Project Proposal

yt2911 Yuchen Teng hc3512 Haoyan Chen yl5757 Yuhan Liu

pw2629 Peiyu Wang yp2726 Yueyan Pang

Title:

Dynamic Characteristics of MAIT Cells for Predicting and Understanding Anti-PD-1 Therapy Response

Background and Goals:

Anti-PD-1 immune checkpoint inhibitors have become an effective therapy for metastatic melanoma, yet their efficacy varies significantly among patients. Mucosal-associated invariant T (MAIT) cells, a unique subset of T cells, play a pivotal role in immune responses, and their functional states may directly reflect patients' responsiveness to immunotherapy. Previous studies have demonstrated that MAIT cells exhibit higher activation and migration capabilities (e.g., CXCR4 expression) in responders.

This study aims to analyze **GSE166181** (standardized gene expression matrices and meta data) using advanced machine learning and trajectory analysis methods to identify critical features related to anti-PD-1 therapy response and provide predictive and biological insights.

Methods and Advantages:

1.Single-cell Data Quality Control and Clustering Analysis: perform quality control, dimensionality reduction, and clustering on scRNA-seq data to identify MAIT cell subpopulations and their distribution across different treatment states (T0, T1, T2).

2. Dynamic Trajectory Analysis: perform **pseudotime analysis** to reveal state transitions of MAIT cells across treatment cycles and compare pseudotime differences between responders and non-responders. **Scanpy** and **Scvelo** can offer a comprehensive view from static pseudotime inference to dynamic RNA velocity analysis. These tools allow for the characterization of MAIT cell state transitions across multiple time points (T0, T1, T2).

3. **Differential Expression Analysis:** perform differential expression analysis by **DESeq2** to compare the gene expression differences between responders and non-responders at different states. Perform functional enrichment analysis **(GO/KEGG)** to elucidate the roles of these genes in immune-related pathways.

4. Classification Model Development: utilize machine learning models **(SVM, XGBoost, Random Forest)** to select important genes based on feature weights. Build and evaluate models to predict therapy response, leveraging the strengths of each algorithm. For example, SVM is effective with high-dimensional data, Random Forest provides robust feature importance scores, and XGBoost handles complex interactions.