SEGMENTATION OF CYTOPLASMS OF CULTURED CELLS

Amin Allalou and Carolina Wählby

Centre for Image Analysis
Uppsala University
Sweden

F. M. van de Rijke, R. Jahangir Tafrechi and A. K. Raap

Department of Molecular Cell Biology
Leiden University Medical Center
the Netherlands

ABSTRACT

Cell cultures as well as cells in tissue always display a certain degree of variability, and measurements based on cell averages will miss important information contained in a heterogeneous population. This paper presents automated methods for segmentation of cells and cytoplasms. The segmentation results are applied to image based measurements of mitochondrial DNA (mtDNA) mutations in individual cells. Three different methods for segmentation of cytoplasms are compared and it is shown that automated cytoplasmic delineation can be performed 30 times faster than manual delineation, with an accuracy as high as 87%, compared to an inter observer variability of 79% at manual delineation.

1. INTRODUCTION

Great improvements in microscopy hardware have made it possible to produce thousands of high resolution cell images in a short period of time. It has led to a great demand for high-throughput automated cell image analysis. Image based analysis of microscopy data and cell segmentation is not new. The interest for high- throughput image based techniques is however growing fast. An image of a cell culture often contains different cells that all possess different characteristics. Taking the average over such an image will not reveal the biologically important differences and variations between the cells. Single cell analysis is clearly the only option to observe the dissimilarities between the cells. To be able to assign a signal to a particular cell it has to be delineated. A common approach is to use a cytoplasm staining as a guide in delineation of the cell at its cytoplasmic borders [5, 10]. Another approach is to use a membrane stain, which binds to the cytoplasmic surface. In combination with nuclear staining individual cells can then be delineated by a combination

of gradient curvature flow techniques and seeded watershed segmentation [6]. In many studies a blue stain is used for the nucleus, and red and green stain for molecular detection. Due to fluorescence spectral overlap this may limit the possibility of using a unique color for a stain that helps segmenting the cytoplasm. Also, a cytoplasmic stain may not be compatible with the molecular stain of interest.

Mitochondrial DNA (mtDNA) is a small genome, present in 100s to 1000s of copies in the cytoplasm of each mammalian cell. The genetic information contained in the human mtDNA is essential for a major energy-generating process of the cell called oxidative phosphorylation. All DNA mutates and so does mtDNA. When the mutation is pathogenic it needs to accumulate to relative large amounts (>80% of all mtDNAs) for the cells energy provision to become so subverted that cell functions are lost and cells die. Such mutation accumulation leads to devastating diseases if the mutation is inherited from the mother or to normal aging phenomena if it is acquired somatically. A major factor in determining cellular mutation loads is the process of mitotic segregation. To understand mtDNA segregation and with it mtDNA mutation accumulation, our research focuses on mtDNA segregation patterns in *in vitro* cultured cells. Experimentally, this requires the determination of the mutation load in hundreds of individual cells in multiple serial cell culture passages of a cloned heteroplasmic founder cell (i.e., a single cell carrying mutant and wildtype-mtDNA molecules). *In situ* genotyping mtDNA with the padlock/rolling circle method [4] provides an elegant approach for detection of mtDNA sequences variants at the (sub-) cellular level. However, in our experience thusfar no cytoplasmic or membrane staining proved compatible with the padlock/rolling circle method. One way to approximate the outline of the cytoplasm is using a fixed radius for each cell [2]. A fixed radius may not always be the best choice since cells are often not spherical. Thus, to analyze thousands of cells, the challenge is to develop for this application a cell segmentation in absence of a cytoplasmic stain. Here we describe development of such an automated image cytometric procedure for fully automated measurements of mtDNA mutation loads of single cells. Preliminary results show that it greatly facilitates the determination of the mutant
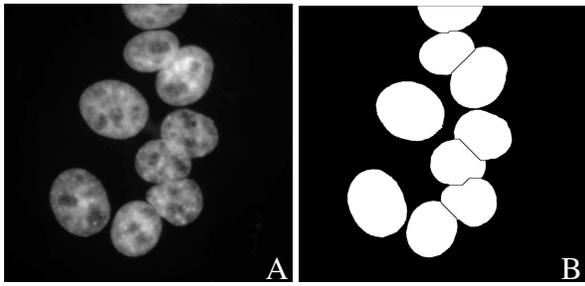
Fig. 1. **A:** Nuclear stain. **B:** Final segmentation result.

