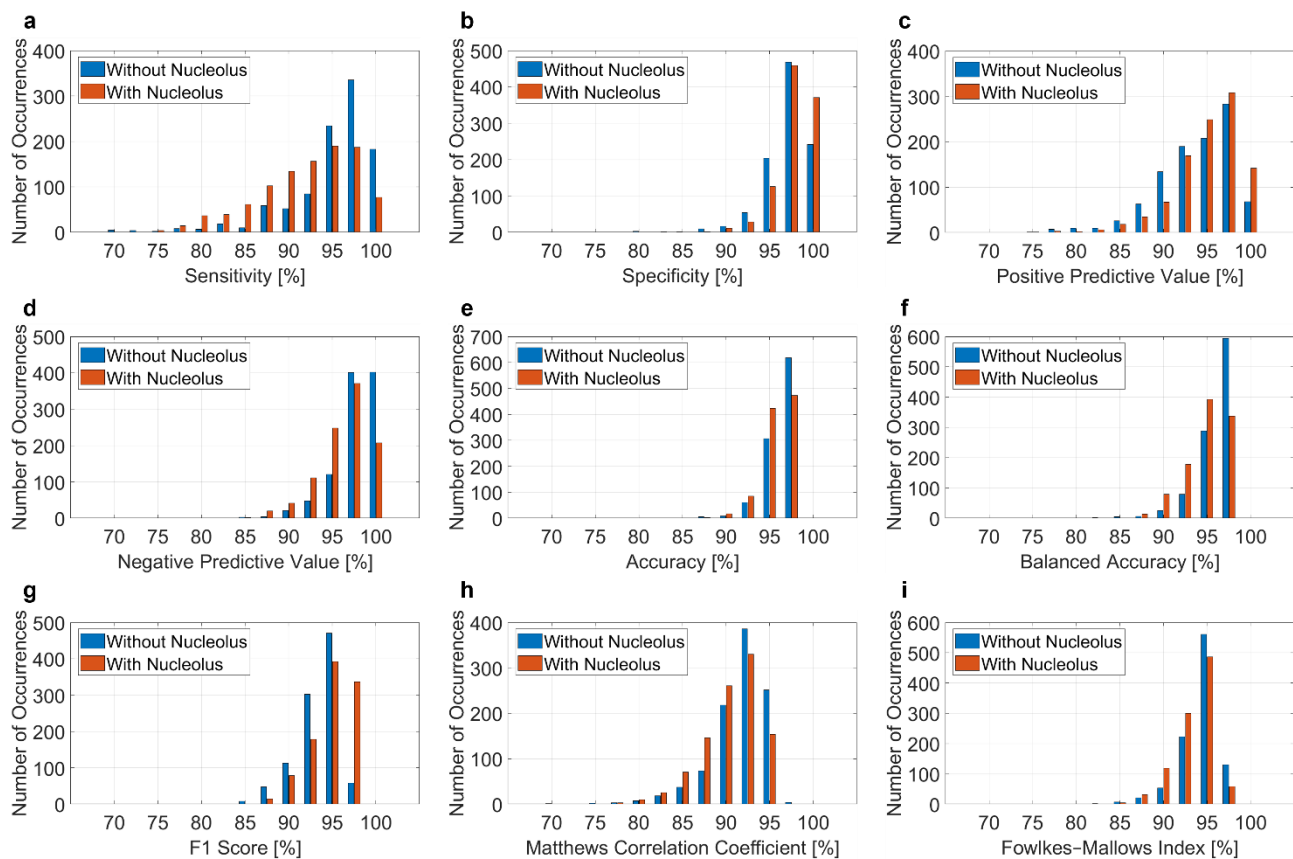


Fig. S4. Histograms of the several metrics in Table S3 used to quantify the performances in computing the nuclear OCH over a dataset of 1000 3D numerical cell phantoms without (blue) and with (orange) the presence of the nucleoli. a Sensitivity (Recall or True Positive Rate). **b** Specificity (True Negative Rate). **c** Positive Predictive Value (Precision). **d** Negative Predictive Value. **e** Accuracy. **f** Balanced Accuracy. **g** F1 Score. **h** Matthews Correlation Coefficient (Phi Coefficient). **i** Fowlkes–Mallows Index.

