

Interactive, Graph-Based Visual Analysis of High-Dimensional, Multi-Parameter Fluorescence Microscopy Data in Toponomics

Steffen Oeltze, *Member, IEEE*, Wolfgang Freiler, Reyk Hillert, Helmut Doleisch, *Member, IEEE*, Bernhard Preim, and Walter Schubert

Abstract—In *Toponomics*, the function protein pattern in cells or tissue (the *toponome*) is imaged and analyzed for applications in toxicology, new drug development and patient-drug-interaction. The most advanced imaging technique is robot-driven multi-parameter fluorescence microscopy. This technique is capable of co-mapping hundreds of proteins and their distribution and assembly in protein clusters across a cell or tissue sample by running cycles of fluorescence tagging with monoclonal antibodies or other affinity reagents, imaging, and bleaching *in situ*. The imaging results in complex multi-parameter data composed of one slice or a 3D volume per affinity reagent. Biologists are particularly interested in the localization of co-occurring proteins, the frequency of co-occurrence and the distribution of co-occurring proteins across the cell.

We present an interactive visual analysis approach for the evaluation of multi-parameter fluorescence microscopy data in toponomics. Multiple, linked views facilitate the definition of features by brushing multiple dimensions. The feature specification result is linked to all views establishing a focus+context visualization in 3D. In a new attribute view, we integrate techniques from graph visualization. Each node in the graph represents an affinity reagent while each edge represents two co-occurring affinity reagent bindings. The graph visualization is enhanced by glyphs which encode specific properties of the binding. The graph view is equipped with brushing facilities. By brushing in the spatial and attribute domain, the biologist achieves a better understanding of the function protein patterns of a cell. Furthermore, an interactive table view is integrated which summarizes unique fluorescence patterns. We discuss our approach with respect to a cell probe containing lymphocytes and a prostate tissue section.

Index Terms—Visual Analytics, Fluorescence Microscopy, Toponomics, Protein Interaction, Graph Visualization.

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 [30]. The *toponome* of a cell describes its function protein pattern, i.e. the location and topological distribution of proteins. In *Toponomics*, the toponome is imaged, explored and analyzed for applications in toxicology, drug development and patient-drug-interaction. In the traditional fluorescence microscopy, a maximum of five proteins may be mapped concurrently. With a sophisticated flow cytometer up to 17 proteins may be mapped [23]. However, this number is still insufficient for mapping protein network features. The most advanced imaging technique is robot-driven multi-parameter fluorescence microscopy [28]. It is capable of co-mapping hundreds of proteins and their distribution across a cell or tissue sample by running cycles of fluorescence tagging with monoclonal antibodies or other affinity reagents, imaging, and bleaching *in situ*. The imaging results in complex multi-parameter data composed of one slice or a 3D volume per affinity reagent. In a preprocessing step, the data is binarized such that 1 encodes protein present and 0 encodes protein absent. Biologists are particularly interested in the localization of co-occurring proteins, i.e. co-occurring affinity reagent bindings, the frequency of co-occurrence and the distribution of co-occurring proteins across the cell. This information is crucial in order to understand how proteins cooperate in cells and tissues in time and space.

We present an interactive visual analysis approach for the evaluation of multi-parameter fluorescence microscopy data in toponomics. Multiple, linked views facilitate the definition of features by brushing

multiple dimensions. The feature specification result is linked to all views establishing a focus+context style of visualization in 3D. In a new attribute view, we integrate techniques from graph visualization. Each node in the graph represents an affinity reagent while each edge represents two co-occurring affinity reagent bindings. The graph visualization is enhanced by glyphs which encode specific properties of the binding. The graph view is equipped with brushing facilities. By brushing in the spatial and attribute domain, the biologist achieves a better understanding of the function protein patterns of a cell. Additional attributes, such as local binding frequency, are derived from the data and made accessible from other attribute views, e.g., histograms and scatter plots. Furthermore, an interactive table view summarizes unique fluorescence patterns existing in the data. We discuss our approach with respect to a cell probe containing lymphocytes and a prostate tissue section. The probe has been stained with 32 protein affinity reagents (PARs) and imaged at 20 different slice locations. The tissue section has been stained with 17 PARs and imaged at 1 slice location. So far, we could test our approach only for the investigation of these two datasets since datasets from other studies were not available in the present cooperation and the temporal and financial costs which are involved in the imaging are high (sample preparation and imaging together often require several days [10]). However, at the end of this paper, we will elaborate on the generality of our approach including its scalability.

2 BIOLOGICAL AND TECHNICAL BACKGROUND

This section gives an overview on the biological background and familiarizes the reader with the imaging technique that we apply for mapping the toponome.

2.1 Toponomics

Toponomics requires new technologies which are able to co-localize a quasi random number of different proteins in the one biological sample in order to map what any cell does in reality: it forms functional protein patterns (assemblies of clusters of different proteins) to generate concrete cell functions. Hence, the cell is a protein pattern formation apparatus [28, 30]. This machinery, the whole functional plan of the cell, still poses many open questions. Neither the really existing pro-

- Steffen Oeltze, Reyk Hillert, Bernhard Preim and Walter Schubert are with the University of Magdeburg, E-mail: {oeltze|preim}@ovgu.de, {reyk.hillert|walter.schubert}@med.ovgu.de.
- Wolfgang Freiler and Helmut Doleisch are with SimVis GmbH, Vienna, E-mail: {freiler|doleisch}@simvis.at.

Manuscript received 31 March 2011; accepted 1 August 2011; posted online 23 October 2011; mailed on 14 October 2011.

For information on obtaining reprints of this article, please send email to: tvcg@computer.org.

tein clusters in given functional states of a cell *in vivo/in situ*, nor the rules of their formation can be derived from genomic or pure molecular protein data. Although many details on the molecular function and structure of many proteins are known, we cannot simply derive the corresponding cellular functions of these proteins, because the latter are dependent on the contextual position of a given protein within a protein network inside the cell [28, 29, 30, 31, 32, 33]. It is therefore essential to distinguish the molecular function from the cellular function of a protein, defined by the spatial protein context cell by cell. The entirety of all protein networks, in which proteins are defined by their protein-to-protein context in any given cell, is defined as the *toponome* [10, 28, 32]. The toponome is a system of proteins. Its inner structure, its biological code, and its semantics are investigated in *toponomics*.

The toponome in cells is hierarchically organized: protein clusters interlocked as a network contain *lead protein(s)* that control the topology of the protein clusters and their function as a whole network [32]. This has been clearly demonstrated by knocking down the lead protein or inhibiting it by a chemical agent which result in a disassembly of the whole protein cluster network and an essential alteration of the cellular function that is encoded by this network [32]. Neither a lead protein nor the protein cluster which is controlled by it can be predicted from molecular data. Thus, the toponome must be mapped in human tissues and human or animal cells to understand how cells encode the myriads of different functionalities both in health and disease. Many investigations have shown that mapping the toponome is essential for finding new drugs in cancer and for finding protein clusters that can be regarded as a new system of biomarkers in disease [2, 30, 34].

2.2 Imaging the Toponome

Imaging the toponome is based on a cyclical imaging procedure in which a *tag library* (specific affinity reagents recognizing proteins) is conjugated to one and the same dye. A toponome imaging system (TIS) robot [10] applies these tags sequentially in the following way: the first dye-conjugated tag is applied to a fixed cell or tissue sample, the resulting fluorescence image is registered by an epifluorescence microscope which is connected to a CCD camera, and the dye is then bleached gently to avoid any energy transfer into the remaining proteins [27, 28, 32]. This first incubation imaging bleaching cycle is followed by a second incubation imaging bleaching cycle using a second tag reacting with the second protein of interest, and so on. This imaging procedure can be performed in 2D or 3D. Up to 100 cycles to label 100 proteins and co-map these proteins at any given sub-cellular data point in a cell have been demonstrated [32].

