Gene Expression Profiling Reveals Potential Biomarkers of Human Hepatocellular Carcinoma

Hu-Liang Jia,[1,3] Qing-Hai Ye,[3] Lun-Xiu Qin,[3] Anuradha Budhu,[1] Marshonna Forgues,[1] Yidong Chen,[2] Yin-Kun Liu,[3] Hui-Chuan Sun,[3] Lu Wang,[3] Hong-Zhou Lu,[4] Fang Shen,[4] Zhao-You Tang,[3] and Xin Wei Wang[1]

Abstract

Purpose: Hepatocellular carcinoma (HCC), a common cancer worldwide, has a dismal outcome partly due to the poor identification of early-stage HCC. Currently, one third of HCC patients present with low serum α-fetoprotein (AFP) levels, the only clinically available diagnostic marker for HCC. The aim of this study was to identify new diagnostic molecular markers for HCC, especially for individuals with low serum AFP.

Experimental Design: We used the microarray technique to determine the expression profiles of 218 HCC specimens from patients with either high or low serum AFP. From the microarray study, we selected five candidate genes (i.e., *GPC3, PEG10, MDK, SERPINI1*, and *QP-C*), which were overexpressed in HCCs. Using quantitative real-time PCR analyses, we validated the expression of these five genes in 50 AFP-normal and 8 AFP-positive HCC specimens and 36 cirrhotic noncancerous hepatic specimens, which include 52 independent specimens not used in microarray analysis.

Results: A significant increase in the expression of the five candidate genes could be detected in most of the HCC samples, including those with normal serum AFP and small tumors. *GPC3, MDK*, and *SERPINI1* encode known serum proteins. Consistently, a significant increase in serum midkine, encoded by *MDK*, was associated with HCC patients, including those with normal serum AFP. Using prediction analysis of microarray, we showed that a combined score of these five genes can accurately classify noncancerous hepatic tissues (100%) and HCC (71%).

Conclusions: We suggest that a diagnostic signature approach using a combined score of these five biomarkers rather than a single marker may improve the prediction accuracy of HCC patients, including those with normal serum AFP and smaller-sized tumors.

Hepatocellular carcinoma (HCC), endemic to Asia and Africa with a rising incidence in Western countries (1–3), is one of the most common and aggressive cancers worldwide. It has been the third cancer killer worldwide and the second cancer killer in China since 1990s (4, 5). Globally, the 5-year survival rate of HCC is <5% and ~598,000 HCC patients die each year (6). The high mortality associated with this disease is mainly attributed to the inability to diagnose HCC patients at an early stage. In fact, most symptomatic HCC patients are diagnosed at an advanced stage, thus precluding their chance for surgical intervention (7). In contrast, HCC patients who were diagnosed at an early stage and received curative resection had a significantly improved survival time (8–10). Thus, early detection and resection have been generally recognized as the most important factors to achieve long-term survival for HCC patients.

Since its detection in the serum of HCC patients in 1970s, α-fetoprotein (AFP) has been the only serologic marker widely used for diagnosing HCC patients. This marker allows the identification of a small set of HCC patients with smaller tumors, and these patients have a relatively long-term survival rate following curative treatment (8–10). Presently, the only approach to screen for the presence of HCC in high-risk populations is the combination of serum AFP and ultrasonography (11–13). However, elevated serum AFP is only observed in about 60% to 70% of HCC patients and, to a lesser extent (33-65%), in patients with smaller HCCs (14). Moreover, nonspecific elevation of serum AFP has been found in 15% to 58% of patients with chronic hepatitis and 11% to 47% of patients with liver cirrhosis (14). Therefore, it is necessary to identify new serologic HCC biomarkers that have a sufficient

Authors' Affiliations: [1]Liver Carcinogenesis Section, Laboratory of Human Carcinogenesis, Center for Cancer Research, National Cancer Institute; [2]Laboratory of Cancer Genetics, National Human Genome Research Institute, Bethesda, Maryland; [3]Liver Cancer Institute and Zhongshan Hospital, Fudan University and [4]Shanghai Public Health Center, Shanghai, China

Received 4/27/06; revised 11/9/06; accepted 11/13/06.

Grant support: National Cancer Institute Intramural Research Program of the Center for Cancer Research and China National Natural Science Foundation for Distinguished Young Scholars grant 30325041, China National "863" R&D High-Tech Key Project grant 2002BA711A02-4, State Key Basic Research Program of China grant N. G1998051210, and Key Program Project of National Natural Science Foundation of China grant 30430720 (H-L. Jia, Q-H. Ye, L-X. Qin, Y-K. Liu, H-C. Sun, L. Wang, and Z-Y. Tang).

The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked *advertisement* in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.

Note: Supplementary data for this article are available at Clinical Cancer Research Online (http://clincancerres.aacrjournals.org/).

H-L. Jia and Q-H. Ye contributed equally to this work.

Requests for reprints: Xin Wei Wang, Laboratory of Human Carcinogenesis, Center for Cancer Research, National Cancer Institute, NIH, 37 Convent Drive, MSC 4258, Building 37, Room 3044A, Bethesda, MD 20892-4258. Phone: 301-496-2099; Fax: 301-496-0497; E-mail: xw3u@nih.gov or Zhao-You Tang, Liver Cancer Institute of Fudan University, Shanghai, China. E-mail: ytang8@yahoo.com.cn.

doi:10.1158/1078-0432.CCR-06-1025

sensitivity and specificity for the diagnosis of HCC patients, especially in AFP-normal and/or smaller HCC cases.

