**Project Title:** Simulation of Correlated Single-Cell RNA-Seq Data and Evaluation of Co-Expression eQTL Mapping Methods

**Objective:** This project aims to evaluate two existing co-expression expression quantitative trait locus (co-eQTL) mapping methods using simulated data. Specifically, I will assess whether current approaches accurately detect co-eQTLs and identify their limitations in capturing co-expression signals influenced by genetic variation.

**Background & Motivation:**

Genes within the same biological pathway often exhibit transcriptional co-regulation, sharing common cis-regulatory elements. Single-cell RNA sequencing (scRNA-seq) enables the construction of gene co-expression networks, offering insights into regulatory mechanisms underlying genetic variation.

Two primary methods have been developed for co-eQTL mapping: one leveraging Spearman correlation (Li et al., 2023) and another using a bootstrap-based approach (Kim et al., 2024). However, their performance in detecting co-eQTLs remains insufficiently explored.

Genome-wide association studies (GWAS) have identified numerous genetic variants associated with disease, yet the regulatory mechanisms remain largely unresolved. A robust statistical framework for co-eQTL analysis is essential for advancing our understanding of gene regulation and its implications in complex traits and diseases.

**Approach:**

1. Simulation Framework

Develop a hierarchical model to simulate the expression of two correlated genes, where the correlation structure is influenced by genotype.

Control key parameters such as gene expression variability, correlation strength, and genotype-dependent effects.

2. Evaluation of Co-eQTL Mapping Methods

Implement and modify existing co-eQTL mapping methods based on the authors’ provided code.

Compare the accuracy and power of each method under various simulation conditions.

Throughout this project, I will utilize the simulation framework and high-performance computing (HPC) techniques learned in class to efficiently simulate data and analyze large-scale results.

**References:**

Kim, M. C., Gate, R., Lee, D. S., Tolopko, A., Lu, A., Gordon, E., Shifrut, E., Garcia-Nieto, P. E., Marson, A., Ntranos, V., & Ye, C. J. (2024). Method of moments framework for differential expression analysis of single-cell RNA sequencing data. *Cell*, *187*(22), 6393-6410.e6316. <https://doi.org/10.1016/j.cell.2024.09.044>

Li, S., Schmid, K. T., de Vries, D. H., Korshevniuk, M., Losert, C., Oelen, R., van Blokland, I. V., Groot, H. E., Swertz, M. A., van der Harst, P., Westra, H.-J., van der Wijst, M. G. P., Heinig, M., Franke, L., & Bios Consortium, s.-e. C. (2023). Identification of genetic variants that impact gene co-expression relationships using large-scale single-cell data. *Genome Biology*, *24*(1), 80. <https://doi.org/10.1186/s13059-023-02897-x>