











Visual Analysis of Tissue Images at Cellular Level

A. Somarakis¹ , Marieke E. Ijsselsteijn² , Boyd Kenkhuis^{3,1} , Vincent van Unen⁴ , Sietse J. Luk⁵ ,
Frits Koning⁶ , Louise van der Weerd¹ , Noel F.C.C. de Miranda² , Boudewijn P.F. Lelieveldt¹ , and T. Höllt^{7,1} 

¹Radiology Department, Leiden University Medical Center, Leiden, The Netherlands

²Pathology Department, Leiden University Medical Center, Leiden, The Netherlands

³Human Genetics Department, Leiden University Medical Center, Leiden, The Netherlands

⁴School of Medicine, Stanford University, Stanford, CA, United States of America

⁵Hematology Department, Leiden University Medical Center, Leiden, The Netherlands

⁶Immunology Department, Leiden University Medical Center, Leiden, The Netherlands

⁷Computer Graphics and Visualization Department, TU Delft, Delft, The Netherlands

Abstract

The detailed analysis of tissue composition is crucial for the understanding of tissue functionality. For example, the location of immune cells related to a tumour area is highly correlated with the effectiveness of immunotherapy. Therefore, experts are interested in presence of cells with specific characteristics as well as the spatial patterns they form. Recent advances in single-cell imaging modalities, producing high-dimensional, high-resolution images enable the analysis of both of these features. However, extracting useful insight on tissue functionality from these high-dimensional images poses serious and diverse challenges to data analysis. We have developed an interactive, data-driven pipeline covering the main analysis challenges experts face, from the pre-processing of images via the exploration of tissue samples to the comparison of cohorts of samples. All parts of our pipeline have been developed in close collaboration with domain experts and are already a vital part in their daily analysis routine.

CCS Concepts

• **Human-centered computing** → *Visualization systems and tools*;

1. Introduction

Spatially resolved omics are currently revolutionizing the understanding of the organization of cells and tissues. The journal *Nature Methods* just proclaimed spatially resolved transcriptomics as *method of the year 2020* [Mar21]. Spatially resolved transcriptomics and other spatially resolved omics methods allow the acquisition of highly multiplexed imaging data of various biochemical properties, such as transcriptomics or proteomics [CB19]. The resulting data enables experts to explore spatial cellular patterns in tissue context in unprecedented detail and as such have the potential to significantly change medicine. For example, the spatial localisation of immune cells with respect to a tumour is important for immune cell recognition. Both immune-excluded and immune-inflamed tumours are characterised by a high number of immune cells but respond differently to immunotherapy due to the localisation of the immune cells [dORZC20]. Spatial characterisation of the tumour immune micro-environment could thus help determine which patients are good candidates for immunotherapy.

However, data analysis methods for this kind of data are still lacking. The large amount of different cell types, in combination with their spatial information creates a complex system to analyze, with limited prior knowledge. Current solutions for the exploration of cellular patterns [SJR*17] or the comparison of cohorts based on them [YFR*12] are limited to statistical hypothesis testing, where a limited amount of combinations can be tested, posing the risk of

missing information and introducing biases. Moreover, the quality of any findings is uncertain due to the insufficient pre-processing of the data, which lacks image de-noising and often utilizes sub-optimal cell segmentation algorithms.

In collaboration with clinical researchers from Leiden University Medical Center (LUMC), we developed a pipeline for the analysis of high-dimensional tissue imaging data sets, focused on the analysis of spatial cellular patterns. Our pipeline covers the major parts of the analysis process. We worked on data normalization and segmentation to provide reliable input to the main analysis components. With ImaCytE, we designed and implemented a visual analytics framework that enables experts to identify cell types, label the segmented cells based on the measured tissue properties in an interactive and data-driven way and to explore the micro-environments the different cell types form. Based on the resulting segmented and labelled cell images, SpaCeCo enables the comparison of two distinct cohorts, based on their tissue characteristics. Throughout the process and for all resulting applications, we put special focus on the sustainability of our pipeline, by supporting standard domain practices and data types for import and export.

