Gene Expression – Based Recurrence Prediction of Hepatitis B Virus – Related Human Hepatocellular Carcinoma

Hyun Goo Woo,[5,6] Eun Sung Park,[7] Jae Hee Cheon,[4] Ju Han Kim,[2] Ju-Seog Lee,[8] Bum Joon Park,[1] Won Kim,[1] Su Cheol Park,[1] Young Jin Chung,[1] Byeong Gwan Kim,[1] Jung-Hwan Yoon,[1] Hyo-Suk Lee,[1] Chung Yong Kim,[1] Nam-Joon Yi,[3] Kyung-Suk Suh,[3] Kuhn Uk Lee,[3] In-Sun Chu,[5] Tania Roskams,[9] Snorri S. Thorgeirsson,[6] and Yoon Jun Kim[1]

Abstract

Purpose: The poor prognosis of hepatocellular carcinoma (HCC) is, in part, due to the high rate of recurrence even after "curative resection" of tumors. Therefore, it is axiomatic that the development of an effective prognostic prediction model for HCC recurrence after surgery would, at minimum, help to identify in advance those who would most benefit from the treatment, and at best, provide new therapeutic strategies for patients with a high risk of early recurrence.

Experimental Design: For the prediction of the recurrence time in patients with HCC, gene expression profiles were generated in 65 HCC patients with hepatitis B infections.

Result: Recurrence-associated gene expression signatures successfully discriminated between patients at high-risk and low-risk of early recurrence ($P = 1.9 \times 10^{-6}$, log-rank test). To test the consistency and robustness of the recurrence signature, we validated its prognostic power in an independent HCC microarray data set. CD24 was identified as a putative biomarker for the prediction of early recurrence. Genetic network analysis suggested that SP1 and peroxisome pro-liferator – activated receptor-α might have regulatory roles for the early recurrence of HCC.

Conclusion: We have identified a gene expression signature that effectively predicted early recurrence of HCC independent of microarray platforms and cohorts, and provided novel biolog-ical insights into the mechanisms of tumor recurrence.

Hepatocellular carcinoma (HCC) is one of the most common cancers in the world, accounting for an estimated 600,000 deaths annually (1). The resistance of HCC to existing treatments and the lack of biomarkers for early detection make it one of most hard-to-treat cancers. Surgical tumor resection, including liver transplantation, remains the only curative modality for HCC. Although the progress of surgery and preoperative or postoperative care have improved the survival of patients with HCC, it is disappointing that the recurrence rate remains high even after curative resection of tumors. HCC recurrence is a serious complication following the resection of the primary tumor and occurs in 75% to 100% of patients within 5 years after surgery (2 – 4). In light of the dismal clinical outcome of this neoplasm, the development of effective systems that can predict the likelihood of recurrence is much needed. This will help in deciding therapeutic strategies for patients with HCC according to the predicted risk of recurrence.

Several attempts have been made to predict recurrence and prognostic outcomes based on single or multiple clinicopath-ologic features such as the severity of the liver function, age, tumor grade, size, microvascular invasion, portal vein throm-bosis, and the presence of microsatellite regions (2, 5 – 7). Prognostic staging systems have also been proposed to stratify patients according to expected survival (8 – 10). However, their prognostic significances and clinical utilities needed to be further validated with large-scale studies (11, 12).

Recent studies on gene expression profiles could successfully predict recurrence, metastasis, or survival prognosis of HCCs (13 – 17). Even though these studies successfully provide prognostic markers for clinical application, the lack of consistency and robustness of predictors generated from different microarray platforms remain one of the major obstacles for the clinical use of microarray-based predictors (18, 19). As the lack of reproducibility mainly comes from the

Fig. 1. Identification of genes responsible for early recurrence of HCC. *A,* hierarchical clustering was done on the expression profile of recurrence genes that were identified by Cox proportional hazards analysis ($P < 0.005$, log-rank test). Before clustering, the average expression levels of each gene were set to 0. HCCs were labeled according to recurrence time as early recurrence (i.e., patients with tumor recurrence within a year after surgery, $n = 15$; *red*) and late recurrence (i.e., patients free of recurrence for >1 year, $n = 25$; *blue*). The patients who had been followed up for less than a year without recurrence were assigned as unclassifiable ($n = 25$, *gray*). *B,* Kaplan-Meier plot of recurrence-free survival of HCCs stratified by hierarchical clustering of expression profile of recurrence genes.
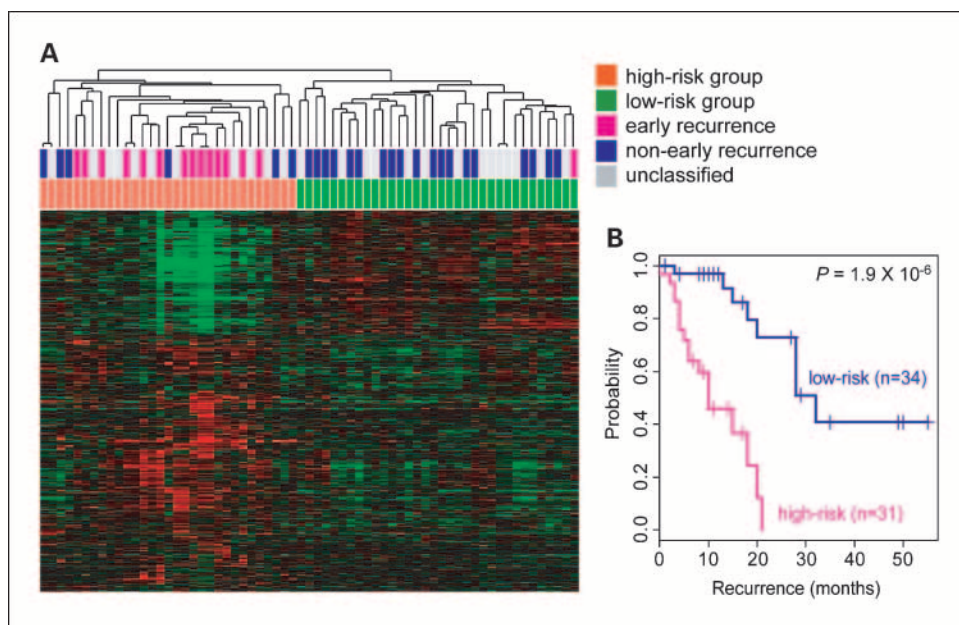
heterogeneity of the patient cohorts and the difference in microarray platforms, it is important to identify a reliable and consistent predictor that is robust enough to overcome the variabilities introduced by different platforms or different patient cohorts.

