nature methods

Article

https://doi.org/10.1038/s41592-025-02960-4

Cell context-dependent in silico organelle localization in label-free microscopy images

Received: 19 November 2024

Accepted: 28 October 2025

Published online: 19 December 2025

Check for updates

Nitsan Elmalam & Assaf Zaritsky

The in silico labeling prediction of organelle fluorescence from label-free microscopy images has the potential to revolutionize our understanding of cells as integrated complex systems. However, out-of-distribution data caused by changes in the intracellular organization across cell types, cellular processes or perturbations can lead to altered label-free images and impaired in silico labeling. Here we demonstrate that incorporating biological meaningful cell contexts, via a context-dependent model that we call CELTIC, enhanced in silico labeling prediction and enabled the downstream analysis of out-of-distribution data such as cells undergoing mitosis and cells located at the edge of the colony. These results suggest a link between cell context and intracellular organization. Using CELTIC to generate single-cell images transitioning between different contexts enabled us to overcome intercell variability toward the integrated characterization of organelles' alterations in cellular organization. The explicit inclusion of context has the potential to harmonize multiple datasets, paving the way for generalized in silico labeling foundation models.

Organelles act in concert to shape and enable cell function. Accordingly, the organization of organelles and the spatial relations between different organelles are remarkably versatile and can change in response to multitude factors including undergoing cellular processes such as proliferation[1], migration[2] or differentiation[3,4] and being influenced by extrinsic factors such as local cell density and different microenvironmental conditions (for example, mechanical stresses, diffusible factors and chemical treatments)[5,6]. For instance, during mitosis, the nuclear envelope disassembles, the nucleus undergoes condensation and separation, the Golgi apparatus is disassembled and then reformed and the microtubules rearrange to form the mitotic spindle[1]. The ability to measure whether and how the intracellular organization of organelles change is fundamental to cell biology but is technically challenging owing to substantial limitations in the simultaneous labeling of multiple organelles within the same cell[7].

The in silico labeling of organelles is the computational cross-modality translation of label-free transmitted light microscopy images to their corresponding organelle-specific fluorescent images[8]. In silico labeling holds the promise of enabling computationally multiplexed live cell imaging toward an integrated understanding of the cell[9]. Attaining an in silico labeling model involves the acquisition of matched label-free and fluorescently labeled images and using them to train a deep learning model that maps the label-free images to their corresponding matched fluorescence images[10]. This training process is repeated for each organelle, producing a set of organelle-specific in silico labeling models (Fig. 1a, left). At inference, the organelle-specific models can be applied to generate a multiplexed image displaying the localization of several organelles simultaneously[11] (Fig. 1a, right). Several recent studies demonstrated that in silico labeling can be applied to reveal how the intracellular organelle organization, cell shape and/or cell behavior alters in response to different cell states and different experimental conditions[11–26]. This forthcoming wave of in silico labeling applications to cell phenotyping raises a major question regarding generalization: are in silico labeling models confounded by cells that are not sufficiently represented during training? In principle, changes in the intracellular organization can alter the cell's optical properties, inducing out-of-distribution label-free images and impaired in silico labeling. For example, alterations in the cell's internal organization due to changes in cell–cell adhesions in densely packed microenvironments may lead to changes in the cell's optical properties that in

Institute for Interdisciplinary Computational Science, Faculty of Computer and Information Science, Ben-Gurion University of the Negev, Beer-Sheva, Israel.
e-mail: assafzar@gmail.com

turn can hamper high-quality in silico labeling. Indeed, a few studies demonstrated inferior performance upon perturbations[15] or upon inference on a cell type different from the one seen on training[27]. This deteriorated accuracy in organelle localization is posing a limitation in generalizing in silico labeling and hampering the possibility of using in silico labeling to understand how intracellular organelle organization is changing across cell types, throughout cellular processes and following perturbations.

Here, we focus on the problem of out-of-distribution label-free images due to rare cellular states and contexts that are underrepresented in the training data. We report a decreased performance of in silico labeling for rare cell populations, and introduce a new method called CEll in silico labeling using Tabular Input Context, or CELTIC, to overcome this limitation. CELTIC is designed to improve the in silico labeling of out-of-distribution cell populations by incorporating biological meaningful cell context (encoded as tabular data) into the in silico labeling model. We show that, by the inclusion of cell context, CELTIC enhances the in silico labeling of rare cell populations, especially organelle localization patterns associated with that context. We demonstrate that context-dependent generative traversal with CELTIC has the potential to reveal alterations in intracellular organization during context transitions. We also demonstrate that CELTIC can be generalized to other types of context by introducing a unified model for in silico labeling with the organelle type as context.

Results

Deteriorated in silico labeling for rare cell populations

We used three-dimensional (3D) spinning-disk microscopy images of genetically edited fluorescent human induced pluripotent stem cell lines (hiPS cell) colonies from the Allen Institute for Cell Science (AICS) WTC-11 hiPS cell Single-Cell Image Dataset v1[28]. The dataset comprises 3D field-of-view (FOV) images, each containing label-free images (brightfield) and a specific genetically edited EGFP-tagged protein representing an organelle. In addition, the dataset contains cell segmentation masks and metadata regarding the individual cells, which includes annotations regarding the mitotic stage (interphase/prophase/early prometaphase/prometaphase–metaphase/anaphase–telophase–cytokinesis) and its location within a colony (interior/edge). We decided to focus on six organelles that span the range of in silico labeling performances reported in ref. 10. Organelles with high performance in silico labeling included the nuclear envelope, actin filaments and microtubules (average pixel-wise Pearson correlation coefficient (PCC) of ~0.78–0.88); organelles with intermediate performance included the mitochondria and endoplasmic reticulum (~0.66–0.73); and organelles with low performance included the Golgi apparatus (0.23). We replicated the U-Net-based[29] in silico labeling model reported in ref. 10 and reproduced their results (Supplementary Table 1).

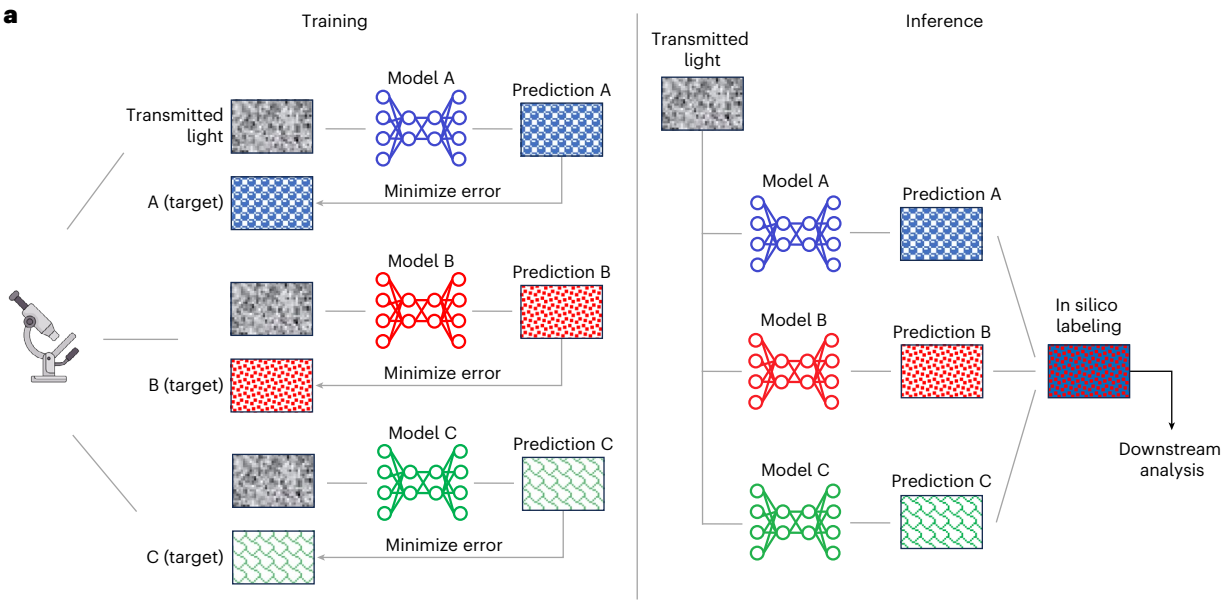
Our focus on the single-cell biologically meaningful context required us to use the cells' segmentation masks to isolate individual cells and assign for each cell whether it was undergoing mitosis (non-dividing–interphase, or one of five mitotic stages), its location in the colony (whether it is located at the colony's interior or edge), its volume and its local density (that is, the number of adjacent neighboring cells in the colony). Overall, we collected a single-cell dataset consisting of 1,116–1,575 single cells per organelle, extracted from 100 FOV images that were never seen by the trained model for each organelle. Cells in interphase accounted for more than 95% of the cells in our dataset, cells located away from the colony's edge (classified as interior as opposed to edge) accounted for over 98%, cells with typical volumes (that is, a z score higher than −1.5 relative to the overall population distribution classified as medium–large versus small) accounted for about 93% and cells in a microenvironment of typical density (that is, five or more adjacent neighbors for a local density z score >−1.5, classified as dense versus sparse) accounted for about 96% (Methods; Fig. 1b and Supplementary Table 2). For each cell, we measured the PCC between the fluorescent ground truth and its corresponding in silico prediction. Overlaying the FOV PCC values onto the single-cell segmentation masks showed poor endoplasmic reticulum in silico labeling for two cells in the prometaphase–metaphase stage of mitosis and for another cell located in a sparse neighborhood (Fig. 1c and Supplementary Fig. 1). Similarly, cells in the prometaphase–metaphase and early prometaphase stages of mitosis showed poor in silico labeling for the nuclear envelope (Fig. 1d and Supplementary Fig. 1). To systematically evaluate this observation, we pooled all single-cell PCC values across organelles and rare populations (Supplementary Table 2). Cells undergoing mitosis displayed declined in silico labeling across all organelles compared with cells in interphase (Fig. 1e). The most dramatic performance deterioration occurred for microtubules and for the nuclear envelope. Cells located at the colony edge exhibited inferior in silico labeling of the microtubules and actin (Supplementary Fig. 2a). Cells with small volumes and cells in sparse neighborhoods demonstrated reduced in silico labeling for most organelles and most prominently for the nuclear envelope and actin (Supplementary Fig. 2b,c). The decrease in in silico prediction quality is probably a consequence of applying the model to underrepresented data, probably due to out-of-distribution intracellular organization and the corresponding changes in these cells' optical properties. We hypothesized that incorporating context information about each cell would increase the performance of in silico label models and address the generalizability issue of these models to underrepresented data.

