ISCI, Volume *13*

Supplemental Information

eDetect: A Fast Error Detection and Correction Tool for Live Cell Imaging Data Analysis

Hongqing Han, Guoyu Wu, Yuchao Li, and Zhike Zi

Supplemental Figures and Legends

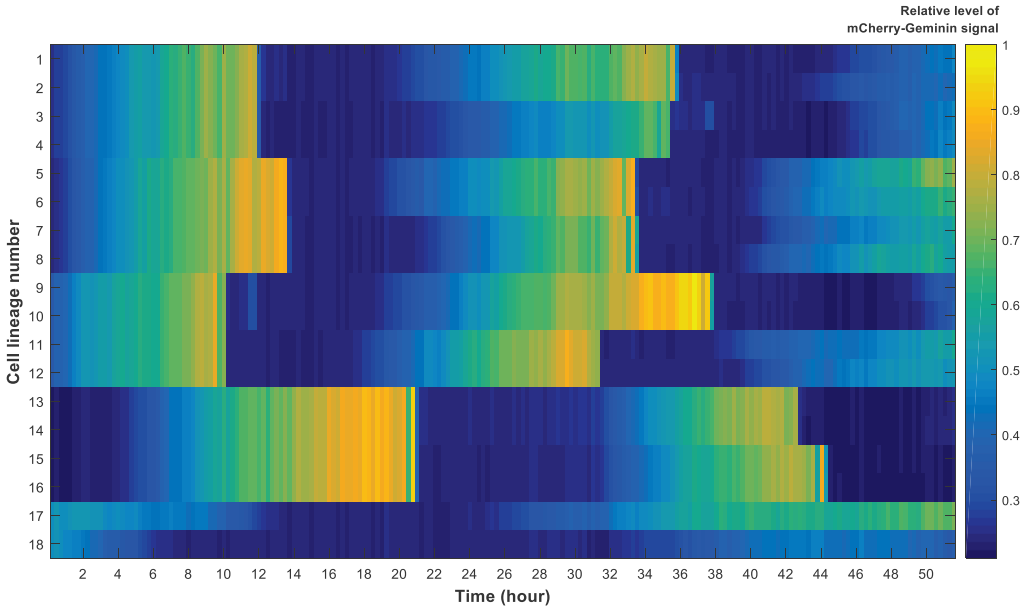

Figure S1: The dynamics of mCherry-Geminin protein in HaCaT single cells, which were quantified with eDetect. Related to Figure 2.

Supplementary Tables

Table S1. Names and IDs of shape and additional features, Related to Figure 1.

ID in Segmentation Gating and Cell Pair Gating	Feature Category	Feature Name
v{1}	shape[1]	Area
v{2}	shape	Eccentricity
v{3}	shape	Orientation
v{4}	shape	Solidity
v{5}	shape	Perimeter
v{6}	shape	MeanIntensity
v{7}	additional[2]	radial distribution

[1] The meanings of shape features can be found at:

https://www.mathworks.com/help/images/ref/regionprops.html

[2] The additional feature "radial distribution" describes how intensity varies as a function of distance from the nearest point in the border of the object.

Transparent Methods

Cell segmentation

eDetect uses an adaptive thresholding approach to segment cell nuclei from nuclear marker images (Bradley and Roth, 2007). To de-clump the clustered nuclei, we apply watershed transformation (Meyer, 1994) on a filtered distance map of nuclei contours (Schneider et al., 2012). Because single nuclei sometimes are wrongly split into smaller objects, we adopt a hierarchical procedure to merge directly adjacent objects based on the areas of minimum enclosing ellipses (Moshtagh).

Feature extraction

eDetect extracts cellular features from nuclei contours and nuclear marker images. It provides the options of computing 5 categories of features, i.e. shape, intensity, Haralick features (Haralick and Shanmugam, 1973), Zernike features (Boland et al., 1998) and additional features. By default, eDetect calculates shape and additional features, which are listed in Table S1. The calculations of intensity, Haralick and Zernike features are implemented using the source code of CellProfiler (Carpenter et al., 2006).

Cell tracking

eDetect adopts a constrained nearest-neighbor strategy for cell tracking. For each object (except the objects in the first frame), eDetect calculates its distances to all the objects in the previous frame. The closest object in the previous frame is the predecessor, unless this shortest distance is still larger than a predefined threshold.

Sometimes, consecutive frames are not perfectly aligned to each other in the x-y plane because of paraxial shifts of the field. We tackle this problem using an optimization-based approach. For each frame, before calculating distances, eDetect subtracts a "shift vector" from the coordinates of all objects in the current frame. Therefore, the assignment of predecessors depends on the value of this "shift vector", and the sum of the aforementioned shortest distances is a function of the "shift vector". eDetect searches for the "shift vector" (within a square whose size is user-defined) that minimizes this function. The assignment that is associated to this minimal is the tracking result.

