Automated Analysis of siRNA Screens of Virus Infected Cells based on Immunofluorescence Microscopy

Petr Matula[1,2], Anil Kumar[3], Ilka Wörz[3], Nathalie Harder[1,2], Holger Erfle[4], Ralf Bartenschlager[3], Roland Eils[1,2], Karl Rohr[1,2]

[1]BIOQUANT, IPMB, University of Heidelberg, [2] Dept. Bioinformatics and Functional Genomics, Biomedical Computer Vision Group, DKFZ Heidelberg
[3] Department of Molecular Virology, University of Heidelberg
[4] BIOQUANT, University of Heidelberg
p.matula@dkfz.de

Abstract. We present an image analysis approach as part of a high-throughput microscopy screening system based on cell arrays for the identification of genes involved in Hepatitis C and Dengue virus replication. Our approach comprises: cell nucleus segmentation, quantification of virus replication level in cells, localization of regions with transfected cells, cell classification by infection status, and quality assessment of an experiment. The approach is fully automatic and has been successfully applied to a large number of cell array images from screening experiments. The experimental results show a good agreement with the expected behavior of positive as well as negative controls and encourage the application to screens from further high-throughput experiments.

1 Introduction

Viruses need to enter cells and exploit their cellular machinery to produce own copies. Classical virology has focused on the virus itself, thereby neglecting critically important processes of the host cells [1]. Our ultimate goal is to develop a high-throughput microscopy screening system for genome-wide identification of cellular genes potentially involved in virus entry and replication, which is expected to lead to improvements in antiviral treatments. The general aim of our study is to develop approaches for automated analysis of images from genome-wide screens (with more than 20,000 genes generating more than 100,000 fluorescence images). Key tasks in this application are cell nucleus segmentation, detection of regions with transfected cells, and quantification of the virus infection level. The approaches should be efficient, robust, and fully automatic.

In recent years, a number of approaches for cell nuclei or whole cell segmentation based on fluorescence microscopy or other microscopy techniques have been reported (e.g., [2, 3, 4]). In *high-throughput* applications, approaches based on adaptive thresholding [5, 6] proved to give good results for cell nuclei segmentation, especially if the nuclei are not clustered. To separate clusters of nuclei, watershed-based approaches are often employed (e.g., [6]).

In this contribution, we describe an image analysis approach as part of a high-throughput system for genome-wide identification of cellular genes that are important for Hepatitis C and Dengue virus replication. Currently, to our best knowledge, there exists no image analysis system for quantifying viral signals given a large number of images on per cell basis, i.e. which measures the status of virus replication in transfected cells only, and combines the results from many different and repeated experiments to produce reliable statistics. In particular, we propose novel approaches (1) for the segmentation of cell nuclei based on gradient thresholding, (2) for the localization of regions with transfected cells within cell array images, and (3) for cell classification based on the infection level with application to quality assessment of an experiment.

2 Materials and Methods

The input data in our application consists of two-channel microscopy images of small interference RNA (siRNA) cell arrays [7] on which transfection reagents and different siRNAs are spotted in a grid pattern (Fig. 1a). We have developed an approach that comprises the following five main steps: (1) cell nuclei segmentation in channel 1 (Fig. 1b), (2) measurement of the level of viral protein expression in channel 2 stained by immunofluorescence (Fig. 1c), (3) detection of siRNA spot areas with transfected cells (Fig. 1d), (4) classification of cells based on their infection status, and (5) quality assessment of an experiment.

– *Segmentation:* To segment cell nuclei we use a gradient thresholding approach. We determine cell nucleus boundary regions by combining information from a gradient magnitude image and an image obtained by applying the Laplacian of Gaussian. This approach was previously applied for the segmentation of characters for text recognition [8]. In comparison to [8] we introduce the following post-processing steps. First, we apply connected component labeling and remove small and large objects. Then, the remaining objects are
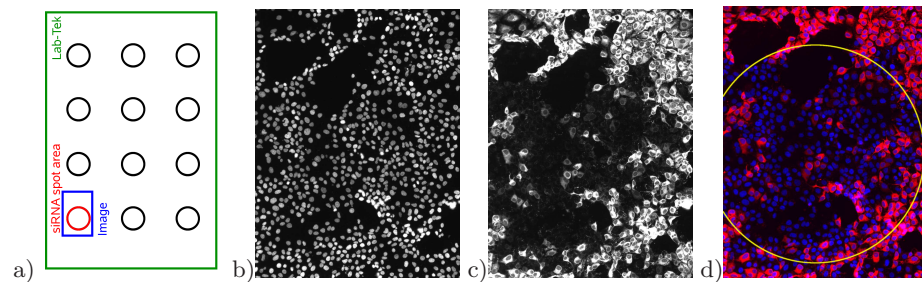

