Visual analytics for spatially resolved omics data at single cell resolution: methods & applications

Somarakis, A.

4

Cohort Comparison

This chapter was adapted from:

A. Somarakis, M. E. Ijsselsteijn, S. J. Luk, B. Kenkhuis, N. F. C. C. de Miranda, B. P. F. Lelieveldt, and T. Höllt, "Visual cohort comparison for spatial single-cell omics-data," *IEEE Transactions on Visualization and Computer Graphics*, vol. 27, no. 2, pp. 733–743, 2021.

4.1. ABSTRACT

Spatially-resolved omics-data enable researchers to precisely distinguish cell types in tissue and explore their spatial interactions, enabling deep understanding of tissue functionality. To understand what causes or deteriorates a disease and identify related biomarkers, clinical researchers regularly perform large-scale cohort studies, requiring the comparison of such data at cellular level. In such studies, with little *a-priori* knowledge of what to expect in the data, explorative data analysis is a necessity. Here, we present an interactive visual analysis workflow for the comparison of cohorts of spatially-resolved omics-data. Our workflow allows the comparative analysis of two cohorts based on multiple levels-of-detail, from simple abundance of contained cell types over complex co-localization patterns to individual comparison of complete tissue images. As a result, the workflow enables the identification of cohort-differentiating features, as well as outlier samples at any stage of the workflow. During the development of the workflow, we continuously consulted with domain experts. To show the effectiveness of the workflow, we conducted multiple case studies with domain experts from different application areas and with different data modalities.

4.2. INTRODUCTION

Omics-data describe biochemical properties, such as genomics, transcriptomics, proteomics, or metabolomics of biological systems [1], such as cells. In recent years, high-resolution spatial measurements of such systems have become available. State of the art spatially-resolved omics modalities [2–6] enable the precise characterization of cellular populations in tissue, enabling the discovery and identification of novel cell types[7] in large cohorts of samples. Information about the cell type, in combination with the specific location of each cell creates many heterogeneous multi-cellular patterns.

With the identification of these multi-cellular patterns, a crucial question arises; are such patterns correlated with clinical information, such as survival rate? Current research findings [8–10] support the clinical importance of analysing spatial multi-cellular interactions. Hence, the development of workflows for the systematic comparison of cohorts consisting of spatially-resolved omics-data with specific clinical characteristics is essential for the understanding of tissue functionality.

In the majority of life-science studies, the comparison of cohorts of samples is based on statistical comparison of predefined finite number of elements [11–14]. However, traditional statistical approaches, based on prior knowledge pose the risk of missing unexpected correlations and cannot capture the vast combinatorial space [15] of spatial configurations for all different cell types. Moreover, they depend on high quality input which often cannot be guaranteed with single-cell omics-data due to uncertainty in cell segmentation and cell type identification. Comparative visualization [16] can provide useful insights into the differentiating factors of two cohorts and enables the interactive, data-driven exploration of the vast combinatorial space while simultaneously investigating the biological relevance and plausibility of findings with regard to the preprocessing.

Here, we extended our previous work focused on the identification and exploration of multi-cellular spatial interactions in single-cell omics-data [17] to enable interactive comparison of cohorts of such data. The main goals are to identify the characteristics that

differentiate a cohort, explore the cohorts' heterogeneity and relate these characteristics directly to the tissue. In some cases, just the comparison of the cell types abundance is adequate to differentiate cohorts. In other cases, a detailed comparison of contained cells and their specific neighborhoods, i.e. microenvironments is needed.

We propose an interactive, data-driven cohort comparison workflow. More specifically the main contributions of this paper are:

1. A workflow for the comparison of cohorts of spatially-resolved single-cell omics-data, specifically addressing the following tasks

   T1 compare cohorts based on the abundance of different cell types,

   T2 compare cohorts based on multi-cellular microenvironments,

   T3 detect outliers within each cohort, and

   T4 relate findings to their spatial position.

2. A protoype implementation of the described workflow

The remainder of this paper is structured as follows. We present related work in Section 4.3, followed by a brief description of target users, input data and tasks in Section 4.4. In Section 4.5, we describe the rationale behind our visual design and implementation in our prototype. We present a set of case studies and user feedback in Section 4.6. Finally, we discuss the limitations of our work and conclude in Section 4.7.

4.3. RELATED WORK

The visual analytics community spent considerable effort on approaches for the exploration of cohorts of medical data combining spatial and non-spatial features. Preim et al. [18] provide an overview of image-centric approaches [19–21] focused on the exploration of large imaging cohorts and derived attributes. For the data analysis, these approaches share linking of attribute views with image views to provide context, visual queries for direct feedback, and interactive definition of groups of attributes. They typically deal with traditional medical imaging databases, such as those acquired by computed tomography (CT) or magnetic resonance imaging (MRI).

Dealing with microscopic images, Screenit [22] offers a system of linked views, similar to our system, to explore the drug screening results of cell cultures at multiple levels of detail. However, only recently, spatially-resolved omics-data [2–4] have become a standard tool for the exploration of tissue structure at the cellular level. Consequently, only few visual analysis tools exist that address the specific needs of such data. Facetto [23] is a scalable framework that allows hierarchical cell type identification in large multiplexed images. histoCAT [24] enables the identification of cell types and the significant pairwise spatial interactions between them. CytoMAP [25] offers an extensive toolbox for the exploration of tissue structure based on the analysis of spatial interactions. In our previous work on ImaCytE [17], we propose an interactive exploratory pipeline for cell type identification and neighborhood analysis in spatial single-cell data. Minerva [26] extends such exploration concepts with storytelling tools, to support communication and sharing of results. All of the above focus on the identification or exploration of cell types or significant multi-cellular

interactions in a single cohort of spatial single-cell data. Here, we use some of the concepts introduced in these works and extend them to introduce the first workflow for comparative analysis of two cohorts of such data, based on the abundance of cell types, as well as colocation patterns.

Based on a survey on existing comparative visualization tools [27], Gleicher et al. define a taxonomy that divides comparative visualization into juxtaposition (side-by-side placement), superposition (layering), and explicit encoding. A large body of work on comparative visualization for individual images exist. For example, Blaas et al. [19] combine superposition with explicit coding of the differences using complementary colors for the comparands, which cancels out in regions without differences. We use the same technique in some of our charts. Lindemann et al. [28], Maries et al. [29] and Ma et al. [30] utilize juxtaposition in an interactive comparative visualization pipeline for one-to-one comparison of segmentation results of brain imaging data. Juxtaposition for the comparison of images is also utilized in our work.

Schmidt et al. [31] facilitate the comparison of images with small differences within an ensemble. Raidou et al. [32] compare volume data and corresponding segmentations of bladders to explore the results of longitudinal radiotherapy treatment studies. Both works focus on all-to-all comparison of (3D) images in a single group, compared to the between-cohort comparison presented in this work. Basole et al. [33] as well as Wagner et al. [34] propose pipelines for the comparison of two cohorts. In their comparison workflow they use the same visual enocdings in order to compare the cohorts as a whole and simultaneously provide information for the intra-cohort heterogeneity, similar to the visual encodings we utilize in our system. Both approaches are limited to non-spatial healthcare data, though. Zhang et al. [35] present a visual analytics approach to compare two cohorts of diffusion tensor images. While we took some inspiration from their work, such as using complementary colors for the two cohorts that cancel each other out when overlapping, ultimately, the solutions described in their work are specific to tensor data and not easily transferrable to the spatial single-cell data described here.

4.4. ABSTRACTION

Recent developments in the spatially-resolved omics field manifest a wide variety of available modalities [3, 5, 36, 37]. These technologies measure transcriptomics or proteomics information at sub-cellular resolution, resulting in high-resolution image data with tens to thousands of values per pixel. Since researchers are interested in this information per cell, rather than per pixel, these images are typically pre-processed by segmenting individual cells and aggregating the values of the segmented pixels. Based on this aggregated information and potentially further features like morphology, the function and type of the segmented cells can be identified [24]. Both, cell segmentation [24, 38], as well as cell type identification [17, 23–25] in this kind of data is an active research topic. Large variations in cellular morphology and different quality of marker staining, among others, can lead to a considerable amount of uncertainty in the result of these preprocessing steps, making the validation, for example by referencing the actual images, during comparison imperative.

4.4.1. TARGET USERS AND GOALS

Our proposed workflow is targeted at clinical researchers who want to analyze their own data, for example to do an initial exploration of the data to form hypotheses. Typical goals when doing comparative analysis of two cohorts of spatial single-cell data could be the identification of cell types that are abundant in one cohort but not the other or cell co-localization patterns that are correlated with one of the cohorts. Such correlations or biomarkers [39] can be used for prognosis, monitoring or therapy of disease. While scripting in python or R is becoming more common in the domain, all our collaborators prefer visual exploration through GUI interfaces. Our proposed workflow is the first such visual exploration system that supports the comparative analysis of two cohorts of spatial single-cell data.

4.4.2. INPUT DATA

The overarching goal of our workflow is the comparison of two cohorts of spatially-resolved omics data as briefly introduced above. A single cohort consists of a set of samples, i.e., segmented and classified images as described above. Depending on the goal of the study, the samples consist of multiple images from a single subject or an arbitrary number of samples from multiple subjects. Typically, the two cohorts describe different populations, for instance, cancer patients who respond well to treatment in one cohort and those who respond worse in the second. A typical cohort consists of tens to hundreds of images, each consisting of thousands of segmented cells.

In a typical study, tens to hundreds of different cell types will be identified. The granularity depends on the goal of the study, as well as the data modality. For example, the Vectra imaging system [40] measures only a few different proteins (i.e. 4 in the case study in Section 4.6.3). Assuming differentiation into only low and high abundance, this results in an upper limit of 2^4 16 differentiable cell types. Other systems, such as Imaging Mass Cytometry, allow the measurement of up to 40 proteins, such that the number of cell types is limited rather by which types are of interest for the given study. A broad study would capture in the order of a hundred different cell types.

For each sample, we store the segmentation mask including a cell type label, i.e. class, for each segmented cell. Based on the cell segmentation mask, we derive the microenvironment for each cell. The microenvironment consists of the cell types and their abundance in the neighborhood of the given cell. We store the corresponding information per cell as a list of all cells that are contained in the microenvironment. The microenvironment of a cell varies according to the resolution of the modality and the type of sample. For example, in a tumor crowded with compact cells we would consider cells belonging to the microenvironment in a smaller distance, compared to brain tissue, where interacting cells can be further apart. Therefore, the distance defining the microenvironment of a cell is specified by the user. Typically, the microenvironment of a cell consists of no more than some tens of cells.