In the present study, we examined the gene expression profiles of 65 patients with HCC associated with the same viral background of hepatitis B virus (HBV) infection and identified molecular markers that predict HCC prognostic subtypes of high-risk and low-risk of early recurrence. The robustness and consistency of predictability was validated when our gene expression signature was applied to a completely independent patient cohort (15). This suggests that the signature would be more accurate and promising for clinical application. Moreover, as all of the 65 patients were HBV positive, these gene expression profiles might chiefly help in the understanding of HBV-related hepatocarcinogenesis. Detailed functional analyses of the prognostic subtypes provide novel molecular insights into HCC recurrence mechanisms.

Patients and Methods

Patients. Between February 2001 and May 2005, we prospectively collected resected HCC specimens with a pathologically proven cirrhotic background from 65 patients who were chronically infected with HBV and had surgical treatment for HCC at Seoul National University Hospital. All patients were preoperatively evaluated with routine blood tests, α-fetoprotein, routine X-ray, abdominal ultrasonography, and two-phase spiral liver computed tomography scan. Space-occupying lesions in the liver remnant were examined by intraoperative ultrasonography; no distant metastases or space-occupying lesions were identified in the nonresected part of the liver of any of the individuals in this study. We excluded subjects who were positive in serologic tests for anti-HCV or anti-HIV (HCV3.2; Dong-A Pharmaceutical Co.; Greencross Life Science Corp.). Patients with other types of liver disease, such as autoimmune hepatitis, toxic hepatitis, primary biliary cirrhosis, or Budd-Chiari syndrome were also excluded. The study protocol was approved by the institutional review board for the use of human subjects

at the Seoul National University School of Medicine, and all participants provided written informed consent. We defined curative resection as complete excision of the tumor with clear microscopic margins and no residual tumors as indicated by computed tomography scan at 1 month after surgery. To assess tumor size and undertake pathologic examination, we sectioned the resected specimens using the slice with the largest diameter, which we then cut at intervals of 5 mm. Two experienced pathologists independently examined all samples for evidence of residual tumors at the surgical margin, tumor differentiation, stage, and presence of vascular invasion. Based on these examinations, all 65 patients were determined to have received "curative resection." Patients were followed up at least once every 3 months after surgery.

Microarray experiments and analysis. Total RNA was extracted from frozen tissues using TRIzol (Invitrogen) and then cleaned using an RNeasy Mini kit (Qiagen). Five micrograms of total RNA from the HCC tissues was used for labeling, and microarray hybridization was carried out on Human Genome U133A 2.0 chips (Affymetrix) according to the manufacturer's protocol. The fluorescent intensities were determined with a GeneChip scanner 3000 (Affymetrix), controlled by GCOS Affymetrix software.

Raw data were normalized using the Robust Multiarray Average method (20) and global median centering. Hierarchical clustering analyses of gene expression profiles were done based on centered correlation metric and average linkage method.

Class prediction and the misclassification rates of the classifiers were estimated by a leave-one-out cross-validation method using different algorithms (compound covariate predictor, linear discriminant analysis, nearest centroid, *k*-nearest neighbor, and support vector machine) implemented in BRB-Array Tools.[10] The probabilities of recurrence-free and overall survival rates were estimated with Kaplan-Meier plots and significance was determined by log-rank test. Statistical analyses were done using R/Bioconductor package.

For data integration with the independent data set, each data set was standardized independently by transforming the expression of each gene to a mean of 0 and SD of 1, pooled the expression profiles together, and then considered them as a single data set. Probes in each data set were matched with Entrez Gene identifiers. For the multiple tagged genes, the probe with the largest magnitude (i.e., sum of the
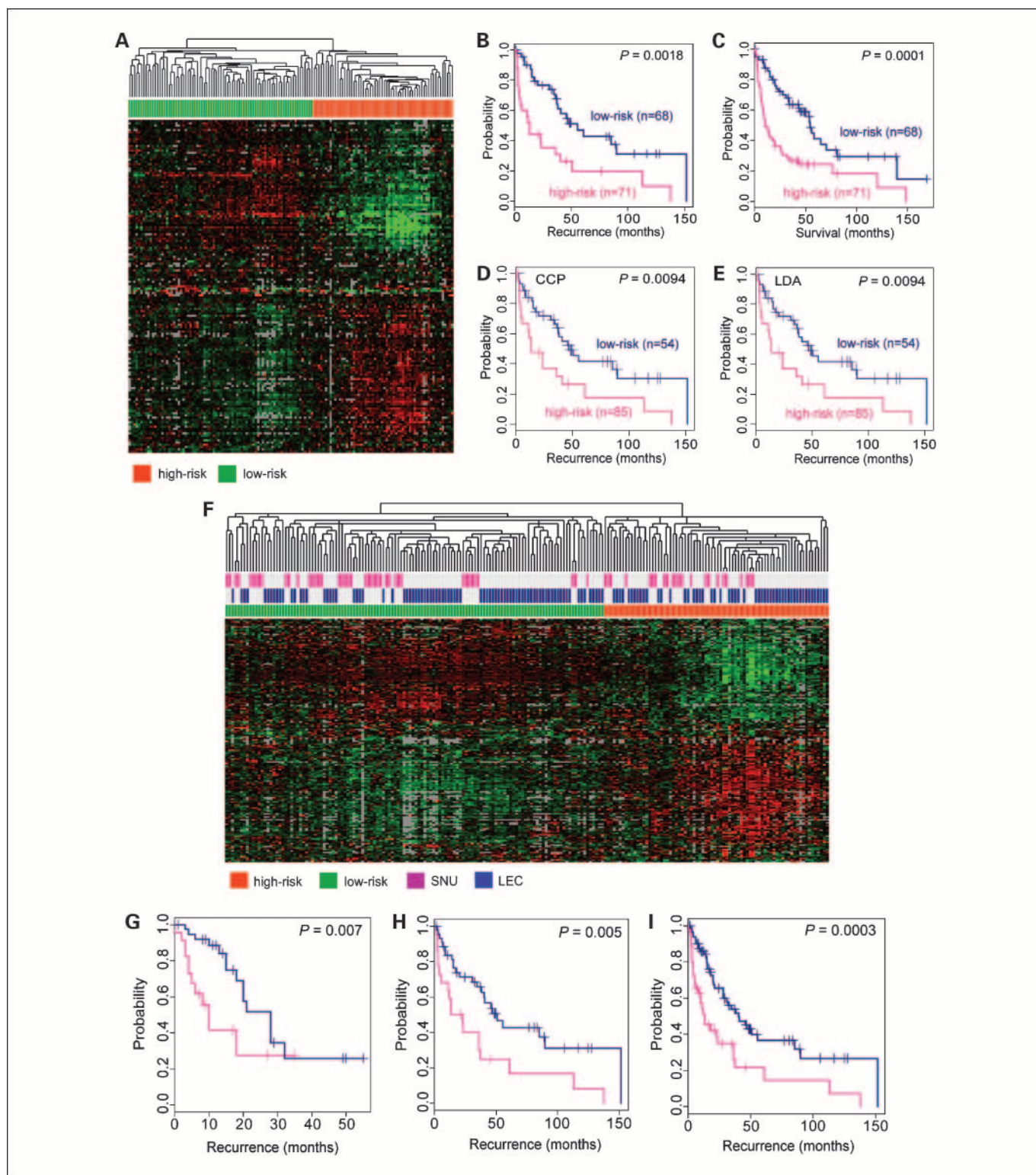
---

