**Trajectory Analysis of Mouse scRNA-seq Data from PRJNA626450**

***Prepared by Tao Sun* *Date: April 17, 2025***

**Project Title:** Single-Cell Trajectory Analysis of Mouse Lung scRNA-seq Data

**Dataset Used:** PRJNA626450 (C1 96-well Open App IFC platform, Fluidigm)

**Analysis Goal:** To investigate cell state transitions across pseudotime and identify genes with dynamic expression patterns during the process using trajectory inference methods.

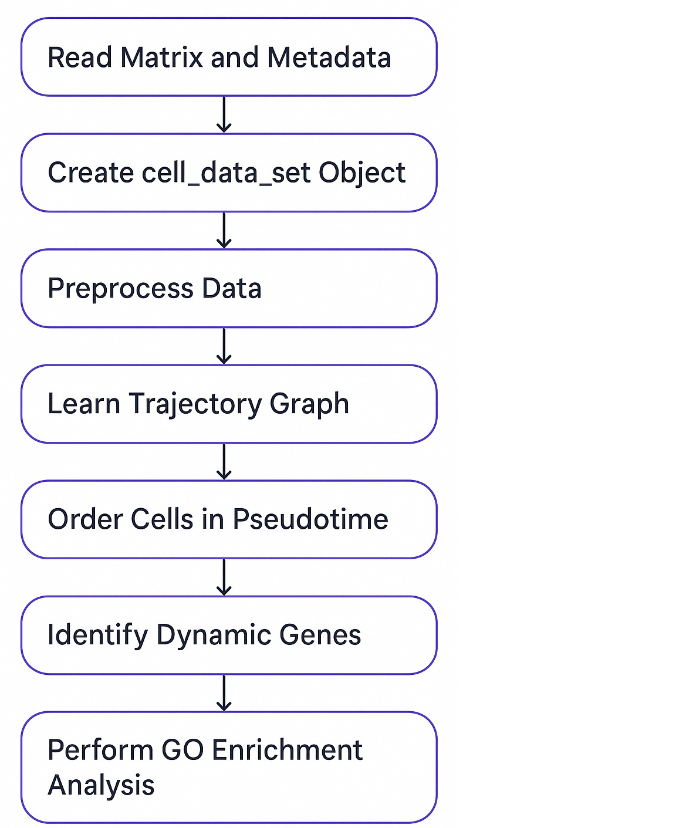
**Overview**

This report documents the trajectory analysis performed on single-cell RNA sequencing (scRNA-seq) data from mouse lung tissue, obtained from the NCBI BioProject [PRJNA626450](https://www.ncbi.nlm.nih.gov/bioproject/626450). The primary objective of this analysis was to infer cell developmental trajectories, identify genes changing along pseudotime, and interpret their biological relevance using functional enrichment analysis.

The analysis was conducted using the Monocle 3 package in R.

