Bull Yamaguchi Med School 52(3-4):37-41, 2005

Prediction of Early Intrahepatic Recurrence of Hepatocellular Carcinoma by Molecular Profiling

Norio Iizuka

Department of Complementary Medicine, Yamaguchi University School of Medicine, 1-1-1, Minami-Kogushi, Ube, Yamaguchi 755-8505, Japan

(Received October 5, 2005)

Abstract Hepatocellular carcinoma (HCC) has a poor prognosis even after curative surgery, due to the high frequency of early intrahepatic recurrence (IHR). Conventional staging systems are almost completely inadequate, and need to be complemented by novel tools. To this end, many investigators have performed DNA microarray analysis on the basis of genome-wide information. However, so far, few studies have been able to truly account for the clinical efficacy of DNA microarray analysis in HCC. To address this dilemma, we used a supervised learning method with information of 7070 genes from 33 HCC samples, to construct a 12-gene predictor for early IHR, and then evaluated its predictive performance in 27 independent HCC samples. Our 12-gene predictor correctly predicted early IHR or non-recurrence in 25 (93%) of the 27 independent samples. This predictive value is higher than that of any other system currently available, suggesting that our system can serve as a robust tool for accurate prediction of early IHR of HCC. I emphasize in this mini-review that, although there are some technical issues to resolve prior to clinical use, DNA microarray technology can provide molecular basis to initiate "bench to bedside" translation, which cannot be easily reached with other methods.

Key words: hepatocellular carcinoma, intrahepatic recurrence, microarray, supervised learning

Introduction

In the post-genomic era, DNA microarray technology has opened a new avenue in genome-wide research in biomedical science. We can now use this state-of-the-art technology, a snapshot of comprehensive genetic alterations, to gain new insights into the molecular basis of cancer biology. The prevailing attitude toward DNA microarray technology is "Don't miss the bus", and thus some may use it to fish for novel genes that may be superior to *p53*, a hallmark for carcinogenesis, while others apply it to clinical practice, such as prediction of patients' outcomes or responses to drugs.

Our project team selected the latter task, and attempted to construct a robust predictive system for early intrahepatic recurrence (IHR) of hepatocellular carcinoma (HCC) within 1 year after surgery. The justifications were that (1) HCC is one of the most common malignancies with increasing incidence in many countries,[1] (2) early IHR is found in more than 30% of HCC patients even when curative surgery is performed, and thus limits the potential of surgery as a cure for HCC, (3) in order to avoid unnecessary overtreatment of patients who have already been cured by surgery alone and to provide them

with more personalized therapeutic options, we must accurately predict early IHR or non-recurrence, and (4) although many staging systems have been applied clinically to HCC patients, there are technological limitations for accurately predicting early IHR using these conventional systems. These problems have long frustrated both trained hepatologists and pathologists.

In this mini-review, I outline our recent work[2-4] as it pertains to signatures of early IHR of HCC. The several issues to resolve prior to the clinical use of microarray technology are also discussed. Finally, I would like to provide perspectives for future study.

Clinical features of IHR in HCC

Hepatocarcinogenesis is attributable to chronic inflammation by two hepatitis viruses (HBV and HCV) in most cases.[5] This unique etiology makes the recurrence of HCC more complicated than that of other cancers.[6] Intrahepatic metastasis and multicentric occurrence are the two possible means of HCC recurrence in the remnant liver after surgery.[1][2] Most of the former (*i.e.,* early IHR) appears to arise from the early spread of tumor cells via the portal venous system within 1 year after surgery, and is closely correlated with poor prognosis of HCC.[2] In contrast, the latter (*i.e.,* late IHR) is a *de novo* primary tumor and is thought to be affected by host factors rather than tumor factors.[2] It is, therefore, impossible to predict late IHR on the basis of the gene signature of the primary HCC. Given the clinicopathologic features of HCC, we focused our investigation on accurate prediction of early IHR.

Molecular portraits of early IHR in HCC

By applying a supervised learning theory (see glossary in Ref. 2) to DNA microarray data, we successfully developed a 12-gene predictor for early IHR of HCC.[2] However, the 12 genes used are involved in a wide range of biological processes, and their roles in early IHR remain to be clarified. To extend this finding, we recently performed a post-planned analysis of the DNA microarray datasets.[2][3] That study identified 46 IHR-related genes, including 10 immune response-related genes,[4] all of which were down-regulated in HCC with early IHR. Among the 10 genes, four (*HLA-DRA, HLA-DRB1, HLA-DG* and *HLA-DQA*) encoding MHC class II antigens were coordinately down-regulated in HCCs with early IHR compared with the levels in HCCs with non-recurrence. It is well known that MHC class II proteins are involved in the antigen presenting function of macrophages, including dendritic cells. Our multivariate analysis showed that low expression of HLA-DR protein in tumors is an independent risk factor for early IHR.[4] Collectively, these results strongly suggest that the immune system plays a central role in early IHR of HCC.

