Identification and Validation of a Novel Gene Signature Associated with the Recurrence of Human Hepatocellular Carcinoma

Suk Mei Wang,[1] London Lucien P.J. Ooi,[2] and Kam M. Hui[1]

Abstract

Purpose: To improve the clinical management of human hepatocellular carcinoma (HCC) by accurate identification, at diagnosis, of patients at risk of recurrence after primary treatment for HCC.

Experimental Design: Two clinicopathologic variables available at diagnosis, vascular invasion and cirrhosis, together with molecular profiling using Affymetrix human HG-U133A and HG-U133B oligonucleotide probe arrays, were used to identify recurrent HCC disease.

Results: HCC patients presented clinically at diagnosis with vascular invasion and cirrhosis showed a high rate (78-83%) of developing recurrent disease within 6 to 35 months. In comparison, most of the HCC patients (80-100%) without vascular invasion and cirrhosis remained disease-free. However, the risk of recurrent disease for HCC patients with either vascular invasion or cirrhosis could not be accurately ascertained. Using a pool of 23 HCC patients with either vascular invasion or cirrhosis as training set, a 57-gene signature was derived and could predict recurrent disease at diagnosis, with 84% (sensitivity 86%, specificity 82%) accuracy, for a totally independent test set of 25 HCC patients with either vascular invasion or cirrhosis. On further analysis, the disease-free rate was significantly different between patients that were predicted to recur or not to recur in the test group ($P = 0.002$).

Conclusion: We have presented data to show that by incorporating the status of vascular invasion and cirrhosis available at diagnosis for patients with HCC after partial curative hepatectomy and a novel 57-member gene signature, we could accurately stratify HCC patients with different risks of recurrence.

Hepatocellular carcinoma (HCC) is one of the most lethal human malignant cancers worldwide. Although the highest incidence rates occur in parts of Asia and Africa (1), recent epidemiologic studies have projected an overwhelming increase in the incidence and mortality rates of HCC for the next decade in North America, Europe, and Japan (2). The carcinogenesis of HCC is associated with multiple risk factors and is a long-term, multistep process that is believed to require many contributing factors, including chronic liver disease, viral hepatitis [hepatitis B virus (HBV) or hepatitis C virus (HCV)], hemochromatosis, abuse of alcohol, and exposure to hepatic carcinogens and aflatoxins (1, 3–6). In Singapore, HCC is etiologically associated with HBV and is the main cause of death for HBV carriers (7).

Although early HCC is potentially curable by partial hepatectomy, unfortunately only a minority of patients is amenable to surgical resection due to the asymptomatic feature of HCC progression. Moreover, HCC has high incidence of metastases and postsurgical recurrence (8) and, consequently, the prognosis of HCC continues to be dismal. The heterogeneous nature of human HCC has also limited the usefulness of conventional clinicopathologic features such as the status of cirrhosis and vascular invasion available at diagnosis for both treatment and prediction of disease outcome. Often, the disease outcomes for patients with similar tumor stage disease (tumor-node-metastasis classification) could be very different. Hence, a better understanding of the clinicopathologic features and molecular aspects of HCC, as well as the ability to assess and stratify patients with different risks of disease recurrence at diagnosis, would be extremely beneficial for the clinical management of HCC. In this report, we have systematically analyzed and combined some of the conventional clinicopathologic features available at diagnosis to segregate HCC patients with different risk of developing recurrence along with the use of molecular expression profiling to identify genes closely associated with recurrent disease after curative hepatectomy to improve the prediction of HCC recurrence at diagnosis for patients who have 50% risk of developing disease recurrence.

Materials and Methods

Patients. Cancerous and some of the corresponding distal noncancerous liver tissues were obtained from 80 patients who underwent

Authors' Affiliations: [1]Bek Chai Heah Laboratory of Cancer Genomics, Division of Cellular and Molecular Research, Humphrey Oei Institute of Cancer Research and [2]Division of Surgical Oncology, National Cancer Centre, Singapore, Singapore

Received 9/11/06; revised 7/5/07; accepted 7/30/07.

Grant support: National Medical Research Council and Biomedical Research Council of Singapore.

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/).

Requests for reprints: Kam M. Hui, Division of Cellular and Molecular Research, National Cancer Centre, 11 Hospital Drive, Singapore 169610, Singapore. Phone: 65-6436-8337; Fax: 65-6226-3843; E-mail: cmrhkm@nccs.com.sg.

doi:10.1158/1078-0432.CCR-06-2236

partial hepatectomy as a treatment for HCC. Tissues were immediately snap frozen and stored in liquid nitrogen. Before gene profiling or real-time PCR analyses, frozen sections of all the tumor samples studied were stained with H&E and evaluated by qualified pathologists to ascertain that the tissues studied were either cancerous (>80% cancerous tissues) or normal.

All tissue samples used in this study were approved and provided by the Tissue Repository of the National Cancer Centre, Singapore, and conducted in accordance with the policies of its ethics committee. Informed consent was obtained from all participating patients. Linked clinical and histopathologic data were collected from medical records for all the patients who contributed tumor specimens and were rendered anonymous to protect patient confidentiality. To assess recurrence, all treated HCC patients were monitored by routine clinical follow up once every 3 months. The level of α-fetoprotein (AFP), for liver function tests, were determined every 3 months, and ultrasound scans were done every 6 months. Computed tomography scans were done when the level of AFP was >400 ng/mL or when the ultrasound results suggested the presence of recurrent disease. Biopsies of histologically normal liver tissues of 10 colorectal cancer patients who have liver metastases were used as reference normal liver tissues.

Detailed clinicopathologic information of the human HCC samples studied was provided in Supplementary Table S1.

Isolation of total RNA and Affymetrix gene chips experiments. Total RNA was extracted from frozen HCC biopsies using TRIzol reagent according to the manufacturer's protocol (Invitrogen). Five micrograms of total RNA were reversibly transcribed and biotin labeled. The final cRNA obtained was used to hybridize to human HG-U133A and HG-U133B oligonucleotide probe arrays (Affymetrix) as described previously (9, 10). The combined HG-U133A and HG-U133B gene chips contain 44,760 probe sets representing 39,000 transcripts derived from ~33,000 well-defined human genes. The signal intensity of all the arrays studied was normalized to 500. Only genes that had at least 70% present calls in all the samples studied were grouped under a particular clinicopathologic factor and were kept for further analyses. All data generated by Affymetrix Microarray Suite version 5.0 in cel file format were refined using Genedata Expressionist Refiner software package (Genedata GmbH).