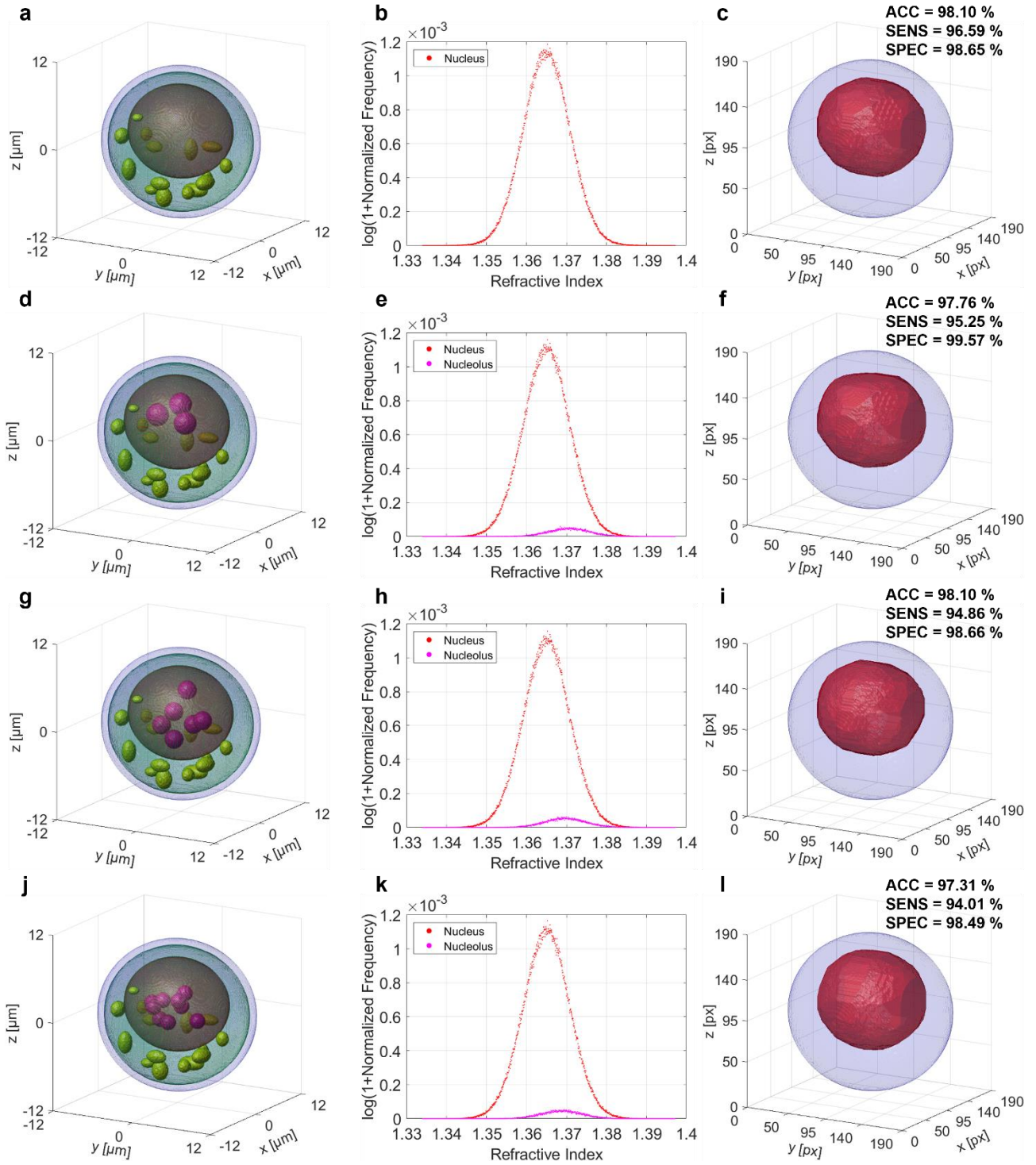


Fig. S5. Numerical assessment of the CSSI algorithm applied to segment the 3D nuclear OCH after the addition of the nucleoli. a, d, g, j Isosurface representation of the 3D cell model, simulated with the same cell membrane, cytoplasm, nucleus, and 15 mitochondria, and with 0, 3, 6, and 9 nucleoli, respectively. **b, e, h, k** Histogram of the RI values assigned to the nucleus and the nucleoli in (a, d, g, j), respectively. **c, f, i, l** Nuclear OCH segmented by the CSSI algorithm within the cell phantoms in (a, d, g, j), respectively. The corresponding CSSI accuracy, sensitivity, and specificity are reported at top right.

