Discussion

- This paper has been composed to compare modeling strategies for detecting subject level associations in single-cell RNA sequencing data sets gathered over multiple subjects.

- In this paper we compared the estimates from five statistical models applied to single-cell RNA sequencing data from 15 subjects from a Lupus Nephritis study.

- We compared estimate values of the fixed effect slope parameter, standard error, and test statistic from each model method. These comparisons allowed us to find models that produced similar estimates and discover models for which there is a large difference. I also performed several nested model comparisons to determine the suitability of each variable under consideration.

- main results

  - As a class of modeling techniques, LM and GEE have population average interpretations for fixed effect slope parameter estimates that tend to be similar to each other. Similarly, as a class of modeling techniques, LM-FE, LMM-RI, and LMM-RS have subject specific fixed effect slope parameter estimate interpretations that tend to be similar to each other. Inter-class comparisons of estimates tend to be more different than Intra-class comparisons.

  - SE of LMM-RS model is higher than all other modeling methods. The LMM-RS method differs from all other methods with the incorperation of subject-specific variability relationships between predictor and response. Incorperation of this effect is reducing the influence (significance) of the fixed-effect slope by instead incorperating the effect into subject-specific variational response related to the covariate magnitude.

  - SE of LMM-RI model is as low or lower than all other modeling methods. The LMM-RI model differs from all other models either by lacking the incorperation of subject-specific variability relationships between predictor and response, or

incorperating subject-specific variability information independent of predictor. In either case, further emphasis is being placed on the fixed-effect slope parameter due to a need to estimate subject-specific variability distributions with a zero-mean.

- The above conclusions were also supported by a similar analysis performed on test statistics

- Nested model comparisons indicated that inclusion of subject-specific terms was advisable at all levels (fixed and random, intercept and slope) with exception of the FBLN1~CD34 variable pairing random slope.

- **why it makes sense for GEE to be different from LMMs**
  - **population average vs conditioning. . .**
  - **but under the conditions of linearity and normality of error we can show that LMMs can also be *described* as marginalizing over the structure**

- Drawbacks
  - The results of this paper have been based on evidence obtained from two scRNA-seq variable paring models.

  - More variable pairings would be needed to establish further conclusions, and reinfornce current findings.
  - scRNA-seq data is heavily influenced by protocol dependencies and measurement inconsistencies. Quality control must be carefully considered and conducted prior to any analysis.

- Future directions
  - The evaluation of more variable pairings is the foremost objective outstanding in this analysis.

- Summary
  - scRNA-seq data sets with measurements taken over multiple individuals will continue to rise in prevalence

– we have presented an initial comparison of methods for detecting subject-level

associations in such data sets

References