**1. Data Acquisition and Processing**

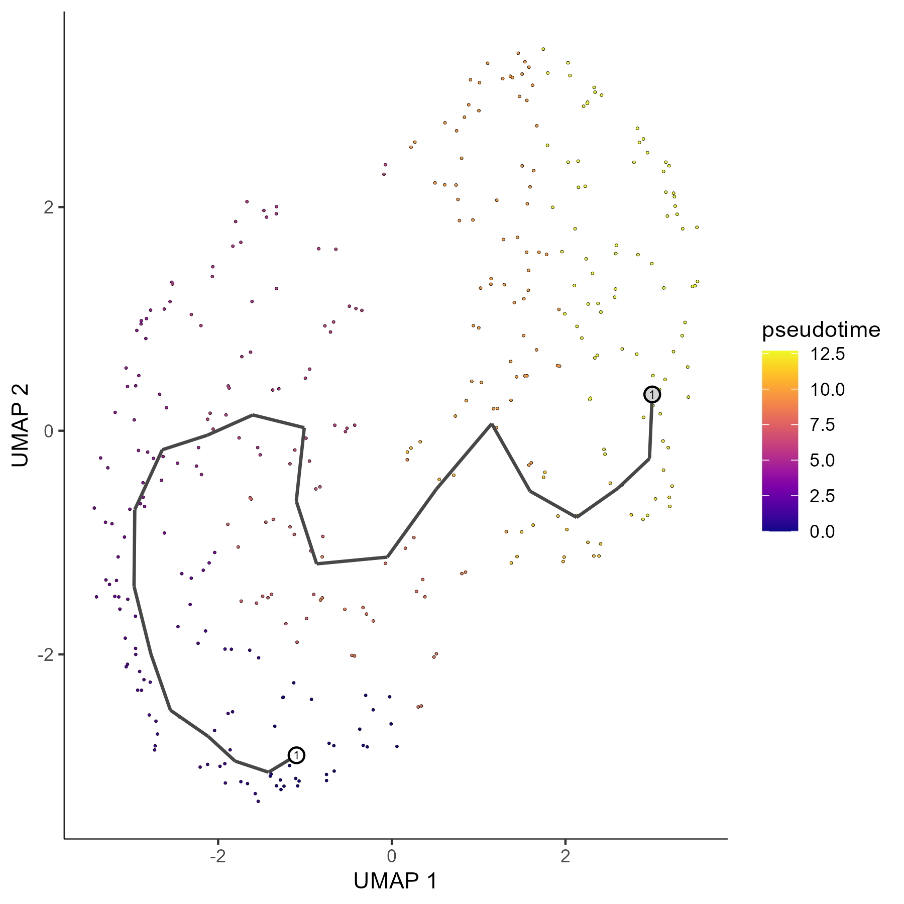
* Raw FASTQ files corresponding to 382 SRR runs from BioProject PRJNA626450 were downloaded via SRA Toolkit.
* Reads were aligned to the mouse genome (GRCm39) using HISAT2.
* Resulting SAM files were converted to BAM, sorted, and indexed using Samtools.
* Gene-level counts were generated with featureCounts using the Ensembl GTF annotation (Release 110).



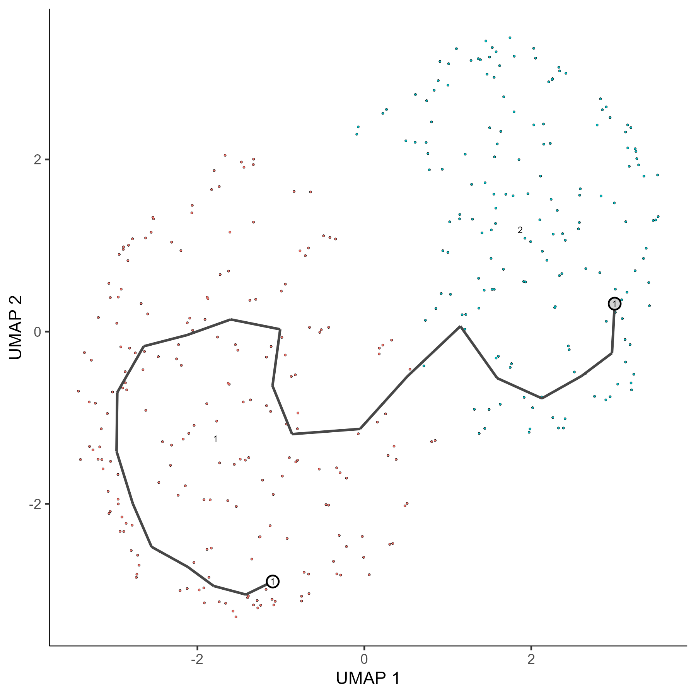
**Figure 1:** workflow figure illustrating the overall pipeline.

**2. Pseudotime Analysis Using Monocle3**

* Count matrix and metadata from SraRunTable.csv were loaded.
* Monocle3 object (cds) was created and preprocessed (50 PCs used).
* Dimensionality reduction was performed with UMAP.
* Clustering was done using cluster\_cells() with resolution = 1e-2, resulting in two clusters:
  + Cluster 1: 211 cells
  + Cluster 2: 171 cells
* Trajectory graph was learned with learn\_graph() and cells were ordered using manually selected root nodes via interactive Shiny GUI.