A wide-spread method for the rapid analysis of toponome data is thresholding each protein fluorescence signal thereby generating a combinatorial binary code, where 0 indicates protein absent and 1 indicates protein present [1, 9, 28, 32]. In case of a 3D dataset, this results in a binary code at each voxel. The size of the code equals the number of applied tags. All binary codes that exist in the data, out of all possible combinations of 0 and 1, can be assembled in a *toponome map* and be referred to as combinatorial molecular phenotypes (CMPs). In a toponome map, a unique color is assigned to each CMP. Since the corresponding binary code frequently occurs at several locations, the map visualizes the location of given protein clusters present in groups of pixels or voxels. We introduce the term *single-1-CMP* which refers to a CMP which contains only a single element equal 1.

3 PRIOR AND RELATED WORK

This section describes prior and related work on the application of graph visualization techniques for investigating biological data and on the visual analysis of such data by employing multiple coordinated views. It starts with a brief discussion on how our collaborating biologists have been analyzing their data so far.

After imaging the toponome and thresholding the data, the CMPs are determined and added as rows to a table. Each row then represents a group of pixels/voxels having the same binary code. The table may be sorted according to the columns which represent the employed affinity reagents. The table is linked to a toponome map such that the user may select individual CMPs and observe their location in the data

or define a region of interest in the toponome map thereby restricting the table to the corresponding CMPs. A limitation of this setup is the ability to explore a 3D dataset as a whole volume instead of slice by slice. Furthermore, it is impossible to select individual co-occurring PAR bindings across different CMPs and the frequency of co-occurrence is very difficult to infer from just the table. The distribution of CMPs across the cell may be recognized in the toponome map whereas the distribution of individual co-occurring proteins remains hidden in the data.

3.1 Graph Visualization of Biological Data

The most frequent use of graphs in biology is the visualization of biological networks representing the metabolism of cells, the regulation of genes and the interaction of proteins. Graph visualizations of such networks encode biochemical processes and provide an important means to understand their complex nature. A survey of visualization tools for biological network analysis is given in [14]. One prominent example is *Cytoscape* [36]. A central component of the software is the graph view which represents molecular species as nodes and intermolecular interactions as edges. Several graph layout algorithms have been integrated including a circular layout. Another popular software for studying pathways, gene regulation and systems biology is *VisANT* [18]. It is implemented as an on-line tool and integrated with standard databases for organized annotation. All information gathered in VisANT may be visualized in a graph. Among others, a circular layout has been integrated. The usefulness of circular graph layouts for visualizing expressions of genes across a set of patients has been investigated in [37]. Genes are clustered and the resulting clusters are visualized as a set of circles. Edges reflecting the correlation between two genes are drawn inside a cluster and across clusters. Special care is taken to generate a visually pleasing layout of the set of circles. An extension of circular graph layouts to 3D is integrated in the software *Arena3D* [22]. Proteins or pathways can be iteratively grouped and the result of each iteration can be arranged on a separate higher layer. This results in a hierarchical 3D layout where for each layer, e.g., a circular layout may be employed. The described tools are very diverse, but each of them provides among others a circular graph layout since this results in a compact view of the investigated network. Furthermore, the modification of a node's and an edge's visual properties according to a biological aspect as well as the ability to interact on the graph, e.g., select a subset of nodes, are essential.

3.2 Visual Analysis of Biological Data

DeLeeuw et al. present the *ARGOS* system for interactive exploration and batch processing of confocal laser microscopy data [5]. They combine views of the image data with attribute views, e.g. parallel coordinates, scatter plots and histograms. The views are realized as separate windows which leads to maximum flexibility at the expense of guidance. Weber et al. present results of a long-term effort of exploring gene regulatory networks by means of relating them to 3D gene expression data [38]. Thus they had to find visualization techniques which convey the spatial position of gene expression data and to relate these visualizations to attribute views. For this purpose, they created different views on the spatial data including those where they unroll the data to a cylindrical shape and project it in 2D. To relate the spatial data to the attribute views, they developed a query-mechanism. There are a number of other techniques and systems where multiple views are coordinated to investigate biological data. However, most of them do not incorporate image data. In [19], a multiscale synteny browser (MizBee) for the comparison of two genomes is presented. MizBee incorporates linked genome, chromosome, and block views. The genome view applies a double circular graph layout that opposes source and destination chromosomes. Schulz et al. developed special graph visualization techniques for bipartite biological networks, a special kind of network which is frequently used to study findings in biology [35]. The *SpRay*-system supports the exploration of gene expression data with various InfoVis views [6]. To the best of our knowledge, no system exists that can readily process our data and link the image data to the derived information.

4 METHODS

In this section, we provide details on our visual analysis framework and on the integration of the graph and the table view. We put special emphasis on the graph view, in particular, the graph layout, the modification of an edge's visual attributes, the integration with glyph drawings, and the equipment with brushing facilities.

4.1 The Visual Analysis Framework

The interactive visual analysis is carried out in a framework employing the SimVis technology [7]. SimVis has been developed for the analysis of multi-parameter data. It has originally been engineered for the analysis of 3D time-dependent flow simulation data but has been extended to cope with data measured on structured grids, e.g., 3D/4D medical data [20, 21]. Recently, a structured grid ray-casting renderer has been integrated. In the framework, multiple views from scientific and information visualization may be linked together, e.g., 3D view, histogram, scatter plot, and parallel coordinates. By linking these views, the framework supports the concurrent exploration of the observation space (3D view) and the attribute space (InfoVis views). The user may choose an attribute for representation in one of the InfoVis views and then, brush an interesting range of attribute values, i.e. specify a feature. The result of such a brushing operation is reintegrated into the data in form of a synthetic data attribute $DOI_j \in [0, 1]$ (*degree of interest (DOI)*, compare to [12]). This DOI attribution is used in the views of the analysis setup to visually discriminate the interactively specified features from the rest of the data in a focus+context visualization style which is consistent in all (linked) views [15].

There are several interactive visual analysis tasks. In a *feature localization*, the user searches for places in the 3D/4D domain where certain feature characteristics are present. The user can brush features in attribute views and concurrently localize the respective feature in the observation space. In a *multi-variate analysis*, multi-variate data properties are investigated by specifying a feature in one attribute view and at the same time analyzing the DOI distribution with respect to other data attributes in other attribute views (through view linking). In a *local investigation*, the user inspects the values of selected data attributes with respect to certain spatiotemporal subsets of the 3D volume domain. In SimVis, the user can also load spatial as well as temporal data references into attribute views—brushing these kinds of data attributes then yields features which are specified according to their spatiotemporal extents. Each feature is described in SimVis by a feature characteristic and organized in a feature set:

- *feature set*: subsumes an arbitrary number of features (logical OR combination of features), only one feature set is used by the framework at the time
- *feature*: specified by one or more feature characteristics (logical AND combination of DOI functions of all feature characteristics)
- *feature characteristic*: simple (attribute+bounds) or complex (AND/OR/NOT of several brushes) feature characteristic defined on an attribute view

Each attribute view is associated to a feature characteristic and hence, to a feature. SimVis uses standard colors for the geometric primitives in an attribute view, e.g., scatters in a scatter plot, thereby indicating the belonging of the corresponding data points to features:

- *red*: belong to the associated feature and all other features in the current feature set
- *green*: do not belong to the associated feature but to at least one other feature in the current feature set
- *yellow*: do neither belong to the associated feature nor to any other feature in the current feature set
- *dark gray*: only exists if ≥ 1 brush is defined for the associated attribute view; then, corresponding data points do not belong to the associated feature

A mixing of color (averaged DOI) may occur when a geometric primitive in an attribute view represents multiple data points. For example, a simple range of values has been brushed on one attribute in a histogram. In a linked scatter plot, two other attributes are opposed. As a result of the brushing, pure red, pure yellow as well as scatters in various shades of orange appear. A mixed color scatter indicates that some of the corresponding data points exhibit values within the brushed range while others do not. The percentage of each separate color is proportional to the respective number of data points. In all attribute views, color mixing may be turned on and off. If it is off, the drawing priority in decreasing order is red, green, yellow, gray (see [20] for more details on the colors and their mixing).

