

Automated Segmentation of *HeLa* Nuclear Envelope from Electron Microscopy Images

Cefa Karabag^{1,3}, Martin L. Jones², Christopher J. Peddie², Anne E. Weston²,
Lucy M. Collinson², and Constantino Carlos Reyes-Aldasoro^{1,3}

¹ Research Centre for Biomedical Engineering, School of Mathematics, Computer Science and Engineering,

City, University of London, Northampton Square EC1V 0HB, UK,

² Electron Microscopy Science Technology Platform, The Francis Crick Institute,
1 Midland Road, London NW1 1AT, UK

³ Corresponding authors cefa.karabag.1@city.ac.uk
constantino.reyes-aldasoro.1@city.ac.uk

Abstract. This paper describes an image-processing pipeline for the automatic segmentation of the nuclear envelope of *HeLa* cells observed through Electron Microscopy. The pipeline was applied to a 3D stack of 300 images. The intermediate results of neighbouring slices are further combined to improve the final results. Comparison with a hand-segmented ground truth reported Jaccard similarity values between 94-98% on the central slices with a decrease towards the edges of the cell where the structure was considerably more complex. The processing is unsupervised and each 2D slice is processed in about 5-10 seconds running on a MacBook Pro. No systematic attempt to make the code faster was made. These encouraging results could be further used to provide data for more complex segmentation techniques like Deep Learning, which require a considerable amount of data to train architectures like Convolutional Neural Networks. The code is freely available from <https://github.com/reyesaldasoro/HeLa-Cell-Segmentation>.

Keywords: Automatic Nuclear Segmentation, HeLa cells, Electron Microscopy, Cell and Nuclear Shape

1 Introduction

All living creatures are made of cells, which are the smallest structural and functional unit of living organisms [1]. Cells have the ability to create copies of themselves by subdivision [2]. Whilst in the majority of cases, cells are able to subdivide correctly, sometimes errors in the DNA replication lead to mutations. Some mutations then can inactivate tumour suppressor genes or activate proto-oncogenes and the result is a cell that subdivides without control and can become a cancerous tumour [2-4].

The change from a normal cell to a malignant one has multiple characteristics, sometimes summarised as six *hallmarks*: self-sufficiency in growth signals, evasion

of apoptosis, insensitivity to anti-growth signals, sustained angiogenesis, tissue invasion and metastasis and limitless replicative potential [3, 5].

Cancer research has utilised malignant cell lines to study numerous processes; perhaps the most important line has been the cervical cancer *HeLa* cells, derived in 1951 [6, 7] and now widely used as model for thousands of biological experiments to contribute to the understanding of the disease [8, 9]. The availability of cell lines together with genetic advances has contributed significantly to cancer research, which has grown significantly as a proportion of all research in biomedical areas [10]. Cervical cancer is of importance in its own right as it contributes, together with breast cancer, to 4.2% of the global causes of death [11]. In the United Kingdom around 3,000 cases are diagnosed every year and just under 1,000 women die from cervical cancer every year [12].

One key element in the study of cancerous cells is the observation of the nucleus, in processes like the formation of the post-mitotic nuclear envelope [13]. However, the organisation of the nucleus and its surrounding structures remains an area largely unexplored [14]. Perhaps the reason behind this lack of research is due to the high resolution required. The resolution required to reveal cells' fine structures can be provided by electron microscopy [15, 16].

Segmentation of cells in microscopy images is a widely studied area [17]. Considerable work has been done on the segmentation of overlapping cells [18–21] and segmentation of nuclei has been performed with different techniques, among them watershed transformations [22–25]. However, at high magnification such as that provided by electron microscopes, the segmentation concentrates not only on separating the cells from each other, but the cell structures apart to better understand the nuclear formation [26], the chromosome arrangement [27] or the *molecular sociology* [28, 29]. At this level of magnification, the segmentation is particularly difficult as there is little regularity as compared with lower magnifications and even more complicated with 3D data sets [30].

Automatic cell segmentation remains one of the most challenging processes in image analysis [18, 31] and despite the significant disadvantages of time and inter- and intra-user variability, manual segmentation remains widely used and in many cases, is still considered as the gold standard. Thus, a considerable amount the work is still done manually [32] or with semi-automated methods [33].

