Bulk RNA-Seq Analysis Report: ATF3 Knockout in SMCs

Prepared by: Tao Sun

Date: April 13, 2025

# Objective

The objective of this project is to perform a bulk RNA-seq analysis using NCBI BioProject PRJNA716327. Specifically, the aim is to compare gene expression between ATF3 knockout Sca1+ smooth muscle cells (SMCs) and control groups, followed by downstream functional enrichment analysis.

# Sample Information

The following BioSamples were analyzed:  
- KO group: SAMN18442667, SAMN18442665  
- CTRL group: SAMN18442666, SAMN18442664

# Pipeline Overview

The RNA-seq analysis pipeline consisted of the following steps:

* 1. Download raw sequencing data using SRA Toolkit (prefetch + fasterq-dump)
* 2. Quality control using FastQC
* 3. Trimming low-quality reads using Trimmomatic
* 4. Aligning reads to the mouse genome (mm10) using HISAT2
* 5. Gene-level quantification using featureCounts
* 6. Differential expression analysis using DESeq2
* 7. Functional enrichment analysis using clusterProfiler (GO and KEGG)
* 8. GSEA analysis based on full ranked gene list

# Software and Tools

- SRA Toolkit (fasterq-dump, prefetch)  
- FastQC (v0.11.9)  
- Trimmomatic (v0.39)  
- HISAT2 (v2.1.0)  
- featureCounts (from Subread package)  
- R/Bioconductor: DESeq2, clusterProfiler, org.Mm.eg.db

# Results Summary

Due to sample limitations and experimental noise, very few differentially expressed genes (DEGs) were detected with stringent FDR thresholds. An exploratory threshold (p-value < 0.05, |log2FC| > 1) identified a small subset of DEGs (n = 8). Gene Ontology enrichment of these genes revealed terms associated with epithelial signaling, calcium ion transport, and morphogenesis.  
  
KEGG enrichment analysis yielded no significantly enriched pathways, likely due to the small gene set size.

# Deliverables

- Source code and scripts for all steps (available as a tar.gz archive)  
- Processed count matrix and DESeq2 results  
- GO enrichment barplots and dotplots

# Limitations

This analysis is limited by the small number of samples and minimal signal observed in differential expression. Results should be interpreted with caution and validated in future experiments with greater replication.