Model-Based Segmentation and Colocalization Quantification in 3D Microscopy Images

Stefan Wörz[1], Petra Sander[2], Martin Pfannmöller[1], Ralf J. Rieker[3,4], Stefan Joos[2], Gunhild Mechtersheimer[3], Petra Boukamp[5], Peter Lichter[2], Karl Rohr[1]

[1]Dept. Bioinformatics and Functional Genomics, Biomedical Computer Vision Group, University of Heidelberg, BIOQUANT, IPMB, and DKFZ Heidelberg
[2]Dept. Molecular Genetics, DKFZ Heidelberg
[3]Dept. of Pathology, University Hospital, Heidelberg
[4]Institute of Pathology, Medical University of Innsbruck
[5]Dept. Genetics of Skin Carcinogenesis, DKFZ Heidelberg
s.woerz@dkfz.de

Abstract. We introduce a new model-based approach for automatic quantification of colocalizations in 3D microscopy images. The approach is based on different 3D parametric intensity models in conjunction with a model fitting scheme to quantify subcellular structures with high accuracy. The central idea is to determine colocalizations between different channels based on the estimated geometry of subcellular structures. Furthermore, we perform a statistical analysis to assess the significance of the determined colocalizations. We have successfully applied our approach to about 400 three-channel 3D microscopy images of soft-tissue tumors.

1 Introduction

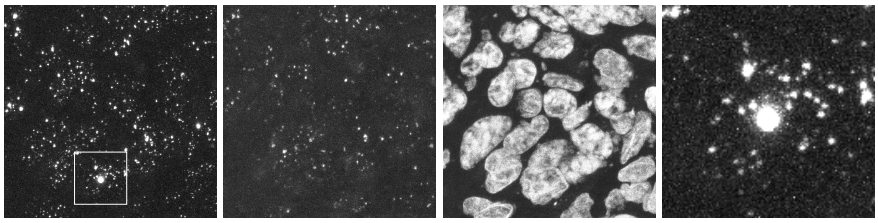
Human telomeres are the physical ends of linear chromosomes that protect them from improper DNA metabolism. Telomere shortening within cell division plays an important role in aging and cancer. The activation of a telomere length maintenance mechanism (TMM) is thus essential for cellular immortalization and long-term tumor growth. Telomerase activity is the most common TMM. However, some tumors have been shown to utilize recombination-based TMMs that have been termed alternative lengthening of telomeres (ALT). Since the TMM can vary significantly in soft tissue sarcomas its characterization is crucial for developing new therapies. Promyelocytic leukemia (PML) bodies are subnuclear entities colocalizing with various proteins in the cell nucleus. ALT is indicated by preferential colocalization of PML bodies with telomeres as well as a heterogeneous telomere lengths with some very long telomeres [1].

This work aims at assessing ALT in human soft tissue sarcomas. Using a confocal laser scanning microscope, about 400 three-channel 3D images of soft-tissue sarcomas were acquired, visualizing telomere spots, PML bodies and DAPI-stained nuclei in the first, second, and third channel, respectively (Fig. 1).

The main task of image analysis is to automatically detect and classify colocalizations of telomeres and PML bodies as well as to detect and quantify very large telomere structures.

In previous work on quantifying the ALT mechanism, colocalizations were estimated manually or semi-automatically, which is subjective, error prone, and time consuming [1]. Automatic approaches for colocalization analysis can be classified into intensity correlation-based and object-based schemes. With intensity correlation-based schemes, global colocalization measures are computed, which directly correlate the intensity information of two channels [2]. Main disadvantages are the sensitivity w.r.t. image contrast and noise. In contrast, with object-based schemes a segmentation of the relevant image structures is obtained, which is used to determine colocalizations [3, 4]. However, so far typically spot-like structures have been considered using, e.g., threshold-based schemes [3] or the wavelet transform [4]. Moreover, these approaches usually only consider distances between spots to determine colocalizations [3, 4].

In this contribution, we introduce a new model-based approach for automatic quantification of colocalizations and very large telomere structures in three-channel 3D microscopy images. Our approach is based on 3D parametric intensity models in conjunction with a model fitting scheme to quantify telomeres and PML bodies with high accuracy. 3D intensity models have been previously used to segment subcellular structures [5, 6]. However, so far these approaches only used 3D Gaussian models and they have not been used for the detection of colocalizations. The central idea of our approach is to determine colocalizations of subcellular structures between different channels based on the estimated geometry. In contrast to previous approaches [3, 4], our approach not only allows to accurately detect colocalizations but also to differentiate between different types of colocalizations. Furthermore, we perform an automatic statistical analysis to assess the significance of the determined colocalizations.



Telomeres PML bodies DAPI Enlarged section

Fig. 1. Maximum intensity projections of 3D microscopy images of a tumor.

