

# Identifying the differentially expressed genes with RNA-Seq data

Hung-Ting Lu, Huey-Miin Hsueh

Department of Statistics, National Chengchi University, Taipei, Taiwan

Corresponding author: Huey-Miin Hsueh, email: hsueh@nccu.edu.tw

April 22, 2013

## **Abstract**

Recently, the RNA-Seq experiment is developed for a high-throughput DNA sequencing method for mapping and quantifying the transcriptomes. The gene expression level obtained from a RNA-Seq experiment is of the count data type and is often fitted by a Negative-Binomial distribution to account for over-dispersion. To find the differentially expressed genes with a binary phenotypic response, we aim to develop a statistical test for comparing the means of two Negative-Binomial distributions. A Wald test statistic based on the pseudo maximum likelihood estimators is proposed. A numerical study is performed for justification of the proposed test. The applicability of the proposed method is demonstrated via the data analysis of two real example data sets.

**Keywords:** Differentially expressed genes, negative-binomial, overdispersion, pseudo maximum likelihood estimator, RNA-Seq experiment.