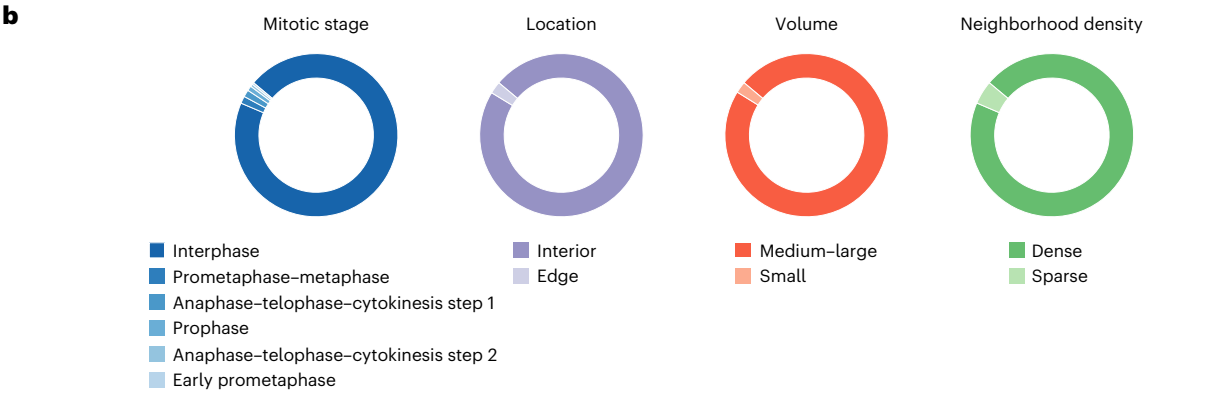
CELTIC, cell context-dependent in silico labeling

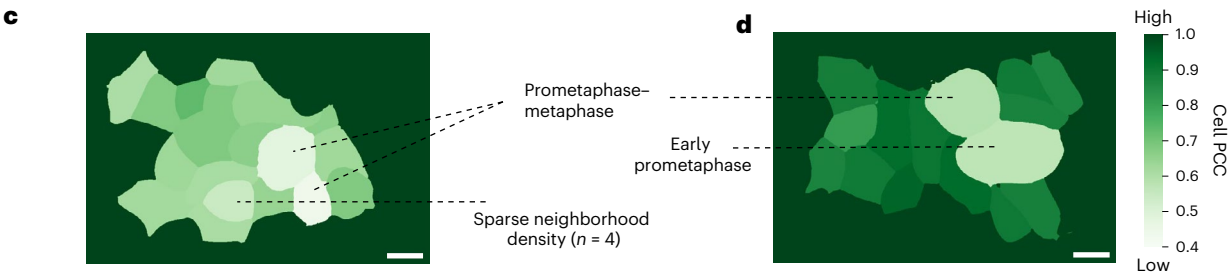
We propose CELTIC, an in silico labeling model that integrates cellular contextual information. CELTIC is an image-to-image translation model, where the innermost layer of the network ('bottleneck layer') explicitly encodes predefined cell context parameters. The single-cell centric approach required us to move from a FOV-based to a single-cell-based in silico labeling (Fig. 2a). Specifically, we cropped single cells according to their corresponding fluorescent plasma membrane-derived

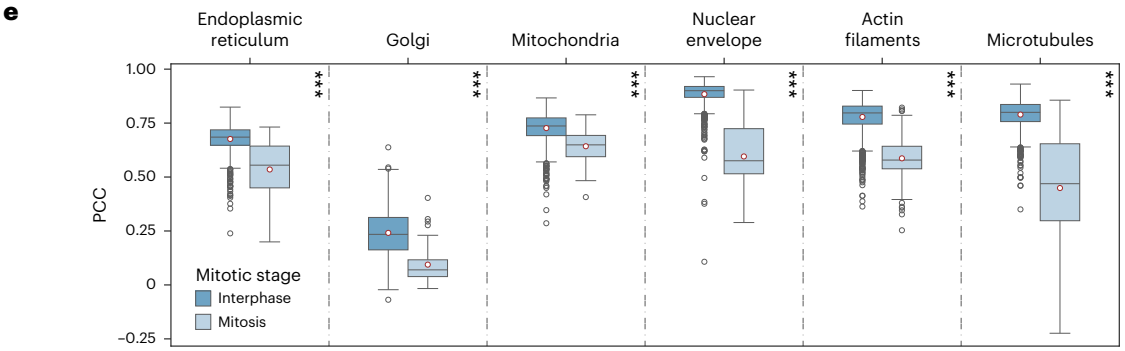
Fig. 1 | Inferior in silico labeling for rare cell populations. a, In silico labeling. Left: training; an organelle-specific model receives label-free transmitted light images and their corresponding fluorescent targets and is trained to minimize the reconstruction error between the model's prediction and the target. Right: inference; each organelle-specific model translates a transmitted light image to its corresponding predicted fluorescence image. The predictions can be combined to an integrated multi-organelle image, which can be used for downstream analyses. **b**, The distribution of rare cell populations in the dataset, comprising 7,622 single cells. Left to right: mitotic stage, 5% of the cells were in one of the five noninterphase mitotic stages; location, 2.4% of the cells were located at the colony edge; volume, 2.2% of the cells had a z score lower than −1.5 relative to the population distribution; and neighborhood density, 4.7% of the cells were in sparse neighborhoods consisting of four or less adjacent cells. **c**, The poor endoplasmic reticulum in silico labeling of three cells. Two cells were in the prometaphase–metaphase stage of mitosis, and the other cell was in a sparse neighborhood. **d**, The poor nuclear envelope in silico labeling of two cells. One cell was in the prometaphase–metaphase, and the other was in the early prometaphase stages of mitosis. In **c** and **d**, the cell-level predictions are shown. Each colored region in the FOV represents the replicated U-Net's[10] average PCC of a cell. The cells not fully contained within the FOV were masked out. Scale bar, 10 µm. Additional examples, along with ground truth and predictions, are shown in Supplementary Fig. 1. **e**, The distribution of the single-cell in silico labeling performance across organelles for cells in interphase (dark blue) versus mitosis (light blue). The box plots show the median (center line), mean (red circle), first and third quartiles (box bounds) and whiskers extending to 1.5× the interquartile range; the individual points beyond the whiskers represent outliers. One-sided Mann–Whitney U test ***$P < 0.001$. The full results including in silico labeling performance, statistical tests and population sizes are available in Supplementary Table 2. Microscope icon in **a** from Freepik.

https://doi.org/10.1038/s41592-025-02960-4









https://doi.org/10.1038/s41592-025-02960-4

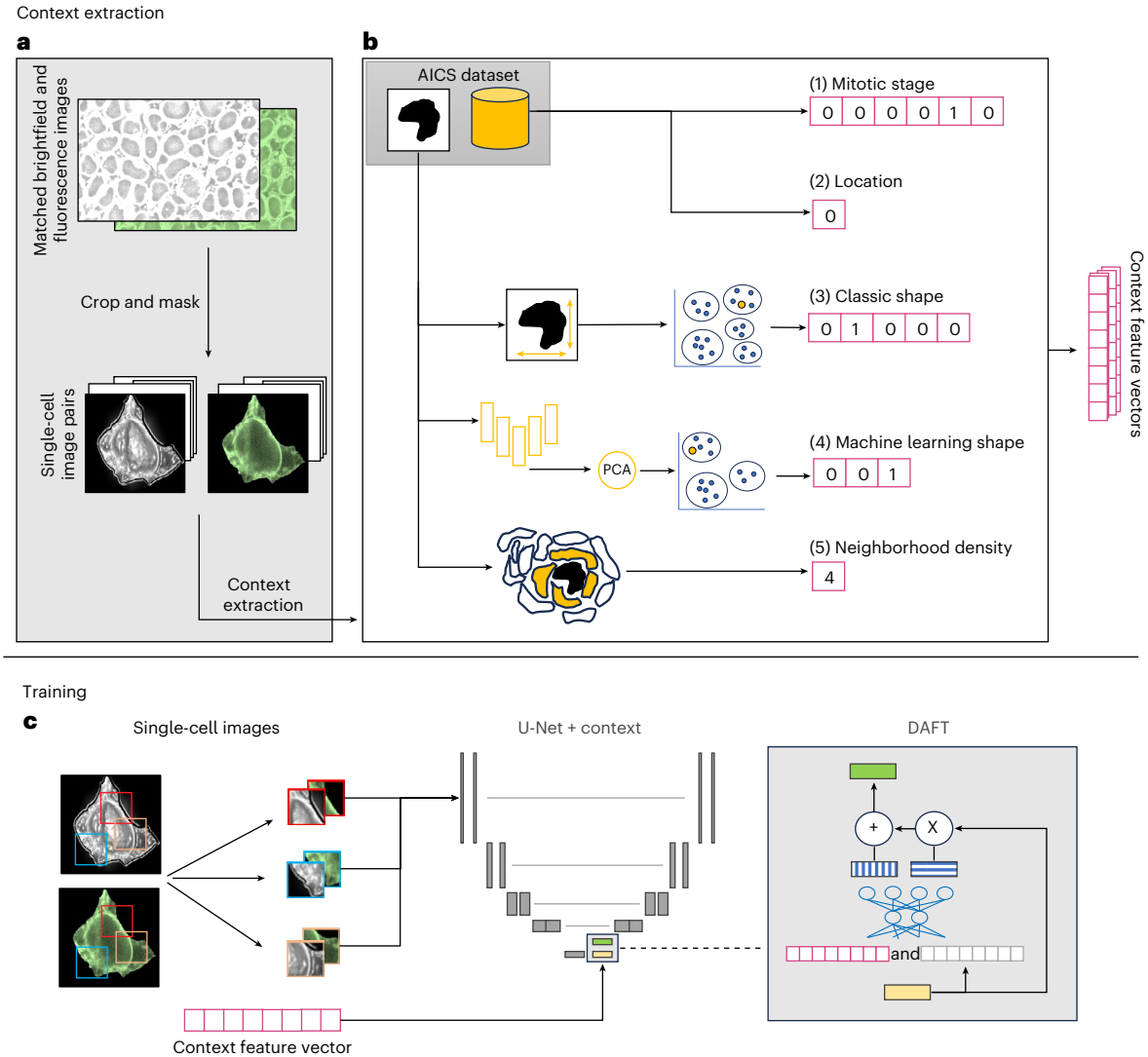


Fig. 2 | CELTIC—incorporating the cell context to the in silico labeling models. **a**,**b**, The context extraction: single-cell images are cropped and masked (**a**), and single-cell context is extracted and represented by a 16-dimensional context feature vector, comprising 5 context types (**b**). The mitotic stage and the edge location indicator contexts were available in the AICS dataset's metadata; other contexts are computationally extracted from the segmented cells and concatenated to define the context feature vector (magenta). PCA, principal component analysis. **c**, CELTIC's architecture. The image patches are fed to CELTIC along with their corresponding context. The cell context is incorporated as an auxiliary input to the U-Net in silico labeling model transforming the U-Net's bottleneck layer (yellow) into a context-enriched feature map (green) via the DAFT[30] block (gray box with a detailed view on the right). DAFT uses its own bottleneck to fuse the image and context, creating a scaler and shifter that adjust the feature map accordingly (see Methods and Supplementary Fig. 4 for more details).

segmentation masks (Methods). We defined five types of context representation per cell, which are briefly described here and are detailed in Methods (Fig. 2b). The first type of context ('mitotic stage') was the mitotic state represented as a one-hot encoding vector. The second type ('location') was a binary indicator representing the cell's location in the colony. The third context type ('classic shape') captured the cell's shape through a one-hot encoded vector derived from the clustering of shape descriptors (height, minimum and maximum width and volume). The fourth type ('machine learning shape') represented the cell shape using autoencoder-compressed binary cell masks that were subsequently clustered, encoding each cell as a one-hot vector on the basis of its shape cluster (Methods; Supplementary Fig. 3, interpretation of clusters). Finally, the fifth context representation ('neighborhood density') was a scalar quantifying the local neighborhood density measured as the number of adjacent cells. These five context representations were concatenated to define a 16-dimensional context feature vector encoding the single-cell context.

CELTIC extends the classic U-Net architecture by incorporating the context vector to its deepest layer. We followed the footsteps of a recent method for fusing image and tabular data called Dynamic Affine Feature Map Transform (DAFT)[30]. DAFT was previously applied with ResNet for classification tasks and, more recently, with U-Net for medical image segmentation[31]. DAFT uses the cellular context vectors to affine-transform the bottleneck image representations (Fig. 2c and Supplementary Fig. 4). In CELTIC, this approach enables the network to learn a unified representation that incorporates both the intrinsic image details encoded at the network deepest layer and the contextual cellular information.

Cell context contributes to the in silico labeling of rare cell populations

We compared CELTIC with the single-cell U-Net to evaluate the contribution of the cell's context to the in silico labeling predictions. We found that context contributed to better predictions of the endoplasmic

reticulum and the nuclear envelope in dividing cells (Fig. 3a and Supplementary Fig. 5). Both models displayed poor performance in predicting the microtubules mitotic spindle, but CELTIC was able to predict the two astral arrays radiating from the spindle poles during mitosis (Fig. 3a and Supplementary Fig. 5). These observations were corroborated by calculating the ΔPCC defined as the signed mean contribution of context to the single-cell-based in silico labeling model, that is, positive ΔPCC values indicated that the cell context contributed to the in silico prediction. Indeed, cell context enhanced the prediction of cells in mitosis for all organelles (Fig. 3b and Supplementary Table 3). The most prominent contribution was measured for microtubules, which is expected given the dramatic changes occurring in microtubule organization during cell division. Similarly, the inclusion of context improved the in silico labeling of cells at the colony edge, most notably for the Golgi apparatus and for microtubules (Fig. 3b). Cells with small volumes also showed improvement in most cases, particularly for the endoplasmic reticulum and for the nuclear envelope. Lastly, adding context enhanced the in silico prediction of cells in sparse neighborhood densities, with the most substantial differences observed for the Golgi apparatus and for the nuclear envelope (Fig. 3b). Intriguingly, context systematically contributed to the in silico labeling of most organelles, even when considering the entire cell population (Fig. 3b, 'all'), but this contribution was marginal because the rare populations constituted only a small fraction of the dataset. It is important to emphasize that these models were trained without the enrichment of rare populations. For instance, only 28 of the 697 (4%) cells that were used to train the endoplasmic reticulum model were mitotic, with some subcategories, such as early prometaphase, containing as few as 2 cells, indicating that CELTIC learns meaningful representations from very limited data.