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

Fig. 2. Validation of recurrence genes by cross-platform comparison with an independent HCC data set. *A,* 150 out of the 628 SNU recurrence genes were found in the LEC data set. Hierarchical clustering was done on these gene expressions in the LEC data set (*n* = 139, recurrence information was available in 67 samples). Before clustering, the average expression levels of each gene were set to 0. *B* and *C,* Kaplan-Meier plot of recurrence-free survival (*B*) and overall survival (*C*) of HCCs (LEC) grouped by hierarchical clustering of the recurrence gene expression profile. *D* and *E,* independent prediction algorithms of compound covariate predictor (*D*) and linear discriminant analysis (*E*) were trained with SNU expression data and then applied to the LEC data set, respectively, and Kaplan-Meier plot analysis and log-rank test were done on the predicted classes. *F,* hierarchical clustering was done on an integrated data set comprised of both SNU and LEC data sets. Before integration, each data set was standardized independently by transforming the expression levels of each gene to a mean of 0 and SD of 1 (see Patients and Methods) *G-I,* Kaplan-Meier plots and log-rank test of recurrence-free survival of HCCs in SNU (*G, n* = 65), LEC (*H, n* = 139), and the overall integrated data (*I,* SNU + LEC, *n* = 204), respectively. HCCs were grouped based on hierarchical clustering of the expression profiles of the integrated data sets. Abbreviations: CCP, compound covariate predictor; LDA, linear discriminant analysis; SNU, data from Seoul National University; LEC, data from Laboratory of Experimental Carcinogenesis, National Cancer Institute, NIH.
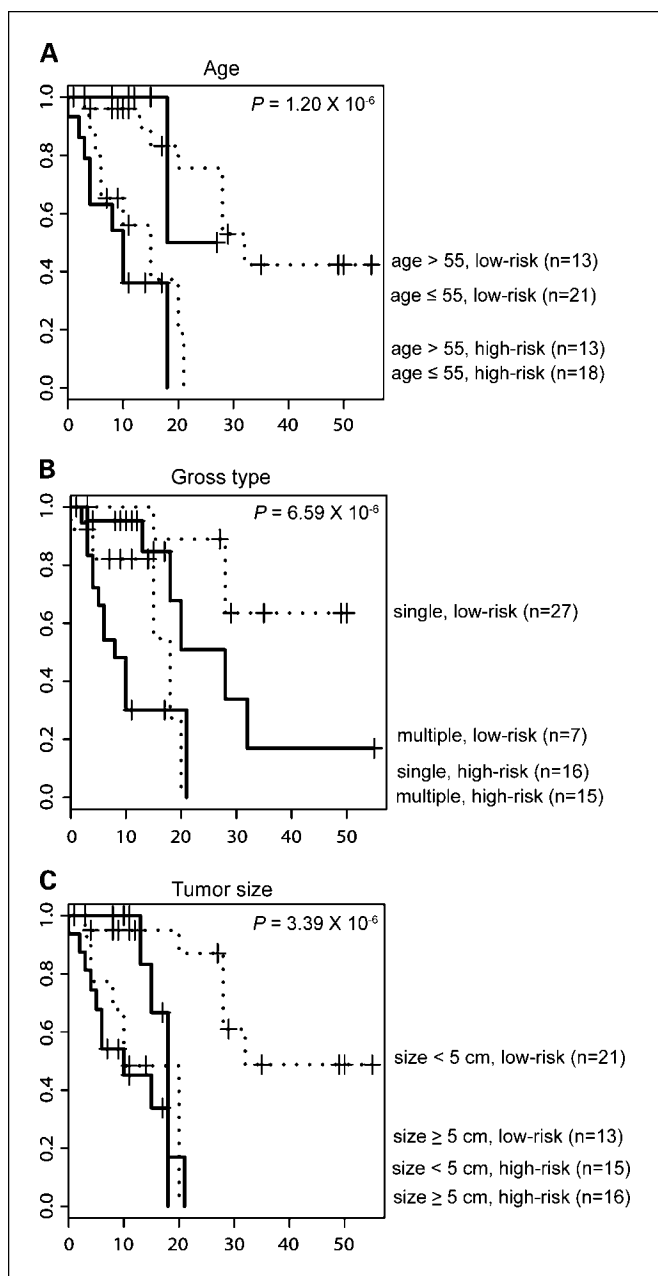
Fig. 3. Prediction of recurrence-free survival by combining the classification of clinicopathologic features and molecular subtypes. *A,* Kaplan-Meier plot of recurrence-free survival of patients with HCC grouped by combining patients' age (age >55 or ≤55 y old) and molecular subtypes. *B,* Kaplan-Meier plot of recurrence-free survival of patients with HCC grouped by combining tumor gross type and molecular subtypes. Tumor gross type was classified into single nodular (*n* = 43) and multiple nodular types (*n* = 25). The single nodular type includes solitary nodular and solitary nodular with perinodal extension types. The multiple nodular type includes confluent multinodular and multinodular discrete types. *C,* Kaplan-Meier plot of recurrence-free survival of patients with HCC grouped by combining variables of tumor size (diameter, >5 or ≤5 cm) and molecular subtypes. *P* values represent the significances of log-rank tests across the subgroups divided by combined variables.

squares of expression values in each sample) was selected as a representative probe.

