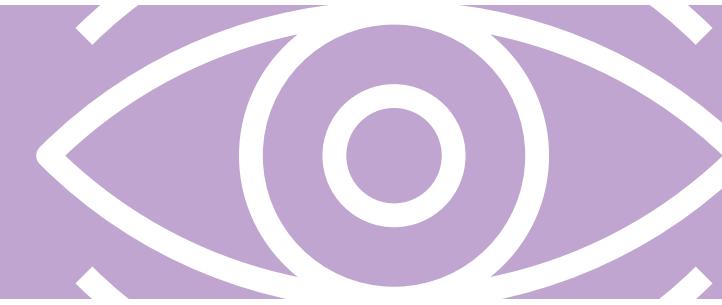


# Biomarkers for breast invasive carcinoma using TCGA data.

- ***ADAM ELÍ DAVÍÐSSON*** -

- ***MARGRÉT DÍS STEFÁNSDÓTTIR*** -



# OVERVIEW

- Introduction/problem relevance
- Aim
- Steps
- Approach
- Results
- Future work
- Conclusions



# INTRODUCTION/PROBLEM RELEVANCE

- Most common type of breast cancer that is characterized by the abnormal growth of cells in the breast tissue.
- Biomarkers can be used as diagnostic tools to identify patients with abnormal conditions.
- RNA sequencing-based differentiation gene expression analysis.
- Compare healthy and cancerous breast tissues.
- Multiple testing.





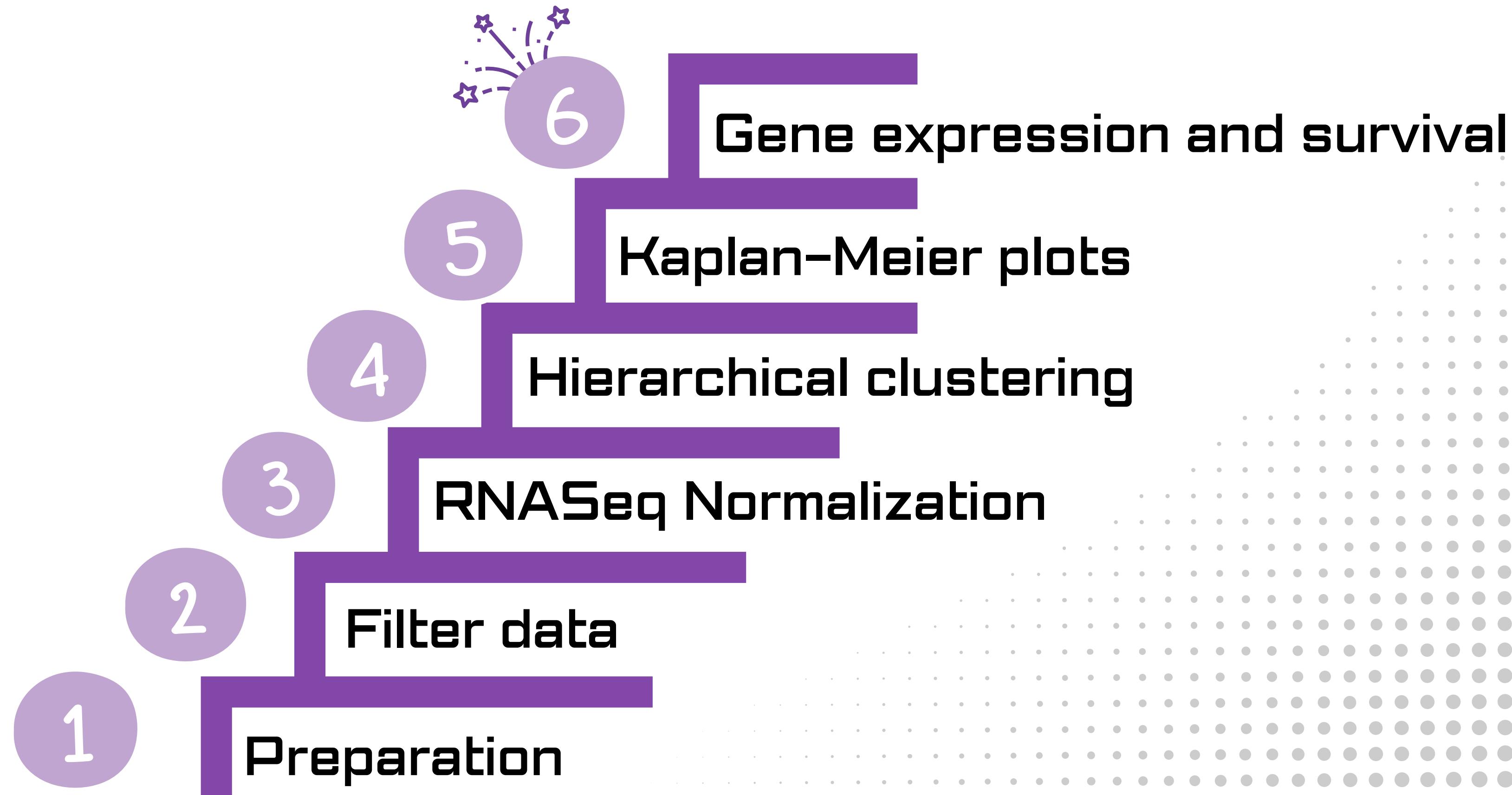
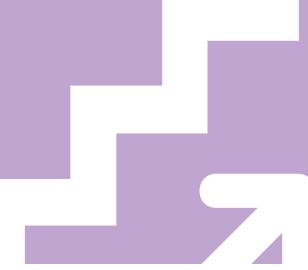
AIM