This study was designed to identify new biomarkers for HCC. In this study, we first conducted microarray analyses to identify genes that were differentially expressed in HCC samples among patients with high serum AFP to those with low serum AFP. We then selected five candidate genes that were highly expressed in HCC samples, including those with low serum AFP and whose protein products could potentially be identified in sera. We did validation analysis by quantitative real-time PCR analyses (qRT-PCR). We show that these five genes separate HCC specimens from cirrhotic and noncancerous liver specimens. A significant increased expression can be found in most of the HCC specimens, including cases with normal serum AFP and small tumor size, suggesting that the products of these five genes may serve as biomarkers to aid HCC diagnosis. Serum midkine was significantly elevated in HCC patients, including those who had normal serum AFP and smaller tumors. These results warrant further efforts in the development of serum-based detection assays for *GPC3* and *SERPINI1*, and possibly *PEG10* and *QP-C*, and an algorithm by combining these five markers as a potential early indicator for HCC.

Materials and Methods

Clinical specimens. Surgical tissue specimens from Chinese HCC patients were collected with informed consent and approved by the Institutional Review Board of the Liver Cancer Institute (Shanghai, China). A total of 206 HCC tumor tissue and paired adjacent nontumor liver tissue samples [Fig. 1, cohort B: 185 cases used in oligo microarray study and additional 21 cases used in qRT-PCR (Fig. 1, cohort B + C)] was collected from patients undergoing surgery during the period August 2002 to November 2003 at the Liver Cancer Institute. An additional 33 HCC samples used in cDNA microarray analysis (Fig. 1, cohort A) and 26 cirrhotic liver specimens from non-HCC chronic liver disease patients (Fig. 1, part of cohort C) were previously described (15, 16). Normal liver tissue samples were obtained from seven disease-free liver donors. Pathologic diagnosis was independently done by two pathologists. In addition, independently archived serum samples from a total of 64 HCC patients (53 males from Liver Cancer Institute) and 26 normal blood donors (24 males from the Department of Transfusion Medicine, NIH, Bethesda, MD) were also included in the study (Fig. 1, cohort D). The median ages for cases and controls are 54 (range, 25-71) and 45 (range, 18-65), respectively. An additional set of serum samples collected at the Shanghai Public Health Center of Fudan University (Shanghai, China) was independently tested by MDK ELISA, which included 32 HCC patients, 20 hospital controls, and 49 cirrhosis patients (Fig. 1, cohort E). Measurements of hepatitis viral status and AFP level were standardized between studies. The clinical characteristics of these samples are included in Table 1 and in Supplementary Tables S1, S2, and S3.

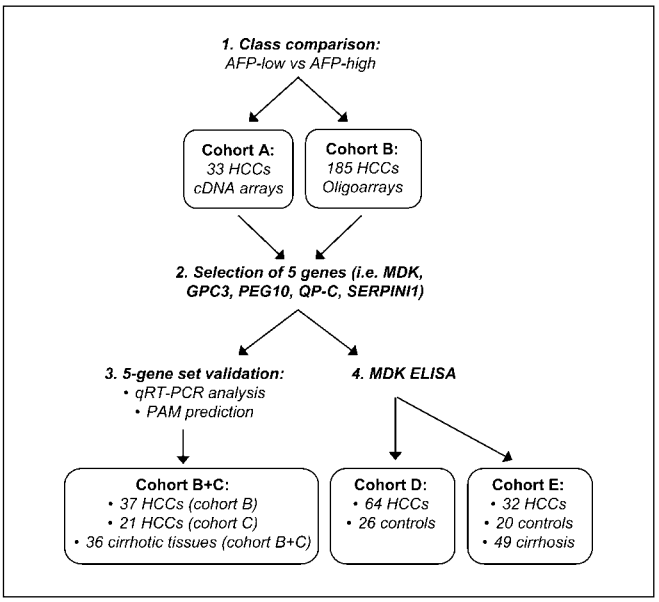

Fig. 1. Schematic of search for a predictive signature for AFP-normal HCC.

AFP grouping. The clinically acceptable normal serum AFP was defined as ≤20 ng/mL. Patients with serum AFP of >300 ng/mL were considered as true positive cases because of a higher specificity by such a cutoff (11, 17). We therefore grouped HCC cases with AFP of ≤20, <300, and ≥300 ng/mL as AFP-normal HCC, low AFP HCC, and high AFP HCC, respectively.

Microarrays, RNA isolation, qRT-PCR, MDK ELISA, and statistical analyses. The microarray slides were generated at the Advanced Technology Center of the National Cancer Institute (Gaithersburg, MD). The Human Array-Ready Oligo Set (version 2.0) containing 70-mer probe of 21,329 genes was obtained from Qiagen, Inc. (Valencia, CA). Total RNA samples were extracted from snap-frozen tissue sections using Trizol reagent (Invitrogen, Carlsbad, CA) according to the manufacturer's protocol. The microarray methodology was similar to that previously described (16), except an indirect labeling approach was used. The BRB ArrayTools software V3.2.2, developed by Simon et al. at the National Cancer Institute Biometric Research Branch,[5] was used for supervised class comparison between high and low AFP groups. The class comparison algorithm with random permutation tests was proven to be superior in identifying significant discriminatory genes with a control for false discovery (15, 16). This method was used to search for genes that were differentially expressed at a significance level of $P < 0.002$ between low and high AFP HCC.

