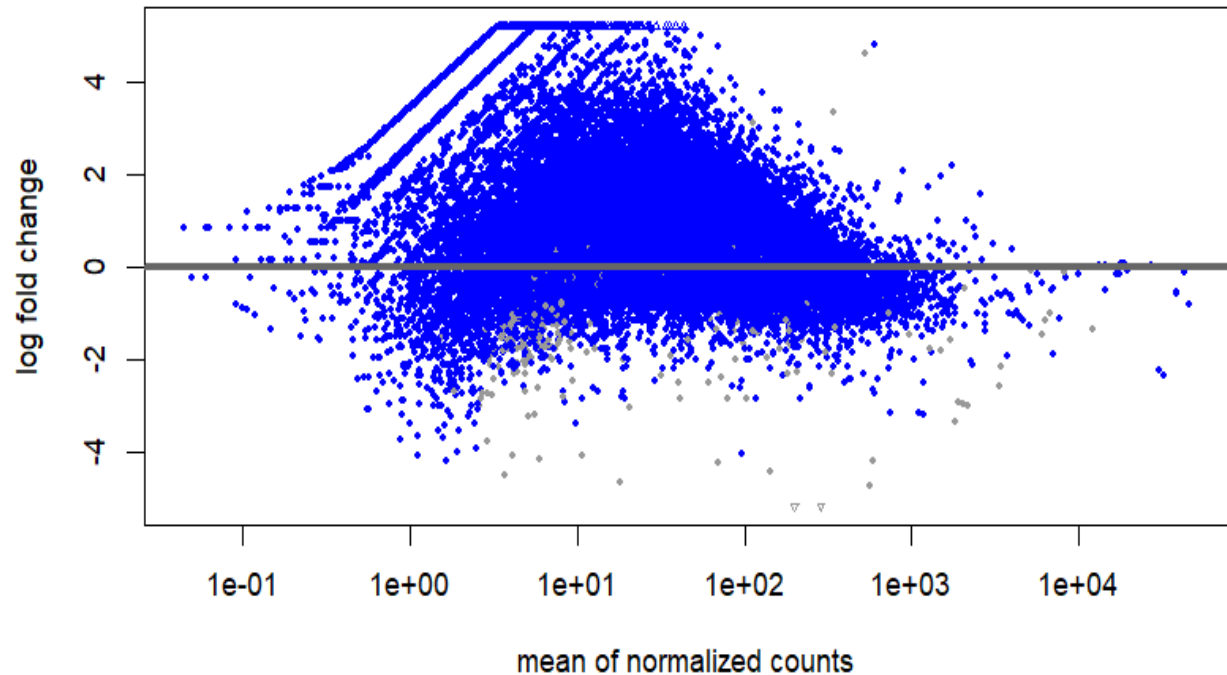# GSE18394 R plots

Based on the extracted data from GEO dataset GSE18394 the following R plots were plot

MA, PCA, Heat Map and Volcano Plot

# Mean vs. Average Plot

Code-
plotMA(res,alpha=0.2,ylim = c(-5, 5))
plotMA(res, alpha = 1) # disables padj threshold

- This MA plot displays the log2 fold changes (Y-axis) versus mean gene expression (X-axis) for all genes analyzed using DESeq2.
- Each point represents a gene. The **blue points** indicate genes that are significantly differentially expressed between **metastatic** and **non-metastatic** breast tissue samples (adjusted p-value < 0.1), after applying fold-change shrinkage for accuracy.
- Most genes are not significantly differentially expressed, clustering around a log2 fold change of 0.
- A subset of genes show strong expression differences