S3. Extension of the CSSI algorithm to other organelles

The CSSI algorithm is based on the computation of statistical similarities among groups of voxels inside the same cell by starting from an initial guess on the organelle location. To implement it, the sole property that must be satisfied consists in having different statistical distributions of the reconstructed quantity among the several intracellular organelles, as occurs for example in the RI case³². For this reason, the same strategy, based on statistics approach, can be also exploited to segment other cell organelles. In particular, the steps of the CSSI algorithm described in Section S1 have been developed on the basis of this statistics working principle at the aim of identifying a single compact organelle having a distinctive RI distribution inside the cell, and it has been demonstrated for segmenting the nucleus. Here we show that the same algorithm can be implemented to segment another single compact organelle, like the nucleolus in case a cell has a single nucleolus (e.g., in slowly cycling cells^{S5}). As regards the nucleolus location, two different situations have to be analyzed, namely case 1 and case 2. To describe them, in Fig. S6 we consider a numerical cell phantom with 15 mitochondria and a spherical nucleolus with a 27 times smaller volume than the surrounding nucleus. In the case 1, the nucleolus is not in the center of the cell, as shown by the violet sphere in Fig. S6a. In the case 2, the nucleolus is located in the center of the cell, as displayed by the violet sphere in Fig. S6g. As reported in the RI histograms in Figs. S6b,h, the RI values assigned to the nucleoli simulated in the cases 1 and 2, respectively, are greater than the corresponding outer nuclei but still included in their RI distributions in order to consider the worst case condition for segmentation. As regards the case 1, the CSSI algorithm starts looking for nuclear voxels from the central cube (i.e., the yellow reference cube C_R in Fig. S6c overlapped to the red simulated nucleus). As the CSSI algorithm is able to segment the nuclear OCH, i.e. the best convex hull that on average overlaps to the cell nucleus, the hole left by the presence of the nucleolus is automatically closed. Consequently, the correct nucleus segmentation shown in Fig. S6d, directly comparable with the simulated nucleus in Fig. S6c, can be obtained. Then, as the nucleolus has the highest RIs inside the nucleus, the CSSI algorithm can be easily adapted to search for its voxels by setting as starting reference cube C_R the cube having the highest average RI within the segmented nucleus (i.e., the yellow cube in Fig. S6e overlapped to the simulated nucleolus). As displayed in violet in Fig. S6f, the CSSI output is the nucleolar OCH, i.e. the best convex hull that on average overlaps to the cell nucleolus simulated in Fig. S6f. Instead, the case 2 is exactly the opposite of case 1. In fact, if the nucleolus is in the center of the cell (violet simulated region in Fig. S6i) and the CSSI algorithm normally starts from the central reference cube C_R (yellow cube in Fig. S6i), the first output is the nucleolar OCH, as reported in violet in Fig. S6j. In such a case, it can be easily inferred that the segmented region is the nucleolus rather than the nucleus due to its much smaller size. Then, the

second step consists in searching for the nucleus by starting from a reference cube C_R that is adjacent to the outer side of the segmented nucleolus (e.g., the yellow cube in Fig. S6k). Again, the nuclear OCH is correctly segmented, as shown in red in Fig. S6l with respect to the red simulated nucleus in Fig. S6k. Therefore, starting from the same 1000 3D numerical cell phantoms described in the Methods section, we simulated 500 phantoms for the case 1 and 500 phantoms for the case 2 by using the same distributions presented in Section S2 by fixing one single nucleolus per cell. In the third column of Table S3, we report the values of 9 metrics to quantify the CSSI performances in segmenting the stain-free nucleolus (the corresponding histograms are displayed in Fig. S7), which are just slightly worse than the stain-free nucleus ones because of the much smaller number of voxels representative of this organelle. In fact, the principal factor that could limit the success of the CSSI algorithm in case of a single compact organelle is a low imaging spatial resolution with respect to the size of the analyzed organelle, which means that the organelle is represented by a very low number of voxels. In such a case, to take into account a lower number of voxels representing a certain organelle, the resolution factor ε can be reduced. However, it cannot be excessively decreased otherwise the opposite effect is obtained, i.e. a bad segmentation of the organelle due to a too low statistical power of the WMW test. The statistical power is indeed the probability of correctly classifying a voxel as non-organelle when it actually does not belong to that organelle, and it is well known that the statistical power of a hypothesis test increases with the number of observations. To demonstrate this property, we simulated a 3D numerical cell phantom with 15 mitochondria and we used it to create 8 possible resolution scenarios, i.e. we simulated inside it the same spherical organelle with a radius changing from 15 to 50 pixels (step of 5 pixels). In Fig. S8a, we show three of these phantoms corresponding to the 50, 30, and 15 pixels radii. For each of these 8 resolution scenarios, we implemented the CSSI algorithm by using three resolution factors, i.e. $\varepsilon=10$ px, $\varepsilon=8$ px, and $\varepsilon=6$ px, which segmentation results are reported in Figs. S8b-d, respectively, for the 50, 30, and 15 pixels radii. In order to quantitatively compare the CSSI segmentation at different organelle sizes (i.e., imaging spatial resolutions) and resolution factors ε , we computed the F1 scores through the formula in Table S3. In Fig. S8e, it is evident that by fixing a too low resolution factor (i.e., $\varepsilon=6$ px), the segmentation performances significantly drop because of the low statistical power of the WMW test. Instead, the segmentation performances are much higher in the $\varepsilon=10$ px and $\varepsilon=8$ px cases, which means that starting from the resolution factor $\varepsilon=8$ px, the statistical power (i.e., the number of observations) of the WMW test is enough to guarantee a good segmentation output. Moreover, as can be expected, the zoom-in of the F1 score plot in Fig. S8f highlights a growing trend of the classification performances with the organelle pixel size (i.e., with the imaging spatial resolution). In particular, the $\varepsilon=8$ px curve shows the best behavior since it has the lowest slope, which means that,