4.4.3. IDENTIFIED TASKS

In the following, we describe a set of tasks that we have identified in close collaboration with our domain expert partners from the pathology department at LUMC (co-authors of this manuscript). In general, we compare the two cohorts, based on the contained samples.

The first step of the workflow is comparing the cohorts according to the abundance of different cell types per sample (**T1**). This allows a simple differentiation of the cohorts based on the contained cells. In the second step, we further want to identify patterns in the cells' microenvironments that differentiate the cohorts. In **T2**, we compare cohorts based on multi-cellular microenvironments. Throughout the process we support visual detection of outliers within each cohort (**T3**), according to the abundance of contained cells and their microenvironments, and relate any findings to their spatial position (**T4**).

In the following, we describe and abstract **T1**-**T4** in more detail using Brehmer and Munzners task typology [41]. For references to this typology, we use a `mono-spaced` font.

T1 Cohort comparison based on the abundance of different cell types and combinations thereof in cohort samples. The relative abundance of a cell type in the samples forming a cohort and how much a specific subject deviates from the distribution within the cohort are important clinical biomarkers. As cell types can be of different granularity, it should also be possible to compare the cohorts, based on combinations of cell types. A trivial example is differentiating a cohort of cancer patients and a cohort of healthy subjects by comparing the abundance of tumor cells in the contained samples, where "tumor cells" can be a single cell type, or a combination of cell types according to a more fine grained definition. In this task **T1**, the user `compares` the two cohorts based on the abundance of different cell types within samples forming the cohort `discovering` and `locating` the cell type(s) that differentiate the two cohorts. The input for **T1** is the abundance of each cell type for each sample that we `summarize` as distributions over all samples in one cohort. The output is a list of cell types that differentiate the two cohorts.

T2 Cohort comparison based on multi-cellular microenvironments. The goal of **T2** is to compare the two cohorts according to the spatial co-localization patterns of each sample, as the comparison only based on cell type abundance is not enough to assess tissue functionality. Domain researchers hypothesize that tissue functionality also depends on the cell's interactions with other cells. While co-localization does not automatically lead to such interactions, it is a pre-condition. We facilitate the identification of such spatial features by breaking this task down into a high-level comparison, based on how often any two cell types are spatially co-located (**T2.a**), and a detail comparison where complex user-defined microenvironments can be explored (**T2.b**). In task **T2.a**, the user `discovers` combinations of two cell types that are most differentiating between the two cohorts. The input for this task is the abundance of each combination of two cell types in a microenvironment within the cohort sample. The output is a combination of two cell types to be used for further `exploration`. In task **T2.b**, the user further `explores` and `compares` the two cohorts based on more complex microenvironment compositions. Therefore, the user `produces` these more complex microenvironments by combining different cell types, typically starting with the combination found in **T2.a**. The input for **T2.b** is the complete set of cell microenvironments, optionally filtered to those including the combination of interest `discovered` in **T2.a**. The output is a set of detailed microenvironments differentiating the two cohorts.

T3 **Outlier detection within each cohort.** Detecting outliers within a cohort can provide additional important clinical information. For example subjects with different stages of a disease in the same cohort might exhibit different cell profiles [42]. Therefore, **T3** consists of `identifying` and `locating` outlying samples and their corresponding features identified in **T1** and **T2**. The input to this task is the abundance of cells and their microenvironments, as `identified` in **T1** and **T2**. The output is a list of outlying samples.

T4 **Relate findings to their spatial position.** As described above, **T1**-**T3** can be carried out based on cell abundance and microenvironment descriptions per sample, without consulting the actual imaging data. However, to verify individual findings we inspect the cells and their neighborhoods in their tissue context. Therefore, **T4** relates any findings to their spatial position. The analyst `locates` the structure of interest in their spatial location and `identifies` issues that were not apparent in the abstract representation. The input to **T4** are the segmented images and a structure of interest found with **T1**-**T3**, the output is a verified or rejected finding from **T1**-**T3**.

4.5. WORKFLOW

We designed a workflow to support the four tasks, identified and described in Section 4.4.3 and implemented it in a multiple-linked-views system, shown in Figure 1. The system is divided in three main blocks, where the left (Figure 1a) and right (Figure 1c) blocks support **T1** and **T2**, respectively by comparing the cohorts based on their cell type abundance and spatial interactions. **T4** relies on the inspection of tissue samples and supports **T1**-**T3**.

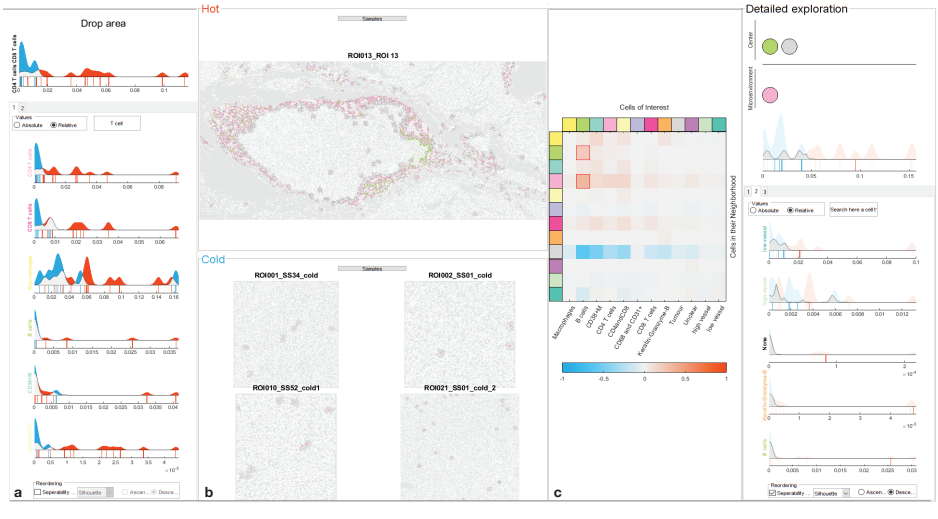

Figure 1: ***Screenshot of our integrated system*** *including the view for the comparison based on the cell abundance using raincloud plots (a), the tissue view, showing selected samples of the two cohorts (b), and the multi-cellular microenvironment comparison view using a difference heatmap and raincloud plots (c).*

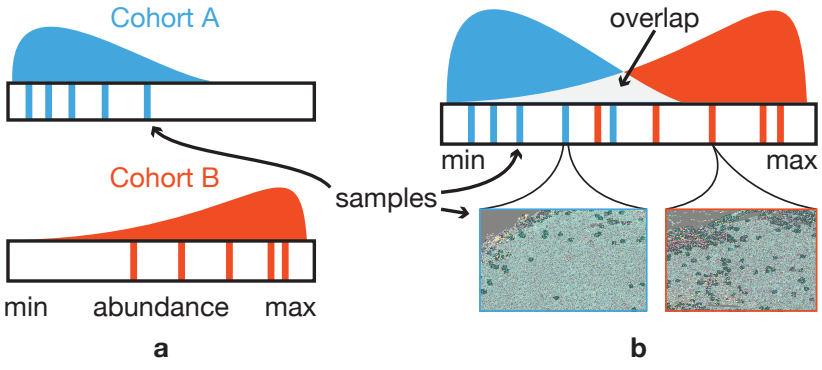

Figure 2: ***Comparison of two cohorts based on a cell type abundance.*** *(a) Individual raincloud plots for two cohorts showing the distribution (cloud) of samples (rain drops) according to the abundance of a contained cell type. (b) Superposition makes the difference visible by the large amount of color and small light-gray overlap area in the area chart.*

Therefore, we show the corresponding images between the views (Figure 1b) for **T1** and **T2** to support the user in directly making the connection for structures identified in any of the tasks to their spatial position. All views allow filtering the data to support visual outlier detection (**T3**).

4.5.1. Comparison Based on Cell Type Abundance

In the first step, we are interested to compare two cohorts according to the abundance of the different existing cell types in each of the contained samples (**T1**) and visually detect possible outliers in each of the cohorts (**T3**). Therefore, we first compute the number of cells of each type within each sample and then visualize the distribution of samples within both cohorts according to this value by superposing two simplified versions of raincloud plots [43]. This plot consists of a density (estimated using a kernel density estimate) plot showing the distribution of samples (the *cloud*) above a one-dimensional scatterplot with vertical lines as marks for the individual samples (*rain*-drops). This combination has proven very effective for our goals in **T1**-**T3**. The superposition of the density plots has shown to be very effective for the comparison of two distributions [44]. Both, the density plot [45] and the one-dimensional scatterplot [46], support visual detection of outliers. Furthermore, individual samples can be efficiently selected in the scatterplot for filtering. Additionally, for easier comparison between samples of different sizes, we enable the user to select whether the x-axis should represent the number of cells either as absolute values, or relative to the number of cells in that sample. As our primary goal is the comparison of the two cohorts, rather than the shape of individual plots, we want to emphasize the differences, rather than the commonalities. Therefore, following the same principle as Blaas et. al. [19], we use complementary colors for the two cohorts, i.e. blue and orange and blend the PDFs additively to receive a neutral light-gray in the overlapping areas as shown in Figure 2b.

The resulting raincloud plot allows the comparison of the composition of the two cohorts, according to the abundance of a single cell type within the contained samples. To allow the inspection of these distributions for all cell types, we use a small multiples approach [47, Chapter 4] and show the raincloud plots for several cell types in the same

view (Figure 3).

As indicated in Section 4.4.2, some studies can contain up to 100 different cell types. Finding a specific type of interest or the types that are the most differentiating for the two cohorts manually is not feasible in such a case. Therefore, we provide the possibility to sort the plots according to how well the corresponding distributions of the two cohorts separate, by default using the Silhouette metric [48], as it is invariant to the range of the input data. For advanced users we provide a set of other metrics, such as Dunn's index [49] which is efficient for compact and well separated clusters. In addition, we provide filtering by means of a textual search box (Figure 3a), based on the cell labels in the input data. Typing, for example *tumor* in this box will bring plots with the term *tumor* in their provided label to the top of the view (Figure 3b).