4.2 The Graph View

The co-occurrence of proteins across the cell and the frequency of co-occurrence are of high interest for biologists. Also the spatial distribution of co-occurring proteins and a comparison of this distribution across cells are crucial. The co-occurrence is represented in the data by pairs of 1s in the binary codes and each such pair may be considered as a relation. The relation is not directional and since PARs are not grouped together, no hierarchy exists in the data. Graphs offer a comprehensive and comprehensible representation of relational data and are frequently used in bioinformatics [14]. Transferred to our binary data, each node in an undirected graph may represent a PAR and each edge may represent two co-occurring PAR bindings, i.e. a pair of 1s in a binary code. The frequency of co-occurrence should be encoded by a visual property of the edge, e.g., edge color or edge width. Understanding the spatial distribution of co-occurring PAR bindings would be supported by integrating the graph into the visual analysis framework as an additional attribute view. Furthermore, the new view may be equipped with brushing facilities, e.g., for node and edge selection, and the brushing result should be reintegrated into the data as DOI thereby linking the graph view to all other views. This allows to select an edge with an associated high co-occurrence frequency and then, show all sites in the 3D view where the two corresponding PARs are concurrently binding. Further, by also updating the graph view according to the modified DOI values, PAR bindings which coincide with the selected PARs are highlighted, i.e., edges connecting these PARs with the selected PARs are emphasized.

Requirements. The requirements on the graph view are derived from many discussions with the biologists. The graph layout should:

- enable a fast recognition of a PAR of interest
- enable a quick inference of which other PARs are connected by an edge to a PAR of interest, i.e., co-occurring PAR bindings
- be static, i.e. node positions must not change during a DOI update in order to guarantee a fast PAR recognition
- avoid visual clutter by minimizing the number of edge crossings

The following information should be communicated by the graph:

- PAR bindings do/do not co-occur
- Co-occurrence frequency in the feature and in the context
- SimVis standard colors (see Sec. 4.1)
- PAR name and frequency of the PAR binding in the feature
- frequency of single-1-CMPs with a 1 for this PAR in the feature
- degree of the node corresponding to the PAR

With regard to interactivity, the graph view should support:

- a repaint of the graph at interactive frame rates when the feature specification has been updated
- brushing in order to include PARs (nodes) in the feature (AND/OR) and to exclude PARs (NOT)

Development Environment. The graph view is implemented in C++ and added to SimVis as a plug-in. The graphical user interface of SimVis and hence, of the plug-in are realized in Qt. The information visualization classes of VTK are used for implementing the graph-related functionality. VTK provides most of the relevant graph features, such as different layout strategies. The most important VTK classes which have been employed are `vtkMutableUndirectedGraph` and `vtkGraphLayoutView`. While the first implements a data structure representing an undirected graph, the latter implements functionality for laying out and drawing a graph. Moreover, a plenty of new functionality has been added for customizing the graph drawing and augmenting its expressiveness by integrating additional information. A crucial step in SimVis is the continuous traversal of all voxels while the feature specification is being changed in order to determine those which do/do not adhere to the specification. In order to offer interactive frame rates, we have parallelized the voxel traversal using OpenMP [3]. Since SimVis processes the data block-wise, the parallelization of the traversal is relatively straightforward by letting each thread process a block. For the graph and the table view, a thread-safe computation of a graph and a list of CMPs per block and a final merging step have been implemented.

Preprocessing. Preprocessing includes the determination of the edges that exist in the graph out of all theoretically possible edges, the determination of the frequency of each PAR binding, i.e., the *edge-weight*, and for each PAR, the computation of the number of single-CMPs. The latter information is not inherently represented by the graph and is therefore added to the view.

At first, the PARs are determined which bind at neither position in the cell probe, i.e. whose corresponding binary images are zero everywhere. These PARs are excluded from further processing. They are presented in a text inset of the graph view (Fig. 3 (b)).

Next, several arrays are initialized. The array $[EW]$ is going to contain the edge-weights. Its size is set to the number of possible graph edges: $n(n-1)/2$, where n is the number of PARs. The arrays $[PAR]$ and $[PAR1]$ are going to contain the PAR binding-related information. The size of the arrays is set to the number of PARs. Finally, an array $[CN]$ is initialized which is going to contain indexes into the nodes connected by an edge in $[EW]$. $[CN]$ is filled with all possible index pairs and its size is set to two times the size of $[EW]$. Next, a nested `for`-loop begins. The outer loop iterates over all voxels (data points will be referred to as voxels in the following) whereas the inner loop iterates over all binary attributes, i.e. PARs. Within the inner loop, the current binary code is reconstructed and all pairs of 1s are determined without repetition and inversion of CMP elements. The elements of $[PAR]$ are incremented at those positions where the elements of the code equal 1. If the code contains just a single 1, $[PAR1]$ is updated accordingly. Next, the index into $[EW]$ is computed for each pair of ones and the corresponding element of $[EW]$ is incremented. At the end of the `for`-loop, $[EW]$ and $[CN]$ are squeezed such that zero elements are deleted.

An additional step, which is carried out within the inner part of the `for`-loop, is the filling of a data structure which will speed up the repaint of the graph on a feature specification update. The new data structure is realized as a Vector, a special type of dynamic array. It is initialized before the outer loop starts and its size is set to the size of the dataset. Within the inner loop, all indexes into $[EW]$ which have been computed for the current voxel, are stored at the corresponding voxel position in the Vector. They are again stored in the form of a Vector v_{sub} . The majority of voxels in the lymphocytes dataset represent protein absent. Here, no indexes are computed resulting in a sparse Vector containing many Vectors v_{sub} of size zero. On the repaint of the graph, all voxels are traversed again and for the non-background voxels, the corresponding indexes into $[EW]$ have to be determined. By means of the Vector, only a readout of each v_{sub} is required. Without the Vector, this would require recomputing the indexes which in turn involves touching every binary attribute as described above. This is particularly ineffective when the voxel represents background which may easily be inferred from v_{sub} by just checking its size.

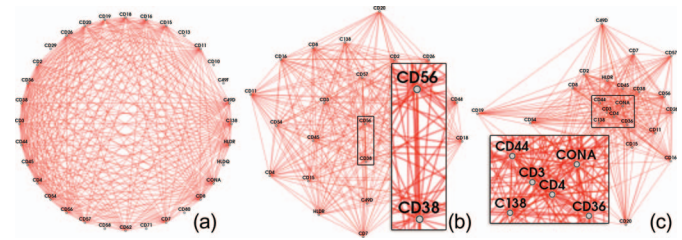


Fig. 1. Different graph layouts: circular (a), force-directed (b), and force-directed under consideration of edge frequency (c). The insets illustrate the strong overlap of edges in (b) and the aggregation of nodes which are connected by edges with a high frequency in (c).

Graph Layout. Two strategies have been tested for laying out the graph in 2D: a circular layout and a force-directed layout [11]. In a circular layout, all nodes are uniformly arranged on a circle (Fig. 1 (a)). In a force-directed layout, the nodes are arranged such that the variation in edge length as well as the number of edge crossings are minimized and a symmetric layout is achieved (Fig. 1 (b)). In addition, edge-weights may be considered such that the nodes connected by an edge with a high weight are positioned close together, e.g., CD3 and CD4 which are connected by the most frequent edge in Figure 1 (c).

Considering the above mentioned requirements on a graph layout, the circular layout performs better than the force-directed layout except for the number of edge crossings. The PARs are easier to find since their position follows an obvious rule and they are ordered alphabetically starting at 3 o'clock. In the force-directed layout, some edges strongly overlap which hampers a separate tracing (inset of Fig. 1 (b)). In an interactive analysis, edge-weights are constantly recomputed with respect to the updated feature characteristics. If edge-weights influence the positioning of nodes, the force-directed layout then constantly changes. This hampers a fast recognition and a visual tracking of nodes representing PARs of interest.

