BIN508 Assignment 1

Due Date: 13 March, Thursday, 13:30 **Cut-off:** 13 March, Thursday, 19:30

Late Policy: After the due time, 5 point deduction will be applied for each extra hour.

- **Uploading your assignment as a single PDF is mandatory.**
- Remember the Win + Shift + S shortcut for screenshots (on Windows).
- Using MS PowerPoint or Google Slides with horizontal format is highly suggested.
- Complete the methods first, then start writing your report.
- You are free to use Galaxy servers (USA, EU) or a Linux/Unix environment.
- **Make your Galaxy history and steps visible, and add a link to your Galaxy history at the top of the assignment. If you use a Linux environment, add your script to the end of the assignment with a mono-space font and include comments.**

Complete the Tutorial on Quality Control (FastQE and Nanoplot & Nanopore steps **ARE NOT INCLUDED**):

1) "**Assess quality with FastQC**" and "**Trim and filter - short reads**": Add screenshots for `FastQC` and `cutadapt` results and comment on them.
    a) What do the graphs tell you? How is the quality of the reads? What kind of changes do you observe after trimming? Explain the story of each graph. You can put raw data plots and trimmed data plots side-by-side to do this. Consider what kind of data is being used while interpreting the plots.
        - *Sequence Duplication Level* plot can be hard to interpret. You can skip it.
    b) Draw a diagram showing your inputs and outputs (you can use pen & paper and take a picture of it).
2) "**Processing multiple datasets**": Add screenshots for `MultiQC` and `cutadapt` results and comment on them.
    a) Same here, but this time interpret the quality by using MultiQC plots. You don't have to compare pre-trimming quality with post-trimming quality. Only comment on the initial quality of the fastq files.
        - Your comments can be more brief than the first question.
    b) Draw a diagram showing your inputs and outputs (you can use pen & paper and take a picture of it).