# Boolean Analysis of Gene Expression in Breast Cancer Relapse


## Abstract

In this project, we applied Boolean logic to analyze gene expression patterns in the GSE2034 breast cancer dataset. Our aim was to identify genes with binary ON/OFF states that are differentially expressed in patients who experienced bone relapse versus those who did not. Using mean-based thresholding, we converted gene expression values into binary states and discovered 62 genes that were ON in >=60% of relapse samples and OFF in <=40% of non-relapse samples. Pathway enrichment analysis using KEGG databases revealed significant involvement in prion disease, toxoplasmosis, and estrogen signaling pathways. These findings provide a basis for future exploration of relapse-specific gene signatures and support further validation studies to refine predictive biomarkers.


## Method Summary

- Dataset: GSE2034 (286 breast cancer samples)
- Preprocessing: Normalized and thresholded using gene-wise mean
- Binary Logic: Converted expression to ON (1) / OFF (0)
- Group Comparison: Bone relapse = YES vs NO
- Marker Selection: Genes ON in >=60% of relapse & OFF in <=40% of non-relapse
- Enrichment: KEGG pathway analysis via gseapy

## Key Results

- Marker Genes Identified: 62
- Top Pathways: Prion disease (adj. p = 0.0035), Toxoplasmosis (adj. p = 0.028), Parkinson disease (adj. p = 0.05)
- Estrogen signaling pathway also enriched (adj. p = 0.128)

## Conclusion

Boolean analysis enabled the identification of discrete gene expression signatures linked to breast cancer relapse. The involvement of neurodegenerative and hormone-related pathways highlights new angles for biomarker discovery. Future work will include validation across additional datasets and integration with clinical outcomes.