

## Challenges of incorporating gene expression data to predict HCC prognosis in the age of systems biology

Yan Du, Guang-Wen Cao

Yan Du, Guang-Wen Cao, Department of Epidemiology, Second Military Medical University, Shanghai 200433, China

**Author contributions:** Du Y collected the materials and drafted the manuscript; and Cao GW supervised and revised the manuscript.

**Supported by** The National Outstanding Youth Fund, No. 81025015; Key Project Fund, No. 91129301; and Creative Research Group Fund of the National Natural Science Foundation of China, No. 30921006

**Correspondence to:** Guang-Wen Cao, Chairman, MD, PhD, Professor of Medicine, Department of Epidemiology, Second Military Medical University, 800 Xiangyin Rd., Shanghai 200433, China. [gcao@smmu.edu.cn](mailto:gcao@smmu.edu.cn)

Telephone: +86-21-81871060 Fax: +86-21-81871060

Received: June 11, 2012 Revised: June 26, 2012

Accepted: June 28, 2012

Published online: August 14, 2012

### Abstract

Hepatocellular carcinoma (HCC) is a leading cause of cancer-related death worldwide. The recurrence of HCC after curative treatments is currently a major hurdle. Identification of subsets of patients with distinct prognosis provides an opportunity to tailor therapeutic approaches as well as to select the patients with specific sub-phenotypes for targeted therapy. Thus, the development of gene expression profiles to improve the prediction of HCC prognosis is important for HCC management. Although several gene signatures have been evaluated for the prediction of HCC prognosis, there is no consensus on the predictive power of these signatures. Using systematic approaches to evaluate these signatures and combine them with clinicopathologic information may provide more accurate prediction of HCC prognosis. Recently, Villanueva *et al.*<sup>[13]</sup> developed a composite prognostic model incorporating gene expression patterns in both tumor and adjacent tissues to predict HCC recurrence. In this commentary, we summarize the current progress in using gene signatures to predict HCC prognosis, and discuss the importance, existing issues and future research directions in this field.

© 2012 Baishideng. All rights reserved.

**Key words:** Gene expression signatures; Hepatocellular carcinoma; Prognosis

**Peer reviewers:** Thomas Kietzmann, Professor, Department of Biochemistry, University of Oulu, FI-90014 Oulu, Finland; Andrzej S Tamawski, MD, PhD, DSc (Med), Professor of Medicine, Chief Gastroenterologist, VA Long Beach Health Care System, University of California Irvine School of Medicine, 5901 E. 7th Street, Long Beach, CA 90822, United States; Yujin Hoshida, MD, PhD, Cancer Program, Broad Institute, 7 Cambridge Center, Cambridge, MA 02142, United States; Ferruccio Bonino, MD, PhD, Professor of Gastroenterology, Director of Liver and Digestive Disease Division, Director of General Medicine 2 Unit, Department of Internal Medicine, University Hospital of Pisa, Via Roma 67, 56124 Pisa, Italy

Du Y, Cao GW. Challenges of incorporating gene expression data to predict HCC prognosis in the age of systems biology. *World J Gastroenterol* 2012; 18(30): 3941-3944 Available from: URL: <http://www.wjgnet.com/1007-9327/full/v18/i30/3941.htm> DOI: <http://dx.doi.org/10.3748/wjg.v18.i30.3941>

### INVITED COMMENTARY ON HOT ARTICLES

Hepatocellular carcinoma (HCC) is the sixth most common cancer type and the third leading cause of cancer-related death worldwide<sup>[1]</sup>. The major risk factor of HCC is chronic infection with hepatitis B virus (HBV) and/or hepatitis C virus (HCV)<sup>[2]</sup>. So far, curative treatments for HCC include orthotopic liver transplantation, surgical resection and percutaneous ablation. However, the recurrence rates remain high and long-term survival is poor.

