Lung Adenocarcinoma Identification via Cellular Phenotyping and Neighborhood Analysis with Imaging Mass Cytometry

Aparajith Kannapiran[1], Pingjun Chen[2], Bo Zhu[3], Lingzhi Hong[2],
Pedro Filipe Simones Da Rocha[4], Jianjun Zhang[3], Jia Wu[2]

Department of Biomedical Engineering, University of Texas at Austin[1]
Departments of Imaging Physics[2], Thoracic/Head & Neck Oncology[3], and Translational Molecular Pathology[4] , The University of Texas MD Anderson Cancer Center

Background

The understanding of the tumor environment at a cellular level facilitates better comprehension of cancer mechanisms and the development of new therapeutic strategies. The advent of imaging-mass cytometry (IMC)[1] empowers researchers to simultaneously analyze dozens of single cell parameters *in situ*, potentially yielding profound insights into cancer biology. Manual analysis is limited in clinical application and significance due to IMC data's high dimensional and heterogenous characteristics. Computational tools like machine learning and deep learning demonstrate great potential to extract knowledge from complex data. This exploratory work adopts the unsupervised clustering and neighborhood analysis techniques to explore cell phenotypes and spatial organization in IMC data, aiming to better understand lung tumor microenvironments.

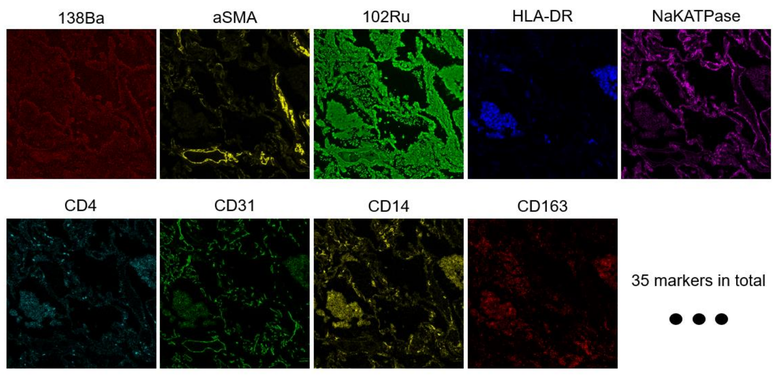

Figure 1: Raw IMC data showing various staining of proteins to show gene expression patterns across lung nodule tissue.

Objective

Due to the exploratory nature of our work, we approached an IMC dataset to better understand the genetic expression present in a tissue microenvironment. Our goal was to compare the tumor microenvironment of adenocarcinoma (ADC) tissue to regular healthy tissue, in search of critical differences cell neighborhoods.

Methods and Results

We first collect IMC data, including 35 lung nodule tissues with 297 regions of interest stained using 35 unique markers. Then, the whole-cell and nuclei are segmented using a convolutional neural network (CNN) model developed at Stanford[2]. Next, we explore the clusters of these segmented cells via their marker staining signals. We then investigate through clustering algorithms, including Phenograph, and t-SNE repeatedly with varying cell numbers[3]. The results provide us with a robust image of the cell clusters, and what features each group exemplifies the most.

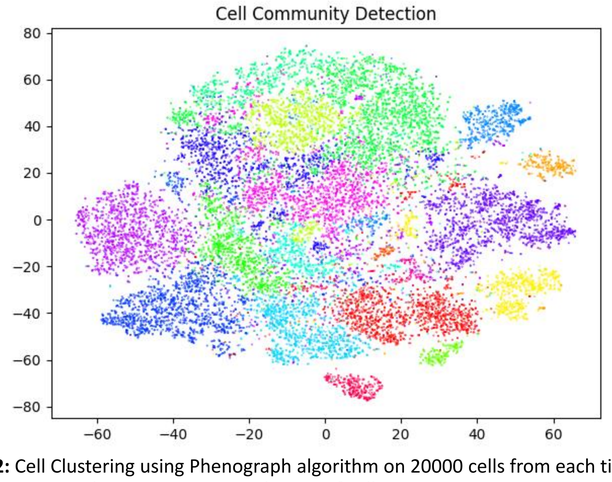

Figure 2: Cell Clustering using Phenograph algorithm on 20000 cells from each tissue sample. Unique colors represent new clusters of cells.