An alternative to a single expert manually delineating a desired region of interest has been to gather the effort of many non-experts, in what has been called *citizen-science* [34]. The main idea is that an army of internet non-experts may provide enough results so that a median result has a comparable accuracy and value than those provided by a limited number of experts. For this purpose, a citizen-science project called *Etch a Cell* [35] was created to provide manual segmentation of the nuclear envelope of HeLa cells.

In this paper, a methodology for the automated segmentation of nuclear envelope of HeLa cells is presented. The methodology consists of several image-processing steps: low-pass filtering, edge detection and determination of super-pixels, distance transforms and delineation of the nuclear envelope. The algo-

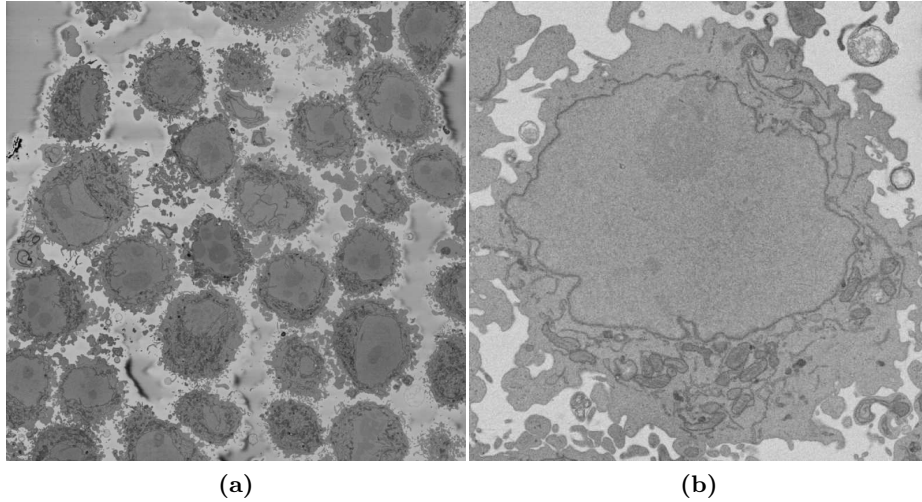


Fig. 1: (a) One representative slice of a 3D stack acquired with an Electron Microscope containing numerous HeLa cells. (b) Region of Interest with one HeLa cell centred. At this resolution, it is easy to distinguish the nuclear envelope, borders of the cell and other structures.

rithm assumes the following: (a) there is a **single HeLa cell** of interest in the field of view, (b) the **centre of the cell** is located at centre of a 3D stack of images, (c) the **nuclear envelope is darker** than the nuclei or its surroundings, and (d) the **background is brighter** than any cellular structure. The centre of the cell is in the centre as individual cells have been hand-located from the whole field of view as will be described in the following section.

2 Materials and Methods

2.1 HeLa Cells preparation and acquisition

Wild type HeLa cells were prepared and embedded in Durcupan resin following the method of the National Centre for Microscopy and Imaging Research (NCMIR) [36].

Serial blockface scanning electron microscopy (SBF SEM) data was collected using a 3View2XP (Gatan, Pleasanton, CA) attached to a Sigma VP SEM (Zeiss, Cambridge). Images were acquired at 8192×8192 pixels over a total of 518 slices, with $10 \times 10 \times 50$ nm voxel size and the intensity of each voxel was between 0 and 255 as the data type was unsigned integer of 8 bits. Individual cells were manually cropped as volumes of interest. The single cell used in this study is a substack of $2000 \times 2000 \times 300$ voxels.