2. Background

From basic science to clinical practice, a wide spectrum of studies are utilizing spatial omics data to answer specific questions regarding tissue functionality. The specific characteristics of each study



Figure 1: Screenshot of *ImaCytE*, our integrated system for the analysis of Imaging Mass Cytometry data.

dictate the type of measured characteristics. We built and tested our pipeline with data originating from two modalities, Imaging Mass Cytometry [GWS*14] and the Vectra imaging system [IBA*19], that are used by our clinical partners.

Both modalities produce highly multiplexed imaging data, i.e., images where every pixel contains multiple scalar values, each representing the abundance of a specific, pre-defined protein at sub-cellular resolution. The main difference between the methods, from a data analysis perspective, is the resulting number of attributes, or dimensionality. Imaging Mass Cytometry can measure in the order of 40 different proteins simultaneously, while Vectra is currently limited to around six. However, Vectra offers much higher throughput and therefore more suited for clinical applications and the analysis of bigger cohorts. In principle, our pipeline can be utilized, as a whole or partially, for other types of spatial omics methods that produce sub-cellular resolution images.

As the acquisition of spatial omics data is a relatively new phenomenon, currently they are often included in larger studies, mainly using related non-spatial data, for qualitative analysis. In such cases only few images are analysed. With the high-throughput Vectra, larger studies centered around spatial analysis have become feasible, requiring analysis capacities for hundreds of images.

3. Contribution

In this project, we aim to support clinical researchers with diverging application and research goals with an holistic analysis pipeline for imaging omics data. More specifically, we worked on three major parts of the analysis pipeline. Clean and reliable data is paramount for proper downstream analysis. In the **Pre-processing** stage (Section 3.1), we introduced a semi-supervised method for noise removal and normalization between samples and dimensions [ISL*20], and developed a segmentation workflow supporting complex cellular morphology such as microglia [KSdH*21]. For in-depth analysis of individual images, including contained cells and the **Micro-Environment Exploration** (Section 3.2), we have developed *ImaCytE* [SvUK*19] (Figure 1), a visual analytics framework, allowing cell type identification within tissue as well as the exploration of formed neighborhoods. Finally, we worked on **Cohort Comparison** (Section 3.3) of labelled images [SIL*20] (Figure 2) for the identification of cell types or cell micro-environments that can be used as biomarkers for health and disease or deteriorating disease.

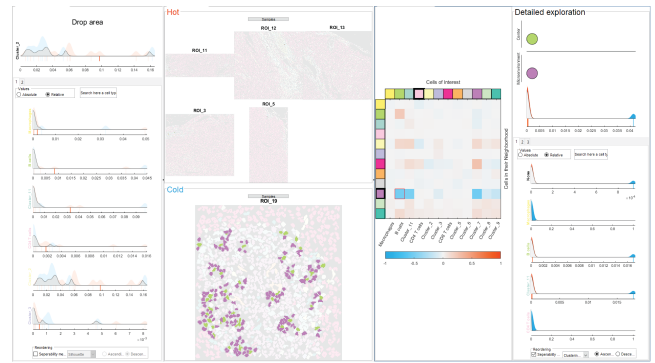


Figure 2: Screenshot of *SpaCeCo*, our system for cohort comparison based on cell type abundance and cellular micro-environments.

3.1. Pre-processing

Normalization. The protein abundance range in Imaging Mass Cytometry data differs among tissue samples from different participants, due to different tissue preservation and acquisition protocols for each sample [BSBT11]. Hence, the combination of data from multiple subjects in a clinical study requires normalization of the samples in order to produce comparable downstream results, free of batch effects. To that end, we propose a semi-supervised workflow [ISL*20] to pre-process the images acquired from Imaging Mass Cytometry for cohort analysis. After percentile based outlier removal, manual annotations of a small subset of images are used to train a classifier to define for actual signal and background pixels using Ilastik [BKK*19]. The result is binary masks for each attribute image, classifying each pixel as expressing the protein or not. We then define relative expression levels per cell as the fraction of positive pixels compared to the total segmented area of the cell. We show that this effectively removes the impact of unreliable quantitative expression values, eliminates non-biological variation in the data and detects previously untraceable cell types [ISL*20]. While normalization is important, this last step of aggregating data per cell also heavily relies on the segmentation of said cells.