Fig. 2. **A:** Result of cytoplasmic segmentation not making use of the cytoplasmic stain (*NCS*). **B:** Result when cytoplasmic stain is included (*CS*). **C:** Result of manual delineation (*O*).

mtDNA fraction of heteroplasmic cells stained for the wild-type and mutant locus at position 3243 of human mtDNA.

2. MATERIALS AND METHODS

2.1. Cell preparation and image acquisition

The nuclei are stained with DAPI (blue). Padlock probes for mutated mtDNA are detected with Cy5 stain (red) and padlock probes for wild-type DNA are detected with FITC stain (green). For visualization of the cytoplasm tubulin is detected with mouse anti-tubulin antibodies and two different secondary antibodies, rabbit anti-mouse FITC (green) and goat anti-mouse Alexa 594 (red). Cytoplasmic stains can not be used together with padlock probes detected with the same color due to spectral overlap. Images are acquired using an epiflourescent microscope (Leica, Leica Microsystems GmbH, Wetzlar, Germany) equipped with a cooled monochrome CCD camera (Quantix, Photometrix, Melbourne, Australia).

2.2. Delineation of nuclei

The cell segmentation is initiated by a segmentation of the image channel representing the nuclear stain (Fig. 1A). Otsu's method of thresholding, which minimizes the variance of the foreground and the background, separates the nuclei from the background [7]. The binary image representing the nuclei is transformed to a landscape-like image using distance transformation. The distance image is produced using the 5-7-11-chamfer distance transform on the binary image [1]. The chamfer distance transform was preferred over the Euclidian distance transform due to lower computational cost and yet sufficient result. Clustered nuclei are thereafter separated by watershed segmentation [3].

Seeds that represent the different nuclei are needed in order to separate clustered nuclei into different objects. Due to imperfect circularity of the nuclei distance transform may lead to multiple seeding points or local maxima, for the same nucleus. This will result in over-segmentation. The h-maxima transform is used to suppress maxima whose depth is smaller than a given threshold t[8]. The value t is directly propor-

tional to the radius of an object and can therefore also be used to remove objects that are too small to be true cell nuclei, and thus have a radius less than a specified value. The result of the watershed segmentation is shown in Fig. 1B.
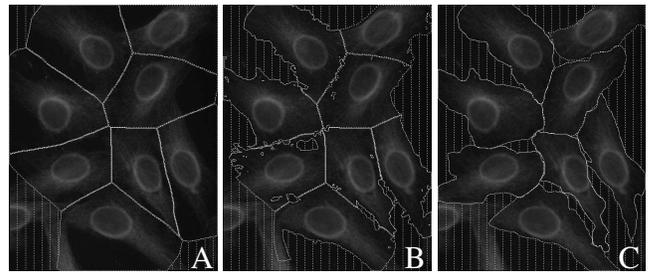
2.3. Delineation of cytoplasms

Once the nuclei have been delineated, a cytoplasm has to be defined for each detected nucelus.

Approach 1: no cytoplasmic stain (*NCS*). If no cytoplasm staining is present the delineation of the cytoplasm is purely based on a fixed distances from the border of the nucleus. A distance transform is applied to the background of the binary image of the nuclei. This results in an image that represents the distance to the nearest nuclei for each pixel. A user defined threshold, corresponding to the maximum radius of the cytoplasm, is applied to the distance transformed background. A watershed is again used to define the borders of the objects, see result in Fig. 2A.

Approach 2: with cytoplasmic stain (*CS*). The second approach to delineation of the cytoplasm makes use of a cytoplasm staining (tubulin stain). Tubulin is present throughout the whole cytoplasm and can therefore be used as a marker for the cytoplasm. A variance filter is applied to the channel representing the tubulin and areas of high intensity variation (tubulin areas) are enhanced. Thereafter, an average filter is applied to smooth the variance. The smoothed image is thresholded by Otsu's method, but to include all of the cytoplasm the threshold is adjusted by multiplying it with 0.25. This may be avoided by using a different thresholding method. A watershed transformation seeded by the nuclei and restricted to the binary image of the cytoplasm is there after applied, see result in Fig. 2B.

Approach 3: manual delineation (*O*). The third method for delineation of the cytoplasm is a manual segmentation. Here, two observers used the software Visiopharm Integrator System (VIS, Visiopharm, Hørsholm, Denmark) to outline the tubulin stained images manually.

2.4. Localization of padlock signals

The image channels containing the padlock signals representing wild-type or mutant mt DNA are filtered with a 5x5 kernel that enhances areas of local maxima. The signals are separated from the image background by a user defined threshold set to a default value that localizes the major proportion of the signals. The same threshold was used for all images, and in order to evaluate the influence of the threshold on the final measure of mutation load the thresholds were increased or decrease by 20%. The variation caused by these changes is shown as error bars in Fig. 3.