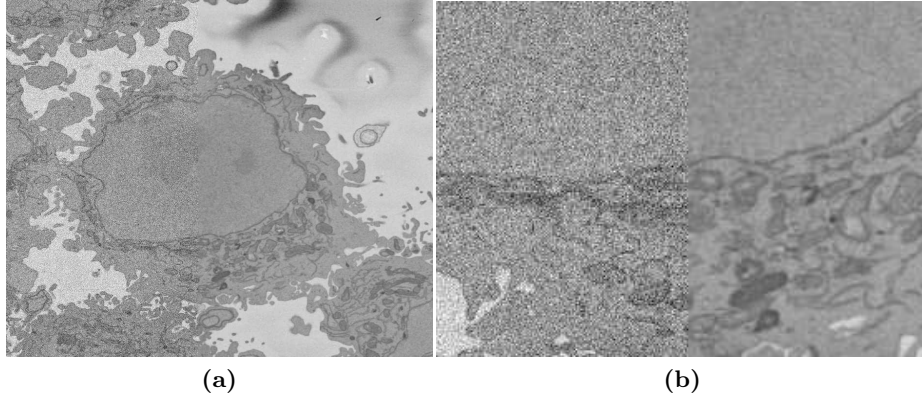


Fig. 2: (a) Illustration of the effects of low-pass filtering on one representative image of a HeLa Cell. Left half: Original image with grainy texture. Right half: image filtered with 5×5 Gaussian low pass filter. (b) Region of interest of (a).

2.2 Algorithms

In this work, a stack of three hundred images acquired as previously described were processed in MATLAB[®] (The Mathworks[™], Natick, USA). The images corresponded to a Region of Interest (ROI) of a larger image (Fig. 1a) which was cropped to contain a single HeLa cell at the centre, but parts of other cells were visible at the sides (Fig. 1b). Each image of the stack had dimensions $(n_h, n_w, n_d) = (2000, 2000, 1)$ and were saved as single channel *TIFF* files. The images of electron microscopy do not contain colour information.

The first step of the pipeline applied a 5×5 Gaussian low-pass filter with kernel size ($h = 7$) and standard deviation ($\sigma = 2$) to remove the high frequency information and to enhance the larger scale edge features. This was required as the images presented a grainy texture probably due to noise in the acquisition (Fig. 2a). The segmentation steps will rely on the intensity thus it was important to have a smooth intensity that would characterise the nuclear region and its boundaries. The region of interest (ROI) and the smoothing effect can be seen in detail in Fig. 2b.

An edge-detection step followed the smoothing. A Canny edge-detection [37] was used in order to detect abrupt changes of intensity. Of particular interest was the variation of the nuclear envelope thus different Gaussians of different values were tested and best results were obtained with $\sigma = 4$ (Fig. 3a). The selection of a large σ is important so that the edges created by the nuclear envelope are not broken but selected as one continuous edge and also to avoid small edges due to relatively small variation of the intensity. It should be noticed, that besides the nuclear envelope a large number of edges are detected. A selection of foreground (the cell and its neighbours) and background (regions without cells) removed those edges. Superpixels were created in the following way: the previ-

ously detected edges were dilated and removed from the image, the remaining unconnected regions were labelled and became superpixels (Fig. 3b). There was no restriction in size of these superpixels as it was important to allow for large superpixels that corresponded to the nucleus and the background. Small superpixels were discarded (Fig. 3c) and then any superpixel that was in contact with the edges of the image was removed. The remaining superpixels were closed with a disk structural element and internal holes filled (Fig. 3d). The superpixel with the centroid closest to the centre of the image was assumed to be the nucleus (Fig. 3e). This is an assumption of the methodology as the average intensity of the large superpixels which were not selected as nucleus were examined to determine which ones corresponded to the background and those with higher intensities were assumed to correspond to the background.

Since the boundary of the nuclear envelope, previously detected with the edge detection, was removed for the creation of the superpixels, it was necessary to dilate the region previously assumed as the nucleus to detect the most accurate location of the envelope. To find the boundary, a distance transform from the region to every pixel of the image was performed (Fig. 3f). The average intensity of every loci with constant distance was calculated (Fig. 3g), where a valley clearly indicated the location of the nuclear envelope. It should be noticed how the intensities inside and outside the nuclear envelope were brighter than the envelope itself. Finally, the envelope was obtained by dilating with the distance previously calculated. Figure 3h shows the envelope overlaid on the filtered image with a magenta line. For comparison purposes, the manual delineation is shown with a cyan line. To illustrate the differences, a ROI is presented in Fig. 3i. Besides the small variations along the boundary, it should be noticed that a separate region was denoted as nuclei by the expert. This decision is taken by moving up and down through the 3D stack as the nucleus has complex shapes.

