

Improvement of Chemosensitivity Prediction by Transcriptional Profiling in Hepatoma Cells

Yujin Hoshida¹

hoshiday-int@h.u-tokyo.ac.jp

Masaru Moriyama¹

moriyamam-int@h.u-tokyo.ac.jp

Motoyuki Otsuka^{1,2}

otsukam-2im@h.u-tokyo.ac.jp

Naoya Kato¹

katon-2im@h.u-tokyo.ac.jp

Yasushi Shiratori¹

Shiratoriy-2im@h.u-tokyo.ac.jp

Naohiko Seki²

nseki@hri.co.jp

Masao Omata¹

omatam-2im@h.u-tokyo.ac.jp

¹ Department of Gastroenterology, Graduate School of Medicine, University of Tokyo, 7-3-1 Hongo, Bunkyo-ku, Tokyo 113-8655, Japan

² Helix Research Institute, 1532-3 Yana, Kisarazu-shi, Chiba 292-0812, Japan

Keywords: gene expression, gene shaving, leave-one-out cross validation, relevance network

1 Introduction

Only limited part of the patients with hepatoma benefits from curative therapy (e.g. surgical resection) owing to its multifocal nature, and chemotherapy is expected to improve the prognosis of such patients. However only limited cases show favorable response to chemotherapy, and we could not have distinguished them until now. In this study, we evaluated a methodology for robust pre-treatment chemosensitivity prediction in hepatoma using gene expression profiles in view of clinical application. Furthermore, we aimed to discover unknown relationship between gene expression and chemosensitivity.

2 Method and Results

In 8 human hepatoma cell lines (HepG2, Hep3B, HLE, HLF, Huh6, Huh7, PLC/PRF/5, and SK-Hep1), gene expression profiles were evaluated using cDNA microarray including 2,244 named genes. The 50% growth inhibitory concentration (GI50) of 9 anti-cancer drugs (5-FU, CDDP, CBDCA, ADM, EpiADM, MMC, ACNU, CPA and MIT) were measured by MTT assay and designated as chemosensitivity. Cell lines were divided into drug-sensitive and insensitive groups according to GI50 for each drug.

To predict chemosensitivity of hepatoma, genes differently expressed between sensitive and insensitive groups were selected using several algorithms: permutation t-test [5], categorization method [4], discrimination score [3], TNoM-Infoscore [1], and permutation rank sum test (U-test) that we developed. Predictive ability of each method was evaluated using leave-one-out cross validation (LOOCV) in various P -level ($P=0.001, 0.01, 0.05, 0.10$) based on 1,000 ~ 10,000 random permutation of class labels. For class prediction (in-silico genotyping), we used “weighted votes” method [3]. In the permutation t-test, we also adopted an originally described classification method based on “compound covariate”. In permutation rank sum test, classification by simple rank was also performed. Sample labels were randomly permuted 1,000 times and 5% significance level of misclassification was calculated.

Gene expression data and GI50 data were combined and comprehensive pair-wise correlations, r^2 , were calculated as square of the Pearson correlation coefficient with original sign (+ or -). Relationships with higher correlation coefficients compared with those calculated from 100 random permutation were regarded as having significant association. These putative functional relationships between gene expression and chemosensitivity of tested drugs were visualized using relevance networks [2].

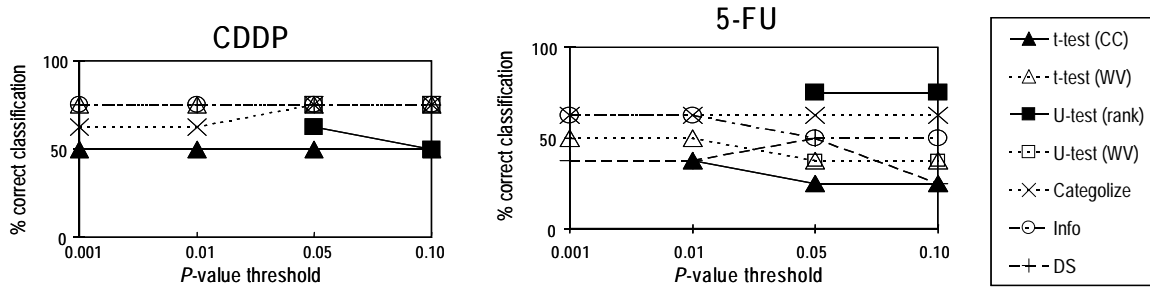


Figure 1: Result of LOOCV using 7 algorithms.

2.1 Results of LOOCV

Each gene shaving algorithm derived 10-500 genes. The core part of each selected gene set was almost identical among different algorithms. In all algorithms, selected gene set could clearly discriminate sensitive and insensitive groups in hierarchical clustering. LOOCV showed that predictive ability of each algorithm is drastically different according to tested data set (Figure 1). Only several algorithms with correct classification rate of 70% or above fulfilled 5% significance level of misclassification.

2.2 Relational Network of Genes and Drugs

There were 737,505 pair-wise relations of genes and drugs. Among them, 1,208 relations with r^2 greater than those generated from 100 random permutation of class labels (P -level of 0.0001) were determined, and relevance networks of potential biological relationships were constructed (Figure 2A-C). Three drugs (ADM, EpiADM and CPA) were included in these networks.

3 Discussions

In this study, we quantitatively evaluated the predictive ability of several gene shaving algorithms with varied P level. Our results showed that the predictive ability of each algorithm was largely affected by tested dataset, even though these datasets were derived from same experiment and had same sample size. This fact indicated that we should choose optimal algorithm specified to each dataset for robust prediction used in daily clinical diagnosis. Using gene expression profile, we can also perform efficient screening of genes which may lead to the clue to the mechanism of functional gene network and drug action. It may facilitate the development of personalized medicine.

References

- [1] Ben-Dor, A., Bruhn, L., Friedman, N., Nachman, I., Schummer, M., Yakhini, Z., Tissue classification with gene expression profiles, *J. Comput. Biol.*, 7:559–583, 2000.
- [2] Butte, A.J., Tamayo, P., Slonim, D., Golub, T.R., and Kohane, I.S., Discovering functional relationships between RNA expression and chemotherapeutic susceptibility using relevance networks, *Proc. Natl. Acad. Sci.*, 97:12182–12186, 2000.
- [3] Golub, T.R., Slonim, D.K., Tamayo, P., Huard, C., Gaasenbeek, M., Mesirov, J.P., Coller, H., Loh, M.L., Downing, J.R., Caligiuri, M.A., Bloomfield, C.D., and Lander, E.S., Molecular classification of cancer: class discovery and class prediction by gene expression monitoring, *Science*, 286:531–537, 1999.
- [4] Ono, K., Tanaka, T., Tsunoda, T., Kitahara, O., Kihara, C., Okamoto, A., Ochiai, K., Takagi, T., and Nakamura, Y., Identification by cDNA microarray of genes involved in ovarian carcinogenesis, *Cancer Res.*, 60:5007–5011, 2000.
- [5] Radmacher, M.D., McShane, L.M., and Simon, R., A paradigm for class prediction using gene expression profiles, *Technical report 001, National Cancer Institute*, July 1, 2001.