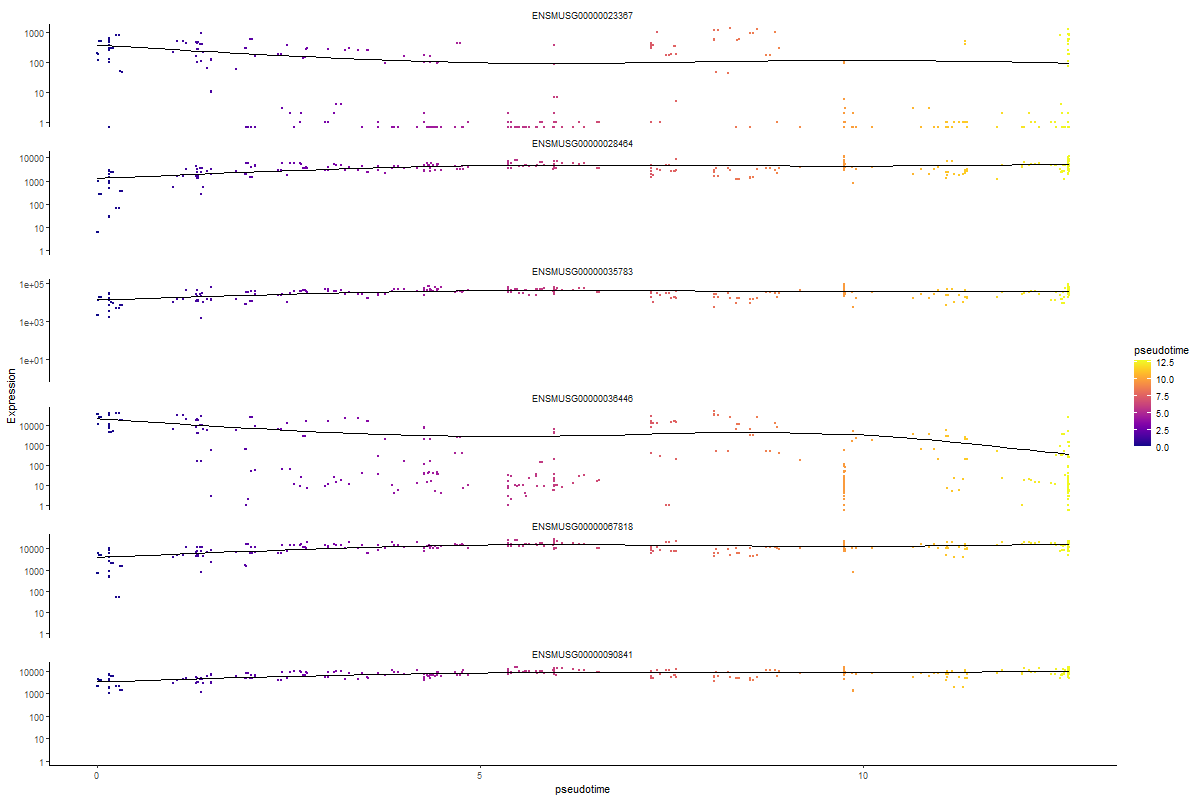
**Figure 2:** UMAP plot with pseudotime



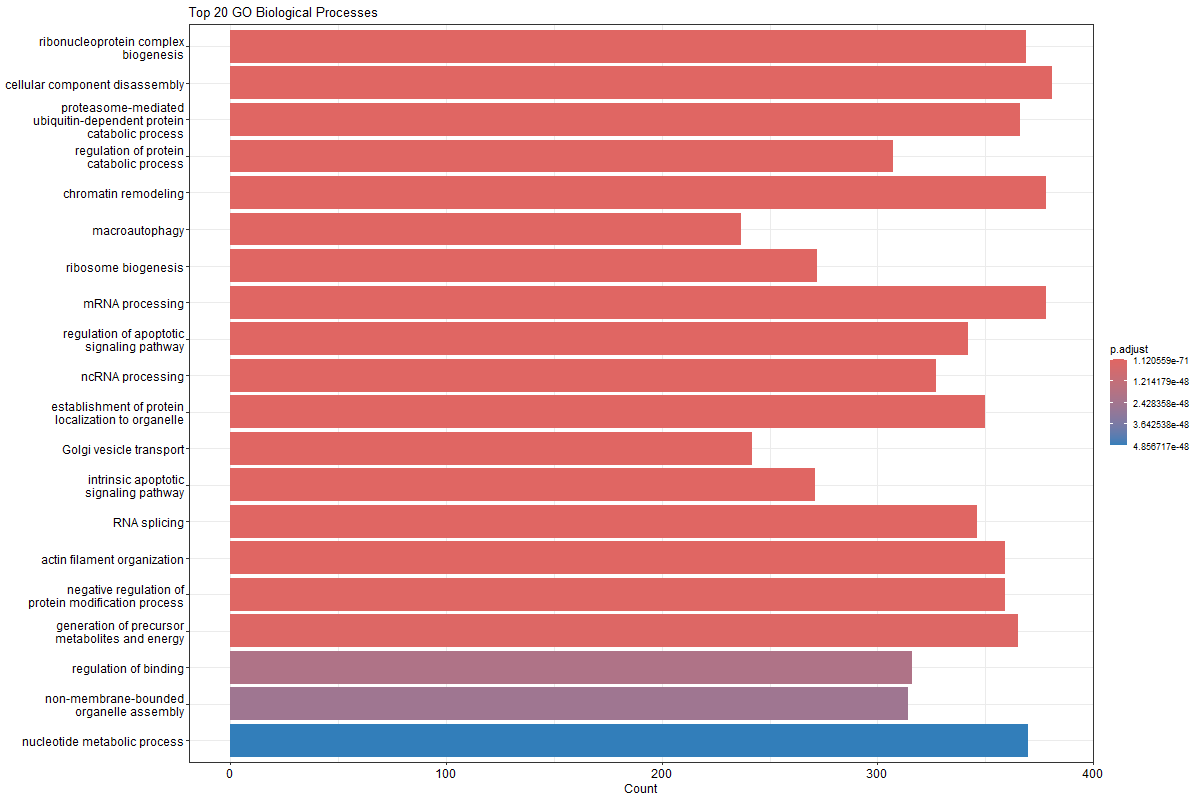
**Figure 3:** UMAP plot by cluster with the resolution to be 1e-2.