There are two types of HCC recurrence: early recurrence and late recurrence with different mechanisms. Early recurrence (< 2 years after the treatment) is mostly caused by metastasis and dissemination of primary HCC; while late recurrence ( $\geq$  2 years after the treatment) mainly

results from *de novo* tumors, as a consequence of field effect in diseased liver which is closely associated with high viral loads and hepatic inflammatory activities<sup>[3,4]</sup>. The treatment after curative therapy varies greatly, depending on individual's profile<sup>[5]</sup>. The traditional prognostic markers of HCC include vascular invasion (both macroscopic and microscopic) which is the most significant factor, tumor size, number of nodules,  $\alpha$ -fetoprotein level, degree of differentiation, and satellites<sup>[6]</sup>. Recent advancement in the field has shown that viral factors and inflammation-related conditions are apparently associated with HCC prognosis. Viral load, genotype C, viral mutations, and expression of inflammatory molecules in HBV-related HCC tissues are significantly associated with poor prognosis. Host-inflammation-related factors such as imbalance between intratumoral CD8<sup>+</sup> T lymphocytes and regulatory T lymphocytes, T helper (Th)1 and Th2 cytokines in peritumoral tissues are also predictors of HBV-related HCC<sup>[7,8]</sup>. In addition, non-coding RNA also plays a significant role in HCC progression<sup>[9]</sup>. However, even after incorporating viral and other factors, the prediction power can not be optimized. Therefore, it is crucial to identify new prognostic markers to better approach opportunities for individualized therapeutics for HCC patients.

The application of high-throughput methods has provided new opportunities for analyzing the diversity and heterogeneity of cancers. Studies of microarray-based gene expression profiling in breast cancer have shown a great success and led to a working model for a breast cancer molecular taxonomy<sup>[10]</sup>. Gene expression signatures succeeded in prognosis prediction and treatment responses for HCC<sup>[11]</sup>, and they are promising in developing personalized cancer medication<sup>[12]</sup>. Gene expression profiles may add new and important prognostic information beyond those provided by the standard clinical predictors. It is important to incorporate molecular information to more accurately predict early and overall recurrence of HCC.

We read with great interest the recent article by Villanueva *et al.*<sup>[13]</sup>. In this article, the authors developed an integrated prognostic model combining genomic and clinicopathologic data to improve outcome prediction in single-nodule early HCC patients. They analyzed the prognostic power of 22 previously reported gene signatures in a cohort of 287 early-stage HCC patients. The analysis showed that the proliferation signature was the most prevalent prediction (number of patients identified with the signature/number of total patients); and there was a substantial association among three groups of signatures: (1) signatures related to increased cell proliferation, progression in cell cycle and activation of specific pathways; (2) signatures generated in the adjacent tissues; and (3) cytokeratin-19 gene signature. They found that G3 (tumoral) signature and poor-survival (non-tumoral) signature, along with satellites were independent predictors of early tumor recurrence and overall recurrence. They also reported that genomic profiles of tumor and adjacent tissues were complementary in refining the prediction.

Advanced imaging techniques such as computed to-

mography and magnetic resonance imaging have been used to detect vascular invasion and conduct satellite evaluation before surgery, which are helpful in the pre-operative prediction of HCC prognosis. Genomic profiling using tumor and adjacent tissues obtained by fine-needle biopsy may provide complementary and/or confirmative information, thus having a great potential when combined with imaging findings in the clinical practice. Many studies have used array-based gene expression profiling obtained from tumoral or non-tumoral tissues to predict HCC prognosis. However, the number and heterogeneity of the signatures hinders their further application. The study of Villanueva *et al.*<sup>[13]</sup> attempted to address these issues. They evaluated the prognostic predictive power of previously reported gene signatures in an independent cohort, and then developed a "composite genomic-based prognostic model". They further validated the stability of the model using samples from different sites of the same tumor nodule to test whether the genomic signature was consistent throughout different sites of a tumor<sup>[13]</sup>. This study presents a unified approach to systematically evaluate and independently validate HCC prognostic gene signatures; and the procedure developed in this study is conducive to the future studies of other complex disease.

Cancer gene signatures may indicate specific biological traits of heterogeneous tumor sub-phenotypes that cannot be identified by traditional methods. They may be associated with tumor biology and tumor microenvironment such as chromosomal instability, wounded stroma, or invasiveness, and possibly also linked to certain signaling pathways<sup>[14]</sup>. Gene signatures may have functional implications and may be predictive of response to specific therapeutic agents such as antiviral medications. Signatures identified in the study of Villanueva *et al.*<sup>[13]</sup> (tumoral G3-proliferation signature and nontumoral poor-prognosis signature) reflect highly relevant biological events for outcome prediction and point out possible pathways to search for biomarkers as therapeutic targets. If used appropriately, gene signatures should be important complementary methods to current clinicopathological risk stratification systems<sup>[15]</sup>. Integrating gene signatures in HCC prognosis prediction may potentially improve patient outcomes, obtain a better understanding of the underlying HCC biology, and identify effective therapeutic options for an individual patient.