Segmentation. Imaging modalities provide per-pixel measurements. To extract the cellular information from the corresponding images, we first need to segment them. For the majority of the cellular types with a circular shape, a standard semi-supervised workflow has been established [SJR*17]. However, for example microglia cells, the immune cells of the brain demand a unique approach for segmentation, due to their complex structure. Microglia cells consist of the main part of the cell (soma) and long branches (processes). To that end, we developed a novel workflow for the segmentation of complex cell shapes as prevalent in microglia cells acquired in high resolution brain imaging with the Vectra imaging modality [KSdH*21].

The identification of the whole microglia cell area, especially when many microglia cells attack a pathogen and consort around it, is a difficult process. We tackle this problem, starting with the identification of the soma, using the sum of four channels (i.e., membrane markers) where the soma shows highest intensity values. Further it overlaps with the cell nucleus, which is highlighted by DAPI nucleus marker in another image channel. We first segment the soma and then keep only the components that overlap with a

nucleus. For both steps, we utilize a level-set-based cell segmentation method [DVCE*10]. For the microglia cells close to pathogens whose soma often overlap between multiple cells, we apply watershed segmentation [Beu79] to identify the unique soma structures. A major common problem in cell segmentation is the exclusion of processes from the microglia cells, as the acquired images are 2D and the processes are detached from the soma as they connect in a different depth. We solve this problem assigning all detached processes which are identified in a close radius of each soma to the corresponding microglia cell. In a brief evaluation [KSdH*21], we achieved significantly better overlap with a manual segmentation, compared to currently used standard cell segmentation algorithms. We utilized the segmentation algorithm in a study covering the role of iron accumulating microglia cells in Alzheimer’s disease [KSdH*21].

3.2. Micro-environment Exploration

Spatial omics data provides the possibility to map identified cells to their position in tissue. This allows to investigate co-localization patterns, or micro-environments, which are hypothesized to play an important role in cell behaviour and functionality. Therefore, we developed a visual analytics framework, ImaCytE [SH19], for interactive micro-environment exploration of Imaging Mass Cytometry. ImaCytE allows quality control, cell type identification and micro-environment exploration in a single, integrated application.

Quality Control. Basically all spatial omics data requires extensive pre-processing, as for example described in Section 3.1. To verify the quality of the processed data, ImaCytE allows for interactive, visual quality control. This allows the identification of markers or proteins that did not stain well (i.e. dimensions that should be excluded from the analysis) or samples that were not normalized correctly. We achieve this by use of a small multiples view to quickly compare multiple dimensions for a single sample or the same dimension for multiple samples, respectively.

Cell Type Identification. After the selection of the samples and markers that meet the quality standards, the first main goal is the identification of the different cell types. The combination of various cell types reveals the identity of the tissue sample. The presence of cancer cells, for example, can characterize a tissue sample as tumour. The type of each cell is defined by the expression levels of the selected proteins. Hence, following previous work on non-spatial cytometry data [HPvU*16], we derive a two-dimensional embedding based on the high-dimensional protein signature of each cell. The embedding enables the visual inspection of the similarities among cells and the clustering of the visually identified groups of cells. To label the identified clusters as biologically relevant cell types, ImaCytE also enables the user to visually inspect the aggregated protein expression values of each cluster via a heatmap view and identify spatial patterns in a tissue view.

Micro-environment Exploration. Having identified the cell types, more complex questions, such as “How often immune cells attack cancer cells?” arise. ImaCytE offers the expert an overview of how frequently any two cell types spatially interact, as well as a detail exploration of existing micro-environment motifs. The overview information is encoded with a heatmap, where each column represents a cell type of interest and each row a cell type that exist

