Single-Nucleus RNA Sequencing Case Study

Pang-Kuo Lo

Title: Deciphering Tumor-Associated Macrophage Heterogeneity Through Single Nucleus RNA Sequencing Analysis

Study Goal/Aim: This study aims to characterize the molecular landscape of tumor-associated macrophages (TAMs) compared to normal macrophages in adjacent tissues using single nucleus RNA sequencing (snRNA-seq). By identifying differentially expressed genes and pathways in TAMs, we seek to enhance our understanding of their role in tumor progression and identify potential therapeutic targets.

Introduction: Tumor microenvironments are complex ecosystems where various cell types, including immune cells like macrophages, interact dynamically with cancer cells. TAMs, in particular, exhibit phenotypic plasticity and can exert either pro- or anti-tumor functions depending on their activation state. Understanding the molecular mechanisms underlying TAM heterogeneity is crucial for developing targeted therapies. Single nucleus RNA sequencing (snRNA-seq) offers a high-resolution approach to dissect the transcriptomic profiles of individual cells within the tumor microenvironment.

Study Plan:

1. Project Overview:

- This project aims to unravel the molecular landscape of tumor-associated macrophages (TAMs) using single nucleus RNA sequencing (snRNA-seq) data. By employing advanced bioinformatics methods, we seek to identify differentially expressed genes and pathways in TAMs compared to normal macrophages, thereby enhancing our understanding of their role in cancer progression and identifying potential therapeutic targets.

2. Data Acquisition and Preprocessing:

- Describe steps for preprocessing snRNA-seq data, including data normalization and quality control.

3. Exploratory Data Analysis (EDA):

- Analyze cell cluster numbers based on health/disease types and sample types using bar graphs and heatmaps.

4. Dimensionality Reduction and Clustering:

- Utilize PCA, tSNE, and UMAP algorithms for dimensionality reduction and cell clustering to identify distinct cell populations, including TAMs.

5. Characterization of TAM Subtypes:

- Identify marker genes associated with different TAM subtypes using differential expression analysis and functional enrichment analysis.

6. Visualization and Interpretation:

- Visualize the spatial distribution of TAM subtypes within the tumor microenvironment using UMAP plots.

7. Comparison with Normal Macrophages:

- Compare the transcriptomic profiles of TAMs with those of normal macrophages in adjacent tissues to identify differentially expressed genes and pathways.

8. Functional Annotation and Pathway Analysis:

- Perform gene set enrichment analysis to elucidate the biological processes and pathways enriched in TAMs compared to normal macrophages.

9. Therapeutic Target Identification:

- Identify potential therapeutic targets among the differentially expressed genes in TAMs, considering their role in tumor progression and immune modulation.

10. Integration with Clinical Data:

- Integrate transcriptomic data with clinical metadata (e.g., patient survival, treatment response) to correlate TAM heterogeneity with clinical outcomes.

11. Interpretation and Conclusion:

- Summarize key findings, implications for cancer biology, and future directions for research.

12. Code Availability:

- Provide access to the R script and associated documentation on GitHub for reproducibility and transparency.

<u>snRNA-Seq Data Analysis Workflow:</u>

1. Engineering Metadata:

- Define relevant information for each sample (e.g., cell type, health status)

2. Data Import:

- Import barcode information (barcodes.tsv)
- Import gene information (genes.tsv)
- Import expression matrix (matrix.mtx) into R environment

3. Create Seurat Object:

- Combine imported data and metadata into a Seurat object

4. Quality Control (QC):

- Analyze number of RNA features (genes)
- Examine expression counts for individual genes
- Assess percentages of mitochondrial RNA per nucleus

5. Data Pre-processing:

- Filter out cells with abnormal RNA counts or mitochondrial RNA percentages

6. Data Normalization:

- Normalize expression data to account for technical variations

7. Find Variable Features:

- Identify genes with high variability across cells

8. Data Scaling:

- Scale data for downstream analysis (e.g., PCA)

9. Principal Component Analysis (PCA):

- Reduce dimensionality of data for visualization

10. Elbow Plot Analysis:

- Determine optimal number of principal components (PCs) to retain

11. Cell Clustering:

- Group cells with similar gene expression profiles

12. Engineering Cell Cluster Labels:

- Assign meaningful names to identified cell clusters

13 & 14. Visualization - UMAP & t-SNE:

- Create UMAP plots to visualize cell clusters in different sample types and health/disease conditions
- Create t-SNE plots for an alternative visualization of cell clusters

15. Exploratory Data Analysis:

- Analyze cell numbers within each cluster across health/disease groups using bar graphs and heatmaps

16. Macrophage Identification:

- Use UMAP plot and Violin plots to identify cell clusters representing macrophages based on M1, M2, and tumor-associated macrophage marker genes

17. Subclassification:

- Subdivide cell clusters into healthy and disease subgroups

18. Differentially Expressed Genes (DEG) Analysis - Macrophages:

- Identify genes with significantly different expression between healthy and disease macrophage subgroups

19-21. Heatmap Analysis:

- ***Create heatmaps with hierarchical clustering to visualize:***
    - All differentially expressed genes (DEG)
    - Differentially expressed lncRNAs (long non-coding RNAs)
    - Differentially expressed protein-coding genes

22. Gene Set Enrichment Analysis:

- Identify biological pathways enriched in differentially expressed genes using ClusterProfiler

23 & 24. Visualization of Enrichment Analysis:

- ***Visualize enriched gene sets using:***
    - Dotplots
    - Ridgeplots

25. Visualize DEGs on UMAP:

- Overlay selected differentially expressed genes onto UMAP plots of tumor-associated macrophage clusters

26. Macrophage Marker Gene Expression:

- Use DoHeatmap to visualize expression of macrophage marker genes in healthy and tumor-associated macrophages at the single-cell level

27. DotPlot Analysis of DEGs:

- Compare differential gene expression patterns across different macrophage cell clusters using DotPlot

28. Phylogenetic Tree Analysis:

- Construct a phylogenetic tree to visualize relationships between different cell clusters

Study Perspective: This study adopts a comprehensive bioinformatics approach to unravel the complexity of TAM heterogeneity in the tumor microenvironment. By integrating advanced data analysis techniques with biological interpretation, we aim to shed light on the molecular mechanisms driving TAM function and identify novel therapeutic avenues for cancer treatment.

Conclusion: This study plan outlines a structured approach to investigate TAM heterogeneity using snRNA-seq data, with the ultimate goal of informing precision medicine strategies for cancer therapy. By elucidating the transcriptomic profiles of TAMs, we aim to pave the way for targeted interventions that harness the immune system's potential to combat cancer.