RNA-Seq Analysis of Breast Cancer vs Normal Tissue

Student Name: Shriya Arora
Date: 04/01/2026

1. Introduction

Breast cancer is one of the most common malignancies worldwide. Understanding differences in gene expression between tumor and normal tissue can reveal key pathways and biological processes involved in cancer progression.

This project aims to identify differentially expressed genes (DEGs) between breast cancer and normal tissue using RNA-Seq data from GSE62944, and to perform functional enrichment analyses (KEGG and GO BP) to uncover significant biological mechanisms.
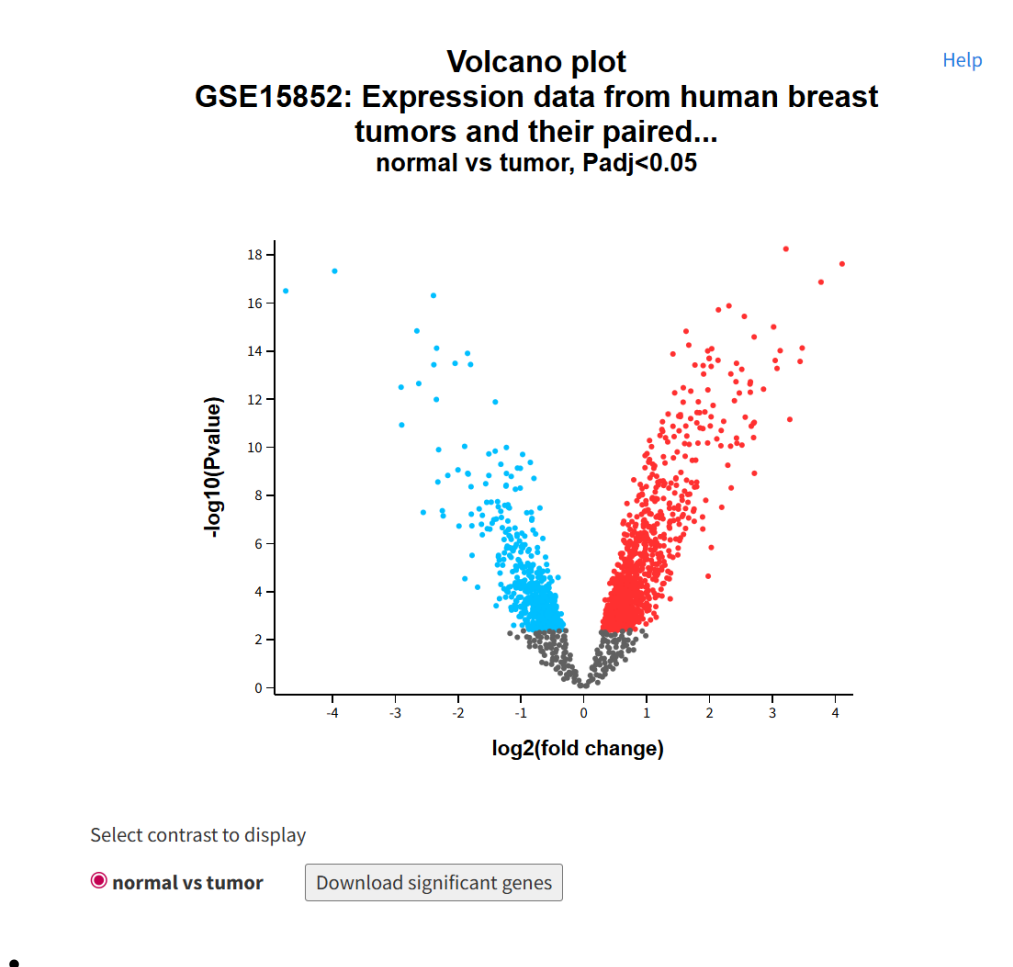
2. Materials and Methods

- **Dataset:** GSE62944 (breast cancer RNA-Seq data from GEO database)
- **Sample groups:**
    - Breast cancer tissue samples
    - Normal breast tissue samples
- **Analysis tool:** GEO2R (web-based tool for differential gene expression)
- **Analysis steps:**

1. Access *GEO2R* for GSE62944.

2. Assign samples to "Breast Cancer" and "Normal" groups.

3. Perform differential expression analysis to generate a list of DEGs (genes with p-value < 0.05 and fold change > 1).

4. Export the list of DEGs for enrichment analyses.

- **Functional enrichment:**
    - *KEGG* pathway analysis
    - GO Biological Process (BP) analysis
- **Visualization:** Bar plots of top enriched pathways and processes were generated in Excel.

