Visualization and Exploration of 3D Toponome Data

Steffen Oeltze*, Paul Klemm, Reyk Hillert, Bernhard Preim, Walter Schubert
Otto-von-Guericke University of Magdeburg

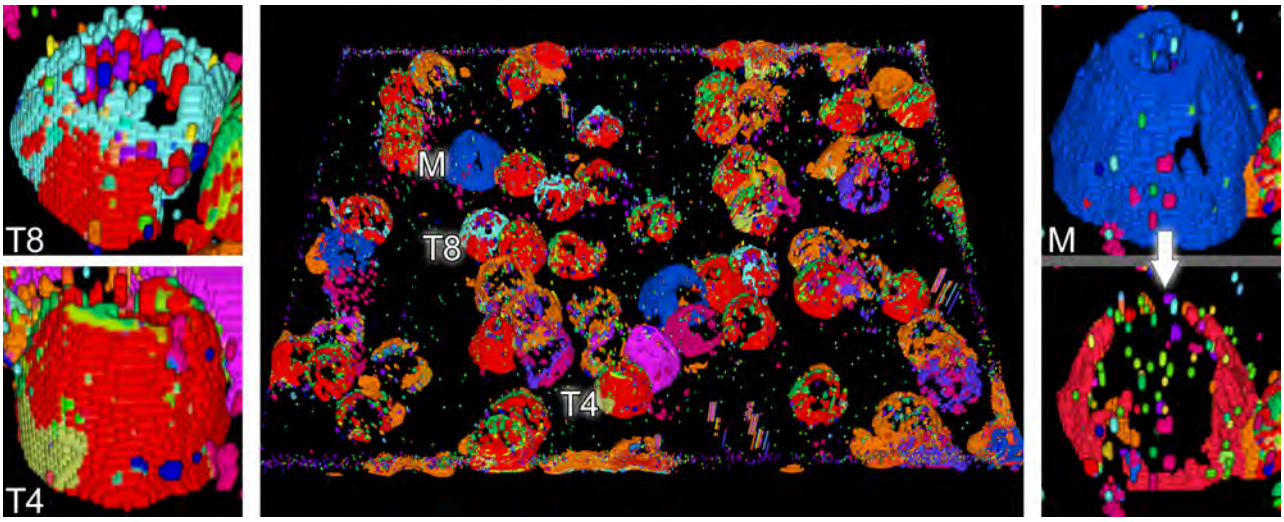


Figure 1: Toponome data showing cell surface protein clusters in a blood sample. Each color represents a different cluster of distinct co-localized proteins. In the sample, for instance T4 and T8 lymphocytes as well as monocytes (M) appear as hemisphere-shaped structures. Interactively "peeling off" clusters provides insight into the composition of a cell's surface, as illustrated for the blue cluster of the monocyte.

Abstract

The *toponome* of a cell describes the location and topological distribution of proteins across the cell. In *toponomics*, the toponome is imaged and its inner structure and its semantics are investigated in order to understand how cells encode different functionalities both in health and disease. Toponome imaging results in complex multi-parameter data composed of a 3D volume per protein affinity reagent. After imaging, the data is binarized such that 1 encodes protein present and 0 encodes protein absent. Biologists are particularly interested in the clustering of these binary protein patterns and in the distribution of clusters across the cell.

We present a volume rendering approach for visualizing all unique protein patterns in 3D. A unique color is dynamically assigned to each pattern such that a sufficient perceptual difference between colors in the current view is guaranteed. We further present techniques for interacting with the view in an exploratory analysis. The biologist may for instance "peel off" clusters thereby revealing occluded cell structures. The 3D view is integrated in a multiple coordinated view system. Peeling off clusters or brushing protein patterns in the view updates all other views. Brushing and linking supports the biologist in deciphering the toponome. We demonstrate our approach for a cell probe containing lymphocytes.