2.5. Comparison of segmentation methods

In the evaluation of the different methods for cytoplasm segmentation, accuracy (agreement with truth), precision (reproducibility) and efficiency (time) are considered, based on the ideas by Udupa et. al. [9]. Define S as being the result of the segmentation method being compared to S_t, the true segmentation. The accuracy makes use of three definitions; False Negative Area Fraction (FNAF), False Positive Area Fraction (FPAF) and True Positive Area Fraction (TPAF). FNAF is the fraction of S_t that was missed by S. FPAF denotes the area that is falsely identified by S as a fraction S_t. In the current case, the parts of the S that overlap with the image background, as defined by S_t, are not counted as falsely identified because the background does not give rise to any signals and will not affect the calculation of signals per cell. TPAF describes the total amount of cytoplasm defined by S that coincides with S_t as a fraction of S_t.

Precision is the ability to reproduce the same result. Naturally, a fully automated method will always reproduce the same result when applied to the same image. To observe true precision, the cells should be imaged repeatedly. This was however not done. At manual delineation, the result will most likely not be fully reproducible, we will have inter- and intra-observer variation.

Two factors must be considered when comparing the efficiency of a segmentation method; the computational time and the human operator time required to complete the segmentation.

3. RESULTS

3.1. Comparison of segmentation methods

For a full comparison of segmentation methods, accuracy, precision, and efficiency should be considered. The comparative study of methods for cytoplasm segmentation was performed on 9 images containing a total of 56 cells. Two fully automated image based segmentation methods, one using information from a cytoplasmic stain (referred to *CS*), and which does not make use of a cytoplasmic stain (referred to

Table 1. Comparison of Accuracy

Method	vs.	Accuracy		
		TPAF	FNAF	FPAF
NCS	O_{1a}	0.87 ± 0.03	0.14 ± 0.03	0.12 ± 0.04
CS	O_{1a}	0.85 ± 0.03	0.16 ± 0.03	0.11 ± 0.03
O_2	O_{1a}	0.84 ± 0.02	0.16 ± 0.02	0.02 ± 0.01
O_{1b}	O_{1a}	0.90 ± 0.02	0.10 ± 0.02	0.03 ± 0.01

Table 2. Comparison of Precision and Efficiency

Method	vs.	Precision (%)	Efficiency Cells/min
NCS	O_{1a}	100	30
CS	O_{1a}	100	30
O_2	O_{1a}	79	1
O_{1b}	O_{1a}	84	1

as *NCS*) were compared to each other and to manual segmentation (referred to as *O*) of the same cytoplasms. Both automated methods are seeded by the same image of the cell nuclei, and the same threshold t for the h-maxima transform (the only input parameter) was used in all images. As no gold standard or ground truth is possible to produce, it is assumed that the manual segmentation method (*O*) results in the true delineation, defined as S_t^O. Manual segmentation was performed three times by two different persons to provide measurements of precision (reproducibility) in terms of inter- and intra- observer variability (refered to as O_{1a}, O_{1b} and O_2). The results can be seen in Table 1 and 2.

First of all, considering the accuracy, *NCS* and *CS* is significantly (α=0.05) less accurate than *O*. Between *CS* and *NCS* no significant (α= 0.05) difference can be seen in terms of accuracy. Furthermore, method *O* has noticeably lower precision than the other methods, as the computer based methods will reproduce the same result if re-run on the same image data, i.e., 100% precision, while manual segmentation varies both between observers (inter-observer precision is 79%) and for the same observer assessing the data at different times (intra-observer precision is 84%). Finally, the efficiency of *NCS* and *CS* is approximately 30 times higher than that of the manual segmentation *O* when using a 2.53 GHz Intel Pentium 4 processor.