Edge Color. After the nodes of the graph have been laid out in a circular fashion, edges need to be constructed. For each edge in $[EW]$, the corresponding nodes are retrieved from $[CN]$ and a line is drawn between them. The edge color is modified according to the SimVis standard colors (Fig. 2 (a), recall Sec. 4.1). The selected color depends on the voxels which contain this edge and their belonging to the features in the feature set. As can be seen, it is difficult to visually group edges of equal color due to the large number of edge crossings. However, early feedback from the biologists indicated that mostly only edges being part of the feature set are of interest, i.e. red and green edges. Furthermore, green edges are only interesting in a comparison of different cells or different cell parts which must be described by different features. In order to account for this feedback, the opacity of the edges is initially modulated and may be further adjusted by the user. Red is assigned full opacity, while green and yellow are assigned an opacity of 50% and 20%, respectively (Fig. 2 (b)).

Edge Width. We decided to map the edge-weights in $[EW]$ to the width of the graph edges. Width is chosen instead of color or opacity since the perception of quantities depicted by area is more accurate [4]. Unfortunately, VTK only offers a uniform change of edge width which requires a workaround. Each line representing an edge is replaced by a ribbon that is generated in 2D by means of `vtkRibbonFilter` (Fig. 2 (c)). The width of the ribbons is modified according to the values in $[EW]$. All values are normalized with respect to an empirically determined minimum and maximum width. Width modulation and normalization are only applied to edges being part of the feature. All other edges are assigned a uniform width being slightly below the minimum. In Figure 2 ((a) and (c)), two problems become obvious. At first, yellow edges are often drawn in front of red feature edges since the drawing order does neither consider importance nor edge width. Second, edges with a strong overlap are sometimes difficult to visually separate when being assigned the same color (inset of Fig. 2 (c)). The first problem could be solved by adapting the drawing order each

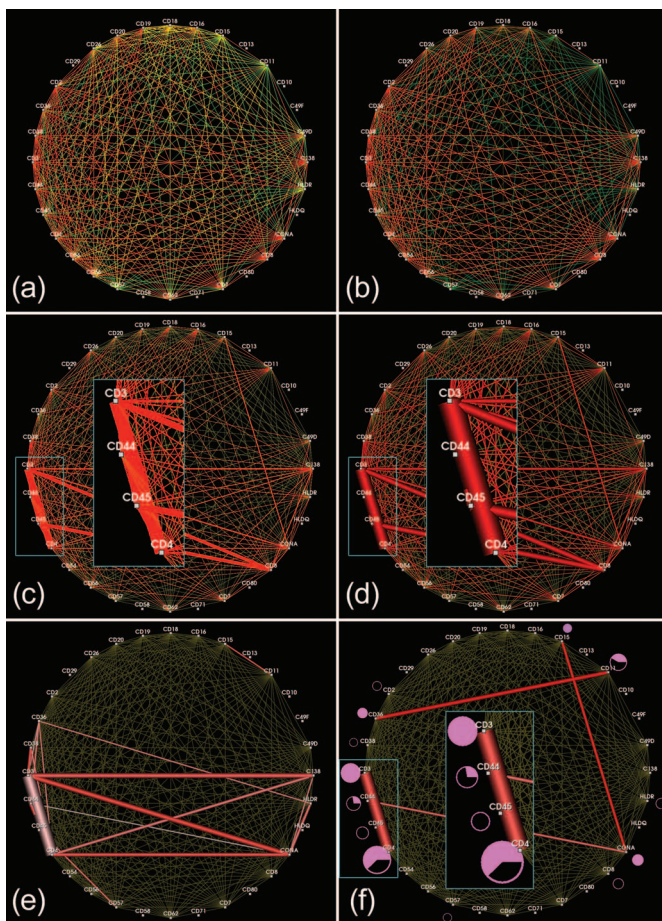


Fig. 2. Development of the graph view. (a): Edge color is modified according to the SimVis standard colors. (b): Opacity modulation to emphasize edges that are part of features. (c): Ribbons encode the edge-weights. (d): Ribbons are replaced by tubes thereby resolving the dependency between edge drawing order and good visibility. (e): A feature has been defined. The different shades of red encode the frequency of each edge inside and outside of the feature. (f): Circular glyphs encode node-specific properties such as binding frequency inside the feature. Please, note the close-up views in (c), (d), and (f).

time the feature specification changes and draw red edges on top of the yellow edges. However, a less expensive solution also solves the second problem as a side effect. The ribbons were replaced by tubes constructed of four faces by means of `vtkTubeFilter` (Fig. 2 (d)). The tube radius is modified according to the edge-weights. To ensure a rounded appearance, surface normals are computed and the tubes are oriented such that a salient edge faces the viewer. Without any additional sorting, edges now automatically appear in increasing order of their associated edge-weight from back to front. The thickest edge, i.e. the tube representing the highest weight is drawn in front of all other edges. Also, feature edges are drawn in front of context edges (yellow) since the latter are guaranteed to have a lower radius. With regard to the second problem, partially overlapping edges are now easier to separate due to shading (inset of Fig. 2 (d)).

Feature Edge Saturation. In a local investigation, the user inspects a certain subset of the observation space, e.g., a single cell. From the current graph visualization it remains unclear whether a PAR binding (edge) only occurs in this cell. In other attribute views of SimVis, such information is conveyed by percentage color mixing of red and yellow (Sec. 4.1). However, a color map with a varying hue component is less effective in encoding variations in magnitude than isomorphic color maps, i.e. maps where saturation or luminance are increased monotonously [25]. Hence, we modify the saturation of the

feature edges (Fig. 2 (e)). If the corresponding PAR binding is fully contained in the feature, the edge is assigned a fully saturated red. If it often occurs in the context, the edge is assigned a very light red. In between, the saturation is interpolated. From our meetings with the biologists, we learned that for a pure exploration of the data, the definition of one feature set including a single feature is sufficient. The biologists then browse through the data by constantly modifying the feature's characteristics. Hence, only mixtures of red and yellow may occur which are replaced here by a varying saturation. The definition of multiple features in a feature set becomes interesting if two cells or two parts of a single cell, i.e. two features shall be compared. In that case, green is used to convey the difference between the two features and other color mixtures, e.g., of green and yellow may occur. However, since the feature edges are of primary interest, we decided to neglect these mixtures and consider only the green color component.

Glyph Drawings. The frequency of an individual PAR binding and the existence and frequency of single-1-CMPs may not be inferred from the graph. In order to integrate this information, the graph view is augmented by glyph drawings (Fig. 2 (f)). Disc sections (`vtkSectorSource`) are employed as a base for the glyph shape and customized such that the outer construction circle is shown as a frame of reference. This simplifies the estimation of sector circumference with respect to the full circumference and better conveys glyph size. One glyph is created per PAR and placed close to the corresponding node. The glyph's inner and outer radius as well as its start and end angle may be modified. All modifications are carried out with respect to the associated feature and neglecting the context. The user is provided with two settings of glyph attribute modification. In the default setting, the inner radius is set to zero and the outer radius is adapted according to the binding frequency of the PAR (values of $[PAR]$). The start angle is set to zero and the end angle is modified according to the percentage of CMPs which exhibit a 1 for this PAR but are not single-1-CMPs (computed based on values of $[PAR]$ and $[PAR1]$). Hence, the missing piece of the full circle represents the percentage of single-1-CMPs.