In order to compare the final automated segmentation with the manually segmented ground truth, Jaccard Similarity Index [38] of intersection over union of areas was calculated.

3 Results

The final segmentation of the HeLa cell is shown as a shaded volume (Fig. 4a,b) and two representative slices (Fig. 4c,d). The algorithm captured well the shape of the cell and it can be seen that one side of the cell is smoother (a) than the other (b). This is also visible on the 2D images. Solid arrows indicate an interesting *notch* that travels along the cell between slices. These variations on the shape of the cell must have biological significance and could be used for further biological experiments. A dashed line indicates another variation; however, this is most likely an artefact, either of the algorithm when contiguous slices were compared or of the acquisition due to a slight shift, and will require further work to determine the cause. The comparison of the areas reported values of the Jaccard similarity index between 94-98% on the central slices (Fig. 5).

6

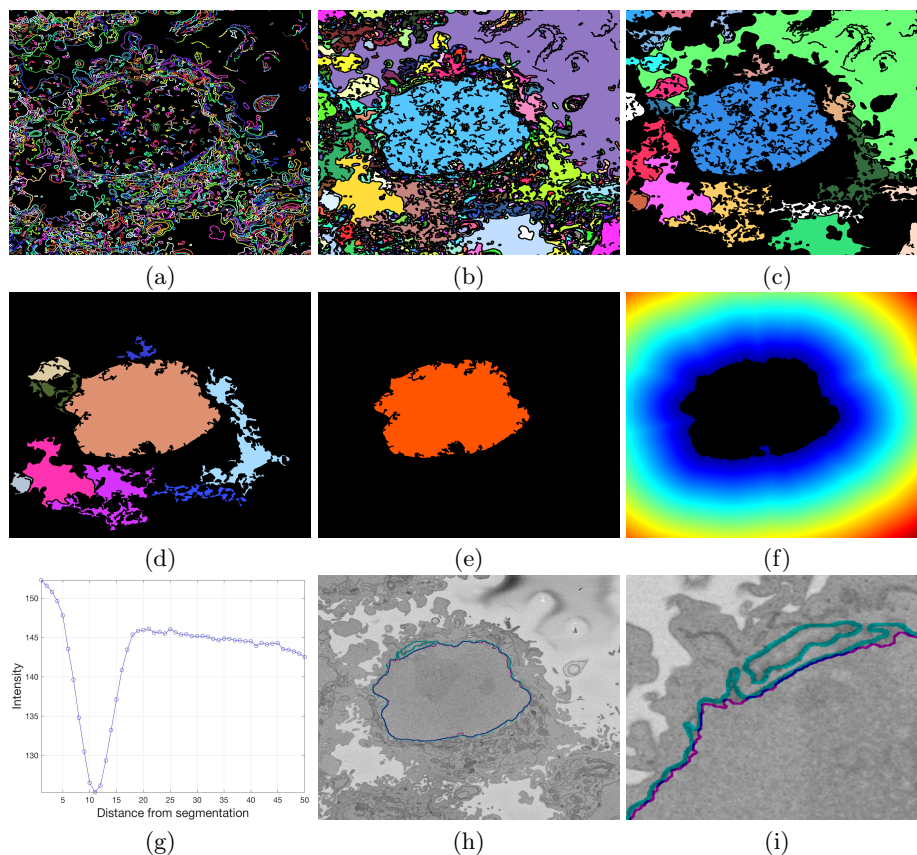


Fig. 3: Illustration of the pipeline to segment the nuclear envelope of the HeLa Cell. (a) Edges detected by Canny algorithm. The edges were further dilated to connect those edges that may belong to the nuclear envelope but were disjoint due to variations of the intensity of the envelope itself. (b) The edges previously detected were dilated, removed from the image and the remaining regions were labelled to become superpixels. (c) Morphological selection of large superpixels. (d) Removal of superpixels that contact the edges of the image and morphological processing to remove internal holes. (e) Selection of the central superpixel. (f) Distance map from the central superpixel. (g) Average intensity from the central superpixel, the minimum corresponds to the nuclear envelope that appears dark as compared with the nuclei and its surrounding structures. (h) Final segmentation of the nuclear envelope shown in purple. For comparison, the hand-segmented segmentation is shown in cyan. (i) Detail of (h) where differences can be appreciated. Notice that the ground truth detected a region that does not connect with the central nuclear envelope. The edges and the labelled areas have been assigned random colours for visualisation purposes.

