Development of an Efficient Data Processing Method for cDNA Microarray and Its Application to Tissue Expression Profiling

Koji Kadota [12] kadota@bi.a.u-tokyo.ac.jp
Rika Miki [13] rika@rtc.riken.go.jp
Yasushi Okazaki [14] okazaki@rtc.riken.go.jp
Kentaro Shimizu [2] shimizu@bi.a.u-tokyo.ac.jp
Yoshihide Hayashizaki [134] yosihide@rtc.riken.go.jp

[1] Genome Science Lab., Genomic Sciences Center, RIKEN, 3-1-1 Koyadai, Tsukuba-city, Ibaraki 305-0074, Japan
[2] Department of Biotechnology, The University of Tokyo, 1-1-1 Yayoi, Bunkyo-ku 113-8657, Tokyo, Japan
[3] Department of Medicine Tsukuba University, Tenno-dai, Tsukuba-city, Ibaraki 305-8577, Japan
[4] CREST, JST, Tsukuba-city, Ibaraki, Japan

1 Introduction

cDNA microarray is a powerful tool in that it can simultaneously monitor massive gene expression of cellular transcripts. Since the experimental procedure consists of multi-steps, there are several factors that can affect the reproducibility of the final results. To increase the accuracy or robustness of the data, it would be better to use the results that are reproducible. We have developed a filtering program to selectively extract genes whose expression pattern showed reasonable reproducibility. We applied this system for the analysis of tissue expression profiling and found that this filtering method is highly feasible for the data processing of cDNA microarray.

2 Materials and Methods

Materials: Mouse full-length RIKEN cDNA 20K sets were printed on the poly-L-lysine coated slide glasses using Stanford-type arrayer. Tissue specific expression profiles were produced using adult tissues and embryonic organ tissues (about 50 tissues).
Methods: Probes (tissue mRNA) were prepared by labeling Cy-3 dye. The cDNA microarray uses the dual dye system. We used Cy-5 labeled embryo17.5 days (whole body) as a reference. The algorithm of this filtering program consists of three steps: (1) omit the results which have flags (flags are built manually when the spot image does not fulfill a certain criteria), (2) eliminate spots whose signal intensity is less than mean(background signal) $+3\sigma$ in both Cy-3 and Cy-5, (3) eliminate spots that are located outside the best fit line (least-mean squares) $\pm 2\sigma$. The final results were subjected to further analysis when the correlation coefficient value of the repeated pair of experiments showed higher than 0.7.

3 Result and Discussion

We applied this program for the analysis of mouse full-length RIKEN cDNA 20K microarray and analyzed the expression profiles of normal adult and embryonic tissues. The results correlated very well with the expression profiles of the source (tissue) of each cDNA clone showing the good feasibility

of this program. Furthermore, our filtered data showed high reproducibility, i.e. most of the repeated pair of experiments were on the same terminal branch in our clustering results.

Several examples of inter-tissue clustering results were as follows: (1) tongue, heart, muscle (muscle-related) organs, (2) cortex, eyeball, brain, cerebellum (head-related) organs, (3) embryo11, 12, 13 head, embryo10, 11 whole body (embryo-related), (4) small intestine, stomach, appendix, colon(digestive system), were clustered very closely. Several examples in inter-gene clustering were as follows: (1) a gene (most homologous gene accession number: L36088), which showed more abundant expression pattern in lung than in other tissues on microarray, had the comparative feature in Northern blot analysis [1], (2) a gene (most homologous gene accession number: X95351), which was expressed in embryo10-13 on microarray, was also reported to be expressed during developmental stages using differential display analysis [2].

Knowing spatial and temporal expression profiles of a gene provides a good clue for the analysis of biological function of the gene. By using the newly developed filtering method, we could have reproducible and high quality results. We compared the results of the microarray with other conventional expression analysis, such as Northern blot. The results were highly comparative showing the feasibility of our system. These data would facilitate the discovery of genes and/or mechanisms which control the tissue or stage specific expression.

References

[1] Panchenko, M.P., Williams, M.C., Brody, J.S., and Yu, Q., Type I receptor serine-threonine kinase preferentially expressed in pulmonary blood vessels, *Am. J. Physiol.*, 270:547–558, 1996.

[2] Heyer, B.S., Warsowe, J., Solter, D., Knowles., B., and Ackerman, S.L., New member of the Snf1/AMPK kinase family, Melk, is expressed in the mouse egg and preimplantation embryo, *Mol. Reprod. Dev.*, 47(2):148–156, 1997.