In the feature represented by Figure 2f (see inset), CD4 is the most frequently binding PAR. About 35% of the CMPs with a 1 for CD4 are single-1-CMPs. CD3 has the second highest binding frequency. The associated filled circle indicates that no single-1-CMPs exists with a 1 for CD3. A special case from a visualization point of view is the existence of only single-1-CMPs for a specific PAR. Without special care, no glyph would be visible here since start and end angle would be equal in this case. To treat those cases, only the outline of the circle is drawn, as for CD45. In a second glyph modification setting, the inner radius and the start angle are set to zero. The outer radius is assigned a uniform value. The end angle is modified according to the percentage of CMPs with a 1 for the associated PAR with respect to the overall number of non-background CMPs in the feature. A very high percentage may indicate that the protein which is bound by this PAR is a lead protein. The glyphs are colored according to the degree of their corresponding node, i.e. the number of incident edges. Finally, the nodes are labeled with the name of the associated PAR.

Brushing Facilities. The graph view is equipped with a brushing facility. Each user-defined brush modifies the DOI attribute which is associated with the view. The modified DOI is then merged with the DOIs of all other views thereby linking the graph to these views. For simplification, we decided to restrict the brushing to nodes. Edge brushing would be hampered by the large number of edge crossings and can also easily be replaced by multiple node brushes. AND, OR, and NOT brushes are provided for the definition of simple and complex brushes. Complex brushes facilitate feature specifications such as 'Show all sites where CD3 AND CD4 bind but CD36 does NOT bind.'. Once a (complex) brush has been defined, a rule is constructed based on the involved brush types and the involved binary attributes corresponding to the selected nodes. Then, the DOI attribute is modified according to this rule thereby triggering a new merge with the DOI attributes of all other views. This again updates these views and also the graph view itself.

The brush type may be chosen from the graphical user interface. A rubber band brushing which is accomplished by pressing the left mouse button and moving the mouse allows for the selection of multiple nodes. The small squares which are drawn at the corresponding node positions are then colored according to the brush type. The colors adhere to the SimVis style: AND = yellow, OR = turquoise, NOT = magenta (not selected = gray). The deselection of nodes is accomplished by rubber band brushing after pressing the right mouse button. An extension of brushing is *Smooth brushing* facilitating the definition of non-discrete DOIs [8]. It has been neglected here since the underlying binarized fluorescence data does not exhibit smooth features.

4.3 The Table View

We integrate a table view in the framework in order to represent the unique CMPs being part of the current feature. Thereby, we take into account the familiarity of the biologists with a table view. Furthermore, the graph view is not able to convey individual CMPs. This information may only be inferred from the graph by a tedious search for individual cliques. A clique is a complete subgraph, i.e. each edge is connected to all other edges. This is inherent to the graph representation of each CMP. The table view has been implemented as a derivation of `vtkQtTableView` which is based on Qt's `QTableView`. It contains one column per PAR and three additional columns showing the absolute frequency of the CMP within the current feature, the percentage frequency and the number of elements which equal 1 (Fig. 3 (e)). Each column may be applied for sorting the table. Alternating row colors simplify the differentiation of adjacent CMPs. Elements which equal 1, are highlighted in the SimVis standard color red since the table view shows only CMPs which are part of the current feature. If only a subset of PARs is of interest, other columns may be hidden by decreasing their width to zero. The table view is equipped with a brushing facility that allows for an OR selection being defined on the table rows (CMPs). This facilitates feature specifications such as 'Show all voxels with CMP_1 or CMP_2 '. AND and NOT brushes are needless here since only a single CMP may exist per voxel and unwanted CMPs may simply not be selected. The table supports the brushing of separate rows, a range of rows or multiple ranges. The brushed rows are highlighted in blue and white.

5 APPLICATION

Our analysis framework has been applied to one cell probe containing lymphocytes and a prostate tissue section. In the following, we discuss how a biologist uses the framework. Please note that for local investigation (Sec. 5.3), the biologist decreased the opacity of the context edges (yellow) to 0% such that visual clutter is reduced and switched off the saturation modulation of the feature edges. When investigating a cell of a specific type, it is less important to know which edges exist in other cells in the data likely being of another type. The situation changes when only cells of the same type are in the probe or can be extracted from the probe and shall be compared or in a sub-cellular examination of a single cell.

5.1 Case Studies

Lymphocytes. In this example, a large PAR library (monoclonal antibodies directed against cluster of differentiation (CD) marker proteins) was used to co-map the cell surface toponome of lymphocytes in a healthy subject. Lymphocytes are a special type of white blood cells. CD marker proteins are proteins that are expressed on the surface of immune cells, such as lymphocytes, but are also expressed in many other cell types of the human and animal organism. We have chosen lymphocytes, since this cell type in the human blood is frequently related to causing chronic inflammatory diseases by entering healthy tissues. Hence, a major biological challenge is to decipher the cell surface toponome code of lymphocytes to detect disease-specific codes, e.g., by a systematic co-mapping of a large number of CD surface proteins [29]. In the present study, we have co-mapped 32 CD surface proteins on a cell probe by using a TIS robot system [2, 10]. 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 an expert and imported into our framework (see [10] for detailed information on the probe preparation, the data acquisition, and thresholding). The initial graph visualization of the entire dataset results in 24 nodes (24 PARs) and 271 edges (out of 276 theoretically possible co-occurring PAR bindings) as illustrated by Figure 3 (a-b). Only 24 of 32 PARs are included in the graph since the remaining PARs, e.g., CD49F and CD10, bind at neither position in the cell probe. Their names are instead displayed in a text inset of the graph view (Fig. 3 (b)). The colors in the 3D view encode the CMP frequency with red corresponding to the highest frequency. The cells appear as ring-shaped structures. In the graph view, it can be seen that CD3 and CD4 co-map most frequently since they are connected by the thickest edge. Also, CD3, CD4 and CONA individually map with a high frequency as indicated by the large, associated glyphs. The color of the glyphs encodes the number of incident edges per node (degree). The highest possible degree in case of 24 binding PARs is 23 (magenta). The high frequency of magenta-colored glyphs reflects the high number of edges. The table view summarizes all existing CMPs and sorts them from top to bottom in order of descending frequency.

Prostate Cancer. In this example, a tissue section has been investigated that was cut from a prostate tissue block of radical prostatectomy. A library of 17 PARs, among which 16 were cell surface proteins, was used to co-map the cell surface toponome in prostate tissue. This example is highly relevant since prostate cancer is the most common noncutaneous malignant neoplasm in men in western countries and its pathogenesis is still unclear [34]. A major biological challenge is to detect disease-specific codes, including the identification of lead proteins which may be candidates for therapeutic intervention: when lead proteins are inhibited, the corresponding protein clusters disassemble and lose their function [30, 34]. The tissue section has been imaged at a single slice location with a matrix of 658×517 pixels and an in-plane-resolution of 216×216 nm. Its analysis results are presented in Section 5.3. For a detailed discussion of the tissue section, see [34].

5.2 CMP-Guided Analysis

In Figure 3 (e), the CMPs with the five highest frequencies have been brushed in the table (bluish rows). The highest frequency has been computed for a single-1-CMP with a 1 at CD36. The corresponding regions appear red in the 3D view (Fig. 3 (c)). Amongst others, CD36 is found on *platelets* (small, solid structures in between the ring-shaped cells). Platelets are crucial in hemostasis since their aggregation causes a bleeding to stop. Since CD36 does not bind to lymphocytes, it may be excluded from the analysis by a NOT-brush in the graph view. Please note that the blank circle, which is drawn at CD36, indicates that this PAR does not co-occur with neither of the other PARs within the feature (Fig. 3 (d)). The CMP with the second highest frequency is again a single-1-CMP with a 1 at CD3 (yellow regions). CD3 may be found in the membrane of *T-lymphocytes* which are a special subgroup of lymphocytes. Different types of T-lymphocytes exist such as *T helper* and *Natural killer T* cells which all participate in the cell-mediated immunity. The third highest frequency is computed for a single-1-CMP with a 1 at CONA (Concanavalin A) which is a sugar-binding protein (green regions). The fourth highest frequency is computed again for a single-1-CMP with a 1 at CD15 (turquoise regions). CD15 binds to *mononuclear immune cells* and *granulocytes* with the latter also being a type of white blood cell. Here, the two ring-shaped, turquoise structures represent mononuclear immune cells since the cell nucleus (missing, inner part of the ring) of granulocytes has a very distinct shape. Mononuclear immune cells represent another type of white blood cell (just as lymphocytes) and are important in immune function since they respond to inflammation signals. The fifth highest frequency is computed for a CMP with two elements equal 1: CD3 and CD4 (blue regions). Hence, an edge exists that connects CD3 and CD4. CD4 binds to *T4 cells* whose function is to activate and direct other immune cells. The depletion of T4 cells is an important characteristic of an infection with the human immunodeficiency virus (HIV). The edge between CD3 and CD4 is colored in