It is generally accepted that angiogenesis and apoptosis play important roles in HCC recurrence. If so, one must ask why angiogenesis- and apoptosis-related genes were not selected in our series. We searched for probes responsible for the relevant genes out of 7070 probe sets. Although numerous probes for angiogenesis- or apoptosis-related genes were present on the array[3] in our study, none of them survived the gene selection procedure consisting of the Fisher ratio and a random permutation test (data not shown). HCC patients have various backgrounds and divergent clinical courses,[1] resulting in much background heterogeneity between tumor samples. The Fisher ratio used in our series measures the difference between two means normalized by the average variance.[2] Namely, a large Fisher ratio indicates genetic change observed commonly in more global cases. In this way, genes with a large Fisher ratio can circumvent such inter-patient heterogeneity.[7-12] Therefore, the finding that no angiogenesis- and apoptosis-related genes survived our gene selection procedure suggests that either the two pathways have few roles in early IHR or that they are related to early IHR in a limited number of cases.

Significance of taking a combination into consideration in constructing a predictive system

Most predictive systems are constructed on the basis of order of magnitude of the

predictive power of individual genes. However, this theory requires numerous gene signatures (*i.e.,* hundreds of gene). Taking a combination into consideration is characteristic of our predictive system,[2] resulting in a decrease in the dimensionality. Namely, the number of genes (*i.e.,* 12 genes) used in our predictive system[2] was much smaller than that in many predictors previously reported (Table). Nevertheless, our system works well and its predictive value is higher than or nearly equal to those of other predictors (Table). Additionally, in the same cohort, our system predicted more accurately early IHR than conventional TNM staging and the support vector machine-based system.[2] These successes are largely attributable to our unique method, by which we searched for an optimal pattern by computing about 200,000,000 combinations. Notably, the 12 genes used in our predictive system had no overlap with those linked to patient backgrounds, such as, virus type, tumor differentiation grade, or coexisting liver diseases (*i.e.,* cirrhosis, hepatitis or normal liver) in our series.[7-10,12] This independency of patient backgrounds also accounts for the high predictive value of our system.

Current issues in DNA microarray technology

Prior to the clinical use of DNA microarray analysis, the most significant is replication of DNA microarray results in a larger cohort. I suggest that failure to exclude "chance" is the most likely explanation for difficulty in the replication of complex diseases such as cancer. Particularly, the failure would be promoted by a situation specific to microarray analysis. The number of samples is usually much lower than that of genes in a microarray study. For example, in our study,[2] the number of genes examined was 7070. By contrast, the number of training samples was 33, $1/20^{th}$ that of the genes. This situation is opposite to that of traditional clinical studies, *i.e.,* 1 gene in 1 experiment. When the number of available training samples is small, gene selection is strongly influenced by the variability of the samples.[13] Thus, a small sample size easily allows chance to play a role in the microarray result.

To avoid or decrease the influence of chance, we must establish the best way to cope with sample variability in a small sample size and assess the constructed predictor in an independent cohort.[13] Thus, resolving the instability of gene selection in a small sample size is critical for future study design. I find it very intriguing that the problem of "small sample size" was proposed in the field of pattern recognition in the 1970s, when insightful works[14,15] by Dr. Okita, one of the great pioneers for HCC research, attracted a great deal of attention in the field.

Most investigators have believed in the central-dogma that transcript levels parallel protein levels in most aspects of cellular biol-

Table Representative DNA microarray-based predictors reported in high impact journals

Institution	Cancer type	Predictor and number of genes used	Predictive accuracy	Publication
Yamaguchi Univ., Japan	Hepatocellular carcinoma	Fisher linear classifier with 12 genes	Prediction of intrahepatic recurrence with 93% accuracy	Lancet. 2003; 361(9361):923-9.
NIH, USA	Hepatocellular carcinoma	Compound co-variate predictor with 153 genes	Prediction of intrahepatic metastasis at surgery with 85% accuracy	Nat Med. 2003; 9(4):416-23.
NIH, USA	Hepatocellular carcinoma	Five-type predictors with 406 genes	Prediciton of long- and short-term survivors by P<0.036	Hepatology. 2004; 40(3):667-76.
NCI, Netherlands	Breast cancer	Cluster analysis with 70 genes	Prediction of poor- and good-prognosis signatures by P<0.001	N Engl J Med. 2002;347(25):1999-2009.
KF-SYS Cancer Centre, Taiwan	Breast cancer	Predictive statistical tree models with 496 genes	Prediction of outcomes of patients with 90% accuracy	Lancet. 2003; 361(9369):1590-6.
University of Michigan, USA	Lung adenocarcinoma	Risk index with 50 genes	Classification of low- and high-risk stage I tumors with high accuarcy	Nat Med. 2002; 8(8):816-24.
MIT, USA	Leukemia	Weighted voting with 50 genes	*Classification of two-type leukemias with 100% accuracy	Science. 1999; 286(5439):531-7.

