Analysis of Gene Expression Profiles Based on Clustering

Yoichi Takenaka and Hideo Matsuda

Genome Res. 2003 13: 1558
Access the most recent version at doi:10.1101/gr.1459703

Analysis of Gene Expression Profiles Based on Clustering

Yoichi Takenaka[1] and Hideo Matsuda

Laboratory for Genome Information Engineering, Department of Bioinformatics Engineering, Graduate School of Information Science and Technology, Osaka University, Osaka 565-0871, Japan

The purpose of analyzing clones using expression profiles is to annotate relationships among clones that are expressed synchronously rather than annotating the function of a single clone. READ/RINGENE in the FANTOM system shows clones whose expression profiles are similar to a query clone. However, it is hard to find appropriate query clones from tens of thousands of the riken clones.

We make clusters of the clones with high correlation coefficients (>0.8) by a QT-cluster (Quality Threshold cluster) algorithm. The QT-cluster algorithm ensures that all the pairs of clones within a generated cluster are more highly correlated than the input quality threshold. We select and improve the generated clusters by GO-terms for molecular function. First, we delete the clones that do not have GO-term annotation from the clusters. Second, if all of the clones in the cluster have similar GO-terms only, we discard it. Third, we add clones to the cluster that have a higher correlation coefficient than 0.7 to a clone in the cluster.

We obtained 19 clusters. One of the clusters includes the clones of four peptidases, two hormones (GO:0005179), and two defense/immune proteins (GO:0006955). Figure 1 shows the cluster. The peptidases include Cathepsin M and Cathepsin J. They have cysteine-type peptidase (GO:0004197) and cysteine-type endopeptidase (GO:0008234) annotations. The two hormones are prolactin-like protein and placental lactogen2 protein (Pl2). The two defense/immune proteins are CEACAM11 (Carcinoembryonic antigen-related cell adhesion molecule) and Psg21 (pregnancy-specific glycoprotein 21), and they have immune response (GO:0006955). CEACAM11 is a member of the immunoglobulin super family and Psg21 has antibody (GO:0003823) annotation.

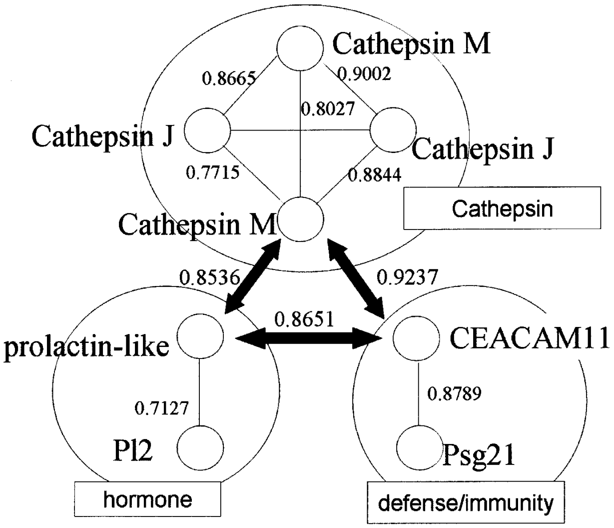

Figure 1 An obtained cluster with seven clones. Each small circle represents a clone, and edges represent the correlation coefficients between clones. The seven clones are categorized into three groups.

---

[1] **Corresponding author.**

E-MAIL takenaka@ics.es.osaka-u.ac.jp; FAX +81-6-6850-6602.

Article and publication are at http://www.genome.org/cgi/doi/10.1101/gr.1459703.

13:1558 ©2003 by Cold Spring Harbor Laboratory Press ISSN 1088-9051/03 $5.00; www.genome.org