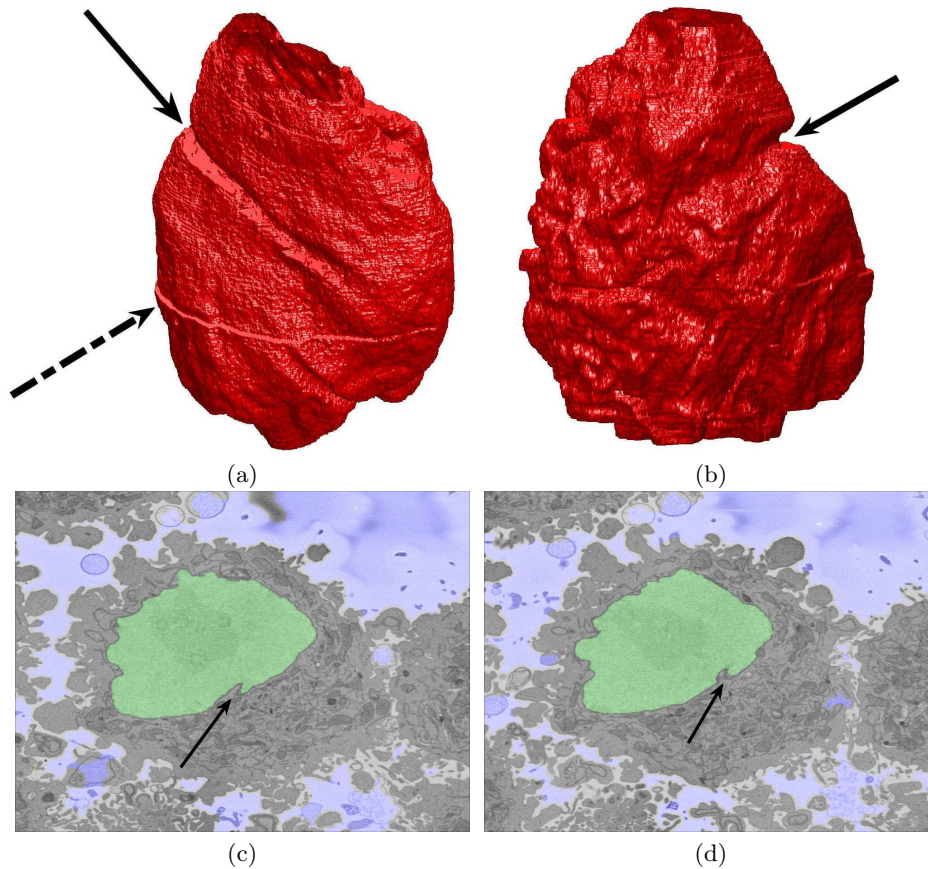


Fig. 4: Final results of the segmentation as a rendered volumetric surface (a,b) and 2D slices (c,d) where the automatic segmentation is shown with a green shade and the background with a light purple shade. Solid line arrows indicate a *notch* that travels along the nucleus. The dashed line arrow indicates an artefact, which could be due to the segmentation algorithm when neighbouring slices are processed or could also be due to a slight shift in the acquisition, which would be resolved with a pre-alignment. Further investigation will be dedicated to this artefact.

4 Discussion

In this work, an automatic segmentation algorithm of the nuclear envelope of HeLa cells acquired with Electron Microscopy was described. The 300 images described in section 2.1 were processed with the proposed automated segmentation algorithm. Promising segmentation results, as indicated by the Jaccard values, were obtained.

