

Interactive Labeling of Toponome Data

S. Oeltze-Jafra¹ and F. Pieper¹ and R. Hillert² and B. Preim¹ and W. Schubert²

¹Department of Simulation and Graphics, University of Magdeburg, Germany

²Molecular Pattern Recognition Research Group, University of Magdeburg, Germany

Abstract

Biological multi-channel microscopy data are often characterized by a high local entropy and phenotypically identical structures covering only a few pixels and forming disjoint regions spread over, e.g., a cell or a tissue section. Toponome data as an example, comprise a fluorescence image (channel) per protein affinity reagent, and capture the location and spatial distribution of proteins in cells and tissues. Biologists investigate such data using a region-of-interest in an image view and a linked view displaying information aggregated or derived from the channels. The cognitive effort of moving the attention back and forth between the views is immense.

We present an approach for the in-place annotation of multi-channel microscopy data in 2D views. We combine dynamic excentric labeling and static necklace maps to cope with the special characteristics of these data. The generated annotations support the biologists in visually exploring multi-channel information directly in its spatial context. A label is generated per unique phenotype included in a flexible, moveable focus region. The labels are organized in a circular fashion around the focus region. On demand, a nested labeling can be generated by displaying a second ring of labels which represents the channels characterizing the focused phenotypes. We demonstrate our approach by toponome data of a rhabdomyosarcoma cell line and a prostate tissue section.

Categories and Subject Descriptors (according to ACM CCS): J.3 [Computer Applications]: Life and Medical Sciences—Biology and genetics

1. Introduction

Proteins are the basic modules of cells performing a huge variety of functions in living organisms. A major challenge in biology is to understand how proteins cooperate in cells and tissues in time and space [Sch10]. The *toponome* of a cell describes its functional protein pattern, i.e. the location and spatial distribution of proteins. In *toponomics*, the toponome is investigated in order to understand how cells encode different functionalities both in health and disease. Robot-driven multi-parameter fluorescence microscopy is employed for imaging the toponome [Sch03]. The imaging may be carried out in 2D or 3D and results in a fluorescence image or volume per protein. Here, we focus on a 2D slice-based analysis of toponome data.

In a post-processing step, the fluorescence data is binarized. For each pixel, a binary code (protein pattern) is constructed over all images, i.e. proteins, which then encodes the local protein co-mapping. Finally, all unique protein patterns are determined and each is assigned a unique color. Biologists are interested in the natural clustering of protein patterns across a cell, in the difference in clustering between

cells or healthy and pathologic tissue, and in the frequencies of proteins and protein patterns. Hence, they visually explore the toponome data piece by piece. They repeatedly define a region-of-interest in an image view and inspect the corresponding unique patterns in a separate table view. The cognitive effort of moving the attention back and forth between the views is immense.

We present an approach to interactively label toponome data in image views facilitating an exploration of the toponome in its spatial context. Labeling the clusters of protein patterns is challenging since (Fig. 1c):

- very small clusters cover only a few pixels,
- the local entropy, i.e. variety of clusters, is high, and
- phenotypically identical clusters form disjoint regions.

To cope with the high local entropy and to account for the piece-wise exploration of the data, we adopt dynamic *excentric labeling* of a focus region [FP99]. Phenotypically identical but disjoint regions, such as the turquoise or red clusters in Figure 1c, require either multiple converging lines (*leaders*) connecting the regions with a single label (*many-to-one*

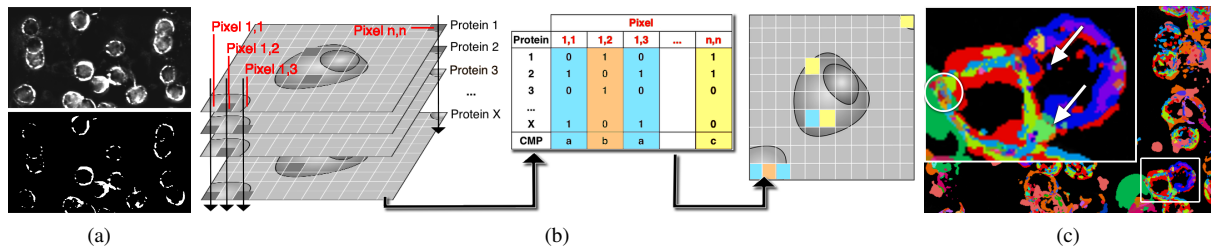


Figure 1: (a) Fluorescence signal of a protein affinity reagent as measured (top) and after binarization (bottom). White pixels indicate protein present. (b) Generation of Combinatorial Molecular Phenotypes (CMPs). For each pixel, the binarized fluorescence signal of all protein affinity reagents is collected in a combinatorial binary code. The set of unique codes, i.e. the CMPs, is computed and visualized in a toponome map (right). Image adapted from [OKH*12]. (c) Inset of an exemplary toponome map illustrating the challenges on labeling. Very small clusters of protein patterns exist (arrows). The local variety of clusters, e.g., inside the circle, is high. Phenotypically identical clusters form disjoint regions, e.g., the turquoise and the red regions.

labeling) or also multiple labels. In order to avoid visual clutter, we combine excentric labeling with static leader-free *necklace maps* [SV10], which line up a single label per unique protein pattern on a curve surrounding the focus region and relate labels to regions, e.g., by matching colors. To the best of our knowledge, we are the first to present a dynamic variant of necklace maps posing special requirements on label update during exploration.

We support multiple labeled focus regions facilitating cell-to-cell comparisons, which so far required the tedious comparison of individual tables. Copies of the labelings are organized in a management view to structure and log the exploration. We demonstrate our approach by a rhabdomyosarcoma cell line and a prostate tissue section. It may be transferred to similar image data, e.g., light microscopy images of differently stained tissue, or maps of geospatial data, e.g., the world-wide distribution of mineral resources.

2. Biological Background

The *toponome* of a cell is defined as the entirety of all protein networks, in which proteins are defined by their protein-to-protein context [Sch03]. It is hierarchically organized and comprises protein clusters which in turn contain *lead proteins* and are interlocked as a network [SBP*06]. The lead proteins control the topology of the clusters and their function as a network. The most advanced toponome imaging technique is robot-driven multi-parameter fluorescence microscopy TISTM [FBKS07]. It is capable of co-mapping hundreds of proteins and their distribution across a cell or tissue sample in situ [SBP*06, SGK*12]. Imaging and analyzing the toponome are essential in finding new drugs, e.g., for cancer treatment, and for detecting protein clusters that can be regarded as a new system of biomarkers in disease [Sch10, SGK*12].