In some cases, the analyst might also be interested in aggregating the information on several cell types. For example, when several different cancer cell sub-types were identified in the original classification, but the analyst is only interested in how the cancer cells are distributed as a whole. To that end, we enabled the user to combine cell types, by gradually dragging and dropping the corresponding plots into a drop area on top of the

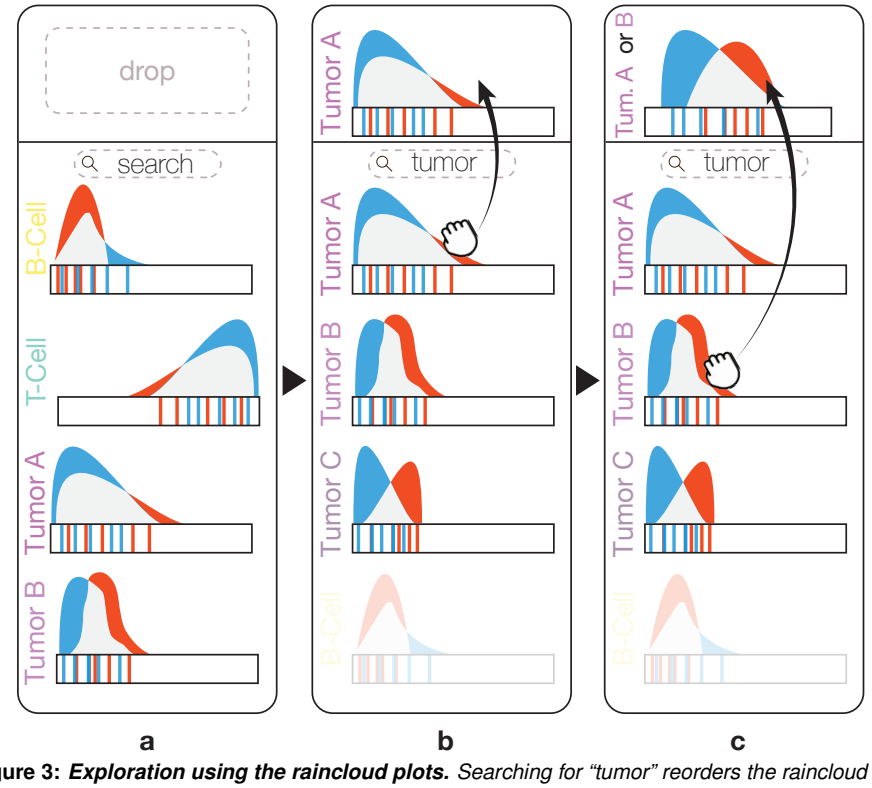

Figure 3: ***Exploration using the raincloud plots.*** *Searching for "tumor" reorders the raincloud plots by placing the plots corresponding to cell types containing the term 'tumor" in their label on top of the list (b). Dragging a raincloud plot and dropping it in the drop area (b,c), creates a new raincloud plot depicting the abundance of the cell types represented from the accumulated dropped raincloud plots.*

view (Figure 3b,c). The abundances of the dropped cell types are then aggregated as if they were a single cell type and a new distribution is created on-the-fly.

All views in our system are linked and allow cross-selection. For example, selection one or more lines in a raincloud plot filters the tissue view (Figure 1b) to show only the corresponding samples, with the cell type corresponding to the raincloud plot emphasized (**T4**). Further, these samples are also highlighted in the other raincloud plots, for example to verify whether a sample that is an outlier for one cell type also shows different behavior for other types (**T3**). To ensure that outliers in one cohort are not occluded by samples of the other cohort, the user can select to fade out one of the cohorts (**T3**).

4.5.2. COMPARISON BASED ON CELLULAR MICROENVIRONMENTS

The comparison of the cohorts based on their spatial interactions patterns, as indicated in task **T2**, is performed in two steps. The first step is to gain a global overview and compare the cohorts based on pairwise co-occurrences of cell types (**T2.a**). In the second step, the analyst can go into detail, explore and built specific, detailed microenvironments, consisting of an arbitrary number of cell types, and compare the distribution of these microenvironments among the two cohorts (**T2.b**). Throughout this process, we allow locating the identified microenvironments with the actual tissue images (**T4**) and in the second step, samples that are outliers in their cohort, according to the created microenvironment can be identified (**T3**).

4.5.2.1. PAIRWISE OVERVIEW

Following ImaCytE [17], we define the microenvironment of a cell, based on a user-defined distance as explained in Section 4.4.2.

We then compute the frequency for each cell type to occur in each other cell type's microenvironment throughout the cohort. For a detailed description we refer to our previous work [17, Section 4.3].

The result of this process is a directed and weighted graph, where each node represents a cell type and the link between two nodes defines the frequency of the target node appearing in the microenvironment of the source node. In ImaCytE, we visualize this frequency graph as a heatmap. Here, instead of showing the frequencies F, we compute the signed differences D in frequency between the two cohorts C_A and C_B. $D_t(C_A, C_B)$ $F(C_A) - F(C_B)$. We encode D using color based on the same heatmap layout, illustrated in Figure 4. The vertical axis shows the cell type of interest and the horizontal axis the cell types in the microenvironments. A large positive value indicates that the combination exists predominantly in Cohort A, while a large negative value means the combination predominantly exists in Cohort B. Based on this, we define a simple color map using the same colors previously assigned to the two cohorts and map the maximum absolute value $max(|D_t|)$ to the color assigned to Cohort A (i.e. blue) and $-max(|D_t|)$ to the color assigned to Cohort B (i.e. orange). Using the same concept of blending between the two colors, described in Section 4.5.1, the middle of this colormap, corresponding to D_t 0, will be a neutral light-grey, indicating both cohorts exhibit similar abundance of the given combination (compare Figure 4).

During one of the case studies (Section 4.6.1), it became clear that using the relative frequencies, used in ImaCytE [17] and the required normalization biased the heatmap

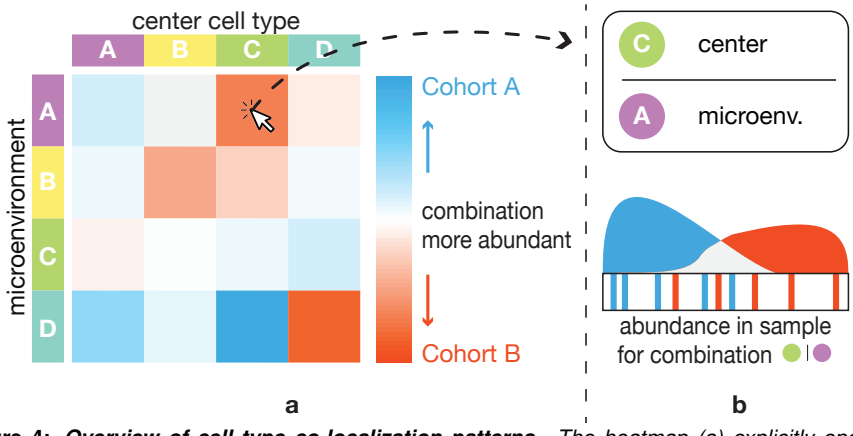

Figure 4: ***Overview of cell type co-localization patterns.*** *The heatmap (a) explicitly encodes differences in the abundance of pairwise combinations of cell types in the two cohorts. Clicking on one of the combinations sets this combination in the detail view (b), showing the distribution of samples according to the abundance of this combination.*

towards differences in small cell populations. To counter this issue, we provide the option to compute the heatmap using the separability metrics, also used for sorting the raincloud plots (Section 4.5.1). As these metrics only provide information on how different the cohorts are, we compute the mean abundance of the given cell type combination for all samples in a cohort and use the sign of the two cohort's difference in combination with the separability metric.

The resulting heatmap effectively shows cell type combinations that differentiate the two cohorts and for which cohort each combination is predominant. The analyst can now further explore individual combinations by clicking the corresponding box in the heatmap. Thereby, the corresponding combination is selected and highlighted in the tissue view (**T3**) and the microenvironment combination tool (Section 4.5.2.2) is pre-populated with the given combination (Figure 5a) for further analysis.

4.5.2.2. DETAIL MICROENVIRONMENTS

Starting with the overview of pairwise co-localization patterns, identified with the heatmap visualization, the analyst can now in detail explore complex microenvironment structures, based on any cell type combination and link those to individual samples along their position in the distribution of the corresponding cohort.

In ImaCytE [17], we used a simple glyph to enable the visual exploration of all the existing unique microenvironments in a sample. Here, the focus is on comparing two cohorts with regard to specific microenvironments, that potentially have already been identified as interesting in a previous analysis of the individual cohorts. Therefore, instead of showing all the existing unique microenvironments, the user can compare the two cohorts based on a specific pattern of spatial interactions. To enable the user to interactively define such a pattern, we utilize an interactive visual query system [50], similar to the one presented in Polaris [51] and further explained by Heer et al. [52]. The comparison of the two cohorts then happens with the same raincloud plots introduced in Section 4.5.1 but instead of the abundance of a single cell type the plot now displays the abundance of the queried microenvironment.

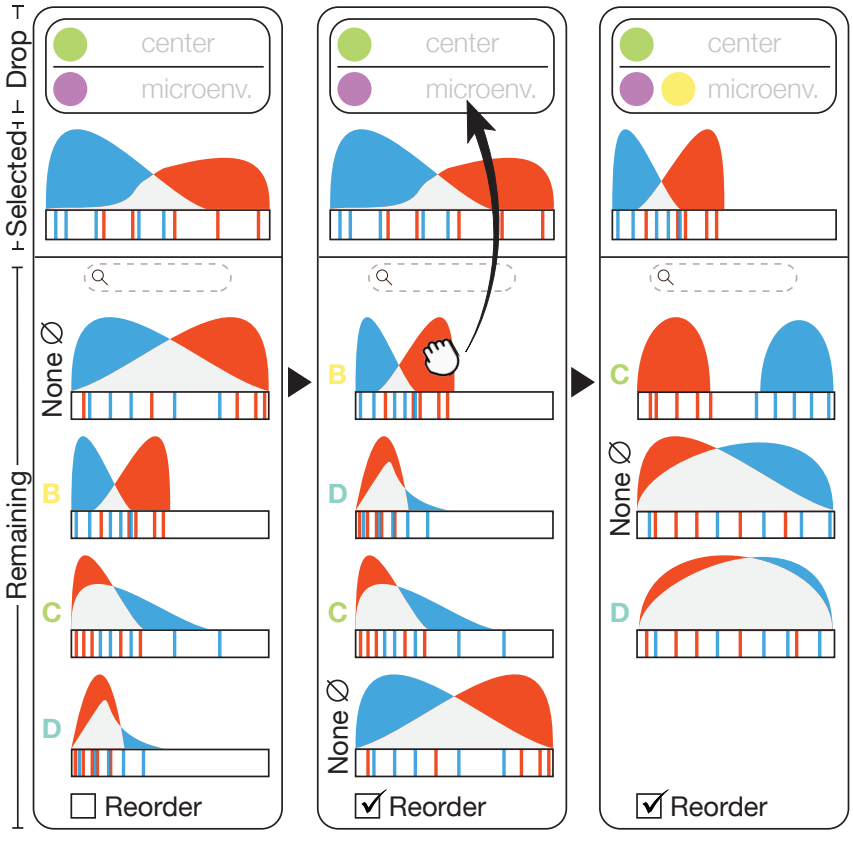

Figure 5: ***Interactive exploration in the detail view.*** *(a) The abundance of the cells fulfilling the cell type pattern in the* Drop *area is illustrated in the* Selected *raincloud plot. (b) The raincloud plots are reordered in the* Remaining *area according to their differentiating ability, the user drags the first raincloud plot and drops it in the* Drop *area. (c) The dropped raincloud plot replaces the previous one. Also, the* Drop *area and the* Remaining *plots are updated for further exploration.*