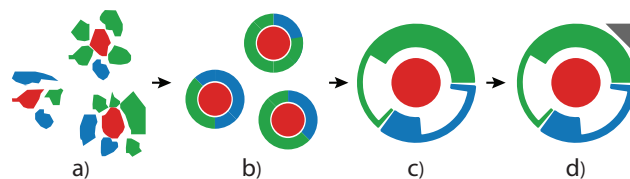


Figure 3: Glyph Design. All cells and their corresponding micro-environment (a) are abstracted to a donut chart with the cell of interest as an extra circle in the center (b). We then combine multiple instances of the same motif to a single glyph, showing the mean frequencies and variation (c). In d) we add an indicator for the significance of the motif in the top right corner.

in its micro-environment. While this overview provides a general idea of tissue functionality, it is not adequate to answer more complex questions. For example, “When immune cells attack cancer cells are they also accompanied by Helper T cells or cells from this newly identified type?”. To answer such questions, ImaCytE uses the concept of motifs to illustrate all possible cell combinations existing in the cellular micro-environments. We designed a simple glyph, inspired by the actual spatial composition, to visualize these motifs (Figure 3). The amount of motifs in each study can vary from hundreds to thousands. To facilitate their exploration, we display the corresponding glyphs in a small multiples view, offering various filtering and sorting options and linking them with the overview exploration. During micro-environment exploration it is crucial to place any spatial interaction into the tissue context. A cancer cell, detached from a tumour is an easy target for other cell types, creating unique and diverse micro-environments. However, these micro-environments are clinically not relevant as they do not exhibit the natural behavior of cancer cells. In order to identify the relevance of spatial interactions, ImaCytE enables the location of any spatial interaction in the tissue.

In ImaCytE, the glyphs are being used in a separate view, typically ordered by significance or frequency. If the images are rather sparsely populated with cells, the glyphs can also be useful for in-place visualization. We have used this in a later study [KSdH*21] on Vectra-captured brain images, where a glyph represents the cellular micro-environment around the identified pathologies (Figure 4).

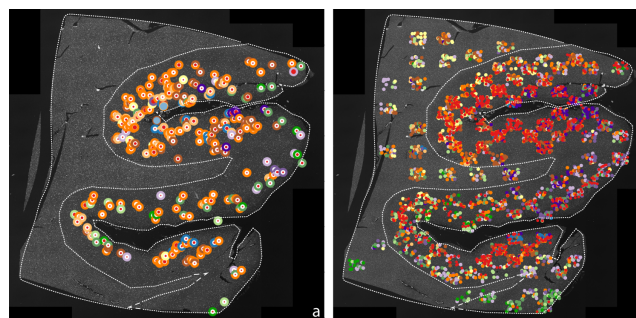


Figure 4: Glyphs in Tissue Context. The described motif glyph shown in context of the tissue (a), compared to showing the full labeled data (b). Here the glyphs are used to aggregate and present information on a large, stitched sample (14560 × 13920 pixels).

In our original publication [SvUK*19], we illustrate the efficacy of our system through a case study with an expert collaborator, analysing the interactions among cancer and immune cells from eight different tissue samples. The expert initially inspected the quality of the data and identified 20 distinct cell types. Then, she filtered from the interaction overview the motifs that include only non-proliferating cells and sorted them according to their frequency. She found out that a motif illustrating a spatial interaction among a specific immune cell type and non-proliferating cancer cells showed unexpected frequency. Hence, she hypothesized that an up-regulated protein of the immune cell type inhibits the proliferation process of cancer cells. Currently, ImaCytE is deployed with our collaborators within LUMC, who use it regularly for the analysis of Imaging Mass Cytometry data on a wide range of studies including synovial sarcoma or colorectal cancer.

3.3. Cohort Comparison

With ImaCytE, we focused on the analysis and exploration of cell types and cellular micro-environments in individual samples. To identify whether these characteristics are correlated with a specific clinical condition or tissue functionality we built an interactive visual system, SpaCeCo [SH20], for the comparison of two distinct cohorts of tissue samples. Additionally, our system enables the outlier detection within a cohort, as it can provide important clinical information, for example on subjects with different stages of a disease within the same cohort. At the same time, it allows the linking of any finding to its spatial position to verify it and place it into the general comparison context.

