Bulk RNA-Seq Analysis Report – ATF3 Knockout in SMCs

Candidate: Tao Sun

Date: April 18, 2025

Task: Question 1 – Bulk RNA-Seq Differential Expression and GO Analysis

GitHub Link: https://github.com/Samuel700712/Bulk\_RNAseq\_ATF3\_KO\_Report

# Objective

To identify differentially expressed genes (DEGs) and perform gene ontology enrichment analysis using bulk RNA-seq data from ATF3 knockout smooth muscle cells (SMCs).

# Dataset

NCBI BioProject: PRJNA716327

Comparison Groups: (SAMN18442667 AND SAMN18442665) vs. (SAMN18442666 AND SAMN18442664)

# Methods and Workflow

A diagram of a process

AI-generated content may be incorrect.

*Figure 1: Analysis workflow*

The overall workflow is summarized as follows:

1. Data Download and Preprocessing  
 - Used `fasterq-dump` and `fastp` in Ubuntu terminal  
 - Trimmed and filtered paired-end FASTQ files

2. Read Alignment  
 - HISAT2 was used to align reads to the mm10 genome  
 - Generated SAM/BAM files and sorted/indexed using SAMtools

3. Gene-Level Quantification  
 - Used `featureCounts` to generate count matrix

4. Differential Expression Analysis  
 - Conducted in R using DESeq2  
 - Shrinkage of log2 fold change with `lfcShrink()`

5. Exploratory Data Visualization  
 - PCA plot, volcano plot, MA plot, and heatmap generated

6. Functional Enrichment  
 - GO enrichment (Biological Process) using clusterProfiler  
 - GSEA-style analysis for KEGG pathways

# Figures

A graph with numbers on it

AI-generated content may be incorrect.

*Figure 2. PCA plot of all samples*

A graph with numbers and a graph

AI-generated content may be incorrect.

A graph with numbers and symbols

AI-generated content may be incorrect.

*Figure 3. Volcano plots highlighting DEGs under different limitations*

A chart with different colors and numbers

AI-generated content may be incorrect.

*Figure 4. Heatmap of top 30 variable genes*

A graph of different colored squares

AI-generated content may be incorrect.

*Figure 5. GO Barplot (Top 10 Terms)*

A graph with text on it

AI-generated content may be incorrect.

*Figure 6. GO Dotplot (Top 10 Terms)*

A graph with a dot in the center

AI-generated content may be incorrect.

*Figure 7. GSEA Dotplot for KEGG Pathways*

# Results Summary

- Number of DEGs (padj < 0.1 & |log2FC| > 1): 0  
 - Number of DEGs (p < 0.05 & |log2FC| > 1): 8  
 - GO terms enriched: 6 (p.adjust < 0.1)  
 - KEGG pathways enriched: 0

# Output Files

Located in GitHub repo or local path E:/UBC\_wang\_qn1/:

- RNA\_Seq\_analysis\_question1\_package.tar.gz  
- gene\_counts.txt  
- DEGs\_KO\_vs\_CTRL.csv  
- GO\_enrichment\_BP.csv  
- GSEA\_KEGG\_results.csv  
- volcano\_plot.png, MA\_plot.png, pca\_plot.png, heatmap\_top30.png  
- GO\_barplot.png, GO\_dotplot.png, GSEA\_dotplot.png

# Conclusion

This bulk RNA-seq analysis revealed exploratory insights into transcriptional changes following ATF3 knockout in mouse smooth muscle cells. While adjusted DEG significance was limited, exploratory thresholds identified several candidate genes and GO terms. Further validation and higher-powered replicates are recommended.