Functional analysis of signatures. Once a gene set was identified as useful in the stratification of a patient's outcome, we attempted to gain insight into molecular mechanisms that might be involved in generating this hierarchy of patient outcome. For the functional

analysis of gene sets, enrichment of the gene set was estimated by the cumulative hypergeometric P values of each biological process provided by Gene Ontology Consortium.[11] In order to obtain representative and significantly enriched terms, those terms with a level higher than two in the Gene Ontology hierarchy, including at least three genes, were considered in our calculation. Statistical significance was determined with a cutoff of $P < 0.01$.

In another approach, we employed PathwayAssist software (Ariadne Genomics, version 3.0) as an independent pathway analysis tool to identify connections between differentially expressed genes. After constructing genetic networks, we sought to identify common regulators or common targets of the differentially expressed gene sets.

Results

Identification of HCC recurrence signature. We examined the gene expression profile of 65 HBV-associated HCCs using Affymetrix U133A 2.0 chips. In order to access the association of the expression variables of each gene feature with recurrence-free survival, a univariate Cox proportional hazard model was applied. A total of 628 gene features were selected as recurrence signatures that were highly correlated with the length of recurrence with strong statistical significance ($P < 0.005$, log-rank test) and differentially expressed across samples at non-trivial levels (SD > 0.3). Hierarchical clustering with these 628 recurrence signature genes subdivided HCC patients into two subtypes that appropriately reflect the difference in recurrence times between patients with HCC (Fig. 1A). Kaplan-Meier plot analysis and log-rank test showed a significant difference of recurrence-free survival between these two HCC subtypes (Fig. 1B, $P = 1.9 \times 10^{-6}$, log-rank test).

The absence of early recurrence during the first year after surgery is the golden standard to determine the success of curative resection. In prior microarray experiments, we subdivided the patients with HCC into early recurrence (i.e., HCCs recurred within a year from curative surgery) and late recurrence (i.e., HCCs free of recurrence for >1 year) groups. As shown in Fig. 1A, most of the early recurrence samples were predicted to be in the high-risk group with 82.5% of accuracy suggesting that our subtype classification might be helpful in planning adequate strategies for patient treatment.

Validation of recurrence signature with independent gene expression data. Having defined two distinct HCC subtypes that reflect significantly different clinical outcomes, we decided to test the robustness of the identified recurrence signature by applying six different class prediction methods (compound covariate predictor, linear discriminant analysis, k-nearest neighbor, nearest centroid, and support vector machine). Prediction of these two risk subtypes by six different class prediction algorithms showed between 83% and 97% mean prediction accuracy rates with significant leave-one-out misclassification rates ($P < 0.01$, based on 100 random permutations; Supplementary Table S1). These results strongly support the robustness of our recurrence signature.

For the validation of the prognostic reproducibility of this recurrence signature, we next applied our recurrence signature directly to an independent gene expression data set of patients with HCC [data from Laboratory of Experimental Carcinogenesis (LEC), National Cancer Institute, NIH; ref. 15]. Hierarchical clustering of gene expression profile of recurrence signature in

---

[11] www.geneontology.org

the LEC data set could subdivide patients into two distinct subgroups with homogeneous expression patterns (Fig. 2A). Kaplan-Meier plot analysis and log-rank test of these HCC subtypes showed a significant difference of overall survival ($P = 0.0001$), as well as recurrence-free survival ($P = 0.0018$; Fig. 2B and C). This suggests that our recurrence signature is well conserved in the independent data set and is able to predict recurrence-free survival regardless of microarray platforms.

In addition to the use of hierarchical clustering, we applied two independent classification algorithms (compound covariate predictor and linear discriminant analysis) to validate the robustness of the gene expression signature that predicted the likelihood of early recurrence. Gene expression data from the Seoul National University (SNU) were used to train classification algorithms and those from LEC were used as the test set. During training, the number of genes used in the prediction was optimized to minimize misclassification errors during leave-one-out cross-validation. When applied to the LEC data set, both algorithms successfully identified early recurrence patients with statistical significance (Fig. 2D and E).

In another approach, we applied data integration by pooling both data sets. Hierarchical clustering of recurrence signatures in overall integrated HCCs ($n = 204$) showed two main clusters with homogeneous expression patterns across platforms (Fig. 3F), suggesting that the expression patterns of the recurrence signatures were well conserved in both data sets. Kaplan-Meier plot analysis of these HCC subtypes showed a significant difference of recurrence between the subgroups of each individual data set (SNU data set, $P = 0.007$; LEC data set, $P = 0.005$, respectively, log-rank test; Fig. 3G and H). Kaplan-Meier analysis on the overall integrated data set also successfully dissected subgroups based on the recurrence rate ($P = 0.0003$; Fig. 3I). These results strongly support the consistency and robustness of this recurrence signature at this independent cohort and experimental platforms of individual studies.

Clinicopathologic features and recurrence signature. The prognostic values of conventional clinicopathologic factors on the risk of recurrence have been widely studied. In agreement with previous reports, we identified patient's age (21), tumor gross type (22), and tumor size (23) to be associated with the likelihood of HCC recurrence in univariate Cox proportional hazards analysis (Table 1). However, other clinicopathologic features such as serum α-fetoprotein, serum platelet count, differentiation, tumor grade, venous invasion, and extranodal invasion and adjuvant therapy (trans-arterial chemoembolization) were not associated with recurrence-free survival. Even though all the patients had a history of HBV infection, the serotype status of HBeAg and anti-HBe were not associated with recurrence-free survival (data not shown). The multivariate analysis, including all the clinicopathologic variables and the molecular subtype, showed that only the molecular subtype was significantly associated with tumor recurrence (hazard rate, 12.54; 95% confidence interval, 3.59-43.76, $P < 7.30 \times 10^{-5}$; Table 1).