2 Materials and Methods

2.1 3D Model-Based Segmentation and Quantification

Telomeres and PML bodies appear as bright spots on a diffuse background (Fig. 1). To segment these structures, we propose a two-step approach. In the first step, we apply different image analysis operations to the 3D images to obtain coarse center positions of the spots [6]. In the second step, we quantify each detected spot based on model fitting. For PML bodies and small and medium sized telomeres the 3D intensity profile can be well represented by a 3D Gaussian

$$g_{\mathrm{Gaussian3D}}(\mathbf{x}) = \exp\left(-\frac{x^2}{2\sigma_x^2} - \frac{y^2}{2\sigma_y^2} - \frac{z^2}{2\sigma_z^2}\right) \tag{1}$$

where $\mathbf{x} = (x, y, z)$ and $(\sigma_x, \sigma_y, \sigma_z)$ are the standard deviations. By including intensity levels a_0 (background) and a_1 (spot) as well as a 3D rigid transform $\mathcal{R}$ with rotation $\boldsymbol{\alpha} = (\alpha, \beta, \gamma)$ and translation $\mathbf{x}_0 = (x_0, y_0, z_0)$, we obtain the 3D parametric intensity model $g_M(\mathbf{x}) = a_0 + (a_1 - a_0)\, g_{\mathrm{Gaussian3D}}(\mathcal{R}(\mathbf{x}, \mathbf{x}_0, \boldsymbol{\alpha}))$. To fit the model to the image data, we use the non-linear optimization scheme of Levenberg/Marquardt. For each spot we obtain estimates of all model parameters with subvoxel resolution. However, for very large telomere structures (cf. the center spot in Fig. 1, right) the 3D intensity profile significantly deviates from a Gaussian shape. We propose to model these telomeres using a 3D superellipsoidal intensity model. Using $\Phi(x) = \int_{-\infty}^{x} (2\pi)^{-1/2}\, e^{-\xi^2/2}\, d\xi$, the model is given by

$$g_{\mathrm{SuperEllipsoid3D}}(\mathbf{x}) = \Phi\left(\frac{\sqrt[3]{r_x r_y r_z}}{\sigma}\left(1 - \sqrt[\epsilon]{\left|\frac{x}{r_x}\right|^{\epsilon} + \left|\frac{y}{r_y}\right|^{\epsilon} + \left|\frac{z}{r_z}\right|^{\epsilon}}\right)\right) \tag{2}$$

where σ is the standard deviation of Gaussian image blur, (r_x, r_y, r_z) are the semi-axes of the superellipsoid, and ϵ describes its roundness. The final 3D parametric intensity model is obtained by including intensity parameters and a 3D rigid transform analogously to the Gaussian model. To segment the telomere channel, we propose a two-pass approach. In the first pass, the above described two-step approach with an adaptation to large spots is applied, i.e., using the superellipsoidal model. In the second pass, the smaller spots are segmented using the two-step approach where the regions of the large spots are masked out to prevent false spot candidates at large spots. To segment the cell nuclei in the third channel, we apply 3D Gaussian smoothing with subsequent Otsu thresholding.

2.2 3D Colocalization Quantification and Classification

To determine colocalizations between subcellular structures in different image channels, we propose the following scheme. The idea is to utilize the quantified geometry of the structures based on the fitting results of the parametric intensity models, i.e., the 3D position, 3D orientation, and 3D shape (specified by the standard deviations and semi-axes of the Gaussian and superellipsoidal model,

respectively). We define a colocalization as an (partial) overlap of two (or more) structures from different channels. Since, in contrast to previous approaches [3, 4], we determine a geometric representation of the structures with subvoxel resolution, we are able to differentiate between different types of colocalizations. In our application, we use in total six different types. The types c_1 to c_5 (Fig. 2) represent colocalizations between PML bodies (grey) and small/medium sized telomeres (black). In addition, c_6 denotes a colocalization of a PML body with a very large telomere signal, which is quantified using the superellipsoid. Overall, we count the number n_{c_i} of colocalizations of type c_i and the total number n_c.

2.3 Statistical Analysis of the Significance of Colocalizations

To assess the significance of the estimated number of colocalizations, we perform a statistical analysis given the null hypothesis that the observed colocalizations are a random result of the spot distribution. This analysis is based on repeated random placements of the segmented telomeres and PML bodies within the segmented volume of the DAPI channel. For each placement, the number of colocalizations is determined as described above. For all r random placements, Σ is the number of placements that yield a number of colocalizations greater than or equal to the experimental result. The empirical p-value is given by $p = \Sigma/r$, which can be compared to a certain level of significance α. Since we apply our approach to a large number of $n \approx 400$ images, we use a Bonferroni correction: given the experiment-wide false positive value π, the applied level of significance is $\alpha = \pi/n$. Thus, we can state for each colocalization type in each image whether the number of detected colocalizations is significant or not.