3. Results

3.1 Differentially Expressed Genes (DEGs)

- Differential expression analysis identified multiple significantly upregulated and downregulated genes in breast cancer tissue compared to normal tissue (adjusted p-value < 0.05).

Volcano plot
GSE15852: Expression data from human breast tumors and their paired...
normal vs tumor, Padj<0.05

Select contrast to display

◉ **normal vs tumor**   [Download significant genes]

The volcano plot represents genes differentially expressed between breast cancer and normal tissue. Genes with high fold change and low p-values are highlighted, indicating significant upregulation or downregulation in cancer samples.

Note/Exception: During differential expression analysis using *GEO2R,* a discrepancy was observed in the number of samples processed. Although the dataset (GSE62944) originally contained 271 samples, only 211 samples were included in the final analysis. This reduction occurred due to missing expression values or incomplete metadata for certain samples, which caused *GEO2R* to automatically exclude them during

preprocessing and normalization. Such automatic filtering is a standard quality-control step in transcriptomic analysis to ensure statistical reliability. The final results were therefore based on the successfully processed 211 samples, and all downstream GO and *KEGG* enrichment analyses were performed using these filtered data.
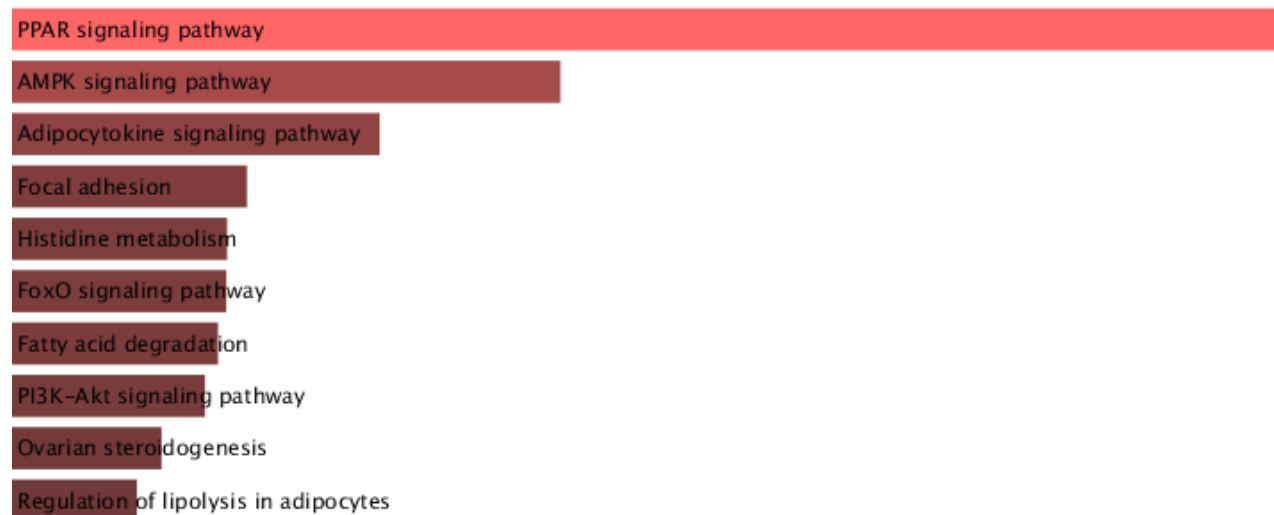
3.2 *KEGG* Pathway Enrichment

Figure 1: Top enriched *KEGG* pathways among DEGs.

KEGG pathway analysis revealed significant enrichment of pathways such as PPAR signaling, AMPK signaling, Adipocytokine signaling, PI3K-Akt signaling, and Focal adhesion, which are involved in metabolic regulation, cellular signaling, and cancer progression. These pathways are known to influence tumor growth, energy metabolism, and cell survival in breast cancer.

