



---

# Introduction to Plots using R and Sequence Alignment(50 points)

## Objective

Introduce students to basic R programming and its applications in bioinformatics. Upon completion, students should be able to use basic R syntax and perform data visualization. Students will be able to perform sequence alignment and discover famous datasets.

## Installation

To begin this lab, you need to first install R binaries and R Studio. You can find the installation instructions in the following [link](#)

## Requirements

- You are obligated to attempt all tasks
- You **MUST** use R Markdown to submit your report
- State your name and ID at the first cell in your markdown report
- The lab should be

## Part 1: Data Visualization (15 points)

NOTE: Depending on the task, you may need to transpose your data frame

### Task 1.1: Visualizing Trend (5 points)

1. Import ggplot2 package. For more information on ggplot2 check the references section.
2. Create a smaller dataframe of the first 150 genes
3. Calculate the mean expression for all genes
4. Create a ggplot2 scatter plot showing expression levels of the first sample and the mean gene expression

## Task 1.2: Advanced Visualization (10 points)

1. Generate a ggplot2 bar plot showing the mean expression levels for the top 10 genes
2. Compare the first gene only for all the cancer types (phenotype) by drawing a box plot (hint: specify the fill as the cancer type)
3. Split the plots on different frames (hint: Facets)

## Part 2: Sequence Alignment (30 Points)

Use the below Sequences in your execution:

1. Sequence A: AGCTGAACTAGCTAGCTGACTGACTGACTAGCTAGCTGACTAGCTG

2. Sequence B: AGCGAACTAGCTGACTGACGACTGACTAGCTGACTAGCTGACTAGC

### Task 2.1: Installing Biostrings (5 Points)

Install biostrings using the following:

```
install.packages("Biostrings")  
library(BiocManager)  
BiocManager::install("Biostrings")  
library(Biostrings)
```

### Task 2.2: Run Pairwise Alignment (25 Points)

Use the 'pairwiseAlignment' function from the Biostrings package to perform alignments.

Hint: The 'pairwiseAlignment' function requires you to specify the sequences you want to align. There are various parameters like scoring scheme and penalty gaps that we encourage you to explore. Refer to [Pairwise Alignment](#) for more clarification.

Hint: Once you've performed the alignment, the result will be an S4 object. You'll need to extract the aligned sequences and scores from this object. Familiarize yourself with the accessor methods for S4 objects.

## Reporting (5 points)

- You will be marked on reporting and code readability and cleanliness
- Write an introduction to the complete report and an introduction for each section of your code
- Write clear comments

## References

- [BrainCancer Dataset](#)
- [About the BrainCancer Dataset](#)
- [ggplot2 Documentation](#)
- [T-Test in R](#)
- [Principal Component Analysis in R](#)
- [Pairwise Alignment](#)
- [S4 Object](#)
- [Markdown Cheatsheet](#)