Transcriptomic analysis of knockout HKDC1

Aung Nyein

To explore how the absence of HKDC1 affects global gene expression, I utilized high-throughput RNA sequencing data from HKDC1 knockout(sgHKDC1-1) and control cell lines, from an original study conducted by Liu et al, 2024. The dataset used in this study is publicly available under the accession number **GSE216107**, which includes comprehensive transcriptomic information from these cell lines, providing valuable insights into the impact of HKDC1 depletion on gene expression. Differential gene expression analysis is performed in this study to see how many genes are upregulated and how many are downregulated, followed by gene set enrichment analysis to identify pathways or gene sets associated with HKDC1.

### Importing Libraries

library(dplyr)  
library(tidyverse)  
library(GEOquery)  
library(DESeq2)

### Importing Data

#Load RNA sequencing data  
RNA\_seq <- read.delim("C:/Users/Duche/Downloads/GSE216107\_gene\_count.txt.gz", header=TRUE)  
  
# get metadata  
gse<- getGEO(GEO='GSE216107', GSEMatrix= TRUE)  
metadata <- pData(phenoData(gse[[1]]))

### Pre-processing data

#Pre-processing data  
metadata\_modified<- metadata%>%  
 select(1,19)%>%  
 mutate(title = sub(" .\*", "", title))%>%  
 arrange(title)  
  
RNA\_seq\_modified<-RNA\_seq %>%  
 select(1:7)

### Differential Gene Expression using DESeq2 package

#Preparing data for Differential Gene Expression  
count\_data <- RNA\_seq\_modified[, -1] # Remove the gene column  
rownames(count\_data) <- RNA\_seq\_modified$gene\_id  
  
rownames(metadata\_modified) <- metadata\_modified$description

#Performing DIfferential Gene Expression  
dds <- DESeqDataSetFromMatrix(countData = count\_data, colData = metadata\_modified, design = ~ title)  
dds <- dds[rowSums(counts(dds)) > 10, ]  
dds <- DESeq(dds)  
  
# Get the results of the differential expression analysis  
res <- results(dds)

### Analyzing the results

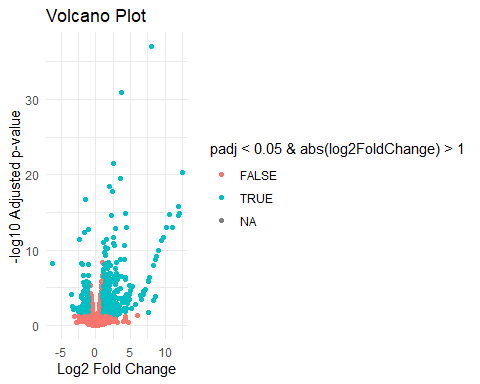
#To view only genes with adjusted p-value of 0.05  
res\_significant <- results(dds, alpha=0.05)  
summary(res\_significant)

out of 20967 with nonzero total read count  
adjusted p-value < 0.05  
LFC > 0 (up) : 893, 4.3%  
LFC < 0 (down) : 652, 3.1%  
outliers [1] : 6, 0.029%  
low counts [2] : 4878, 23%  
(mean count < 8)  
[1] see 'cooksCutoff' argument of ?results  
[2] see 'independentFiltering' argument of ?results

Out of 20,967 genes tested for differential gene expression, 893 genes are upregulated and 652 genes are downregulated in HKDC1 knockout celll lines at adjusted p-value threshold of 0.05.

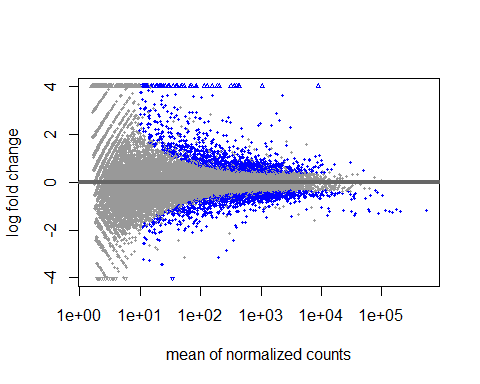
### Visualizing data

# Volcano plot  
library(ggplot2)  
ggplot(res, aes(x=log2FoldChange, y=-log10(padj))) +  
 geom\_point(aes(color = padj < 0.05 & abs(log2FoldChange) > 1)) +  
 theme\_minimal() +  
 labs(title = "Volcano Plot", x = "Log2 Fold Change", y = "-log10 Adjusted p-value")



In this volcano plot, genes with positive Log2 Fold Change are upregulated and those with negative values are downregulated. On the y-axis, genes with higher values are significantly different in their expression.

#MA plot  
plotMA(res)



The MA plot is another way of visualizing differential gene expression data. Genes that are significantly different in their expression between normal samples and HKDC1 knockout samples are colored in blue. Genes with positive log fold change are upregulated and those with negative values are downregulated in HKDC1 knockout cell lines.

### Gene Set Enrichment Analysis

library(fgsea)  
library(msigdbr)  
  
gene\_list <- res$log2FoldChange  
names(gene\_list) <- rownames(res)  
gene\_list <- sort(gene\_list, decreasing = TRUE) # Sort in decreasing order  
  
# Load the gene sets   
gene\_sets <- msigdbr(species = "Homo sapiens", category = "C7") # C7 = Immunological signatures  
  
# Convert the gene sets to a list format  
gene\_sets\_list <- split(gene\_sets$gene\_symbol, gene\_sets$gs\_name)  
  
# Run GSEA  
fgsea\_result <- fgsea(pathways = gene\_sets\_list, stats = gene\_list, nperm = 1000)

#Filtering only gene sets with adjusted p-value of less than 0.1   
sig\_fgsea\_result <- fgsea\_result[fgsea\_result$padj < 0.1, ]  
sig\_fgsea\_result[, c("pathway", "padj")]

pathway padj  
 <char> <num>  
1: GSE17974\_IL4\_AND\_ANTI\_IL12\_VS\_UNTREATED\_24H\_ACT\_CD4\_TCELL\_DN 0.08807588  
2: GSE2706\_UNSTIM\_VS\_2H\_R848\_DC\_DN 0.08807588  
3: GSE2706\_UNSTIM\_VS\_8H\_LPS\_DC\_DN 0.08807588  
4: GSE29615\_CTRL\_VS\_DAY7\_LAIV\_FLU\_VACCINE\_PBMC\_UP 0.08807588  
5: GSE43863\_NAIVE\_VS\_TH1\_EFF\_CD4\_TCELL\_D6\_LCMV\_DN 0.08807588  
6: GSE46242\_TH1\_VS\_ANERGIC\_TH1\_CD4\_TCELL\_DN 0.08807588

Using gene set enrichment analysis with gene sets related to immune response pathways from MSigDB, 6 gene sets are identified to be significantly associated with HKDC1 knockouts. This implies the role of HKDC1 in immune response pathways.

### References:

* Liu P, Luo Y, Wu H, Han Y et al. HKDC1 functions as a glucose sensor and promotes metabolic adaptation and cancer growth via interaction with PHB2. Cell Death Differ 2024 Dec;31(12):1595-1610. PMID: [39375512](https://www.ncbi.nlm.nih.gov/pubmed/39375512)