"Our aim is to use RNA sequencing data to identify and cluster differences in gene expression between healthy and cancerous breast tissue and to identify highly expressed genes that may be potential biomarkers for breast invasive carcinoma prognosis."





# STEPS



# APPROACH



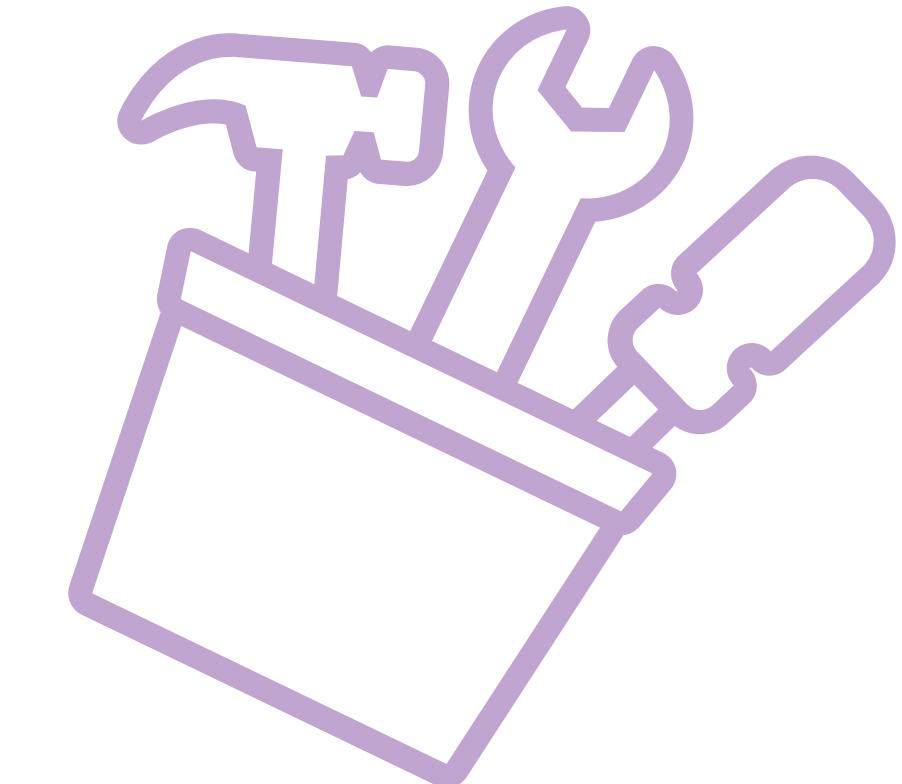
## Tools:

- TCGAbiolinks package
- PCA
- differential expression analysis
- Elastic Net classifier model.