Statistical analysis. Correlation between clinicopathologic features and recurrence was done using the Statistical Package for the Social Sciences for Windows (version 15.0). The association of recurrence and the various clinical variables, including sex, age, presence of HBV or HCV, tumor size, capsulation, AFP level, single or multiple lesions, histologic grading, presence or absence of cirrhosis, and vascular tumor invasion, was examined by univariate regression analyses. Recurrence was entered as a dependent factor whereas the other variables studied were entered as covariate factors into a forward stepwise logistic regression model. A P value of <0.05 was taken as a statistically significant association between two factors studied.

Support vector machines (SVM), sparse linear discriminant analyst (SLD), and K nearest neighbor (KNN) were also used for cross-validation of the training sets and prediction of the independent test sets. Kaplan-Meier survival plots and the log-rank test were incorporated to assess the statistical significance of the predicted groups on recurrence by considering time of recurrence as a variable factor simultaneously. Unsupervised hierarchical clustering algorithm was done with the CLUSTER and TREEVIEW software (11) using mean centered correlation as measurements of similarity and average linkage. These two kinds of software can be accessible through Michael Eisen's laboratory.[3] Genes identified were annotated using the Affymetrix NetAffx GeneOntology analysis system[4] and FATIGO data mining tool.[5]

Quantitative real-time PCR. A total of 17 genes (11 up-regulated and 6 down-regulated) that gave the highest median fold changes from the expression studies with the Affymetrix HG-U133A and HG-U133B probe arrays were chosen from the 57-member gene signature for validation with quantitative real-time PCR. A total of seven histologically normal liver tissues from colorectal patients with liver metastases, seven distal histologically normal surrounding tissues of HCC patients, cancerous tissues from 10 nonrecurrent patients, and 11 recurrent HCC patients were used for quantitative real-time PCR. The recurrent and nonrecurrent tumor specimens were from HCC patients with either vascular invasion or cirrhosis (characteristic of subgroups 2 and 3) and were selected randomly from both the training and test sets of subgroups 2 and 3. All PCR primers were designed to have melting temperatures ranging from 58°C to 60°C.

Each sample was amplified in duplicates using Rotor-Gene 2000 Real Time Cycler (Corbett Research). All reactions were carried out with 45 cycles (94°C, 15 s; 55°C, 30 s; 72°C, 30 s). The relative level of expression of each gene studied was determined by normalizing the average copy concentration of each gene against the corresponding average concentration of the 18S rRNA. Statistical analyses were done with parametric test (unpaired t test) and nonparametric test (Mann-Whitney test) using GraphPad Prism 3.0 software.

Additional microarray information. The details of the microarray study followed the Minimum Information About a Microarray Experiment guidelines issued by the Microarray Gene Expression Data Group (12). The microarray data have been deposited in the European Bioinformatics Institutes of the European Molecular Biology Laboratory[6] and are accessible through ArrayExpress public database with accession numbers E-MEXP-84 and E-TABM-292. Information is also available from the authors on request.

Results

Cirrhosis and vascular invasion are the only two clinicopathologic features found to be associated with the recurrence of HCC. HCC is a malignant disease that is commonly fatal mainly because of its high incidence of metastases and postsurgical recurrence. In an attempt to identify clinicopathologic variables that could be of significant diagnostic value to predict disease recurrence, we have evaluated several common clinicopathologic features that are available at diagnosis for HCC and correlated them with the recurrence status of 80 HCC patients. All clinically well-characterized HCC patients were included in our study, without focusing on selected subgroups of patients, to ensure that the results obtained after the analyses were representative for the overall HCC patients in our population. The detailed clinicopathologic information of the human HCC patients of whom the liver tissues were used in this study was summarized in Supplementary Table S1.

It was observed that male patients were predominant, with its ratio to female of 4.3:1 (81.2% males compared with 18.8% females). Of all the HCC patients studied, 77.5% were HBV^+, 3.7% were HCV^+, and 18.8% were negative for HBV and HCV infection. The tumor of 39 patients (48.75%) was either partially or completely encapsulated; 53.7% (n = 43) of patients had cirrhosis; and the majority of the patients (86.2%, n = 69) had a single nodule. The median size of all resected tumor was 5 cm (range 1.2-17 cm), and the median AFP level was 25.6 ng/mL (range 1.2-70,700 ng/mL). According to the Edmondson grading system, half of the cases (55%; n = 44) were classified as grade 2; 12.5% (n = 10) were classified as

[3] http://rana.Stanford.EDU/software/
[4] http://www.affymetrix.com/site/login/login.affx
[5] http://fatigo.bioinfo.cnio.es/
[6] http://www.ebi.ac.uk/arrayexpress/

grade 1; 25% (n = 20) were classified as grade 3; and the rest (6.25%; n = 5) had mixed histologic grading from grade 1 to grade 4. Vascular invasion occurred in 36.2% (n = 29) of the cases and the median time of recurrence was 4.5 months (range <1-35 months).

The clinicopathologic features were used to determine if they are significantly associated with the recurrence status of 80 HCC patients divided into the duration of 6, 12, 18, and 35 months after hepatectomy resection, and the results were summarized in Table 1. Multivariate analysis ascertained that vascular invasion and cirrhosis are the two independent prognostic factors available at diagnosis that are significantly associated with the recurrence of HCC irrespective of the various durations of recurrence examined (Table 1). The detailed statistical analyses of the association between all the above clinicopathologic features and the recurrence of HCC for the 80 patients in various durations using univariate analysis were provided in Supplementary Table S2A to S2D. These results also showed that the recurrence of HCC was strongly and significantly associated with vascular invasion irrespective of the duration of recurrence, and was weakly associated with cirrhosis. No other clinicopathologic variable was found to be associated with the recurrence of HCC although patients with recurrence apparently had larger tumor size (median 6 cm) and elevated serum concentration of AFP (median 41.35 ng/mL) compared with patients without recurrence (Supplementary Table S2A).