qRT-PCR was used to validate the microarray results. The RT-PCR products collected at the last PCR cycle were analyzed by agarose gel electrophoresis to ensure the unique amplification of targeted genes (Supplementary Fig. S1). Human 18S rRNA labeled with VIC reporter dye was used as an endogenous control for each sample. A RNA pool from seven disease-free normal liver donors was used as a reference. Only AFP-normal HCC cases (≤20 ng/mL of serum AFP) and high AFP HCC cases (≥300 ng/mL of serum AFP) were analyzed by qRT-PCR. The relative mRNA expression levels were analyzed by the $2\text{-}\Delta\Delta C_t$ method (18). Prediction analysis of microarray developed by Tibshirani et al. (19) was used to classify HCC samples based on the qRT-PCR data. Multidimensional scaling analysis based on Euclidean distance was used to visualize the classification outcome. The positive predictive value and negative predictive value were calculated as follows: positive predictive value = n11 / (n11 + n21) and negative predictive value = n22 / (n12 + n22), respectively, where n11 = number of HCC predicted as HCC, n12 = number of HCC predicted as non-HCC, n21 = number of non-HCC predicted as HCC, and n22 = number of non-HCC predicted as non-HCC.

Serum midkine concentrations were determined by ELISA using a commercial kit (BioVendor, LLC, Candler, NC). The assay was done according to the manufacturer's instructions.

Results

We used the microarray technique to search for potential biomarkers that can aid the diagnosis of HCC patients, including those with low serum AFP. Our strategy used a

[5] http://linus.nci.nih.gov/BRB-ArrayTools.html

Table 1. Clinical characteristics of 84 patients used in the independent validation study by qRT-PCR

Clinical variable	HCC samples (N = 58)		Adjacent noncancerous hepatic samples (n = 10)		Other cirrhosis samples (n = 26)	
	n (%)	Range	n (%)	Range	n (%)	Range
Sex						
Male	52 (89.7)		10 (100)		7 (26.9)	
Female	6 (10.4)		0 (0)		18 (69.2)	
Unknown	0 (0)		0 (0)		1 (3.9)	
Age (y)						
<50	20 (34.5)	33-49	6 (60)	33-49	8 (30.8)	19-48
≥50	38 (65.5)	50-74	4 (40)	57-74	18 (69.2)	50-68
Hepatitis viral infection						
HBV^+	58 (100)		10 (100)		0 (0)	
HBV and HCV^+	1 (17)		1 (10)		0 (0)	
HBV infection status						
AVR-CC	14 (24.1)		2 (20)		0 (0)	
CC	44 (75.9)		8 (80)		0 (0)	
Cirrhosis (pathologic)						
Yes	55 (94.8)		10 (100)		26 (100)	
No	3 (5.2)		0 (0)		0 (0)	
Child-Pugh staging						
A	52 (94.5)		10 (100)		0 (0)	
B	3 (5.5)		0 (0)		13 (50)	
C	0 (0)		0 (0)		10 (38.5)	
Unknown	0 (0)		0 (0)		3 (11.5)	
AFP (ng/mL)						
<20	50 (86.2)	1-18	8 (80)	3-8	NA	
≥300	8 (13.8)	329-4,620	2 (20)	329-1,794	NA	
Tumor size (cm)						
<3	20 (34.5)	1-3	5 (50)	2-3	NA	
3-5	20 (34.5)	3-5	3 (30)	3-4	NA	
>5	18 (31.0)	5-15	2 (20)	7-8	NA	
Tumor-node-metastasis staging						
I	25 (43.1)		4 (40)		NA	
II	22 (37.9)		5 (50)		NA	
III	10 (17.3)		1 (10)		NA	
IV	1 (1.7)		0 (0)		NA	

NOTE: A total of 58 HCC samples, 10 randomly selected adjacent noncancerous and cirrhotic hepatic tissues from these 58 HCC samples, and 26 other cirrhosis liver samples from noncancer patients was used for cross-validation analysis by qRT-PCR. The etiologies of the 26 other cirrhosis cases are autoimmune hepatitis (n = 7), primary biliary cirrhosis (n = 15), and alcoholic liver disease (n = 4).
Abbreviation: NA, not available.

primary search for differentially expressed HCC-associated genes among AFP-low and AFP-high groups and a secondary selection of those with a ≥2-fold expression ratio in AFP-low HCC cases. We first analyzed our recent cDNA microarray expression data (containing ~9,000 human genes) of 33 HCC specimens (Fig. 1, cohort A; ref. 16). Among them, 20 cases (61%) had a high AFP serum level (≥300 ng/mL), whereas 13 cases (39%) had a low serum AFP (<300 ng/mL). Supervised class comparison with 1,000 random permutation tests and a false discovery rate control resulted in 90 significant genes (P < 0.002) that could discriminate these two groups. However, only three genes (*MDK, GPC3*, and *QP-C*) had an average expression ratio of ≥2-fold in low AFP cases (Supplementary Table S4). Due to this low success rate in identifying enough candidates for further screening and validation, we analyzed the gene expression profiles of an additional 185 HCC samples compared with their matched noncancerous liver tissues using a microarray platform consisting of 21,329 expression features (Fig. 1, cohort B). Among them, 88 (48%) were high AFP and 97 (52%) were low AFP. Similar supervised analysis resulted in a total of 1,317 significant genes (P < 0.002), of which 35 had an average expression ratio of ≥2-fold in AFP-low cases (Supplementary Table S4). Notably, the *GPC3* gene was commonly identified from both microarray studies.