1 Introduction

While the human genome project has revealed, among other things, the code for all proteins, the next big challenge is to understand how proteins cooperate in cells and tissues in time and space [3]. Although many details on the molecular function and structure of many proteins are known, their corresponding cellular functions cannot simply be derived. This is due to a dependence of the function on the contextual position of a given protein within a protein network. The *toponome* of a cell describes the location and topological distribution of proteins across the cell. It has been shown that the toponome is hierarchically organized [4]. It comprises protein clusters which contain *lead proteins* and are interlocked as a network. The lead proteins control the topology of the clusters and their function as a network. In *Toponomics*, the toponome is imaged, explored and analyzed for applications in toxicology, drug development and patient-drug-interaction. The most advanced toponome imaging technique is robot-driven multi-parameter fluorescence microscopy TIS^{TM}[1]. It is capable of co-mapping hundreds of proteins in a single a cell or tissue sample. The imaging procedure can be performed in 3D and then, results in complex multi-parameter data composed of a 3D volume per protein affinity reagent. In a post-processing step, the data is binarized by an expert such that 1 encodes protein present and 0 encodes protein absent. For each data voxel, a binary vector can be constructed over all volumes, which encodes the local co-mapping of distinct proteins. The unique binary codes that exist in the data, out of all possible codes, are referred to as *combinatorial molecular phenotypes* (CMPs). The binary code that corresponds to a certain CMP very often exists at several voxel positions. In reality, these positions are not randomly spread over the data but clustered at certain locations of the cells. The biologist is interested in these protein clusters since they correspond to functional units of the cells controlling cellular functionalities. He is further interested in answering questions such as: Which proteins co-map?, What is the frequency of co-mapping?, Where across the cell surface do the binary protein patterns cluster?, and How does the clustering differ from cell to cell? These questions guide our research.

*e-mail: stoeltze@isg.cs.uni-magdeburg.de

2 Data

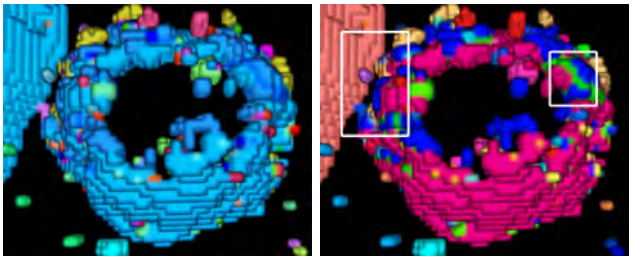
Our methods have been applied to a cell probe containing blood lymphocytes of a healthy subject. In the imaging of this probe, a large tag library, containing monoclonal antibodies directed against cluster of differentiation (CD) marker proteins, was used. CD marker proteins are expressed, e.g., on the surface of immune cells such as lymphocytes. In this study, 32 CD surface proteins have been co-mapped on the probe by using a TIS^{TM}robot system [1]. The probe has been imaged at 20 different slice locations with a matrix of 658×517 pixels, an in-plane-resolution of 216×216 nm, and a slice distance of 200 nm. The data has been binarized by the collaborating biologist and 2167 CMPs have been derived in the subsequent post-processing.

3 Visualization and Exploration Methods

We present a volume rendering approach that generates an integrated 3D visualization of all binary protein affinity reagent volumes. The integration is accomplished by deriving all unique binary vectors (CMPs) from the data and then, assigning a unique color to each CMP. Instead of using a color transfer function, this color is pre-computed and applied directly in the rendering process. A toponome dataset is likely to contain several thousand CMPs. While the generation of a different color for each CMP is technically possible, visually discriminating these colors by far exceeds the capabilities of the human visual system. Hence, we offer a dynamic color range distribution which may be triggered by the biologist for the current view on the scene, e.g., a close-up view of a specific cell (Fig. 2). The algorithm then generates a set of unique colors having a sufficient perceptual difference based on the currently visible CMPs. The colors are chosen iteratively from a precomputed color pool such that there difference in the perceptually linear CIELab color space is maximum. Thus, the visual differentiability of the corresponding protein patterns is improved.