Cell Type Abundance. In Section 3.2, we described the importance of cell type abundance as a defining characteristic of tissue identity. In SpaCeCo, we allow the comparison of the cohorts based on the abundance of contained cell types. We accomplish this by illustrating the distribution of samples according to the abundance of each cell type within each cohort by superposing two parameterized versions of the raincloud plots, using two complementary colors (blue and orange) [APW*19](Figure 5). The density plot allows the identification of differences among the cohort distributions and together with the scatterplot clinically important outliers. The compact size of the design enables us to illustrate the plots for multiple cell types in the same view. To support large numbers of cell types, we implemented a dynamic filtering and search feature, to quickly focus the visual exploration on cell types of interest.

Cellular Micro-environments. Even though the cell type abundance is an important tissue characteristic it is not enough to characterize tissue functionality. Cohorts of samples with similar cell type abundances may exhibit different behavior, caused by the spatial micro-environments and resulting potential interactions between cells. As described in Section 3.2, the number of possible different micro-environments is extremely large and as a result the number of micro-environments that differentiate the two cohorts can be large as well. Thus, we designed a two-step approach for the comparison of the cohorts, providing first an *overview* of the main pairwise spatial interactions and afterwards *detailed* comparison of more complex micro-environments through an interactive visual query system. In the overview, we calculate the pairwise local co-occurrence of all contained cell types in each sample for both cohorts. The difference

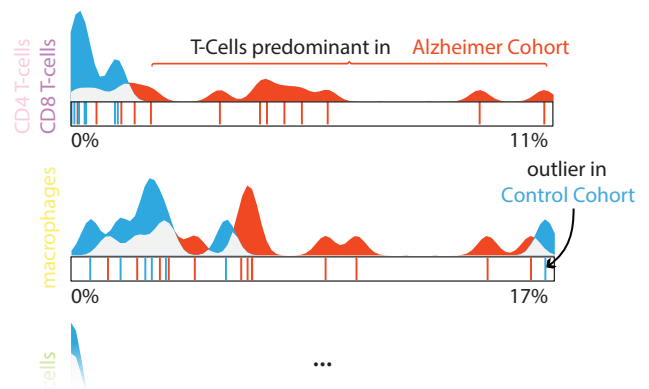


Figure 5: Examples of raincloud plots, illustrating the abundance of different cell types in the samples of two cohorts.

of these co-occurrences between the two cohorts are shown in a heatmap by an explicit encoding with a diverging colormap. Besides the diverging colormap, this view is similar to the one described in Section 3.2, providing a familiar tool for users experienced with ImaCytE. In order to compare the two cohorts based on more complex micro-environments, we implemented an interactive visual query system inspired by Polaris [SH02] that allows the expert to select any combination of cell types and explore their occurrence in each sample. The occurrence of the selected micro-environment is then shown with the presented raincloud plots.

We present several case studies in the original paper [SIL*20], illustrating the efficiency and versatility of the system. In a first published study [KSdH*21], our collaborators utilized our system for the comparison of Alzheimer patients and control individuals. Through initial exploration with SpaCeCo, our collaborators identified a positive correlation among the spatial proximity of amyloid plaques and iron-loaded microglia cells in Alzheimer patients. The results of this initial exploration were then later statistically verified.

3.4. Discussion

In this project, we have designed and implemented a complete pipeline for the analysis of spatial omics-data, which spans from the segmentation of complex cellular structures and image pre-processing over the exploration of cellular micro-environments to the detailed comparison of clinically distinct cohorts of samples.

Besides the scientific output, the project resulted in two open-source software applications [SH19, SH20] that have attracted the attention of the domain science community. Both tools are deployed with and used by our collaborators within LUMC, as well as domain researchers world-wide [Gar21]. The advent of high multiplexed images which can capture tissue functionality at sub-cellular resolution is a very recent but highly relevant development [Mar21]. Currently their analysis is still in an early stage, often limited to the identification of cell types. The use of the clinical information derived from the detailed exploration of the cellular micro-environment is still lacking, but is highly awaited [dVMKdM20, GvUI*20]. As such, we regard the presented holistic pipeline that enables clinical experts to formulate important clinical hypothesis a crucial step for the analysis of spatial omics data.

