

## **LFCseq: a nonparametric approach for differential expression analysis of RNA-seq data**

Bingqing Lin 1,3, Li-Feng Zhang 1 and Xin Chen 2

1 School of Biological Sciences, Nanyang Technological University, Singapore

2 School of Physical and Mathematical Sciences, Nanyang Technological University, Singapore

3 Institute of Statistical Science, Shenzhen University, China

### **Abstract**

#### **Background**

With the advances in high-throughput DNA sequencing technologies, RNA-seq has rapidly emerged as a powerful tool for the quantitative analysis of gene expression and transcript variant discovery. In comparative experiments, differential expression analysis is commonly performed on RNA-seq data to identify genes/features that are differentially expressed between biological conditions. Most existing statistical methods for differential expression analysis are parametric and assume either Poisson distribution or negative binomial distribution on gene read counts. However, violation of distributional assumptions or a poor estimation of parameters often leads to unreliable results.

#### **Results**

In this paper, we introduce a new nonparametric approach called LFCseq that uses log fold changes as a differential expression test statistic. To test each gene for differential expression, LFCseq estimates a null probability distribution of count changes from a selected set of genes with similar expression strength. In contrast, the nonparametric NOISEq approach relies on a null distribution estimated from all genes within an experimental condition regardless of their expression levels.

#### **Conclusion**

Through extensive simulation study and RNA-seq real data analysis, we demonstrate that the proposed approach could well rank the differentially expressed genes ahead of non-differentially expressed genes, thereby achieving a much improved overall performance for differential expression analysis.