Specific Aims

Cancer heterogeneity within solid tumors results in a spectrum of patient outcomes and emphasizes the need for personalized treatment specific to an individual's cancer. For cancers with traditionally poor outcomes such as glioblastoma multiforme, with a median survival time of 15-16 months, individual treatment and understanding how the tumor transcriptome and microenvironment plays an important role in developing an effective treatment. One method of developing personalized treatment plans for such patients is to examine immunological data and the tumor microenvironment consisting of normal cells, tumor cells, macrophages, T-cells, B-cells, and more. Previous research has demonstrated that in the cancer immunological landscape is highly predictive of patient outcome, yet this data is masked in bulk RNA-seq. Currently there is the need to unpack this immunological data and identify tumor heterogeneity by immune cell type from bulk RNA-seq data, thus allowing further analysis across cancer archetypes to better predict and improve patient outcomes.

By better understanding differentially expressed genes in clusters from solid tumor bulk RNAseq data, we aim to analyze how immune cell types in a tumor microenvironment can predict patient outcome and the relationship between immune cell composition and survival time. This will involve applying deconvolution methods on quantified transcripts (STAR and HTseq-count and Kallisto) from solid tumor bulk RNA sequencing data. The retrieved estimated cell type abundances will then be clustered using two separate methods, k-means and Louvain clustering. For each of the generated clusters, we will perform differential gene expression using DESeq2 and Sleuth to compare gene regulation as well as validate gene ontology terms that match immune cell type enrichment. Previous studies have shown that characterizing immune cells based on density, type, and location can be used to predict patient survival and prognosis. By applying these pipelines we hope to not only examine the immunological landscape of tumors, but to also compare pipelines by combining the previously mentioned methods.

Aim 1: Examine the immunological landscape of solid tumor bulk RNAseq. Transcript quantification will undergo deconvolution to achieve immune cell composition, which will then undergo clustering. Differential gene expression will then be applied to these clusters to retrieve gene expression profiles across various cell types.

Aim 2: Compare various tools in incorporated pipelines to achieve these results. Clustering methods will include k-means clustering and Louvain clustering, while DESeq2 and Sleuth will be used for differential gene expression analysis. The results from these methods will be evaluated based on accuracy.

The proposed project will improve our understanding of the relationships between immune cell composition and survival time for patients with cancer. Not only will the identified immune cell types be characterized by their differential gene expression and clustering, but the pipelines used for this analysis will be evaluated. These outcomes can inevitably be applied to research within the scientific community with the intention for physicians to identify and develop personal treatment plans for cancer patients to improve clinical outcomes.