Although each of the 37 candidates may serve as a potential diagnostic biomarker, initially, we subjectively selected five genes (i.e., *MDK, GPC3, QP-C, PEG10*, and *SERPINI1*) for further study by qRT-PCR based on the following considerations: (*a*) all three genes identified by cDNA array with a relatively high expression ratio in AFP-low HCC compared with the matched noncancerous liver tissues (*MDK, GPC3*, and *QP-C*) and (*b*) an additional three genes from the oligo array analysis, two of which had the highest AFP-HCC expression fold change [*GPC3* (also identified by cDNA array) and *PEG10*], and an additional gene whose encoded protein is secreted and could potentially be detected in serum or plasma samples (*SERPINI1*). For example, 92% of low AFP HCC cases had a ≥2-fold expression in the *MDK* gene, whereas 69% to 81% of low AFP HCC cases had a ≥2-fold expression in the *GPC3* gene (Supplementary Table S4).

It seemed that most low AFP HCC samples had an increased expression of *MDK*, *GPC3*, *QP-C*, and *PEG10*.

Next, we quantified the expression of these five genes by qRT-PCR on 36 noncancerous liver tissues and 58 HCC tumors, of which 21 HCC tumors and 31 noncancerous liver tissues were not used in the microarray studies described above (Table 1; Fig. 1, cohort B + C). Fifty HCC samples had normal serum AFP (≤20 ng/mL of serum AFP) and most of them were derived from cirrhotic livers (95%). The average serum levels of AFP were 6 ng/mL (range, 1-18) in the AFP-normal group and 1,758 ng/mL (range, 329-4,620) in the high AFP group (Table 1). Among 36 noncancerous liver tissues, 10 [8 AFP normal, 5 ng/mL (1–8); 2 high AFP HCC, 1,061 ng/mL (range, 329-1,794)] were randomly selected from the 58 HCC patients described above, and 26 were from cirrhotic and noncancer patients [i.e., primary biliary cirrhosis (n = 15),

