

# Gene Expression Analysis Refining System (GEARS) via Statistical Approach: A Preliminary Report\*

**Chen-Hsin Chen**<sup>1</sup>  
chchen@stat.sinica.edu.tw

**Henry Horng-Shing Lu**<sup>2</sup>  
hslu@stat.nctu.edu.tw

**Chen-Tuo Liao**<sup>3</sup>  
ctliao@ccms.ntu.edu.tw

**Chun-houh Chen**<sup>1</sup>  
cchen@stat.sinica.edu.tw

**Ueng-Cheng Yang**<sup>4</sup>  
yang@ym.edu.tw

**Yun-Shien Lee**<sup>5</sup>  
bojack@mail.cgu.edu.tw

<sup>1</sup> Institute of Statistical Science, Academia Sinica, Taipei 11529, Taiwan, R.O.C.

<sup>2</sup> Institute of Statistics, National Chiao Tung University, Hsinchu 30010, Taiwan, R.O.C.

<sup>3</sup> Department of Agronomy, National Taiwan University, Taipei 10617, Taiwan, R.O.C.

<sup>4</sup> Institute of Bioinformatics, National Yang-Ming University, Taipei 11221, Taiwan, R.O.C.

<sup>5</sup> Graduate Institute of Clinical Medicine Sciences, Chang Gung University, Tao-Yuan 33302, Taiwan, R.O.C.

**Keywords:** bioinformatics, database, functional genomics, microarray experiments

## 1 Introduction

Microarray technology has provided powerful and cost-effective tools for large-scale analysis of gene expression in functional genomics. Some major projects of the National Research Program for Genomic Medicine in Taiwan are making efforts to maneuver microarray experiments. We recently established the Statistical Core for Gene Expression Studies to implement the statistical consulting procedures and a software platform for design and analysis of microarray experiments. The statistical analysis system aims at integrating up-to-date statistical and information visualization tools into a software platform for design, quality assessment, and analysis for microarray experiments with adequate biomedical hypotheses. Experimental design, microarray image analysis, normalization for microarray gene expression data, identification of differentially expressed genes, and analyses for prediction of clinical phenotypes using microarray data will be efficiently stringed together as a coherent suite.

## 2 Method and Results

### 2.1 A Microarray Database Platform

The microarray database repository in our Bioinformatics Services Core Facility adopts the MIAMExpress and ArrayExpress of the European Bioinformatics Institute. Following guidelines of the Minimum Information About a Microarray Experiment (MIAME) [1], the MIAMExpress handles the data submissions from local experiments and the ArrayExpress performs data management for public microarray databases and external bioinformatic databases. This integrated database platform produces structured MAGE-ML files for the subsequent gene expression data analysis.

---

\*for the Statistical Core for Gene Expression Studies Bioinformatics Services Core Facility National Research Program for Genomic Medicine in Taiwan

## 2.2 Gene Expression Analysis Refining System (GEARS)

Based on the established microarray database platform, we are interested in developing a statistical platform GEARS to thoroughly conduct more extensive analysis of microarray experiment data. We construct web interface using the scripting language PHP (Hypertext Preprocessor) to integrate relevant statistical packages and information visualization environments designed by R [4] or C++ programming languages into the GEARS. Currently the GEARS has linked to some Bioconductor packages [3] in the R environment, in which novel statistical methods for normalization and identification of differentially expressed genes [6, 7] are included. The menu-driven GEARS, which is more structured and user friendlier than Rweb [8], facilitates the statistical analysis pipeline for microarray gene expression data.

Though we in this report focus on the two-dye cDNA microarrays, many statistical methods and implementations in the GEARS are also applied to the analysis of oligonucleotide chips.

## 3 Discussion

The preliminary version of GEARS enhances the quality of bioinformatic analysis for the microarray data via the integration of useful statistical tools. This prototype will be continuously improved by modifying, customizing and up-scaling statistical and information visualization techniques developed by our team's investigators, for example, as in [2, 5], and make them public accessible. An on-line user's guide for these software tools in the GEARS will be implemented. In addition, statistical consultation can be provided for more sophisticated microarray studies. Through a successful launch of the GEARS, in the near future, we will expand and transfer our statistical platform to promote research of functional genomics in the international academic community.

## References

- [1] Brazma A., Hingamp P., Quackenbush J., Sherlock G., Spellman P., Stoeckert C., Aach J., Ansorge W., Ball C.A., Causton H.C., Gaasterland T., Glenisson P., Holstege F.C.P., Kim I.F., Markowitz V., Matese J.C., Parkinson H., Robinson A., Sarkans U., Schulze-Kremer S., Stewart J., Taylor R., Vilo J., and Vingron M., Minimum information about a microarray experiment (MIAME) - toward standards for microarray data, *Nature Genetics*, 29:365–371, 2001.
- [2] Chen, C.H., Generalized association plots for information visualization: the applications of the convergence of iteratively formed correlation matrices, *Statistica Sinica*, 12:1–23, 2002.
- [3] Dudoit, S., Yang, Y.H., and Bolstad, B., Using R for the analysis of DNA microarray data, *R News*, 2:24–32, 2002.
- [4] Ihaka, R. and Gentleman, R., R: a language for data analysis and graphics, *Journal of Computational and Graphical Statistics*, 5:299–314, 1996.
- [5] Lu, H.H.-S. and Wu, H.-M., Visualization, screening, and classification of cell cycle-regulated genes in yeast by multidimensional scaling, nonlinear dimension reduction and wavelet transform, *Technical Report, Institute of Statistics, National Chiao Tung University*, 2003.
- [6] Wu, H., Kerr, M.K., Cui, X., and Churchill, G.A., MAANOVA: a software package for the analysis of spotted cDNA microarray experiments, *Springer*, in press.
- [7] Yang, Y.H., Dudoit, S., Luu, P., Lin, D.M., Peng, V., Ngai, J., and Speed, T.P., Normalization for cDNA microarray data: a robust composite method addressing single and multiple slide systematic variation, *Nucleic Acids Research*, 30:e15, 2002.
- [8] <http://www.math.montana.edu/Rweb/>