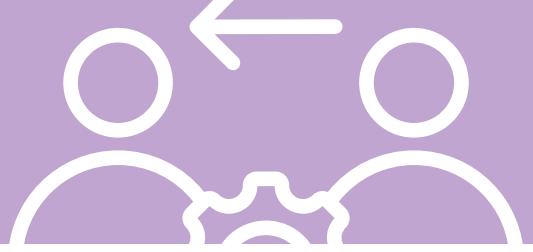


## Data:

- The cancer Genome Atlas (TCGA).
- 1224 samples (113 normal solid tissue & 1111 primary tumor samples).
- Raw expression data 60.660 elements.



# APPROACH



## Methods:

- PCA->assess patterns between normal and cancer tissue samples.
- limma-pipeline function and Elastic Net classifier model->classification.
- Relevant genes are hierarchically clustered.

## Purpose:

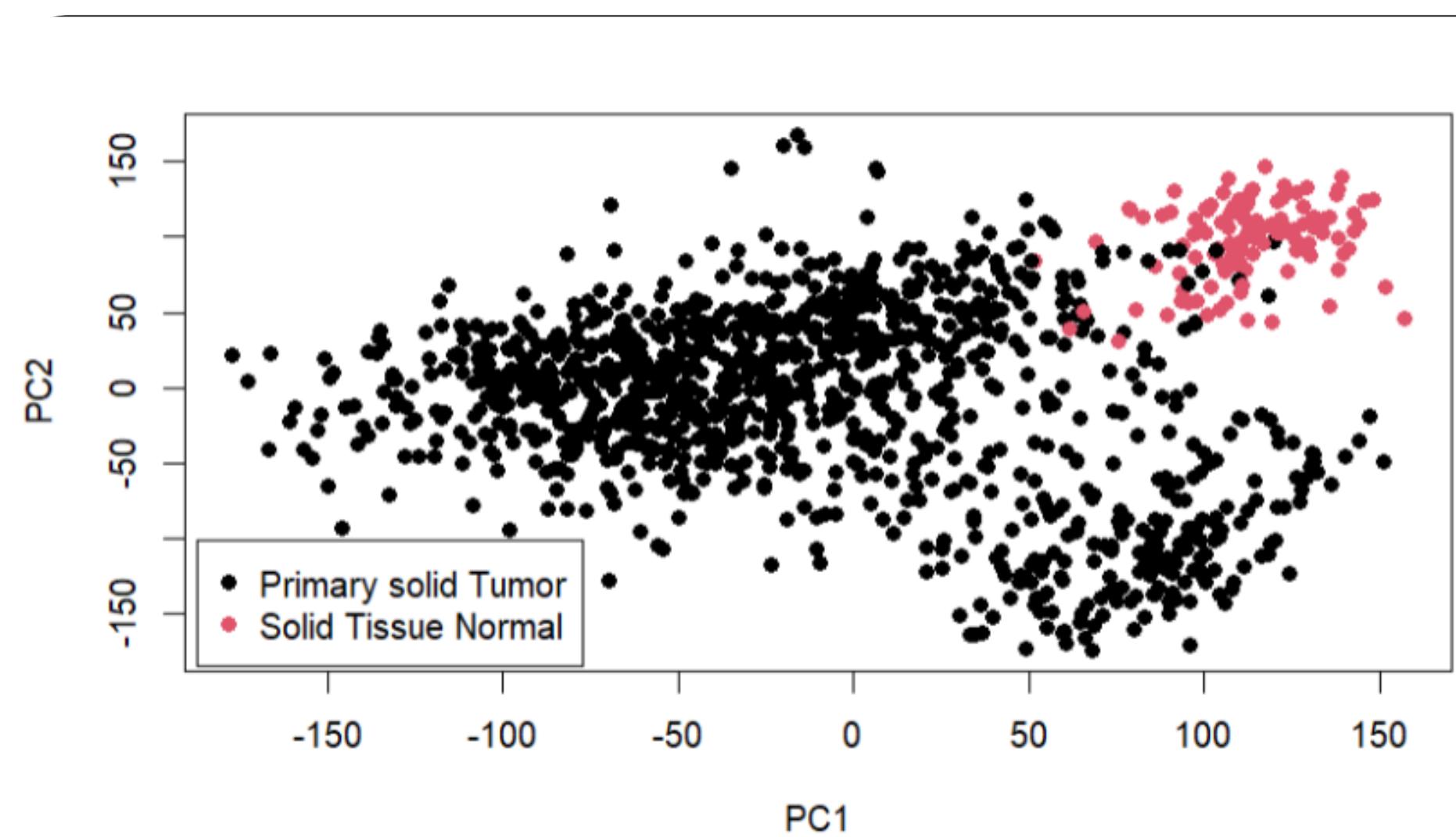
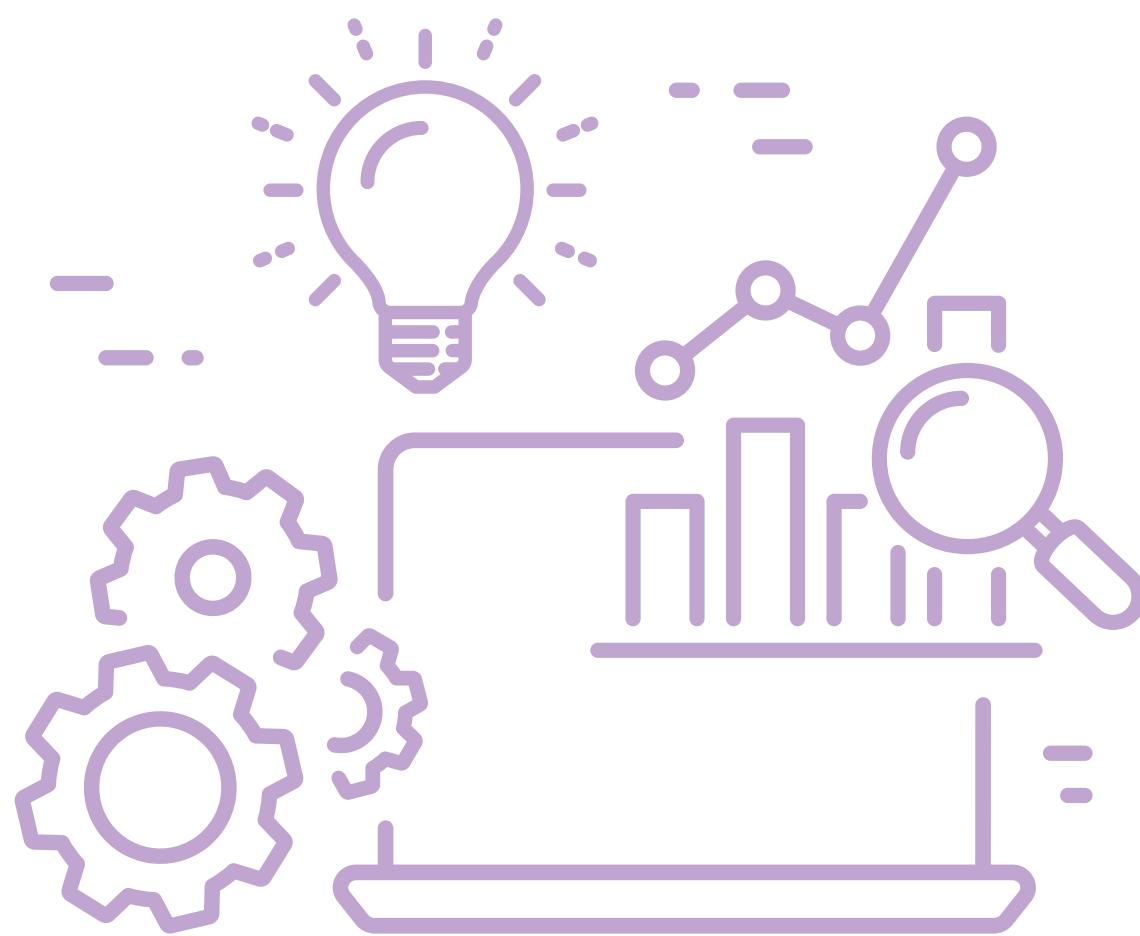
- Identify and cluster differences in gene expression.
- Identify potential biomarkers for breast invasive carcinoma.
- Survival rates.





