Study Questions for "Tissue-specific enhancer-gene maps from multimodal single-cell data identify causal disease alleles" by Sakaue et al.

NOTE: Answering some of these questions will require looking at supplementary figures.

1. Why did the authors choose to use Poisson regression as the initial statistical model for associating gene expression with chromatin accessibility in SCENT? What advantages does this approach offer compared to other regression methods for single-cell data?

2. The paper compares SCENT to other methods like ArchR and Signac. How do the statistical approaches in SCENT differ from these other methods, particularly in terms of regression modeling, significance testing, and controlling for confounders? What are the relative strengths and limitations of each approach?

3. Explain the nonparametric bootstrapping procedure that the SCENT method implements. What is the null hypothesis? What is actually being resampled with replacement?

4. In evaluating the overlap of GWAS and eQTL variants with enhancer-gene (E-G) links prioritized by SCENT and other E-G prediction methods, how was the enrichment statistic calculated? What are some factors accounted for and not accounted for by the method?

5. While SCENT shows improved precision, it lags in recall in comparison to other E-G prediction methods. What do precision and recall mean in this context? Are there applications when it could be desirable to prioritize recall over precision in E-G mapping?

6. How does the integration of ATAC-seq and RNA-seq data in multiome analysis enhance *statistical power* compared to single-omics approaches? Consider aspects such as feature correlation, effect size, sample size, and biological versus technical variation. How does integrative analysis enhance *biological inference* regarding biological relationships?

7. **Challenge Question:** The authors claim that the nature of multi-ome data currently available is a limiting factor, and that if more cells were available, then discovery using the SCENT method would surpass discovery for competing methods. What technique do the authors use to try to prove this? Are you convinced by this argument?