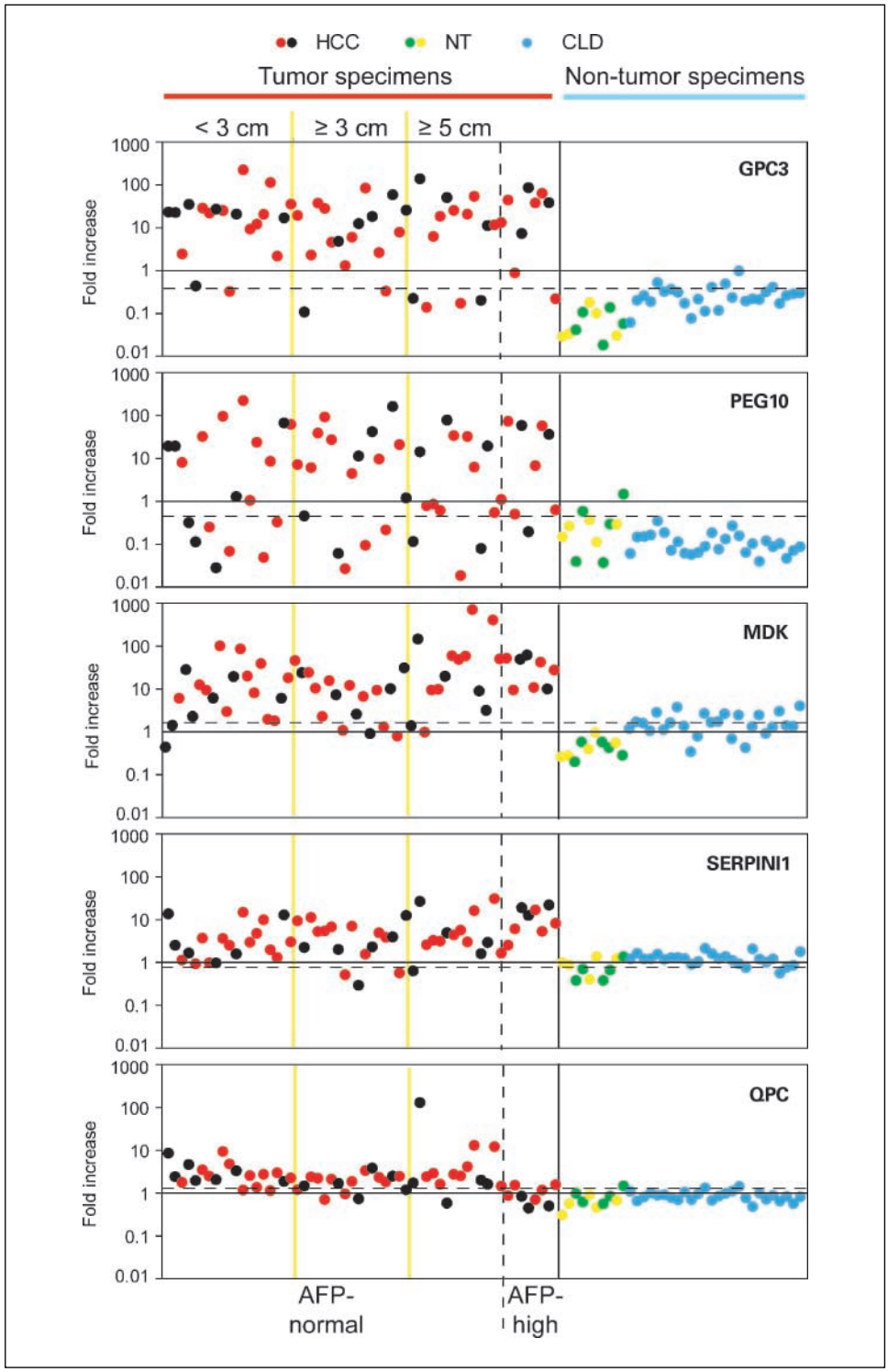

Fig. 2. Relative abundance of the mRNA levels of *GPC3, PEG10, MDK*, *SERPINI1*, and *QP-C* genes in HCC and non-HCC liver specimens. The relative mRNA levels of 58 HCC specimens and 36 non-HCC liver specimens, determined by qRT-PCR, are expressed as fold increase after normalization to 18S ribosomal gene and then to a pool of seven disease-free normal liver samples from autopsy donors. Each dot (average of triplicate ± SD) represents individual cases. Black or yellow dots, new HCC cases (tumor or nontumor, respectively); red or green dots, HCC cases also used in microarray (tumor or nontumor, respectively); blue dots, cirrhotic liver specimens from chronic liver disease and noncancerous patients. Horizontal dashed lines inside each graph, normal threshold established from the mean of 31 new non-HCC cases plus two SDs above the mean. Red and black dots on the left side of the vertical dashed lines, AFP-normal HCC cases (AFP normal, ≤20 ng/mL serum AFP). Yellow lines, divide HCC cases into three groups based on their tumor sizes (i.e., <3 cm, 3-5 cm, and ≥5 cm).

Table 2. Expression of *GPC3, PEG10, MDK, SERPINI1*, and *QP-C* in HCC determined by qRT-PCR

	Range	Total (*N*)	HCC cases that are positive for the expression*				
			GPC3, *n* (%)	*PEG10*, *n* (%)	*MDK*, *n* (%)	*SERPINI1*, *n* (%)	*QP-C*, *n* (%)
AFP status (ng/mL)							
AFP high	329-4,620	3	3 (100)	2 (67)	3 (100)	3 (100)	0 (0)
AFP normal	1-18	18	15 (83)	12 (67)	12 (67)	16 (89)	12 (67)
Tumor size (cm in diameter)							
<3	1.1-2.9	7	7 (100)	4 (57)	5 (71)	7 (100)	7 (100)
3-5	3-4.7	8	8 (100)	6 (75)	7 (100)	7 (88)	3 (38)
≥5	5.2-15	6	3 (50)	4 (67)	4 (67)	5 (83)	4 (67)
Tumor-node-metastasis staging							
Early	I	9	8 (89)	5 (56)	7 (78)	7 (78)	8 (89)
Intermediate	II	9	7 (78)	7 (78)	7 (78)	7 (100)	6 (67)
Advanced	III-IV	3	3 (100)	2 (67)	3 (100)	3 (100)	0 (0)

*A HCC case is considered positive when the lowest value of triplicate readings of a given gene is above the defined threshold. The normal threshold is defined as the mean expression level of a given gene in 31 independent noncancerous liver tissues plus two SDs. Only independent 21 HCC samples and 31 noncancerous liver specimens that were not used in microarray analysis were included in this table.

autoimmune hepatitis (n = 7), and alcoholic liver disease (n = 4)] described previously (15). The average size of primary tumor lesions from the 58 HCC patients was 4.6 cm in diameter (range, 1.1-15.3). The relative mRNA abundance of each gene was obtained by normalization to the 18S ribosomal gene and then to a pool of seven disease-free normal liver tissues. Although a total of 94 specimens was used in qRT-PCR to validate the microarray results, only 51 independent samples (21 HCC samples and 31 noncancerous hepatic samples) not used in microarray analyses were included in computing the prediction accuracy. A "normal" threshold (noncancer) for each gene was established as the mean level of each gene in 31 independent noncancerous liver samples plus two SDs of the variance. Thus, a HCC sample was considered to have an elevated expression of a given gene when its lowest level in a triplicate measurement was above the normal threshold. As shown in Fig. 2, most HCC cases had an elevated expression of these five genes above the defined normal threshold and their increased expressions seemed to be specific to HCC samples, including those with AFP normal (<20 ng/mL) and smaller tumors (<3 or 5 cm in diameter; Table 2; Fig. 2). Furthermore, the expression levels of *MDK* and *SERPINI1* were significantly higher in AFP-high samples than that of AFP-normal samples, whereas the level of *QP-C* was significantly lower in AFP-high cases (Supplementary Fig. S2). Only a few samples had an expression below the threshold, regardless of whether they were from new HCC samples (Fig. 2, *black circles*) or those originally analyzed by microarray (Fig. 2, *red circles*). Similar to the microarray data, the expression levels of *GPC3* and *PEG10* were also higher in AFP-positive than that of AFP-normal samples, although the differences were statistically insignificant (Supplementary Fig. S2). It seems that the qRT-PCR data largely correlate with the microarray data.

To further test if these five genes provided a sufficient combined expression score to separate HCC from nontumor liver specimens, we used qRT-PCR expression data to determine the probability of correctly classifying only 21 independent HCC samples and 31 independent nontumor liver specimens by the prediction analysis of microarray algorithm (Fig. 1, only cohort C). Prediction analysis of microarray analysis, using nearest shrunken centroid classification with 10-fold cross-validation, resulted in 100% correct classification of nontumor samples and 71% correct classification of independent HCC samples (Fig. 3). Similarly, we did prediction analysis of all 94 samples and found that the five-gene signature could provide 84% sensitivity (58 HCC) and 100% specificity (36 non-HCC; data not shown). In addition, similar results were obtained when all 94 samples were included and randomly partitioned into 50% training and 50% testing samples, a process that was repeated five times (Supplementary Fig. S3A). It seemed that *GPC3*, *PEG10*, and *MDK* provided the most weight to discriminate HCC from nontumor liver specimens (Supplementary Fig. S3B and C). Consistently, the same HCC samples were misclassified with these two approaches, suggesting that misclassification did not occur by random chance. Thus, this five-gene signature provided high sensitivity and specificity to classify new HCC samples. The positive predictive value and negative predictive value were 100% and 80%, respectively.