Methods and Results (continued)

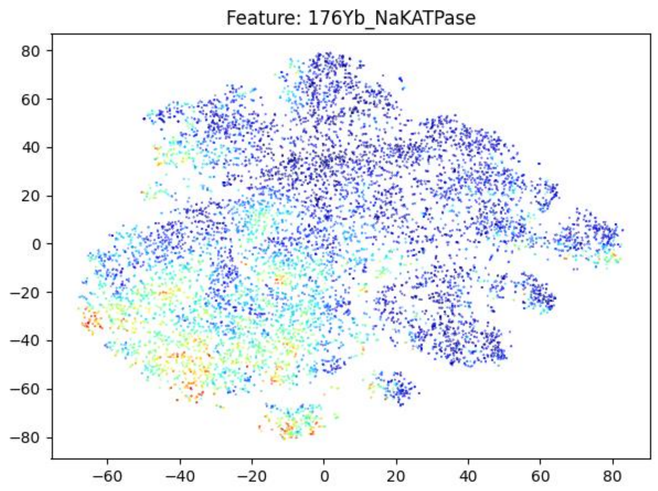

Figure 3: Example of a t-SNE results showing the spread of the NaKATPase staining across all the clusters. This visualization provides a grasp of which clusters exibit which features.

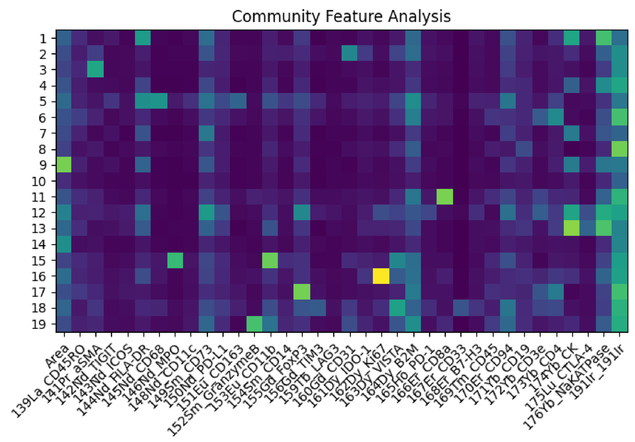

Figure 4: Heatmap demonstrating the level of expression seen through 35 unique staining methods on 19 unique clusters of lung nodule tissue. Clusters created through 20,000 sample cells per tissue. More expression is seen by green-yellow intensities.

Then, we focus on the spatial analysis of cells belonging to different clusters, to inspect cell neighborhood interactions and tumor immune borders. To see if there were any differences between ADC tissue and normal tissue, we used histoCAT's neighborhood analysis tool across the staining's of the ADC and normal tissue interpedently[4]. Due to computational limitations, we used the first region of interest (ROI) across each of the 22 samples that were either normal or ADC. This provides us with a clear visual on spatial variability between the two samples.

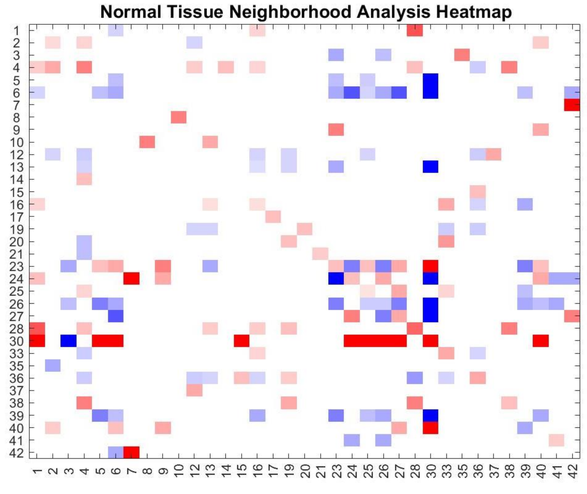

Figure 5: Heatmap of Neighborhood Analysis on normal lung nodule tissues across 42 clusters created through an independent Phenograph heatmap on histoCAT. Blue represents significantly lower levels of proximity between two clusters, while red demonstrates significantly increased levels of proximity.