by reducing the imaging spatial resolution, the segmentation performance decreases more slowly than the $\epsilon=10$ px. In fact, after reaching a sufficient statistical power by selecting $\epsilon>6$ px, the $\epsilon=8$ px resolution factor allows following better the external organelle shape due to a better spatial resolution despite the number of observations is lower than the $\epsilon=10$ px case. It is worth underlying two other important points. First, by using a resolution factor ϵ at least equal to 8 px, an organelle with an equivalent radius less than 15 pixels cannot be examined since the organelle would be covered by very few distinct cubes. Furthermore, the CSSI computational time increases in a nonlinear way with the resolution factor ϵ , therefore we selected $\epsilon=10$ px as a good trade-off to analyze our simulated and experimental tomograms. In summary, remarkably the CSSI algorithm can be used to segment other single compact organelles different from the nucleus but, in case they are very small, the imaging spatial resolution must be able to provide an equivalent radius >15 pixels to implement the CSSI algorithm with at least a resolution factor $\epsilon=8$ px. Otherwise, the statistics of the WMW test are not robust and this greatly worsens the segmentation performances. Instead, in case of multiple organelles, the same working principle based on the statistical similarities can be exploited. However, the algorithm implemented in Section S1 must be slightly modified to avoid the constraints herein used to gather all the candidate cubes at the aim of segmenting a single compact region, i.e. the nucleus.