Measurements

In eDetect, the following measurements are calculated for different fluorescent channels (the channels for signal extraction): nuclei median and mean intensities, cytoplasm median and mean intensities, and nuclear-cytoplasmic ratios of median and mean intensities.

To approximate nuclei regions, we shrink the nuclear segmentations by a user-defined distance to exclude nuclear membrane pixels. The cytoplasm regions are approximated using peri-nuclear ring regions, whose inner and outer contours are expanded from nuclear segmentations. In order to compensate medium illumination, background intensity is subtracted from both nuclei and cytoplasm intensities.

To generate the foreground region, we expand the nuclear segmentations by a large distance that is the maximum of the following two values: (1) the largest object diameter and (2) twice of the distance by which nuclear segmentations are expanded to generate the outer contour of the peri-nuclear ring regions. Afterwards, we derive the background region by removing the foreground region from the whole image.

In eDetect, to shrink an object by a distance of n, we remove the contour pixels of the object n times iteratively. Similarly, to expand an object by a distance of n, we add a layer of pixels to the exterior of the object n times iteratively. We implemented both shrinkage and expansion operations with an algorithm of morphological operations on binary images in Matlab.

Segmentation gating and cell pair gating

eDetect segmentation gating and cell pair gating employ principal component analysis (PCA) to project high dimensional cellular features onto a 2D plane, where the data is visualized with scatter plots. PCA converts a data table, in which the observations are described by a set of possibly correlated variables, into a new one, where the new variables (*i.e.:* principal components) are uncorrelated (Abdi and Williams, 2010). From the first principal component (PC) to the last one, each PC maximizes the data variance under the constraint that it is orthogonal to all the preceding ones. In eDetect, PC1 and PC2 are used as the x- and y-axis variables in the scatter plots. The meanings of imaging features for the PCA input are listed in Table S1.

In cell pair gating, apart from the variables specified in the formula, the input to PCA additionally includes 3 default variables: the 3 sharp angles between 3 lines. The 3 lines are: the major axes of the 2 objects and the line that passes through the centroids of the 2 objects.

Outlier detection in cell lineages display

In eDetect, abrupt changes of cellular measurements are detected based on the median absolute deviation (Leys et al., 2013). eDetect calculates relative difference of measurements between each cell and its predecessor. Among all the relative differences, those whose distances to the median value are larger than a certain predefined number of standard deviations are defined as outliers.

The calculation of performance scores

Segmentation accuracy (SEG), tracking accuracy (TRA) and complete tracks scores (CT) are calculated using the evaluation software packages provided in the Cell Tracking Challenge website. Recall of complete lineages (RCL) is calculated with a customized MATLAB script.

Complete lineages are lineages that start in the first frame and end in the last frame of the entire video. The formula of recall (or sensitivity) is given below (TP: true positive, FP: false positive, FN: false negative.):

$$Recall = \frac{TP}{TP + FN}$$

The ground truth of HaCaT-FUCCI dataset was produced by manually correcting the result of eDetect automatic analysis until the tracks were correct and segmentations were accurate under human inspection.

Supplemental References

Abdi, H., and Williams, L.J. (2010). Principal component analysis. Wiley interdisciplinary reviews: computational statistics *2*, 433-459.

Boland, M.V., Markey, M.K., and Murphy, R.F. (1998). Automated recognition of patterns characteristic of subcellular structures in fluorescence microscopy images. Cytometry *33*, 366-375.

Bradley, D., and Roth, G. (2007). Adaptive thresholding using the integral image. Journal of graphics tools *12*, 13-21.

Carpenter, A.E., Jones, T.R., Lamprecht, M.R., Clarke, C., Kang, I.H., Friman, O., Guertin, D.A., Chang, J.H., Lindquist, R.A., Moffat, J.*, et al.* (2006). CellProfiler: image analysis software for identifying and quantifying cell phenotypes. Genome Biol *7*, R100.

Haralick, R.M., and Shanmugam, K. (1973). Textural features for image classification. IEEE Transactions on systems, man, and cybernetics, 610-621.

Leys, C., Ley, C., Klein, O., Bernard, P., and Licata, L. (2013). Detecting outliers: Do not use standard deviation around the mean, use absolute deviation around the median. Journal of Experimental Social Psychology *49*, 764-766.

Meyer, F. (1994). Topographic distance and watershed lines. Signal processing *38*, 113-125.

Moshtagh, N. Minimum volume enclosing ellipsoid.

Schneider, C.A., Rasband, W.S., and Eliceiri, K.W. (2012). NIH Image to ImageJ: 25 years of image analysis. Nature methods *9*, 671.