Combinatorial Molecular Phenotypes. After imaging the toponome, the fluorescence data is binarized [BDS10]

(Fig. 1a). This generates a combinatorial binary code (protein pattern) for each pixel where 0 indicates protein absent and 1 protein present. The unique binary codes in the data are referred to as *Combinatorial Molecular Phenotypes* (CMPs). A simple technique for visualizing CMPs is their color-coded representation in a *toponome map*. The computation of a unique color per CMP is described in [OKH*12]. The generation of binary codes, the concept of CMPs, and the toponome map are illustrated by Figure 1(b,c). The binary code corresponding to a CMP very often exists at many pixel positions, which are clustered at several locations of a cell or tissue sample. These protein clusters correspond to functional cell units and are of crucial interest.

3. Biological Workflow and Requirement Analysis

The analysis of toponome data starts with a hypothesis-free visual exploration of the CMPs and involves the following biological tasks:

- (1) detection of selective CMP patterns,
- (2) comparison of CMP patterns, and
- (3) identification of lead proteins.

In (1), patterns characteristic for a particular cell type, a developmental stage of cells or a pathology are searched for. Such patterns support an understanding of cell composition and function, protein interaction, and may serve as biomarkers in disease. The comparison of patterns (2) is crucial, e.g., in comparing healthy and pathologic tissue or cells in different developmental stages for understanding stage transition. The detection of lead proteins (3) may be the first step in drug development. Inhibiting a lead protein causes a disassembly and function loss of the associated protein network, which may eventually stop the disease [SBP*06].

Workflow. The biologists perform these tasks following a specific workflow implemented by their in-house, multiple coordinated view framework (see [OFH*11, OKH*12] for

details on the framework). Here, we focus on the 2D view showing the image data and the toponome map and on the table view listing the CMPs as rows and the proteins as columns. Together, they are the main vehicles of initial toponome exploration (Fig. 2a,b).

After toponome data have been acquired, the biologists browse the morphology in the 2D view to orient themselves in the spatial domain of the data. This step is carried out, e.g., based on a phase contrast image facilitating a good visual separation between cells and background (Fig. 2b). Next, the biologists investigate the CMP data at morphologically interesting locations and search for selective CMP patterns. For this purpose, a focus region is defined on the morphology image. This region is neither draggable nor resizable. After its definition, the corresponding part of the toponome map is superimposed. Note that the corresponding CMPs are not only superimposed on the focus region but on the entire image (Fig. 2b). This is necessary to assess whether a CMP pattern is selective or appears anywhere in the data.

Once a selective pattern has been detected, its CMPs and their contributing proteins are investigated in the table view, which is often shown on a second screen (Fig. 2a). The table lists the CMPs of the entire dataset sorted according to each CMP's overall frequency. The rows corresponding to the focused CMPs are colored. A CMP's unique color is employed to establish visual correspondence between table and toponome map. Comparing cells or cell parts regarding their CMP pattern and proteins requires multiple focus regions. Since this was not supported so far, multiple instances of the framework were created or screenshots were compared.

Requirement Analysis. To investigate the CMPs of an interesting pattern, the user browses the table, which may list hundreds or thousands of CMPs. *All* columns must be checked to retrieve the present proteins. This is essential, e.g., for detecting lead proteins. If all CMPs of a pattern contain a specific protein, it represents a lead protein candidate.

The exploration requires the user to constantly move the focus of attention back and forth between table and 2D view. The static focus region prevents a fluent sampling of the toponome and a comprehension of pattern changes between neighboring image regions. The missing support for multiple focus regions hampers the comparison of CMP patterns.

The primary requirement of our collaborators on a novel approach is the embedding of information derived from the table into the 2D view such that the toponome may be explored directly in its spatial context. Further requirements are the support of multiple focus regions and the management of these regions and their respective CMP pattern, e.g., capture, show, hide, and store.

4. Related Work

This section is based on a survey of labeling techniques in medical visualizations [OJP14]. Ali et al. studied handmade

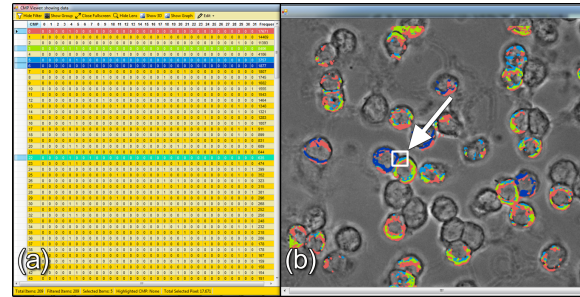


Figure 2: Toponome analysis framework. (a) The table view lists all CMPs as rows and proteins as columns. (b) The 2D view shows a grayscale phase contrast image as spatial context. The ring-shaped structures represent cells. Each CMP within a user-defined focus region (arrow), i.e., the corresponding part of the toponome map, is superimposed in color and the respective table row is colored likewise.

illustrations in scientific and technical textbooks and identified two types of labels: internal and external [AHS05].

Internal Labels. Labels being superimposed on the structure of interest are referred to as internal labels. They have been applied, e.g., to virtual bronchoscopy images [MHST00], medical surface [RPRH07] and volume rendered data [JNH*13]. Their application to toponome data is challenging since clusters of the same CMP do not form a single, continuous region in the 2D toponome map (e.g., the turquoise or red clusters in Fig. 1c). The problem might be tackled by multiple identical labels as shown for annotating vascular structures in volume rendered images [JNH*13]. Here, a vessel is often partially occluded by other vessels or organs. However, another problem prevents the application of internal labels. Often, CMP clusters cover only a few pixels, which would be largely occluded by the label.

