Leiden
Computational Biology Center

AUGMENTING HEATMAPS

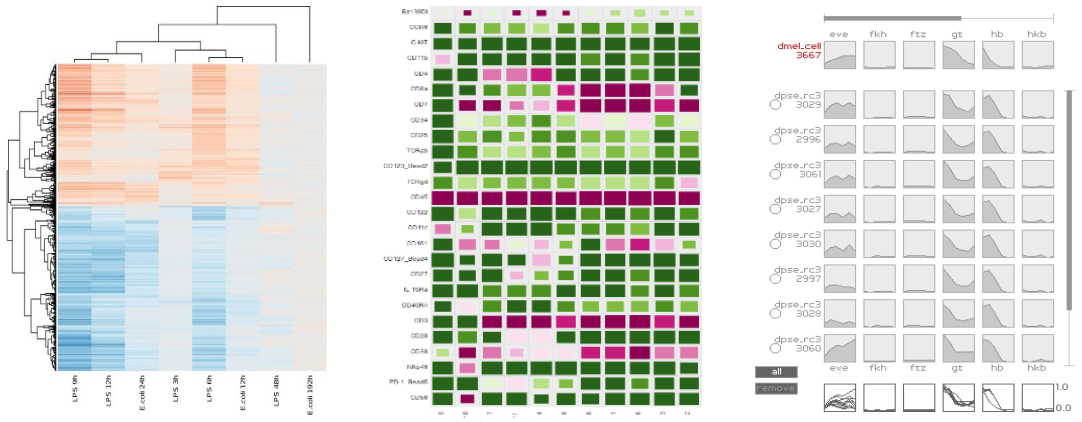
Contact: Thomas Höllt — t.hoellt@lumc.nl

Focus: Research/Practical

Duration: 3–6 Months

Requirements: Programming skills in javascript and D3.js are required. Implementation will be in the Cytosplore platform, based on C++, Qt, OpenGL, and javascript. Some interaction with C++ code might be necessary.

Heatmaps are a common way to visualize high dimensional data. In this project, the data visualized is gathered from single-cell mass spectometry (CyToF) by immune biologists. These data contain immune system signatures of thousands to millions of cells gathered from blood and tissue samples. The signatures consist of high dimensional vectors containing expressions for different antibodies which were attached to each cell. One goal of the immune biologists is to classify and group the cells into so called immune subsets, that describe the functionality of the contained cells. In the example in Figure 1b every column represents such a set of cells and every row one dimension of the high dimensional expression vector, describing these cells. Color is used to show the median value of the expression for each cluster and dimension.



(a) (b) [HPvU*16] (c) [MMDP10]

Figure 1: **Different Heatmap Visualizations.**

While heat maps can be used to visualize very large numbers of dimensions and/or data points directly (Figure 1a) we merely see global trends in this visualization (i.e. red/blue), rather than single data points.

The goal of this project is to augment rather small heat maps, such as the ones in Figure 1b and Figure 1c with additional information of the subset. For example, from a biological point of view, a cluster should contain similar cells, i.e. the expression is homogeneous. A very basic example how information on the homogeneity could be visualized can be seen in Figure 1b. Here we indicate the variation within a cluster, for each dimension of each data point by reducing the amount of paint in the box with increasing variation. As long as boxes in the heat map are large enough they could also be used to show small multiples, for example of the distribution within the cluster (Figure 1c).

Exemplary starting points for your research can be, but are not limited to Meyer et al. *MulteeSum: A Tool for Comparative Spatial and Temporal Gene Expression Data* [MMDP10] or Holzhüter et al. *Visualizing Uncertainty in Biological Expression Data* [HLS*12].

REFERENCES

[HLS*12] HOLZHÜTER C., LEX A., SCHMALSTIEG D., SCHULZ H.-J., SCHUMANN H., STREIT M.: Visualizing uncertainty in biological expression data. In *Proceedings of the SPIE Conference on Visualization and Data Analysis (VDA '12)* (2012), vol. 8294, pp. 82940O–82940O–11. doi:10.1117/12.908516.

[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.

[MMDP10] MEYER M., MUNZNER T., DEPACE A., PFISTER H.: Multeesum: A tool for comparative spatial and temporal gene expression data. *IEEE Transactions on Visualization and Computer Graphics (InfoVis '10) 16*, 6 (2010), 908–917. doi:10.1109/TVCG.2010.137.