To elucidate the contribution of each context type to the in silico labeling of rare populations, we conducted an ablation study. Specifically, we randomized each context type by shuffling the corresponding context values across the single-cell population and measured the reduction in the in silico labeling with the shuffled context (Supplementary Table 3). This analysis revealed that several context types were contributing to the in silico labeling of cells undergoing mitosis, unsurprisingly with mitotic stage ranking first among them (Supplementary Fig. 6). The inclusion of the edge context predominantly impacted cells situated at the periphery of the colony for the in silico labeling of the endoplasmic reticulum, nuclear envelope and the microtubules. Cells with small volumes were primarily influenced by the shape contexts, particularly for the endoplasmic reticulum, Golgi apparatus, nuclear envelope and actin. Intriguingly, the neighborhood density context had a minimal effect on most rare populations. To simulate a 'real world' inference of context, for situations when ground truth labels are not available, we measured the effect of context errors introduced by the imperfect computational inference of the mitotic state context from the raw label-free images (Methods). Gradually introducing errors to the mitotic stage context at inference (Supplementary Fig. 7a) and computational inference of the mitotic stage (Supplementary Fig. 7b,c) showed gradual decay for the in silico labeling of microtubules in mitotic cells, emphasizing the importance of high-quality context extraction at inference. Overall, we conclude that cell context, especially one that associates with a corresponding population, such as the cell mitotic stage for mitotic cells, contributes to in silico labeling by 'guiding' context-aware representations that adapt to the different intracellular organizations associated with rare cell populations.

Predicting spindle axis in mitotic cells enabled by CELTIC

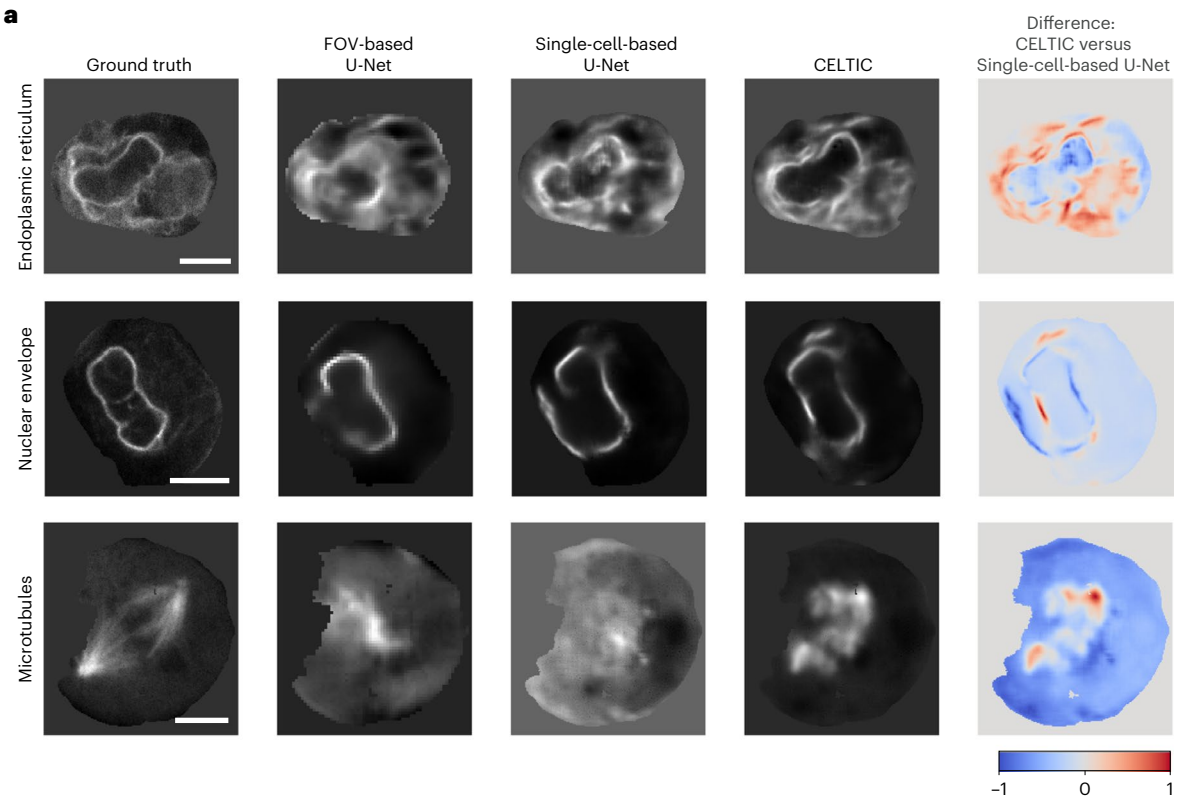
The performance of an in silico labeling model should be evaluated on the basis of its performance for specific application-appropriate downstream analyses, such as organelle localization, counts and shape[8,32,33]. During mitosis, the spindle axis connects the two centrosomes and is a critical determinant of cell division orientation and outcome[34]. We decided to focus on the application-appropriate downstream analysis of determining the location and orientation of the spindle axis during mitosis from label-free images. We applied the single-cell-based in silico model for microtubules, without and with the mitotic state context, for a set of prometaphase–metaphase cells that had not been seen before by the model. For consistency, we selected the middle z slice from each image stack and resized the images to a standard size. Next, we performed a threshold-based segmentation of the in silico labeling predictions and set a line connecting the centers of mass of the two main contours in the image as the predicted spindle axis (Methods). We defined two measurements to evaluate the predicted spindle axis: (1) location error, the distance between the center of the ground truth spindle axis line and the center of the predicted line (ΔC), and (2) orientation error, the angle between these two lines ($\Delta\theta$) (Fig. 4a). The visual assessment indicated that although both models could not perfectly reconstruct the microtubules spindle axis, CELTIC was able to provide reasonable predictions regarding the spindle location and orientation (Fig. 4b). The quantification of ΔC and $\Delta\theta$ reported that the spindle axis was predicted by CELTIC with a median location error ΔC of 3.8 pixels (0.15–0.27 µm per pixel before resizing) and with a median orientation error $\Delta\theta$ of 13°. A comparison with the single-cell-based model (without context) showed deterioration of over 2 fold in the predicted location (Fig. 4c) error and over 1.5 fold in the predicted orientation error (Fig. 4d). To confirm that these prediction errors were not achieved by chance, we conducted a permutation test by keeping the ground truth spindle orientations fixed while randomly reassigning the predicted orientations among different cells. This tests whether our predictions were meaningfully matched to the correct cells rather than just capturing the general distribution of orientations in the population. More specifically, we randomly shuffled all single-cell predictions and calculated the median $\Delta\theta$ prediction errors across the shuffled population. This process was repeated 250,000 times and was used to calculate the statistical significance—the fraction of times that the permuted mean errors were smaller or equal to the observed (unpermuted) mean error, reaching statistical significance P value <0.00002 (Fig. 4e). These results suggest that the inclusion of cell context in CELTIC can enhance the in silico labeling of (application-appropriate) organelle localization patterns that are associated with that context.

A unified model for in silico labeling of multiple organelles

We trained and evaluated different CELTIC models, one model for each one of six different organelles. Thus, although the input images were from the same in-distribution, the outputs were different according to the corresponding organelles' models. We next trained a single CELTIC model with the organelles as the context (Supplementary Fig. 8), that is, this one model was able to predict each of the organelles according to the context—a one-hot encoding of the organelle. This experiment was motivated by three aspects. First, it could further demonstrate the generalizability of CELTIC with a new type of context. Second, this experiment would demonstrate the potential of using CELTIC to harmonize different datasets by incorporating their metadata, in our case organelle type, as context. Third, we speculated that such a multiple-organelle in silico labeling model would benefit from shared representations learned across different organelle contexts. We trained a new CELTIC model with the same FOVs used to replicate the results of Ounkomol et al. (Supplementary Tables 1 and 5a). This unified model achieved a mean PCC of 0.700, exceeding the 0.683 mean PCC of the single-organelle models ($P = 7.12 \times 10^{-8}$, Wilcoxon). The largest improvements were observed for the Golgi apparatus and microtubules (~0.04 PCC), followed by a modest gain for the endoplasmic reticulum (~0.01 PCC), with minimal changes for other organelles (Supplementary Table 5b).

As an additional experiment, we also trained CELTIC Multi-Organelle for the same number of iterations as the single-organelle

https://doi.org/10.1038/s41592-025-02960-4



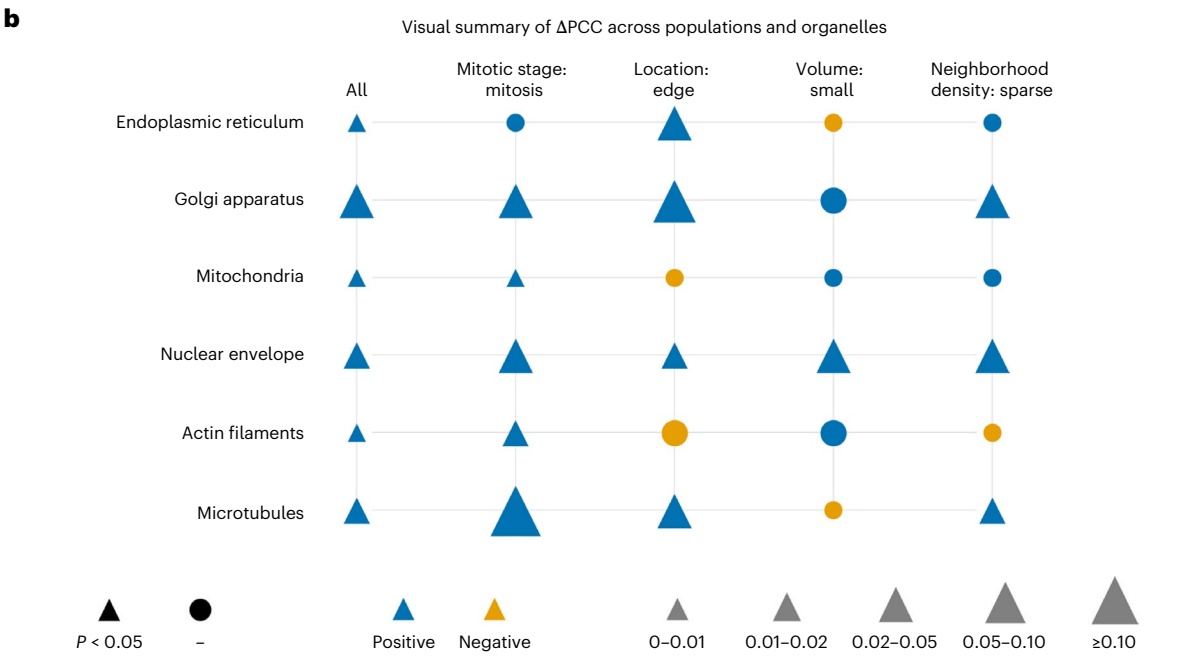
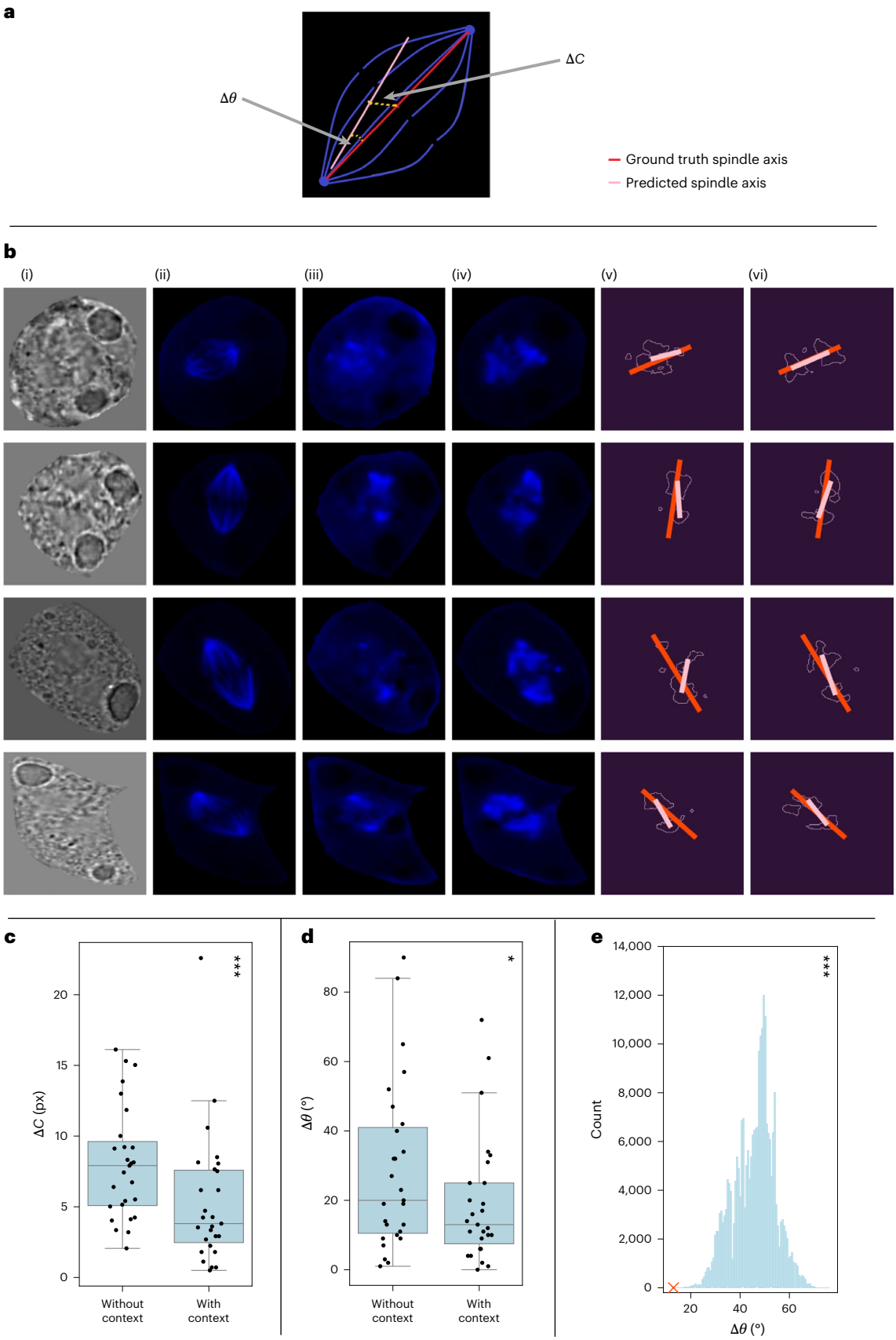