External Labels. The occlusion problem is solved by external labels. They are positioned outside the structure of interest and connected to it by a line. This so-called *leader* connects an *anchor point* on the structure and a point on the label box holding the label's textual representation. Ali et al. proposed a variety of real-time label layout algorithms for anatomical 3D models [AHS05]. Labels are arranged in a circular fashion around the model or along its silhouette. Mühler et al. demonstrated the labeling of 3D medical structures located inside a transparent structure or being currently hidden but still of importance for surgical planning [MP09]. Mogalle et al. presented the automatic optimal placement of external labels representing findings in 2D radiological slice data [MTSP12]. They focused on avoiding leader crossings and labels occluding crucial image parts. Their approach is limited to ≈ 10 labels, which is realistic for radiological data. However, the number of CMPs even in a small subregion of the toponome map is often higher.

Boundary Labeling. In early work, Preim et al. presented a system for exploring anatomical models which combines zooming techniques, fisheye views, and interactive labels [PRS97]. The labels are aligned on the left and right boundary of a virtual rectangle enclosing the model. Bekos et al. later coined the term “boundary labeling” in the context of annotating static maps [BKSW07]. A virtual rectangle containing the map is constructed and external labels are placed outside the rectangle. They are connected by leaders to the map areas of interest. Crossings of leaders are avoided and total leader length is minimized. Boundary labeling is generally applied to the entirety of data. Labeling the entire toponome map is however, neither feasible due to the hundreds or thousand of CMPs nor required by the biologists who explore the data piece by piece.

Excentric Labeling. The cell-wise or subcellular piece-wise exploration of the data is very well related to excentric labeling by Fekete and Plaisant [FP99]. Their dynamic approach aims at labeling dense maps interactively by means of a moveable, flexible focus region. The labels are displayed in stacks to the left or right of the focus region and connected to the structure of interest inside the region by a leader. Fink et al. extended the approach by various techniques for creating a visually pleasing annotation, e.g., the use of straight lines or Bézier curves instead of zigzagging polylines [FHS*12]. Luboschik et al. presented a fast point-feature labeling approach, which avoids the placement of labels over other labels or visual representatives such as leaders and icons [LSC08]. They coupled the approach with a moveable label lens. Transferring excentric labeling to toponome data is not straightforward. Several leaders originating either from a single label (*many-to-one labeling* [Lin10]) or from multiple identical labels would be necessary to annotate multiple clusters of the same CMP. Even with minimized leader crossings, this would cause a cluttered visualization for a larger number of CMPs.

Necklace Maps. A static labeling approach abandoning leaders has been proposed by Speckmann and Verbeek for visualizing statistical data on geographical maps [SV10]. Glaßer et al. have applied necklace maps to labeling clusters of breast tumor tissue with cluster-specific perfusion information [GLP14]. In a necklace map, the labels are related to structures of interest by matching colors – the unique CMP color in our case – and spatial proximity. They are organized on a one-dimensional curve (the necklace) that surrounds the map or a subregion.

Consequences. We choose *external labels* over internal ones since the latter would occlude very small CMP clusters. To cope with the high local entropy of toponome data and to account for its piece-wise exploration, we adopt *excentric labeling* of a focus region [FP99]. Disjoint regions, such as the turquoise clusters in Figure 1c, require either multiple converging leaders connecting the regions with a single label

(*many-to-one labeling*) or also multiple labels. In order to avoid visual clutter, we adopt the leader-free *necklace maps* [SV10], which line up a single label per CMP or protein on a curve surrounding the focus region. The combination of excentric labeling and necklace maps meet our requirements on a visual exploration of toponome data (Sec. 3).

5. Interactive Labeling of Toponome Data

We discuss our visual encoding, aspects of label position, order, and count, and we emphasize modifications to the original static necklace map approach. After describing the necklace composition, we elaborate on interaction facilities and introduce a view for managing multiple necklaces.

5.1. Basic Approach

Initialization. At first, the user defines a focus region (region-of-interest, abbrev. ROI) on the toponome map by means of a flexible lens. We have implemented three lens shapes: circle, rectangle, and lasso. Circular and rectangular lenses are adjustable with respect to size and position. Both are meant for a quick inspection of the CMP distribution. The lasso is employed for a more targeted inspection of separate cellular subregions. It does not need to be adjustable since it is aligned with a particular shape. In an early prototypical implementation, our collaborators favored the circular lens since it adheres to the metaphor of exploring a dark room by means of a flashlight. For a recent survey on interactive lenses in visualization, see [TGK*14].

Nested Necklaces. After ROI definition, all pixel positions within the ROI and their associated CMPs are determined. Then, a one-dimensional curve (the necklace) surrounding the ROI is constructed. Currently, our implementation is restricted to a circular necklace since it best matches the circular lens shape (see [SV10] for arbitrary necklace shapes). The CMPs are represented by graphical symbols, which are strung on the necklace (inner necklace in Fig. 3). Following Speckmann and Verbeek [SV10], we provide circular and bar-shaped symbols (Fig. 4). In the remainder, we use the terms *symbol* and *label* interchangeably.

On demand, a second necklace enclosing the former is displayed. One symbol per protein present in the focused CMPs is drawn (outer necklace in Fig. 3). This nested labeling facilitates the concurrent exploration of CMPs and proteins. While dragging the focus region, the protein necklace is hidden by default to avoid mental overload.

5.2. Visual Encoding

Label Text. When the CMPs of a new toponome dataset have been determined, each is assigned a unique name which simply equals its place in a frequency ranking of all CMPs. This name is typeset within the corresponding symbol. The

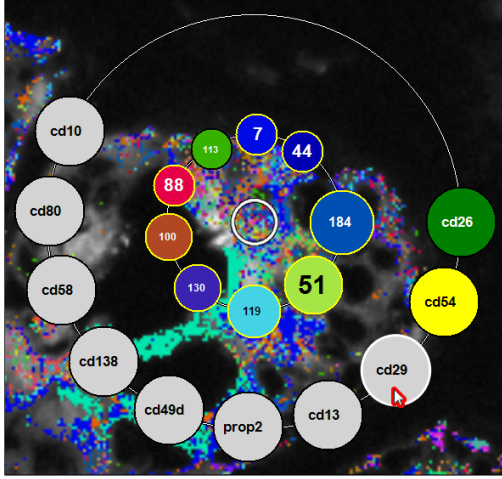


Figure 3: Nested necklace map. Two one-dimensional curves (the necklaces) surround a focus region (white center circle). The CMPs in the focus region and their present proteins are represented by circular symbols strung on the inner and the outer necklace, respectively. The CMP symbol colors match the unique CMP colors while the colors of the protein symbols indicate lead protein likelihood. Please see the text for all other encodings and interaction facilities.

name of the protein affinity reagent is typeset in the symbol of the corresponding protein. The names relate the symbols to the table view since the latter consists of columns listing the ranking place and the proteins (Fig. 2a).