In practice, the analyst would typically start with a combination of two cell types picked from the heatmap. This simple microenvironment is illustrated on top of the detail view as illustrated in Figure 5a, where it is divided into the cell type of interest in the center of the microenvironment (i.e., cell type A, green circle, Figure 5a) and the microenvironment (i.e., cell type B, purple circle, Figure 5a). For the remainder of the paper we will denote microenvironments as ● | ●, where the circle(s) to the left of the vertical line represents the center cells combined with *or* type and the circle(s) to the right the microenvironment combined with *and*. I.e., a cell from either of the types left of the line must appear in the center and all the types to the right must appear in the surrounding of this cell. Below this (*Selected*, Figure 5a) we show the raincloud plot corresponding to the abundance of all microenvironments with at least the selected combination of cell types. Finally, further below (*Remaining*, Figure 5a) we depict the raincloud plots corresponding to the combination of the defined microenvironment plus any of the remaining cell types (here

●|●∅, ●|●●, ●|●●, ●|●●). The example in Figure 5a starts with *None* (indicated as ∅). At first glance, it might seem surprising that the corresponding raincloud plot is different from the initial plot above it. *None*, here means that no other additional cell type must exist in the microenvironment, whereas the initial plot shows all microenvironments that at least contain the given types. We denote this as ●|●∅. Below the *None* plot the remaining combinations are shown with the resulting raincloud plots. As described in Section 4.5.1, these plots can be ordered according to how strongly the corresponding microenvironment separates the two cohorts. Figure 5b illustrates the example after reordering. With this information the analyst can now continue exploring the microenvironments, for example by dragging the plot corresponding to cell type B (yellow) to the drop area, creating ●|●●, (Figure 5b). As the original plot already corresponded to the new microenvironment, we can now simply replace the "Selected" plot with the dragged plot (Figure 5c). The remaining raincloud plots (●|●●●, ●|●●∅, ●|●●●) are re-computed on-the-fly and shown below. Following this procedure, the user can progressively explore all interesting cell type combinations and evaluate their ability to discriminate the two cohorts and as such their potential as biomarkers.

As described in Section 4.5.1, the raincloud plots make it easy to identify samples that are outliers in their corresponding cohort (**T3**). Further, we provide the same linking and brushing features for selecting samples, as described in Section 4.5.1, to link the microenvironment patterns to the tissue view (**T4**).

4.5.3. TISSUE VIEW

In Section 4.4.2 we have described the importance of enabling the linking of any finding to its spatial location (**T4**). Therefore, we provide the tissue view (Figure 6), which shows the original segmented images and, linked to the other views, allows the inspection of selected cell types or microenvironments in the corresponding samples and their spatial context. The tissue view shows the images using color-coding for the different cell types. As we only consider the labeled segmentations as input (Section 4.4.2), we use a categorical colormap to assign a color to each label and thus cell type. We have chosen the qualitative *12 class Set 3* from colorbrewer [53] and have excluded blue and orange hues to avoid interference with the cohort colors. Colors are initially assigned based on the order of the cell type labels, but we allow the user to assign them manually by clicking on a cell type label. As typical studies have more cell types than the available ten colors, they can assign the same color to semantically grouped types. We then automatically adjust the saturation of hues that were selected multiple times to enable differentiation. While not described in detail in previous sections, this color scheme is used throughout the application to represent the different cell types and allow for easy mental linking between views. We have previously used a similar color scheme in ImaCytE [17]. To enable comparison between the cohorts, we divide the tissue view into two parts, one for each cohort. The name and color corresponding to the cohorts is shown on top of each view (Figure 6).

As described before, all views are linked. Therefore, the tissue view can be filtered to only show samples selected in other views. Further, selecting cell types or microenvironments in other views highlights them in the images by fading non-selected structures out, resulting in a light-grey for all unselected areas (Figure 6). Moreover, the tissue view supports zooming and panning across tissue samples to further assist the exploration of the

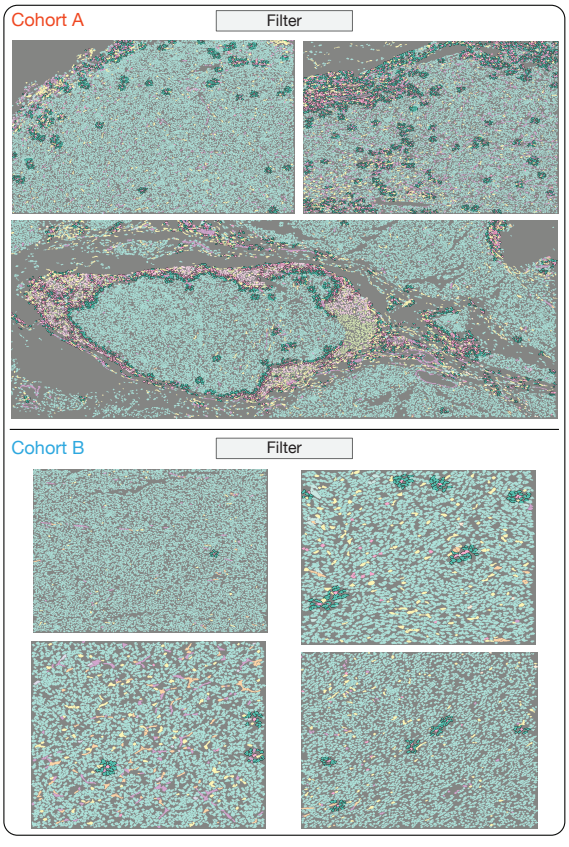

Figure 6: ***Tissue view,*** *highlighting a spatial interaction fading out the non-selected tissue structures. In the tissue samples of Cohort A, the spatial interactions form a compact structure, whereas the spatial interaction of Cohort B tissue samples are distributed all over the samples.*

(highlighted) tissue areas.

4.5.4. IMPLEMENTATION

As described in Section 4.4.1, our target users are clinical researchers with little programming experience. Therefore, we implemented the described workflow in a stand-alone GUI application. The application is implemented in MATLAB, as it allowed us to quickly build a stand-alone prototype. Source code and binaries are available on GitHub [54].

4.6. VALIDATION

In order to show the effectiveness of our workflow, we conducted three case studies with collaborators (P1-P3) at Leiden University Medical Center. P1 was also our main contact during the development of the workflow. After conducting the case studies and collecting feedback, we invited the collaborators to participate in the write up of the case-studies, and hence they are all co-authors of this manuscript. All collaborators acquired their own data with varying biological goals, using two different modalities as indicated in Table 4.1. For

Table 4.1: ***Summary of the case study characteristics.***

	Case Study	Modality	Samples in Cohort		Cell Types
			1	2	
P2	Sarcoma	IMC [3]	13	7	12
P1	Tumor	IMC [3]	19	28	60
P3	Alzheimer's	Vectra [40]	12	9	16

the case studies, we gave participants a hands-on introduction and answered any questions regarding the tool. After that, we observed the participants performing their analysis independently and reproduced their workflows for presentation in Sections. 4.6.1-4.6.3. As described in Section 4.5, for all the case studies the segmentation masks and the cell type identification had been performed as a pre-processing step by the participants. An overview of the study parameters with regard to imaging modality, numbers of samples, and numbers of included cell types is given in Table 4.1. As can be seen, the studies cover three different application areas, contain data from two different modalities, between 20 and 47 samples, and between 12 and 60 cell types. Finally, we asked the participants, as well as a fourth user of the software (P4, not a co-author of this manuscript), to fill out a short questionnaire (available in the supplemental material) via google forms [55]. The questionnaire consists of the ten standard System Usability Scale (SUS) statements [56], an additional nine statements specific to our tool, answered on a 5-point Likert scale, and five questions for open feedback. The individual plots presented in the case study have been exported directly from our tool and laid out with adjusted labels and annotations for the printout.

4.6.1. CASE STUDY I: SYNOVIAL SARCOMA (P2)

Synovial sarcoma is a rare form of cancer. During the immune response, T-cells infiltrate the sarcomas. Previous work has shown that synovial sarcomas can have areas with abundant T-cell infiltration (*hot areas*) and areas with very little T-cell infiltration (*cold areas*), in the same tumor[57]. The goal of this case study was to explore differences in the immune cell composition between these two types of areas. A total of 20 areas from 7 different tumors were imaged, of which 7 were cold (*Cold Cohort*, blue) and 13 were hot (*Hot Cohort*, orange). The size of the samples varied, with the number of cells in each image ranging from 2,678 to 23,774 cells. In the pre-processing step, cells were segmented and 12 different cell types were identified, based on the original data. While the number of cell types is relatively low, they cover a large range of available types, with rather coarse specificity.

4.6.1.1. CELL TYPE ABUNDANCE

In the first step of the analysis the expert was mostly interested in identifying cell type(s) that differentiate the cohorts, matching **T1** of our task analysis. Given the large variation in the number of cells per sample, he used the relative cell type abundance for comparison. First, he wanted to explore the uniformity of each cohort. As indicated above, the samples

were sorted into the two cohorts based on the infiltration of *T-cells* in the contained tumor tissue. Consequently the T-cells should exist predominantly in the Hot Cohort. As a first step, the expert wanted to verify this using the system. As there are two different types of T-cells in the dataset (*CD4* ● and *CD8 T-cells* ●) he first queried for these two cell types and created a combined raincloud plot by dragging the CD4 T-cell and CD8 T-cell plots to the combined drop area (subsection 4.5.1). The resulting combined plot (Figure 7a) confirmed that T-cells were largely non-existent in all seven samples of the Cold Cohort (blue peak close to 0, Figure 7a) but more widely distributed in the Hot Cohort (even spread of the orange distribution, Figure 7a). After navigating among the plots, he discovered the raincloud plot for *B-cells* ● (Figure 7b). This plot caught the expert's interest. Even though most samples from both cohorts hardly contain any B-cells, there are a few samples in the Hot Cohort that contain some B-cells, indicated by the orange lines to the right of the plot in Figure 7b. Given the generally low values, approximately 3 percent, even for the sample with the largest abundance, the expert decided to not further investigate these samples at this point and proceeded with other cell types. Therefore, he ordered the raincloud plots according to the Dunn's index [49]. The first plot illustrating *macrophages* ● showed a pattern similar to the T-cells (Figure 7c). Strikingly, there is an outlier (**T3**) clearly visible in the plot (highlight in Figure 7c). The corresponding sample from the Cold Cohort consists of over 16% macrophages, compared to no more than 5% for all other samples of the same cohort. Selecting the corresponding line in the plot also revealed that this sample has the highest abundance of T-cells in this cohort (though only at around 1% of cells in this sample).