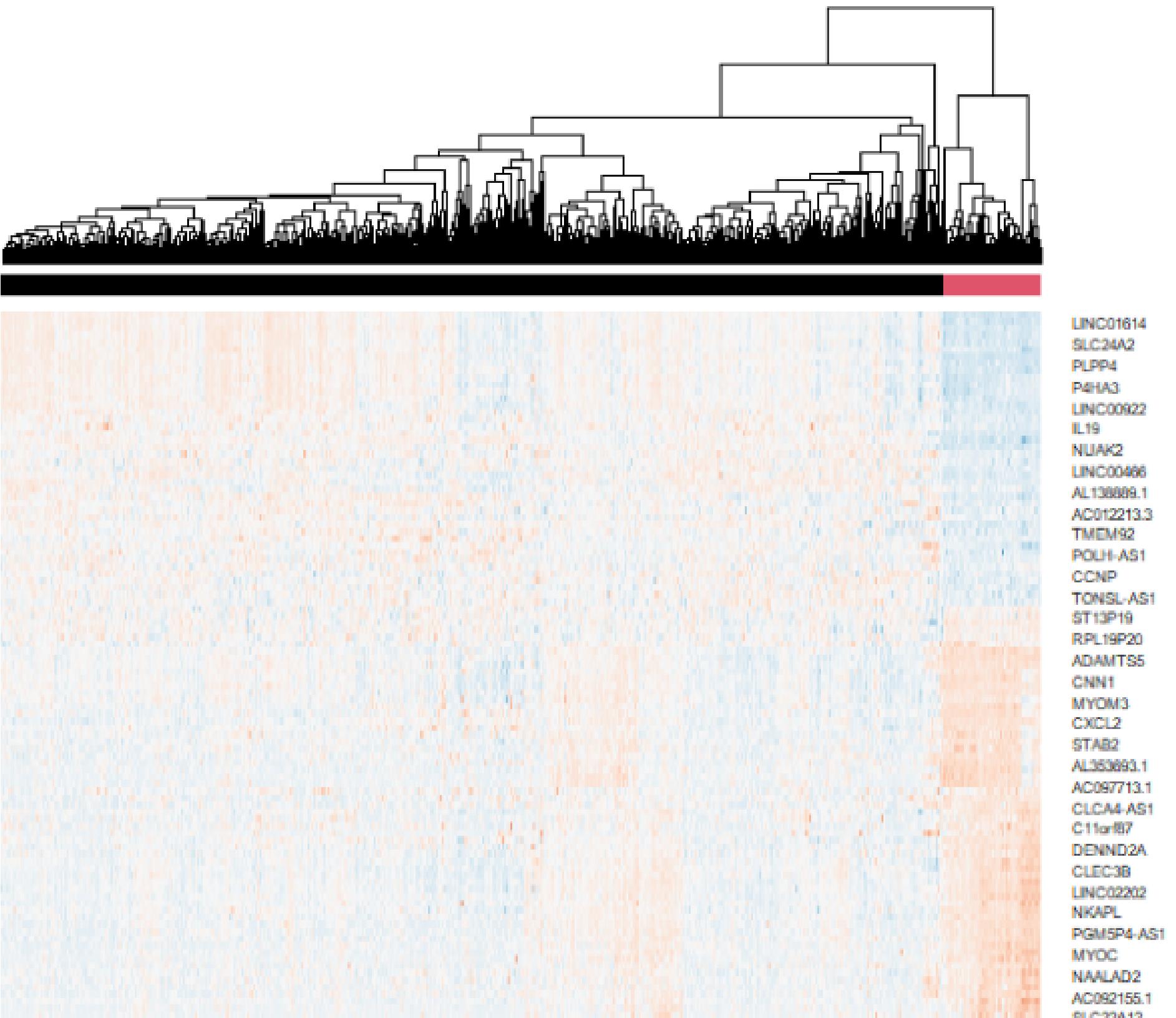