Symbol Size. The relative frequency of a CMP inside the ROI f_{cmp} is of particular interest to the biologists. It is defined as the number of ROI pixels being associated with the CMP normalized by the overall number of ROI pixels. We map the CMP frequencies to the area of the circular symbols and to the length of the bar-shaped symbols, respectively. In accordance with [SV10] and Tufte who demands to “tell the truth about data” [Tuf01], we employ *mathematical scaling*, which directly relates the symbol area/length to the underlying data. However, for the circular symbols, we offer *perceptual scaling* by Flannery’s compensation which aims at com-

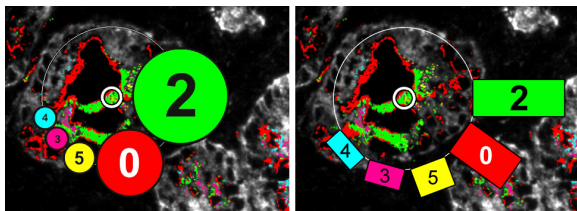


Figure 4: Circular and bar-shaped labels are implemented. Circles encode CMP frequency by area and bars by length.

pensating for the non-linear relationship between an increase in circular area and the perceived increase [Fla71]. Our collaborators prefer circular symbols due to their orientation-independent encoding of frequency and the more symmetric and aesthetic appearance of the resulting necklaces (Fig. 4). Hence, we show circular symbols in the remainder.

The biologists are also interested in the relative frequencies of the proteins inside the ROI f_{prot} . A protein’s relative frequency is independent of the number of pixels. It is defined as the number of focused CMPs with this protein present normalized by the overall number of focused CMPs (except for the background zero-CMP). The biologists categorize the frequencies rather than considering individual values. For the detection of lead proteins, it is sufficient to know whether a protein is present in (nearly) all CMPs inside the ROI or only in a small subset. Hence, we assign a uniform size to the protein symbols and employ color to encode the frequency category (outer necklace in Fig. 3).

Symbol Color. A necklace map communicates the relation between a symbol and its corresponding pixels within the ROI by matching colors and spatial proximity. Hence, we color each symbol on the CMP necklace according to the CMP’s unique color in the toponome map (Fig. 3). For the symbols on the protein necklace, we use a segmenting color scale. Symbols of proteins with a relative frequency $f_{prot} < 80\%$ are shaded in gray, $80\% \leq f_{prot} < 100\%$ in yellow, and $f_{prot} = 100\%$ in green. This facilitates an easy detection of lead protein candidates (green; recall Sec. 3) and of such near the mark (yellow).

5.3. Label Position, Order, and Count

The following methods are straightforward to implement based on simple trigonometry facilitating an update of the necklaces at interactive frame rates during exploration.

Position and Order. Besides color, necklace maps employ spatial proximity to relate image or map regions and their corresponding symbol. Optimizing spatial proximity is a hard problem having received special attention in [SV10]. For toponome data, this problem is even aggravated. Often, multiple clusters of the same CMP exist in a focus region (Fig. 3,5) and also a protein may be scattered across the entire region. Optimization with respect to one cluster is not reasonable in particular for similar-sized, equally distributed clusters. Generating multiple symbols would require their mental integration during exploration. The integration is particularly cumbersome if symbol attributes encode data variables, e.g. size encoding CMP frequency. Finally, very small clusters may exist in the center of a focus region where spatial proximity is hard to achieve by means of a standard convex necklace shape. Discussing these problems with our collaborators revealed that in an initial exploration of toponome data, they are rather interested in the relative frequency of

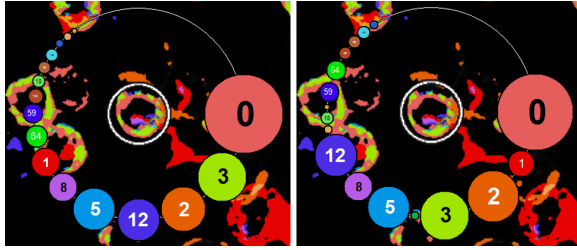


Figure 5: Local vs. global sorting of CMP symbols. (Left) In local mode, the symbols are sorted clockwise according to their CMP's frequency inside the focus region (inner circle). Note that symbol size encodes local CMP frequency while the label text equals the CMP's place in a ranking of global frequencies. (Right) In global mode, the frequency inside the entire dataset is employed for sorting. The symbols are not ordered anymore according to size, but the label texts are ordered now.

the CMPs than in their exact location inside the focus region. Hence, we decided to sacrifice the spatial proximity criterion in favor of a sorted symbol line-up along the necklace starting at 3'o clock with the most frequent CMP and proceeding in clockwise order. Due to the sorted line-up, simple comparisons of CMP frequency within a necklace are even possible when symbols sizes are visually not distinguishable.

The symbols on the CMP necklace may be sorted according to the CMP frequency inside the ROI (*local*) or the total frequency in the dataset (*global*). An exploration in local mode supports the detection and tracking of a CMP's place in a local frequency ranking (Fig. 5a). An exploration in global mode simplifies the tracking of a CMP's presence and frequency inside the focus region (Fig. 5b). This is due to the rather stable place of its corresponding symbol in the order of symbols, which is fix as long as the more frequent CMPs also remain in focus. Note that in global mode, the symbol sizes are not ordered since they still represent the local frequency which often differs from the global one. Please also see our supplemental video for an illustration of the modes.

In order to simplify the search for a specific protein, the symbols on the protein necklace may be arranged alphabetically. Alternatively, the symbol order may be chosen to reflect each protein's place in a ranking of the number of associated ROI pixels. The latter is set by default and also shown in all figures of the remainder.

The necklace radii and the arc length distance between neighboring symbols are chosen such that labels do neither overlap the focus region nor each other. The latter is guaranteed along the necklace and across inner and outer necklace.