Fig. 3 | Qualitative and quantitative assessment of CELTIC's contribution to the in silico labeling of rare cell populations. a, The in silico labeling visualization of mitotic cells. Each row visualizes a different organelle. Top: the endoplasmic reticulum of a cell in late mitosis. Middle: the nuclear envelope of a cell in late mitosis. Bottom: the microtubules of a cell in the prometaphase–metaphase stage. Left to right: the ground truth fluorescence, the U-Net replicating[10], the single-cell-based U-Net, CELTIC and the pixel-wise difference between the CELTIC model and the single-cell-based model. The red regions indicate positive intensity differences, and the blue regions indicate negative intensity differences. The difference image is scaled from −1 to 1. Shown are *z* slices that have been selected by an expert on the basis of the ground truth full *z* stack images. Scale bar, 5 µm. See Supplementary Fig. 5 for additional examples. **b**, The quantification of the contribution of context, via CELTIC, to the in silico labeling of six organelles in the rare cell populations. Each marker represents the mean ΔPCC for a given population and organelle. The marker's color indicates the direction of the effect (blue (positive), better in silico labeling by the inclusion of context; orange (negative), deteriorated performance after the inclusion of context); the marker's shape encodes significance (circle, not significant; triangle, significant; one-sided Wilcoxon signed-rank test, $P < 0.05$); and the marker's size reflects the magnitude of the absolute difference (|ΔPCC|) binned into five discrete ranges. The full population ('all') is shown for reference. See Table 3 for the full results.

https://doi.org/10.1038/s41592-025-02960-4



models (that is, one sixth of the total original training time). Despite the reduced training time, the model achieved a mean PCC of 0.689, closely matching the single-organelle models (0.683). Several organelles, such as the Golgi apparatus, mitochondria and microtubules, maintained improvements, while others (endoplasmic reticulum, nuclear envelope and actin filaments) exhibited slight PCC decreases of up to 0.014 (Supplementary Table 5c). These results suggest that shared representations learned across the different organelle contexts

https://doi.org/10.1038/s41592-025-02960-4

Fig. 4 | Application-appropriate downstream analysis: in silico prediction of the spindle axis location and orientation. a, The measurements of the predicted spindle axis. The location error (ΔC) denotes the distance between the centers of the predicted (pink) and the ground truth (red) spindle axes. The orientation error ($\Delta\theta$) denotes the angle between the predicted and the ground truth spindle axes. **b**, The four representative cells (rows) in the prometaphase–metaphase mitotic stage. Left to right: the columns represent (i) the brightfield label-free image; (ii) the ground truth fluorescent microtubules image from the *z* stack's middle slice; (iii,iv) the single-cell-based (iii) and CELTIC (iv) in silico labeling of microtubules; and (v–vi) the threshold-based segmentation of the single-cell-based (v) and CELTIC (vi) predictions, the spindle axis prediction (pink) and the ground truth (red). The in silico labeling was performed in three dimensions, whereas the spindle axis prediction and segmentation were conducted using a single two-dimensional slice. Note that the *x*-axis and *y*-axis scaling are different owing to resizing (see 'Application-appropriate assessment of the predicted spindle axis' in Methods). **c**,**d**, The distribution of the location (**c**) and orientation (**d**) errors of the single-cell-based versus CELTIC prediction of the spindle axis. Each data point corresponds to a single cell. The box plots show the median (center line), first and third quartiles (box bounds) and whiskers extending to 1.5× the interquartile range; all individual data points are overlaid. The one-sided Wilcoxon signed-rank test was used to reject the null hypothesis that the values with CELTIC are similar or larger than with the single-cell-based model. $N = 27$ cells, P value (location) <0.00017, P value (orientation) <0.028. px, pixels. **e**, The permutation test. CELTIC's observed orientation error (the red X, median $\Delta C = 13°$) and the histogram of random shuffling ($N = 250{,}000$). $P < 0.00002$. $^{*}P < 0.05$, $^{***}P < 0.001$.

can surpass the in silico labeling models of single organelles, with reduced data and computational resources. Moreover, this highlights the potential of harmonizing different datasets with a single CELTIC model by incorporating their metadata as context.

Context-dependent generative organelle localization with CELTIC

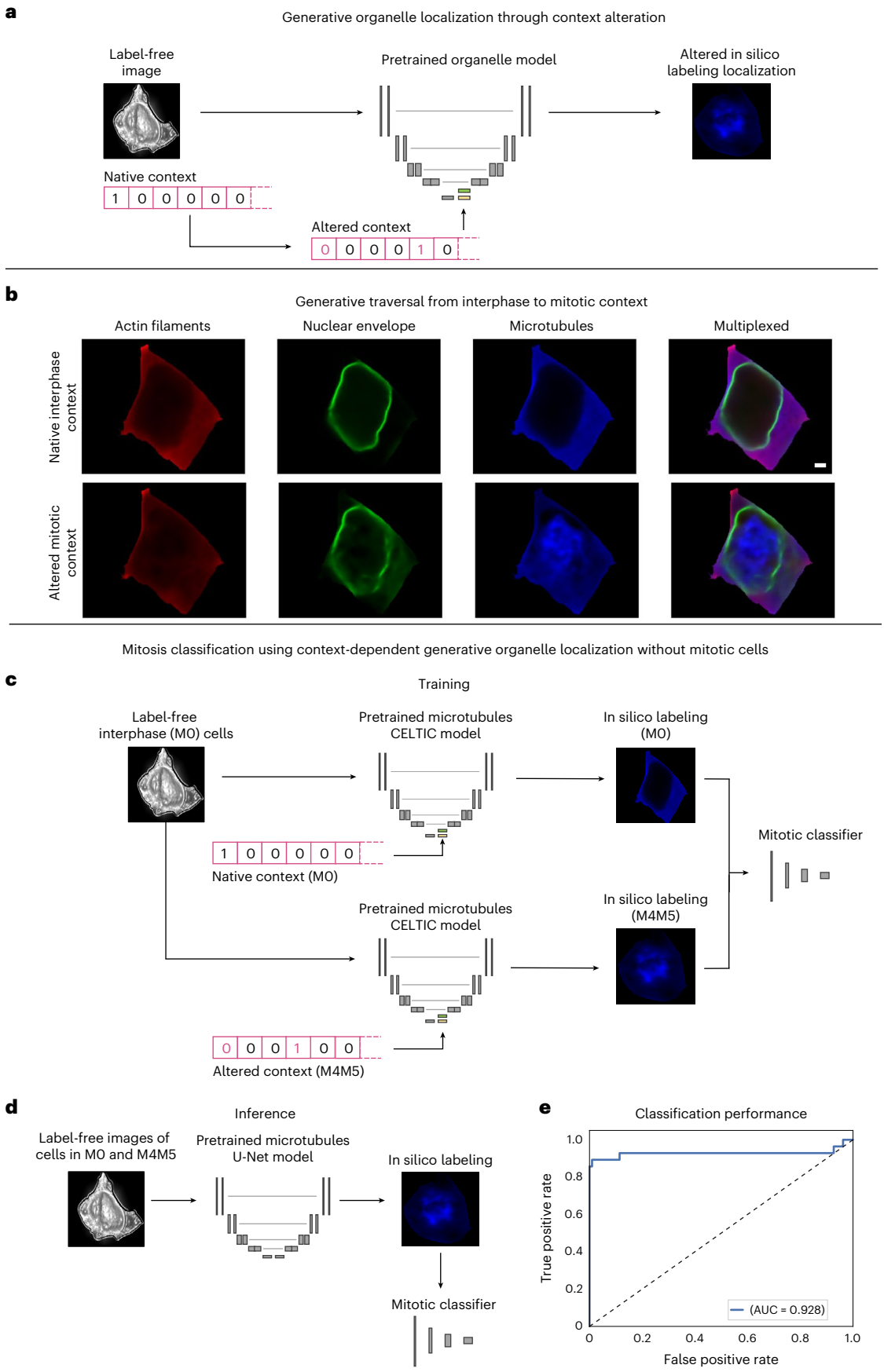
The CELTIC representation incorporates the label-free image along with the cell context, which are jointly used to make the in silico prediction. This representation where the context can be manually manipulated can be used to generate a series of in silico labeling images of the same label-free image under varying contexts (Fig. 5a). Specifically, it does this by manipulating the context and assessing the corresponding alterations in the CELTIC-generated in silico labeling images to infer context-dependent changes in the intracellular organization. While the output of such context manipulations cannot be validated against ground truth images, they can provide insights into context-dependent changes in cellular organization. To illustrate the potential of this approach, we generated the integrated in silico labeling of the actin filaments, the nuclear envelope and the microtubules of four nondividing cells where the actin filaments and the microtubules formed widespread networks throughout the cytoplasm, with a solid nuclear envelope surrounding the nucleus (Fig. 5b, top, and Supplementary Fig. 9a, top rows). Upon manipulating the cells to a prometaphase–metaphase mitotic context, for the same label-free images, CELTIC generated an integrated in silico labeling where (1) the nuclear envelope partially disassembled, (2) the actin filaments dispersed in the cytoplasm following the nuclear envelope disassembly and reorganized to form a ring at the cell equator in preparation for anaphase and (3) the microtubules reorganized to form the aligned mitotic spindle (Fig. 5b, bottom, and Supplementary Fig. 9a, bottom rows). Repeating the same process in the opposite direction, we used the CELTIC generative organelle localization to transition three cells from prometaphase–metaphase to interphase. This process produced integrated in silico labeling images resembling the interphase phenotype with dispersed actin filaments and microtubules in the cytoplasm and a more rigid-appearing nuclear envelope (Supplementary Fig. 9b). Next, upon manipulating the localization context of three cells from 'interior' to 'edge', the same label-free image induced a pronounced polarization of cytoplasmic structures, whereas the nuclear envelope largely retained its positioning, consistent with the observations reported in ref. 28 (Supplementary Fig. 10). As a final demonstration, we altered the different contexts of the same cell. First, we manipulated the cell's location context from interior to edge inducing an altered localization of the actin filaments and microtubules toward the cell's periphery (Supplementary Fig. 11a, top versus middle), in concurrence with ref. 28. Second, we manipulated the same cell's mitotic stage context from interphase to mitotic (Supplementary Fig. 11a, top versus bottom), demonstrating the generation of mitotic localizations, resembling the ground truth phenotypes (Supplementary Fig. 11b).