The clinicopathologic features of cirrhosis and vascular invasion are not sufficient to predict recurrent disease for all patients with HCC. Initially, we selected 396 genes that were statistically different between all the recurrent (n = 39) and nonrecurrent (n = 41) HCC samples and tested their ability to hierarchically cluster the 39 recurrent HCC samples according to their clinicopathologic features. The results showed that these genes failed to segregate the recurrent HCC samples in accordance with their clinicopathologic features (Supplementary Fig. S1).

Because vascular invasion and cirrhosis are the two clinicopathologic features that are strongly associated with the recurrence of HCC (Table 1), we therefore exploited their ability to predict HCC recurrence. The pool of 80 HCC patients were therefore randomly divided into training and test sets according to their status of recurrence, vascular invasion, or cirrhosis, and were analyzed independently. Three independent gene sets derived separately from statistical analysis between test sets of recurrent (n = 20) and nonrecurrent (n = 22) HCC samples, of HCC samples with (n = 18) and without (n = 33) vascular invasion, and of cirrhotic (n = 25) and noncirrhotic (n = 20) HCC samples were unable to yield high predictive power for recurrent disease for corresponding independent test sets (Supplementary Table S3).

Next, using the clinicopathologic features of vascular invasion and cirrhosis, we arbitrarily divided the 80 HCC patients studied into four subgroups by jointly considering the presence or absence of cirrhosis and vascular invasion. Specifically, HCC patients who have vascular invasion and cirrhosis at diagnosis are subgroup 1, whereas HCC patients who have no vascular invasion and are noncirrhotic at diagnosis are subgroup 4. Patients belonging to subgroups 2 and 3 were individuals, at diagnosis, with either vascular invasion but no cirrhosis, or with cirrhosis but no vascular invasion, respectively.

For HCC patients who recurred within 6 months (n = 63), it was observed that patients who had vascular invasion and cirrhosis at diagnosis (subgroup 1) were most likely to recur (78%), whereas none of the patients who had no vascular invasion and were noncirrhotic at diagnosis (subgroup 4) had recurrence disease (Fig. 1). Similar patterns could be observed when the duration of recurrence was extended to within 12 months (n = 74), 18 months (n = 76), and 35 months (n = 80; Fig. 1). Patients in subgroup 1 (positive for vascular invasion and cirrhosis at diagnosis) had a high chance of developing recurrent disease compared with patients in subgroup 4 that were negative for vascular invasion and cirrhosis (Fig. 1). Therefore, the combined clinicopathologic features of vascular invasion and the status of cirrhosis were adequately informative for recurrent disease for HCC patients belonging to subgroup 1 (78-83% of patients developed recurrence) and subgroup 4 (89-100% of patients remained disease-free; Fig. 1).

However, as evident from Fig. 1, it was apparently insufficient to estimate the risk of recurrence for HCC patients in subgroups 2 and 3 with the clinicopathologic features of vascular invasion and cirrhosis. From the pool of 80 patients,

Table 1. Multivariate analyses of the clinicopathologic variables available at diagnosis associated with HCC recurrence with different durations of disease recurrence

	P*			
	***R* ≤ 6 mo**	**R ≤ 12 mo**	***R* ≤ 18 mo**	***R* ≤ 35 mo**
Sex (men/women)	0.196	0.096	0.073	0.040
Age (median; y)	0.179	0.919	0.914	0.782
Viral infection (HBV/HCV/non-B non-C)	0.922	0.872	0.968	0.760
Capsulation (yes/no)	0.548	0.698	0.656	0.866
Tumor size (median; cm)	0.609	0.755	0.796	0.379
AFP level (median; ng/mL) {low/medium/high}	0.092	0.177	0.215	0.215
Lesion (single/multiple nodules)	0.930	0.572	0.487	0.659
Histologic grading (G1/G2/G3/G2 to G4)	0.983	0.913	0.837	0.567
Cirrhosis (yes/no)	0.038	0.018	0.014	0.011
Invasion (yes/no)	0.001	0.0001	0.0001	0.0001

NOTE: Non-B non-C, negative for both HBV and HCV antigen; {low}, <10 ng/mL; {medium}, 10 to 300 ng/mL; {high}, >300 ng/mL; G1, well differentiated; G2, moderately differentiated; G3, poorly differentiated; G4, extremely poor differentiation; *R*, duration of recurrence.
*P values were calculated using multivariate Cox regression analysis.