Count. In a toponome dataset, hundreds to thousands of CMPs may exist depending on the investigated biology and the number of applied protein affinity reagents. Even in a

small focus region, the number of CMPs can be quite high. However, the number of labels that can be drawn on the CMP necklace is restricted by the minimum size of a symbol down to which it is readable and by the necklace perimeter. Since the necklace should closely adhere to the focus region rather than exploiting the entire available screen space, its perimeter is bounded above. Instead of predefining the perimeter, we first map each CMP's relative frequency f_{cmp_i} to symbol size. Based on $f_{cmp_i} \in [0, 1]$, the diameter ϕ_{s_i} of the corresponding symbol s of the necklace map is computed:

$$\phi_{s_i} = \phi_{base} \cdot \sqrt{f_{cmp_i}}, i \in [1, n_{cmp}] \quad (1)$$

The number of CMPs inside the ROI is denoted by n_{cmp} . The global scaling factor ϕ_{base} corresponds to an adjustable maximum symbol size which is initially set to 150 pixels. Note that this high value is only achieved in the rare case of a single CMP covering the entire ROI ($f_{cmp_i} = 1$). If necessary, ϕ_{s_i} is clamped to the minimum value of four pixels to guarantee the readability of its symbol color. The mathematical scaling in Equation 1 directly relates the symbol area – not the radius/diameter – to the underlying data by employing the square root.

Based on the maximum of ϕ_{s_i} , we then compute the necklace diameter such that this symbol does no overlap the focus region. We then draw the symbols starting at 3'o clock and proceeding clockwise until a new symbol would intersect the first one. Following this strategy, the most frequent CMPs inside the ROI are labeled. This has been agreed upon with the biologists, since very small CMP clusters might represent noise not being eliminated in the course of binarization (Fig. 1a,b). However, special care must be taken when the labels shall be ordered according to global CMP frequency. If for instance only 20 out of 30 CMPs can be labeled, the 20 locally most frequent CMPs do not necessarily coincide with the 20 globally most frequent ones. To guarantee that always the former are labeled, we first determine them and then, sort only these in descending order according to global CMP frequency. A more fine-granular inspection of the CMP distribution can be accomplished by capturing the necklace and labeling all CMPs in an enlarged separate widget (Sec. 5.5).

The number of labels that can be drawn on the protein necklace is also limited by the same factors but the number of proteins is small as compared to the number of CMPs. The most comprehensive toponome study hereof, employed 100 protein affinity reagents [SBP*06]. Furthermore, only a subset of all proteins is included in a reasonably sized focus region. So far, we have been able to draw a label for each protein inside a ROI employing a symbol size that guarantees good readability and at the same time a perimeter that is not far off the perimeter of the CMP necklace. Drawing all symbols is crucial here since otherwise lead protein candidates may remain unnoticed.

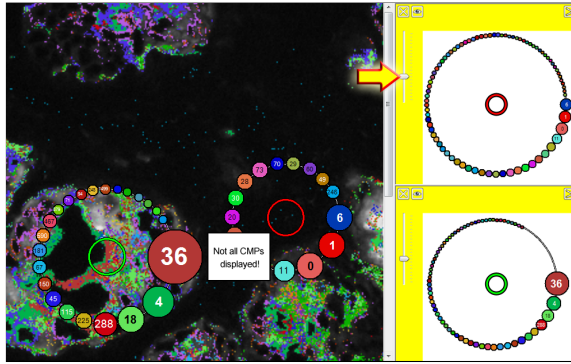


Figure 6: 2D view of image data and toponome map (left) and necklace management view (right). The management view organizes the necklaces of the two focus regions as widgets. Both widgets have been enlarged by means of a slider (arrow) to gain space for more CMP symbols.

5.4. Necklace Interaction

The user can drag the focus region across the toponome map and modify its size by scrolling the mouse wheel. The necklace of a selective CMP pattern can be captured via mouse-click causing an interactive copy to be added to the necklace management view (Sec. 5.5). Another necklace map may be initialized, causing a fade-out of the old map. For orientation purposes, the old focus region remains visible. If multiple necklaces have been defined, any of them can be reactivated by clicking the respective focus region. Note that during interaction, only the CMPs inside the focus region of the active necklace map are colored in the toponome map.

A tooltip listing the relative and the absolute CMP frequency inside the ROI is shown during mouse hover of a CMP's symbol. If the symbol is clicked, the CMP's pixels are highlighted by a temporary blinking. This is particularly useful in cases of CMPs with barely distinguishable colors. Furthermore, if the protein necklace is visible, the symbols of the proteins present in the CMP are highlighted.

The protein necklace is by default only visible on demand. Hovering the mouse pointer over a symbol causes an emphasis of the symbols of all CMPs with this protein present by means of a yellow contour. On clicking the symbol, the CMPs' pixels are highlighted by a temporary blinking.

5.5. Necklace Management View

The necklace management view facilitates the organization of multiple necklaces and helps to structure the exploration. It is attached to the 2D view of the toponome map (Fig. 6). The view is based on requests by the biologists for having a means to record their exploration results. Such records illustrate the daily work and are integrated in the laboratory book. They support scientific reporting of research results

and simplify the communication with other biologists. Furthermore, the management view arranges the necklaces in a non-overlapping fashion thereby simplifying a comparison of the associated CMPs. Superimposing all necklace maps on the toponome map would lead to overlapping necklaces and considerable occlusions of the image data.

In the management view, each necklace is presented in a resizable widget. If a widget is enlarged, the necklace diameter is increased causing previously neglected CMPs to be displayed (recall paragraph "Count" in Sec. 5.3). This facilitates a more fine-granular inspection of the CMP distribution. The background of the widget may be set to the corresponding part of the toponome map. For comparing necklaces, a plain color background causes less distraction.

A necklace map may be shown/hidden in the toponome map by selecting/deselecting its widget. Note that the focus region of a hidden map remains visible. The background color of a selected widget switches from white to yellow. Multiple selections are supported. For each necklace map, the user may choose whether the corresponding CMPs, i.e. their pixels in the toponome map, are shown in color. In Figure 6, the coloring is restricted to the left necklace.