Three of the five genes encode known proteins found in sera. The level of serum midkine (*MDK*) was shown to be elevated in several types of human cancers, such as esophageal cancer, gastric cancer, and neuroblastoma (20). In addition, increased *MDK* expression was also reported in HCC (21). Because a commercial MDK ELISA assay was available, we determined serum *MDK* level in 64 additional HCC patients (cases) and 26 normal blood donors (controls; Fig. 1, cohort D). The medians (25th and 75th percentiles) of serum *MDK* were 442.8 (288.0 and 666.5) pg/mL in the cases and 11.7 (3.8 and 42.3) pg/mL in the controls. A statistically significant difference ($P < 0.0001$, two tailed) was observed between cases and controls as determined by unpaired t test with Welch's correction (Fig. 4A). No significant difference was observed in the cases when stratified by overall AFP ($P = 0.96$) or tumor size ($P = 0.22$) as determined by one-way ANOVA. However, when using the median tumor size (5 cm in diameter) as the cutoff, we found that larger tumors (≥5 cm) had significantly higher serum AFP than that of smaller tumors (<5 cm; $P = 0.02$; Fig. 4B). In contrast, the

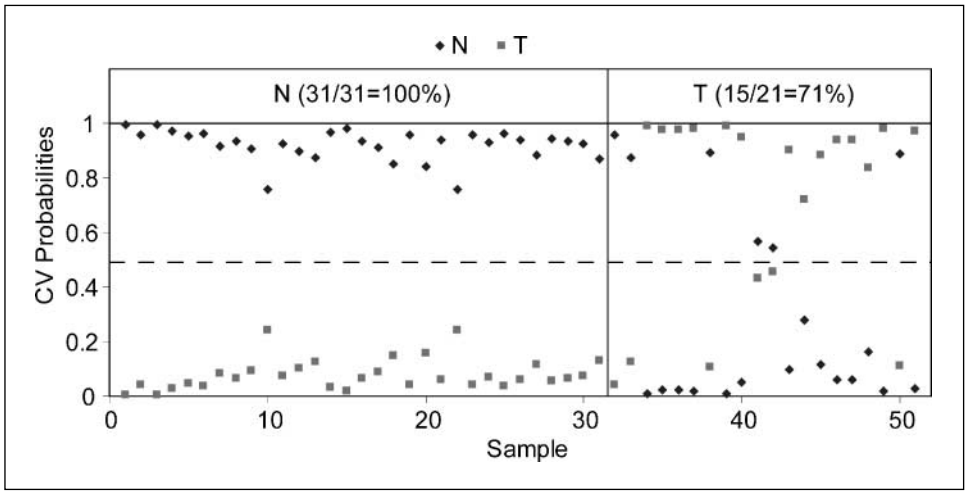

Fig. 3. qRT-PCR – based differential expression of *GPC3, PEG10, MDK, SERPINI1*, and *QP-C* genes can distinguish HCC samples from non-HCC liver tissues. The expression ratio data described in Fig. 2 were converted in a log 2 scale and subjected to prediction analysis of microarray analysis. N, 31 noncancerous liver specimens; T, 21 HCC specimens. Y axes, probabilities of correct classification with 10-fold cross-validation (*CV*). Pink squares, individual HCC cases; blue diamonds, individual non-HCC cases.

difference in serum *MDK* between small and large tumors (Fig. 4C) was insignificant ($P > 0.05$). These results suggest that serum AFP level may reflect the degree of tumor burden, whereas serum *MDK* is independent of the tumor size. Similar results were obtained in a separate cohort (Fig. 1, cohort E). We found that HCC patients had a statistically significant higher serum *MDK* than patients with cirrhosis or hospital controls ($P < 0.0001$ and $P = 0.0002$, respectively; Supplementary Fig. S4) regardless of the AFP level. Thus, it is feasible to diagnose AFP-normal and small HCC with serum samples.

Discussion

HCC is a common and aggressive malignant tumor worldwide with a dismal outcome. Early detection and resection may offer an opportunity to improve the long-term survival for HCC patients. Unfortunately, with current diagnostic approaches, only about 10% to 20% of HCC patients are eligible for resection (22). Currently, AFP is the only serologic biomarker for HCC in clinical practice. With its low sensitivity and nonspecific elevation in nonmalignant hepatic diseases, there is an urgent need to identify additional diagnostic biomarkers that can assist HCC early diagnosis, especially in AFP-normal cases and in tumors with smaller sizes.

