Dear Editor

Enclosed is our submission entitled “Explicit Scale Simulation for analysis of RNA-sequencing with ALDEx2” which we believe is well-suited for publication in NAR Genomics and Bioinformatics. The work shows how a generalization of compositional data (CoDa) approaches that are used to analyze high throughput sequencing data can be applied to transcriptome and meta-transcriptome datasets to solve long-standing issues with their analysis.

The approach builds on work by Dr. Justin Silverman who found that normalizations in widespread use made implicit, but wrong, assumptions about the scale (total size) of the sampled environment. Furthermore, his group found that incorporating scale information could better control both Type 1 and 2 statistical errors. The centred log-ratio often used in CoDa methods was found to be useful and interpretable under this framework, and the method was incorporated into the ALDEx2 R package which builds a Bayesian model of the data during analysis.

While the original scale models were developed for 16S rRNA microbiome analysis, we found that they have useful properties in transcriptomes and the work presented here shows two new findings. The first major finding shows how the incorporation of scale by ALDEx2 alters the variance structure of the modelled data, and that this change in the variance structure accounts for the better error control. The second major finding is that incorporating scale assumptions into meta-transcriptome dataset analysis enhances our ability to analyze these problematic datasets. For this latter case we provide guidance on how to choose appropriate scale parameters and show that any reasonable scale estimate is better than none.

We believe that the new scale approaches will be as transformative for the analysis of sequencing data as was the introduction of CoDa approaches. Including scale estimates makes explicit a formerly implicit parameter, allows for much better error control and enables the analysis of formerly problematic datasets.

We look forward to your response.

Sincerely

Greg Gloor, on behalf of the authors