8

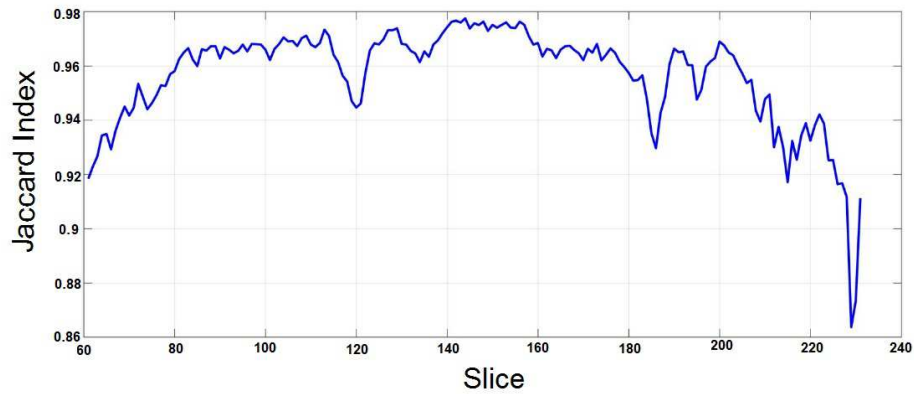


Fig. 5: Jaccard index computed against ground truth for slices between 60 and 230 of the stack. The values for the central slices ranges between 94-98%. For slices on the top and on the bottom the values decrease due to the complexity of the shape.

The decrease in similarity towards the bottom and top of the cell can be understood as the shape becomes less regular and more complex and the nucleus is formed by more than one region. Two sources of error were thus identified. First, the detected boundary did not match that of the manual observer (dashed lines in Fig. 6a,b). Further examination is required to notice the exact contribution of these errors to the total index. It should also be noticed that errors of the manual expert are possible. Second, there were more than one region corresponding to the nucleus (solid lines in Fig. 6a,b). This is a limitation of the algorithm and further work can rely on the segmentation of contiguous slices.

Future work will consider testing this algorithm in a larger number of cells and comparing against other methodologies like active contours.

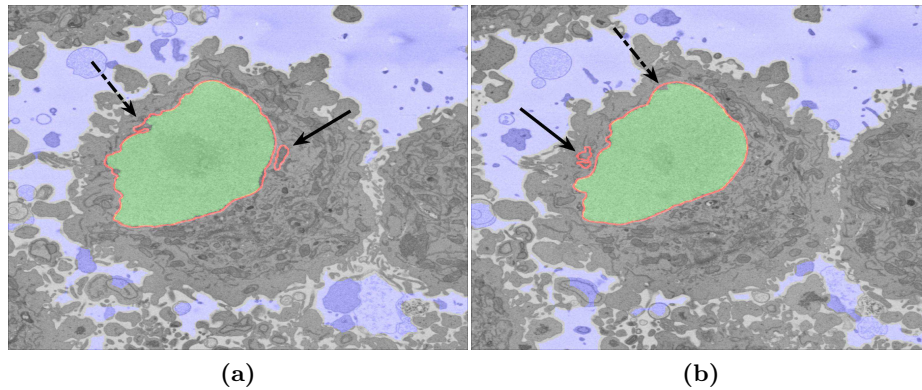


Fig. 6: Two representative slices where the manual segmentation of the nucleus is indicated with a red line, the automatic segmentation with a green shade and the background with a light purple shade. In both images, dashed arrows show regions of nucleus where the manual and automatic boundaries diverge, and solid arrows indicate small regions of nucleus that were not contiguous to the main region of the nucleus.

5 Acknowledgements

This work was supported by the *Francis Crick Institute* which receives its core funding from Cancer Research UK (FC001999), the UK Medical Research Council (FC001999), and the Wellcome Trust (FC001999).

The authors acknowledge the support of the *Alan Turing Institute* through the Data Study Groups organised by Dr Sebastian Vollmer where initial study of this data was made.