**3. Identification of Dynamic Genes**

* Moran's I test (graph\_test) identified genes significantly associated with pseudotime.
* 18,749 genes were significant with FDR q < 0.05.
* Top 6 and top 50 genes were visualized.



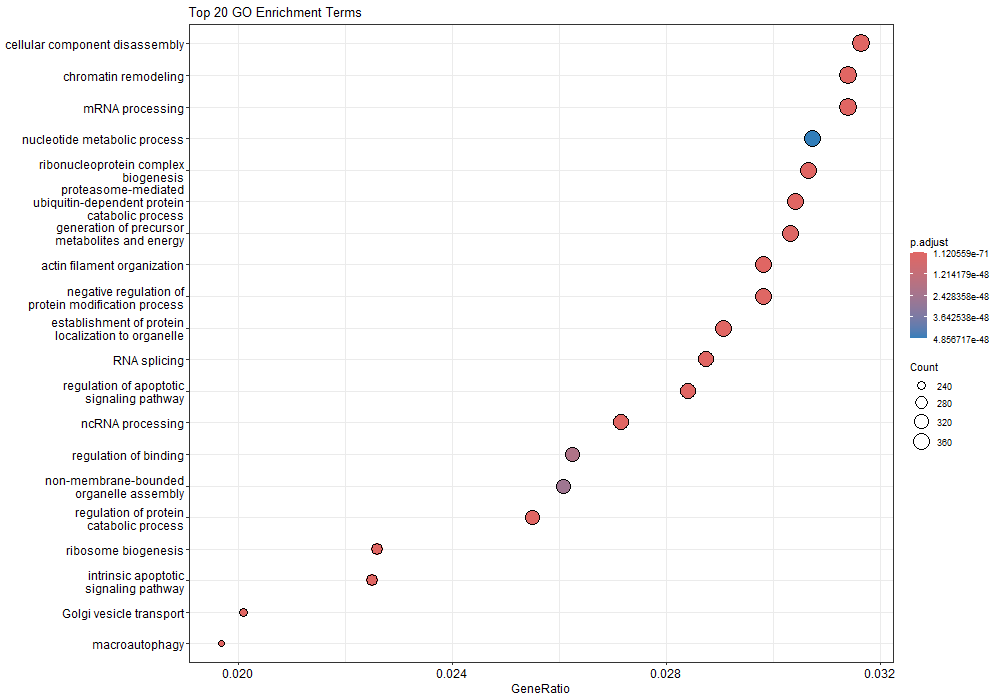
**Figure 4:** Line plot showing gene expression of top 6 genes along pseudotime