In this study, we first did microarray analyses of HCC samples from patients with different levels of serum AFP. The class comparison analysis used in this study successfully identified AFP, the current HCC diagnostic biomarker, as a significant differentially expressed gene, further validating the utility of this approach. The microarray analysis allows us to identify 37 potential diagnostic biomarkers that are differentially expressed in AFP-positive and AFP-negative HCC samples. Five selected genes were validated by qRT-PCR as biomarkers to distinguish tumor samples from noncancerous hepatic tissues (Supplementary Table S5). A traditional HCC diagnosis has relied on the use of a single biomarker approach (e.g., AFP). We propose that the use of multiple markers in combination may improve the accuracy of identifying HCC cases. Among the five candidate genes, *GPC3*, *MDK*, and *SERPINI1* encode known extracellular proteins (i.e., glypican-3, midkine, and serine/cysteine proteinase inhibitor 12), which can be found in sera. We have tested serum *MDK* and have found it can distinguish normal and cirrhosis individuals from HCC patients, including those with normal AFP and small tumors from two independent cohorts, and could therefore be used as a diagnostic marker. In principle, such an approach can also be used to classify HCC by combining readings from several serum markers. Thus, this

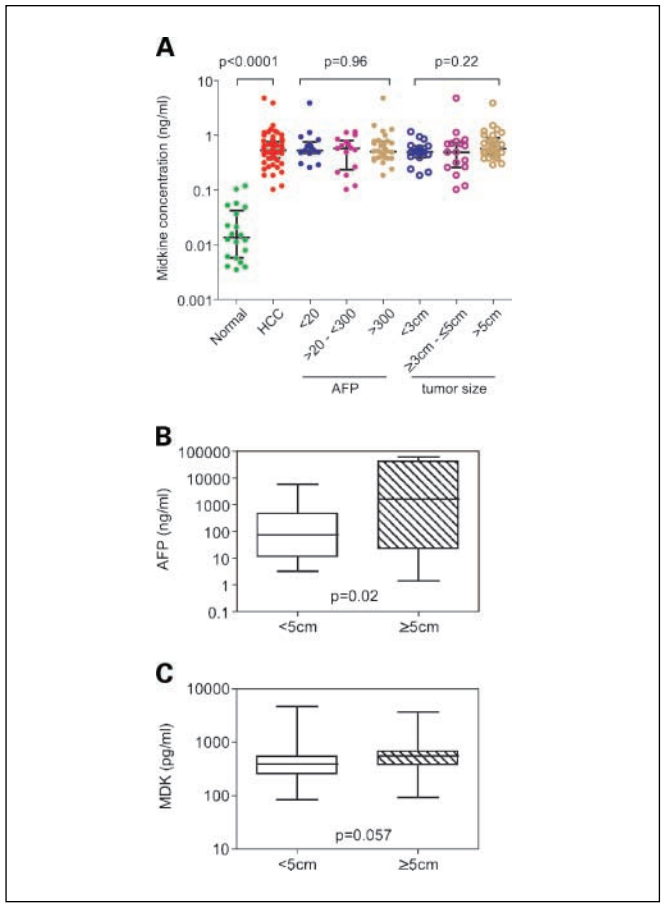

Fig. 4. Serum *MDK* is elevated in HCC patients, including those with normal serum AFP and small tumors. *A*, serum level of *MDK* in healthy volunteers ($n = 26$) and HCC patients ($n = 65$) stratified by either serum AFP (<20 ng/mL, 20-300 ng/mL, and >300 ng/mL) or tumor size. *B*, serum level of AFP stratified by the median tumor size. *C*, serum level of *MDK* stratified by the median tumor size. Median values (25th and 75th percentiles).

study warrants further development of simple methodologies, such as an ELISA-based assay, to quantify these proteins in patient serum samples. It should be noted that most of HCC cases from this study are only positive for hepatitis B. Thus, the applicability of these markers for HCC with other etiologies, such as hepatitis C virus, remains to be determined.

Our qRT-PCR analyses indicate that *GPC3* is the lead gene in distinguishing normal from HCC patients, followed by *PEG10* and *MDK*. *GPC3* encodes glypican-3, a member of the heparin sulfate proteoglycan family, which binds to the cell membrane through glycosylphosphatidylinositol anchors (23). Recent studies indicate that the serum *GPC3* level is elevated in HCC patients and support its use for HCC patients (24–26). In addition, *PEG10* is a novel imprinted gene located on human chromosome 7q21, whereas *MDK* encodes a novel heparin-binding growth factor originally identified in embryonal carcinoma cells that is involved in the early stage of retinoic acid–induced differentiation (27). Analogous to AFP, *MDK* mRNA is highly expressed during embryogenesis, but is undetectable in adult tissues, except kidney (28). The serum *MDK* level has been reported to be elevated in patients with various types of carcinomas, but not in normal individuals (20). Similarly, an increased expression of *MDK* and *PEG10* has been reported to be associated with HCC (21, 29). Moreover, the expression of *GPC3* is independent of the differentiation status and size of HCC (26). We have found that an increased expression of *GPC3*, *PEG10*, *MDK*, *SERPINI1*, and *QP-C* is associated with most HCC samples, including those with normal serum AFP and small tumor size. A combined expression score of AFP and these five genes may prove useful in the clinical setting and will be the subject of future studies. These findings warrant extensive validation of serum *GPC3*, *MDK*, and *SERPINI1* as well as the potential presence of serum *PEG10* and *QP-C* as clinical diagnostic tools by further investigation in a large case-control study with matched clinical variables, followed by a prospective study. Taken together, our results and published data indicate that these five genes may serve as diagnostic aids of HCC, including those with normal AFP and small HCC.

Acknowledgments

We thank Kei-Feng Lei, Yan-Fang Wang, Hong-Hai Wang, Xin Wu, Nicholas Younes, and Xiao-Qun Zhu for their technical support and Jia Fan, Zeng-Chen Ma, Zheng-Gang Ren, Zhi-Quan Wu, Sheng-Long Ye, and Xin-da Zhou for their help in collecting clinical specimens. We also thank Dorothea Dudek-Creaven for editorial assistance.