HCC is not a single disease at the molecular level. Using gene signatures to classify HCC into molecular subtypes with similar prognostic implication can guide clinical decision-making, particularly regarding therapy. However, these signatures lack prognostic power. The assignment of a given patient to a subgroup is strongly dependent on the gene signature used and the results from studies of a specific/single gene signature cannot necessarily be generalized. Furthermore, there are few genes overlapped among gene expression signatures which reflect common cellular phenotypes and yield similar predictions. Therefore, it is not appropriate to use overlapping in gene identity to measure the reproducibility of gene-expression profiles<sup>[16]</sup>. Thus, systematic evaluation of different gene

expression datasets and validation in independent cohorts provide basis for identifying true genomic signatures that are associated with oncogenic pathway, tumor biology and its microenvironment. Nevertheless, there are problems of using gene signatures to classify sub-phenotypes and predict HCC prognosis. In the following section, we take the paper of Villanueva *et al.*<sup>[13]</sup> for an example to discuss several imposing issues in the field.

First, the paper does not mention whether evaluation on the quality of the different gene signatures was used. These signatures were generated from different samples with different biological background. Different studies may vary greatly in study quality, such as patient selection criteria, RNA quality, follow-up criteria, definition of prognosis, treatment after surgery, *etc.* Patient differences including different staging and underlying conditions may reflect etiological differences, thus resulting in the heterogeneity of gene signatures. Prognostic accuracy might differ in tumors with different stages. Additionally, multiple end points, such as overall recurrence, early recurrence, late recurrence, overall survival, or metastasis-free survival, used in the analyses are also the source of heterogeneity. There is also the possibility of stromal contamination, namely, gene signatures derived from analysis of tumor specimens with a high proportion of adjacent tissue contamination, and vice versa. The general reproducibility of these signatures stands out as an important issue.

Second, it is inappropriate to directly combine datasets from different platforms and different experiments because of the non-biological experimental variation or batch effects. In the study of Villanueva *et al.*<sup>[13]</sup>, gene expression data were obtained from 3 high-throughput genomic platforms, and these datasets cannot be readily put together because of their heterogeneity. Again, the authors did not mention whether any standardization procedures were applied. In addition, the method used for integration and/or standardization of different platforms is also a challenge. How to choose a robust normalization method according to the features of the dataset to reduce the batch effect is essential for further computational analyses<sup>[17]</sup>.

Third, the authors did not describe whether they applied the gene mapping procedure. Gene database updates with time, with the accumulation of information, the platform used several years ago may not be comparable to the gene database in service now. Without mapping, the genes in the 22 signatures produced at different time points may not correspond well. Accurately mapping and matching a gene across different signatures generated by different platforms at different time points is an important quality control step to enable the finding of true signatures.

Last but not least, the quality of survival analyses used to generate these signatures differs. The frequently used statistical methods, such as the significant analysis of microarray tool, the trend filter tool, and Cox's proportional Hazard model, may contribute to the great variety of gene expression signatures<sup>[17]</sup>. Different studies also vary in terms of follow-up information collected, covariates

adjusted in multivariate analysis, and non-informative censoring. These directly affect the gene signatures generated.

For gene signatures to be used in clinical practice to accurately predict HCC prognosis, the following procedures are required. For a start, there should be a standardization of tissue composition. Without appropriate and standardized samples, the further experiments to determine a robust signature will be difficult. For example, the variable selection procedure is crucial in developing reliable and reproducible gene signature because pre-analytical variables such as stromal component and tissue processing will directly affect gene expression profiles. In addition, to enable the usage of data by different researchers and future investigators, a detailed description of data processing and analytical methods is required. A further step is to establish unified high criteria for generating gene expression signatures. Moreover, it is also important to identify gene signatures to predict early and late recurrence of HCC. HCCs are a group of diverse and heterogeneous diseases. Gene expression patterns can provide a basis to distinguish sub-phenotypes within the heterogeneity subgroups characterized by conventional clinicopathological variables, and also present important information about individualization of therapy<sup>[18]</sup>. Viral mutations in the preS and the basal core promoter regions of HBV are significantly associated with HCC risk<sup>[19-23]</sup>. The HBV mutations including A1762T/G1764A, preS deletion at nt.107-141, and preS2 mutations in adjacent hepatic tissues and the HCV mutation such as M91L are significantly associated with poor prognosis of HCC<sup>[24-26]</sup>. The viral mutations should be reasonably integrated into the HCC prognosis-related gene signature.