**Figure 5:** Heatmap of top 50 genes ordered by pseudotime

**4. GO Enrichment Analysis**

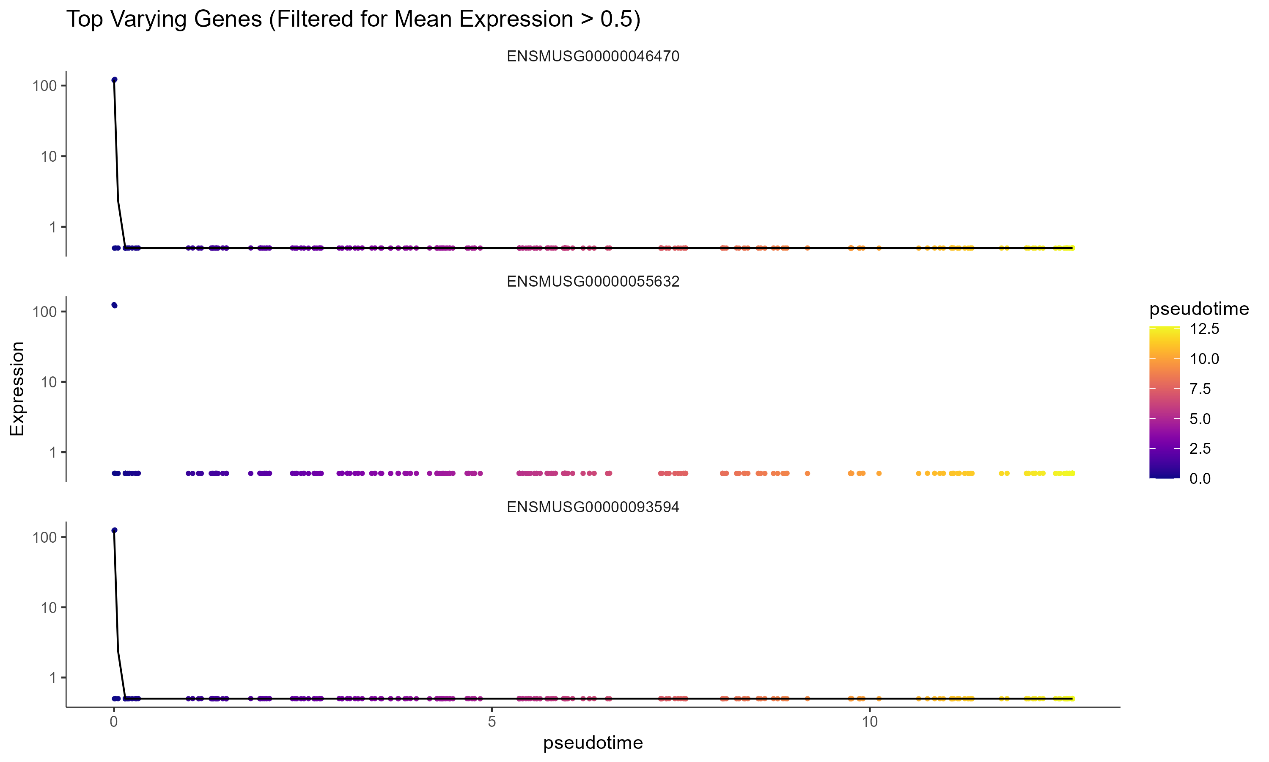
* Ensembl gene IDs from significant pseudotime-related genes (q < 0.01) were converted to Entrez IDs using bitr().
* 13,335 out of 16,847 genes were successfully mapped.
* GO enrichment analysis was conducted using enrichGO() with the org.Mm.eg.db database.



**Figure 6:** Dot plot of top 20 enriched GO terms

**5. Summary of Results**

* The analysis successfully inferred a developmental trajectory and clustered cells into two groups.
* Key genes exhibiting dynamic expression changes across pseudotime were identified.
* Functional enrichment analysis revealed biological processes relevant to lung tissue and immune responses.



**Figure 7:** Filter genes significantly associated with pseudotime

**6. Files Saved**

* CSV of pseudotime DEGs: deg\_pseudotime\_filtered.csv
* CSV of GO enrichment results: go\_results\_all\_filtered.csv
* Summary table: go\_top\_summary.csv
* Figures: All saved in E:\UBC\_wang\_qn2 (named accordingly)

**7. Future Directions**

* Perform module-level analysis to group genes by co-expression patterns along pseudotime.
* Integrate spatial transcriptomics for spatial-aware pseudotime inference.
* Apply downstream gene regulatory network (GRN) reconstruction methods.