In an attempt to improve the prognostic usefulness of clinical features, we combined classifications of molecular subtypes with clinical features, hoping to predict tumor recurrence much more precisely. Of the clinical features that showed significant association with recurrence in univariate Cox regression analysis, the combined application of patient age or tumor gross type with recurrence signature improved the predictability of the patients' outcomes better than the recurrence signature alone (Fig. 3A and B). In addition, combining tumor size with molecular subtype could predict patients' outcome more precisely by stratifying the patients assigned to the low-risk group (Fig. 3C). These results suggest that a combined application of certain clinical features with molecular subtypes would be a practical approach to define stratified recurrence-free survival groups.

Biological insights of HCC recurrence signature. In order to get a biological insight on the mechanisms reflecting the

Table 1. Clinicopathologic features of patients with HCC

Clinical features	Recurrence risk group				Univariate analysis		Multivariate analysis	
Class (I/II)	Low-risk ($n = 34$)		High-risk ($n = 31$)		Hazard rate (95% CI)	*P*	Hazard rate (95% CI)	*P*
	I	II	I	II				
Gender (woman/man)	5	29	8	23	0.948 (0.38-2.35)	0.909	1.319 (0.40-4.28)	0.640
Age (≤55/>55 y)	21	13	18	13	0.421 (0.17-0.99)	0.042	0.495 (0.75-5.39)	0.160
α-Fetoprotein (≤400/>400 ng/mL)	24	10	23	8	0.608 (0.20-1.76)	0.357	0.273 (0.06-1.04)	0.067
Platelet (≤100/>100 × 1,000/μL)	14	20	8	23	0.992 (0.44-2.21)	0.985	0.420 (0.15-1.15)	0.092
Differentiation (well/moderate, poor)	6	28	2	29	1.144 (0.23-5.02)	0.791	0.254 (0.06-1.04)	0.058
Tumor size (5/≥5 cm)	21	13	15	16	3.149 (1.27-7.76)	0.009	2.180 (0.74-6.36)	0.150
Gross type (single/multiple)*	27	7	16	15	2.361 (1.00-5.52)	0.042	1.087 (0.30-3.90)	0.900
T stage (T_0, T_1/T_2, T_3)	17	17	14	17	0.804 (0.37-1.71)	0.573	0.483 (0.10-2.26)	0.360
Venous invasion (no/yes)	20	14	16	15	1.349 (0.62-2.89)	0.440	3.145 (0.66-14.99)	0.150
Extranodal invasion (no/yes)	20	14	10	21	1.864 (0.85-4.05)	0.111	3.995 (0.90-17.65)	0.068
Adjuvant TACE (no/yes)	28	6	24	7	0.708 (0.26-1.88)	0.486	0.832 (0.23-2.96)	0.780
Molecular subtype (low-risk/high-risk)					8.896 (3.14-25.13)	1.90×10^{-6}	12.539 (3.59-43.76)	7.30×10^{-6}

NOTE: For univariate and multivariate analyses, hazard ratio and log-rank *P* values were calculated by Cox proportional hazards analysis.
Abbreviations: TACE, trans-arterial chemoembolization; 95% CI, 95% confidence interval.
*Tumor gross type was classified into two groups as single nodular type ($n = 43$) and multiple nodular types ($n = 25$). Single nodular type includes solitary nodular and solitary nodular with perinodal extension type. Multiple nodular type includes confluent multinodular and multinodular discrete type.
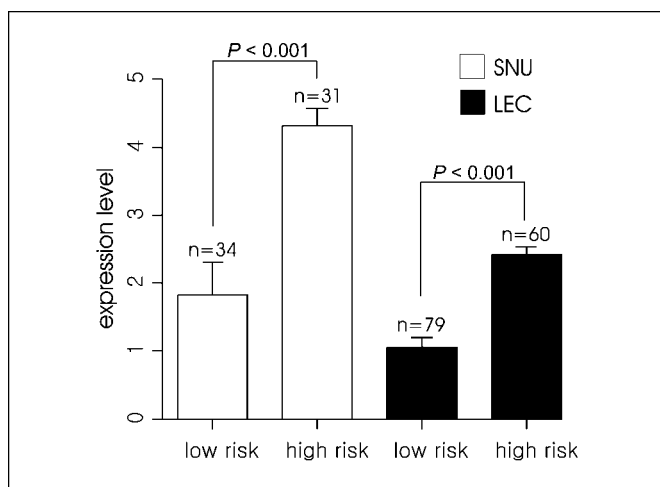
Fig. 4. Expression levels of *CD24* in high-risk and low-risk groups of early recurrence. High-risk and low-risk groups of early recurrence were assigned by hierarchical clustering recurrence signature in SNU (Fig. 1) and LEC data sets (Fig. 2A). The average expression levels of *CD24* between high-risk and low-risk groups were compared in both data sets, respectively. Expression levels in the SNU data set represent relative values to the overall median of normalized log-transformed expression profile, whereas the expression levels of the LEC data set represent the fold changes of log-transformed expression levels between tumor samples and normal liver tissues. Statistical significance was estimated by two-tailed Student's *t* test (*P* < 0.001). Columns, mean; bars, SE.

differences of prognostic outcomes of these two molecular subtypes, the genes showing significant differences in expression between these two subtypes were selected with a two-sample *t* test. A total of 937 genes showing significant differences in 10,000 permuted two-sample *t* tests (*P* < 0.001, false discovery rate < 1.46%) with fold difference between subtypes greater than 1.4-fold were selected for the analysis of Gene Ontology composition.

Functional enrichment analysis with Gene Ontology categories (Supplementary Table S2) showed a significant enrichment of metastasis-related functions including actin filament organization, regulation of cell migration, and cell motility. As expected, proliferation-related functions (cell proliferation, regulation of progression through cell cycle) and differentiation/development-related functions (cytoskeleton organization and biogenesis, cell fate determination, skeletal development) showed significant enrichment in the high-risk group. Of interest, notch signaling genes (*JAG1*, *JAG2*, and *NOTCH2*) were significantly up-regulated in the high-risk group, implying their functional roles in HCC recurrence. Inflammation-related functions (i.e., chemotaxis, humoral immune response) were also highly enriched in the high-risk group. Inflammation/immune response–related genes were reported to have an association with noncancerous hepatic tissues from patients with metastatic HCC (17), suggesting that its enrichment in the recurrence signature might be derived from noncancerous stromal cells promoting surveillance for HCC recurrence.