To summarize, this paper drew our interests because gene expression signatures have shown great promise in classifying cancer subtypes and predicting prognosis. The Villanueva team has introduced an effective approach to systematically integrate different types of data for HCC prognosis prediction. With the increasing amount of data produced, there is an urgent need of standardized methods in systems biology to integrate descriptive data from cohort studies and other sources such as clinicopathological features, massive DNA and RNA parallel sequencing, and proteomics, along with functional data to guide therapeutic decisions. In addition, data on vascular features of HCC from imaging techniques may help select and validate true gene expression signatures associated with HCC prognosis. Future studies should also correlate these two non-invasive and innovative methods. It is still premature to use the current gene signatures for predicting HCC prognosis in the context of clinical practice. There is enormous work to be done for these gene signatures to be used in routine clinical practice and treatment decision making.

## REFERENCES

- 1 Siegel R, Naishadham D, Jemal A. Cancer statistics, 2012. *CA Cancer J Clin* 2012; **62**: 10-29



- 2 **Tang ZY.** Hepatocellular carcinoma--cause, treatment and metastasis. *World J Gastroenterol* 2001; **7**: 445-454
- 3 **Imamura H,** Matsuyama Y, Tanaka E, Ohkubo T, Hasegawa K, Miyagawa S, Sugawara Y, Minagawa M, Takayama T, Kawasaki S, Makuuchi M. Risk factors contributing to early and late phase intrahepatic recurrence of hepatocellular carcinoma after hepatectomy. *J Hepatol* 2003; **38**: 200-207
- 4 **Wu JC,** Huang YH, Chau GY, Su CW, Lai CR, Lee PC, Huo TI, Sheen IJ, Lee SD, Lui WY. Risk factors for early and late recurrence in hepatitis B-related hepatocellular carcinoma. *J Hepatol* 2009; **51**: 890-897
- 5 **Du Y,** Su T, Ding Y, Cao GW. Effects of antiviral therapy on the recurrence of hepatocellular carcinoma after curative resection or liver transplantation. *Hepat Mon* 2012; In press
- 6 **Llovet JM,** Schwartz M, Mazzaferro V. Resection and liver transplantation for hepatocellular carcinoma. *Semin Liver Dis* 2005; **25**: 181-200
- 7 **Chen L,** Zhang Q, Chang W, Du Y, Zhang H, Cao G. Viral and host inflammation-related factors that can predict the prognosis of hepatocellular carcinoma. *Eur J Cancer* 2012
- 8 **Han YF,** Zhao J, Ma LY, Yin JH, Chang WJ, Zhang HW, Cao GW. Factors predicting occurrence and prognosis of hepatitis-B-virus-related hepatocellular carcinoma. *World J Gastroenterol* 2011; **17**: 4258-4270
- 9 **Zhang Q,** Pu R, Du Y, Han Y, Su T, Wang H, Cao G. Non-coding RNAs in hepatitis B or C-associated hepatocellular carcinoma: potential diagnostic and prognostic markers and therapeutic targets. *Cancer Lett* 2012; **321**: 1-12
- 10 **Mackay A,** Weigelt B, Grigoriadis A, Kreike B, Natrajan R, A'Hern R, Tan DS, Dowsett M, Ashworth A, Reis-Filho JS. Microarray-based class discovery for molecular classification of breast cancer: analysis of interobserver agreement. *J Natl Cancer Inst* 2011; **103**: 662-673
- 11 **Faivre S,** Raymond E, Boucher E, Douillard J, Lim HY, Kim JS, Zappa M, Lanzalone S, Lin X, Deprimo S, Harmon C, Ruiz-Garcia A, Lechuga MJ, Cheng AL. Safety and efficacy of sunitinib in patients with advanced hepatocellular carcinoma: an open-label, multicentre, phase II study. *Lancet Oncol* 2009; **10**: 794-800
- 12 **van't Veer LJ,** Bernards R. Enabling personalized cancer medicine through analysis of gene-expression patterns. *Nature* 2008; **452**: 564-570
