Multiplicative Decomposition of Time- and Dose-Dependent Gene Expression Changes

Yukitaka Tani[1] tani3@is.titech.ac.jp
Takeshi Kamimura[1] kamimur1@is.titech.ac.jp
Takeshi Nagashima[2] nagasima@gsc.riken.jp
Kaori Ide[2] kaoide@gsc.riken.jp
Mariko Hatakeyama[2] marikoh@gsc.riken.jp
Hidetoshi Shimodaira[1] shimo@is.titech.ac.jp

[1] Department of Mathematical and Computing Sciences, Tokyo Institute of Technology, Ookayama, Meguro, Tokyo 152-8552, Japan
[2] Cellular Systems Biology Team, Computational and Experimental Systems Biology Group, RIKEN Genomic Sciences Center

Keywords: gene expression, breast cancer cells, two-way table, analysis of variance

1 Introduction

Time- and dose-dependent change in gene expression profiles of ligand-stimulated cancer cells is promising source of information to unwire a relationship between biological activities of the molecules and disease states [1]. In this study, we employed a multiplicative model [2] to decompose the time and dose effects, instead of the ordinary additive effect model in the analysis of variance (ANOVA). We tried to pick-up genes which are directly regulated by growth hormone stimulation. Human breast cancer cells were stimulated with four different dose levels of growth hormone (0.1, 0.5, 1 and 10 nM; $K = 4$), and expression values of genes ($I = 22277$) were obtained using Affymetrix human genome arrays for designated time points (5, 10, 15, 30, 45, 60 and 90 min; $J = 7$). Control was set as the one without growth hormone-treatment. This work is supported in part by Grant KAKENHI-17700276 from MEXT of Japan.

2 Methods

2.1 Multiplicative Decomposition Model

Let X_{ijk} be the expression value of gene i at time point j and dose level k for $i = 1, \ldots, I$, $j = 0, 1, \ldots, J$, and $k = 1, \ldots, K$. For each gene i, we consider a multiplicative model of the two-way table of gene expression changes

$$\log_2(X_{ijk}) - \log_2(X_{i0k}) = A_{ij} \times B_{ik} + E_{ijk},$$

where $(A_{i1}, \ldots, A_{iJ})$ is the time-course pattern and $(B_{i1}, \ldots, B_{iK})$ is the dose-response pattern of gene i. A_{ij} and B_{ik} are estimated by minimizing the squared sum of the error term E_{ijk} with the constraint that $\sum_{j=1}^{J} A_{ij}^2/J = 1$, or equivalently via the singular value decomposition. The signs of $A_{i1}, \ldots, A_{iJ}$ are carefully chosen so that they are approximately proportional to the averaged time-course pattern of gene i.

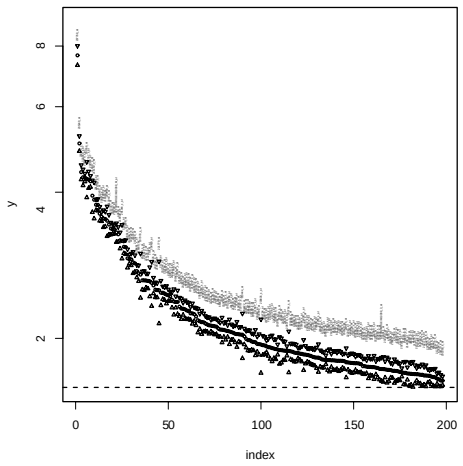

Figure 1: Score Y_i (circle) with confidence interval (triangles). 198 out of 22277 genes are selected.

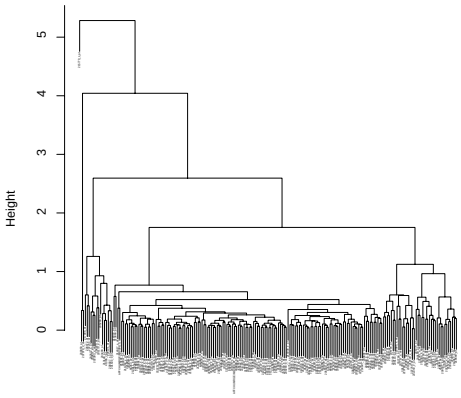

Figure 2: Hierarchical clustering of time-course pattern (A_{ij}) for the selected genes.

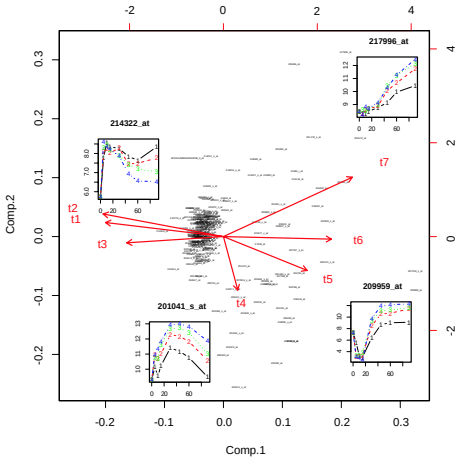

Figure 3: Biplot of time-course pattern (A_{ij}).

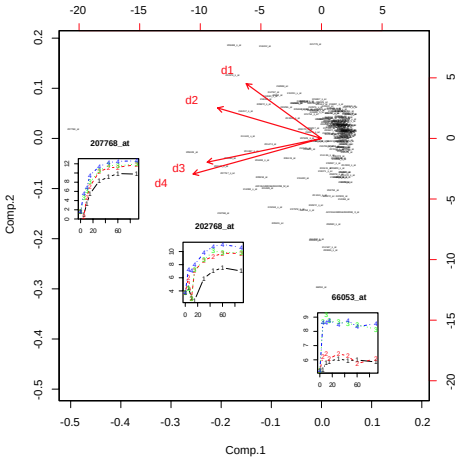

Figure 4: Biplot of dose-response pattern (B_{ik}).

2.2 Gene Selection and Classification

The averaged response of gene i at dose level k is B_{ik}. We employ the mean absolute value $Y_i = \sum_{k=1}^{K} |B_{ik}|/K$ as a score statistic of gene i for selection. The confidence interval of Y_i is computed by the bootstrap residual method. In Fig. 1, genes with Y_i being significantly larger than a specified threshold value $\log_2(3)$, i.e., three-fold change, are selected.

Genes are classified by A_{ij} for finding genes with similar time-course pattern (Fig. 2). The expression patterns are visualized using the biplot of the principal component analysis (PCA) in Fig. 3 (Arrows indicate expression changes at the time points and typical gene expression profiles are inserted). On the other hand, genes can also be classified by B_{ik} for finding genes with similar dose-response pattern in Fig. 4 (Arrows indicate expression changes at the hormone dose levels, whereas the time-course pattern is completely ignored).

References

[1] van Erk, M.J., Teuling, E., Staal, Y.C., Huybers, S., *et al.,* Time- and dose-dependent effects of curcumin on gene expression in human colon cancer cells, *Journal of Carcinogenesis*, 3:8, 2004.

[2] Mandel, J., A new analysis of variance model for non-additive data, *Technometrics*, 13:1-18, 1971.