Notably, *CD24* showed the highest fold difference of geometric mean (6.84-fold) and all six probes for *CD24* (i.e., 208650_s_at, 208651_x_at, 216379_x_at, 209771_x_at, 209772_s_at, and 266_s_at) were significantly overexpressed in the high-risk group (*P* < 0.001; Supplementary Table S3). As shown in Fig. 4, *CD24* expression levels between high-risk and low-risk groups were significantly different in the LEC as well as

in the SNU data set (*P* < 0.001, two-tailed Student's *t* test for each data set). This concordant observation in both data sets identifies CD24 as a putative biomarker for the prediction of early recurrence.

Once the molecular subtype of HCC was validated for robustness of prognostic capacity, we examined the genetic network of differentially expressed genes between the subtypes. To identify the most prominent common regulatory genes, we carried out pathway analysis using PathwayAssist (Ariadne Genomics, version 3.0). We found that SP1 was the most prominent common regulator for the genes overexpressed in the high-risk group compared with the low-risk group (Fig. 5). Of these SP1 targets, many genes, e.g., *PLAUR* (14, 16, 24, 25), *FGFR1* (26), *VIM* (27), *PDGFA* (28), and *HK2* (29) had previously been reported to be associated with HCC prognosis or metastasis, and it strongly implicates SP1 as a critical factor in HCC recurrence. Contrary to SP1, peroxisome proliferator–activated receptor α (PPARα) was identified as a prominent common regulator for many of the down-regulated genes in the high-risk group. From these results, we suggest that SP1 and PPARα play critical roles in HCC recurrence.

Discussion

Many previous studies have shown successful analyses of gene expression profiles for the prognostic prediction of patients with cancer, but their clinical applications have been overoptimistic and still premature. The lack of consistency and the robustness of expression profiles is thought to be one of major obstacles for its clinical application. In this regard, external validation by comparing totally independent cross-platform and cross-site studies will help to identify robust predictors reducing data set–derived systematic biases. The application of cross-platform comparison of independent studies has its own limitations due to the heterogeneity and unavailability of the data sets to be combined. However, this approach is now at the stage of one of reliable solutions to overcome with the overfitting problem of microarray data.

In the present study, we examined the gene expression profiles of 65 patients with HCC to generate a genetic classifier that could identify the patients with a high-risk of early recurrence following curative resection. The 628 gene features selected as genetic classifiers by a univariate Cox proportional hazard model could classify HCC patients into high-risk (*n* = 31) and low-risk (*n* = 34) subtypes of early recurrence of HCCs using hierarchical clustering analysis. Cross-platform analysis of this recurrence signature with independent data sets showed consistent stratification of HCC patients which appropriately reflects the risk of early recurrence, suggesting that it might be less prone to false findings and is independent of individual studies. Moreover, the HCC samples in our data set were collected from a homogenous patient population with the same viral exposure (i.e., HBV), ethnicity, hospital care, and postoperative follow-up; therefore, it would be less confounded and more informative for the understanding of recurrence mechanisms.

CD24 was identified as a putative biomarker for classifying low-risk and high-risk groups of early recurrence in both SNU and LEC data sets (Fig. 4). Congruent with this finding, previous studies showed that the CD24 expression level is prognostic in many cancers (30–36), including HCC (37),
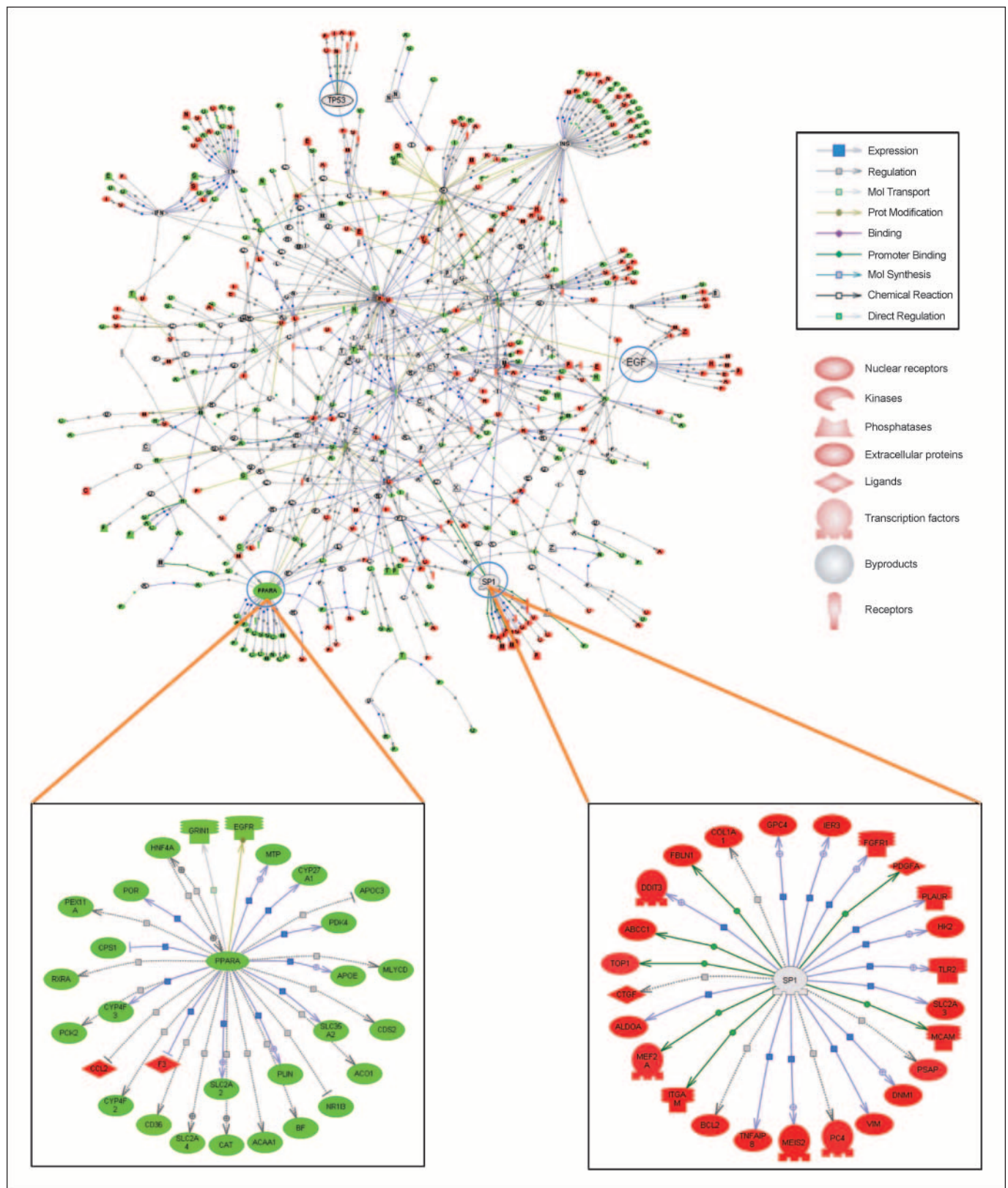
Fig. 5. Common regulatory genes of differentially expressed genes between molecular subtypes of HCCs. Gene regulatory network of differentially expressed genes ($P < 0.001$, two-tailed Student's t test) between high-risk and low-risk groups were constructed using PathwayAssist software. Up-regulated (*red*) and down-regulated (*green*) differentially expressed genes in the high-risk group compared with nondifferentially expressed genes in the low-risk group (*gray*). Enlarged pictures of SP1 and PPARα indicated their common regulations of recurrence genes. Common regulators of TP53 and EGF (*blue circle*; details in Supplementary Fig. S1). References for genetic interactions for SP1 and PPARα are listed in the Supplementary Notes.

