

# Differential Expression Analysis Report

Dataset: GSE174431 (Breast Cancer PBMCs)

Prepared by: Hanieh Roodashty

GitHub: [https://github.com/hanyroze/GSE174431\\_DESeq2\\_BreastCancer](https://github.com/hanyroze/GSE174431_DESeq2_BreastCancer)

Date: August 30, 2025

## Introduction

Dataset: GSE174431

Context: RNA-seq analysis of PBMCs from metastatic breast cancer patients. Samples were separated into lineage-positive (Lin+) and lineage-negative (Lin-) populations.

Objective: Identify differentially expressed genes between Lin+ (differentiated) and Lin- (stem-like/progenitor) cells.

## Methods

- Downloaded raw exon count files from GEO (GSE174431).
- Processed and merged counts into a single matrix in R.
- Used DESeq2 for normalization and differential expression analysis.
- Groups defined from sample metadata (LinPositive vs LinNegative).
- Significance threshold: adjusted p-value (FDR) < 0.05.

## Results

- PCA shows clear separation between Lin+ and Lin- samples.
- Volcano plot highlights significantly up- and down-regulated genes.
- Heatmap of the top 20 DE genes reveals distinct clustering by condition.



Figure 2. Volcano Plot

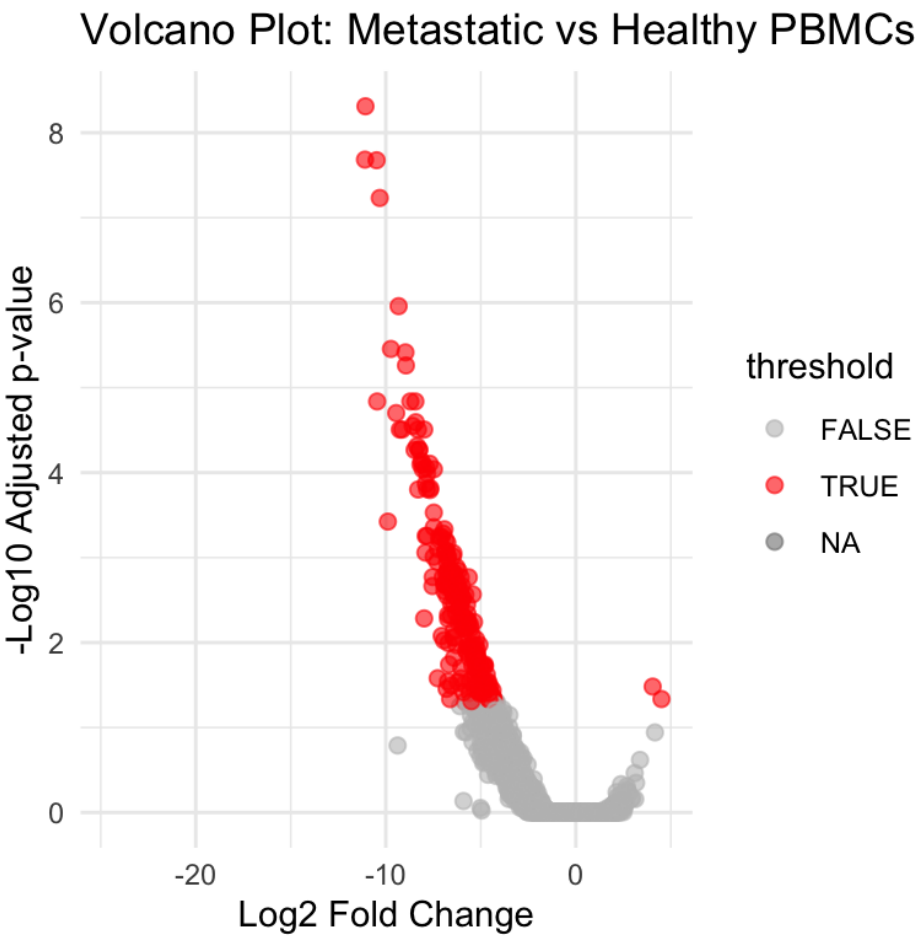
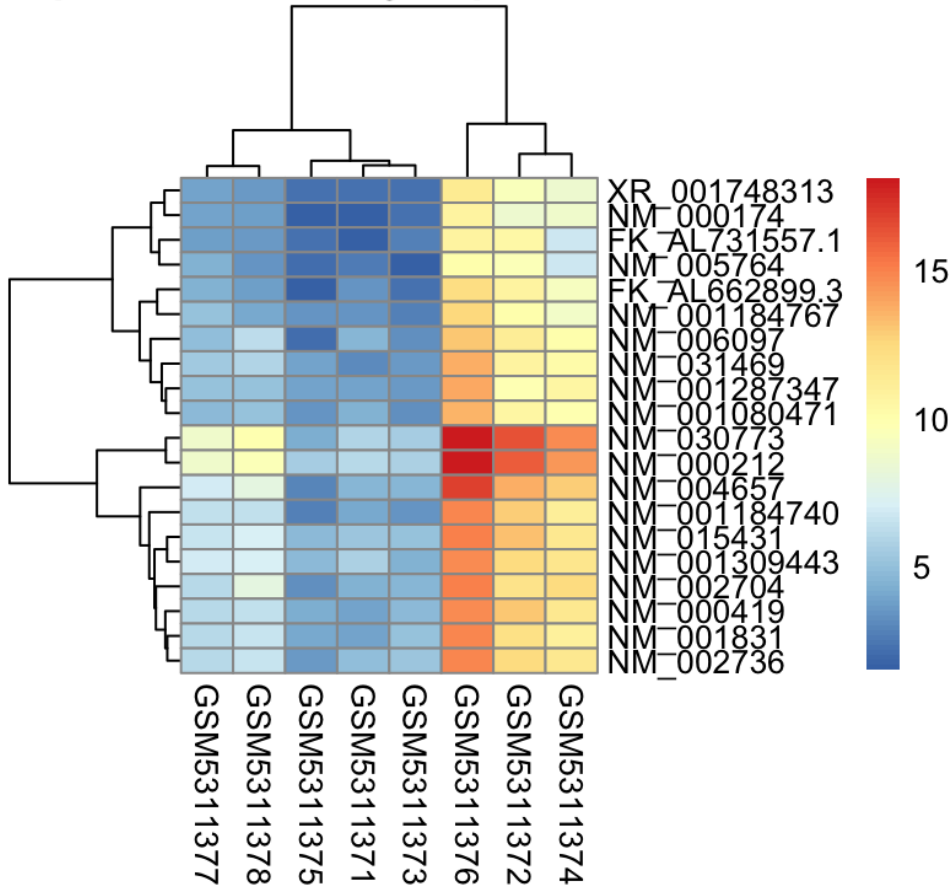


