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

Somarakis, A.

Citation

Somarakis, A. (2022, January 20). *Visual analytics for spatially resolved omics data at single cell resolution: methods & applications*. Retrieved from https://hdl.handle.net/1887/3250550

6

DISCUSSION

6.1. DISCUSSION

In each chapter of this thesis, we discussed in detail possible limitations and future directions specifically for each part of our end-to-end pipeline. Here, we will place this discussion in a wider context, and reflect on the overall strengths and weaknesses of our work to fully understand the possibilities and limitations and propose directions for future research.

As we have defined in Chapter 1, the cell is the basic unit of life. Therefore, the whole pipeline is based on the exploration of cellular patterns and characteristics. Furthermore, the high-resolution of the raw, highly multiplexed images offers information about the extracellular matrix and the distribution of molecular characteristics inside the cell. All this information is omitted in favor of the cellular analysis, introducing a limitation of our work. A "cell-free" analysis approach could replace the error-prone cell segmentation step and add information to the analysis pipeline. On the other hand, as the amount of high-dimensional cellular data extracted from an image is two to three orders of magnitude smaller than its contained high-dimensional pixel values, it would cause scalability issues. Fortunately, current developments [1] have enabled the interactive exploration of millions of data points enabling the possibility of a "pixel-based" analysis, circumventing the need for cell segmentation [2]. An adaptation of our proposed Visual Analytics methods for the inclusion of "pixel-based" analysis in the pipeline is an interesting direction for further development.

A factor that limits the broad applicability of our method is related to the novelty and consequently shortage of (annotated) data. The well-defined nature of the preprocessing problems make them ideal candidates for the efficient machine learning algorithms, given there are plenty of annotated data. With the current boom in the use of spatial-omics modalities [3] a huge volume of annotated data, ideal for the efficient development of fully-supervised solutions, is expected. Such solutions could increase the generality and efficiency of the presented preprocessing algorithms.

For the detailed in-depth exploration of the highly multiplexed cellular images, the cellular microenvironment characteristics are crucial. The exploration of the cellular microenvironment, though, requires first its definition. Before starting exploring the cellular microenvironment characteristics, in both Visual Analytics tools; ImaCytE and SpaCeCo, the microenvironment is defined. A cell's microenvironment is composed of cells that are below a specific distance. This way each cell of a cell's microenvironment, regardless of its distance, contributes the same in its microenvironment. However, the biological significance of a cell to its neighbor varies regarding their distance. Therefore, developing a weighted, or a multi-level approach of the algorithms proposed in this thesis may be more suitable for the exploration of cellular microenvironments.

Our visual analytics tools enable experts to generate hypotheses for a tissue's functionality. Mere hypotheses are not enough for the experts draw safe conclusions. Therefore, integrating hypotheses testing methods into the pipeline would facilitate the interpretation of our observation and relate them to specific clinical information.

The introduction of novel visual designs for the exploration of cellular characteristics, such as the glyphs for the motifs representation in ImaCytE and the visual query system in SpaCeCo, is challenging. Their perception and usefulness follows a steep learning curve and is highly correlated with the complexity of the analyzed data. Hence, it is important for the designer to provide to the users continuous support and explanation of the design concepts and be willing to "transfer" the design concepts according to the needs of the experts. A characteristic example is the utilization of the glyphs designed for the representation of the motifs for the representation of Abeta plaques microenvironment on the tissue image.

The above example highlights a second challenge that we should deal with during visual design process, the discrimination of visualization design and visual data analysis. A difficult task to be accomplished in the context of a PhD project, where close and continuous collaboration with clinical researchers is part of the daily routine. However, during the phase of visual design it is important not to slip under the pressure of experts for the analysis of specific datasets and abstract as much as possible the design solutions. To deal with this challenge and meet both the imminent and more generic needs of our collaborators, we developed two pure Visual Analytics projects (ImaCytE and SpaCeCo), where the exploration of spatial omics data can take place and contributed equally with our clinical collaborators to two other projects related with the analysis of specific data(sets).

In general, multiplexed cellular images have not been utilized to their full potential. Until now, experts analyze such data solely to identify the contained cell types or to observe in a qualitative manner the proximity of two cellular types. A detailed quantitative analysis of cellular microenvironments for the extraction of important information about tissue functionality is still missing, but it is highly anticipated [2, 4] and the development of high quality open-source software applications is necessary for developments in this direction. To that end, the daily usage of the open-source tools developed for our holistic pipeline by our collaborators in LUMC and other research labs [5] indicate that our work has been established as an important tool for the generation of significant biological hypotheses from the analysis of multiplexed cellular images. Moreover, the awarding of the third Dirk Bartz Prize for Visual Computing in Medicine 2021 and the honourable mention for the Karl-Heinz Höhne Award for Medical Visualization 2021 indicate the recognition of our solutions for the Medical Visualization community as well.

REFERENCES

[1] V. 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, p. 1740, 2017.

[2] N. L. de Vries, A. Mahfouz, F. Koning, and N. F. C. C. de Miranda, “Unraveling the complexity of the cancer microenvironment with multidimensional genomic and cytometric technologies,” *Frontiers in Oncology*, vol. 10, p. 1254, 2020.

[3] V. Marx, “Method of the year: spatially resolved transcriptomics,” *Nature Methods*, vol. 18, no. 1, pp. 9–14, 2021.

[4] N. Guo, V. van Unen, M. E. Ijsselsteijn, L. F. Ouboter, A. E. van der Meulen, S. M. Chuva de Sousa Lopes, N. F. C. C. de Miranda, F. Koning, and N. Li, “A 34-marker panel for imaging mass cytometric analysis of human snap-frozen tissue,” *Frontiers in Immunology*, vol. 11, p. 1466, 2020.

[5] J. O. Garcia, “Data analysis at Newcastle Unicersity flow cytometry core facility.” `https://www.ncl.ac.uk/fccf/research-development/data-analysis/`, 2021. [Online; Accessed: 2021-02-05].