**Gene Expression Analysis in Autism Spectrum Disorder (ASD)**

**📌 Project Overview**

This project was developed **for learning purposes** with the objective of exploring gene expression data analysis techniques in Autism Spectrum Disorder (ASD) studies. Using publicly available data from GEO (GSE102741), we conducted a comprehensive exploratory analysis, from preprocessing to the identification of differentially expressed genes.

**🧠 Scientific Context**

The study analyzed transcriptome data from:

* **13 samples from individuals diagnosed with ASD**
* **39 samples from healthy controls**

Focusing on:

* Differential expression patterns
* Pathways related to synaptic plasticity
* Potential neuroinflammation biomarkers

**🛠️ Methodology**

**1. Data Preprocessing**

* Loading and cleaning expression data (log2-RPKM matrix)
* Metadata annotation (condition: autism/control)
* Data quality and integrity verification

**2. Exploratory Analysis**

* PCA for clustering visualization
* Heatmaps of the most variable genes
* Expression distribution analysis

**3. Differential Analysis**

* Statistical tests (t-test and Mann-Whitney)
* Multiple comparison correction (FDR)
* Log2 Fold Change calculation
* Visualization with Volcano Plots

**4. Functional Analysis**

* Annotation of significant genes
* Expression visualization for candidate genes

**💡 Key Findings**

We identified **1 gene with significant differential expression**:

* **SNORA74A (small nucleolar RNA, H/ACA box 74A)**
  + **Log2FC: -0.738** (reduced in ASD group)
  + **Adjusted p-value: 0.018**

**📚 Educational Value**

This project served as:

* A practical introduction to bioinformatics
* A learning experience in omics data analysis
* Development of skills in:
  + **Python for data science**
  + **Applied statistics**
  + **Scientific visualization**
  + **Biological interpretation of results**