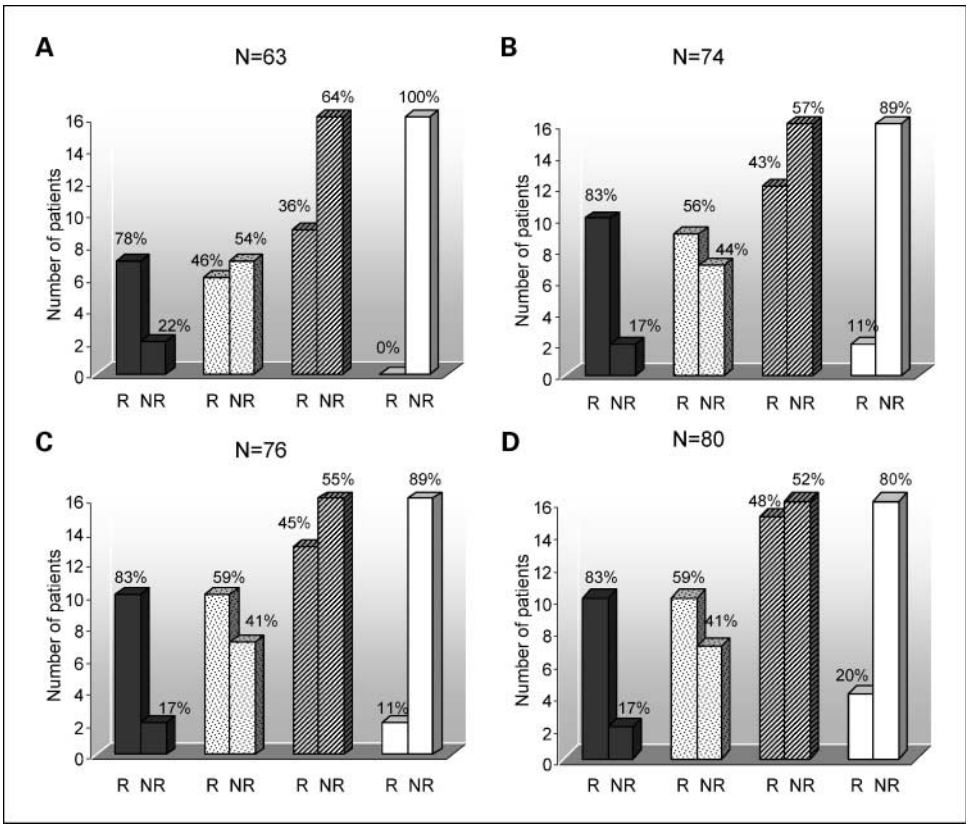

Fig. 1. The use of two clinicopathologic risk factors, cirrhosis and vascular invasion, to segregate HCC patients with recurrent disease less than 6 mo (*A*), 12 mo (*B*), 18 mo (*C*), and 35 mo (*D*), after surgical resection. R, recurrence; NR, nonrecurrence. ■, HCC patients with vascular invasion and cirrhosis; ▨, HCC patients with vascular invasion but without cirrhosis; ▨, HCC patients without vascular invasion but with cirrhosis; □, HCC patients who have neither vascular invasion nor cirrhosis.

subgroup 2 composed of 17 patients with vascular invasion but without cirrhosis, and subgroup 3 consisted of 31 cirrhotic HCC patients without vascular invasion (Fig. 1D). Hence, the clinicopathologic features of vascular invasion and the status of cirrhosis were not clinically useful, at diagnosis, to estimate the risk of recurrence for HCC patients in subgroups 2 and 3.

Identification of a novel molecular gene signature to accurately predict the recurrence of HCC for patients with vascular invasion and no cirrhosis (subgroup 2) and patients with cirrhosis and no vascular invasion (subgroup 3). Forty-eight of the 80 HCC patients that we studied belonged to subgroups 2 and 3. In an attempt to derive a molecular signature to predict recurrent disease for these HCC patients, these group of 48 patients were randomly divided into training (n = 23) or test (n = 25) sets. Parametric (Student's t test) and nonparametric (Wilcoxon test) statistical tests of the Genedata Expressionist Analyst

Table 2. The 57-member gene signature identified is able to predict recurrent HCC disease for HCC patients in subgroups 2 and 3

Duration of recurrence (mo)	No. HCC patients in training set	Accuracy of estimating training set	No. HCC patients in test set	Accuracy of predicting test set (sensitivity, specificity)
6	7R, 12NR	SVM-100%	8R, 11NR	SVM-79% (75%, 82%)
		SLD-100%		SLD-74% (75%, 73%)
		KNN-100%		KNN-84% (87.5%, 82%)
12	8R, 12NR	SVM-100%	13R, 11NR	SVM-83% (85%, 82%)
		SLD-100%		SLD-75% (77%, 73%)
		KNN-100%		KNN-83% (85%, 82%)
18	10R, 12NR	SVM-100%	13R, 11NR	SVM-83% (85%, 82%)
		SLD-100%		SLD-75% (77%, 73%)
		KNN-100%		KNN-83% (85%, 82%)
35	11R, 12 NR	SVM-100%	14R, 11NR	SVM-84% (86%, 82%)
		SLD-100%		SLD-76% (78.6%, 73%)
		KNN-96%		KNN-84% (86%, 82%)

NOTE: Subgroup 2 comprises patients with invasion but without cirrhosis. Subgroup 3 comprises patients with cirrhosis but without invasion. The 57-gene set was selected based on two main steps: (*a*) *P* value was generated from both parametric (Student's t test) and nonparametric (Wilcoxon test) statistical tests that determined gene ranking. (*b*) Each top-ranked gene set (from top 50, 60, until 500) were further filtered by progressively increasing the median fold changes from 1.25 to 5. The 57 genes were derived from the top 130 ranked genes resulting from analyses with fold >1.5.
Abbreviations: R, recurrence; NR, nonrecurrence.

(version 1.0) were used to select genes that could discriminate between recurrent disease (n = 11) and nonrecurrent disease (n = 12) in the training set. The genes obtained were ranked using their P values. The top 500 probe sets with the most significant P values were selected. To generate additional gene sets to cross-validate the training set, probe sets were further filtered from this gene list by progressively selecting those with increased median fold changes from 1.25 to 5. The gene set that gave the lowest error rate for the cross-validation in the training group was eventually used to estimate the probability of recurrence for an independent test group (n = 14 for recurrent disease and n = 11 for nonrecurrent disease) of HCC patients. The 61 probe sets (57 genes), selected from the 130 top-ranked probe sets with median fold change >1.5, could cross-validate the training set with the best estimation score ranging from 96% to 100% using all three different classifier algorithms including SVM, SLD, and KNN (K = 3; Table 2). Most importantly, when this 57-member gene signature was used to predict recurrent disease for a totally independent test set of 14 HCC patients with recurrence and 11 HCC patients without recurrence, the accuracy was 84% (sensitivity 86%; specificity 82%) using SVM and KNN (K = 3) as classifier algorithms (Table 2). This gene cluster of 61 probe sets (57 genes) identified is tentatively designated as the HCC recurrence prediction molecular gene signature.

When Kaplan-Meier plots and log-rank tests were used to assess the outcome of the patients from these two groups that were predicted to have recurrence and nonrecurrence (14 recurrent,