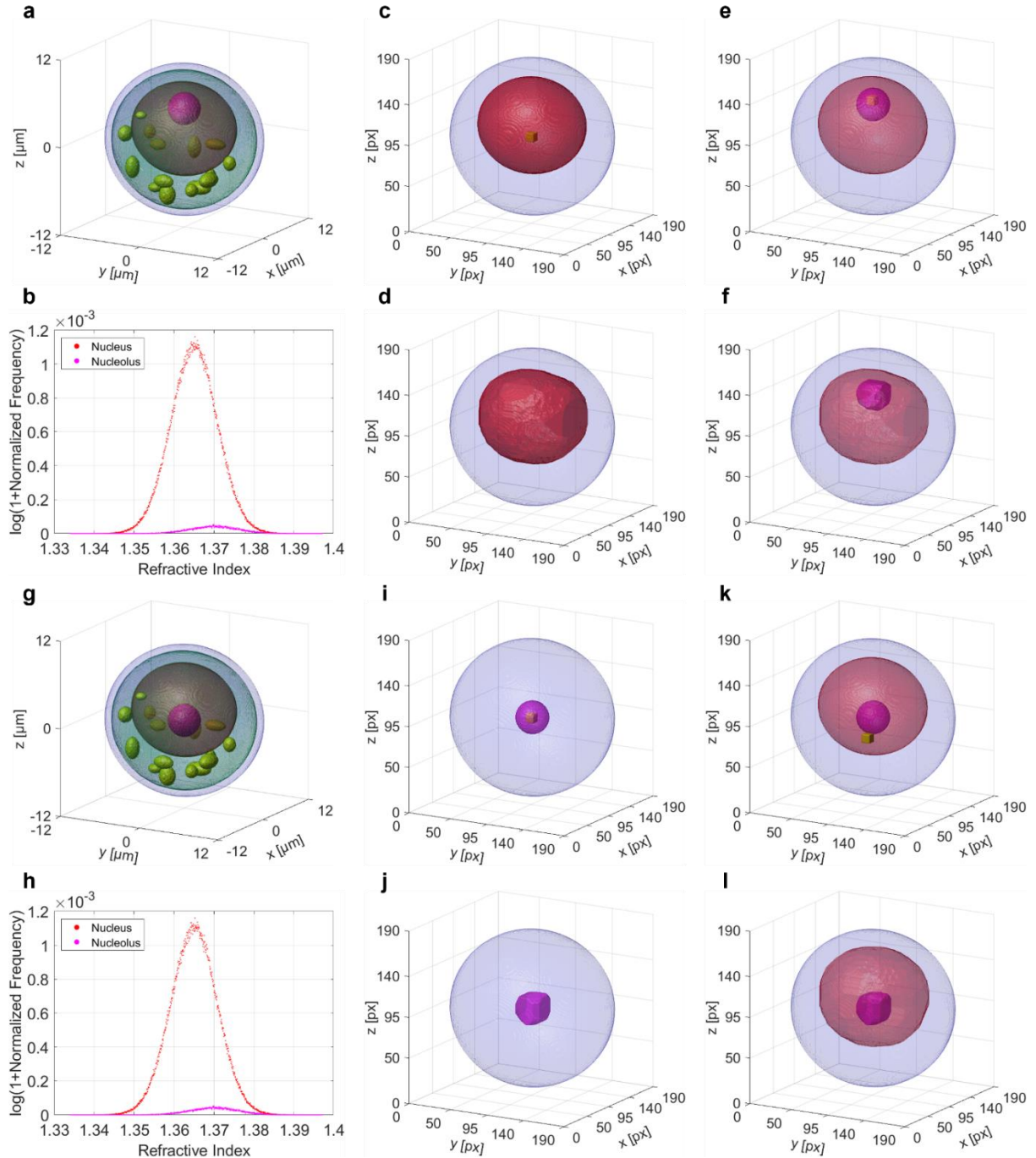


Fig. S6. Numerical assessment of the CSSI algorithm applied to segment a single 3D nucleolar OCH when far from the center (case 1 in a-f) or in the center (case 2 in g-l) of the 3D numerical cell phantom. a, g Isolevels representation of the 3D cell model, simulated with five sub-cellular components, i.e., cell membrane, cytoplasm, nucleus, nucleolus, and 15 mitochondria. **b, h** Histogram of the RI values assigned to the nucleus and the nucleolus in (a, g), respectively. **c, d** Respectively, simulated nucleus and corresponding nuclear OCH (red) segmented by the CSSI algorithm despite the presence of the not-centered nucleolus in (a). In (c), the starting central reference cube C_R of the CSSI algorithm is overlapped in yellow. **e, f** Respectively, simulated nucleolus and corresponding nucleolar OCH (violet) segmented by the CSSI algorithm after the nucleus identification (red) in (d). In (e), the starting reference cube C_R of the CSSI algorithm is overlapped in yellow in the zone of the segmented 3D nuclear OCH in (d) having the highest RIs. **i, j** Respectively, simulated nucleolus and corresponding nucleolar OCH (violet) segmented by the CSSI algorithm because, by starting from the central reference cube C_R overlapped in yellow in (i), the centered nucleolus in (g) is found. **k, l** Respectively, simulated nucleus and corresponding nuclear OCH (red) segmented by the CSSI algorithm after recognizing that the region segmented in (j) is the nucleolus due its small size. In (k), the starting reference cube C_R of the CSSI algorithm is overlapped in yellow in the adjacent zone to the outer side of the segmented 3D nucleolar OCH in (j).