although its prognostic role for recurrence has not been noted. CD24 is known to participate in the regulation of cell-to-cell and cell-to-matrix interactions, and its ligand, P-selectin is associated with tumor metastasis by increasing cell spreading, adhesion, and proliferation (30). Therefore, we suggest that CD24 might be a good putative biomarker for the prediction of early recurrence of HCC.

Genetic network analysis of this recurrence signature revealed SP1 and PPARα as prominent common regulators of genes that differed in expression between high-risk and low-risk groups. Many genes that regulate the cell cycle frequently contain proximal GC-rich promoter sequences, and their interactions with SP proteins and other transcription factors are critical for their expression (38). SP1 is associated with the prognosis of cancers, including pancreatic cancer (39), breast cancer (40), and gastric cancer (41), although the potential roles of SP1 in HCC prognosis remains unclear. In line with these studies, it is likely that SP1 would be a good candidate for further studies to elucidate its regulatory role in the recurrence mechanism of HCC.

PPARα agonists have been known to cause hepatocarcinogenesis in rats and mice, whereas humans seem to be resistant (42, 43). Human PPARα mRNA and functional receptor is expressed at a level <10% of that found in rats and mice, which may contribute to a difference in susceptibility to agonists between rodents and humans (44). Lower expression of human PPARα might be due to variant human PPARα mRNA species, in which exon 6 was deleted by alternative splicing, and in which the amounts of variants were reported to be up to 20% to 50% of the total PPARα mRNA in human tissues (44). These studies imply that the variants of PPARα might lead to different expressions of PPARα and its target genes between high-risk and low-risk groups. In tumor progression, the PPARα agonist, fenofibrate, was revealed to have antimetastatic potential in both human and mouse melanoma cells (45), suggesting that PPARα has a tumor suppressor role, at least in humans, besides its tumorigenic potential in rodents. From these findings, we could hypothesize that lower expression of PPARα (possibly related to variant PPARα) may affect the differences of recurrence potential between HCC subtypes. However, we cannot rule out the possibility that the deleterious loss of hepatic functions and subsequent depletion of lipid metabolism in the high-risk group could be related to the lower expression of PPARα (see Supplementary Table S2).

In addition to SP1 and PPARα, close examination of genetic networks revealed several prominent common regulators such as EGF and PTGS2, which have previously been well studied in association with HCC progression (refs. 46, 47; Fig. 5; Supplementary Fig. S1A and B). When we constructed a genetic network with the target genes of the differentially expressed genes between the two groups, FOS and JUN were revealed as common downstream targets of the overexpressed genes in the high-risk group (Supplementary Fig. S1C), which is consistent with a previous study that shows its regulatory role in HCCs with poor prognosis (15). When all the regulators and targets of the differentially expressed genes were pooled in the network, TP53 and TGFB1 were identified as commonly regulated and targeted genes, which have previously been shown to play critical roles in cancer progression (refs. 47, 48; Supplementary Fig. S1D and E). Taken together, these results suggest that concomitant disruption of multiple gene expression networks is required for HCCs to adopt an aggressive phenotype.

In conclusion, we generated a consistent and robust recurrence predictor independent of platforms and cohorts, which could successfully predict molecular subtypes of HCC that reflect the likelihood of early recurrence after curative resection. We also showed that the combined analysis of the molecular subtypes with clinicopathologic features could improve their prognostic utilities. In addition, our study provides substantial biological insights that prioritize the functional significance of SP1 and PPARα in HCC recurrence mechanisms and CD24 as a putative biomarker. We believe that our predictor profile can be helpful to clinicians in choosing a treatment modality for HCC patients who have a high risk of early recurrence.

Acknowledgments

References

1. El-Serag HB, Mason AC. Rising incidence of hepatocellular carcinoma in the United States. N Engl J Med 1999;340:745–50.

2. Tung-Ping Poon R, Fan ST, Wong J. Risk factors, prevention, and management of postoperative recurrence after resection of hepatocellular carcinoma. Ann Surg 2000;232:10–24.

3. Hanazaki K, Kajikawa S, Shimozawa N, et al. Survival and recurrence after hepatic resection of 386 consecutive patients with hepatocellular carcinoma. J Am Coll Surg 2000;191:381–8.

