# Gene Expression Analysis

This repository contains code for analysing gene expression data from microarray experiments.

## Data Files

The following data files are required for running the analysis:

1. `Gene\_Expression\_Data.xlsx`: Excel file containing gene expression data.

2. `Sample\_Information.tsv`: Tab-separated file containing sample information.

3. `Gene\_Information.csv`: CSV file containing gene information.

## R Code

The R code performs similar tasks as the Python code, including:

1. Loading the data files into R data frames.

2. Changing the sample names based on the phenotype information.

3. Splitting the data into tumor and normal samples.

4. Computing the average gene expression for tumor and normal samples.

5. Calculating the fold change for each probe between the two groups.

6. Identifying genes with fold change magnitude greater than 5.

7. Adding a column indicating whether the gene is upregulated or downregulated in tumor samples.

8. Creating visualizations like histograms and bar charts.

## Output Files

The code generates the following output files:

- `Updated\_Gene\_Expression\_Data.xlsx`: Excel file with updated sample names.

- `df\_tumor.csv`: CSV file containing tumor sample data.

- `df\_normal.csv`: CSV file containing normal sample data.

- `average\_tumor\_in\_df\_tumor.csv`: CSV file with average gene expression for tumor samples.

- `average\_normal\_in\_df\_normal.csv`: CSV file with average gene expression for normal samples.

- `fold\_change\_probes.csv`: CSV file with fold change values for each probe.

- `merged\_data.csv`: CSV file containing merged data from gene information and fold change values.