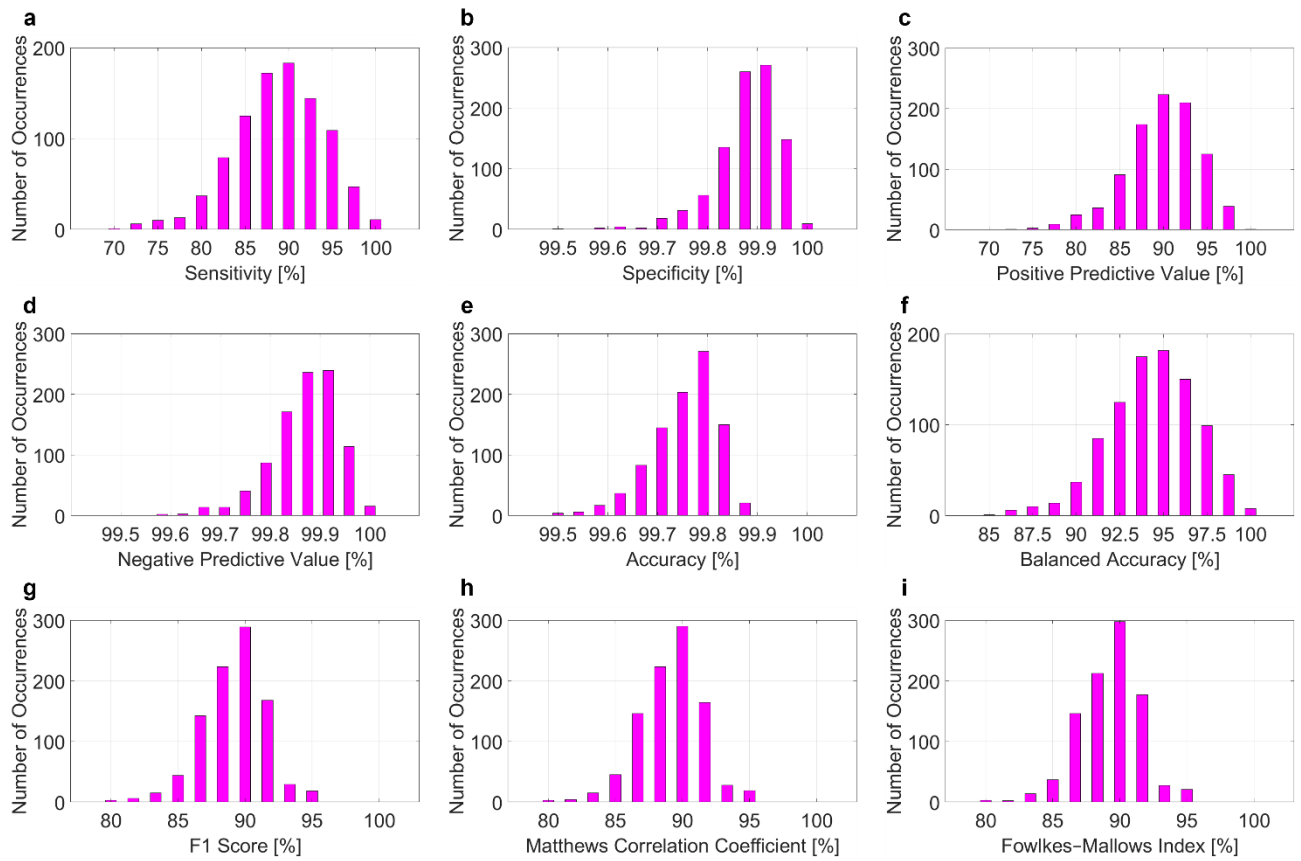


Fig. S7. Histograms of the several metrics in Table S3 used to quantify the performances in computing a single nucleolar OCH over a dataset of 1000 3D numerical cell phantoms. a Sensitivity (Recall or True Positive Rate). **b** Specificity (True Negative Rate). **c** Positive Predictive Value (Precision). **d** Negative Predictive Value. **e** Accuracy. **f** Balanced Accuracy. **g** F1 Score. **h** Matthews Correlation Coefficient (Phi Coefficient). **i** Fowlkes–Mallows Index.