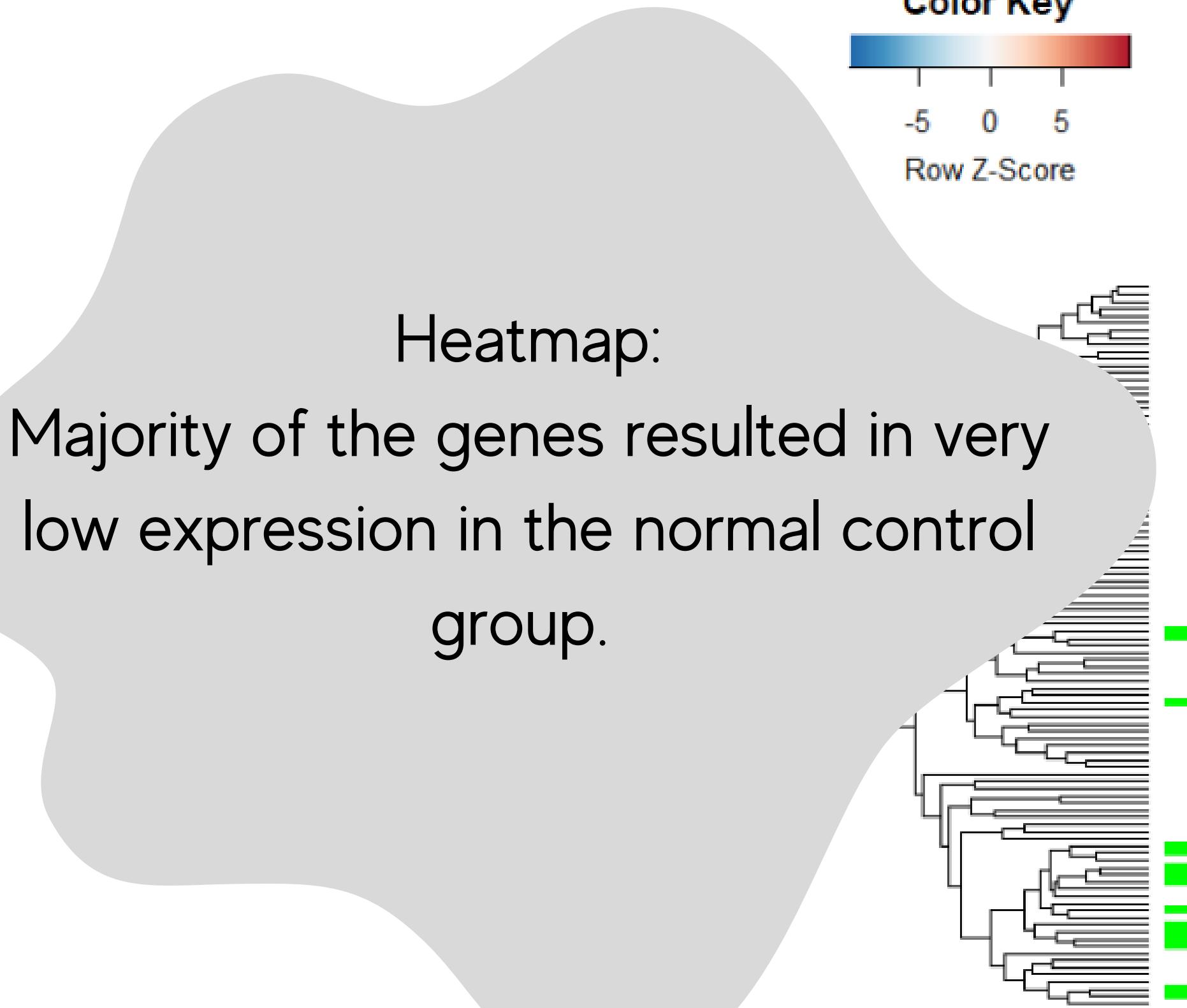
# RESULTS

