

MULTIPLICATIVE DECOMPOSITION OF TIME- AND DOSE-DEPENDENT GENE EXPRESSION CHANGES

Yukitaka Tani¹

tani3@is.titech.ac.jp

Takeshi Kamimura¹

kamimur1@is.titech.ac.jp

Takeshi Nagashima²

nagasima@gsc.riken.jp

Kaori Ide²

kaoide@gsc.riken.jp

Mariko Hatakeyama²

marikoh@gsc.riken.jp

Hidetoshi Shimodaira¹

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

INTRODUCTION

Time- and dose-dependent change in gene expression profiles of ligand-stimulated cancer cells is promising source of information to unwind 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.

GENE SELECTION

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.

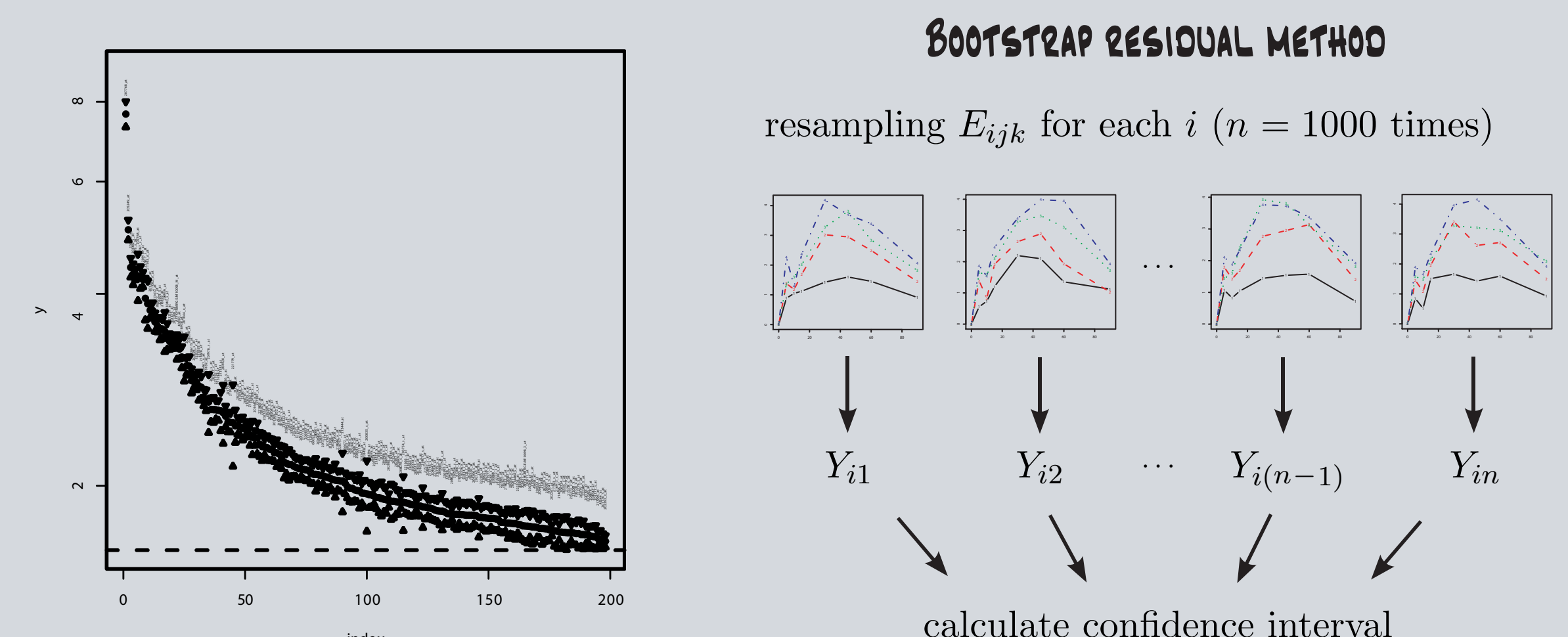


Figure 1: Score Y_i (circle) with confidence interval (triangles). 198 out of 22277 genes are selected.

MULTIPLICATIVE DECOMPOSITION MODEL

Let X_{ijk} be the expression value of gene i at time point j and dose level k for $i = 1, \dots, I$, $j = 0, 1, \dots, J$, and $k = 1, \dots, K$. 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}, \dots, A_{iJ}) is the time-course pattern and (B_{i1}, \dots, 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}, \dots, A_{iJ} are carefully chosen so that they are approximately proportional to the averaged time-course pattern of gene i .

