

KEGG/EXPRESSION: A Database for Browsing and Analysing Microarray Expression Data

Susumu Goto¹

goto@kuicr.kyoto-u.ac.jp

Tomomi Kamiya¹

tomomi@scl.kyoto-u.ac.jp

Shuichi Kawashima¹

shuichi@kuicr.kyoto-u.ac.jp

Satoshi Miyazaki²

miyazaki@fns.se.fujitsu.co.jp

Yoshinori Okuji¹

okuji@kuicr.kyoto-u.ac.jp

Youjiro Numata²

numata@fns.se.fujitsu.co.jp

Minoru Kanehisa¹

kanehisa@kuicr.kyoto-u.ac.jp

¹ Institute for Chemical Research, Kyoto University, Uji, Kyoto 611-0011, Japan

² Fujitsu Nagano System Engineering, 1403-3 Tsuruganabeyaden, Nagano 380-0813, Japan

Keywords: microarray, expression, database, KEGG, pathway

1 Introduction

The recent progress of DNA chip and microarray technologies produces a large amount of expression data with various experiments. A new database management system that integrates large-scale study of gene expression with other molecular biology databases and various analysis tools is required for management and analysis of the data [2]. In the Japanese molecular biology community, several laboratories started microarray projects for yeast, cyanobacteria, bacillus, and *Escherichia coli*, as well as mouse and human. Therefore a repository that enables integrated analysis of the data is highly required.

We have developed a prototype of KEGG/EXPRESSION system for storing publicly available yeast microarray data and analysed behavior of aminoacyl-tRNA synthetase using the data [1]. As a repository of the microarray data, it should provide database search facility and analysis tool accessed via WWW. We integrated KEGG/EXPRESSION into the DBGET/LinkDB system and developed new Java applets for browsing and analysing the data.

In this paper, we report recent development and improvement of the KEGG/EXPRESSION database.

2 Results and Discussion

Each entry of the KEGG/EXPRESSION in the DBGET/LinkDB system corresponds to each array and contains the information on entry name, accession ID, brief description of the experiment, conditions of the control and target arrays, experimentors, date and organism. This entry is in the HTML format and is a starting point to the browsing methods. The numerical data, such as intensity of each gene is stored in another format (tab delimited text files) and used when the browsing and analysis methods are invoked.

Fig. 1 shows the system overview of the KEGG/EXPRESSION database. The expression data from each experiment can be browsed as either a microarray image or a scatter plot. We provide three types of plots; 1) A plot for whole genes, 2) classified plots according to the KEGG functional classification, 3) classified plots according to the functional classification by the sequencing group of

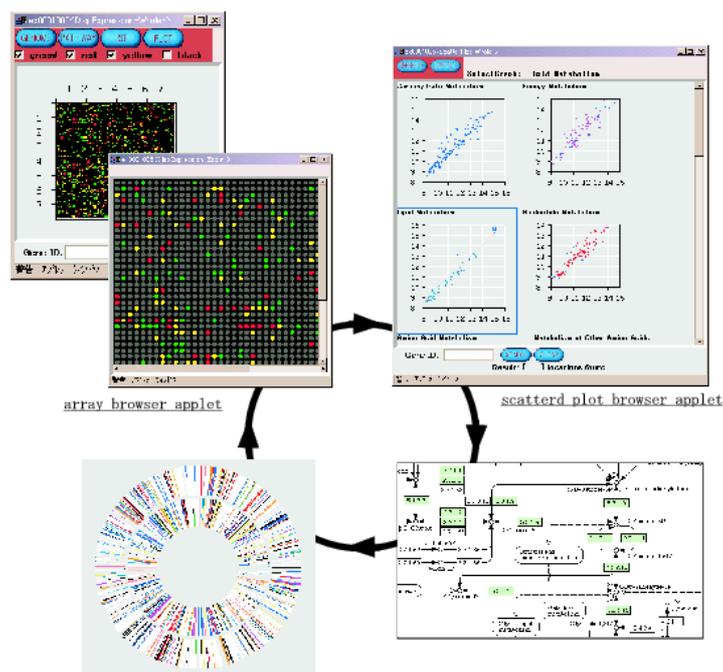


Figure 1: System overview of the KEGG/EXPRESSION.

the organism. By clicking 'Up' ('Down') button in the scatter plot window, the list of genes up (down) regulated in the experiment will be shown.

Another feature of this database is links to various analysis tools from the result of the browser and the list of genes. For example, one can map the genes to the KEGG/PATHWAY database or KEGG/GENES database (chromosomal view). It is also possible to search and reconstruct network using protein-protein binary relations, e.g. from yeast two hybrid system, stored in the BRITE database.

This system is already used by the Cyanobacteria DNA Chip Consortium.

Acknowledgements

This work was supported by grants from the Ministry of Education, Science, Sports and Culture in Japan, the Science and Technology Agency of Japan, and the Japan Society for the Promotion of Science. The computation time was provided by the Supercomputer Laboratory, Institute for Chemical Research, Kyoto University.

References

- [1] Nakao, M., Bono, H., Kawashima, S., Kamiya, T., Sato, K., Goto, S., and Kanehisa, M., Genome-scale gene expression analysis and pathway reconstruction in KEGG, *Genome Informatics*, 10:94–103, 1999.
- [2] Ermolaeva, O., Rastogi, M., Pruitt, K.D., Schuler, G.D., Bittner, M.L., Chen, Y., Simon, R., Meltzer, P., Trent J.M., and Boguski, M.S., Data management and analysis for gene expression arrays, *Nature Genetics*, 20:19–55, 1998.
- [3] Kanehisa, M. and Goto, S., KEGG: Kyoto Encyclopedia of Genes and Genomes, *Nucleic Acids Res.*, 28:27–30, 2000.