**BMI 706 Project: Dataset and Tasks**

Nkambule, Lethukuthula

Melnikas, Max

Wan, Guihong

**1. Identify a set for your visualization project.** Your dataset should be from a biomedical domain and contain multiple entities of interest, different types of variables (quantitative,  
categorical, ordinal, relations, temporal, geospatial, genomic, etc.), and publicly  
accessible.

We are interested in spatially resolved multiplexed imaging data at single-cell resolution, generated by CyCIF[[1]](#footnote-1). Specifically, we are going to use the publicly available data from a melanoma study [1]. The data can be downloaded from [here](https://www.tissue-atlas.org/atlas-datasets/nirmal-maliga-vallius-2021/).

**2. Summarize the variables, data types** (temporal, geospatial, networks, multivariate  
matrices, etc.), and key statistics (# of elements, # of attributes, # of timepoints, etc.) of  
your data set.

Same as the paper [1], we are going to focus on a melanoma sample, which contains 1,110,585 cells with (X, Y) coordinates, 30 antibody markers, and 17 phenotypes, including keratinocytes, tumor cells, T cells, macrophages, myofibroblast, and so on. We can build networks based on the spatial location. It also has the original multiplexed image, along with the H&E image of the same tumor. We will add them into our visualization if needed.

**3. Describe what kind of information can be derived through exploratory visualization  
analysis of the data set**.

We plan to produce visualizations of results from spatial analyses of this data, with the goal that biologists can use our tool to explore the same type of new samples and visualize results for a better understanding of the tissue and preparing for manuscripts.

**4. Identify the target audience for the visualization tool that you will build.**

Biomedical researchers and cancer biologists.

**5. Develop a list of visualization tasks for the data set.**

[1] Distribution of cell phenotypes across the whole slide using bar plot.

[2] Distribution of cell phenotypes stratified by four regions (intratumor, tumor margin, stroma margin, and stroma).

[3] Clustering the cells and visualization of the top antibody markers in each cluster.

[4] Visualization of spatial neighborhoods on cell phenotypes.

**References**

[1] Nirmal, Ajit J., Zoltan Maliga, Tuulia Vallius, Brian Quattrochi, Alyce A. Chen, Connor A. Jacobson, Roxanne J. Pelletier et al. "The spatial landscape of progression and immunoediting in primary melanoma at single-cell resolution." Cancer Discovery 12, no. 6 (2022): 1518-1541.

1. https://www.cycif.org/methods/ [↑](#footnote-ref-1)