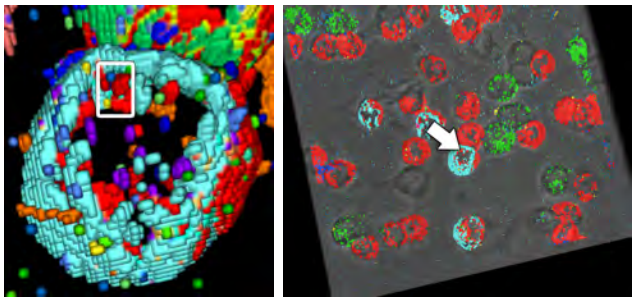
Our approach has been implemented employing the open-source Visualization Toolkit (VTK) and the included GPU-based raycasting. The 3D view is integrated via *linking* in a multiple coordinated view system, which is used by our collaborators in their daily routine. The system consists of a filter view for defining a template CMP by selecting and deselecting proteins, a table view, which lists all CMPs as rows, a 2D slice view, a recently added graph view [2], and our new 3D view. All views are linked with each other and equipped with brushing facilities.

To support the biologist in an exploratory analysis of the toponome data, the 3D view is equipped with several interaction techniques. For instance, the biologist may inspect the binary patterns of a cell by hovering the mouse pointer over its 3D representation. During mouse hovering, a ray is cast into the scene, the first hit non-background voxel is reported, the corresponding CMP is extracted,



(a) (b)

Figure 2: *(a)* Close-up view of a cell and a neighboring cell (upper left). A large CMP cluster appears turquoise in both cells. *(b)* A color range distribution for the same view reveals the true variety of CMPs (insets). It shows that the visible CMP cluster of the neighboring cell differs from the largest cluster of the focused cell.



(a) (b)

Figure 3: Similarity brushing of a T8 lymphocyte. *(a)* The biologist brushes an interesting part of the cell surface in screen space. *(b)* The corresponding CMPs are determined across all slices and the 3D visualization is restricted to these CMPs. A reference volume (grayish) serves as spatial context.

and the names of its mapped CD marker proteins are displayed at the mouse pointer position.

While the biologist samples the cell's surface, he mentally constructs a molecular "face" of the cell which he later compares to the "face" of other cells, e.g., for detecting (pathological) variations. We further support this comparison by similarity brushing (Fig. 3). Here, a cell surface part is brushed in screen space by the biologist, the corresponding CMPs are determined via ray-casting and sampling across all slices, and the visualization in all views (including the 3D view) is restricted to these CMPs. Thus, regions containing at least one of the brushed CMPs are revealed.

A common problem in 3D data visualization is occlusion. Transferred to toponome data, cell regions corresponding to one CMP occlude others that correspond to another CMP. Hence, we implemented a CMP peeling interaction (Fig. 1, two rightmost images). The desired CMP is selected by means of point-and-click and then, "peeled off" by rendering it fully transparent. Further, it is automatically deselected in all other views of the framework. The peeling step may be carried out any number of times. We enhance the usability of this technique by offering an undo/redo mechanism which is operated via the graphical user interface of the 3D view.

4 Summary

The presented approach for visualizing 3D toponome data enables our collaborators to produce an overview visualization in half a minute, as compared to 1-2 workdays before. We integrated the 3D visualization into their in-house analysis framework. In this framework, a unique color serves as CMP identifier. We improved the existing coloring scheme by dynamically computing a perceptually optimized set of colors. To support an exploratory analysis, the 3D view was equipped with several interaction techniques.

References

[1] M. Friedenberger, M. Bode, A. Krusche, and W. Schubert. Fluorescence Detection of Protein Clusters in Individual Cells and Tissue Sections by Using Toponome Imaging System: Sample Preparation and Measuring Procedures. *Nat Protoc*, 2(9):2285–2294, 2007.

[2] S. Oeltze, W. Freiler, R. Hillert, H. Doleisch, B. Preim, and W. Schubert. Interactive, Graph-Based Visual Analysis of High-Dimensional, Multi-Parameter Fluorescence Microscopy Data in Toponomics. *IEEE Trans Visual Comput Graph*, 17(12):1882–1891, 2011.

[3] W. Schubert. On the Origin of Cell Functions Encoded in the Toponome. *J Biotechnol*, 149(4):252–259, 2010.

[4] W. Schubert, B. Bonnekoh, A. J. Pommer, L. Philipsen, R. Bckelmann, Y. Malykh, H. Gollnick, M. Friedenberger, M. Bode, and A. W. M. Dress. Analyzing Proteome Topology and Function by Automated Multidimensional Fluorescence Microscopy. *Nat Biotechnol*, 24(10):1270–1278, 2006.