

# ASIAN on Personal Computer

**Shigeru Saito**<sup>1</sup>

sh.saito@infocom.co.jp

**Kousuke Goto**<sup>1</sup>

k.gotoh@infocom.co.jp

**Sachiyo Aburatani**<sup>2</sup>

sachiyo@ims.u-tokyo.ac.jp

**Katsuhisa Horimoto**<sup>2</sup>

khorimot@ims.u-tokyo.ac.jp

<sup>1</sup> Bioscience Department, INFOCOM CORPORATION, Mitsui Sumitomo Insurance Surugadai Annex Building, 3-11, Kanda-surugadai, Chiyoda-ku, Tokyo 101-0062, Japan

<sup>2</sup> Laboratory of Biostatistics, Institute of Medical Science, University of Tokyo, 4-6-1 Shirokane-dai, Minato-ku, Tokyo 108-8639, Japan

**Keywords:** cluster analysis, graphical Gaussian modeling, network inference, personal computer

## 1 Introduction

Recently, we have developed a system, named ASIAN (Automatic System for Inferring A Network), to infer a network from a large amount of data. The system was the first application of graphical Gaussian modeling (GGM) to gene expression profiles on a genomic scale, by the combination of hierarchical clustering [3,5]. The feasibility of the system was validated by its application to several data sets of gene expression profiles [1,4,6]. The system is also provided as a website (<http://eureka.ims.u-tokyo.ac.jp/asian>) for enhanced utilization of the ASIAN system [2]. In the ASIAN website, therefore, the user can analyze a wider range of data, especially data including redundant information.

Here, we introduce the ASIAN that is operated on the personal computer. The PC ASIAN runs on WINDOWS and LINUX operation systems. In addition to the various options in the ASIAN website, furthermore, the PC ASIAN is equipped with the newly developed functions.

## 2 Method and Results

### 2.1 Outline of ASIAN

The system is composed of four parts: the calculation of a correlation coefficient matrix for the raw data, the hierarchical clustering, the estimation of cluster boundaries, and the application of GGM to the clusters. In the GGM, a network is inferred by the calculation of a partial correlation coefficient matrix from the correlation coefficient matrix [7]. The partial correlation coefficient matrix can only be obtained if the correlation coefficient matrix is regular. Since the gene expression profiles on a genomic scale often include many profiles sharing similar expression patterns, the correlation coefficient matrix is not always regular. Therefore, the first three parts are prerequisite to analyze the redundant data, including many similar patterns of expression profiles, by the last part, the network inference by the GGM.

### 2.2 New Functions on PC ASIAN

One of the remarkable improvements in the PC ASIAN is that the PC version can handle an enormous size of data that requires large physical memory for dissimilarity matrix to perform hierarchical clustering. For this purpose, we add the Ward's method using the reciprocal nearest neighbor data structure to the set of hierarchical clustering procedures. Even in other parts, we further improve algorithms and data structures to operate stably the processing of data at high speed. Another high-light of improvement is the visual aspect of operation. As shown in Fig.1, a user can use various graphical tools such as a correlation viewer, a dendrogram viewer, and a network viewer for interactive analyses.