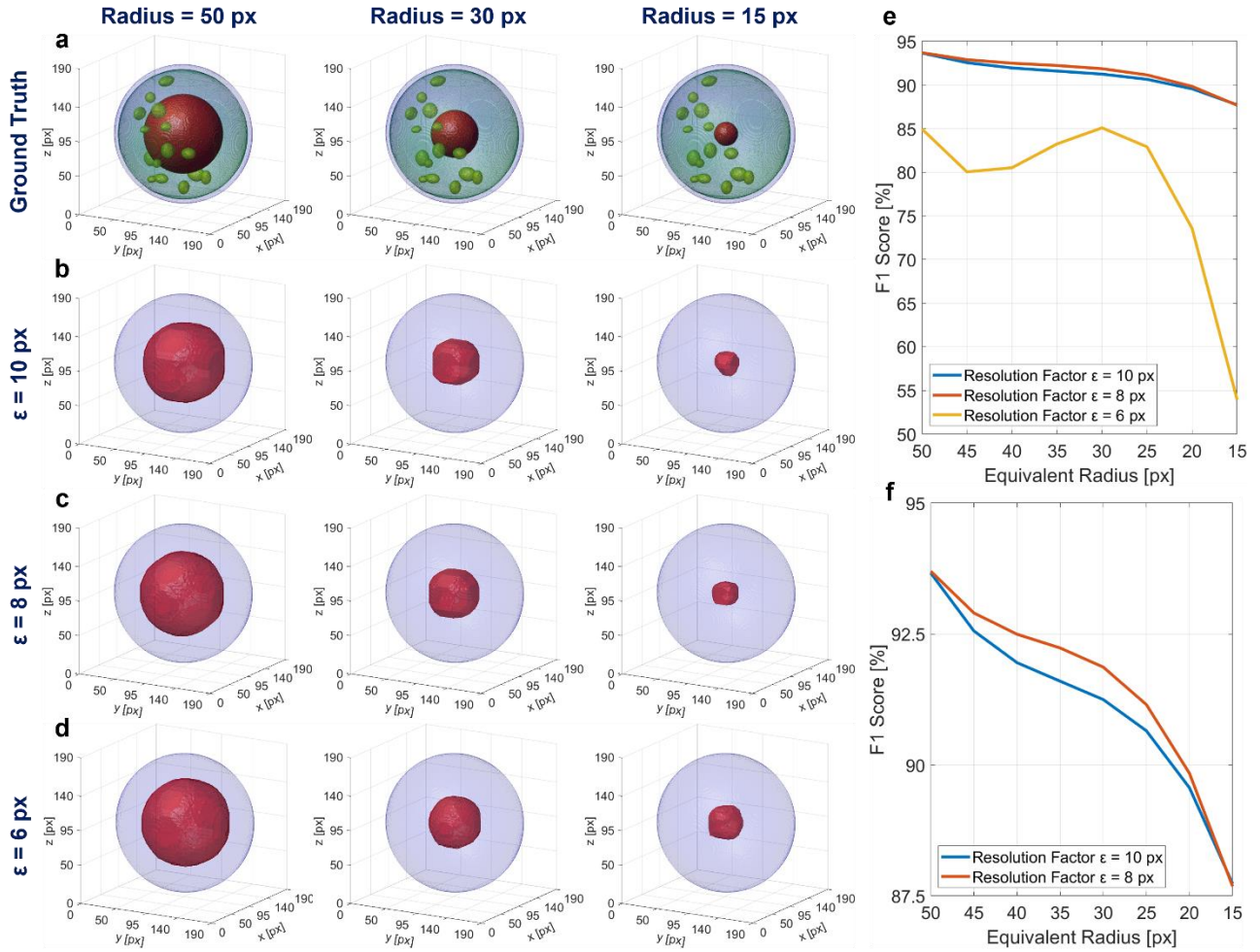


Fig. S8. CSSI performances in segmenting a generic organelle with respect to the imaging spatial resolution (i.e., its number of voxels) and the CSSI resolution factor ϵ . **a** Three 3D numerical cell phantoms with 15 mitochondria (green) and a spherical organelle (red) having an equivalent radius of 50 px (left), 30 px (center), and 15 px (right). **b-d** CSSI segmentation of the red organelles in (a) by using the resolution factors $\epsilon = 10$ px, $\epsilon = 8$ px, and $\epsilon = 6$ px, respectively. **e** F1 score in segmenting the red organelle in (a) by using the resolution factors $\epsilon = 10$ px, $\epsilon = 8$ px, and $\epsilon = 6$ px while changing its equivalent radius (i.e., the imaging spatial resolution) between 15 px and 50 px. **f** Zoom-in of the $\epsilon = 10$ px and $\epsilon = 8$ px curves in (e).

Table S1. Setting of the parameters involved in the CSSI algorithm to segment the stain-free 3D nuclear OCH from in-flow TPM reconstructions ($\lfloor \cdot \rfloor$, $\lceil \cdot \rceil$, and $\lceil \cdot \rceil$ are the floor, ceil, and nearest integer operators, respectively).				
$\varepsilon = 10 \text{ } px$	$V_{1x} = \frac{L_x - \varepsilon + 2}{2}$	$V_{2x} = \frac{L_x + \varepsilon}{2}$	$V_{1y} = \frac{L_y - \varepsilon + 2}{2}$	
$V_{2y} = \frac{L_y + \varepsilon}{2}$	$V_{1z} = \frac{L_z - \varepsilon + 2}{2}$	$V_{2z} = \frac{L_z + \varepsilon}{2}$	$\tau = 0.99$	
$M = 10$	$n = \left\lfloor \frac{2L_x + L_y + L_z}{\varepsilon} \right\rfloor$	$T_5 = \frac{1}{2} [\mu(\bar{p}) + p_{max}]$	$T_4 = \mu(\bar{p})$	
$T_3 = \begin{cases} 0.7 & \text{if } \frac{m+1}{n} < 0.15 \\ 0.95 & \text{if } \frac{m+1}{n} > 0.85 \\ 0.7 + \frac{5}{14} \left(\frac{m+1}{n} - 0.15 \right) & \text{otherwise} \end{cases}$		$T_V = \begin{cases} \frac{\lfloor 5\mu(\bar{V}^P) \rfloor}{5} & \text{if } \mu(\bar{V}^P) > 50 \% \\ \frac{\lfloor 5\mu(\bar{V}^P) \rfloor}{5} & \text{if } \mu(\bar{V}^P) < 50 \% \\ \mu(\bar{V}^P) & \text{if } \mu(\bar{V}^P) = 50 \% \end{cases}$		
$T_2 = T_3 - 0.1$	$T_1 = T_2 - 0.1$	$\alpha = 0.9$	$\beta = 0.5$	$K = 20$

Table S2. Parameters used for simulating the 3D numerical cell phantoms.			
$\mu_{\mu_1} = 1.352$	$\mu_{\mu_3} = 1.368$	$\sigma = 0.005$	$a_1 = 10$
$\mu_{\mu_2} = 1.365$	$\mu_{\mu_4} = 1.370$	$\sigma_{\mu} = 0.003$	$b_1 = 20$
$a_2 = 1$	$a_3 = 1$	$a_4 = 25$	$a_5 = \mu_N + \frac{1}{3}[q_{0.95}(N) - \mu_N]$
$b_2 = 3$	$b_3 = 10$	$b_4 = 30$	$b_5 = \mu_N + \frac{2}{3}[q_{0.95}(N) - \mu_N]$

Table S3. Performances in computing the nuclear OCH (without and with the presence of nucleoli) and a single nucleolar OCH over the same dataset of 1000 3D numerical cell phantoms. TP (True Positive) is the number of voxels that are correctly classified as organelle, TN (True Negative) is the number of voxels that are correctly classified as non-organelle, FP (False Positive) is the number of voxels that are wrongly classified as organelle, and FN (False Negative) is the number of voxels that are wrongly classified as non-organelle.