References

1. Ellinger, I., Ellinger, A.: Smallest Unit of Life: Cell Biology 2. Comparative Medicine: Anatomy and Physiology 9(24), 19–33 (2014)
2. Alberts, B., Bray, D., Johnson, A., Lewis, J., Raff, M., Roberts, K., Walter, P.: Essential Cell Biology: An Introduction to the Molecular Biology of the Cell. Garland Publishing Inc., 1st edn. (1998)
3. Hanahan, D., Weinberg, R.A.: The Hallmarks of Cancer. Cell 100, 57–70 (2000)
4. Malumbres, M., Barbacid, M.: Cell cycle, CDKs and cancer: a changing paradigm. Nature 9(24), 153–166 (2009)
5. Hanahan, D., Weinberg, R.A.: Hallmarks of cancer: The next generation. Cell 144(5), 646674 (Mar 2011), pMID: 21376230
6. Verma, R.P., Hansch, C.: Chemical Toxicity on HeLa Cells. Current Medicinal Chemistry 13(4), 423–448 (2006)
7. Masters, J.R.: HeLa cells 50 years on: the good, the bad and the ugly. Nature Reviews Cancer 2, 315–319 (2002)

10

8. Lucey, B.P., Nelson-Rees, W.A., Hutchins, G.M.: Henrietta Lacks, HeLa Cells and Cell Culture Contamination. *Archives of Pathology & Laboratory Medicine* 133, 1463–1467 (2009)
9. Lacroix, M.: Persistent use of “false” cell lines. *International Journal Cancer* 122, 1–4 (2008)
10. Reyes-Aldasoro, C.C.: The proportion of cancer-related entries in PubMed has increased considerably; is cancer truly the emperor of all maladies? *PLOS ONE* 12(3), e0173671 (Mar 2017)
11. Lozano, R., Naghavi, M., Foreman, K., Lim, S., Shibuya, K., Aboyans, V., Abraham, J., Adair, T., Aggarwal, R., Ahn, S.Y., et al.: Global and regional mortality from 235 causes of death for 20 age groups in 1990 and 2010: a systematic analysis for the global burden of disease study 2010. *Lancet (London, England)* 380(9859), 20952128 (Dec 2012), PMID: 23245604
12. NHS, UK. <https://www.nhs.uk/conditions/cervical-cancer/>, accessed: 2018-01-17
13. Lu, L., Ladinsky, M.S., Kirchhausen, T.: Formation of the postmitotic nuclear envelope from extended ER cisternae precedes nuclear pore assembly. *The Journal of Cell Biology* 194(3), 425440 (Aug 2011), PMID: 21825076 PMID: PMC3153650
14. Mahamid, J., Pfeffer, S., Schaffer, M., Villa, E., Danev, R., Cuellar, L.K., Frster, F., Hyman, A.A., Plitzko, J.M., Baumeister, W.: Visualizing the molecular sociology at the hela cell nuclear periphery. *Science (New York, N.Y.)* 351(6276), 969972 (Feb 2016), PMID: 26917770
15. Haguenuau, F., Hawkes, P.W., Hutchison, J.L., Satiat-Jeunemaitre, B., Simon, G.T.: Key Events in the History of Electron Microscopy. *Microscopy and Microanalysis* 9(4), 96–138 (2003)
16. Masters, B.R.: History of the Electron Microscope in Cell Biology. *Encyclopedia of Life Sciences*, John Wiley & Sons, Ltd (2009)
17. Wang, Z., Li, H.: Generalizing cell segmentation and quantification. *BMC bioinformatics* 18(1), 189 (Mar 2017), PMID: 28335722 PMID: PMC5364575
18. Cloppet, F., Boucher, A.: Segmentation of overlapping/aggregating nuclei cells in biological images. In *Proceedings IEEE International Conference Pattern Recognition* pp. 1–4 (2008)
19. Solis-Lemus, J.A., Stramer, B., Slabaugh, G., Reyes-Aldasoro, C.C.: Segmentation and Shape Analysis of Macrophages Using Anglegram Analysis. *Journal of Imaging* 4(2) (2017)
20. Zhou, X., Li, F., Yan, J., Wong, S.T.C.: A Novel Cell Segmentation Method and Cell Phase Identification Using Markov Model. *IEEE Transactions on Information Technology in Biomedicine* 13(2), 152–157 (2009)
21. Jung, C., Kim, C., Chae, S.W., Oh, S.: Unsupervised Segmentation of Overlapped Nuclei Using Bayesian Classification. *IEEE Transactions on Biomedical Engineering* 57(12), 2825–2832 (2010)
22. Cheng, J., Rajapakse, J.C.: Segmentation of clustered nuclei with shape markers and marking function. *IEEE Transactions on Biomedical Engineering* 56(3), 741–748 (2009)
23. Yang, X., Li, H., Zhou, X.: Nuclei Segmentation Using Marker-Controlled Watershed, Tracking Using Mean-Shift, and Kalman Filter in Time-Lapse Microscopy. *IEEE Transactions on Circuits and Systems I: Regular Papers* 53(11), 2405–2414 (2006)
24. Kaya, B., Goceri, E., Beker, A., Elder, B., Puduvalli, V., Winter, J., Gurcan, M., Otero, J.J.: Automated fluorescent microscopic image analysis of PTBP1 expression in glioma. *PLoS One* 12(3):e0170991 (2017)