We estimate A_{ij} and B_{ik} via single value decomposition.

$$\log_2(X_{ijk}) - \log_2(X_{i0k}) = U_i \times D_i \times V_i$$

$$U_i = \begin{pmatrix} u_{i11} & \dots & u_{i1K} \\ \vdots & \ddots & \vdots \\ u_{iJ1} & \dots & u_{iJK} \end{pmatrix}, \quad \sum_{j=1}^J u_{ij1}^2 = 1 (k = 1, \dots, K),$$

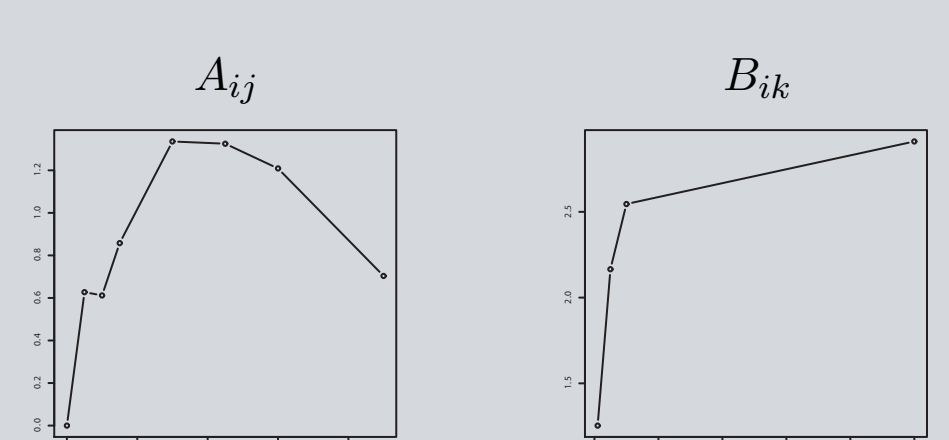
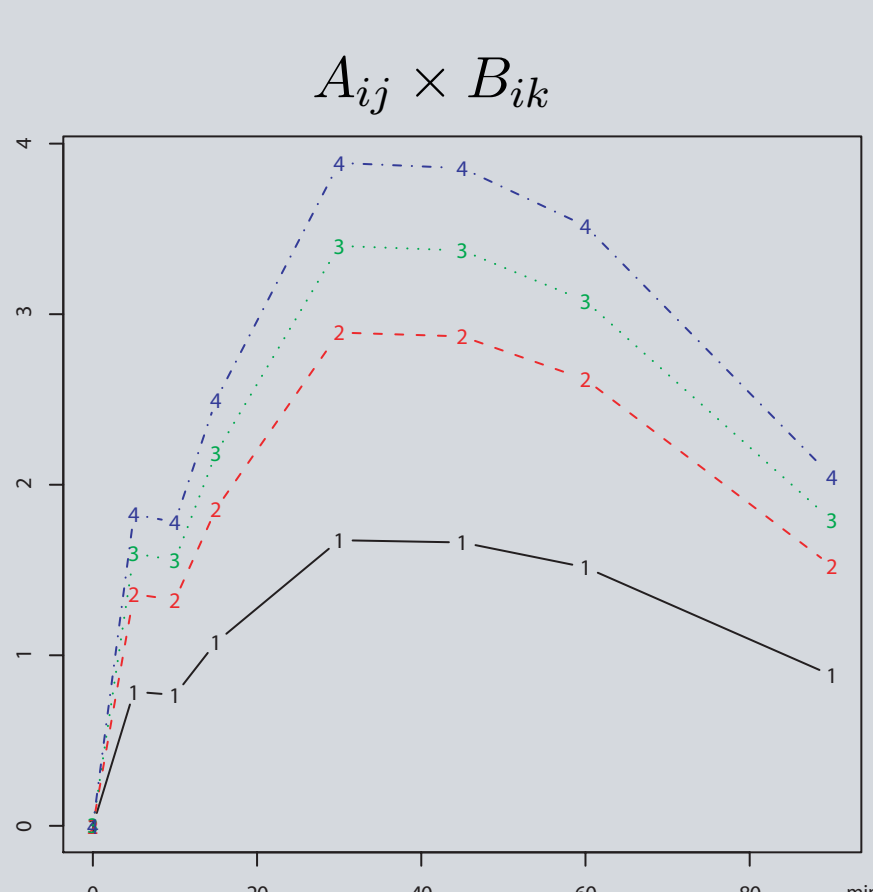
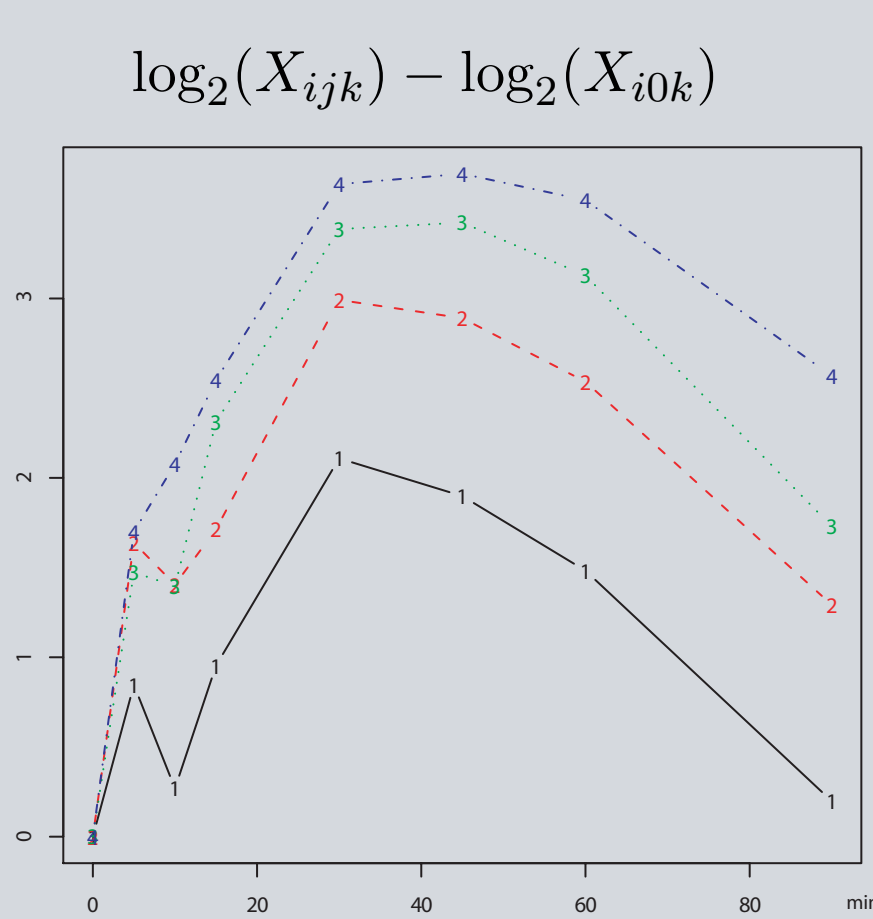
$$D_i = \begin{pmatrix} d_{i1} & & 0 \\ & \ddots & \\ 0 & & d_{iK} \end{pmatrix}, \quad d_{i1} \geq \dots \geq d_{iK} \geq 0,$$