* This system can use a rejection option. Indeed, the accuracy rate is 29 (85%) of 34.
Note that our predictor with only 12 genes shows high accuracy similar to that of other predictors.

ogy. According to this scenario, we and many investigators have studied the transcriptome. However, the levels of transcripts were not always consistent with those of the encoded proteins in our series,[16)17)] and vice versa. Naturally, we then asked what the transcriptome we observed actually indicated. The discrepancy between transcript and proteins levels might be explained partly by post-transcriptional regulation and/or the time-lag between sampling points. For the former, up-to-date research reveals the presence of micro RNA, which may be responsible for translational modulation of genes.[18)] Thus, the central-dogma has been invalidated, and the relevance of the observed transcriptome is now a complete mystery. Molecular biological techniques appear inadequate for resolving this issue, and bioinformatics may be the only useful strategy for the task.

Perspectives of DNA microarray technology

I propose empirically that two types of predictive systems will be available for use by hepatologists early in the 21st century. One is a clinical standard system that can classify HCC patients into several subclasses according to their outcomes and show a similar predictive value worldwide. Such a system would promise precise information or evidence regarding the degree to which a newly developed drug or therapy benefits HCC patients.

The other is an individualizing system. Our current system[2)] makes it possible to stratify, but not to individualize, HCC patients for early IHR. We do also hope to individualize HCC patients to provide them with more personalized therapeutic options. In my opinion, combining a clinical staging system with molecular profiling will be the best way to individualize them. For example, such a combined system might calculate correctly the outcome of each patient within a range of N months +/- 3 months after surgery. To reach this goal, we must evaluate the predictive performance of our current system[2)] in a larger cohort of hundreds of patients. Unfortunately, the Affymetrix microarrays used in our study[2)] are too expensive for this purpose. To allow daily clinical use, our strategy must be replaced by more easy-to-use low cost analyses, such as, custom arrays or real-time quantitative RT-PCR. Studies of miniature arrays containing only 30 genes and low-cost PCR assay are currently underway. Our project team expects that these systems will become the mainstay for personalizing HCC patients in the near future.

Acknowledgements

I thank Prof. M. Oka, the project conductor, for his guidance in performing this research. I thank Prof. Y. Hamamoto, Dr. S. Uchimura, and Dr. T. Miyamoto for their bioinformatics support. I thank Dr. H. Yamada-Okabe and Dr. H. Ishitsuka for analysis of the DNA microarrays. I also thank successive liver-team surgeons (Dr. T. Tamesa, Dr. R. Shimizu, Dr. K. Yano, Dr. M. Nishida, Dr. T. Yagyu, Dr.Y. Maeda, Dr. K. Takao, Dr. S. Hiraki, Dr. N. Mori, Dr. T. Takao, Dr. T. Okada, Dr. N. Takemoto, Dr. K. Matoba, Dr. M. Takashima, and Dr. K. Sakamoto) of Surgery II, Yamaguchi University School of Medicine, Japan, for supplying us with HCC samples and the follow-up data.

References

1) Llovet, J.M., Burroughs, A. and Bruix, J. : Hepatocellular carcinoma. *Lancet*, **362** : 1907-1917, 2003.
2) Iizuka, N., Oka, M., Yamada-Okabe, H., Nishida, M., Maeda, Y., Mori, N., Takano, T., Tamesa, T., Tangoku, A., Tabuchi, H., Hamada, K., Nakayama, H., Ishitsuka, H., Miyamoto, T., Hirabayashi, A., Uchimura, S. and Hamamoto, Y. : Oligonucleotide microarray for prediction of early intrahepatic recurrence of hepatocellular carcinoma after curative resection. *Lancet,* **361** : 923-929, 2003.
3) http://surgery2.med.yamaguchi-u.ac.jp/research/DNAchip/hcc-recurrence/index.html.
4) Matoba, K., Iizuka, N., Gondo, T., Ishihara, T., Yamada-Okabe, H., Tamesa, T., Takemoto, N., Hashimoto, K., Sakamoto, K., Miyamoto, T., Uchimura, S., Hamamoto, Y. and Oka, M. : Tumor HLA-DR