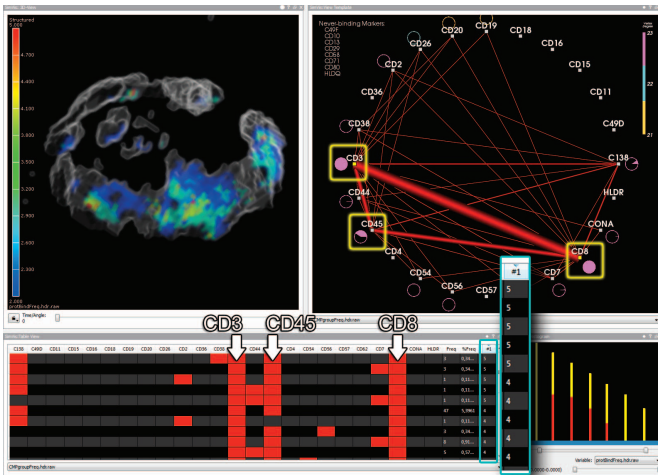


Fig. 5. Local investigation of a T8-cell topome. CD3 and CD8 have been brushed in the graph by means of an AND-brush. The resulting cell parts are colored in the 3D view according to the number of PAR bindings per voxel (blue represents 2, red 5). The remaining parts of the cell are rendered in gray and serve as context information. The histogram plots the number of PAR bindings (logarithmic scaling). The CMPs in the table are sorted according to their number of elements equal 1 in a decreasing order from top to bottom.

our new system and illustrate how the analysis may benefit from the linked views (Fig. 6). At first, we exclude the marker propidium iodide (PROP2) from the analysis by means of a NOT brush on the graph view. PROP2 is excluded since it is no cell surface marker, but instead binds to the cell nucleus. Then, we select the 30 most frequent CMPs in the table (not illustrated here) and superimpose a context visualization of the PAR CD138 (inset of the 3D view). CD138 has been chosen since it is a marker for prostate cancer progression. The 3D view is colored according to CMP frequency. The context appears as a gray silhouette. Next, the focus is on a single cell by brushing x- and y-coordinates of the dataset in the scatter plot (compare to Figs. 4-5 in [34]). This lens function was not available to the biologists in [34], requiring a manual cropping of the dataset. The inset of the 3D view shows the location of the cell within the entire tissue section. Then, we apply an AND brush on CD26 in the graph since CD26 is known to recognize prostate epithelium [34]. Two very frequent co-mappings are revealed: CD54 and CD26 as well as CD26 and CD29. In [34], the biologists had to extract this information from a table with 17 columns and 4217 rows (number of unique CMPs) which was a cumbersome task. After further investigation of other cells in the dataset, they concluded that CD26 and CD29 are lead proteins in a *motif* that is specific to epithelial cells with features of neoplasia inside prostate acini.

5.4 Comparative Analysis

The comparison of cell samples from healthy subjects and patients is crucial in detecting and understanding disease-related protein patterns. We mimic this procedure by comparing two cells in a sample, in particular, we compare two T8-cells: $T8_1$ and $T8_2$ (Fig. 7). Two features are defined in a feature set, one for each cell. The separation of each cell is carried out as described in Section 5.3. The result is visualized in the 3D view and colored according to CMP frequency with red representing the highest frequency. For each cell, a graph has been generated. The glyph encoding in the graph views has been switched to the default setting (see Sec. 4.2, paragraph *Glyph Drawings*). Apart from red edges, green edges appear indicating that the corresponding PAR pair is not co-mapped in this cell but in the other cell (see Sec. 4.1 for a detailed discussion of the edge colors). Thus, similarities and differences between the graphs can easily be inferred. At a first glance, the graphs and the glyphs look similar which is also reflected by similar colors in the 3D view. However, an interesting observation is the binding of CD36 at $T8_2$ in a single-1-CMP. Essentially, CD36 only binds

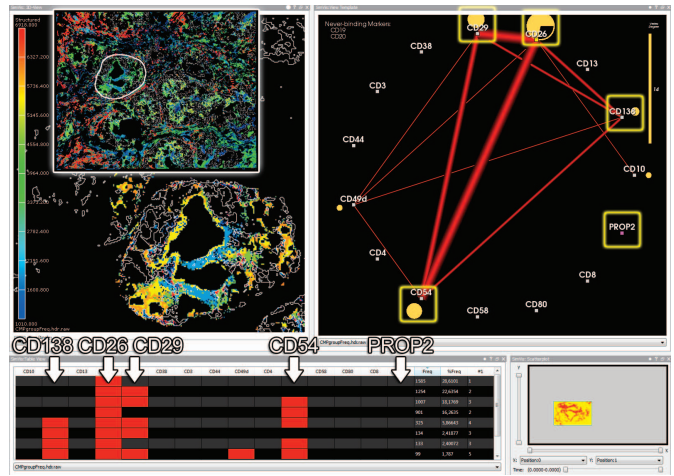


Fig. 6. Local investigation of a prostate cell topome. The analysis has been restricted to the cell by brushing the x- and y-coordinates of the dataset in the scatter plot (lower right view). The cell is colored in the 3D view according to CMP frequency. The PAR CD138 is superimposed as gray isolines for providing context information. The inset shows the location of the cell in the tissue section. CD26 has been brushed in the graph by means of an AND-brush restricting the visualization to prostate epithelium. PROP2 has been excluded from the analysis by means of a NOT-brush. The table shows all CMPs that adhere to the final feature specification.

to platelets and mononuclear immune cells. For further investigation, CD36 has been brushed in the graph view leading to a separation of the small, solid, red region in $T8_2$ (not illustrated here). It was concluded that this region represents a platelet which superimposes $T8_2$. Another interesting observation is that CD19 is not binding at $T8_1$ but at $T8_2$. This may indicate that $T8_2$ is in a different transition stage. Living cells undergo different stages during their life span. The two histograms plot the number of PAR bindings per voxel. In $T8_2$, the highest number of co-mappings is five whereas it is four in $T8_1$.

5.5 User Feedback

We gathered anecdotal feedback from two users, an experienced biologist and a computer scientist who has been working in the biology domain for several years. Both appreciated the possibility to explore the 3D data as a whole instead of slice by slice. The latter is a serious limitation of their homemade system which requires tedious manual post-processing in order to merge the analysis results from individual slices. They commended the interactive selection of parts of the cell probe and the interactive update of graph and table view which greatly simplifies the data exploration. Also, the graph brushing was appreciated as a fast method for defining a template CMP that can be searched for in the data. The graph provides a fast and easy to comprehend overview of a cell's topome and clearly outperforms the table. As shown in Section 5.4, it can be used as a symbolic description of a cell's topome in a cell-by-cell comparison. Information that is cumbersome to infer from a table but can easily be observed in the graph is the frequency of individual PAR bindings and co-occurring PAR bindings. Another advantage of the new system is the possibility to integrate histograms and scatter plots displaying derived parameters, e.g., the number of PAR bindings per voxel.

A drawback of the system is the way of defining a region of interest. For this purpose, a scatter plot opposes the x- and y-coordinates of the data points and the user may define a region by means of a rectangular brush. This requires the user to focus on the 3D view while moving the brush in order to select the desired cell. However, guidance is provided by a context visualization of the entire dataset. The biologist commented that a more flexible brush shape is desirable in order to adhere to more complex cell shapes. Furthermore, measurement units and the current scale should be included in the 3D view.

