**Transcriptomic Analysis Assignment: Utilizing GEO Database for Differential Gene Expression and Pathway Analysis**

Maulana A. Empitu, dr., MSc, Ph.D

**Objective:**  
This assignment is designed to develop your expertise in formulating a research question, mining the Gene Expression Omnibus (GEO) database for relevant datasets, and performing bioinformatic analyses. You will conduct differential gene expression analysis, generate volcano plots and heatmaps, and perform Gene Set Enrichment Analysis (GSEA) to identify key pathways and biological functions. This exercise aims to integrate clinical insights with molecular data analysis, promoting a translational research approach.

**Instructions:**

1. **Define the Research Question:**
   * Develop a clear and focused research question relevant to a biomedical context (e.g., "How does gene expression differ in [specific disease] compared to healthy controls?").
2. **Identify Key Words for GEO Database Search:**
   * Compile a list of relevant keywords and phrases that capture the critical elements of your research question.
3. **Select an Appropriate GEO Dataset:**
   * Use the GEO database to identify a dataset that best addresses your research question, taking into account factors such as sample size, experimental design, tissue or cell type, and overall relevance to your hypothesis.
4. **Basic Data Analysis and Volcano Plot Generation:**
   * Preprocess and analyze the selected dataset to compare the groups of interest.
   * Generate a volcano plot to visualize significantly upregulated and downregulated genes. Document the software tools and parameters used in your analysis.
5. **Identification of Differentially Expressed Genes (DEGs):**
   * List and interpret the top differentially expressed genes based on predetermined thresholds (e.g., p-value < 0.05, fold change criteria).
   * Discuss the potential clinical or biological relevance of these DEGs.
6. **Conduct Gene Set Enrichment Analysis (GSEA) using EnrichR:**
   * Run GSEA to identify significantly enriched pathways and gene functions.
   * Provide a summary of your findings, highlighting the most relevant biological processes and pathways that correlate with your research question.
   * Discuss the implications of these enriched pathways in the context of disease mechanisms or therapeutic targets.
7. **Generation of a Heatmap:**
   * Create a heatmap to depict the expression patterns of the identified key genes across all samples.