expression linked to early intrahepatic recurrence of hepatocelluar carcinoma. *Int. J. Cancer.,* **115** : 231-240, 2005.
5) Okita, K., Sakaida, I. and Hino, K. : Curent strategies for chemoprevention of hepatocellular carcinoma. *Oncology,* **62** : S24-S28, 2002.
6) Iizuka, N., Hamamoto, Y. and Oka, M. : Predicting individual outcomes in hepatocellular carcinoma. *Lancet,* **364** : 1837-1839, 2004.
7) Iizuka, N., Oka, M., Yamada-Okabe, H., Mori, N., Tamesa, T., Okada, T., Takemoto, N., Sakamoto, K., Hamada, K., Ishitsuka, H., Miyamoto, T., Uchimura, S. and Hamamoto, Y. : Self-organizing-map-based molecular signature representing the development of hepatocelluar carcinoma. *FEBS Lett.,* **579** : 1089-1100, 2005.
8) Iizuka, N., Oka, M., Yamada-Okabe, H., Mori, N., Tamesa, T., Okada, T., Takemoto, N., Tangoku, A., Hamada, K., Nakayama, H., Miyamoto, T., Uchimura, S. and Hamamoto, Y. : Comparison of gene expression profiles between hepatitis B virus- and hepatitis C virus-infected hepatocellular carcinoma by oligonucleotide microarray data based on a supervised learning methods. *Cancer Res.,* **62** : 3939-3944, 2002.
9) Iizuka, N., Oka, M., Yamada-Okabe, H., Mori, N., Tamesa, T., Okada, T., Takemoto, N., Hashimoto, K., Tangoku, A., Hamada, K., Nakayama, H., Miyamoto, T., Uchimura, S. and Hamamoto, Y. : Differential Gene Expression in Distinct Virologic Types of Hepatocellular Carcinoma: Association with Liver Cirrhosis. *Oncogene,* **22** : 3007-3014, 2003.
10) Iizuka, N., Oka, M., Yamada-Okabe, H., Hamada, K., Nakayama, H., Mori, N., Tamesa, T., Okada, T., Takemoto, N., Matoba, K., Takashima, M., Sakamoto, K., Tangoku, A., Miyamato, T., Uchimura, S. and Hamamoto, Y.: Molecular signature in three types of hepatocellular carcinoma with different viral origin by oligonucleotide microarray. *Int. J. Oncol.,* **24** : 565-574, 2004.
11) Okada, T., Iizuka, N., Yamada-Okabe, H., Mori, N., Tamesa, T., Takemoto, N., Tangoku, A., Hamada, K., Nakayama, H., Miyamoto, T., Uchimura, S., Hamamoto, Y. and Oka, M.: Gene expression profile linked to p53 status in hepatitis C virus-related hepatocellular carcinoma. *FEBS Lett.,* **555** : 583-590, 2003.
12) Takemoto, N., Iizuka, N., Yamada-Okabe, H., Hamada, K., Tamesa, T., Okada, T., Hashimoto, K., Sakamoto, K., Takashima, M., Miyamoto, T., Uchimura, S., Hamamoto, Y. and Oka, M. : Sex-based molecular profiling of hepatitis C virus-related hepatocellular carcinoma. *Int. J. Oncol.,* **26** : 673-678, 2005.
13) Iizuka, N., Hamamoto, Y. and Oka, M. : Prediction of cancer outcome with microarrays. *Lancet,* **365** : 1683-1684, 2005.
14) Okita, K., Gruenstein, M., Klaiber, M. and Farber, E. : Localization of alpha-fetoprotein by immunofluorescence in hyperplastic nodules during hepatocarcinogenesis induced by 2-acetylaminofluorene. *Cancer Res.,* **34** : 2758-2763, 1974.
15) Okita, K., Kligman, L.H. and Farber, A. : A new common marker for premalignant and malignant hepatocytes induced in the rat by chemical carcinogens. *J. Natl. Cancer Inst.,* **54** : 199-202, 1975.
16) Takashima, M., Kuramitsu, Y., Yokoyama, Y., Iizuka, N., Toda, T., Sakaida, I., Okita, K., Oka, M and Nakamura, K.: Proteomic profiling of heat shock protein 70 family members as biomarkers for hepatitis C virus-related hepatocellular carcinoma. *Proteomics,* **3** : 2487-2493, 2003.
17) Yokoyama, Y., Kuramitsu, Y., Takashima, M., Iizuka, N., Toda, T., Terai, S., Sakaida, I., Oka, M., Nakamura, K. and Okita, K. : Proteomic profiling of proteins decreased in hepatocellular carcinoma from patients infected with hepatitis C virus. *Proteomics,* **4** : 2111-2116, 2004.
18) Croce, C.M. and Calin, G.A. : miRNAs, cancer, and stem cell division. *Cell,* **122** : 6-7, 2005.