Exceptions: Some pathways may appear enriched due to overlapping genes or dataset-specific biases. Only pathways passing statistical thresholds should be considered highly reliable.

3.3 GO Biological Process (BP) Enrichment

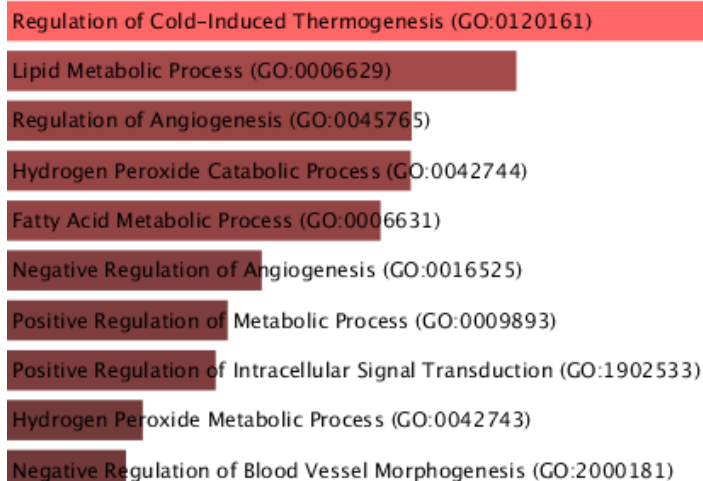| Process |
|---|
| Regulation of Cold-Induced Thermogenesis (GO:0120161) |
| Lipid Metabolic Process (GO:0006629) |
| Regulation of Angiogenesis (GO:0045765) |
| Hydrogen Peroxide Catabolic Process (GO:0042744) |
| Fatty Acid Metabolic Process (GO:0006631) |
| Negative Regulation of Angiogenesis (GO:0016525) |
| Positive Regulation of Metabolic Process (GO:0009893) |
| Positive Regulation of Intracellular Signal Transduction (GO:1902533) |
| Hydrogen Peroxide Metabolic Process (GO:0042743) |
| Negative Regulation of Blood Vessel Morphogenesis (GO:2000181) |

Figure 2: Top enriched GO biological processes among DEGs.

GO Biological Process analysis identified significant enrichment of processes including regulation of cold-induced thermogenesis, lipid metabolic process, fatty acid metabolism, angiogenesis regulation, and hydrogen peroxide metabolic processes. These processes suggest altered metabolic activity, oxidative stress response, and vascular regulation in breast cancer tissue.

Exceptions: Enrichment analyses depend on the gene set input and the database version. Some processes may be underrepresented if certain genes were not annotated.

4. Discussion

- The differential expression analysis revealed several genes significantly upregulated or downregulated in breast cancer tissue compared to normal tissue.

- Functional enrichment indicates key pathways (e.g., **metabolic and signaling pathways such as PPAR, AMPK, PI3K-Akt signaling**) and biological processes (e.g., **cell proliferation, DNA repair, apoptosis**) that are dysregulated in breast cancer.

- These results are consistent with previous studies highlighting the importance of these pathways in cancer progression.

- **Limitations / Notes:**

  - *GEO2R* provides a web-based, user-friendly tool but lacks advanced options for batch correction or complex normalization.

- - - Results may vary with dataset quality, sample size, or choice of thresholds.
    - Enrichment analyses may include false positives due to overlapping gene annotations or incomplete database information.
- Insights from this analysis could guide further experimental studies or potential therapeutic targets.

5. Conclusion

This project successfully identified differentially expressed genes between breast cancer and normal tissue and performed functional enrichment analysis to highlight key pathways and processes involved in cancer. While some exceptions exist due to tool and dataset limitations, *KEGG* and GO BP analyses revealed significant mechanisms contributing to tumorigenesis, providing a foundation for further research in breast cancer biology.

6. References

1. GEO Accession: **GSE62944** – https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE62944
2. *GEO2R* Tool: https://www.ncbi.nlm.nih.gov/geo/geo2r/
3. *KEGG* Pathway Database: https://www.genome.jp/kegg/
4. Gene Ontology Database: http://geneontology.org/