A NEW SCORE FOR ROBUST DETECTION OF DIFFERENTIALLY EXPRESSED GENES

Yufei Xiao[*]
Department of Epidemiology and Biostatistics, University of Texas Health Science Center at San Antonio
San Antonio, TX 78229, USA
[*]*Email: xiaoy4@uthscsa.edu*

Xiaowu Wu and Steven E. Wolf
Department of Surgery, University of Texas Health Science Center at San Antonio
San Antonio, TX 78229, USA
Email: wux2, wolfs@uthscsa.edu

Yidong Chen
Department of Epidemiology and Biostatistics, University of Texas Health Science Center at San Antonio
San Antonio, TX 78229, USA
Email: cheny8@ uthscsa.edu

We have developed a new score, π(pi)-value, for detecting differentially expressed (DE) genes. Identifying DE genes from microarray data usually entails the use of p-value obtained from t-test or its variants. A statistical challenge arises when genes with biologically significant expression changes are undetected by t-test because of the high variation of measured mRNA levels across samples, while genes with biologically non-significant changes are tested falsely as DE if their mRNA levels have small variation. Existing methods usually tackle the latter problem by making adjustment to the small variance, but the former problem, which leads to missed biologically important genes, are seldom addressed. We address the two problems by introducing a gene score, π, defined as LFC×($-\log_{10}$ p), product of log-fold-change and log-transformed p-value. π-value is defined such that small p-value is penalized by low fold-change, while large p-value is made more favorable by high fold change. Genes with π larger than a given threshold will be considered DE, and the threshold can be determined from the empirical distribution of π. Simulation results show that when false discovery rate (FDR) is fixed, using π achieves higher sensitivity than p-value, while with fixed sensitivity, the former has lower FDR. Moreover, the modified ROC (receiver operating characteristic, using sensitivity as y-axis, FDR as x-axis) curve of π has the largest area under curve (AUC) in comparison with p-value and fold-change criteria. We conclude that π is a robust score for identifying DE genes.

---

[*] Corresponding author.