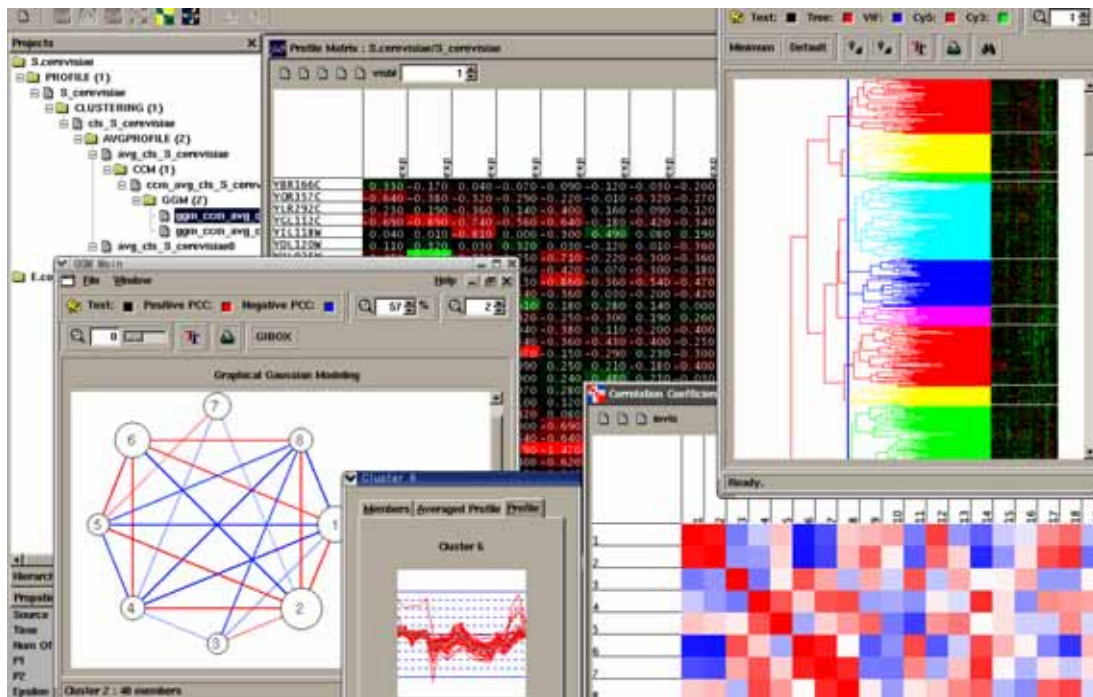


Figure 1: Snapshot of PC ASIAN

### 2.3 Computational Performance

The performance of PC ASIAN was tested for two practical sets of expression profile data; one is yeast data of 2,467 genes measured under 79 conditions, and another is human data of 43,033 genes measured under 33 conditions. The computational time is measured by a machine with the CPU of PentiumIV 2.4GHz and memory of 1GB under the LINUX system. By the test, the computational time was 24 sec and 1,584 sec (ca. 27 min), respectively. The present performance promises to analyze the data set containing tens of thousands genes within several tens of minutes.

## 3 Discussion

The present version of PC ASIAN shows high performance in computing data and in presenting visually the analyzed results. The practical analyses of some data will be demonstrated, and the further extension of ASIAN will be discussed. In near future, PC ASIAN Macintosh OSX and Solaris version will be released.

## References

- [1] Aburatani, S., Kuhara, S., Toh, H., and Horimoto, K., Deduction of a gene regulatory relationship framework from gene expression data by the application of graphical Gaussian modeling, *Signal Processing*, 83: 777-788, 2003.
- [2] Aburatani, S., Goto, K., Saito, S., Fumoto, M., Imaizumi, A., Sugaya, N., Murakami, H., Sato, M., Toh, H. and Horimoto, K., ASIAN: A web site for network inference, *Bioinformatics*, (in press).
- [3] Horimoto, K. and Toh, H., Statistical estimation of cluster boundaries in gene expression profile data, *Bioinformatics*, 17:1143-1151, 2001.
- [4] Imaizumi, A., Sugaya, N., Murakami, H., Sato, M., Aburatani, S. and Horimoto, K. Network reconstruction of genetic relationship in tryptophan metabolism from gene expression profiles, *Proceedings of the 8th World Multi-Conference on Systemics, Cybernetics and Informatics*, 7:4-11, 2004.
- [5] Toh, H. and Horimoto, K., Inference of a genetic network by a combined approach of cluster analysis and graphical Gaussian modeling, *Bioinformatics*, 18:287-297, 2002.
- [6] Toh, H. and Horimoto, K. System for automatically inferring a genetic network from expression profiles. *J. Biol. Phys.*, 28: 449-464, 2002.
- [7] Whittaker, J., *Graphical Models in Applied Multivariate Statistics*, John Wiley, 1990.