Visualization methods for RNA-sequencing data analysis

Summary:

It was initially claimed that RNA-seq produced unbiased data that did not require sophisticated normalization. However, numerous studies have since revealed that RNA-seq data is replete with biases and that accurate detection of differentially expressed genes is not a trivial task. In light of these complications, researchers should analyze RNA-seq data like they would any other biased multivariate data. The most effective approach to modern data analysis is to iterate between models and visuals, and to enhance the appropriateness of applied models based on feedback from visuals. Unfortunately, researchers do not commonly use models and visuals in a complimentary fashion when analyzing RNA-seq data. Here, we use real RNA-seq data to demonstrate that our visualization tools can detect normalization problems, DEG designation problems, and common errors in the RNA-seq analysis pipeline. We also show that our tools can identify genes of interest that cannot otherwise be obtained by any models. We emphasize that interactive graphics should be an indisposable component of RNA-seq analysis. In this paper, we do not propose that users radically change their approach to RNA-seq analysis. Instead, we propose that users simply modify their approach to RNA-seq analysis by assessing the sensibility of their models with multivariate graphical tools, namely parallel coordinate plots, scatterplot matrices, and replicate point plots.

Key words: Data visualization; Exploratory data analysis; Interactive graphics; RNA-sequencing; Statistical graphics