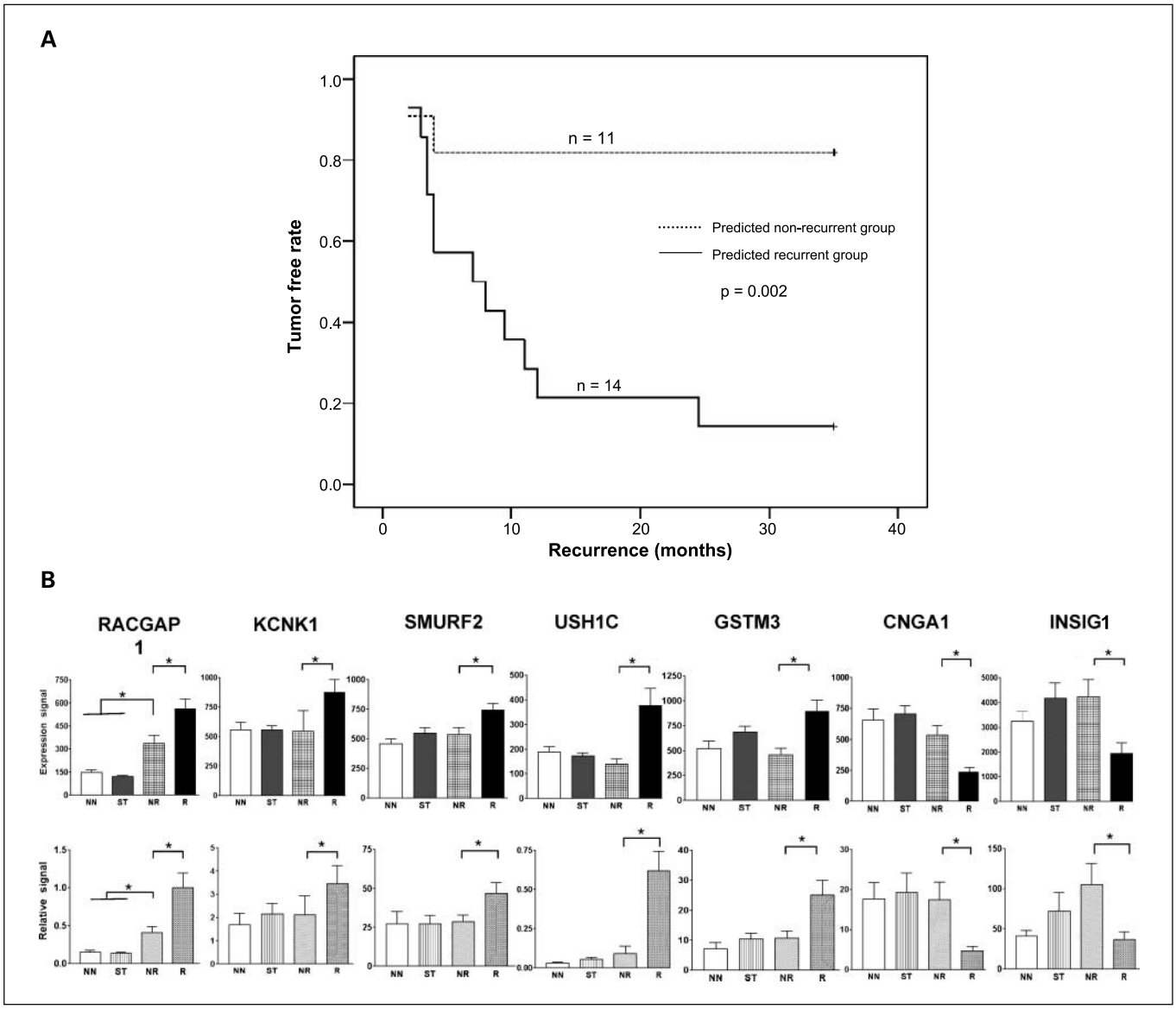

Fig. 2. *A,* Kaplan-Meier analysis done to validate the statistical significance of the outcomes on prediction of recurrence. A log-rank test of predicted groups of recurrence of HCC patients in subgroups 2 and 3 in an independent test set classified by SVM revealed statistically significant difference (P = 0.002) between patients predicted to have recurrence and no recurrence. *B,* level of gene expression of five representative up-regulated genes and two representative down-regulated genes selected from the 57-member HCC recurrence-associated gene set studied by Affymetrix gene chips (*top*) and quantitative real-time PCR (*bottom*). ■ and ▨, HCC recurrence; ▦ and ▧, nonrecurrence; ▨ and ▥, matched normal surrounding tissue (*ST*); □, normal liver tissues from patients with colon cancer metastases to liver (*NN*). Columns, mean; bars, SE. *, $P < 0.05$.

11 nonrecurrent) using the duration of being disease-free after resection as time factor, statistically significant difference between these two groups could be obtained ($P = 0.002$; Fig. 2A). In addition, when multivariate analysis was done, this 57-member gene signature was found to be confounded by the outcome of the predicted recurrence but not by other clinical factors (Supplementary Table S5). These results further support that the 57-member gene set could effectively distinguish patients with and without recurrence in the cohort of HCC patients that present clinically, at diagnosis, with either vascular invasion but no cirrhosis, or have cirrhosis but no invasion.

This gene signature was further used to predict disease recurrence for the independent sets of HCC patients in subgroups 2 and 3 after segregating them according to the duration when recurrent disease occurred within 6 months (8 recurrent HCC), 12 months (13 recurrent HCC), and 18 months (13 recurrent HCC). The accuracy achieved was consistently within the range of 74% and 83% irrespective of the use of any of the three classifier algorithms, SVM, SLD or KNN (Table 2). The sensitivity (rate of recurrence cases correctly predicted) and specificity (rate of nonrecurrence cases correctly predicted) of all the tests done also consistently ranged between 73% and >80% (Table 2). Furthermore, when the predictability of this gene signature was investigated for HCC patients in subgroups 1 and 4, the prediction scores achieved using all three classifiers were below 80% (Supplementary Table S4), and were not as accurate compared with the extreme distribution of HCC recurrence in each of subgroup 1 and 4 segregated by using the two clinicopathologic features, cirrhosis and vascular invasion (Fig. 1).

Validation of representative genes within the gene signature by real-time PCR. To further validate the differential expression pattern of genes in the HCC recurrence prediction molecular signature, a total of 17 representative genes, 11 up-regulated and 6 down-regulated, which gave the highest median fold changes as detected by Affymetrix gene chips, were chosen from the gene signature and validated by quantitative real-time PCR. The HCC samples tested by real-time PCR consisted of seven samples of distal normal surrounding tissues of cancer, seven samples of liver tissues from non-HCC patients, 10 samples of liver tissues from nonrecurrent HCC patients, and 11 samples of liver tissues from recurrent HCC patients were also used in Affymetrix gene chip analysis. Results obtained by real-time PCR analyses correlated well with those of the Affymetrix gene chips for all the 17 genes tested. Figure 2B showed the quantitative real-time PCR results obtained and showed that expression of all the seven genes tested could be validated to be significantly different ($P \leq 0.05$) between recurrent and nonrecurrent HCC samples. Similar to results obtained with the Affymetrix gene chips, *RACGAP1*, *KCNK1*, *SMURF2*, *USH1C*, and *GSTM3* were up-regulated in the recurrent HCC liver samples whereas *CNGA1* and *INSIG1* were down-regulated (Fig. 2B).