# Principle Component Analysis Plot – Sample Clustering by Metastasis Status
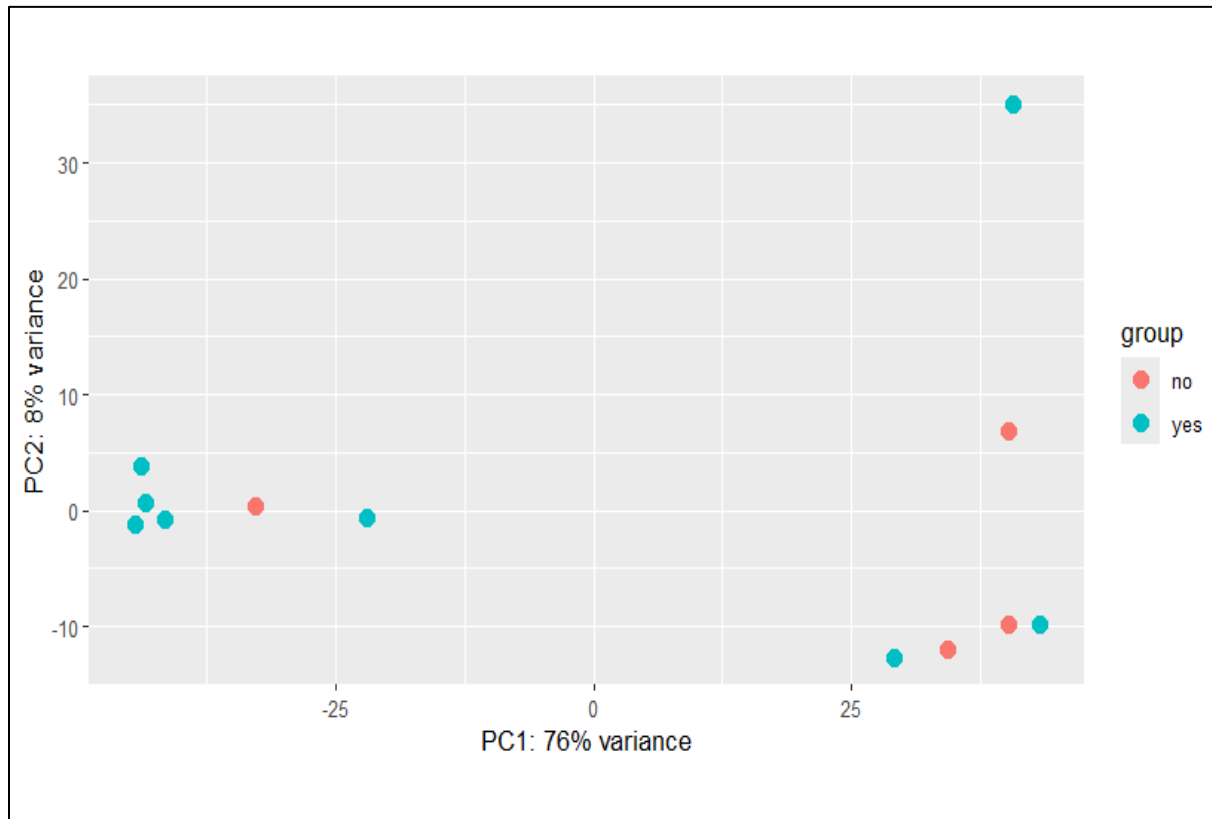
- This PCA plot shows the first two principal components (PC1 and PC2) based on variance-stabilized gene expression across all samples.
- PC1 explains 76% of the variance in the dataset
- PC2 explains 8% of the varianceSamples are colored by metastasis status:
- Blue = Metastatic
- Red = Non-metastatic
- The strong separation along PC1 suggests that metastasis status is the dominant source of variation in global gene expression. This indicates that metastatic and non-metastatic samples have distinct transcriptional profiles.

Code-
vsd <- vst(dds, blind = FALSE)
plotPCA(vsd, intgroup = "Metastasis")

Heat Map

Top 30 Differentially Expressed Genes
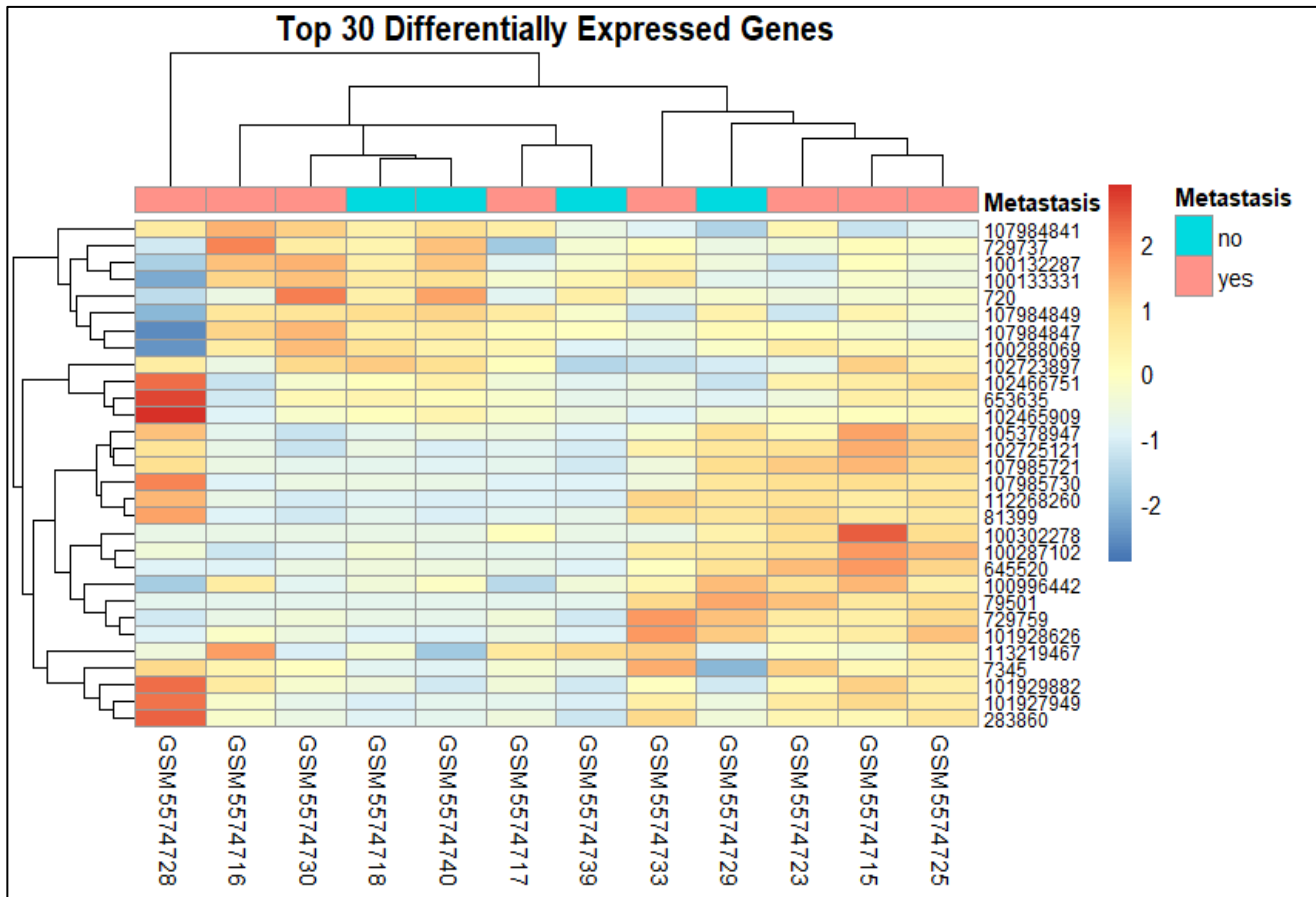
This heatmap displays the expression levels of the top 30 most differentially expressed genes (based on adjusted p-value) across all samples.
- Each row = one gene
- Each column = one sample
- Red = high expression,
- Blue = low expression (Z-score normalized)

