A Mixed Model Expression Index to Summarize Affymetrix GeneChip Probe Level Data

Shibing DENG, *Pfizer, Inc., USA*, E-mail: shibing.deng@pfizer.com
Tzu-Ming Chu, *SAS Institute, USA*
Russ Wolfinger, *SAS Institute, USA*

Key words: Genetics, Microarray, Mixed Model

Mathematical Subject Classification: 62-07, 62P10

Abstract: Microarray technology has been widely used in functional genomics research in recent years. It provides a powerful tool to study gene functionality, predict clinical outcomes and discover gene expression patterns. One of the most widely used microarray platforms is the single-channel Affymetrix GeneChip array. It uses multiple probes to detect the expression level of a gene. How to summarize the probe level data to obtain a gene expression index has been an active research area in the analysis of Affymetrix GeneChip gene expression data. Affymetrix's MAS5.0 Signal provides a direct summarization of probe level data for each individual chip. Li and Wong (2001) proposed a model-based approach to summarize the probe data. Using an ANOVA model, Irizarry et al. (2003) proposed a robust multi-array average (RMA) as a summarization index to measure the level of gene expression. We adopt a linear mixed model approach to analyze the probe level data by modeling the array effect as a random effect and propose a mixed model expression index (MMEI) as our new summarization index. Modeling array as a random effect provides us additional degree of freedoms to account for treatment effect and the treatment by probe interaction effect. To evaluate the performance of MMEI, it is compared with three commonly used expression indices: MAS5.0 Signal, Li-Wong's (2001) model-based expression index (MBEI) and RMA index with applications to two Affymetrix array datasets. We find the MMEI is very consistent with other expression indices but offers lower variability, thus higher reproducibility. By using a Receiver Operating Characteristic (ROC) curve, we find MMEI also has the highest sensitivity for a given specificity in detecting expressed genes among the compared indices. The performance of our MMEI is further assessed at Affycomp's website (http://affycomp.biostat.jhsph.edu). Among all submitted methods (more than 30 of them) it has the best performance in several assessment criteria by far, especially, in reducing variability and detecting true signals.

References

[1] Li C. and Wong W.H. (2001). Model-based Analysis of Oligonucleotide Arrays: Expression Index Computation and Outlier Detection, *Proc Natl Acad Sci USA*, 98, 31-36.

[2] Irizarry R.A., Hobbs B., Collin F., Beazer-Barclay Y.D., Antonellis K.J., Scherf U. and Speed T.P. (2003). Exploration, Normalization, and Summaries of High Density Oligonucleotide Array Probe Level Data, *Biostatistics*, 4, 249-264.