6. Application

We demonstrate our approach by a rhabdomyosarcoma cell line and a prostate tissue section. Both probes have been imaged by means of the TIS robot system with an in-plane-resolution of $216 \times 216 \text{ nm}$ (Sec. 2). Protein affinity reagents, more precisely, monoclonal antibodies directed against cluster of differentiation (CD) surface marker proteins, were co-mapped on the probes. The resulting fluorescence images were binarized according to [BDS10] (Sec. 2). We conclude the section by providing anecdotal user feedback.

6.1. Rhabdomyosarcoma Cell Line

Rhabdomyosarcoma (RMS) is the most common peripheral malignant tumor of soft tissue in children and adolescents and its causes are unclear [HJC*13]. RMS is made up of cells which normally develop into skeletal muscles. To research RMS, muscle cells were extracted from the RMS cell line TE671. Cell lines are populations of cells which have been cultivated from a single cell thus held to contain the same genetic makeup. The cell sample has been imaged in a single transection with a matrix of 693×552 pixels employing 23 protein affinity reagents. 958 CMPs were derived from the binarized data. Sample preparation, data acquisition, and binarization are detailed in [SBP*06].

RMS cells enter two different evolutionary states characterized by a specific cell shape: spherical and elongated with spindle-shape extensions [SBP*06] (Fig. 7a). Spherical cells spontaneously enter an exploratory state in which they form three spindle-shaped extensions. Once a promising direction

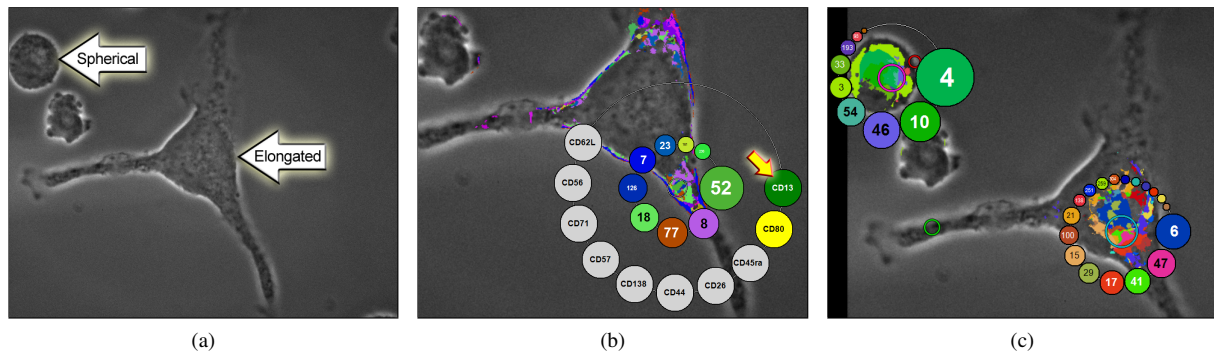


Figure 7: Necklace maps for visually exploring the toponome of Rhabdomyosarcoma (RMS) cells. (a) Phase-contrast image of RMS cells in two different states of their evolution: spherical and elongated with spindle-shaped extensions. (b) A necklace map at one of the extensions confirms CD13's function as a lead protein [SBP*06] as indicated by the green symbol (arrow). (c) Two focus regions have been defined in the cell bodies. Note the strikingly different toponome despite the same cell type.

has been detected by the cell, it proceeds to a migratory state characterized by a withdrawal of one of the extensions. The whole process is targeted at metastasis formation.

Previous toponome decoding work has shown that the proteolytic enzyme CD13 functions as a lead protein driving and directing the formation of the cell extensions [SBP*06]. Based on the same cell type and a similar dataset, we recapitulate this finding (Fig. 7b). While previous work required a time-consuming investigation of the CMP table view, the necklace facilitates a quick identification of CD13 as a lead protein. Its corresponding symbol is colored in green and appears at the starting position of symbol drawing (arrow).

Furthermore, we show that the protein network controlled by CD13 across the cell body shows strikingly different variations for cells in the spherical as compared to the exploratory state (Fig. 7c). Two focus regions were placed within the cell bodies. The toponomes represented by the corresponding necklaces are completely disjoint. Furthermore, the CMPs included in the focus region of the spherical cell barely occur in the elongated cell and vice versa. An investigation of the protein necklaces of both focus regions revealed an omnipresence of CD13 (not illustrated here to simplify a comparison of the CMP patterns). This provides further evidence that CD13 functions as a control element steering the transformation from the spherical to the exploratory state by a recombination with other proteins. It was shown in [SBP*06], that inhibiting CD13 prevents the transformation from the spherical to the exploratory state.

6.2. Prostate Tissue Section

The tissue section was cut from a prostate tissue block of radical prostatectomy — the surgical removal of the entire prostate gland in the therapy of prostate cancer. This type of cancer is the most common noncutaneous malignant neoplasm in men in western countries and its pathogenesis is

still unclear [SGKH09]. The tissue section has been imaged in a single transection with a matrix of 658×517 pixels employing 17 protein affinity reagents. 2100 CMPs were derived from the binarized data. Sample preparation, data acquisition, and binarization are detailed in [SGKH09].

The tissue section contains several *prostate acini* — many-lobed, berry-shaped terminations of the prostate glands lined by secretory epithelial cells — and the fibromuscular *stroma* between the acini. The protein affinity reagent CD138, which is a marker for prostate cancer progression, singles out the acini in its fluorescence image (Fig. 8a). For clarification, one acinus has been encircled. Its epithelial cells appear white in the image while their nuclei and the lumen of the acinus show no response to CD138 and hence, appear as small black circular and large black centered regions, respectively. The encircled acinus drew the interest of the biologists since a fraction of its epithelial cells exhibits features of *prostate intraepithelial neoplasia (PIN)* [SGKH09].