Fig. 1. Image data. (a) Cell array with $M \times N$ siRNA spot areas, (b) Image of cell nuclei for one spot area (channel 1), (c) Corresponding viral protein expression (channel 2), (d) Overlay of both channels and marked spot area

conditionally dilated while preventing merging with other components. Afterwards, holes in objects are filled (Fig. 2a). Finally, cell nuclei are identified in segmented objects by applying size and circularity criteria.

- *Quantification:* The virus replication level is quantified for each nucleus in its neighborhood by computing the mean of the intensities in the virus channel (channel 2). We have implemented and compared three different approaches for defining the neighborhood of nuclei: simple dilation (Fig. 2b), restricted dilation by influence zones (IZ) (Fig. 2d), and region growing in IZ (Fig. 2e). The latter two approaches rely on a partition of an image into IZ of each nucleus (Fig. 2c), which was computed using a seeded watershed transform of the inverted virus channel with the nuclei as initial seeds.

- *Localization:* Only those cells that come in contact with siRNA molecules should be quantified and therefore a circular region with transfected cells of known diameter must be localized in each image. Our approach consists of three steps: (1) finding the position of a circle of known diameter that has maximal difference between the mean virus expression level of cells inside and outside the circle, (2) selecting those positions that have expression level differences within a valid range, and (3) fitting of a grid with known parameters. Spot diameter and grid parameters are known from the siRNA spotting process. The valid range in step 2 was determined by simulations.

- *Classification:* To detect differences in virus infection level we use a measure denoted as infection rate ratio: $IRR = IR_i/IR_N$, where IR_i is the percentage of virus infected cells in siRNA spot area i and IR_N is the normal virus infection rate, i.e. the percentage of infected cells without knocked-down genes. To compute infection rates, the cells must first be classified as infected or non-infected. For classification we use the mean intensity in the virus channel in the neighborhood of a cell nucleus. The classification threshold was determined by maximizing the difference of infection rates in positive and negative controls. In positive controls the protein production level is blocked and therefore the signal is reduced. In negative controls the level of virus replication is not altered.

- *Quality Assessment:* A large number of images need to be analyzed to derive sound statistics. Since some of the images may be of poor quality (e.g., out-
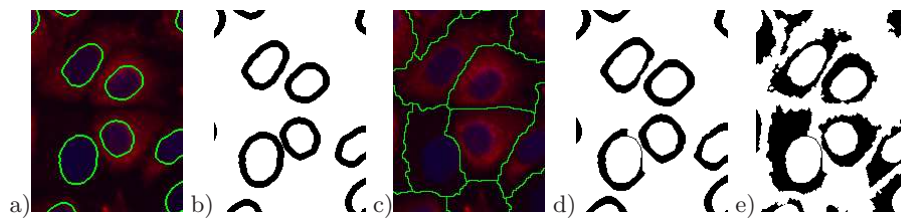

Fig. 2. Segmentation results for an image section. (a) Cell nuclei (blue), viral protein expression (red) with overlaid segmented nuclei, (b) Neighborhoods obtained by dilation, (c) Influence zones (IZ), (d) Constrained dilation by IZ, (e) Region-growing inside IZ

of-focus, no cells in certain areas, image artifacts) we need to exclude them. We perform quality checks on two levels: (1) on the whole experiment level and (2) on the single image level. On the whole experiment level, the main criterion is the difference between infection rates in positive and negative controls. If the difference is too small, the whole experiment is excluded. On the single image level, images with too small or a too large number of cells as compared to the average number of the whole experiment are excluded.

3 Results

Prior to applying the overall approach, we performed an evaluation of our approach for cell nuclei segmentation. To obtain ground truth, two experts marked cell nuclei in 8 randomly selected real images from 4 different experiments (2243 nuclei in total). We compared the results of our approach with that of adaptive thresholding by Otsu's method [5]. With our approach we were able to segment more than 94% of all nuclei whereas Otsu's method yielded only about 80%. We also studied the performance of the three approaches for defining the cell nucleus neighborhood. For each approach, we computed infection rate ratios (IRRs) for two experiments with 384 images each. The results of all three approaches were similar. Therefore we use the simplest approach based on dilation.