- Principal Component Analysis (PCA):
- Significant differences in gene expression patterns between normal tissue and breast cancer tissue.





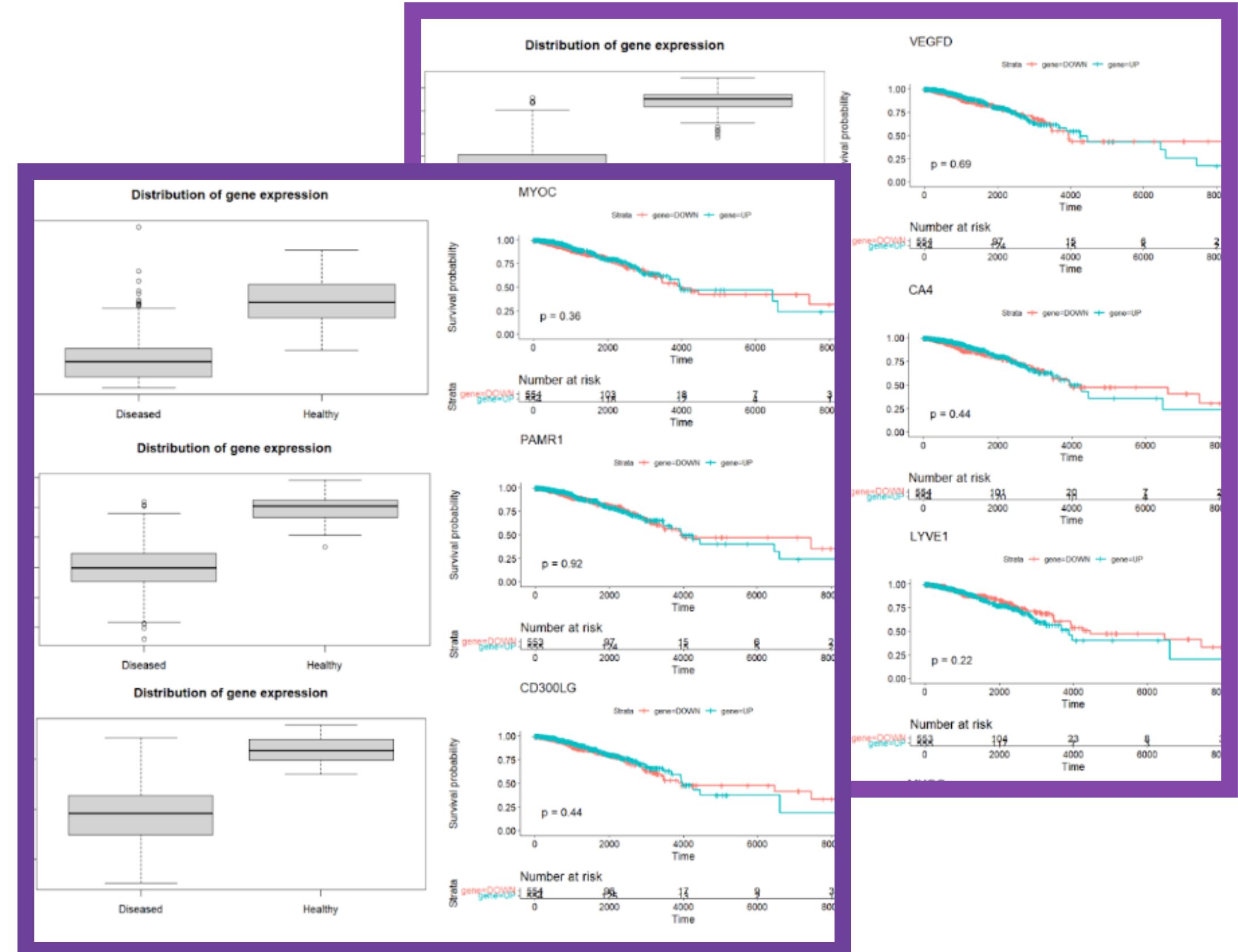
# RESULTS



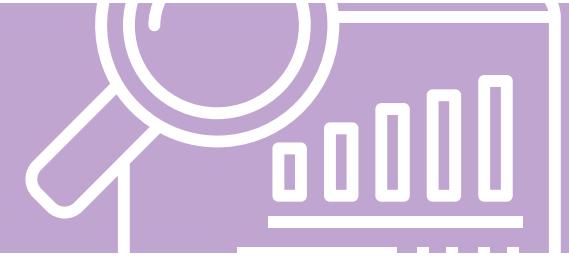


# RESULTS

- Two images visualized for each gene selected.
- The top six most differentially expressed genes are used.
- Kaplan-Meier Survival plot on prognosis for clustered genes.
- Up-regulated --> patient has high gene expression
- Down-regulated --> patient has low gene expression.
- No difference for prognosis.



# FUTURE WORK



- Analyze more gene expressions to find effective biomarkers.
- Compare different types of breast cancer.
- Check if age plays any role when it comes to survival, and analyze that further.



# CONCLUSIONS



- Six most differentially expressed genes:
  1. VEGFD
  2. CA4
  3. LYVE1
  4. MYOC
  5. PAMR1
  6. CD300LG
- None of the expression genes appeared to make any difference for prognosis.





# Thank You