Functional annotation of genes in the HCC recurrence prediction molecular gene signature. The HCC recurrence prediction molecular gene signature comprises of 57 genes that are associated with a wide variety of cellular functions, including cell growth and maintenance, DNA replication and cellular metabolism, transcription and protein processing, cellular signaling, transport, immune regulators, and apoptosis (Table 3). When the expression of the 57 genes was compared with distal normal tissues of the same HCC patient and liver tissues from non-HCC patients, the 57 genes could be conveniently clustered into four distinct groups according to whether they are up-regulated or down-regulated in recurrence HCC disease (Fig. 3). Figure 3A shows the 23 probe sets (22 genes) that are differentially up-regulated in recurrent HCC patients of subgroups 2 and 3 but not in nonrecurrent HCC patients and normal controls. Figure 3B shows seven genes that are specifically down-regulated in nonrecurrent HCC patients in subgroups 2 and 3. Figures 3C and D, on the other hand, show genes (19 probe sets and 16 genes) that are differentially down-regulated in HCC patients with recurrence, and the 12 genes that are specifically up-regulated in HCC patients without recurrence of subgroups 2 and 3, respectively.

Discussion

Although great advancements have been made in the surgical treatment of HCC, recurrence after surgery remains a key challenge for the clinical management of HCC patients (8, 13).

Table 3. Functional annotation of the 57 genes identified in the HCC recurrence-associated gene signature

Up-regulated genes	Gene symbol	Down-regulated genes	Gene symbol
Cell cycle	*PARD3*	Cell proliferation	*INSIG1*
	MPHOSPH9		
	CUL4B	Cell cycle	*SKP2*
	CETN2		
		Cell growth and maintenance	*ETS2*
Signaling	*USH1C*		
	PTPN11	Immune/defense	*OASL*
	ARL5		
	SMURF2	Signaling	*CDC42SE1*
			CSNK1G3
Transport	*RACGAP1*	Transport	*CYP17A1*
	DNAJC10		*CDO1*
	KCNK1		*CNGA1*
			NUDT9
Development	*GSTM3*		*SLC16A5*
			SARA2
Translation	*SR3*		
		Development	*ID2*
Metabolism/ biosynthesis	*MTHFD2L*		
	SYNGR2	Proteolysis/ peptidolysis	*RNF130*
	PRL10A		
		Transcription	*MAFB*
Ubiquitination	*FLJ23749*		*ELL2*
Nucleic acid binding	*LARP6*	Protein processing	*PTP4A1*
	R3HDM		*PGGT1B*
	ZNF652	Metabolism/ biosynthesis	*HEXB*
			SLC27A2
ATP/ion/protein binding	*ZBTB38*		*AGPAT3*
	SH3GLB2		*GCH1*
			SPACA3
Miscellaneous	*LOC159090*		*AS3MT*
	FLJ11016		
	LOC285550	Nucleic acid binding	*TIGD2*
	NARG2	ATP/ion/protein binding	*KIAA0676*
	FAM33A	Miscellaneous	*LMBRD2*
			MARVELD

