

PathAct: a novel method for pathway analysis using gene expression profiles

Kaoru Mogushi¹
mogushi@bioinfo.tmd.ac.jp

Hiroshi Tanaka¹
tanaka@bioinfo.tmd.ac.jp

¹ Department of Bioinformatics, Medical Research Institute, Tokyo Medical and Dental University, 24F M&D Tower Bldg., 1-5-45 Yushima, Bunkyo-ku, Tokyo, Japan

Keywords: DNA microarray, Gene expression profile, Pathway analysis

Gene expression profiling by microarray analysis provides a huge amount of biological information and has been widely used in biological and clinical research. We developed PathAct¹, a novel method for pathway analysis to investigate the biological and clinical implications of the gene expression profiles. The advantage of PathAct in comparison with the conventional pathway analysis methods is that it can estimate pathway activity levels for individual patient quantitatively in the form of a pathway-by-sample matrix using the median polish algorithm². This matrix can be used for further analysis such as hierarchical clustering and other analysis methods.

To evaluate the feasibility of PathAct, comparison with frequently used gene-enrichment analysis methods was conducted using two public microarray datasets. The dataset #1 was that of breast cancer patients, and we investigated pathways associated with triple-negative breast cancer by PathAct, compared with those obtained by gene set enrichment analysis (GSEA)³. The dataset #2 was another breast cancer dataset with disease-free survival (DFS) of each patient. Contribution by each pathway to prognosis was investigated by our method as well as the Database for Annotation, Visualization and Integrated Discovery (DAVID)⁴ analysis. In the dataset #1, four out of the six pathways that satisfied $p < 0.05$ and $FDR < 0.30$ by GSEA were also included in those obtained by the PathAct method. For the dataset #2, two pathways ("Cell Cycle" and "DNA replication") out of four pathways by PathAct were commonly identified by DAVID analysis.

Thus, we confirmed a good degree of agreement among PathAct and conventional methods. Moreover, several applications of further statistical analyses such as hierarchical cluster analysis by pathway activity, correlation analysis and survival analysis between pathways were conducted. Therefore, PathAct is a promising tool for pathway-level investigation and interpretation of the comprehensive gene expression data.

- [1] Mogushi K, Tanaka H. PathAct: a novel method for pathway analysis using gene expression profiles. *Bioinformatics*. 2013;9(8):394-400.
- [2] Hoaglin DC, Mosteller F, Tukey JW, eds. *Understanding robust and exploratory data analysis*. Vol. 3. New York: Wiley, 1983.
- [3] Subramanian A, Tamayo P, Mootha VK, Mukherjee S, Ebert BL, Gillette MA, Paulovich A, Pomeroy SL, Golub TR, Lander ES, Mesirov JP. Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. *Proc Natl Acad Sci U S A*. 2005;102(43):15545-50.
- [4] Dennis G Jr, Sherman BT, Hosack DA, Yang J, Gao W, Lane HC, Lempicki RA. DAVID: Database for Annotation, Visualization, and Integrated Discovery. *Genome Biol*. 2003;4(5):P3.