Samples are annotated by **metastasis status**. The consistent expression patterns across groups suggest strong transcriptional signatures associated with metastasis.

Exploratory Volcano Plot of Differential Gene Expression
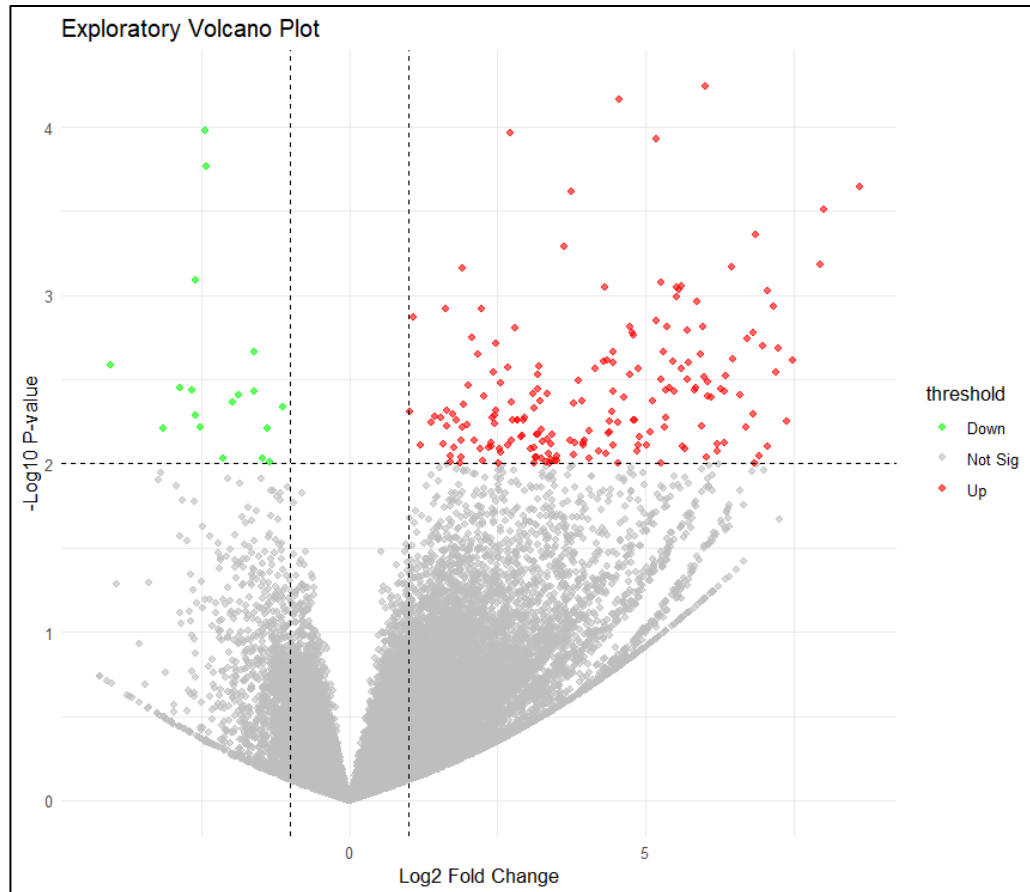
- Here, we see a volcano plot showing changes in gene expression between metastasis-positive and -negative samples.
- Although no genes reached statistical significance after multiple testing correction, we relaxed the threshold to p < 0.01 and log2 fold change > 1 to explore potential trends.
- Genes in red are upregulated in metastatic samples, and green are downregulated.
- These trends could represent biologically meaningful differences that didn't reach significance due to limited sample size or variability, and they suggest candidates for future validation studies.