Acknowledgements. This work was partially funded through the Leiden University Data Science Research Programme and the H2020-Marie Sklodowska-Curie Action Research and Innovation Staff Exchange (RISE) Grant 644373-PRISAR.

References

- [APW*19] ALLEN M., POGGIALI D., WHITAKER K., MARSHALL T. R., KIEVIT R.: Raincloud plots: a multi-platform tool for robust data visualization. *Wellcome open research* 4 (2019). doi:10.12688/wellcomeopenres.15191.1.4
- [Beu79] BEUCHER S.: Use of watersheds in contour detection. In *Proceedings of the International Workshop on Image Processing* (1979), CCTT. 3
- [BKK*19] BERG S., KUTRA D., KROEGER T., STRAEHLE C. N., KAUSLER B. X., HAUBOLD C., SCHIEGG M., ALES J., BEIER T., RUDY M., ET AL.: Ilastik: interactive machine learning for (bio) image analysis. *Nature Methods* 16, 12 (2019), 1226–1232. doi:10.1038/s41592-019-0582-9. 2
- [BSBT11] BUCHWALOW I., SAMOILOVA V., BOECKER W., TIEMANN M.: Non-specific binding of antibodies in immunohistochemistry: fallacies and facts. *Scientific reports* 1, 1 (2011), 1–6. doi:10.1038/srep00028. 2
- [CB19] CONESA A., BECK S.: Making multi-omics data accessible to researchers. *Scientific data* 6, 1 (2019), 1–4. doi:10.1038/s41597-019-0258-4. 1
- [dORZC20] DE OLZA M. O., RODRIGO B. N., ZIMMERMANN S., COUKOS G.: Turning up the heat on non-immunoreactive tumours: opportunities for clinical development. *The Lancet Oncology* 21, 9 (2020), e419–e430. doi:10.1016/S1470-2045(20)30234-5. 1
- [DVCE*10] DZYUBACHYK O., VAN CAPPELLEN W. A., ESSERS J., NIESSEN W. J., MEIJERING E.: Advanced level-set-based cell tracking in time-lapse fluorescence microscopy. *IEEE transactions on medical imaging* 29, 3 (2010), 852–867. doi:10.1109/TMI.2009.2038693. 3
- [dVMKdM20] DE VRIES N. L., MAHFOUZ A., KONING F., DE MIRANDA N. F. C. C.: Unraveling the complexity of the cancer microenvironment with multidimensional genomic and cytometric technologies. *Frontiers in Oncology* 10 (2020), 1254. doi:10.3389/fonc.2020.01254. 4
- [Gar21] GARCIA J. O.: Data analysis at Newcastle University flow cytometry core facility. <https://www.ncl.ac.uk/fccf/research-development/data-analysis/>, 2021. Accessed: 2021-02-05. 4
- [GvUI*20] GUO N., VAN UNEN V., IJSSELSTEIJN M. E., OUBOTER L. F., VAN DER MEULEN A. E., CHUVA DE SOUSA LOPES S. M., DE MIRANDA N. F. C. C., KONING F., LI N.: A 34-marker panel for imaging mass cytometric analysis of human snap-frozen tissue. *Frontiers in Immunology* 11 (2020), 1466. doi:10.3389/fimmu.2020.01466. 4
- [GWS*14] GIESEN C., WANG H. A., SCHAPIRO D., ZIVANOVIC N., JACOBS A., HATTENDORF B., SCHÜFFLER P. J., GROLIMUND D., BUHMANN J. M., BRANDT S., VARGA Z., WILD P. J., GÜNTHER D., BODENMILLER B.: Highly multiplexed imaging of tumor tissues with subcellular resolution by mass cytometry. *Nature Methods* 11, 4 (2014), 417–422. doi:10.1038/nmeth.2869. 2
- [HPvU*16] HÖLLT T., PEZZOTTI N., VAN UNEN V., KONING F., EISEMANN E., LELIEVELDT B. P. F., VILANOVA A.: Cytosplore: Interactive immune cell phenotyping for large single-cell datasets. *Computer Graphics Forum (Proceedings of EuroVis)* 35, 3 (2016), 171–180. doi:10.1111/cgf.12893. 3
