BT4110: Computational Biology Laboratory
Transcriptome Data Analysis Assignment

Download the gene expression dataset from Moodle

This dataset includes expression values for 41287 genes measured across 4 different human tissues with two replicates for each tissue.

1. **Principal Component Analysis (PCA):**
   - Apply Principal Component Analysis (PCA) to this dataset to reduce the dimensionality of the gene expression data.
   - Ensure the data is scaled so that each gene (row) has a mean of zero before performing PCA.
   - Provide the percentage of variance explained by the 8 principal components.
   - Plot the first two principal components (PC1 and PC2) and observe any clustering patterns between tissue groups using PCA Score plot.

2. **Differential Gene Expression Analysis (DESeq2):**

   - Using DESeq2, determine which genes are differentially expressed (upregulated or downregulated) in human tissues when comparing ovary tissue against other tissue types.
   - Summarize the results by reporting the number of differentially expressed genes and include a volcano plot to illustrate the findings

3. **Functional Enrichment Analysis (KEGG Pathways):**

   - Perform a functional enrichment analysis using the KEGG pathway database.
   - Present the significant pathways associated with the upregulated and downregulated genes, highlighting any biological processes of interest.

Please summarize your findings in a detailed report, addressing each question with corresponding figures and explanations. Be sure to clearly state any assumptions or decisions made during the analysis.