3.2. Validation of image based measurements

Mutation load is the proportion of mutated mtDNA (number of red padlock signals) compared to wild type mtDNA (number of green padlock signals) per cell. The image based analysis was first performed on a padlock probed co-culture, meaning that cells with 100% wild type mtDNA were mixed and
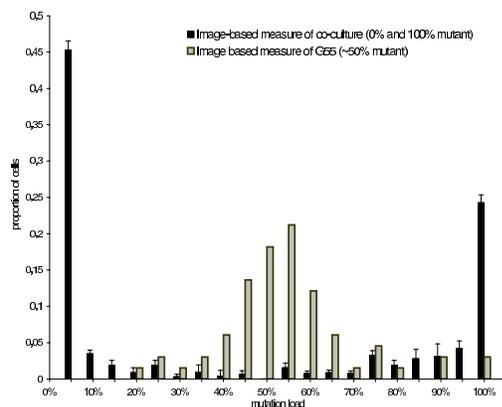
Fig. 3. Histogram of proportion of cell population against mutation load as achieved by image based measures. Error bars show variation caused when varying the threshold for signal detection.

cultured together with cells having 100% mutated mtDNA. This data set consisted of 29 images containing a total of 178 cells. A histogram of mutation load per cell measured from image data is shown in Fig. 3. The data from the co-culture shows distinct distributions at the extremes, i.e., cells with 100% and 0% mutation load. The automated analysis performed very well considering hardly any intermediate levels were found. A large amount of intermediate levels would have been an indication of a high degree of error in the analysis method. Second, an analysis of a padlock probed culture (G55) of cells with a ~50% mtDNA mutation load was made on 66 cells in 10 images. As predicted, the analysis from G55 has a clear peak close to 50% mutation load.

4. CONCLUSIONS

Comparison of methods for cytoplasmic segmentation show that presence of a cytoplasmic stain does not result in a significant increase in accuracy. This may seem strange as a cytoplasmic stain will guide the segmentation mask to the true edges of the cytoplasm. However, in the presented analysis, inclusion of parts of the image background does not affect the measurement of mutation load, as no signals are present in the background. Therefore, we have chosen not to count the inclusion of background as part of the false positive area fraction (FPAF). For other applications, e.g., if cytoplasmic area is to be measured, a segmentation method making use of the information from a cytoplasmic stain may be necessary. It is also worth mentioning that the agreement between manual cytoplasm segmentation and either of the fully automated methods is about the same as the agreement between manual cytoplasm segmentation performed by two different persons.

The automated method not including a cytoplasmic stain turned out to be a sufficiently accurate and fast method for analysis of single cell mutation load. The fact that no cytoplsmic stain was included also allows the use of two different colors for mutant and wild type mtDNA without problems with overlapping fluorescence spectra.

In combination with automated image acquisition and batch processing of image files, the presented methods opens the possibility of high-throughput analysis of large numbers of cells with little human interaction or observer bias. This is also the next step to take within this project.

5. REFERENCES

[1] G. Borgefors. Distance transformations in digital images. *Computer Vision, Graphics and Image Processing*, 34:344–371, 1986.

[2] A. E. Carpenter, T. R. Jones, M. R. Lamprecht, C. Clarke, I. H. Kang, O. Friman, D. A. Guertin, J. H. Chang, R. A. Lindquist, J. Moffat, P. Golland, and D. M. Sabatini. Cellprofiler: image analysis software for identifying and quantifying cell phenotypes. *Genome Biology*, 7(10), 2006.

[3] C. Lantuéjoul and S. Beucher. On the use of geodesic metric in image analysis. *Journal of Microscopy*, 121:39–49, 1981.

[4] C. Larsson, J. Koch, A. Nygren, G. Janssen, A. K. Raap, U. Landegren, and M. Nilsson. In situ genotyping individual DNA molecules by target- primed rolling- circle amplification of padlock probes. *Nature Methods*, 1:227–232, 2004.

[5] J. Lindblad, C. Wählby, E. Bengtsson, and A. Zaltsman. Image analysis for automatic segmentation of cytoplasms and classification of Rac1 activation. *Cytometry*, 57(1):22–33, January 2004.

[6] C. Ortiz de Solorzano, R. Malladi, S. A. Lelievre, and S. J. Lockett. Segmentation of nuclei and cells using membrane related protein markers. *Journal of Microscopy*, 201(3):404–415, 2001.

[7] N. Otsu. A threshold selection method from gray-level histograms. *IEEE Trans. on System Man and Cybernetics*, 9(1):62–69, 1979.

[8] P. Soille. *Morphological Image Analysis: Principles and Applications*. Springer-Verlag, 1999.

[9] J. K. Udupa, V. R. LeBlanc, Y. Zhuge, C. Imielinska, H. Schmidt, L. M. Currie, B. E. Hirsch, and J. Woodburn. A framework for evaluating image segmentation algorithms. *Computerized Medical Imaging and Graphics*, 30:75–87, 2006.

[10] C. Wählby, J. Lindblad, M. Vondrus, E. Bengtsson, and L. Björkesten. Algorithms for cytoplasm segmentation of fluorescence labelled cells. *Analytical Cellular Pathology*, 24:101–111, 2002.