3 Results

We have applied our approach to about 400 three-channel 3D fluorescence microscopy images of human soft-tissue tumors. The images have a size of about $512 \times 512 \times 25$ voxels. For all images, we used the same parameter settings. Overall, we achieved quite good segmentation results. For example, Fig. 2 shows a maximum intensity projection of the segmentation results for the region marked

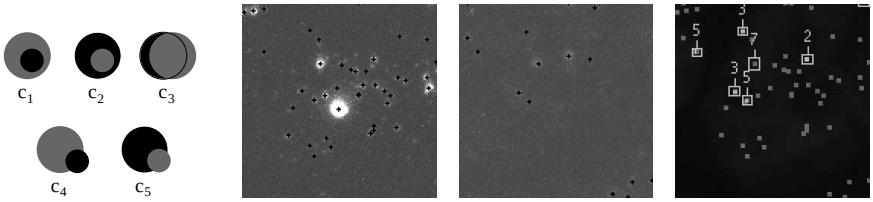


Fig. 2. Colocalization types c_1 to c_5 and results for the region shown in Fig. 1.

Table 1. Number of telomeres (n_{tel}, normal + very large) and PML bodies (n_{pml}) as well as total number n_c and numbers $n_{c_1}, \ldots, n_{c_6}$ of different colocalization types.

	n_{tel}	n_{pml}	n_c	n_{c_1}	n_{c_2}	n_{c_3}	n_{c_4}	n_{c_5}	n_{c_6}
Image 1	458+8	231	89	3	(1)	34	29	16	6
Image 2	479+5	181	22	(0)	(0)	(3)	(11)	(1)	7
Image 3	742+1	331	(16)	(0)	(0)	(0)	(16)	(0)	(0)

in Fig. 1 (left). Shown are the center positions of the segmented telomeres and PML bodies as well as the detected colocalizations together with their type.

For the image shown in Fig. 1 and for two additional images, Tab. 1 shows the number n_{tel} of segmented telomeres (normal + very large), the number n_{pml} of PML bodies, the total number n_c of colocalizations, as well as the numbers $n_{c_1}, \ldots, n_{c_6}$ of different types of colocalizations. Colocalization results which are not significant have been put in parenthesis. It can be seen that the number of colocalizations as well as the significance of these colocalizations is very different in the three images. Interestingly, whereas the number of segmented telomeres and PML bodies in the first two images is similar, the total number of colocalizations n_c differs by a factor of about 4. Still, the total number of colocalizations $n_c = 22$ of image 2 is significant. Considering individual types of colocalizations, it can be seen that the level of significance varies between different types. For example, for image 2 a number of $n_{c_4} = 11$ colocalizations is not significant, whereas a number of $n_{c_6} = 7$ colocalizations is significant. As a consequence, the statistical analysis of the significance of the colocalizations is important to allow an accurate analysis of the colocalization results.

4 Discussion

We introduced a new approach for automatic quantification of colocalizations and very large telomere structures in 3D microscopy images. The approach is based on 3D parametric intensity models in conjunction with a model fitting scheme to quantify telomeres and PML bodies with high accuracy. With our approach we determine colocalizations based on the estimated geometry (position and shape) of subcellular structures and differentiate between different types of colocalizations. Furthermore, we performed a statistical analysis to assess the significance of the colocalizations. We have successfully applied our approach to about 400 three-channel 3D fluorescent microscopy images of soft-tissue tumors. Future work is the analysis of the colocalization results w.r.t. biological meaning.

References

1. Fasching, CL et al. DNA damage induces alternative lengthening of telomeres (ALT)-asociated promyelocytic leukemia bodies that preferentially associate with linear telomeric DNA. Cancer Res. 2007;67:7072–7.

2. Manders EMM, Verbeek FJ, Aten JA. Measurement of co-localization of objects in dual color confocal images. J Microsc. 1993;169:375–82.
3. Fay FS, Taneja KL, Shenoy S, et al. Quantitative digital analysis of diffuse and concentrated nuclear distributions of nascent transcripts, SC35 and Poly(A). Exp Cell Res. 1997;231:27–37.
4. Zhang B, Chenouard N, Olivo-Marin JC, et al. Statistical colocalization in biological imaging with false discovery control. In: Proc ISBI; 2008. p. 1327–30.
5. Thomann D, Rines DR, Sorger PK, et al. Automatic fluorescent tag detection in 3D with super-resolution: application to the analysis of chromosome movement. J Microsc. 2002;208:49–64.
6. Heinzer S, Wörz S, Kalla C, et al. A model for the self-organization of exit sites in the endoplasmic reticulum. J Cell Sci. 2008;121(1):55–64.