We have applied the overall approach to more than 20,000 images of Hepatitis C (HCV) and Dengue Virus (DV) infected screens. The overall results of 10 repeated experiments of an HCV screen (grid 6×6) and of 4 repeated experiments of a DV screen (grid 12×32) are presented in Fig. 3. Positive and negative controls are marked in Fig. 3a with light and dark gray bars and in Fig. 3b,c
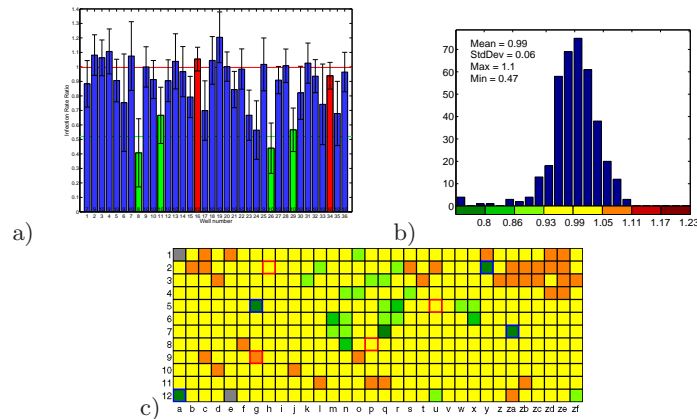

Fig. 3. Infection rate ratios computed from screening experiments. Results averaged (a) from 10 repeated experiments of an HCV screen, and (b, c) from 4 repeated experiments of a DV screen. The infection rate ratio is represented by gray tones (see the gray tone scale below the histogram)

they are marked at positions (a12, g5, q7, za7) and (g9, h2, p8, u5), respectively. It turned out that we obtain a good agreement with a reduced IRRs in positive controls as compared to IRRs in negative controls. Besides positive controls, reduced IRRs were also observed in other siRNA spot areas. This indicates the applicability of the whole approach.

4   Discussion

We have described an approach for the identification of genes involved in Hepatitis C and Dengue virus replication based on high-throughput screening experiments. The whole approach relies on (1) a novel gradient-based approach for segmentation of cell nuclei in fluorescence microscopy images, which increases the number of correctly segmented objects compared to adaptive thresholding, (2) combination of model-based circular region fitting and grid fitting for the localization of siRNA spot regions, which allows to exclude non-transfected cells, and (3) cell classification based on the infection level. The overall approach allows to fully automatically quantify a large number of images on single cell basis. The obtained results are in good agreement with the expected behavior and encourage the application to images from other high-throughput experiments, in particular, from genome-wide screens.

References

1. Dam EM, Pelkmans L. Systems biology of virus entry in mammalian cells. Cell Microbiol. 2006;8(8):1219–27.
2. Wählby C, Lindblad J, Vondrus M, et al. Algorithms for cytoplasm segmentation of fluorescence labelled cells. Anal Cell Pathol. 2002;24:101–11.
3. Würflinger T, Stockhausen J, Meyer-Ebrecht D, et al. Robust automatic coregistration, segmentation, and classification of cell nuclei in multimodal cytopathological microscopic images. Comput Med Imaging Graph. 2004;28(1–2):87–98.
4. Elter M, Daum V, Wittenberg T. Maximum-intensity-linking for segmentation of fluorescence-stained cells. Proc MIAAB. 2006; p. 46–50.
5. Harder N, Mora-Bermúdez F, Godinez WJ, et al. Automated analysis of the mitotic phases of human cells in 3D fluorescence microscopy image sequences. Lect Note Comp Sci. 2006;4190:840–8.
6. Li F, Zhou X, Ma J, et al. An automated feedback system with the hybrid model of scoring and classification for solving over-segmentation problems in RNAi high-content screening. J Microsc. 2007;226(2):121–32.
7. Erfle H, Simpson JC, Bastiaens PIH, et al. siRNA cell arrays for high-content screening microscopy. Biotechnol. 2004;37(3):454–62.
8. Gonzalez RC, Woods RE. Digital Image Processing. Prentice Hall; 2002.