At this point, the expert was curious whether the microenvironments of the macrophages

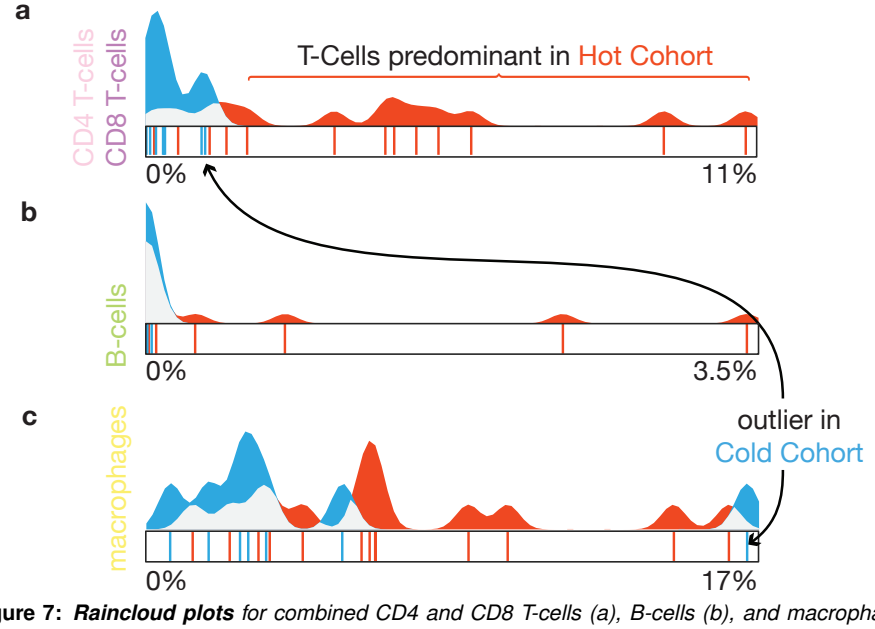

Figure 7: ***Raincloud plots*** *for combined CD4 and CD8 T-cells (a), B-cells (b), and macrophages (c). An outlier for macrophages in the cold cohort is clearly visible in (c). Selecting it showed it also contained slightly more T-cells than other samples in the cold cohort (a).*

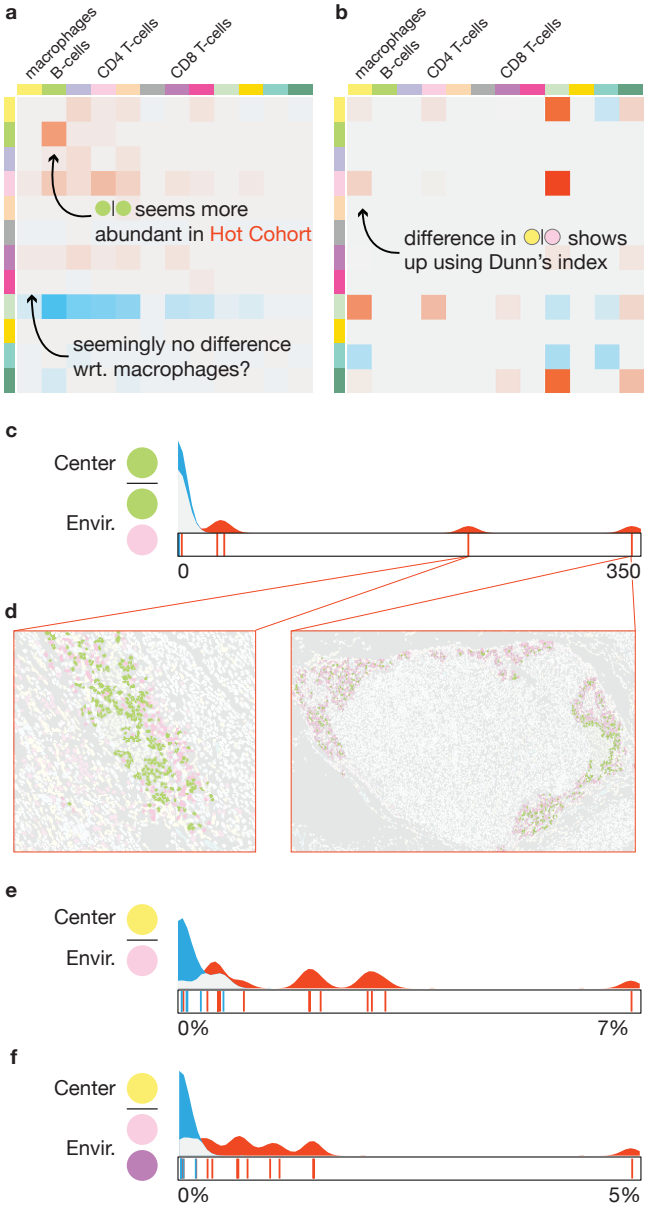

Figure 8: ***Multi-cellular microenvironment cohort comparison.*** *(a) A heatmap depicting the difference of the amount of pairwise spatial interaction between two cohorts normalized according to the abundance of each cell type. (b) A heatmap depicting the Dunn index for the samples of each cohort for each pairwise co-localization pattern. (c) The amount of B-cells having in their microenvironment B-cells and CD4 T-cells, depicting that the occured differentiation in (a) was due to the two outlier samples, which exist in a tertiary lymphoid structure, an interesting biological structure (d). The amount of macrophages having in their microenvironment CD4 T-cells (e) and CD8 T-cells (f).*

and B-cells could provide further clues on differentiating factors between and within the cohorts.

4.6.1.2. Micorenvironments

The exploration of the differences between the two cohorts, with regard to the contained microenvironments (**T2**) starts with the overview provided by the difference heatmap (Figure 8a). The difference heatmap (Figure 8a) indicated that combinations of B-cells and B-cells ● | ● and B-cells and T-cells ● | ● were more prevalent in the Hot Cohort (highlighted orange boxes). With this information, the expert created the combined mircoenvironment ● | ● ● using the drag and drop interface. The corresponding raincloud plot showed two clear outliers in the Hot Cohort showing a larger abundance of this combination (Figure 8c). Using the linked tissue view, the expert could highlight the microenvironments in the corresponding samples (Figure 8d). The expert observed that the highlighted microenvironments were mostly present in so-called tertiary lymphoid structures [57]. While not directly relevant for the cohort comparison, he noted the two outlier samples for later detailed inspection in his standard workflow.

In the previous step, the expert had also identified macrophages ● for further exploration. Curiously, the heatmap did not show any strong differences between the two cohorts with regard to the microenvironments of this cell type. After the case study, we analyzed the data and came to the conclusion that the normalization applied to create the heatmap (subsubsection 4.5.2.1) strongly biased the heatmap in favor of small cell populations such as the B-cells in this study (subsubsection 4.6.1.1). As a result, we added the option to use the same cluster separation metrics used for sorting the raincloud plots according to their power to separate the cohorts for the heatmap as described in subsubsection 4.5.2.1. Figure 8b shows the heatmap using the Dunn's index as an example. Here, the ● | ● microenvironment is more clearly visible, while the small values of the B-Cell microenvironments are suppressed. The expert selected the corresponding box from the heatmap and examined the distribution of the samples for each cohort in the detail view. The blue area around zero (Figure 8e) indicated the absence of ● | ● microenvironment in the Cold Cohort, verifying the heatmap findings. Then, the expert having already identified the correlation among CD8 T-cells and macrophages navigated among the plots of the "Remaining" area of the detail view and located the CD8 T-cell raincloud plot. The addition of CD8 T-cells in the microenvironment of macrophages ● | ● ● further differentiated the two cohorts, shown by the restriction of the blue area to almost zero (Figure 8f). Even the strong outlier in the Cold Cohort that contained the largest amount macrophages of all samples did not show any significant co-localization of macrophages and T-cells. On the other hand, several samples in the Hot Cohort showed significant amounts of both combinations. Therefore, the expert concluded that both T-cell sub-types seems to better differentiate the hot and cold tumor areas, than their one-to-one spatial interaction or even their abundances.

4.6.2. Case Study II: Tumor Metastasis (P1)

In this case study, the expert wanted to explore the differences in the cellular microenvironments of tumors with different clinical characteristics. In particular, she had acquired a data set, consisting of a total of 47 images taken from different tumor samples. Based on other clinical parameters she divided the set in two cohorts. The first one contains

19 images of non-metastatic tumors (*Non-Metastatic Cohort*, orange), the second 28 images of metastatic tumors (*Metastatic Cohort*, blue). She had segmented the images in a pre-processing step and identified 60 different cell types, among a total of 393,727 cells.

4.6.2.1. CELL TYPE ABUNDANCE

First, the expert was interested to discover cell type(s) which exist predominantly in one of the cohorts. Given the large amount of cell types, she ordered the raincloud plots according to the Silhouette metric in descending order, to assist her exploration. The first few plots consisted mostly of different subsets of T-cells, which had been defined in great detail in the preprocessing step. All of the corresponding plots showed a similar pattern of very small abundances for the Metastatic Cohort, indicated by a large blue peak to the left of the plot and a varying, but generally larger abundance in the Non-Metastatic Cohort. Searching for all cell types containing "T-cell" in their label showed a similar pattern for all of the remaining types (Figure 9a). This pattern is not completely surprising, as T-cells are a major factor in the immune response to cancer. For further exploration, in particular the relation of the identified T-cells to cancer cells, the expert aggregated all T-cell subsets using the drag and drop interface. The resulting raincloud plot (Figure 9b) confirmed that the T-cells ● clearly differentiate the two cohorts. There were, however, three samples from the Metastatic Cohort visible (blue lines, labeled A,B,C in Figure 9b) that showed a somewhat increased abundance compared to the remaining samples in that cohort. Next, the expert was interested, whether the increased amount of T-cells in the Non-Metastatic Cohort would correlate to differences in contained tumor cells. The expert searched for "tumor", to bring up the raincloud plot, corresponding to *Proliferating Tumor Cells* ●. However, as shown in Figure 9c, no clear separation between the two cohorts can be made, based on these cells. Finally, selecting the three outliers samples (A,B,C) in the T-cell plot did not show a specific differentiation with regard to the tumor cells.