Figure 3. Heatmap of Top 20 DE Genes

Top 20 Differentially Expressed Genes



**Table 1. Top 20 Differentially Expressed Genes**

Gene	log2FC	Adj.P.Val
NM_030773	-11.07	4.89e-09
NM_004657	-11.10	2.07e-08
NM_000212	-10.48	2.11e-08
NM_001184740	-10.32	5.86e-08
NM_002704	-9.32	1.10e-06
XR_001748313	-9.73	3.51e-06
NM_000419	-8.97	3.85e-06
NM_031469	-8.94	5.48e-06
NM_000174	-10.45	1.45e-05
NM_001831	-8.71	1.45e-05
NM_015431	-8.44	1.45e-05
FK_AL662899.3	-9.45	1.99e-05
NM_002736	-8.42	2.55e-05
NM_001287347	-8.58	2.82e-05
NM_001184767	-8.31	3.11e-05
NM_005764	-9.26	3.11e-05
NM_001309443	-7.98	3.11e-05
FK_AL731557.1	-9.14	3.11e-05
NM_001080471	-8.35	4.88e-05
NM_006097	-8.23	5.35e-05

## Conclusion

This analysis identified significant transcriptional differences between Lin<sup>+</sup> and Lin<sup>-</sup> breast cancer cell populations.

- Lin<sup>-</sup> cells (lineage negative) show expression patterns consistent with stem-like/progenitor features.
- Lin<sup>+</sup> cells (lineage positive) display more differentiated gene expression signatures.

These findings align with the hypothesis that lineage-negative cells may represent cancer stem-like populations contributing to tumor progression and metastasis.