- 13 **Villanueva A,** Hoshida Y, Battiston C, Tovar V, Sia D, Alsinet C, Cornella H, Liberzon A, Kobayashi M, Kumada H, Thung SN, Bruix J, Newell P, April C, Fan JB, Roayaie S, Mazzaferro V, Schwartz ME, Llovet JM. Combining clinical, pathology, and gene expression data to predict recurrence of hepatocellular carcinoma. *Gastroenterology* 2011; **140**: 1501-12.e2
- 14 **Nevins JR,** Potti A. Mining gene expression profiles: expression signatures as cancer phenotypes. *Nat Rev Genet* 2007; **8**: 601-609
- 15 **Utsunomiya T,** Okamoto M, Wakiyama S, Hashimoto M, Fukuzawa K, Ezaki T, Aishima S, Yoshikawa Y, Hanai T, Inoue H, Barnard GF, Mori M. A specific gene-expression signature quantifies the degree of hepatic fibrosis in patients with chronic liver disease. *World J Gastroenterol* 2007; **13**: 383-390
- 16 **Fan C,** Oh DS, Wessels L, Weigelt B, Nuyten DS, Nobel AB, van't Veer LJ, Perou CM. Concordance among gene-expression-based predictors for breast cancer. *N Engl J Med* 2006; **355**: 560-569
- 17 **Cavalieri D,** Dolara P, Mini E, Luceri C, Castagnini C, Toti S, Maciag K, De Filippo C, Nobili S, Morganti M, Napoli C, Tonini G, Baccini M, Biggeri A, Tonelli F, Valanzano R, Orlando C, Gelmini S, Cianchi F, Messerini L, Luzzatto L. Analysis of gene expression profiles reveals novel correlations with the clinical course of colorectal cancer. *Oncol Res* 2007; **16**: 535-548
- 18 **Acharya CR,** Hsu DS, Anders CK, Anguiano A, Salter KH, Walters KS, Redman RC, Tuchman SA, Moylan CA, Mukherjee S, Barry WT, Dressman HK, Ginsburg GS, Marcom KP, Garman KS, Lyman GH, Nevins JR, Potti A. Retraction: Acharya CR, et al. Gene expression signatures, clinicopathological features, and individualized therapy in breast cancer. *JAMA*. 2008; 299(13): 1574-1587. *JAMA* 2012; **307**: 453
- 19 **Liu S,** Xie J, Yin J, Zhang H, Zhang Q, Pu R, Li C, Ni W, Wang H, Cao G. A matched case-control study of hepatitis B virus mutations in the preS and core promoter regions associated independently with hepatocellular carcinoma. *J Med Virol* 2011; **83**: 45-53
- 20 **Yin J,** Xie J, Liu S, Zhang H, Han L, Lu W, Shen Q, Xu G, Dong H, Shen J, Zhang J, Han J, Wang L, Liu Y, Wang F, Zhao J, Zhang Q, Ni W, Wang H, Cao G. Association between the various mutations in viral core promoter region to different stages of hepatitis B, ranging of asymptomatic carrier state to hepatocellular carcinoma. *Am J Gastroenterol* 2011; **106**: 81-92
- 21 **Yin J,** Xie J, Zhang H, Shen Q, Han L, Lu W, Han Y, Li C, Ni W, Wang H, Cao G. Significant association of different preS mutations with hepatitis B-related cirrhosis or hepatocellular carcinoma. *J Gastroenterol* 2010; **45**: 1063-1071
- 22 **Cao GW.** Clinical relevance and public health significance of hepatitis B virus genomic variations. *World J Gastroenterol* 2009; **15**: 5761-5769
- 23 **Liu S,** Zhang H, Gu C, Yin J, He Y, Xie J, Cao G. Associations between hepatitis B virus mutations and the risk of hepatocellular carcinoma: a meta-analysis. *J Natl Cancer Inst* 2009; **101**: 1066-1082
- 24 **Tsai HW,** Lin YJ, Lin PW, Wu HC, Hsu KH, Yen CJ, Chan SH, Huang W, Su IJ. A clustered ground-glass hepatocyte pattern represents a new prognostic marker for the recurrence of hepatocellular carcinoma after surgery. *Cancer* 2011; **117**: 2951-2960
- 25 **Yeh CT,** So M, Ng J, Yang HW, Chang ML, Lai MW, Chen TC, Lin CY, Yeh TS, Lee WC. Hepatitis B virus-DNA level and basal core promoter A1762T/G1764A mutation in liver tissue independently predict postoperative survival in hepatocellular carcinoma. *Hepatology* 2010; **52**: 1922-1933
- 26 **Toyoda H,** Kumada T, Kaneoka Y, Maeda A. Amino acid substitutions in the hepatitis C virus core region are associated with postoperative recurrence and survival of patients with HCV genotype 1b-associated hepatocellular carcinoma. *Ann Surg* 2011; **254**: 326-332

S- Editor Cheng JX L- Editor Ma JY E- Editor Xiong L