References

1. Taylor-Robinson SD, Foster GR, Arora S, Hargreaves S, Thomas HC. Increase in primary liver cancer in the UK, 1979-94. Lancet 1997;350:1142–3.
2. Deuffic S, Poynard T, Buffat L, Valleron AJ. Trends in primary liver cancer. Lancet 1998;351:214–5.
3. El-Serag HB, Mason AC. Rising incidence of hepatocellular carcinoma in the United States. N Engl J Med 1999;340:745–50.
4. Parkin DM, Pisani P, Ferlay J. Global cancer statistics. CA Cancer J Clin 1999;49:33–64.
5. Yang L, Parkin DM, Ferlay J, Li L, Chen Y. Estimates of cancer incidence in China for 2000 and projections for 2005. Cancer Epidemiol Biomarkers Prev 2005;14:243–50.
6. Parkin DM, Bray F, Ferlay J, Pisani P. Global cancer statistics, 2002. CA Cancer J Clin 2005;55:74–108.
7. Yuen MF, Cheng CC, Lauder IJ, Lam SK, Ooi CG, Lai CL. Early detection of hepatocellular carcinoma increases the chance of treatment: Hong Kong experience. Hepatology 2000;31:330–5.
8. Zhou XD, Tang ZY, Yang BH, et al. Experience of 1000 patients who underwent hepatectomy for small hepatocellular carcinoma. Cancer 2001;91:1479–86.
9. Yamamoto J, Okada S, Shimada K, et al. Treatment strategy for small hepatocellular carcinoma: comparison of long-term results after percutaneous ethanol injection therapy and surgical resection. Hepatology 2001;34:707–13.
10. Poon RT, Fan ST, Lo CM, Liu CL, Wong J. Long-term survival and pattern of recurrence after resection of small hepatocellular carcinoma in patients with preserved liver function: implications for a strategy of salvage transplantation. Ann Surg 2002;235:373–82.
11. Johnson PJ. The role of serum α-fetoprotein estimation in the diagnosis and management of hepatocellular carcinoma. Clin Liver Dis 2001;5:145–59.
12. Nakakura EK, Choti MA. Management of hepatocellular carcinoma. Oncology (Huntingt) 2000;14:1085–98.
13. Zhang BH, Yang BH, Tang ZY. Randomized controlled trial of screening for hepatocellular carcinoma. J Cancer Res Clin Oncol 2004;130:417–22.
14. Taketa K. α-Fetoprotein: reevaluation in hepatology. Hepatology 1990;12:1420–32.
15. Kim JW, Ye Q, Forgues M, et al. Cancer-associated molecular signature in the tissue samples of patients with cirrhosis. Hepatology 2004;39:518–27.
16. Ye QH, Qin LX, Forgues M, et al. Predicting hepatitis B virus-positive metastatic hepatocellular carcinomas using gene expression profiling and supervised machine learning. Nat Med 2003;9:416–23.
17. Soresi M, Magliarisi C, Campagna P, et al. Usefulness of α-fetoprotein in the diagnosis of hepatocellular carcinoma. Anticancer Res 2003;23:1747–53.
18. Livak KJ, Schmittgen TD. Analysis of relative gene expression data using real-time quantitative PCR and the 2(-ΔΔC(T)) method. Methods 2001;25:402–8.
19. Tibshirani R, Hastie T, Narasimhan B, Chu G. Diagnosis of multiple cancer types by shrunken centroids of gene expression. Proc Natl Acad Sci U S A 2002;99:6567–72.
20. Ikematsu S, Yano A, Aridome K, et al. Serum midkine levels are increased in patients with various types of carcinomas. Br J Cancer 2000;83:701–6.
21. Kato M, Shinozawa T, Kato S, Awaya A, Terada T. Increased midkine expression in hepatocellular carcinoma. Arch Pathol Lab Med 2000;124:848–52.
22. Lai EC, Fan ST, Lo CM, Chu KM, Liu CL, Wong J. Hepatic resection for hepatocellular carcinoma. An audit of 343 patients. Ann Surg 1995;221:291–8.
23. Bernfield M, Gotte M, Park PW, et al. Functions of cell surface heparan sulfate proteoglycans. Annu Rev Biochem 1999;68:729–77.
24. Capurro M, Wanless IR, Sherman M, et al. Glypican-3: a novel serum and histochemical marker for hepatocellular carcinoma. Gastroenterology 2003;125:89–97.
25. Hippo Y, Watanabe K, Watanabe A, et al. Identification of soluble NH_2-terminal fragment of glypican-3 as a serological marker for early-stage hepatocellular carcinoma. Cancer Res 2004;64:2418–23.
26. Yamauchi N, Watanabe A, Hishinuma M, et al. The glypican 3 oncofetal protein is a promising diagnostic marker for hepatocellular carcinoma. Mod Pathol 2005;18:1591–8.
27. Kadomatsu K, Tomomura M, Muramatsu T. cDNA cloning and sequencing of a new gene intensely expressed in early differentiation stages of embryonal carcinoma cells and in mid-gestation period of mouse embryogenesis. Biochem Biophys Res Commun 1988;151:1312–8.
28. Muramatsu H, Shirahama H, Yonezawa S, Maruta H, Muramatsu T. Midkine, a retinoic acid-inducible growth/differentiation factor: immunochemical evidence for the function and distribution. Dev Biol 1993;159:392–402.
29. Tsou AP, Chuang YC, Su JY, et al. Overexpression of a novel imprinted gene, PEG10, in human hepatocellular carcinoma and in regenerating mouse livers. J Biomed Sci 2003;10:625–35.