To systematically analyze how altering the cell's context changes the corresponding CELTIC-generated image, we manipulated each of the five context types and calculated the PCC between the two in silico-labeled generated images—before and after the context alteration for 230 cells from the dataset (Supplementary Fig. 12). A high correlation implies that the context does not alter the in silico labeling, and a low correlation indicates that the context change induced an alteration in the in silico labeling, suggesting that the corresponding organelle is changing with the change of the corresponding context. Altering the cell mitotic stage context made the most dramatic change in the generated image space, particularly for the microtubules, the nuclear envelope and the Golgi apparatus. These changes in the computationally generated images were aligned with the phenotypes in the ground truth fluorescence images (Supplementary Fig. 13). The alteration of the location context was mapped to changes in the endoplasmic reticulum. The alteration of other contexts did not change the corresponding generated images much. Given the superiority of CELTIC on the corresponding cell populations (Fig. 3), these results showing context changes that do not translate to altered generated images can be explained by CELTIC's (pre-DAFT) representations, which better generalize to the corresponding rare cell populations. In other words, we suggest that, in these cases, the model layers preceding the DAFT layer learned more effective representations of the rare populations. Altogether, the explicit representation of cell context enables a guided traversal along the context axis of the same cell, overcoming the vast variability between different cells, for easier visual interpretation[35]. An analysis of the alteration in cell organization along these context traversals may teach us about the gradual change in integrated cell state along a physiological process.

Can we use context-dependent generative organelle localization to train a mitosis classification model without showing it any mitotic cells during training? To test this, we assembled a training dataset of 4,080 brightfield images of interphase (M0) cells from 360 FOVs. Each of these images was used to in silico label two microtubules

Fig. 5 | Context-dependent generative in silico labeling. a, The approach: the native context feature vector is modified and used to generate an altered CELTIC in silico labeling. **b**, The in silico labeling of a cell in interphase: the native interphase context (top) versus mitotic context (bottom). Left to right: actin filaments (red), nuclear envelope (green), microtubules (blue) and a multiplexed representation of all three organelles together. Shown are the central *z* slices. Scale bar, 2 µm. $N = 4$ replicates (Supplementary Fig. 9a). **c**, Training a mitosis classifier using interphase cells by altering their mitotic stage context. Each interphase brightfield image is used twice as an input to the CELTIC microtubules model with a native M0 context and with an altered M4M5 context. The resulting in silico-labeled images are used to train a binary classifier to distinguish between the two contexts. **d**, The mitosis classifier is evaluated on U-Net-generated in silico images of interphase and mitotic cells that were experimentally imaged. **e**, The AUC of the classifier on the held-out test microtubules dataset (Supplementary Table 3). In **a** and **c**, pink digits in the context vector denote the altered context.

https://doi.org/10.1038/s41592-025-02960-4



images using our CELTIC microtubules model (Fig. 5c). The first image was in silico labeled with interphase as the mitotic context and the second with prometaphase–metaphase (M4M5) as the context. This training dataset did not include any cells in mitosis but did include CELTIC synthetic-generated mitotic M4M5 cells that exceeded the overall number of experimentally imaged M4M5 cells in the 18,100 FOVs of the entire WTC-11 dataset. We used this dataset to train a binary classifier that discriminated between CELTIC microtubules images generated with the mitotic stages context of M0 versus M4M5. This classifier was tested on an independent dataset that included cells in interphase and cells in prometaphase–metaphase, in silico labeled by the (context-free) microtubules U-Net model (Fig. 5d), achieving an area under the curve (AUC) score of 0.928 (Fig. 5e). These results confirmed that the context-dependent generative organelle localization can generate in silico images that resemble the altered context and demonstrated how context alteration can be used to generate synthetic data to overcome the scarcity of training data for mitosis classification models.

Discussion

Measuring and interpreting how organelles adjust their internal structure and organization with respect to one another under different experimental conditions or during a physiological process is the 'holy grail' of cell biology. In silico labeling is a promising method to overcome several of the technical hurdles that currently prevent us from reaching this goal. Here, we report that in silico labeling is confounded by rare cell contexts owing to alterations in the cells' optical properties that lead to out-of-distribution label-free images. CELTIC guides context-aware representations by incorporating the explicit cell context into the in silico labeling model. We show that CELTIC enhances the in silico labeling of rare cell populations, especially for organelle localization patterns associated with the corresponding context, and highlights its potential for modeling context transitions through context-dependent generative capabilities. Our results emphasize the strong link between the cell's context and its intracellular organization.

In this study, we demonstrated that the integration of intrinsic cell contexts, such as the cell's mitotic stage, can enhance the in silico labeling prediction. However, cell context is a very broad term and can include extrinsic properties such as the cell type, perturbations, disease state, assay, microscope, fluorescent marker and imaging parameters. The inclusion of extrinsic context descriptors could be used to harmonize datasets from multiple resources to one large dataset. Thus, context-dependent in silico labeling has the potential to be the enabler toward training general in silico labeling 'foundation models'. Christiansen et al. made the first step in this direction by arguing for the benefit of training one model for the in silico labeling of multiple fluorescent channels[12], which in our case was translated to providing the organelle type as context. An exciting possibility is to integrate the cell's continuous state during the progression of a physiological process as intrinsic context for in silico labeling, for example, using the FUCCI system as a rich cell cycle context[36] or by the computational prediction of the continuous cell state[37–44]. Notably, while cell-extrinsic contexts should be available via the experimental metadata, intrinsic contexts are commonly computationally derived from the image data, thus introducing measurement errors that would affect the in silico labeling. Specifically, the cell's mitotic state, periphery location and segmentation were available to us in the AICS dataset. However, as demonstrated in this study, in the 'wild' these cell contexts would be computationally extracted from the raw images, introducing errors in the context and limiting the full potential of context-dependent in silico labeling. Practical demonstrations of data harmonization toward training foundation models remain to be proven in future research.

An alternative approach for incorporating context is through weakly supervised representation learning, with the cell context as the 'weak' label. Previous recent studies guided the representations of protein localization patterns using the protein[45,46] or the perturbation[47] as the weakly supervised label. In principle, cell context can also be used as a label for weakly supervised representation learning. CELTIC's explicit 'injection' of the cell context to the representations has two major benefits. First, the possibility to control the context to morph a specific cell along the context trajectory, while fixing the other factors of cell–cell variability. This generative capacity along a context trajectory has the potential to serve as a hypothesis generation method to follow how the different organelles and their spatial interorganelle dependencies are changing as a function of their contexts. The second benefit of explicitly injecting context over weak supervision is the exponential growth of the combinatorial context space with the number of different contexts. There are many possible contexts, and thus, training representations to simultaneously encode multiple weak context labels is not feasible[48]. CELTIC bypasses this limitation by avoiding the technical challenge of learning representations that encode context by providing the context explicitly. Accordingly, the explicit context representation enables learning out-of-distribution label-free images from a very small set of examples consisting of tens of cells per rare population. One advantage of weak supervision over CELTIC is the simpler inference that does not require the weak context label as input.

Context does not improve in silico labeling for all organelles to the same extent. For example, in silico labeling of mitochondria was not improved much with the inclusion of context, suggesting that the corresponding brightfield patterns used to localize the mitochondria do not depend much on the cell's context. The systematic characterization of which contexts contribute to the in silico labeling of different cell populations can be used as a phenotypic signature of these populations, indicating how organelles reorganize upon physiological context alteration. Moreover, screening for the contexts that contribute most to the in silico labeling in CELTIC, compared with a model without context, can be used as a method to identify the cell's context. This approach for context prediction is especially relevant for application-appropriate measurements, such as predicting the spindle axis, which should be dramatically improved during mitosis owing to the strong association between intracellular organizational patterns and the context. Thus, the same approach can be used to discover unknown patterns that are associated with a specific cell context. Of course, this type of analysis must be carefully validated to rule out the possibilities that the pre-DAFT CELTIC representations generalized to the cellular context or that the model was not able to encode the organelle-specific context alteration owing to a lack of sufficient training data of rare populations.

Online content

Any methods, additional references, Nature Portfolio reporting summaries, source data, extended data, supplementary information, acknowledgements, peer review information; details of author contributions and competing interests; and statements of data and code availability are available at https://doi.org/10.1038/s41592-025-02960-4.

References

1. Carlton, J. G., Jones, H. & Eggert, U. S. Membrane and organelle dynamics during cell division. *Nat. Rev. Mol. Cell Biol.* **21**, 151–166 (2020).
2. Kroll, J. & Renkawitz, J. Principles of organelle positioning in motile and non-motile cells. *EMBO Rep.* **25**, 2172–2187 (2024).
3. Lee, A., Hirabayashi, Y., Kwon, S.-K., Lewis, T. L. Jr & Polleux, F. Emerging roles of mitochondria in synaptic transmission and neurodegeneration. *Curr. Opin. Physiol.* **3**, 82–93 (2018).
4. Ahmed, R. E., Tokuyama, T., Anzai, T., Chanthra, N. & Uosaki, H. Sarcomere maturation: function acquisition, molecular mechanism, and interplay with other organelles. *Philos. Trans. R. Soc. Lond. B* **377**, 20210325 (2022).