4.6.2.2. MICORENVIRONMENTS

The last findings of subsubsection 4.6.2.1 intrigued the interest of the expert to further explore whether the tumor cells are present in the same amounts also in the microenvironment of T-cells. She quickly combined T-cells and proliferating tumor cells to a microenvironment ● | ● to bring up the corresponding raincloud plot (Figure 9d) in the detail view. The plot shows a clear differentiation among the two cohorts. In fact, this combination differentiates the two cohorts even stronger than only the T-cells. Even for the samples (Samples A,B,C) that showed increased abundance in T-cells, compared to the rest of the Metastatic Cohort, there was only a very small abundance of the ● | ● microenvironment. This strongly indicates that tumor cells exist in the microenvironment of T-cells in the Non-Metastatic Cohort, whereas in the Metastatic Cohort there is no spatial interaction between tumor and T-cells regardless their abundance. This lead the expert to hypothesize that the co-localization between the tumor and T-cells needs to be taken into account in tumor analysis, rather than the abundance of T-cells alone.

4.6.3. CASE STUDY III: ALZHEIMER'S DISEASE (P3)

The accumulation of *amyloid plaques* in the brain is an important characteristic of Alzheimer disease. These amyloid plaques are infiltrated by microglial cells, the resident immune cells of the brain. In this final case study, the expert wanted to verify the

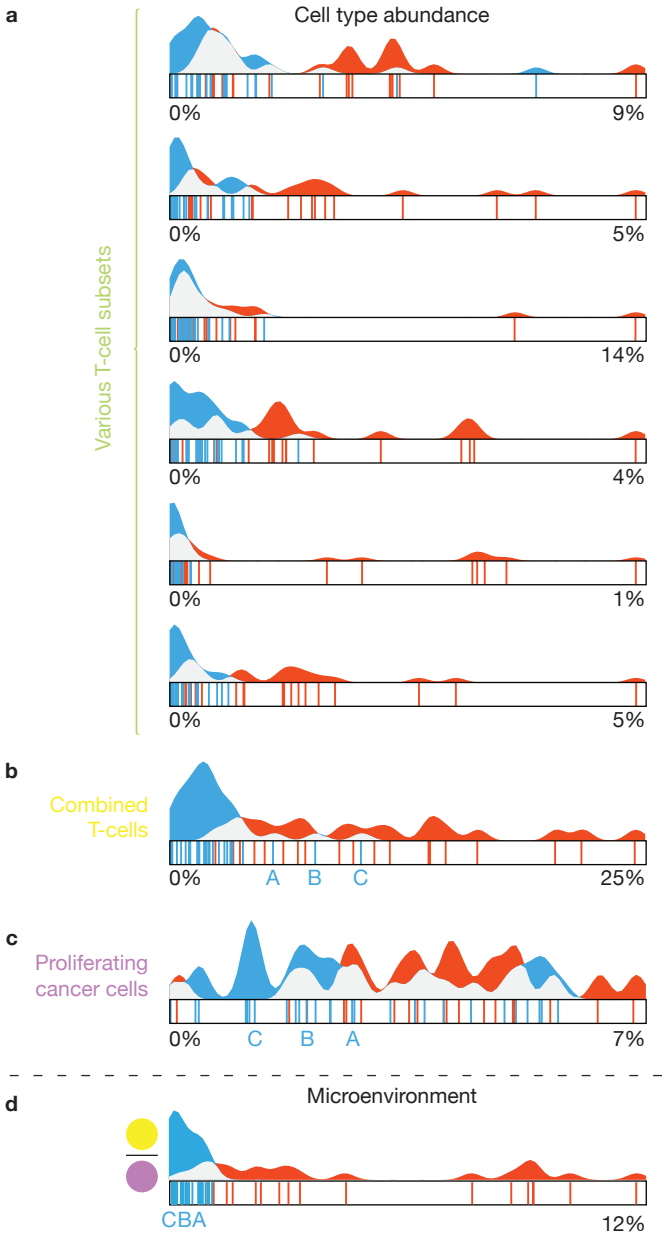

Figure 9: ***Raincloud plots*** *for various T-cell subsets (a), the aggregated plot combined from those subsets (b), and proliferating cancer cells (c). (d) shows the amount of the aggregated T-cells with proliferating cancer cells in their microenvironment. Even though the samples A-C, of the Metastatic Cohort, had a significant amount of T-cells and proliferating cancer cells (b,c) they did not spatially interact (d).*

hypothesis that the microglia cells close to and potentially attacking amyloid plaques are different from the microglia cells in healthy individuals.

The data used in this case study are somewhat different from the first two cases. The number of samples is comparable. Here, each sample represents one subject, for a total of 12 patients in the Alzheimer's Cohort (orange) and 9 healthy subjects in the Control Cohort (blue). However, each subject is described by up to 150 images, acquired with the Vectra 3.0 [40] machinery. 16 different cell types were identified and segmented in the pre-processing step. The identified cell types consist mostly of different subsets of microglia cells and as a result, the segmentation of the images is rather sparse, containing only in the order of 25 cells per image, plus the separately segmented amyloid plaques. As such, the individual images were not as important in this study as in the previous two and the data set only contained aggregated information of cell type abundance and microenvironments for all images per subject.

4.6.3.1. DATA ANALYSIS

As the experts goal was to verify a specific hypothesis, the data analysis in this study was much more targeted, compared to the rather explorative nature of the previous case studies. First, he brought up the raincloud plots corresponding to two microglia subtypes with contradictory patterns (Figure 10a,b). As can be seen in the plots *Subtype 1* ● was prevalent in the Control Cohort (blue), whereas Subtype 2 ● was mostly found in samples of the Alzheimer's Cohort (orange) but there was still some overlap between the samples from the two cohorts. This differentiation was already an indicator to verify the original hypothesis of the expert. Going back to the original data, the expert noted that the microglia Subtype 2

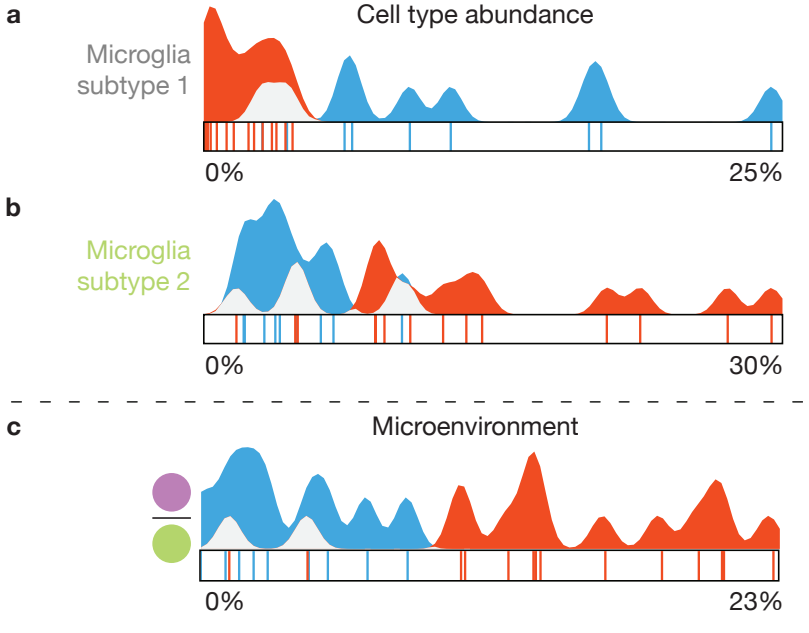

Figure 10: ***Raincloud plots*** *for microglia subtypes (a,b), and amoyloid plaques with microglia subtype 2 in their microenvironment (c).*

did not express two proteins that were expressed by Subtype 1 and hypothesized that these proteins might be suppressed when in the vicinity of the amyloid plaques ● in Alzheimer's disease patients. Consequently, he brought up the raincloud plot of the corresponding microenvironment ● | ● (Figure 10c). Here, the distinction between the two cohorts is even clearer, with only two samples from the Alzheimer's Cohort in the range of the Control Cohort. The distribution further indicates that Subtype 2 seems to co-localize with amyloid plaques, supporting the generated hypothesis.

4.6.4. Feedback

After the case studies, we collected feedback from the participants using a short questionnaire (available in the supplemental material) via google forms [55]. The questionnaire consists of the ten standard System Usability Scale (SUS) statements [56] (Q1–Q10), an additional nine statements specific to our tool (Q11–Q19), answered on a 5-point Likert scale, and five questions for open feedback. After the case studies, a fourth collaborator started working with the tool. After she got acquainted with it, we asked her to fill out the same questionnaire.

The average SUS-score, based on all four questionnaires was 76.25 with a standard deviation of 3.23 resulting in a *good* rating [58]. In the following we briefly summarize the feedback of the custom block of the questionnaire (Q11–Q19), for the complete set of responses we refer to the supplemental material. An overview of the responses is provided in Table 4.2. The custom part of the questionnaire is divided into three blocks. The first block (Q11–Q14) corresponds to the identified tasks (subsection 4.4.3). The second block (Q15–Q18) targets the interaction with the raincloud-based views in the cell abundance and microenvironment exploration. Finally, in the third block, we ask about general feedback.

With statements Q11–Q14 we queried whether **T1**–**T4** (subsection 4.4.3) could be carried out efficiently. (Q11; *The tool allows me to efficiently compare two cohorts, according to the abundance of contained cell types per sample* relates to **T1**, Q12 to **T2**, and so on). Generally, responses were clearly positive with strongly agree (++) or agree (+) with the exception of a neutral (○) response to Q11 and Q12, each. From the open feedback (Q20: *What functionality was missing to fully accomplish all goals?*) we could gather that participants would like to be able to "*correct[ion] cell abundance*" with regard to the amount of cells from user-defined area. Further, "*statistical testing of differences found between cohorts*" was requested, related to **T1** and **T2**.