Researching PIN is crucial since it is considered to be a pre-malignancy of the prostatic glands. In order to investigate the toponome of PIN, we have dragged a focus region across the epithelial cells. A representative necklace map including the protein necklace is shown in Figure 8b. The CMP pattern is selective for epithelial cells since none of the CMPs appear in the stroma surrounding the acini. The protein necklace reveals CD26 and CD29 as lead protein candidates indicated by the yellow colored symbols. Both contribute to all but one CMP, which in both cases is the one with only the respective other protein present. For instance, only CMP 6 does not exhibit CD29 but instead solely contains CD26 (Fig. 8b). Similar to the role of CD13 in tackling rhabdomyosarcoma (Sec. 6.1), inhibiting CD26 and CD29 may contribute to preventing the transformation of PIN to prostate adenocarcinoma [SGKH09]. CD26 and CD29 were already identified as lead proteins in [SGKH09]

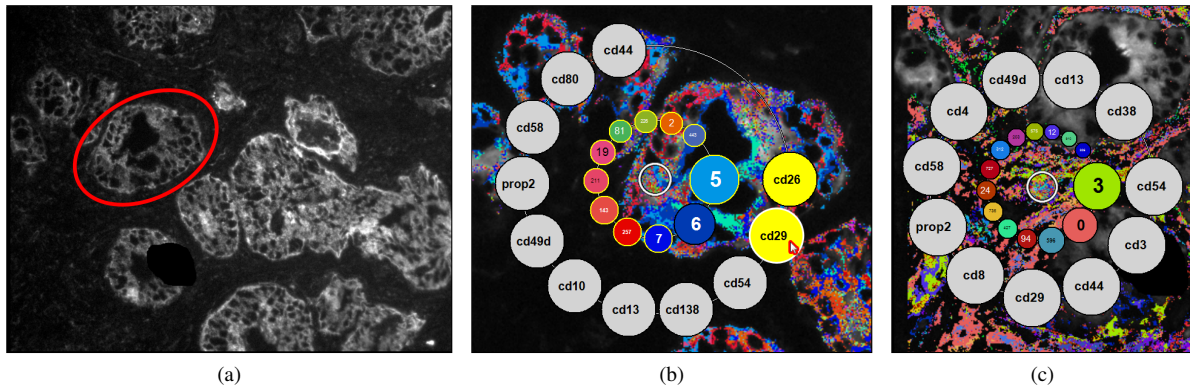


Figure 8: Necklace maps for visually exploring the toponome of a prostate tissue section. (a) Fluorescence image of protein affinity reagent CD138 with one acinus encircled. (b) A necklace map at epithelial cells of the acinus from (a) indicates CD26 and CD29 as lead protein candidates (yellow circles). CD29 is mouse hovered causing all symbols of CMPs containing CD29 to be highlighted (yellow border). (c) A focus region is defined below the acinus in the stroma. Note the strikingly different CMP pattern compared to (b) despite the overlap of contributing proteins (7 out of 11).

and [OFH*11] however, by means of a more complex and time-consuming pipeline of analysis and interaction steps involving additional views.

A second necklace has been positioned over a part of the stroma (Fig. 8c). The corresponding CMP pattern is selective for the stroma and considerably differs from the one in the acinus (Fig. 8b). In Figure 8c, the acinus is located in the upper right corner. The protein necklace reveals again a high frequency of CD29 but also no mapping of CD26. Since the latter specifically recognizes prostate epithelium, this may be seen as a validation of our labeling algorithm. Furthermore, the necklace shows a mapping of CD4 and CD8 indicating the presence of T4 and T8 lymphocytes both participating in the cell-mediated immunity. This in turn, substantiates the presence of inflammatory cells.

6.3. User Feedback

We gathered anecdotal feedback from a biologist with a long-term, strong background in oncology and a computer scientist who has been working in his laboratory for many years. Both are co-authors of the paper. They used our necklace map approach and we simultaneously recorded their comments. They appreciated the in-place annotation of CMPs and proteins as a great cognitive relief since it avoids the tiresome shifting of attention back and forth between table and 2D view (Sec. 3). The comprehensive and sorted display of CMPs along the necklace obviates the search for the focused CMPs in the table. The display of the protein necklace and the interaction with it simplify the identification of present proteins, the detection of lead proteins, and the determination of cell types. Retrieving this information from the table view requires scrolling through the rows and examining each selected row for 1s (Fig. 2a).

The interaction with the necklace map was considered simple and effective. Merely, the temporary blinking of CMP pixels after clicking a symbol causes distraction and should be replaced by a less discomposing highlighting technique. The necklace management view was considered useful. It was heavily used for hiding and showing individual necklace maps. In contrast, the scalability of the necklace widgets was barely utilized due to a common focus of the CMP analysis on the most frequent ones, which were always visible.

7. Summary and Discussion

We have presented an approach to interactively label toponome data in 2D views thereby supporting biologists in visually exploring the data. The approach may be readily transferred to other image data exhibiting a very high local entropy, phenotypically identical structures forming multiple disjoint regions, and very small structures.

We have combined the dynamic excentric labeling of a focus region [FP99] with the static leader-free labeling of necklace maps [SV10]. The user may place a single or multiple focus regions in the image view causing the contained protein patterns to be displayed as symbols strung on a necklace surrounding each focus region. On demand a second necklace illustrating the proteins present in the focused patterns can be displayed. A focus region may be dragged and adjusted causing an update of the necklace(s) at interactive frame rates. For the use cases in Section 6 and larger test images (1600×1200 pixels), no restricted interactivity even for unreasonably large focus regions was observed.

A necklace management view has been implemented for organizing multiple necklaces and structuring the exploration. While necklaces may overlap in the toponome map, the management view arranges them in a non-overlapping

fashion subserving a comparison of the represented toponomes. We have demonstrated our approach for the visual exploration of a rhabdomyosarcoma cell line and a prostate tissue section. We plan to integrate the approach into volume rendered views of 3D toponome data [OKH*12].

