Introduction

Traditional methods of sequencing the human transcriptome involve analyzing the combined genetic material of thousands or even millions of cells. These so called "bulk" techniques provide information about the average gene expression across the cell sample but often fail to capture the underlying variability in expression profiles within the sample of cells [1].

Conversely, single-cell RNA sequencing (scRNA-seq) obtains measurements of transcriptomic information specific to individual cells. Hundreds or even thousands of RNA-sequencing profile measurements, each specific to a single-cell, can be used to estimate expression variability across the cells within the sample. This feature of single-cell data analysis is suited for research applications that seek to identify rare cellular subpopulations or characterize expressions that are differentially expressed across conditions [2]. Additionally, technological developments have made generating single-cell data more cost effective, and easier to obtain on multiple sample-sources, most notably on multiple individuals.

The utility of single-cell data, and the feasibility of single-cell data measurements across multiple subjects motivates a need to compare methods that can adequately model single-cell data while accounting for the correlation of repeated measures within subjects (many single-cell observations within each subject).

Here, I compare five methods for modeling scRNA-seq expression profiles that account for within-subject correlation: linear modeling (LM), linear modeling with subjects as fixed effects (LM-FE), linear mixed effects models with subjects only as random intercepts (LMM-RI) or as both random intercepts and random slopes (LMM-RS), and generalized estimating equations (GEE). I first present the overall framework for each method. Then I compare the results for each model using single-cell data from a study of 27 Lupus Nephritis cases.

References

1. Macaulay IC, Voet T (2014) Single cell genomics: Advances and future perspectives. *PLoS genetics* 10: e1004126.

2. Bacher R, Kendziorski C (2016) Design and computational analysis of single-cell rna-sequencing experiments. *Genome biology* 17: 63.