In Q15–Q18 we were interested whether the raincloud plots were helpful to compare the distributions (Q15, **T1-T2**) and to find outliers (Q16, **T3**) as well as whether the drag and drop interaction made it easy to combine cell types (Q17) and build microenvironments (Q18). Q15–Q17 were overwhelmingly positive, with Q18 getting neutral responses by majority. The different response to Q17 and Q18 is rather unclear to us, as the interaction for combining cell types and building the detailed microenvironments is essentially the same. Unfortunately, there is also no further feedback on this in the open part of the questionnaire.

In the open feedback we can see that Participant 3 was missing "*Within subject distribution of cell types/clusters.*" As described in Figure 4.6.3, we had aggregated the very large amount of images in this study to a single dataset per subject. It might be interesting to provide a hierarchical approach in the future, that allows drilling into these subjects.

Table 4.2: ***Summary of participants' answers*** *to statements of our questionnaire on a 5-point Likert scale from very positive (++) to negative (-). No very negative (–) responses were given.*

Q	11	12	13	14	15	16	17	18	19
++	●●		●●	●	●●	●	●		
+	●	●●●	●●	●●●	●●	●●	●●●	●	●●
○	●	●				●		●●●	
-									●●

Participant 4 mentioned "*the option to compare 3 cohorts*" as a missing feature in the open feedback. While we focus on the comparison between two cohorts this is a possible future extension.

Finally, in the open feedback the "*possibility to detect outliers (and directly identify the subject*" (**T3**) was specifically mentioned as a positive aspect. The link between the abstract views and the actual images (**T4**) was highlighted by one participant: "*The rainbowplots are really cool, especially because you can go up and down to the images again.*" Particularly positive was a comment by Participant 1, that "*with the tool I already discovered a very nice thing in my existing data!*".

4.7. Discussion and Conclusion

We presented a workflow for the interactive visual comparison of two cohorts comprising single-cell omics-data, based on the cell abundance and their cell microenvironments.

The presented case studies contained up to 47 samples and up to nearly 400.000 cells. Our sorting and filtering options allow effective exploration of datasets of such sizes, however, increasing numbers to hundreds of samples will pose new challenges. In the Alzheimer's disease case study we accommodated a much larger original dataset (3286 images) by aggregating the information per patient and imaged region to a single larger image, resulting in the dataset described in Figure 4.6.3. Extending this to a hierarchical approach, facilitating the exploration of such aggregated regions and then individual images within a region might be a worthwhile extension.

At this point, our workflow is focused on two-dimensional images, as our partners currently only acquire such data. However, image stacks or volumetric measurements are becoming more readily available. Assuming a three-dimensional definition of microenvironments, the views based on abstract information, such as the raincloud plots and heatmap, would readily adapt to such data. Extensions to the spatial view, for example by volume rendering, would be necessary to inspect findings in the tissue context.

We have implemented the drag and drop interface to create simple center-neighborhood microenvironments. Nevertheless, the approach would support more advanced microenvironments through more drop targets, intuitively. For example, the neighborhood could be divided into multiple segments to allow a microenvrionment definition that has cell type A to the left and cell type B to the right of the center cell. A more traditional user interface, such as checkboxes, to assign cell types to each of those segments would be less flexible and quickly require a large amount of additional user interface elements.

Our workflow is designed to compare two clearly defined separate cohorts such as control vs. disease. Extending it to support more cohorts, or including more continuous features such as age or trial dose are open questions that certainly warrant future research.

ACKNOWLEDGEMENTS

This work received funding through Leiden University Data Science Research Programme. B.P.F.Lelieveldt received partial funding from H2020-Marie Skodowska-Curie Action Research and Innovation Staff Exchange (RISE) Grant 644373-PRISAR. N.F.C.C.de Miranda has received funding from the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation program (grant agreement No. 852832)

REFERENCES

[1] A. Conesa and S. Beck, "Making multi-omics data accessible to researchers," *Scientific data*, vol. 6, no. 1, pp. 1–4, 2019.

[2] R. Ke, M. Mignardi, A. Pacureanu, J. Svedlund, J. Botling, C. Wählby, and M. Nilsson, "In situ sequencing for rna analysis in preserved tissue and cells," *Nature Methods*, vol. 10, no. 9, pp. 857–860, 2013.

[3] C. Giesen, H. A. Wang, D. Schapiro, N. Zivanovic, A. Jacobs, B. Hattendorf, P. J. Schüffler, D. Grolimund, J. M. Buhmann, S. Brandt, Z. Varga, P. J. Wild, D. Günther, and B. Bodenmiller, "Highly multiplexed imaging of tumor tissues with subcellular resolution by mass cytometry," *Nature Methods*, vol. 11, no. 4, pp. 417–422, 2014.

[4] N. Crosetto, M. Bienko, and A. Van Oudenaarden, "Spatially resolved transcriptomics and beyond," *Nature Reviews Genetics*, vol. 16, no. 1, pp. 57–66, 2015.

[5] J. H. Lee, E. R. Daugharthy, J. Scheiman, R. Kalhor, T. C. Ferrante, R. Terry, B. M. Turczyk, J. L. Yang, H. S. Lee, J. Aach, K. Zhang, and G. M. Church, "Fluorescent in situ sequencing (FISSEQ) of rna for gene expression profiling in intact cells and tissues," *Nature Protocols*, vol. 10, no. 3, pp. 442–458, 2015.

[6] Y. Goltsev, N. Samusik, J. Kennedy-Darling, S. Bhate, M. Hale, G. Vazquez, S. Black, and G. P. Nolan, "Deep profiling of mouse splenic architecture with codex multiplexed imaging," *Cell*, vol. 174, no. 4, pp. 968–981, 2018.

[7] V. Van Unen, T. Höllt, N. Pezzotti, N. Li, M. J. Reinders, E. Eisemann, F. Koning, A. Vilanova, and B. P. Lelieveldt, "Visual analysis of mass cytometry data by hierarchical stochastic neighbour embedding reveals rare cell types," *Nature Communications*, vol. 8, no. 1, pp. 1–10, 2017.

[8] H. R. Ali, H. W. Jackson, V. R. T. Zanotelli, E. Danenberg, J. R. Fischer, H. Bardwell, E. Provenzano, O. M. Rueda, S.-F. Chin, S. Aparicio, C. Caldas, and B. Bodenmiller, "Imaging mass cytometry and multiplatform genomics define the phenogenomic landscape of breast cancer," *Nature Cancer*, vol. 1, no. 2, pp. 163–175, 2020.

[9] L. Keren and M. Angelo, "Mapping cell phenotypes in breast cancer," *Nature Cancer*, vol. 1, no. 2, pp. 156–157, 2020.

[10] H. W. Jackson, J. R. Fischer, V. R. Zanotelli, H. R. Ali, R. Mechera, S. D. Soysal, H. Moch, S. Muenst, Z. Varga, W. P. Weber, and B. Bodenmiller, "The single-cell pathology landscape of breast cancer," *Nature*, vol. 578, no. 7796, pp. 615–620, 2020.

[11] C. J. Newschaffer, K. Otani, M. K. McDonald, and L. T. Penberthy, "Causes of death in elderly prostate cancer patients and in a comparison nonprostate cancer cohort," *Journal of the National Cancer Institute*, vol. 92, no. 8, pp. 613–621, 2000.

[12] Y. Yuan, H. Failmezger, O. M. Rueda, H. Raza Ali, S. Gräf, S. F. Chin, R. F. Schwarz, C. Curtis, M. J. Dunning, H. Bardwell, N. Johnson, S. Doyle, G. Turashvili, E. Provenzano, S. Aparicio, C. Caldas, and F. Markowetz, "Quantitative image

analysis of cellular heterogeneity in breast tumors complements genomic profiling," *Science Translational Medicine*, vol. 4, no. 157, pp. 157ra143–157ra143, 2012.

[13] A. Nagaishi, M. Takagi, A. Umemura, M. Tanaka, Y. Kitagawa, M. Matsui, M. Nishizawa, K. Sakimura, and K. Tanaka, "Clinical features of neuromyelitis optica in a large japanese cohort: Comparison between phenotypes," *Journal of Neurology, Neurosurgery and Psychiatry*, vol. 82, no. 12, pp. 1360–1364, 2011.

[14] C. Robert, A. Ribas, J. D. Wolchok, F. S. Hodi, O. Hamid, R. Kefford, J. S. Weber, A. M. Joshua, W.-J. Hwu, T. C. Gangadhar, A. Patnaik, R. Dronca, H. Zarour, R. W. Joseph, P. Boasberg, B. Chmielowski, C. Mateus, M. A. Postow, K. Gergich, J. Elassaiss-Schaap, X. N. Li, R. Iannone, S. W. Ebbinghaus, S. P. Kang, and A. Daud, "Anti-programmed-death-receptor-1 treatment with pembrolizumab in ipilimumab-refractory advanced melanoma: A randomised dose-comparison cohort of a phase 1 trial," *The Lancet*, vol. 384, no. 9948, pp. 1109–1117, 2014.

[15] L. Cibulski and B. Preim, "Visual analytics support for analysis of cohort study data: Requirements and concepts," tech. rep., Otto-Von-Guericke University Magdeburg, 2016.

[16] H.-G. Pagendarm and F. H. Post, "Comparative visualization: Approaches and examples," in *Visualization in Scientific Computing* (M. Göbel, H. Müller, and B. Urban, eds.), Springer, 1995.

[17] A. Somarakis, V. van Unen, F. Koning, B. P. Lelieveldt, and T. Höllt, "ImaCytE: visual exploration of cellular microenvironments for imaging mass cytometry data," *IEEE Transactions on Visualization and Computer Graphics*, vol. 27, no. 1, pp. 98–110, 2021.

[18] B. Preim, P. Klemm, H. Hauser, K. Hegenscheid, S. Oeltze, K. Toennies, and H. Völzke, "Visual analytics of image-centric cohort studies in epidemiology," in *Visualization in Medicine and Life Sciences III. Mathematics and Visualization.* (L. Linsen, B. Hamann, and H. C. Hege, eds.), pp. 221–248, Springer, 2016.

[19] O. Dzyubachyk, J. Blaas, C. P. Botha, M. Staring, M. Reijnierse, J. L. Bloem, R. J. Van Der Geest, and B. P. Lelieveldt, "Comparative exploration of whole-body MR through locally rigid transforms," *International Journal of Computer Assisted Radiology and Surgery*, vol. 8, no. 4, pp. 635–647, 2013.

[20] M. D. Steenwijk, J. Milles, M. Buchem, J. Reiber, and C. P. Botha, "Integrated visual analysis for heterogeneous datasets in cohort studies," in *Proceedings of the Workshop on Visual Analytics in Healthcare (VAHC)*, 2010.