5. Miroshnikova, Y. A. & Wickström, S. A. Mechanical forces in nuclear organization. *Cold Spring Harb. Perspect. Biol.* **14**, a039685 (2022).
6. Ye, Z. et al. Burst of hopping trafficking correlated reversible dynamic interactions between lipid droplets and mitochondria under starvation. *Exploration* **3**, 20230002 (2023).
7. Garini, Y., Young, I. T. & McNamara, G. Spectral imaging: principles and applications. *Cytometry A* **69**, 735–747 (2006).
8. Elmalam, N., Ben Nedava, L. & Zaritsky, A. In silico labeling in cell biology: potential and limitations. *Curr. Opin. Cell Biol.* **89**, 102378 (2024).
9. Sullivan, D. P. & Lundberg, E. Seeing more: a future of augmented microscopy. *Cell* **173**, 546–548 (2018).
10. Ounkomol, C., Seshamani, S., Maleckar, M. M., Collman, F. & Johnson, G. R. Label-free prediction of three-dimensional fluorescence images from transmitted-light microscopy. *Nat. Methods* **15**, 917–920 (2018).
11. Cheng, S. et al. Single-cell cytometry via multiplexed fluorescence prediction by label-free reflectance microscopy. *Sci. Adv.* **7**, eabe0431 (2021).
12. Christiansen, E. M. et al. In silico labeling: predicting fluorescent labels in unlabeled images. *Cell* **173**, 792–803.e19 (2018).
13. Kandel, M. E. et al. Phase imaging with computational specificity (PICS) for measuring dry mass changes in sub-cellular compartments. *Nat. Commun.* **11**, 6256 (2020).
14. Kandel, M. E. et al. Multiscale assay of unlabeled neurite dynamics using phase imaging with computational specificity. *ACS Sens.* **6**, 1864–1874 (2021).
15. Jo, Y. et al. Label-free multiplexed microtomography of endogenous subcellular dynamics using generalizable deep learning. *Nat. Cell Biol.* **23**, 1329–1337 (2021).
16. Ben-Yehuda, K. et al. Simultaneous morphology, motility, and fragmentation analysis of live individual sperm cells for male fertility evaluation. *Adv. Intell. Syst.* **4**, 2100200 (2022).
17. Gu, S. et al. Label-free cell tracking enables collective motion phenotyping in epithelial monolayers. *iScience* **25**, 104678 (2022).
18. Cross-Zamirski, J. O. et al. Label-free prediction of cell painting from brightfield images. *Sci. Rep.* **12**, 10001 (2022).
19. Somani, A. et al. Virtual labeling of mitochondria in living cells using correlative imaging and physics-guided deep learning. *Biomed. Opt. Express* **13**, 5495–5516 (2022).
20. Wang, R., Butt, D., Cross, S., Verkade, P. & Achim, A. Bright-field to fluorescence microscopy image translation for cell nuclei health quantification. *Biol. Imaging* **3**, e12 (2023).
21. Noy, L. et al. Sperm-cell DNA fragmentation prediction using label-free quantitative phase imaging and deep learning. *Cytometry A* **103**, 470–478 (2023).
22. Atwell, S. et al. Label-free imaging of 3D pluripotent stem cell differentiation dynamics on chip. *Cell Rep. Methods* **3**, 100523 (2023).
23. Asmar, A. J. et al. High-volume, label-free imaging for quantifying single-cell dynamics in induced pluripotent stem cell colonies. *PLoS ONE* **19**, e0298446 (2024).
24. Ivanov, I. E. et al. Mantis: high-throughput 4D imaging and analysis of the molecular and physical architecture of cells. *PNAS Nexus* **3**, gae323 (2024).
25. Liu, Z. et al. Robust virtual staining of landmark organelles with Cytoland. *Nat. Mach. Intell.* **7**, 901–915 (2025).
26. Follain, G. et al. Fast label-free live imaging reveals key roles of flow dynamics and CD44–HA interaction in cancer cell arrest on endothelial monolayers. Preprint at *bioRxiv* https://doi.org/10.1101/2024.09.30.615654 (2024).
27. Tonks, S. et al. Can virtual staining for high-throughput screening generalize? In *Proc. European Conference on Computer Vision (ECCV) Workshops* 34–50 (2024).
28. Viana, M. P. et al. Integrated intracellular organization and its variations in human iPS cells. *Nature* **613**, 345–354 (2023).
29. Ronneberger, O., Fischer, P. & Brox, T. U-Net: convolutional networks for biomedical image segmentation. In *International Conference on Medical Image Computing and Computer Assisted Intervention (MICCAI)* 241 (2015).
30. Wolf, T. N., Pölsterl, S. & Wachinger, C. DAFT: a universal module to interweave tabular data and 3D images in CNNs. *Neuroimage* **260**, 119505 (2022).
31. Rickmann, A.-M., Xu, M., Wolf, T. N., Kovalenko, O. & Wachinger, C. HALOS: hallucination-free organ segmentation after organ resection surgery. *Inf. Process. Med. Imaging* **667**, 678 (2023).
32. LaChance, J. & Cohen, D. J. Practical fluorescence reconstruction microscopy for large samples and low-magnification imaging. *PLoS Comput. Biol.* **16**, e1008443 (2020).
33. Chen, J., Viana, M. P. & Rafelski, S. M. When seeing is not believing: application-appropriate validation matters for quantitative bioimage analysis. *Nat. Methods* **20**, 968–970 (2023).
34. Yamashita, Y. M. Regulation of asymmetric stem cell division: spindle orientation and the centrosome. *Front. Biosci.* **14**, 3003–3011 (2009).
35. Rotem, O. & Zaritsky, A. Visual interpretability of bioimaging deep learning models. *Nat. Methods* **21**, 1394–1397 (2024).
36. Sakaue-Sawano, A. et al. Visualizing spatiotemporal dynamics of multicellular cell-cycle progression. *Cell* **132**, 487–498 (2008).
37. Gut, G., Tadmor, M. D., Pe'er, D., Pelkmans, L. & Liberali, P. Trajectories of cell-cycle progression from fixed cell populations. *Nat. Methods* **12**, 951–954 (2015).
38. Eulenberg, P. et al. Reconstructing cell cycle and disease progression using deep learning. *Nat. Commun.* **8**, 463 (2017).
39. Rappez, L., Rakhlin, A., Rigopoulos, A., Nikolenko, S. & Alexandrov, T. DeepCycle reconstructs a cyclic cell cycle trajectory from unsegmented cell images using convolutional neural networks. *Mol. Syst. Biol.* **16**, e9474 (2020).
40. Yang, K. D. et al. Predicting cell lineages using autoencoders and optimal transport. *PLoS Comput. Biol.* **16**, e1007828 (2020).
41. Szkalisity, A. et al. Regression plane concept for analysing continuous cellular processes with machine learning. *Nat. Commun.* **12**, 2532 (2021).
42. Zaritsky, A. et al. Interpretable deep learning uncovers cellular properties in label-free live cell images that are predictive of highly metastatic melanoma. *Cell Syst.* **12**, 733–747.e6 (2021).
43. Stallaert, W. et al. The structure of the human cell cycle. *Cell Syst.* **13**, 230–240.e3 (2022).
44. Shakarchy, A. et al. Machine learning inference of continuous single-cell state transitions during myoblast differentiation and fusion. *Mol. Syst. Biol.* **20**, 217–241 (2024).
45. Kobayashi, H., Cheveralls, K. C., Leonetti, M. D. & Royer, L. A. Self-supervised deep learning encodes high-resolution features of protein subcellular localization. *Nat. Methods* **19**, 995–1003 (2022).
46. Razdaibiedina, A. et al. PIFiA: self-supervised approach for protein functional annotation from single-cell imaging data. *Mol. Syst. Biol.* **20**, 521–548 (2024).
47. Moshkov, N. et al. Learning representations for image-based profiling of perturbations. *Nat. Commun.* **15**, 1594 (2024).
48. Frénay, B. & Verleysen, M. Classification in the presence of label noise: a survey. *IEEE Trans. Neural Netw. Learn. Syst.* **25**, 845–869 (2014).

Methods

Data

We used the AICS WTC-11 hiPS cell Single-Cell Image Dataset v1[28]. From the FOV spinning-disk confocal microscopy section, we used the 16-bit *z* stack images, acquired with a 100× objective, with a resolution of 624 pixels × 924 pixels and a physical pixel size of 0.108 µm × 0.108 µm. Each *z* stack consisted of 50–75 slices with 0.29 µm between consecutive slices. Specifically, we used the brightfield channel and the EGFP-tagged cellular structure channel for the following proteins: alpha-tubulin (microtubules), β-actin (actin filaments), lamin B1 (nuclear envelope), sec61B (endoplasmic reticulum), STGAL1 (Golgi apparatus) and Tom20 (mitochondria). For the FOV segmentation, we used the cell segmentation channel, a 3D integer labelmap representing cell locations within the FOV. Each cell's unique index, as provided in the WTC-11 dataset metadata, was located in the labelmap. The FOV was cropped to include only the pixels corresponding to the identified cell, with all other pixels outside this region set to 0. Lastly, we used the 'edge_flag' and 'cell_stage' features from the metadata CSV file. Note that we did not use the single-cell crops as provided in the WTC-11 dataset, as they included extraneous margins that would have required additional processing. Instead, we worked with the full FOV images. This approach allowed us to develop a pipeline from the FOV segmentation to the single-cell analysis, which was essential for extracting certain features, such as neighbor cell quantification, that depend on the complete visual field.

We selected 80 FOV images per organelle to train and validate all models, that we call the 'development set'. The images were selected from the list provided in the code repository of ref. 10. The repository is FOV-based and lacks the single-cell-level metadata that we aimed to leverage from the WTC-11 single-cell dataset. While the WTC-11 dataset mostly overlaps with the FOV-based dataset, it is not identical. In cases where images from this list were not available in the single-cell dataset, we randomly replaced them with FOV images that closely matched in terms of their cellular properties, such as cell count, mitotic percentage and edge characteristics. For example, for the endoplasmic reticulum, WTC-11 contained only 73 of the 80 specified FOVs. We characterized these 73 FOVs by collecting statistics on average cell count, imaging days and mitotic stage distribution. We then loaded seven additional WTC-11 FOVs with similar statistics, specifically matching the average cell count (±1 cell), imaging days (preferentially from the same day) and qualitatively consistent proportional distribution across mitotic stages. In addition, we selected 100 FOV images for independent evaluations on the basis of the same criteria. We call these images the 'test set'.

We used the FOV cell segmentation label maps to extract individual cells from the FOV images. We excluded cells that were not entirely inside the FOV owing to the lack of metadata and inability to reliably extract shape descriptors. For each cell, we constructed a 3D bounding box and applied the segmentation masks to the brightfield and EGFP FOV images, resulting in a two-channel single-cell dataset. The pixel intensities of all images were normalized to have a mean of 0 and a standard deviation of 1 to account for variations in illumination intensity.

We categorized each cell into populations on the basis of the following criteria: mitotic stage (mitosis versus interphase) by using the metadata cell_stage field; location (edge versus interior) by using the metadata edge_flag field; and volume, where small cells were those with a volume lower than −1.5 standard deviations from the mean. For the neighborhood density analysis, the cells were considered in sparse neighborhoods if they had fewer than five adjacent cells. This categorization defines rare cell populations (Fig. 1b) and is technically and biologically justified in the 'Context representations' section below. Throughout the paper, we use the term 'out-of-distribution' to describe rare events in agreement with the definitions in refs. 49,50.

In silico labeling replication

To replicate the in silico U-Net model in ref. 10, we used the development set. For training, we used 64 FOV images (56 for training, 8 for validation), and for testing, we used the remaining 16 FOVs. We followed the preprocessing steps that were reported in the paper: resizing the *z* slices to 244 pixels × 366 pixels and normalizing each image to a mean of 0 and a standard deviation of 1. We refer to these models as FOV-based models and report their replication results in Supplementary Table 1.

We used the test set to assess the in silico labeling performance of the FOV-based models at the single-cell level. The procedure for extracting individual cells was adjusted ('Data' section). Specifically, we resized the segmentation channel to match the size of the FOV images, without smoothing, to ensure consistency of the segmentation at the resolution of a single pixel. We then applied a binary mask to both the predictions and their corresponding fluorescent targets and calculated the PCC between them, considering only the pixels belonging to the cell, as defined by the segmentation mask.

To assess the null hypothesis that the distributions of the rare and non-rare populations are identical, with no substantial difference in their medians, we used the one-sided Mann–Whitney *U* test (see Supplementary Table 2 for results).

Context representations

We defined five types of context representation per cell. The mitotic stage and location in the colony are provided in the WTC-11 dataset. The shape representations and neighborhood density were computed.

Mitotic stage. The mitotic stage was considered, given the well-documented changes in organelle localization during mitosis[1]. The representation was extracted from the single-cell metadata CSV file, where the cell mitotic stage is indicated as one of six stages: 'M0' (interphase), 'M1M2' (prophase), 'M3' (early prometaphase), 'M4M5' (prometaphase–metaphase), 'M6M7_single,' and 'M6M7_complete' (anaphase–telophase–cytokinesis in two stages). These annotations are provided in the AICS metadata, where they were generated by a deep learning-based classifier and rule-based criteria[28]. We represented this information numerically as a six-column one-hot vector, with each column corresponding to one of the six cell cycle stages.

Location in the colony. The location of cells at the colony edge was informed by previous work[28], which observed the polarization of organelles toward the outer periphery of edge cells. The representation was extracted from the edge_flag column in the single-cell metadata CSV file ('true' for edge cells, 'false' for interior cells). We represented this context as a Boolean.

Shape representations. The cell shape and volume were included on the basis of the observation that organelle size and morphology scale with cell size[51,52]. We used two types of shape representation. The first shape representation was of predefined shape features capturing classic shape attributes such as cell volume, termed classic shape. The second shape representation was machine learning-derived shape features that captured more subtle and complex attributes, which are not necessarily encoded in the 'classic' representation, and termed machine learning shape. These two representations are described next.

Classic shape. We used the segmentation masks to calculate a predefined set of cell shape features. The cell height was defined as the difference between the minimal and maximal *z* coordinates within the mask (measured in pixels). The cell volume was computed as the total number of foreground pixels in the segmentation mask. The cell width was defined using two measures: min_xy, the shorter axis length in pixels between the *x* and *y* axes and max_xy, and the longer axis length in pixels between the *x* and *y* axes. Following the minimum–maximum scaling of these measurements, we applied *k*-means clustering with $k = 5$, where *k* was selected according to the elbow method, which evaluates the relationship between the number of clusters and the

within-cluster sum of squares. Each cell was assigned to a cluster, which was represented as a one-hot encoding vector. We also attempted incorporating the distances of each cell to the *k*-means centroids but found this to be less effective.

