Dear PLOS Editors,

We would like to submit a revised version of our manuscript, “Dispersion Estimation and Its Effect on Test Performance in RNA-seq Data Analysis: A Comparison of Methods”, to PLOS One. Please find the revised manuscript, the latex source file with tracked changes, figures, and supplementary code in the online submission.

Next generation sequencing technology provides a powerful tool for measuring gene expression (mRNA) levels in the form of RNA-seq data. In order to test for differentially expressed genes, a major goal in the RNA-seq data analysis, many researchers apply a negative binomial model that assigns a dispersion parameter to each gene. Estimation of these dispersions is important because underestimation leads to false discovery and overestimation lowers the rate of true detection.

We briefly review several popular negative binomial dispersion estimation methods and tests for differential expression in RNA-seq data. More importantly, we develop an RNA-seq data simulation method that draws key information from real datasets, and we use simulated pseudo-datasets to ascertain the relative strengths and weaknesses of the dispersion estimation methods according to multiple distinct standards of quality. For example, we select the methods that optimize the performance of tests for differentially expressed genes.

This work is supported in part by the National Science Foundation Grant IOS-1127017. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript. No additional external funding received for this study.

We feel that our paper is appropriate for PLOS One. Please have it refereed. Thank you very much for handling our submission. We look forward to hearing from you.

Best regards,

Will Landau and Peng Liu