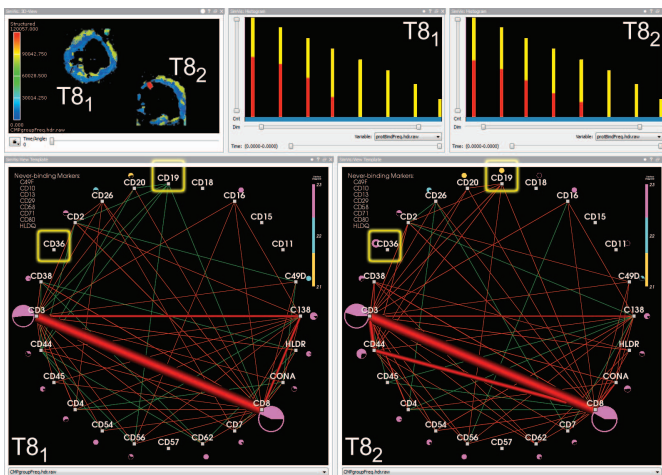


Fig. 7. Comparative analysis of two T8-cells: $T8_1$ and $T8_2$. Red edges in the graph represent PAR bindings that exist in both cells whereas green edges represent bindings that exist only in the other cell. For example, CD19 binds only to $T8_2$. The histograms plot the number of PAR bindings per voxel (logarithmic scaling). In the 3D view, the cells are colored according to CMP frequency.

Feature requests of our users include a picking facility in the 3D view that allows the selection of cell parts thereby inducing an update of all other views with respect to the selected part. Furthermore, it should be possible to crop the data such that only a single cell is considered in the analysis. At the moment, a cell can be focused but the entire dataset is considered as context. This hampers, e.g., a sub-cellular examination in which cell parts shall be focused while the rest of the cell serves as context. If this was possible, the users would fade in again the yellow context edges and apply the modulation of feature edge saturation (Sec. 4.2, paragraphs *Edge Color* and *Feature Edge Saturation*).

5.6 Scalability

In order to assess the scalability of our approach, we created two synthetic datasets: D_1 and D_2 . D_1 has the characteristics: $1056 \times 1027 \times 32(\text{slices}) \times 31(\text{PARs})$. At the moment of writing, a dataset of this size is being acquired by the biologists and represents their largest dataset with respect to voxel count. D_2 resembles the skin biopsy data analyzed in [32] which contains the highest number of PARs that have ever been applied: $658 \times 517 \times 1(\text{slices}) \times 100(\text{PARs})$. The original data could not be provided for the present collaboration. Hence, we randomly distributed 155,000 unique CMPs over both datasets while treating 50% of the voxels as background. This is the highest number of CMPs that the biologists ever observed in one of their datasets. The maximum number of 1's in a CMP has been set to 37% of the PAR number corresponding to the maximum for the lymphocytes dataset.

All tests were carried out on a 3.07GHz Intel 8-core PC with 8GB RAM and a 64bit Windows operating system. Our view setup consisted of a 3D view, a table view, a graph view and a scatter plot. Both datasets were rendered at interactive frame rates in the 3D view. After defining the entire dataset D_1 (values of D_2 in parentheses) as a feature, the table view with 155,000 rows and 31 columns (155,000 rows, 100 columns) needed $\approx 240\text{ms}$ and $\approx 360\text{ms}$ ($\approx 80\text{ms}$ and $\approx 200\text{ms}$) rendering time with and without sorting according to one of its columns. The graph view with 32 nodes and 496 edges (100 nodes, 4950 edges) needed $\approx 2.44\text{s}$ (395ms) rendering time, 2.4s (51ms) for the voxel traversal and the update of the data structures and 40ms (344ms) for repainting the graph and the glyphs. For comparison, the lymphocytes dataset which contains 2228 unique CMPs needed $\approx 10\text{ms}$ and $\approx 11\text{ms}$ for table update and $\approx 70\text{ms}$ (32ms+38ms) for graph update. The long time for voxel traversal of D_1 when updating the graph view does not primarily result from the high voxel

count. It is related to the high number of edges (10) which are stored on average per non-background voxel (17,352K voxels) in the custom data structure described in Section 4.2 (paragraph *Preprocessing*). For comparison, 1 edge is stored on average for the lymphocytes dataset and 3 edges are stored for the prostate dataset with the latter being significantly smaller than D_1 . Further investigation and more real world examples are required to study this effect more thoroughly. From the timings, it can be concluded that even data with a high voxel count and a high number of PARs may be processed. However, interactive frame rates with 25 updates per second are not achieved anymore.

We also explored the visual scalability. The table of D_2 is very wide requiring ineffective horizontal scrolling. In a two screen solution, the table may span both screens. A more appropriate solution could be the application of focus+context techniques, e.g., table lenses [24]. The graph of D_2 is very cluttered and the glyphs may cover only a few pixels which hampers their readability. The situation is improved slightly when the graph is displayed in full screen mode on a second screen. In addition, focus+context techniques such as graphical fish eye views may be applied [26]. A more sophisticated solution to edge cluttering would be *edge bundling* [16]. The original method requires a hierarchy being defined on the nodes. However, such a hierarchy is not defined on the PARs we applied. An extension proposed by the same authors does not require a hierarchy but applies a flexible spring model [17]. An additional advantage of this approach is a parameter that controls the bundling strength. By reducing the bundling strength from full to a lower value, edges become slightly separated which is a prerequisite for the recognition of a varying edge width. Additional solutions for reducing edge cluttering are the reordering of nodes (in contrast to dynamic node arrangements, node positions on the circle are fixed but the assignment of PARs to nodes is flexible) and the exterior routing of edges [13, 37]. It should be noted that the visual clutter is already significantly reduced once only individual cells are investigated and context edges are faded out. However, in future scenarios fading out the edges may not be an option (recall Sec. 5.5).

6 SUMMARY AND DISCUSSION

We presented the integration of techniques from graph visualization in an interactive visual analysis framework for the investigation of topome data. Each node in the graph represents an affinity reagent while each edge represents two co-occurring affinity reagent bindings. The frequency of co-occurring affinity reagent bindings is encoded in the edge's width. The graph visualization is enhanced by glyphs which encode specific properties of the binding. The graph view is equipped with brushing facilities and linked to all other views of the framework. Furthermore, an interactive table view is integrated which summarizes unique fluorescence patterns existing in the data.

We applied the framework to a cell probe containing lymphocytes and to a prostate tissue section. By brushing in the spatial and attribute domain of the corresponding datasets, the biologist achieves a better understanding of the function protein patterns of a cell. In a local investigation, a single cell may be separated from the probe and inspected. By browsing the table view, individual combinatorial molecular phenotypes may be localized. Interesting co-mappings of individual affinity reagents may be localized by brushing the corresponding nodes in the graph view. Clusters of protein mappings with a different number of protein bindings per voxel were observed in a mononuclear immune and a T8 cell. By defining two features (cells), the biologist is able to compare two cells. Edges in the graph of one cell are then colored differently if they occur only in the other cells.

In the future, we will implement more advanced methods and brush shapes for cell separation which will more strongly adhere to the actual cell shape. So far, a reliable analysis of the cell depends on its separability from other cells in the probe.

ACKNOWLEDGMENTS

Technological and biological topome studies were supported by BMBF grants Biochance, CELLECT, NBL3, NGFN2, NGFNplus, and through DFGschu627/10-1, the Innovationskolleg INK15, and the Klaus Tschira foundation (project topome atlas).

