**BIOS 785 Project Proposal**

**Team members:** Seowoo Kim, Taeim Kwon, Arthi Hariharan

**Biological questions of interest:**

The tumor microenvironment (TME) consists of immune cells, blood vessels, stromal cells and ECM surrounding the cancer cells [1]. It is essential to understand the tumor heterogeneity across breast cancer subtypes to determine treatment response. CXCR4 is chemokine receptors involved in T- Cell and B-Cell antigen recognition and is known to shape the tumor environment. CXCR4 plays a diverse role in tumor progression including metastasis and angiogenesis [2]. We aim to identify the differential expression of CXCR4 and VEGF across breast cancer subtypes to determine if CXCR4 interaction with Immune cells varies across tumor subtypes. In addition, Papanicolaou et al. analyzed single-cell RNA seq data of human breast cancer atlas and mice mammary tumor, demonstrating the result of *Col12a1* expression captured in cancer-associated fibroblasts (CAFs) is consistent for both the datasets [3]. Furthering the analysis, we aim to find more sets of gene markers and cell types related to collagen, one of the major compositions of the ECM, across different species and see concordance between human breast tumors and mice mammary tumors in terms of their microenvironment. Finally, the interaction between each gene could have another effect. For example, each of them could not affect cancer but they can inhibit expression when both exist. Our purpose is to find out the difference between the situation when we do one-way analysis and when we do two-way analysis. One-way analysis is analyzing each gene and two-way analysis is analyzing the pair of genes.

**Datasets to be adopted:**

| **Title** | **GSE of Dataset** | **PMID of Paper** |
|---|---|---|
| A single-cell and spatially resolved atlas of human breast cancers [4] | 176078 | 34493872 |
| Single cell transcriptomics reveals involution mimicry during the specification of the basal breast cancer subtype [3] | 158677 | 35933466 |

**Planned analysis framework:**

The scRNA-seq data of human breast cancer atlas [4] and mice mammary tumor [3] will be loaded and processed for them to be created as a Seurat object separately. Each Seurat object will undergo quality control, normalization, feature selection, and dimensionality reduction to reproduce a cell-type clustering, using a UMAP plot. The markers expressed in Immune cells will be identified using FindMarkers function. The differential expression of CXCR4 and VEGF will be analyzed across the immune cells. Recreated Seurat objects of human breast cancer atlas and mice mammary merged to one Seurat object for Canonical Correlation Analysis (CCA) developed in Seurat v4.0. Anchors will be identified from the list of the two Seurat objects after splitting the Seurat object by species for data integration. A UMAP plot will show if the cells cluster by species. Then the conserved cell type markers will be found with differential gene expression testing. Those associated with the structure of the ECM will be visualized across a UMAP plot, and for comparison of the results across species, feature plots split by species will be made after cell type annotation based on the gene markers. Also, we can use heatmap to identify correlation between gene-gene interaction and cancer. We can calculate average gene expression within a cell type and plot the average expression between immune cells and CAFs to check their correlation.

**Initial task assignment among group members:**

| | |
|---|---|
| **Seowoo Kim** | Initially processing the dataset such as Quality control to apply to Seurat directly. |
| **Taeim Kwon** | Analysis of the two datasets based on our biological questions of interest: differential expression testing, CCA alignment, and plotting interaction effects |
| **Arthi Hariharan** | Reproduce the analysis conducted in the original paper. |

**BIOS 785 Project Proposal**

**Team members:** Seowoo Kim, Taeim Kwon, Arthi Hariharan

**References:**

1. Anderson, N. M., & Simon, M. C. (2020). The tumor microenvironment. *Current Biology : CB*, *30*(16), R921–R925. https://doi.org/10.1016/j.cub.2020.06.081

2. Bruno, A., Pagani, A., Pulze, L., Albini, A., Dallaglio, K., Noonan, D. M., & Mortara, L. (2014). Orchestration of angiogenesis by immune cells. *Frontiers in Oncology*, *4*, 131. https://doi.org/10.3389/fonc.2014.00131

3. Papanicolaou, M., Parker, A. L., Yam, M., Filipe, E. C., Wu, S. Z., Chitty, J. L., Wyllie, K., Tran, E., Mok, E., Nadalini, A., Skhinas, J. N., Lucas, M. C., Herrmann, D., Nobis, M., Pereira, B. A., Law, A. M. K., Castillo, L., Murphy, K. J., Zaratzian, A., … Cox, T. R. (2022). Temporal profiling of the breast tumour microenvironment reveals collagen XII as a driver of metastasis. *Nature Communications*, *13*(1), 4587. https://doi.org/10.1038/s41467-022-32255-7

4. Wu, S. Z., Al-Eryani, G., Roden, D. L., Junankar, S., Harvey, K., Andersson, A., Thennavan, A., Wang, C., Torpy, J. R., Bartonicek, N., Wang, T., Larsson, L., Kaczorowski, D., Weisenfeld, N. I., Uytingco, C. R., Chew, J. G., Bent, Z. W., Chan, C.-L., Gnanasambandapillai, V., … Swarbrick, A. (2021). A single-cell and spatially resolved atlas of human breast cancers. *Nature Genetics*, *53*(9), 1334–1347. https://doi.org/10.1038/s41588-021-00911-1