

Metabolic reconstruction of mouse transcriptome using the RIKEN set of 18,816 mouse cDNA arrays

Hidemasa Bono

Laboratory for Genome Exploration
Research Group, RIKEN Genomic
Sciences Center (GSC), RIKEN
Yokohama Institute

1-7-22 Suehiro-cho, Tsurumi
Yokohama 230-0045, Japan

bono@gsc.riken.go.jp

Hideo Matsuda

Department of Informatics and
Mathematical Science, Graduate
School of Engineering Science,
Osaka University

1-3 Machikaneyama, Toyonaka
Osaka 560-8531, Japan

matsuda@ics.es.osaka-u.ac.jp

Takeya Kasukawa

Laboratory for Genome Exploration
Research Group, RIKEN Genomic
Sciences Center (GSC), RIKEN
Yokohama Institute

1-7-22 Suehiro-cho, Tsurumi
Yokohama 230-0045, Japan

kasukawa@gsc.riken.go.jp

ABSTRACT

In order to integrate gene function information in the genomic scale, we have studied methods to extract such information systematically that can characterize the function of genes more definitely. To develop such methods, data in the genomic scale are required. For the analyses, we focused on gene expression data. This is because it would be possible to get functional inference of genes that cannot be functionally annotated from existing sequence analysis. We analyzed the data set of the gene expression patterns in 49 adult and embryonic mouse tissues by using cDNA microarray with 18,816 mouse cDNA (denotes the RIKEN 19K set), one of the largest microarray dataset at that moment.

As the systematic reconstruction of mouse metabolic pathways with their gene expression patterns in 49 tissues was reported, the ways of the assignment of enzymes were refined to draw better results. One improvement is the use of Gene Ontology (GO). All enzymes are coded in EC numbers, but now all genes in eukaryotes are going to be associated with GO terms and GO IDs. Thus we assign GO IDs first and then convert GO IDs to EC numbers automatically in order to reconstruct metabolic pathways in the KEGG database. Another improvement is the quality of gene association between GO IDs and RIKEN cDNA cloneIDs. Two strategies were taken to assign GO IDs to cDNA clones on chips. One is the evidence based ID matching, and the other is the prediction utilizing SwissProt and InterPro. The use of GO ID allows us to facilitate the analyses in other pathways (signal transductions and other biological systems) and species (*Drosophilla* and others).

By reconstructing metabolic pathways with gene expression pattern, these gene association methods can be compared. Then the coverage of the RIKEN 19K set in comparison with human set in KEGG can also be calculated toward the complete set of metabolome array.

ADDITIONAL AUTHORS

Itoshi Nikaido (RIKEN GSC GER Group, email address: itoshi@gsc.riken.go.jp), Yoshihide Hayashizaki (RIKEN GSC GER Group leader, email address: yoshihide@gsc.riken.go.jp) and Yasushi Okazaki (RIKEN GSC GER Group, email address: okazaki@gsc.riken.go.jp).

ACKNOWLEDGEMENT

This study has been supported by Grant-in-Aid for Scientific Research on Priority Areas (C) "Genome Information Science" and Research Grant for the RIKEN Genome Exploration Research Project from the Ministry of Education, Culture, Sports, and Technology of the Japanese Government and ACT-JST (Research and Development for Applying Advanced Computational Science and Technology) of Japan Science and Technology Corporation (JST).

REFERENCES

- [1] Miki, R. et al. Delineating developmental and metabolic pathways in vivo by expression profiling using the RIKEN set of 18,816 full-length enriched mouse cDNA arrays. *Proc Natl Acad Sci USA*, **98**, 2199-2204 (2001).
- [2] READ (RIKEN Expression Array Database). <http://genome.gsc.riken.go.jp/READ/>.
- [3] The RIKEN Genome Exploration Research Group PhaseII Team and the FANTOM Consortium. Functional annotation of a full-length mouse cDNA collection. *Nature*, **409**, 685-690 (2001).
- [4] The Gene Ontology Consortium. Gene ontology: tool for the unification of biology. *Nat Genet*, **25**, 25-29 (2000).
- [5] Kanehisa, M. and Goto, S. KEGG: Kyoto encyclopedia of genes and genomes. *Nucleic Acids Res*, **28**, 27-30 (2000).