REFERENCES

- [1] A. Barysenka, A. W. M. Dress, and W. Schubert. An information theoretic thresholding method for detecting protein colocalizations in stacks of fluorescence images. *J Biotechnol*, 149(3):127–131, 2010.
- [2] S. Bhattacharya, G. Mathew, E. Ruban, D. B. A. Epstein, A. Krusche, R. Hillert, W. Schubert, and M. Khan. Toponome imaging system: in situ protein network mapping in normal and cancerous colon from the same patient reveals more than five-thousand cancer specific protein clusters and their subcellular annotation by using a three symbol code. *J Proteome Res*, 9(12):6112–6125, 2010.
- [3] B. Chapman, G. Jost, and R. v. d. Pas. *Using OpenMP: Portable shared memory parallel programming*. The MIT Press, 2007.
- [4] W. S. Cleveland and R. McGill. Graphical perception: Theory, experimentation, and application to the development of graphical methods. *J Am Stat Assoc*, 79(387):531–554, 1984.
- [5] W. de Leeuw, P. J. Verschure, and R. van Liere. Visualization and analysis of large data collections: a case study applied to confocal microscopy data. *IEEE Trans Visual Comput Graph*, 12(5):1251–1258, 2006.
- [6] J. Dietzsch, J. Heinrich, K. Nieselt, and D. Bartz. SpRay: A visual analytics approach for gene expression data. In *Proc. IEEE Symp. Visual Analytics Science and Technology VAST*, pages 179–186, 2009.
- [7] H. Doleisch, M. Gasser, and H. Hauser. Interactive feature specification for focus+context visualization of complex simulation data. In *Proc. of IEEE TCVG - EUROGRAPHICS Symp. on Vis.*, pages 239–248, 2003.
- [8] H. Doleisch and H. Hauser. Smooth brushing for focus+context visualization of simulation data in 3D. *Journal of WSCG*, 10(1):147–154, 2002.
- [9] A. Dress, T. Lokot, L. Pustyl'nikov, and W. Schubert. Poisson numbers and poisson distributions in subset surprisology. *Annals of Combinatorics*, 8:473–485, 2005.
- [10] 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.
- [11] T. M. J. Fruchterman and E. M. Reingold. Graph drawing by force-directed placement. *Softw. Pract. Exper.*, 21:1129–1164, 1991.
- [12] G. Furnas. Generalized fisheye views. In *Proc. of the ACM CHI Conf. on Human Factors in Computing Systems*, pages 16–23, 1986.
- [13] E. Gansner and Y. Koren. Improved circular layouts. In M. Kaufmann and D. Wagner, editors, *Graph Drawing*, volume 4372 of *Lecture Notes in Computer Science*, pages 386–398. Springer, 2007.
- [14] N. Gehlenborg, S. I. O'Donoghue, N. S. Baliga, A. Goesmann, M. A. Hibbs, H. Kitano, O. Kohlbacher, H. Neuweger, R. Schneider, D. Tenenbaum, and A.-C. Gavin. Visualization of omics data for systems biology. *Nat Methods*, 7(3 Suppl):S56–S68, 2010.
- [15] H. Hauser. Generalizing focus+context visualization. In *Scientific Visualization: The Visual Extraction of Knowledge from Data*, pages 305–327, 2005.
- [16] D. Holten. Hierarchical edge bundles: Visualization of adjacency relations in hierarchical data. *IEEE Trans Visual Comput Graph*, 12(5):741–748, 2006.
- [17] D. Holten and J. J. van Wijk. Force-directed edge bundling for graph visualization. *Comput Graph Forum*, 28(3):983–990, 2009.
- [18] Z. Hu, J. Mellor, J. Wu, and C. DeLisi. VisANT: an online visualization and analysis tool for biological interaction data. *BMC Bioinformatics*, 5:17, 2004.
- [19] M. Meyer, T. Munzner, and H. Pfister. MizBee: A multiscale synteny browser. *IEEE Trans Visual Comput Graph*, 15:897–904, 2009.
- [20] P. Muigg, J. Kehrer, S. Oeltze, H. Piringner, H. Doleisch, B. Preim, and H. Hauser. A four-level focus+context approach to interactive visual analysis of temporal features in large scientific data. *Comput Graph Forum*, 27(3):775–782, 2008.
- [21] S. Oeltze, H. Doleisch, H. Hauser, P. Muigg, and B. Preim. Interactive visual analysis of perfusion data. *IEEE Trans Visual Comput Graph*, 13(6):1392–1399, 2007.
- [22] G. A. Pavlopoulos, S. I. O'Donoghue, V. P. Satagopam, T. G. Soldatos, E. Pafilis, and R. Schneider. Arena3D: visualization of biological networks in 3D. *BMC Syst Biol*, 2:104, 2008.
- [23] S. P. Perfetto, P. K. Chattopadhyay, and M. Roederer. Seventeen-colour flow cytometry: unravelling the immune system. *Nat Rev Immunol*, 4(8):648–655, 2004.
- [24] R. Rao and S. K. Card. The table lens: merging graphical and symbolic representations in an interactive focus + context visualization for tabular information. In *Proc. of the ACM CHI Conf. on Human Factors in Computing Systems*, pages 318–322, 1994.
- [25] B. E. Rogowitz, L. A. Treinish, and S. Bryson. How not to lie with visualization. *Comput. Phys.*, 10:268–273, 1996.
- [26] M. Sarkar and M. H. Brown. Graphical fisheye views of graphs. In *Proc. of the ACM CHI Conf. on Human Factors in Computing Systems*, pages 83–91, 1992.
- [27] W. Schubert. Multiple antigen-mapping microscopy of human tissue. In *Advances in analytical cellular pathology*, Excerpta Medica, pages 97–98. Elsevier, 1990.
- [28] W. Schubert. Topological proteomics, toponomics, melk-technology. *Adv Biochem Eng Biotechnol*, 83:189–209, 2003.
- [29] W. Schubert. A three-symbol code for organized proteomes based on cyclical imaging of protein locations. *Cytometry A*, 71(6):352–360, 2007.
- [30] W. Schubert. On the origin of cell functions encoded in the toponome. *J Biotechnol*, 149(4):252–259, 2010.
- [31] W. Schubert, M. Bode, R. Hillert, A. Krusche, and M. Friedenberger. Toponomics and neurotoponomics: a new way to medical systems biology. *Expert Rev Proteomics*, 5(2):361–369, 2008.
- [32] 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.
- [33] W. Schubert, M. Friedenberger, M. Bode, A. Krusche, and R. Hillert. Functional architecture of the cell nucleus: towards comprehensive toponome reference maps of apoptosis. *Biochim Biophys Acta*, 1783(11):2080–2088, 2008.
- [34] W. Schubert, A. Gieseler, A. Krusche, and R. Hillert. Toponome mapping in prostate cancer: detection of 2000 cell surface protein clusters in a single tissue section and cell type specific annotation by using a three symbol code. *J Proteome Res*, 8(6):2696–2707, 2009.
- [35] H.-J. Schulz, M. John, A. Unger, and H. Schumann. Visual analysis of bipartite biological networks. In *VCBM'08: Proc. of the Eurographics Workshop on Visual Computing for Biomedicine*, pages 135–142, 2008.
- [36] P. Shannon, A. Markiel, O. Ozier, N. S. Baliga, J. T. Wang, D. Ramage, N. Amin, B. Schwikowski, and T. Ideker. Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome Res*, 13(11):2498–2504, 2003.
- [37] A. Symeonidis and I. Tollis. Visualization of biological information with circular drawings. In *Proc. Int'l Symp. on Biological and Medical Data Analysis (ISBMDA)*, pages 468–478, 2004.
- [38] G. H. Weber, O. Rubel, M.-Y. Huang, A. H. DePace, C. C. Fowlkes, S. V. E. Keranen, C. L. Luengo Hendriks, H. Hagen, D. W. Knowles, J. Malik, M. D. Biggin, and B. Hamann. Visual exploration of three-dimensional gene expression using physical views and linked abstract views. *IEEE/ACM Trans Comput Biol Bioinformatics*, 6(2):296–309, 2009.