[21] Z. Zhang, D. Gotz, and A. Perer, "Interactive visual patient cohort analysis," in *Proceedings of the Workshop on Visual Analytics in Healthcare (VAHC)*, 2012.

[22] K. Dinkla, H. Strobelt, B. Genest, S. Reiling, M. Borowsky, and H. Pfister, "Screenit: Visual analysis of cellular screens," *IEEE Transactions on Visualization and Computer Graphics*, vol. PP, no. 99, pp. 1–1, 2017.

[23] R. Krueger, J. Beyer, W. D. Jang, N. W. Kim, A. Sokolov, P. K. Sorger, and H. Pfister, "Facetto: Combining unsupervised and supervised learning for hierarchical phenotype analysis in multi-channel image data," *IEEE Transactions on Visualization and Computer Graphics*, vol. 26, no. 1, pp. 227–237, 2020.

[24] D. Schapiro, H. W. Jackson, S. Raghuraman, J. R. Fischer, V. R. T. Zanotelli, D. Schulz, C. Giesen, R. Catena, Z. Varga, and B. Bodenmiller, "histoCAT: analysis of cell phenotypes and interactions in multiplex image cytometry data," *Nat Methods*, vol. 14, no. 9, pp. 873–876, 2017.

[25] C. R. Stoltzfus, J. Filipek, B. H. Gern, B. E. Olin, J. M. Leal, Y. Wu, M. R. Lyons-Cohen, J. Y. Huang, C. L. Paz-Stoltzfus, C. R. Plumlee, *et al.*, "CytoMAP: A spatial analysis toolbox reveals features of myeloid cell organization in lymphoid tissues," *Cell reports*, vol. 31, no. 3, p. 107523, 2020.

[26] R. Rashid, Y.-A. Chen, J. Hoffer, J. L. Muhlich, J.-R. Lin, R. Krueger, H. Pfister, R. Mitchell, S. Santagata, and P. K. Sorger, "Interpretative guides for interacting with tissue atlas and digital pathology data using the minerva browser," *bioRxiv*, 2020.

[27] M. Gleicher, D. Albers, R. Walker, I. Jusufi, C. D. Hansen, and J. C. Roberts, "Visual comparison for information visualization," *Information Visualization*, vol. 10, no. 4, 2011.

[28] F. Lindemann, K. Laukamp, A. H. Jacobs, and K. Hinrichs, "Interactive comparative visualization of multimodal brain tumor segmentation data," in *Proceedings of Vision, Modeling & Visualization (VMV)*, 2013.

[29] A. Maries, N. Mays, M. Hunt, K. F. Wong, W. Layton, R. Boudreau, C. Rosano, and G. E. Marai, "GRACE: a visual comparison framework for integrated spatial and non-spatial geriatric data," *IEEE Transactions on Visualization and Computer Graphics*, vol. 19, no. 12, pp. 2916–2925, 2013.

[30] C. Ma, F. Pellolio, D. A. Llano, K. A. Stebbings, R. V. Kenyon, and G. E. Marai, "RemBrain: exploring dynamic biospatial networks with mosaic matrices and mirror glyphs," *Journal of Imaging Science and Technology R*, vol. 61, no. 6, pp. 0–1, 2017.

[31] J. Schmidt, M. E. Gröller, and S. Bruckner, "VAICo: visual analysis for image comparison," *IEEE Transactions on Visualization and Computer Graphics*, vol. 19, no. 12, pp. 2090–2099, 2013.

[32] R. Raidou, O. Casares-Magaz, A. Amirkhanov, V. Moiseenko, L. Muren, J. Einck, A. Vilanova, and M. Gröller, "Bladder Runner : Visual analytics for the exploration of rt-induced bladder toxicity in a cohort study," *Computer Graphics Forum*, vol. 37, no. 3, pp. 205–216, 2018.

[33] R. C. Basole, H. Park, M. Gupta, M. L. Braunstein, D. H. Chau, M. Thompson, V. Kumar, R. Pienta, and M. Kahng, "A visual analytics approach to understanding care process variation and conformance," in *Proceedings of the Workshop on Visual Analytics in Healthcare (VAHC)*, 2015.

[34] M. Wagner, D. Slijepcevic, B. Horsak, A. Rind, M. Zeppelzauer, and W. Aigner, "KAVAGait: knowledge-assisted visual analytics for clinical gait analysis," *IEEE Transactions on Visualization and Computer Graphics*, vol. 25, no. 3, pp. 1528–1542, 2019.

[35] C. Zhang, T. Höllt, M. W. A. Caan, E. Eisemann, and A. Vilanova, "Comparative visualization for diffusion tensor imaging group study at multiple levels of detail," in *Proceedings of Visual Computing for Biology and Medicine (VCBM)*, pp. 53–62, 2017.

[36] A. M. Femino, F. S. Fay, K. Fogarty, and R. H. Singer, "Visualization of single rna transcripts in situ," *Science*, vol. 280, no. 5363, pp. 585–590, 1998.

[37] L. Keren, M. Bosse, S. Thompson, T. Risom, K. Vijayaragavan, E. McCaffrey, D. Marquez, R. Angoshtari, N. F. Greenwald, H. Fienberg, J. Wang, N. Kambham, D. Kirkwood, G. Nolan, T. J. Montine, S. J. Galli, R. West, S. C. Bendall, and M. Angelo, "MIBI-TOF: a multiplexed imaging platform relates cellular phenotypes and tissue structure," *Science Advances*, vol. 5, no. 10, p. eaax5851, 2019.

[38] P. J. Schüffler, D. Schapiro, C. Giesen, H. A. Wang, B. Bodenmiller, and J. M. Buhmann, "Automatic single cell segmentation on highly multiplexed tissue images," *Cytometry Part A*, vol. 87, no. 10, pp. 936–942, 2015.

[39] R. Mayeux, "Biomarkers: Potential uses and limitations," *NeuroRX*, vol. 1, no. 2, pp. 182–188, 2004.

[40] M. E. Ijsselsteijn, T. P. Brouwer, Z. Abdulrahman, E. Reidy, A. Ramalheiro, A. M. Heeren, A. Vahrmeijer, E. S. Jordanova, and N. F. de Miranda, "Cancer immunophenotyping by seven-colour multispectral imaging without tyramide signal amplification," *The Journal of Pathology: Clinical Research*, vol. 5, pp. 3–11, 2019.

[41] M. Brehmer and T. Munzner, "A multi-level typology of abstract visualization tasks," *IEEE Transactions on Visualization and Computer Graphics*, vol. 19, no. 12, pp. 2376–2385, 2013.

[42] V. van Unen, N. Li, I. Molendijk, M. Temurhan, T. Höllt, A. E. van der Meulen-de Jong, H. W. Verspaget, M. L. Mearin, C. J. Mulder, J. van Bergen, B. P. Lelieveldt, and F. Koning, "Mass cytometry of the human mucosal immune system identifies tissue- and disease-associated immune subsets," *Immunity*, vol. 44, no. 5, pp. 1227–1239, 2016.

[43] M. Allen, D. Poggiali, K. Whitaker, T. R. Marshall, and R. Kievit, "Raincloud plots: a multi-platform tool for robust data visualization," *Wellcome open research*, vol. 4, 2019.

[44] M. Blumenschein, L. J. Debbeler, N. C. Lages, B. Renner, D. A. Keim, and M. El-Assady, "v-plots: Designing hybrid charts for the comparative analysis of data distributions," *Computer Graphics Forum*, vol. 39, no. 3, pp. 565–577, 2020.

[45] M. Correll, M. Li, G. Kindlmann, and C. Scheidegger, "Looks good to me: Visualizations as sanity checks," *IEEE Transactions on Visualization and Computer Graphics*, vol. 25, no. 1, pp. 830–839, 2019.

[46] P. Kampstra, "Beanplot: A boxplot alternative for visual comparison of distributions," *Journal of Statistical Software*, vol. 28, no. 1, pp. 1–9, 2008.

[47] E. R. Tufte, *Envisioning information*. Graphics Press, 1990.

[48] P. J. Rousseeuw, "Silhouettes: A graphical aid to the interpretation and validation of cluster analysis," *Journal of Computational and Applied Mathematics*, vol. 20, no. C, pp. 53–65, 1987.

[49] J. C. Bezdek and N. R. Pal, "Cluster validation with generalized dunn's indices," in *Proceedings of Artificial Neural Networks and Expert Systems (ANNES)*, pp. 190–193, 1995.

[50] B. Shneiderman, "Dynamic queries for visual information seeking," *IEEE Software*, vol. 11, no. 6, pp. 70–77, 1994.

[51] C. Stolte and P. Hanrahan, "Polaris: A system for query, analysis and visualization of multi-dimensional relational databases," *IEEE Transactions on Visualization and Computer Graphics*, vol. 8, pp. 52–65, 2002.

[52] J. Heer and B. Shneiderman, "Interactive dynamics for visual analysis," *Communications of the ACM*, vol. 55, no. 4, p. 45–54, 2012.

[53] C. A. Brewer, G. W. Hatchard, and M. A. Harrower, "ColorBrewer in print: A catalog of color schemes for maps," *Cartography and Geographic Information Science*, vol. 30, no. 1, pp. 5–32, 2003.

[54] A. Somarakis and T. Höllt, "SpaCeCo open source software." https://www.doi.org/10.5281/zenodo.3885814, 2020. [Online; Accessed: 2021-07-14].

[55] GF, "Google forms." `https://www.google.com/forms/about/`. [Online; Accessed: 2020-04-20].

[56] J. Brooke, "SUS: a "quick and dirty" usability scale," in *Usability Evaluation in Industry* (P. W. Jordan, B. Thomas, B. A. Weerdmeester, and I. L. McClelland, eds.), pp. 189–194, Taylor and Francis, 1996.

[57] S. J. Luk, D. M. der Steen, R. S. Hagedoorn, E. S. Jordanova, M. W. Schilham, J. V. Bovée, A. H. Cleven, J. F. Falkenburg, K. Szuhai, and M. H. Heemskerk, "PRAME and HLA Class I expression patterns make synovial sarcoma a suitable target for PRAME specific t-cell receptor gene therapy," *OncoImmunology*, vol. 7, no. 12, p. e1507600, 2018.

[58] A. Bangor, P. Kortum, and J. Miller, "Determining what individual SUS scores mean: Adding an adjective rating scale," *Journal of Usability Studies*, vol. 3, pp. 114–1234, 1996.