Acknowledgements

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

References

- [AHS05] ALI K., HARTMANN K., STROTHOTTE T.: Label Layout for Interactive 3D Illustrations. *Journal of WSCG* 13, 1 (2005), 1–8. [3](#)
- [BDS10] BARYSENKA A., DRESS A. W. M., SCHUBERT W.: An Information Theoretic Thresholding Method for Detecting Protein Colocalizations in Stacks of Fluorescence Images. *J Biotechnol* 149, 3 (2010), 127–131. [2, 7](#)
- [BKS07] BEKOS M. A., KAUFMANN M., SYMVONIS A., WOLFF A.: Boundary Labeling: Models and Efficient Algorithms for Rectangular Maps. *Comp Geom-Theor Appl* 36, 3 (2007), 215–236. [4](#)
- [FBKS07] FRIEDENBERGER M., BODE M., KRUSCHE A., SCHUBERT W.: 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 (2007), 2285–2294. [2](#)
- [FHS*12] FINK M., HAUNERT J.-H., SCHULZ A., SPOERHASE J., WOLFF A.: Algorithms for Labeling Focus Regions. *IEEE Trans. Vis. Comput. Graphics* 18, 12 (2012), 2583–2592. [4](#)
- [Fla71] FLANNERY J. J.: The Relative Effectiveness of Some Common Graduated Point Symbols in the Presentation of Quantitative Data. *Cartographica* 8, 2 (1971), 96–109. [5](#)
- [FP99] FEKETE J.-D., PLAISANT C.: Excentric Labeling: Dynamic Neighborhood Labeling for Data Visualization. In *SIGCHI Conference on Human Factors in Computing Systems* (1999), pp. 512–519. [1, 4, 9](#)
- [GLP14] GLASSER S., LAWONN K., PREIM B.: Visualization of 3D Cluster Results for Medical Tomographic Image Data. In *Conference on Computer Graphics Theory and Applications (GRAPP)* (2014), pp. 169–176. [4](#)
- [HJC*13] HINSON A. R., JONES R., CROSE L. E., BELYEA B., BARR F. G., LINARDIC C. M.: Human Rhabdomyosarcoma Cell Lines for Rhabdomyosarcoma Research: Utility and Pitfalls. *Frontiers in Oncology* 3, 183 (2013), eCollection. [7](#)
- [JNH*13] JIANG Z., NIMURA Y., HAYASHI Y., KITASAKA T., MISAWA K., FUJIWARA M., KAJITA Y., WAKABAYASHI T., MORI K.: Anatomical annotation on vascular structure in volume rendered images. *Comput. Med. Imag. Grap.* 37, 2 (2013), 131–141. [3](#)
- [Lin10] LIN C.-C.: Crossing-Free Many-to-One Boundary Labeling With Hyperleaders. In *Pacific Visualization Symposium (PacificVis)* (2010), pp. 185–192. [4](#)
- [LSC08] LUBOSCHIK M., SCHUMANN H., CORDS H.: Particle-Based Labeling: Fast Point-Feature Labeling Without Obscuring Other Visual Features. *IEEE Trans. Vis. Comput. Graphics* 14, 6 (2008), 1237–1244. [4](#)
- [MHST00] MORI K., HASEGAWA J., SUENAGA Y., TORIWAKI J.: Automated Anatomical Labeling of the Bronchial Branch and its Application to the Virtual Bronchoscopy System. *IEEE Trans. Med. Imag.* 19, 2 (2000), 103–114. [3](#)
- [MP09] MÜHLER K., PREIM B.: Automatic Textual Annotation for Surgical Planning. In *Vision, Modeling, and Visualization (VMV)* (2009), pp. 277–284. [3](#)
- [MTSP12] MOGALLE K., TIETJEN C., SOZA G., PREIM B.: Constrained Labeling of 2D Slice Data for Reading Images in Radiology. In *Eurographics Workshop on Visual Computing for Biology and Medicine (VCBM)* (2012), pp. 131–138. [3](#)
- [OFH*11] OELTZE S., FREILER W., HILLERT R., DOLEISCH H., PREIM B., SCHUBERT W.: Interactive, Graph-Based Visual Analysis of High-Dimensional, Multi-Parameter Fluorescence Microscopy Data in Toponomics. *IEEE Trans. Vis. Comput. Graphics* 17, 12 (2011), 1882–1891. [2, 9](#)
- [OJP14] OELTZE-JAFRA S., PREIM B.: Survey of Labeling Techniques in Medical Visualizations. In *Eurographics Workshop on Visual Computing for Biology and Medicine (VCBM)* (2014), p. this volume. [3](#)
- [OKH*12] OELTZE S., KLEMM P., HILLERT R., PREIM B., SCHUBERT W.: Visualization and Exploration of 3D Toponome Data. In *Eurographics Workshop on Visual Computing for Biology and Medicine (VCBM)* (2012), pp. 115–122. [2, 10](#)
- [PRS97] PREIM B., RAAB A., STROTHOTTE T.: Coherent zooming of illustrations with 3d-graphics and text. In *Graphics Interface* (1997), pp. 105–113. [4](#)
- [RPRH07] ROPINSKI T., PRASSNI J.-S., ROTERS J., HINRICHS K.: Internal Labels as Shape Cues for Medical Illustration. In *Vision, Modeling, and Visualization (VMV)* (2007), pp. 203–212. [3](#)
- [SBP*06] SCHUBERT W., BONNEKOH B., POMMER A. J., PHILIPSEN L., BÖCKELMANN R., MALYKH Y., GOLLNICK H., FRIEDENBERGER M., BODE M., DRESS A. W. M.: Analyzing Proteome Topology and Function by Automated Multi-dimensional Fluorescence Microscopy. *Nat Biotechnol* 24, 10 (2006), 1270–1278. [2, 6, 7, 8](#)
- [Sch03] SCHUBERT W.: Topological proteomics, toponomics, MELK-technology. *Adv Biochem Eng Biotechnol* 83 (2003), 189–209. [1, 2](#)
- [Sch10] SCHUBERT W.: On the origin of cell functions encoded in the toponome. *J Biotechnol* 149, 4 (2010), 252–259. [1, 2](#)
- [SGK*12] SCHUBERT W., GIESELER A., KRUSCHE A., SEROCKA P., HILLERT R.: Next-generation biomarkers based on 100-parameter functional super-resolution microscopy TIS. *New Biotechnology* 29, 5 (2012), 599–610. [2](#)
- [SGKH09] SCHUBERT W., GIESELER A., KRUSCHE A., HILLERT R.: 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 (2009), 2696–2707. [8](#)
- [SV10] SPECKMANN B., VERBEEK K.: Necklace Maps. *IEEE Trans. Vis. Comput. Graphics* 16, 6 (2010), 881–889. [2, 4, 5, 9](#)
- [TGM*14] TOMINSKI C., GLADISCH S., KISTER U., DACHSELT R., SCHUMANN H.: A Survey on Interactive Lenses in Visualization. In *EuroVis State-of-the-Art Reports* (2014), pp. 43–62. [4](#)
- [Tuf01] TUFTE E. R.: *The Visual Display of Quantitative Information*, 2nd ed. Graphics Press, 2001. [5](#)