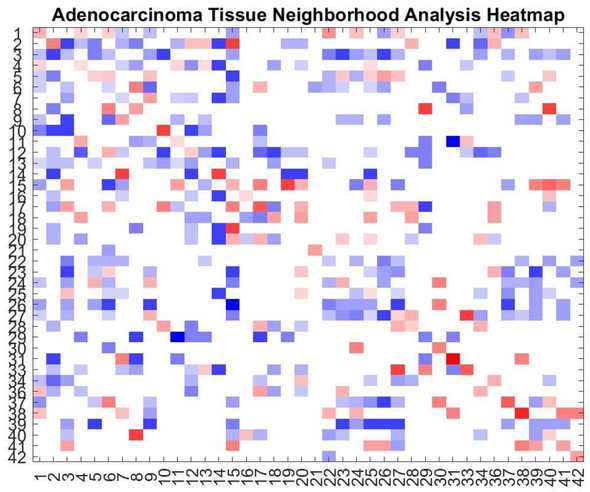

Figure 6: Heatmap of Neighborhood Analysis on adenocarcinoma lung nodule tissues across 42 clusters created through an independent Phenograph heatmap on histoCAT. Colors represent the same proximities as Figure 3.

Discussion

The clustering stage has been completed (Figure 2) with various key features already present. First, when clustering with 10,000 cells, there appears to be 16 groups, but as this number increases to 20,000 or above, we see the clusters plateau at 19. These 19 individual clusters share some unique features that can classify their cell type and behavior more accurately. The categorization of the features in the clusters allowed us to create a heat map (Figure 4), and mark mean trait intensities in each cluster. One important thing to consider is that at the current stage we are unable to note what type of cell each cluster is with great accuracy; we can only make predictions based on the gene expressions visualized through the staining's. Furthermore, the change in the number of clusters as we increase the number of cells indicates that some cells might be quite similar to one another.

Next, we hypothesized that the tumor microenvironment should be different form that of normal tissue. Using qualitative analysis on both the heatmaps produced from the neighborhood analysis we can there are some clear differences. First, based on the ADC heatmap (Figure 6) we can there is dramatic increase in significant interactions, many of these indicating that cells clusters are not spaced near each other. Additionally, we can see one cluster of calls that interacts highly with other clusters seen in the normal tissue (Figure 5) has shifted it proximity in relation to other clusters. There is a stark increase in the number of cells that do not share any proximity to one another.

Conclusion

To conclude, by utilizing unsupervised clustering and neighborhood analysis techniques on IMC data, we have been able to cluster cells into unique categories, providing a considerable amount of insight into what may be present in the tissue microenvironment and the differences between lung adenocarcinoma and normal tissue. This project is ongoing as we plan to continue deeper levels of analysis to elucidate patterns of cellular interactions among the lung tumor microenvironments.

References

1. Giesen, Charlotte, et al. "Highly Multiplexed Imaging of Tumor Tissues with Subcellular Resolution by Mass Cytometry." *Nature Methods*, vol. 11, no. 4, 2014, pp. 417–422., doi:10.1038/nmeth.2869.
2. Greenwald, Noah F., et al. "Whole-Cell Segmentation of Tissue Images with Human-Level Performance Using Large-Scale Data Annotation and Deep Learning." 2021, doi:10.1101/2021.03.01.431313.
3. Levine, Jacob H., et al. "Data-Driven Phenotypic Dissection of Aml Reveals Progenitor-like Cells That Correlate with Prognosis." *Cell*, vol. 162, no. 1, 2015, pp. 184–197., doi:10.1016/j.cell.2015.05.047.
4. Schapiro, Denis, et al. "HistoCAT: Analysis of Cell Phenotypes and Interactions in Multiplex Image Cytometry Data." *Nature Methods*, vol. 14, no. 9, 2017, pp. 873–876., doi:10.1038/nmeth.4391.