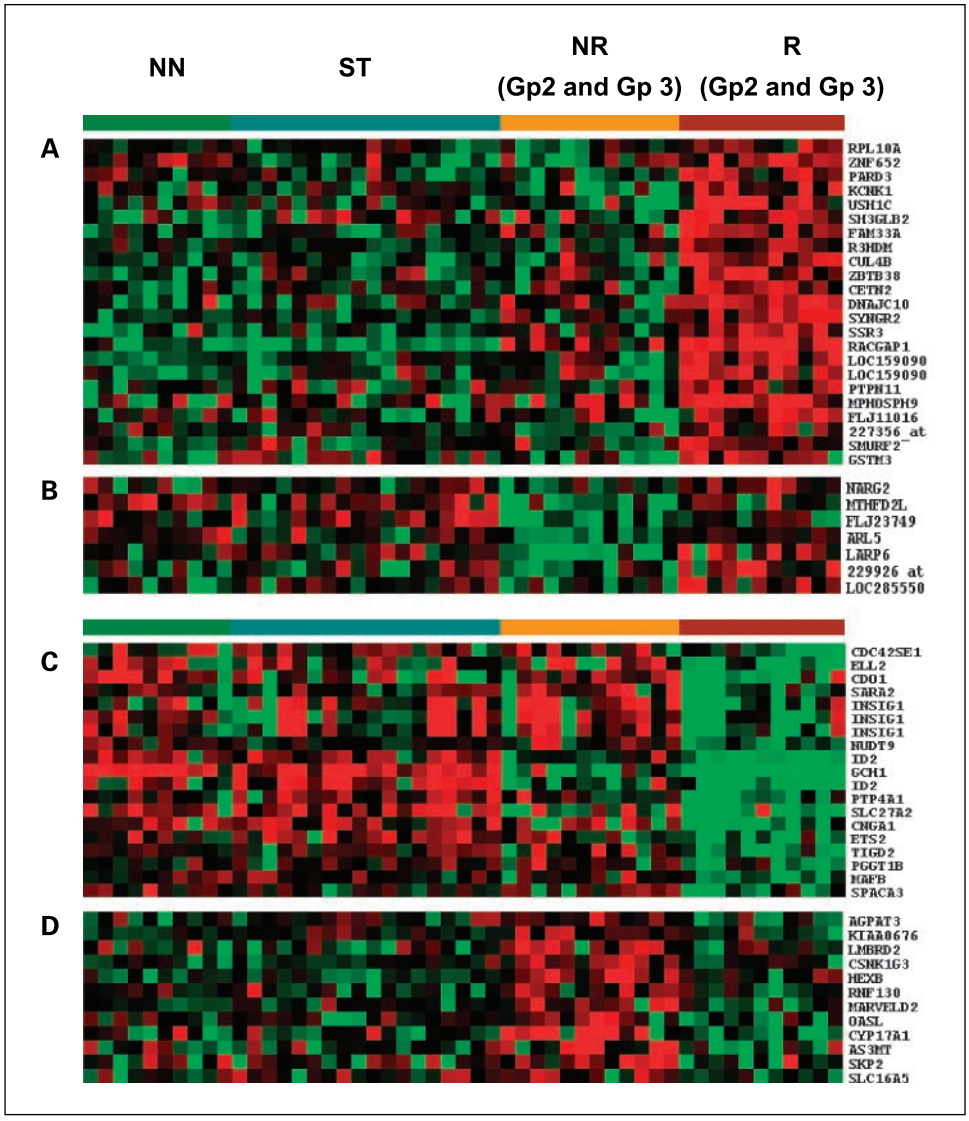

Fig. 3. Hierarchical clustering of the 57-member gene set that could best predict recurrent disease for HCC patients in groups 2 and 3. *A,* 22 genes with the highest level of expression in recurrent HCC samples. *B,* seven genes with distinctive lower level of expression in nonrecurrent HCC samples. *C,* 16 genes that are down-regulated in the recurrent HCC. *D,* 12 genes with higher expression in the nonrecurrent HCC. R, HCC patients with recurrence; NR, HCC patients without recurrence; Gp2, subgroup 2 patients with invasion but without cirrhosis; Gp3, subgroup 3 patients without invasion but with cirrhosis; ST, matched normal surrounding tissue; NN, normal liver tissues from patients with colon cancer metastases to liver. Red, up-regulation; black, no change; green, down-regulation.

There have been several recent reports on applying molecular gene profiling to segregate HCC from normal tissues (14) and to discriminate patients according to etiologic factors (15) and disease state (16, 17). However, the clinical usefulness of these studies is limited, and what is urgently needed is the ability to accurately identify, at diagnosis, patients at risk for recurrence after curative hepatectomy for primary HCC. The ability to predict the clinical course of HCC would enable the implementation of optimal clinical treatments and follow-up strategies before clinical manifestation of recurrent disease.

When the clinicopathologic features of recurrence, vascular invasion, and cirrhosis were used independently to estimate the risk of recurrence of all the HCC patients studied, the gene signatures derived failed to yield a satisfied prediction score of disease recurrence for independent test sets of HCC patients (Supplementary Table S3). Although we could show that HCC patients with vascular invasion and cirrhosis at diagnosis (subgroup 1) are most likely (78-83%) to develop recurrent disease and, in contrast, HCC patients who had no vascular invasion and were noncirrhotic at diagnosis (subgroup 4) are unlikely (0-20%) to have recurrent disease (Fig. 1), the two clinicopathologic features of vascular invasion and the status of cirrhosis on their own were not sufficient to identify recurrent disease for all HCC patients. It is evident from Fig. 1 that the recurrence status for HCC patients belonging to subgroups 2 and 3 who were, at diagnosis, with either vascular invasion but no cirrhosis, or with cirrhosis but no vascular invasion, respectively, could not be ascertained at high confidence. We have therefore further studied 48 HCC patients belong to subgroups 2 and 3 using Affymetrix HG-U133A and HG-U133B gene chips along with some of their corresponding normal tissues to identify biomarkers associated with recurrence. The combined Affymetrix HG-U133A and HG-U133B gene chips contain 44,760 probe sets representing 39,000 transcripts derived from ~33,000 well-defined human genes. By processing the probe level data obtained and analyzed with multiple statistical analyses, a 57-member molecular signature was derived from a training set consisting of 23 HCC patients with either vascular invasion or cirrhosis. This gene signature could subsequently identify, with 84% accuracy, recurrent disease for

a totally independent test set of 25 HCC patients with either vascular invasion or cirrhosis (Table 2). The predictability of this gene signature was specifically derived for patients bearing either cirrhosis or vascular invasion at diagnosis (subgroups 2 and 3). In comparison, its ability to predict recurrent disease for patients in subgroups 1 and 4 yielded relatively poor accuracy as shown in Supplementary Table S4. The reproducibility of this gene signature to predict recurrent HCC disease, however, awaits a second study using a large cohort of patients.

Recently, molecular signatures associated with premalignant lesions (18, 19) and intrahepatic metastatic or recurrent potentials (20–23) of HCC have been reported. Specifically, there were 621, 231, 29, and 37 reported probe sets representing 406, 153, 12, and 20 genes from the studies of Lee et al. (20), Ye et al. (21), Iizuka et al. (22), and Kurokawa et al. (23), respectively. The comparison between these gene sets and the molecular signature in this study revealed that one single gene, the *insulin-induced gene 1*, consistently exist among our 57-member gene signature and that reported by Lee et al. (20), whereas there is no overlapping gene between our 57-member gene signature to that reported by Ye et al., Iizuka et al., and Kurokawa et al.

Within the HCC recurrence-associated gene signature, the expression of 22 genes is specifically elevated in the recurrent HCC samples. These genes are associated with a wide variety of cellular functions (Table 3). The most up-regulated gene in recurrent HCC samples, as determined by both Affymetrix gene chips and real-time PCR assays, was *USH1C*. *USH1C* encodes the gene product harmonin, a PDZ-containing protein. PDZ domains are modular protein interaction domains that play a role in protein targeting and the assembly of large protein complexes involved in signaling or subcellular transport and was the first mutation in a PDZ-encoding gene linked to a human disease (24, 25).

Another gene that is also consistently up-regulated in the recurrent HCC samples is Rac GTPase–activating protein 1. Rho GTPases, which include Rho, Rac, and CDC42, play pivotal roles in regulating the actin cytoskeleton necessary for cell motility, cell-cell contact, and malignant transformation in many cell types (26), are responsible for the regulation of many downstream kinases such as CDC42. Interestingly, CDC42 small effector 1 was also found to be down-regulated in the recurrent HCC samples studied. Moreover, casein kinase that regulates the cytoskeletal organization through small GTPases via the Wnt signaling pathway (27) is also down-regulated in the recurrent HCC samples.

There are seven genes with distinctive lower expression in nonrecurrent HCC patients compared with recurrence groups and these include *NARG2*, *MTHFD2L*, *FLJ23749*, *ARL5*, *LARP6*, *LOC285550*, and *229926_at* (Fig. 3B). Twelve genes that are markedly up-regulated in the nonrecurrent HCC patients include *CYP17A1*, *LMBRD2*, *SKP2*, *OASL*, *AS3MT*, *MARVELD2*, *AGPAT3*, *KIAA0676*, *HEXB*, *CSNK1G3*, *SLC16A5*, and *RNF130* (Fig. 3D). Many of these genes presently do not have sufficient annotation (NARG2, 229926_at, LOC285550, LMBRD2, MARVELD2, and KIAA0676). Expression of the detoxification enzyme cytochrome P450 and enzymes involved in metabolism (AS3MT and HEXB) that are predominantly expressed in differentiated hepatocytes were down-regulated in recurrent HCC samples. The suppression of these genes could reflect tumor dedifferentiation after the progression of malignancy. On the other hand, the gene that is noticeably up-regulated in recurrent HCC samples is La ribonucleoprotein domain family, member 6 (*LARP6*), a RNA-binding protein.