Machine learning shape. Each cell's binary segmentation mask was first zero-padded to a fixed-size 3D box, centered within the largest bounding box observed for the organelle and then resized to a standardized shape of (32, 64, 64) voxels (z, y, x) using nearest-neighbor interpolation. This ensured uniform input dimensions for training the autoencoder. We trained an autoencoder to encode representations of these single-cell segmentations. The encoder consisted of two 3D convolution layers (depths 16 and 32), each followed by ReLU activation and max-pooling with kernel size 2 and stride 2. The decoder included two 3D transposed convolution layers, with ReLU activation for the first and sigmoid activation for the last. The autoencoder architecture was deliberately kept simple, with the aim of capturing broad morphological differences while avoiding overfitting to fine-grained shape variation. The autoencoder was trained for ten epochs to minimize the mean squared error (MSE) on the cell segmentation images from the training set. The final latent representation had 32 feature maps, each of spatial dimensions $8 \times 16 \times 16$, yielding a latent space shaped as (8, 16, 32). This was reshaped into vectors of size 65,536. A principal component analysis reduced the dimensionality to five main components. We then performed *k*-means clustering, determining the optimal number of clusters ($k = 3$) using the elbow method. The shape clusters were represented by cluster membership, with a binary indicator for cluster belongingness.

Neighborhood density. Local cell density was included on the basis of the observation that confinement influences contractility, cytoskeletal organization and organelle anchorage[2], for example, in the context of cell cycle progression[37]. For each cell, we identified and counted the neighboring cells according to the segmentation image of the FOV. The number of neighboring cells was minimum–maximum scaled according to the minimum and maximum number of neighbors in the training set.

CELTIC architecture

We adopted the four-level U-Net architecture from ref. 10, selected for its demonstrated effectiveness on this dataset. The model consists of convolutional layers, batch normalization and ReLU activations. The complete architecture is detailed in Supplementary Fig. 4. Similarly to the original model, which was trained on FOV images, we used patching, as even a single cell in 3D was too large to fit into the GPU. We used random patches of size $32 \times 64 \times 64$ (z, y, x), excluding those that contained no signal within the segmentation mask. We integrated the DAFT[30] to the U-Net's architecture. DAFT extends the FiLM[53] method to fuse image data with its complementary tabular information. Integrating image and tabular data is challenging owing to their dimensionality mismatch, and naive approaches such as concatenating latent representations often result in networks prioritizing image features, with minimal improvements over traditional CNNs[53,54]. The DAFT method addresses this limitation by facilitating a dynamic exchange of information between the 3D image and the tabular data through an auxiliary neural network, enhancing their interaction and integration capacity. Specifically, the U-Net's bottleneck produced a feature map tensor (of size of $512 \times 2 \times 4 \times 4$) that was provided as input to the DAFT architecture. This input underwent 3D adaptive average pooling, reducing each feature map to a single scalar by averaging across the spatial dimensions. The result was a 512-dimensional global context vector, with each value summarizing the overall strength of a learned feature, independent of its spatial location. This tensor was then concatenated with the 16-dimensional context feature vector and passed through a fully connected layer, reducing its size through a bottleneck architecture. The compression factor, a hyperparameter, determined the degree of compression, and the output was followed by a ReLU activation function. Several factors were tested (2, 4, 12, 32, 48, 128), and the final chosen factor for each organelle is reported in Supplementary Table 3. A subsequent linear layer decoded the representation to twice the original feature map size. Half of this decoded representation was activated by a sigmoid function and used to scale the feature map values, whereas the other half shifted them. The adjusted feature map was then propagated through the U-Net's upstream layers. The alternative configurations of applying DAFT to different or additional layers showed inferior results and thus were not further pursued.

Model training and performance evaluation

To evaluate the contribution of cell context, we trained and evaluated in silico models for each organelle dataset: CELTIC (with the DAFT block and context) versus U-Net. We randomly split the development set into three subsets: 'train', 'validation 1' and 'validation 2', with a ratio of 7:1:2. The split was performed at the FOV level (rather than the single-cell level) to ensure that cells from the same FOV appeared in only one data subset, thereby preventing the models from learning batch effects. The models were trained for 60,000 iterations, where each iteration corresponds to a mini-batch update. The images were loaded cyclically into a buffer, with patches randomly cropped to form mini-batches. The MSE was used as the loss function, calculated on the masked signal area. During training, MSE was evaluated on the 'validation 1' subset once every 100 iterations, and the best model was selected on the basis of the iteration that achieved the lowest validation score. To determine the optimal bottleneck size hyperparameter for DAFT, we trained both the U-Net and the CELTIC models using three different split seeds of the development set to assess the robustness of model selection. The optimal bottleneck size was selected on the basis of the average PCC across the three 'validation 1' subsets, ensuring the generalization across different data partitions. From the three models with this optimal bottleneck size, we then selected the CELTIC model that performed best on the 'validation 2' subsets, along with the U-Net model corresponding to the same random seed. This process resulted in a total of 12 models, a matched CELTIC and U-Net models for each one of six organelles. All training was conducted on an NVIDIA RTX 2080 GPU using PyTorch. The hyperparameters used for training included a learning rate of 0.001, betas of (0.5, 0.999) and a batch size of 24.

These six pairs of matched organelle-specific CELTIC-versus-U-Net models were evaluated using the test set. For each cell, we calculated the in silico labeling PCC for CELTIC versus the matched U-Net prediction. We evaluated the contribution of context via CELTIC for five cell populations: one population containing all cells ('all'), and four rare populations (mitosis, edge, small volume and sparse neighborhood density). For each population, we calculated the $\Delta\text{PCC} = \text{CELTIC}_{\text{PCC}} - \text{U-Net}_{\text{PCC}}$ and the corresponding Wilcoxon signed-rank test to reject the null hypothesis that the ΔPCC values are distributed around zero. The *P* values, population sizes and distribution of PCC values across organelles and rare populations are detailed in Supplementary Table 3.

Ablation study

To systematically assess the contribution of context to the in silico labeling of rare cell populations, we performed bootstrapping analysis by permuting the context between cells and evaluating the resulting deterioration in the in silico labeling performance. Specifically, for each organelle, context type and rare population, we shuffled the values of the context type across all cells such that each cell was randomly assigned a context value from another cell in the population. We then used the pretrained model to generate predictions on the test set images with permuted context and calculated the mean PCC across shuffles. This process was repeated ten times for each representation using different random seeds. The results, reported in Supplementary Table 4 and in Supplementary Fig. 6, were compared with the mean PCC obtained with true (unshuffled) context. Statistical significance was assessed using paired *t*-tests between the true and shuffled contexts.

Sensitivity to 'real world' inference of context

The success of CELTIC depends on the accurate incorporation of contextual information to guide its predictions. Although our study benefited from the high-quality annotations in the AICS dataset—including mitotic state, cell location and segmentation—real-world applications would require the computational extraction of these contexts from raw images. This process is prone to noise, which could degrade the model performance and impact downstream analyses. To assess the effect of such errors on CELTIC's predictions, we focused on the context using the CELTIC microtubules model and measured the effect of mitotic state context errors introduced by the imperfect computational inference from the raw label-free images. Namely, by gradually simulating erroneous mitotic stage context at inference and by computationally inferring the mitotic stage as described next.

Gradually introducing errors to the mitotic stage context at inference

We systematically introduced noise to the mitotic state labels of the test set samples by simulating their alteration between similar morphological mitotic states that are probable in human annotations as well as in computational inferences. Specifically, we grouped the six original AICS stages into three pairs on the basis of their similarity in microtubule morphology: M0 with M1M2, M3 with M4M5 and M6M7_single with M6M7_complete. For a given noise ratio, that fraction of samples had their contextual label swapped with its paired stage, mimicking consistent errors that could occur owing to similar visual appearance. We incrementally increased the proportion of swapped labels from 5% (noise ratio of 0.05) to 100% (noise ratio of 1). For each noise level, we measured the resulting PCC of the in silico labeling microtubules model's predictions. These results are reported in Supplemenary Fig. 7a.

Computational inference of the mitotic stage

We trained a multiclass classifier to predict the mitotic stage context using the in silico-labeled microtubule images generated by the U-Net model, using brightfield images from the test sets of five organelles (excluding microtubules). All 3D images were resized to $32 \times 64 \times 64$ voxels using trilinear interpolation to ensure consistent input dimensions. The classifier was supervised using the original AICS mitotic stage annotations. To ensure class balance, we retained all mitotic samples and randomly sampled a matched number of interphase (M0) cells. The final dataset included 720 single cells images split into stratified training (70%), validation (20%) and test (10%) subsets. A 3D convolutional neural network was trained using PyTorch with the following structure: four to five stacked convolutional blocks, each consisting of two 3D convolution layers, batch normalization, ReLU activations and a down sampling layer. Fully connected layers followed the final feature map, with optional dropout applied. The classifier was trained using the cross-entropy loss function and optimized with Adam. A grid search was performed over combinations of batch size (8–128), learning rate (1×10^{-6} to 1×10^{-3}), dropout rate (0–0.5) and network depth (4 or 5), each run for 100 epochs. The best-performing model, selected by validation F1 score, used a batch size of 16, learning rate of 5×10^{-5}, a dropout rate of 0.4 and a depth of 5. The evaluation was conducted on the held-out test set using the macro-averaged F1 score, both across the six fine-grained mitotic classes and across three grouped-stage categories corresponding to the visually confusable pairs defined above ('Gradually introducing errors to the mitotic stage context at inference' section). The classifier accurately recovered the mitotic stage in most cases, with errors typically involving confusion between adjacent stages along the mitotic progression (Supplemenary Fig. 7b). The classifier achieved an F1 score of 0.53 for the six-stage classification task and 0.807 for the grouped-stage classification task. We then retrained the CELTIC microtubules model from scratch, using the same training and test sets as in the original experiment but replacing the AICS-based stage context with the noisy classifier-predicted mitotic stage labels. These results are reported in Supplemenary Fig. 7c.

Application-appropriate assessment of the predicted spindle axis

As an application-appropriate assessment, we evaluated the ability to determine the location and orientation of the spindle axis during mitosis from brightfield images. We used all 27 cells in the mitotic prometaphase–metaphase stage from the microtubule test set data. We performed the in silico labeling of microtubules using the corresponding CELTIC and U-Net model. To segment these in silico predictions and calculate the spindle axis, we devised the following pipeline. First, resize the middle z slice of the 3D images to 128 pixels × 128 pixels using standard resizing techniques available in OpenCV (cv2.resize). Second, erode the microtubules prediction according to the cell's segmentation mask with a kernel size of 5 pixels to remove any residual predictions on the cell's border. Third, threshold the microtubules prediction according to the 90th pixel intensity percentile. Fourth, to define the spindle axis, we selected the two largest connected regions in the segmented image, calculated their moments and their center of mass and defined the spindle axis as the line connecting the two centers of mass. In cells with a single region, we defined the spindle axis according to the longest chord of the region, that is, according to the longest line that could be accommodated within this region. To define the ground truth spindle axis, we manually annotated the 27 cells. The location error (ΔC) was calculated as the distance between the predicted and the ground truth spindle axes. The orientation error ($\Delta\theta$) was calculated as the cosine of the angle between the predicted and the ground truth spindle axes. Statistical significance was determined with a permutation test, where we kept the ground truth spindle orientations fixed while randomly reassigning the predicted orientations among different cells. The P value represents the fraction of 250,000 permutations that achieved errors smaller than or equal to the unpermuted predictions.

Using the organelle label as context toward a unified multi-organelle in silico labeling model

We trained a single CELTIC model to predict each of the organelles according to the context—a one-hot encoding of the organelle (Supplemenary Fig. 8). This model was trained with the same FOVs used to replicate the results of Ounkomol et al. (Supplemenary Tables 1 and 5a), where the train–validation–test splits for each organelle were pooled together to form a single dataset. The organelle context was encoded as a six-dimensional one-hot vector. The model architecture mirrored that of the CELTIC model described previously, with a DAFT bottleneck factor of four. Given the sixfold increase in data volume, we trained the model for sixfold iterations used for training single-organelle models (Supplemenary Table 5b). In a second experiment, we trained the same model with one sixth of the training iterations to match the number of iterations used to train a single-organelle model (Supplemenary Table 5c).

