Initial Paper Outline

Abstract:
Genomic mapping methods such as Illumina Dye Sequencing and Sequencing by Ligation have demonstrated utility in a wide variety of scientific research and practical applications. These methods for genomic sequencing have recently been focused to single-cell resolution(s) in areas where population-averaged (bulk) data would not be useful (e.g. studying microbiomes and targeting cancer cells). The collection of single-cell sequencing data involves error and protocol dependencies that need to be addressed before an analysis can be conducted. Single-cell data is high-dimensional and clustered, which can make analysis difficult. In this paper, we demonstrate a practical application of quality control, and we also compare models generated for the filtered data using three methods: Fixed Effect Models, Mixed Effects Models, and Generalized Estimating Equations.

Introduction:
- What is single-cell data?
  - Data attributes: dimensionality, clustering, and sampling
  - scRNA-seq: Why is it important? What type of outcome? How are samples acquired?
  - Case Study: Lupus Paper
- Quality Control Principles
  - Parameters of interest-why they matter
  - How to find good QC thresholds
  - Case Study: Lupus Paper
- Modeling Principles
  - Fixed Effect Models
    - Definition/nomenclature
    - Motivation (why use this type of model?)
    - Estimation methods, Structure, assumptions, transformations, parameter(s) of interest (why)
  - Mixed Effects Models
    - Definition/nomenclature
    - Motivation (why use this type of model?)
    - Structure, assumptions, transformations, parameter(s) of interest (why)
    - What's different?
  - GEE
    - Definition/nomenclature
    - Motivation (why use this type of model?)
    - Structure, assumptions, transformations, parameter(s) of interest (why)
    - What's different?
- Modeling Results
  - Multiple versions of data/Multiple outcome-predictor pairs.
- Discussion
  - What is the interpretation of the parameters of interest (differential)

- What are some applications of this analysis?
- What are the current limitations of this analysis?
- What should be considered in the future?