In summary, our present study describes an effective strategy to identify HCC recurrence with an overall accuracy above 80% for a cohort of HCC patients that is 77.5% HBV^+, 3.7% HCV^+, and 18.8% non-HBV, and non-HCV at diagnosis by incorporating a novel 57-member molecular gene signature to predict recurrent disease for HCC patients that had either vascular invasion or cirrhosis and the use of clinicopathologic features to estimate extreme HCC recurrence outcome. This 57-member gene set would serve as a pool of lead gene targets for the identification and development of novel diagnostic and therapeutic biomarkers to greatly improve the clinical management of HCC patients with different risks of recurrence after curative partial hepatectomy.

Acknowledgments

We thank the National Cancer Centre Tissue Repository for providing human tissue specimens for this study and Yu Kun for his help with the statistical analysis.

References

1. Schafer DF, Sorrell MF. Hepatocellular carcinoma. Lancet 1999;353:1253–7.
2. Bonn D. Hepatocellular carcinoma on the increase in USA. Lancet 1999;353:989.
3. Evans AA, Chen G, Ross EA, et al. Eight-year follow-up of the 90,000-person Haimen city cohort: I. Hepatocellular carcinoma mortality, risk factors, and gender differences. Cancer Epidemiol Biomarkers Prev 2002; 11:369–76.
4. Block TM, Mehta AS, Fimmel CJ, et al. Molecular viral oncology of hepatocellular carcinoma. Oncogene 2003;22:5093–107.
5. Iizuka N, Oka M, Yamada-Okabe H, et al. Comparison of gene expression profiles between hepatitis B virus- and hepatitis C virus-infected hepatocellular carcinoma by oligonucleotide microarray data on the basis of a supervised learning method. Cancer Res 2002;62: 3939–44.
6. Mas VR, Maluf DG, Stravitz R, et al. Hepatocellular carcinoma in HCV-infected patients awaiting liver transplantation: genes involved in tumor progression. Liver Transpl 2004;10:607–20.
7. Seow A, Koh WP, Chia KS, Shi LM, Lee HP, Shanmugaratnam K. Trends in cancer incidence in Singapore 1968-2002. Singapore Cancer Registry 2004; Report No.6:100–1.
8. Llovet JM, Burroughs A, Bruix J. Hepatocellular carcinoma. Lancet 2003;362:1907–17.
9. Linn YC, Wang SM, Hui KM. Comparative gene expression profiling of cytokine-induced killer cells in response to acute myeloid leukemic and acute lymphoblastic leukemic stimulators using oligonucleotide arrays. Exp Hematol 2005;33:671–81.
10. Tan MG, Ooi LL, Aw SE, Hui KM. Cloning and identification of hepatocellular carcinoma down-regulated mitochondrial carrier protein, a novel liver-specific uncoupling protein. J Biol Chem 2004;279: 45235–44.
11. Eisen MB, Spellman PT, Brown PO, Botstein D. Cluster analysis and display of genome-wide expression patterns. Proc Natl Acad Sci U S A 1998;95: 14863–8.
12. Brazma A, Hingamap P, Quackenbush J, et al. Minimum information about a microarray experiment (MIAME)—toward standards for microarray data. Nat Genet 2001;29:365–71.
13. Thomas MB, Zhu AX. Hepatocellular carcinoma: the need for progress. J Clin Oncol 2005;23:2892–9.
14. Xu XR, Huang J, Xu ZG, et al. Insight into hepatocellular carcinogenesis at transcriptome level by comparing gene expression profiles of hepatocellular carcinoma with those of corresponding noncancerous liver. Proc Natl Acad Sci U S A 2001;98:15089–94.
15. Okabe H, Satoh S, Kato T, et al. Genome-wide analysis of gene expression in human hepatocellular carcinomas using cDNA microarray: identification of genes involved in viral carcinogenesis and tumor progression. Cancer Res 2001;61:2129–37.
16. Chen X, Cheung ST, So S, et al. Gene expression patterns in human liver cancers. Mol Biol Cell 2002; 13:929–39.
17. Chuma M, Sakamoto M, Yamazaki K, et al. Expression profiling in multistage hepatocarcinogenesis: identification of HSP70 as a molecular marker of early hepatocellular carcinoma. Hepatology 2003;37: 198–207.

18. Kim JW, Ye QH, Forgues M, et al. Cancer-associated molecular signature in the tissue samples of patients with cirrhosis. Hepatology 2004;39:518–27.
19. Zindy P, Andrieux L, Bonnier D, et al. Upregulation of DNA repair genes in active cirrhosis associated with hepatocellular carcinoma. FEBS Letters 2005; 579:95–9.
20. Lee JS, Chu IS, Heo J, et al. Classification and prediction of survival in hepatocellular carcinoma by gene expression profiling. Hepatology 2004;40:667–76.
21. 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.
22. Iizuka N, Oka M, Yamada-Okabe H, et al. Oligonucleotide microarray for prediction of early intrahepatic recurrence of hepatocellular carcinoma after curative resection. Lancet 2003;361:923–9.
23. Kurokawa Y, Matoba R, Takemasa I, et al. Molecular-based prediction of early recurrence in hepatocellular carcinoma. J Hepatol 2004;41:284–91.
24. Hung AY, Sheng M. PDZ domains: structural modules for protein complex assembly. J Biol Chem 2002; 277:5699–702.
25. Bitner-Glindzicz M, Lindley KJ, Rutland P, et al. A recessive contiguous gene deletion causing infantile hyperinsulinism, enteropathy and deafness identifies the Usher type 1C gene. Nat Genet 2000;26:56–60.
26. Caceres M, Guerrero J, Martinez J. Overexpression of RhoA-GTP induces activation of the epidermal growth factor receptor, dephosphorylation of focal adhesion kinase and increased motility in breast cancer cells. Exp Cell Res 2005;309:229–38.
27. Cong F, Schweizer L, Varmus H. Casein kinase Iε modulates the signaling specificities of disheveled. Mol Cell Biol 2004;24:2000–11.