**LASSO-Based Gene Signature: Prognostic Score Construction and Validation**

This repository contains the full R pipeline used to derive and validate a prognostic gene expression signature using LASSO Cox regression. The model was trained on z-score normalized expression data and survival outcomes (Overall Survival and Progression-Free Interval).

**Overview**

- Selects genes using LASSO Cox models (OS and PFI)  
- Builds a composite prognostic score from overlapping genes  
- Dichotomizes patients into High vs. Low risk groups  
- Generates Kaplan–Meier curves (OS and PFI)  
- Compares performance of 10-gene and 9-gene (CGB8-excluded) scores  
- Outputs plots and prints the final score formula

# Folder Structure

lasso\_signature\_pipeline/  
├── data/ # Place your dataset here  
│ └── LASSO\_definitive\_dataset.xlsx  
├── outputs/ # Output folder for KM plots (auto-generated)  
├── lasso\_signature\_full\_pipeline.R # Main R script  
├── README.md # This file

# Required Data File

Important: The dataset `LASSO\_definitive\_dataset.xlsx` is included in this repository under `data/`. It contains anonymized or simulated data and is provided for academic use.

# How to Run

1. Open R or RStudio  
2. Install required packages (only once):  
install.packages(c("readxl", "glmnet", "survival", "survminer", "gridExtra"))  
3. Place the dataset under ./data/  
4. Run the script:  
source("lasso\_signature\_full\_pipeline.R")  
All output plots will be saved under the ./outputs/ folder.

# Output

The script generates:  
- A printed list of genes selected by both OS and PFI models  
- The final formula used to compute the prognostic score  
- Kaplan–Meier survival plots:  
 - OS\_10gene.pdf  
 - OS\_9gene.pdf  
 - PFI\_10gene.pdf  
 - PFI\_9gene.pdf  
 - KM\_OS\_comparison.pdf  
 - KM\_PFI\_comparison.pdf

# Reproducibility

- Tested on R 4.1.0 or higher  
- Script includes set.seed(12345) for consistent LASSO results  
- Final sessionInfo() is printed for transparency

# Contact

For questions, data access, or collaboration:  
  
Mauricio Cuello Fredes  
Pontificia Universidad Católica de Chile  
Email: mcuello@uc.cl

# License

This project is licensed for academic and non-commercial use. Please cite the corresponding manuscript when reusing this code or methodology.