



A new method for estimating the number of non-differentially expressed genes

J. Wu¹, C.Y. Liu², W.T. Chen², W.Y. Ma¹ and Y. Ding¹

¹Department of Mathematics and Computer Science, Nanjing Medical University, Nanjing, China

²Department of Biomedical Engineering, Nanjing Medical University, Nanjing, China

Corresponding author: Y. Ding

E-mail: yongdingcn@126.com

Genet. Mol. Res. 15 (1): gmr.15017402

Received August 8, 2015

Accepted November 26, 2015

Published March 28, 2016

DOI <http://dx.doi.org/10.4238/gmr.15017402>

ABSTRACT. Control of the false discovery rate is a statistical method that is widely used when identifying differentially expressed genes in high-throughput sequencing assays. It is often calculated using an adaptive linear step-up procedure in which the number of non-differentially expressed genes should be estimated accurately. In this paper, we discuss the estimation of this parameter and point out defects in the original estimation method. We also propose a new estimation method and provide the error estimation. We compared the estimation results from the two methods in a simulation study that produced a mean, standard deviation, range, and root mean square error. The results revealed that there was little difference in the mean between the two methods, but the standard deviation, range, and root mean square error obtained using the new method were much smaller than those produced by the original method, which indicates that the new method is more accurate and robust. Furthermore, we used real microarray data to verify the conclusion. Finally we provide a suggestion when analyzing differentially expressed genes using statistical methods.

Key words: Differentially expressed gene; False discovery rate; Multiple testing