

A two parameter generalized Poisson model to improve the analysis of RNA-seq data

Keywords: RNA-Seq – Generalized Poisson – Gene Expression – Normalization – Differential Expression – Differential Splicing –

Abstract: Deep sequencing of RNAs (RNA-seq) has been a useful tool to characterize and quantify transcriptomes. However, there are significant challenges in the analysis of RNA-seq data, such as how to separate signals from sequencing bias and how to perform reasonable normalization. Here we focus on a fundamental question in RNA-seq analysis: the distribution of the position-level read counts. Specifically, we propose a two-parameter generalized Poisson (GP) model to the position-level read counts. We show that the GP model fits the data much better than the traditional Poisson model. Based on the GP model, we can better estimate gene or exon expression, perform a more reasonable normalization across different samples, and improve the identification of differentially expressed genes and the identification of differentially spliced exons. The usefulness of the GP model was demonstrated by applications to multiple RNA-seq data sets

Authors:

- Sudeep Srivastava sudeepsr@usc.edu United States University of Southern California
- Liang Chen liang.chen@usc.edu United States University of Southern California