- [IBA*19] IJSSELSTEIJN M. E., BROUWER T. P., ABDULRAHMAN Z., REIDY E., RAMALHEIRO A., HEEREN A. M., VAHRMEIJER A., JORDANOVA E. S., DE MIRANDA N. F.: Cancer immunophenotyping by seven-colour multispectral imaging without tyramide signal amplification. *The Journal of Pathology: Clinical Research* 5 (2019), 3–11. doi:10.1002/cjp2.113. 2
- [ISL*20] IJSSELSTEIJN M. E., SOMARAKIS A., LELIEVELDT B. P., HÖLLT T., DE MIRANDA N. F.: Semi-automated background removal limits loss of data and normalises the images for downstream analysis of imaging mass cytometry data. *bioRxiv* (2020). doi:10.1101/2020.11.26.399717. 2
- [KSdH*21] KENKHUIS B., SOMARAKIS A., DE HAAN L., DZYUBACHYK O., IJSSELSTEIJN M. E., DE MIRANDA N. F., LELIEVELDT B. P., DIJKSTRA J., VAN ROON-MOM W. M., HÖLLT T., ET AL.: Iron loading is a prominent feature of activated microglia in alzheimer's disease patients. *Acta neuropathologica communications* 9, 1 (2021), 1–15. doi:10.1186/s40478-021-01126-5. doi:10.1186/s40478-021-01126-5. 2, 3, 4
- [Mar21] MARX V.: Method of the year: spatially resolved transcriptomics. *Nature Methods* 18, 1 (Jan 2021), 9–14. URL: <https://doi.org/10.1038/s41592-020-01033-y>, doi:10.1038/s41592-020-01033-y. 1, 4
- [SH02] STOLTE C., HANRAHAN P.: Polaris: A system for query, analysis and visualization of multi-dimensional relational databases. *IEEE Transactions on Visualization and Computer Graphics* 8 (2002), 52–65. doi:10.1109/INFVIS.2000.885086. 4
- [SH19] SOMARAKIS A., HÖLLT T.: ImaCytE open source software. doi:10.5281/zenodo.3350018. 3, 4
- [SH20] SOMARAKIS A., HÖLLT T.: SpaCeCo open source software. doi:10.5281/zenodo.3885814. 4
- [SIL*20] SOMARAKIS A., IJSSELSTEIJN M. E., LUK S. J., KENKHUIS B., DE MIRANDA N. F. C. C., LELIEVELDT B. P. F., HÖLLT T.: Visual cohort comparison for spatial single-cell omics-data. *IEEE Transactions on Visualization and Computer Graphics* (2020). doi:10.1109/tvcg.2020.3030336. 2, 4
- [SJR*17] SCHAPIRO D., JACKSON H. W., RAGHURAMAN S., FISCHER J. R., ZANOTELLI V. R., SCHULZ D., GIESEN C., CATENA R., VARGA Z., BODENMILLER B.: HistoCAT: analysis of cell phenotypes and interactions in multiplex image cytometry data. *Nature Methods* 14, 9 (2017), 873–876. doi:10.1038/nmeth.4391. 1, 2
- [SvUK*19] SOMARAKIS A., VAN UNEN V., KONING F., LELIEVELDT B. P., HÖLLT T.: ImaCytE: visual exploration of cellular microenvironments for imaging mass cytometry data. *IEEE Transactions on Visualization and Computer Graphics* (2019), 1–1. doi:10.1109/tvcg.2019.2931299. 2, 4
- [YFR*12] YUAN Y., FAILMEZGER H., RUEDA O. M., RAZA ALI H., GRÄF S., CHIN S. F., SCHWARZ R. F., CURTIS C., DUNNING M. J., BARDWELL H., JOHNSON N., DOYLE S., TURASHVILI G., PROVENZANO E., APARICIO S., CALDAS C., MARKOWETZ F.: Quantitative image analysis of cellular heterogeneity in breast tumors complements genomic profiling. *Science Translational Medicine* 4, 157 (2012), 157ra143–157ra143. doi:10.1126/scitranslmed.3004330. 1