Using context for image generation

To generate how an organelle may change for the same cell in a different context, we used CELTIC to predict the in silico labeling of the same cell with different context inputs. For example, to generate the images in Fig. 5a, we in silico-labeled an interphase cell with the actin filaments, nuclear envelope and microtubules CELTIC models. The top image is a result of the prediction with the native interphase context. For the bottom result, we manipulated the mitotic stage representation in the context vector from interphase (mitotic context representation of '100000') to prometaphase–metaphase ('000100'). To quantify how altering context affects the in silico labeling, we used 230 cells from 16 unseen FOVs. For each cell and for each organelle, we used CELTIC with the appropriate context and then repeated the inference by manipulating the context vector. We measured the PCC between these context-altered in silico-labeled predictions.

Training a mitosis classifier from a dataset of interphase cells

We composed a dataset of 4,080 brightfield images of interphase cells, taken in full from both the development and test sets of the endoplasmic reticulum and Golgi datasets. We generated in silico labeling for each image using the CELTIC microtubules model in two passes: first, using the original context, corresponding to cell stage M0, and second, by modifying only the cell stage context from M0 to M4M5 (Fig. 5c). This process yielded a balanced in silico labeling dataset of 8,160 images. We split the data into training (70%), validation (20%) and test (10%) subsets.

We then trained a binary classifier on these images to predict the mitotic stage (Fig. 5c). All 3D images were resized to $32 \times 64 \times 64$ voxels using trilinear interpolation. We used the same 3D convolutional neural network architecture as described in the 'Computational inference of the mitotic stage' section. The model was trained using the binary cross-entropy loss function and optimized with Adam. A grid search was conducted over hyperparameter combinations similar to those described in that section, with the batch size fixed at eight. Each configuration was trained for 20 epochs. The best-performing model, selected on the basis of validation F1 score, used a learning rate of 1×10^{-4}, a dropout rate of 0.5 and a depth of 4. A final evaluation was conducted on the held-out test set using the F1 score, yielding a score of 0.994.

The binary classifier was then used to predict the mitotic stage context in images generated by the microtubules U-Net model (a model without context) (Fig. 5d). These images were taken from the microtubules test set and included 1,244 images labeled as stage M0 and 28 images as stage M4M5. The classifier achieved a macro-averaged F1 score of 0.870 and an AUC of 0.928 (Fig. 5e).

Computational efficiency and preprocessing overhead

To quantify the computational demands of CELTIC, we compared training, inference and preprocessing times with a baseline U-Net model without context. For CELTIC, we report results using a DAFT bottleneck size of 48, the most commonly used configuration. Training CELTIC required 0.427 s per iteration (batch size of 24 patch samples) compared with 0.418 s for the baseline model. The average inference time per single cell required 0.285 s for CELTIC and 0.265 s for the baseline. Preprocessing was performed, one time per cell, before the training and inference. Masking, normalization and saving the image to the disk required 0.767 s per cell. Computing the classic shape feature using the *k*-means classifier required 0.06 ms per cell, whereas computing the machine learning shape feature—which involves autoencoder inference, PCA transformation and *k*-means assignment—required 6.88 ms per cell. Calculating the neighborhood density required 0.26 s per cell. All measurements were performed on a compute node equipped with an x86_64 CPU, 503 GB RAM and an NVIDIA GeForce RTX 3090 GPU.

Reporting summary

Further information on research design is available in the Nature Portfolio Reporting Summary linked to this article.

Data availability

Single-cell and processed data are available from the BioImage Archive[55,56] at ref. 57. The single-cell data consists of six datasets, one per organelle used in the paper. Each organelle dataset contains between 2,052 and 2,993 3D cell images cropped from 180 FOVs from the Allen Institute WTC-11 dataset[28]. Each cell is represented by three aligned images: brightfield, the EGFP-tagged organelle and a segmentation mask. Metadata for each cell include the cell cycle stage and the edge flag annotations from the WTC-11 dataset, along with the additional computed features of neighbor counts and cell shape descriptors.

Code availability

CELTIC source code is available via GitHub at https://github.com/zaritskylab/CELTIC.

References

49. Olteanu, M., Rossi, F. & Yger, F. Meta-survey on outlier and anomaly detection. *Neurocomputing* **555**, 126634 (2023).
50. Carreño, A., Inza, I. & Lozano, J. A. Analyzing rare event, anomaly, novelty and outlier detection terms under the supervised classification framework. *Artif. Intell. Rev.* **53**, 3575–3594 (2019).
51. Heald, R. & Cohen-Fix, O. Morphology and function of membrane-bound organelles. *Curr. Opin. Cell Biol.* **26**, 79–86 (2014).
52. Rafelski, S. M. et al. Mitochondrial network size scaling in budding yeast. *Science* **338**, 822–824 (2012).
53. Perez, E., Strub, F., de Vries, H., Dumoulin, V. & Courville, A. FiLM: visual reasoning with a general conditioning layer. In *Proc. AAAI Conference on Artificial Intelligence* Vol. 32 (2018).
54. Pölsterl, S., Wolf, T. N. & Wachinger, C. Combining 3D image and tabular data via the dynamic affine feature map transform. In *International Conference on Medical Image Computing and Computer Assisted Intervention (MICCAI)* 688–698 (2021).
55. Sarkans, U. et al. The BioStudies database—one stop shop for all data supporting a life sciences study. *Nucleic Acids Res.* **46**, D1266–D1270 (2018).
56. Hartley, M. et al. The bioimage archive—building a home for life-sciences microscopy data. *J. Mol. Biol.* **434**, 167505 (2022).
57. Elmalam N, Zaritsky A. Cell-context dependent in silico organelle localization in label-free microscopy images. *BioStudies* https://doi.org/10.6019/S-BIAD2156 (2025).

Acknowledgements

This research was supported by the Israel Science Foundation (ISF, grant no. 2516/21 to A.Z), by the Israeli Council for Higher Education (CHE) via the Data Science Research Center, Ben-Gurion University of the Negev, Israel (to A.Z.) and by an Allen Distinguished Investigator Award, a Paul G. Allen Frontiers Group advised grant of Allen Family Philanthropies (to A.Z.). We thank M. Viana, N. Gaudreault, J. Chen, S. Rafelski, M. Feldman and N. Elia for discussions, suggestions and critical reading of the paper.

Author contributions

A.Z. and N.E. conceived of the study. N.E. developed the computational method and analyzed the data. N.E. and A.Z. interpreted the data and drafted and edited the paper and approved its content. A.Z. mentored N.E.

Competing interests

N.E. and A.Z. are inventors on a US Provisional Application, number 63/718,647, filed by B. G. Negev Technologies and Applications, which covers the methods described in this paper.

Additional information

Supplementary information The online version contains supplementary material available at https://doi.org/10.1038/s41592-025-02960-4.

Correspondence and requests for materials should be addressed to Assaf Zaritsky.

Peer review information *Nature Methods* thanks Sharmishtaa Seshamani and the other, anonymous, reviewer(s) for their contribution to the peer review of this work. Peer reviewer reports are available. Primary Handling Editors: Rita Strack and Madhura Mukhopadhyay, in collaboration with the *Nature Methods* team.

nature portfolio

Corresponding author(s): Assaf Zaritsky

Last updated by author(s): Oct 15, 2025

Reporting Summary

Nature Portfolio wishes to improve the reproducibility of the work that we publish. This form provides structure for consistency and transparency in reporting. For further information on Nature Portfolio policies, see our Editorial Policies and the Editorial Policy Checklist.

Statistics

For all statistical analyses, confirm that the following items are present in the figure legend, table legend, main text, or Methods section.

n/a	Confirmed
☐	☒ The exact sample size (*n*) for each experimental group/condition, given as a discrete number and unit of measurement
☐	☒ A statement on whether measurements were taken from distinct samples or whether the same sample was measured repeatedly
☐	☒ The statistical test(s) used AND whether they are one- or two-sided *Only common tests should be described solely by name; describe more complex techniques in the Methods section.*
☐	☒ A description of all covariates tested
☐	☒ A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons
☐	☒ A full description of the statistical parameters including central tendency (e.g. means) or other basic estimates (e.g. regression coefficient) AND variation (e.g. standard deviation) or associated estimates of uncertainty (e.g. confidence intervals)
☐	☒ For null hypothesis testing, the test statistic (e.g. *F*, *t*, *r*) with confidence intervals, effect sizes, degrees of freedom and *P* value noted *Give P values as exact values whenever suitable.*
☒	☐ For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings
☒	☐ For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes
☐	☒ Estimates of effect sizes (e.g. Cohen's *d*, Pearson's *r*), indicating how they were calculated

Our web collection on statistics for biologists contains articles on many of the points above.

Software and code

Policy information about availability of computer code

Data collection	No software was used for data collection.
Data analysis	Software for analysis can be found at https://github.com/zaritskylab/CELTIC .

For manuscripts utilizing custom algorithms or software that are central to the research but not yet described in published literature, software must be made available to editors and reviewers. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Portfolio guidelines for submitting code & software for further information.

Data

Policy information about availability of data

All manuscripts must include a data availability statement. This statement should provide the following information, where applicable:

- Accession codes, unique identifiers, or web links for publicly available datasets
- A description of any restrictions on data availability
- For clinical datasets or third party data, please ensure that the statement adheres to our policy

Single cell and processed data are available from the BioImage Archive55,56 at 57. The single cell data consists of six datasets, one per organelle used in the manuscript. Each organelle dataset contains between 2052 to 2993 3D cell images cropped from 180 fields of view from the Allen Institute WTC-11 dataset28. Each cell is represented by three aligned images: brightfield, the EGFP-tagged organelle, and a segmentation mask. Metadata for each cell includes the cell cycle stage and the edge flag annotations from the WTC-11 dataset, along with the additional computed features of neighbor counts and cell shape descriptors.

Human research participants

Policy information about studies involving human research participants and Sex and Gender in Research.

Reporting on sex and gender	N/A
Population characteristics	N/A
Recruitment	N/A
Ethics oversight	N/A

Note that full information on the approval of the study protocol must also be provided in the manuscript.

Field-specific reporting

Please select the one below that is the best fit for your research. If you are not sure, read the appropriate sections before making your selection.

☒ Life sciences ☐ Behavioural & social sciences ☐ Ecological, evolutionary & environmental sciences

For a reference copy of the document with all sections, see nature.com/documents/nr-reporting-summary-flat.pdf

Life sciences study design

All studies must disclose on these points even when the disclosure is negative.

Sample size	For each organelle, the test sample size was determined to include 100 fields of view (FOVs), which were randomly chosen from the pool of FOVs in the AICS dataset. These FOVs closely matched the characteristics of those presented in the original paper by Ounkomol et al., in terms of cellular properties such as cell count and mitotic distribution. The total number of cells in these FOVs typically ranged from 1,100 to 1,600, ensuring that rare populations could achieve a minimum size of n>10 to allow significance and to adequately represent rare cell populations within the dataset.
Data exclusions	After randomly selecting the FOVs per organelle as stated above, no data were excluded.
Replication	The replication and reproducibility of the results are supported by the large amount of data used in the analysis—100 FOVs per organelle, representing a dataset substantially larger than the entire training set. The tested data were acquired from a mix of imaging days and are not related to the training data, ensuring independent validation of the results. Each result was replicated at least 3 times independently.
Randomization	The development set was chosen following the original set published by Ounkomol et al., while the test set was randomly selected as described above. The organelles were chosen from the list provided in the AICS dataset, focusing on six organelles that span the range of in silico labeling performances reported in Ounkomol et al.
Blinding	Blinding was not relevant because all sample allocations were computationally randomly selected.

Reporting for specific materials, systems and methods

We require information from authors about some types of materials, experimental systems and methods used in many studies. Here, indicate whether each material, system or method listed is relevant to your study. If you are not sure if a list item applies to your research, read the appropriate section before selecting a response.

Materials & experimental systems

n/a	Involved in the study
☒	☐ Antibodies
☒	☐ Eukaryotic cell lines
☒	☐ Palaeontology and archaeology
☒	☐ Animals and other organisms
☒	☐ Clinical data
☒	☐ Dual use research of concern

Methods

n/a	Involved in the study
☒	☐ ChIP-seq
☒	☐ Flow cytometry
☒	☐ MRI-based neuroimaging