$$V_i = \begin{pmatrix} v_{i11} & \dots & v_{i1K} \\ \vdots & \ddots & \vdots \\ v_{iK1} & \dots & v_{iKK} \end{pmatrix}$$

We use only first component and estimate that $\sum_{j=1}^J A_{ij}^2/J = 1$

$$A_{ij} = u_{ij1} \sqrt{J} \quad (\text{time-course pattern})$$

$$B_{ik} = d_{i1} v_{i1k} / \sqrt{J} \quad (\text{dose-response pattern})$$



CLASSIFICATION

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).

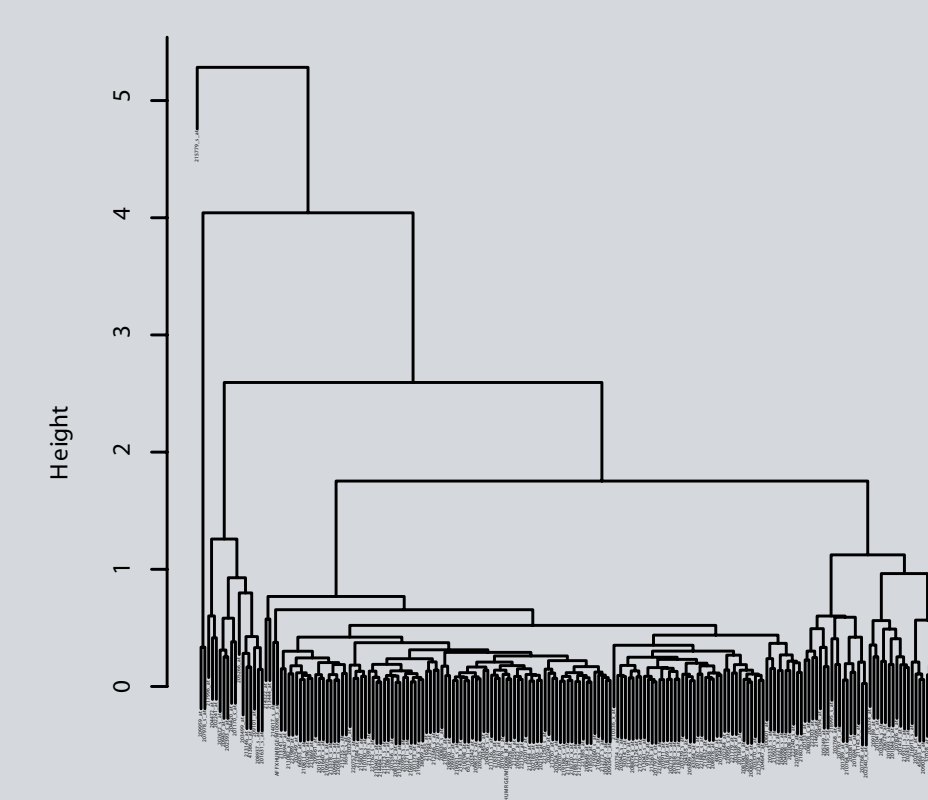


Figure 2: Hierarchical clustering of time-course pattern (A_{ij}) for the selected genes.

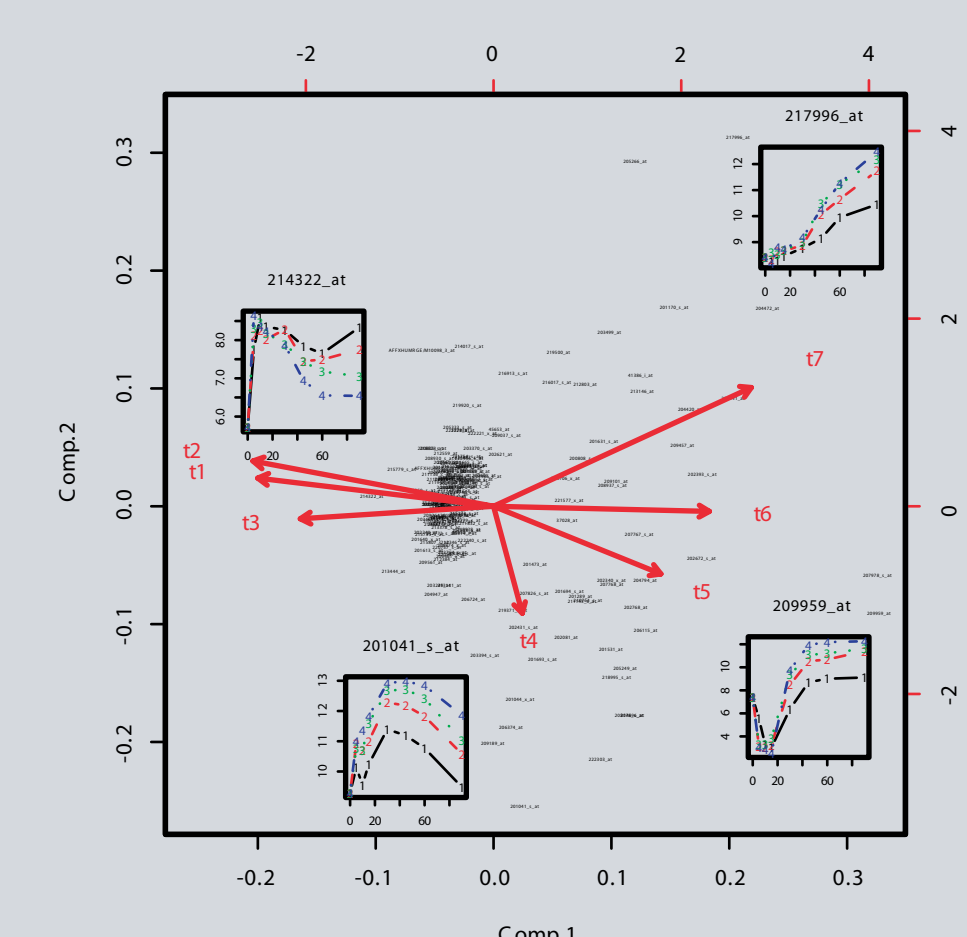


Figure 3: Biplot of time-course pattern (A_{ij}).

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.