25. Goceri, E., Goksel, B., Elder, J.B., Puduvali, V.K., Otero, J.J.: Quantitative validation of anti-PTBP1 antibody for diagnostic neuropathology use: Image analysis approach. *Journal of Numerical Methods in Biomedical Engineering* 33:e2862 (2017)
26. Lu, L., Ladinsky, M.S., Kirchhausen, T.: Formation of the postmitotic nuclear envelope from extended ER cisternae precedes nuclear pore assembly. *The Journal of Cell Biology* 194(3), 425–440 (2011)
27. Walter, J., Schermelleh, L., Cremer, M., Tashiro, S., Cremer, T.: Chromosome order in HeLa cells changes during mitosis and early G1, but is stably maintained during subsequent interphase stages. *The Journal of Cell Biology* 160(5), 685–697 (2003)
28. Mahamid, J., Pfeffer, S., Schaffer, M., Villa, E., Danev, R., Cuellar, L.K., Forster, F., Hyman, A.A., Plitzko, J.M., Baumeister, W.: Visualizing the molecular sociology at the HeLa cell nuclear periphery. *Science* 351(6276), 969–972 (2016)
29. Robinson, C.V., Sali, A., Baumeister, W.: The molecular sociology of the cell. *Nature* 450, 973–982 (2007)
30. Peddie, C.J., Collinson, L.M.: Exploring the third dimension: Volume electron microscopy comes of age. *Micron* 61, 919 (Jun 2014)
31. Bajcsy, P., Cardone, A., Chalfoun, J., Hlater, M., Juba, D., Kociolek, M., Majurski, M., Peskin, A., Simon, C., Simon, M., Vandecreme, A., Brady, M.: Survey statistics of automated segmentations applied to optical imaging of mammalian cells. *BMC Bioinformatics* 16(330) (2015)
32. Coelho, L.P., Shariff, A., Murphy, R.F.: Nuclear Segmentation in Microscope Cell Images: A Hand-Segmented Dataset and Comparison of Algorithms. In *Proceeding IEEE International Symposium Biomedical Imaging* pp. 518–521 (2009)
33. Belevich, I., Joensuu, M., Kumar, D., Vihinen, H., Jokitalo, E.: Microscopy image browser: A platform for segmentation and analysis of multidimensional datasets. *PLOS Biology* 14(1), e1002340 (Jan 2016)
34. Schnoor, J.L.: Citizen science. *Environmental Science & Technology* 41(17), 5923–5923 (Sep 2007)
35. Etch A Cell. <https://www.zooniverse.org/projects/h-spiers/etch-a-cell>, accessed: 2018-01-17
36. Deerinck, T.J., Bushong, E., Thor, A., Ellisman, M.H.: NCMIR - National Center for Microscopy and Imaging Research. NCMIR methods for 3D EM: A new protocol for preparation of biological specimens for serial block-face SEM microscopy (2010), <https://ncmir.ucsd.edu/sbem-protocol>, accessed: 2018-02-09
37. Canny, J.: A Computational Approach to Edge Detection. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 8(6), 679–698 (1986)
38. Jaccard, P.: Étude comparative de la distribution florale dans une portion des Alpes et des Jura. *Bulletin del la Société Vaudoise des Sciences Naturelles* 37, 547–579 (1901)