<i>Metric</i>	<i>Mathematical Definition</i>	<i>Mean \pm Standard Deviation [%]</i>		
		<i>Nuclear OCH</i>		<i>Single Nucleolar OCH</i>
		<i>Without Nucleoli</i>	<i>With Nucleoli</i>	
<i>True Positive Rate (Sensitivity or Recall)</i>	$TPR = SENS = \frac{TP}{TP + FN}$	94.97 ± 5.01	92.08 ± 5.47	88.84 ± 5.11
<i>True Negative Rate (Specificity)</i>	$TNR = SPEC = \frac{TN}{TN + FP}$	96.92 ± 2.59	97.69 ± 2.05	99.88 ± 0.06
<i>Positive Predictive Value (Precision)</i>	$PPV = \frac{TP}{TP + FP}$	93.75 ± 4.28	95.07 ± 3.74	89.86 ± 4.19
<i>Negative Predictive Value</i>	$NPV = \frac{TN}{TN + FN}$	97.61 ± 2.46	96.29 ± 2.82	99.87 ± 0.07
<i>Accuracy</i>	$ACC = \frac{TP + TN}{TP + TN + FP + FN}$	96.28 ± 1.71	95.86 ± 1.65	99.75 ± 0.07
<i>Balanced Accuracy</i>	$BA = \frac{TPR + TNR}{2}$	95.95 ± 2.17	94.89 ± 2.31	94.36 ± 2.54
<i>F1 Score</i>	$F1 = \frac{2TP}{2TP + FP + FN}$	94.17 ± 2.42	93.36 ± 2.33	89.16 ± 2.34
<i>Matthews Correlation Coefficient (Phi Coefficient)</i>	MCC $= \frac{TP \times TN - FP \times FN}{\sqrt{(TP + FP)(TP + FN)(TN + FP)(TN + FN)}}$	91.60 ± 3.36	90.53 ± 3.28	89.13 ± 2.30
<i>Fowlkes–Mallows index</i>	$FM = \sqrt{PPV \times TPR}$	94.26 ± 2.27	93.47 ± 2.22	89.25 ± 2.27

Table S4. 2D morphological parameters of SK-N-SH cells measured in labelled nuclei segmented from 11549 2D FM cyto-fluorimetric images and in unlabeled nuclei segmented by CSSI algorithm from 90 2D reprojections of five in-flow TPM reconstructions.

		<i>Mean Value</i>		<i>p-value</i> <i>Hotelling's T^2 test</i>
		<i>TPM</i>	<i>FM</i>	
<i>nucleus size</i>	<i>nucleus-cell area ratio</i>	0.404	0.406	0.962
<i>nucleus shape</i>	<i>nucleus aspect ratio</i>	0.839	0.838	
<i>nucleus position</i>	<i>normalized nucleus-cell centroid distance</i>	0.116	0.118	

Table S5. 3D morphological parameters of MCF-7 cells measured in labeled nuclei segmented from FM confocal images³⁷ and in unlabeled nuclei segmented by CSSI algorithm from three in-flow TPM reconstructions.

		<i>FM</i>	<i>TPM</i> <i>cell 1</i>	<i>TPM</i> <i>cell 2</i>	<i>TPM</i> <i>cell 3</i>
<i>nucleus size</i>	<i>nucleus-cell volume ratio</i>	0.3396 ± 0.0939	0.296	0.415	0.315
<i>nucleus shape</i>	<i>nucleus surface-volume ratio [μm^{-1}]</i>	0.713 ± 0.103	0.702	0.531	0.635
<i>nucleus position</i>	<i>normalized nucleus-cell centroid distance</i>	0.152 ± 0.108	0.053	0.141	0.172

References

- S1. Tsai, R. Y. & Pederson, T. Connecting the nucleolus to the cell cycle and human disease. *FASEB J.* **28**(8), 3290-3296 (2014).
- S2. Núñez Villacís, L. et al. New roles for the nucleolus in health and disease. *Bioessays* **40**(5), 1700233 (2018).
- S3. Guttman, P. H. & Halpern, S. Nuclear-nucleolar volume ratio in cancer. *Am. J. Cancer* **25**(4), 802-806 (1935).
- S4. Farley, K. I., Surovtseva, Y., Merkel, J. & Baserga, S. J. Determinants of mammalian nucleolar architecture. *Chromosoma* **124**(3), 323-331 (2015).
- S5. Schwarzacher, H. G. & Wachtler, F. The nucleolus. *Anat. Embryol.* **188**(6), 515-536 (1993).