4. Yang Y, Nagano H, Ota H, et al. Patterns and clinicopathologic features of extrahepatic recurrence of hepatocellular carcinoma after curative resection. Surgery 2007;141:196–202.

5. Tsai TJ, Chau GY, Lui WY, et al. Clinical significance of microscopic tumor venous invasion in patients with resectable hepatocellular carcinoma. Surgery 2000; 127:603–8.

6. Zhou L, Rui JA, Wang SB, et al. Clinicopathological features, post-surgical survival and prognostic indicators of elderly patients with hepatocellular carcinoma. Eur J Surg Oncol 2006;32:767–72.

7. Poon RT, Fan ST, Ng IO, Lo CM, Liu CL, Wong J. Different risk factors and prognosis for early and late intrahepatic recurrence after resection of hepatocellular carcinoma. Cancer 2000;89:500–7.

8. Okuda K, Ohtsuki T, Obata H, et al. Natural history of hepatocellular carcinoma and prognosis in relation to treatment. Study of 850 patients. Cancer 1985;56: 918–28.

9. Llovet JM, Bruix J. Prospective validation of the Cancer of the Liver Italian Program (CLIP) score: a new prognostic system for patients with cirrhosis and hepatocellular carcinoma. Hepatology 2000;32: 679–80.

10. Kudo M, Chung H, Osaki Y. Prognostic staging system for hepatocellular carcinoma (CLIP score): its value and limitations, and a proposal for a new staging system, the Japan Integrated Staging Score (JIS score). J Gastroenterol 2003;38:207–15.

11. Talwalkar JA, Gores GJ. Diagnosis and staging of hepatocellular carcinoma. Gastroenterology 2004;127: S126–32.

12. Llovet JM, Burroughs A, Bruix J. Hepatocellular carcinoma. Lancet 2003;362:1907–17.

13. 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.

14. 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.

15. Lee JS, Heo J, Libbrecht L, et al. A novel prognostic subtype of human hepatocellular carcinoma derived from hepatic progenitor cells. Nat Med 2006;12:410–6.

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. Budhu A, Forgues M, Ye QH, et al. Prediction of venous metastases, recurrence, and prognosis in hepatocellular carcinoma based on a unique immune response signature of the liver microenvironment. Cancer Cell 2006;10:99–111.

18. Tinker AV, Boussioutas A, Bowtell DD. The challenges of gene expression microarrays for the study of human cancer. Cancer Cell 2006;9:333–9.

19. Bammler T, Beyer RP, Bhattacharya S, et al. Standardizing global gene expression analysis between laboratories and across platforms. Nat Methods 2005;2:351–6.

20. Irizarry RA, Hobbs B, Collin F, et al. Exploration,

normalization, and summaries of high density oligonu-cleotide array probe level data. Biostatistics 2003;4: 249 – 64.

21. Hanazaki K, Matsushita A, Nakagawa K, Misawa R, Amano J. Risk factors of intrahepatic recurrence after curative resection of hepatocellular carcinoma. Hepatogastroenterology 2005;52:580 – 6.

22. Inayoshi J, Ichida T, Sugitani S, et al. Gross appearance of hepatocellular carcinoma reflects E-cadherin expression and risk of early recurrence after surgical treatment. J Gastroenterol Hepatol 2003;18:673 – 7.

23. Hanazaki K, Matsushita A, Nakagawa K, Misawa R, Amano J. Risk factors of long-term survival and recurrence after curative resection of hepatocellular carcinoma. Hepatogastroenterology 2005;52:552 – 7.

24. Zhou L, Hayashi Y, Itoh T, Wang W, Rui J, Itoh H. Expression of urokinase-type plasminogen activator, urokinase-type plasminogen activator receptor, and plasminogen activator inhibitor-1 and -2 in hepatocellular carcinoma. Pathol Int 2000;50:392 – 7.

25. Itoh T, Hayashi Y, Kanamaru T, et al. Clinical significance of urokinase-type plasminogen activator activity in hepatocellular carcinoma. J Gastroenterol Hepatol 2000;15:422 – 30.

26. Huang X, Yu C, Jin C, et al. Ectopic activity of fibroblast growth factor receptor 1 in hepatocytes accelerates hepatocarcinogenesis by driving proliferation and vascular endothelial growth factor-induced angiogenesis. Cancer Res 2006;66:1481 – 90.

27. Hu L, Lau SH, Tzang CH, et al. Association of Vimentin overexpression and hepatocellular carcinoma metastasis. Oncogene 2004;23:298 – 302.

28. Gotzmann J, Fischer AN, Zojer M, et al. A crucial function of PDGF in TGF-β-mediated cancer progression of hepatocytes. Oncogene 2006;25:3170 – 85.

29. Yasuda S, Arii S, Mori A, et al. Hexokinase II and VEGF expression in liver tumors: correlation with hyp-oxia-inducible factor 1α and its significance. J Hepatol 2004;40:117 – 23.

30. Baumann P, Cremers N, Kroese F, et al. CD24 expression causes the acquisition of multiple cellular properties associated with tumor growth and metastasis. Cancer Res 2005;65:10783 – 93.

31. Kristiansen G, Winzer KJ, Mayordomo E, et al. CD24 expression is a new prognostic marker in breast cancer. Clin Cancer Res 2003;9:4906 – 13.

32. Su MC, Hsu C, Kao HL, Jeng YM. CD24 expression is a prognostic factor in intrahepatic cholangiocarcinoma. Cancer Lett 2006;235:34 – 9.

33. Kristiansen G, Pilarsky C, Pervan J, et al. CD24 expression is a significant predictor of PSA relapse and poor prognosis in low grade or organ confined prostate cancer. Prostate 2004;58:183 – 92.

34. Sagiv E, Memeo L, Karin A, et al. CD24 is a new oncogene, early at the multistep process of colorectal cancer carcinogenesis. Gastroenterology 2006;131: 630 – 9.

35. Kristiansen G, Schluns K, Yongwei Y, Denkert C, Dietel M, Petersen I. CD24 is an independent prognostic marker of survival in nonsmall cell lung cancer patients. Br J Cancer 2003;88:231 – 6.

36. Kristiansen G, Denkert C, Schluns K, Dahl E, Pilarsky C, Hauptmann S. CD24 is expressed in ovarian cancer and is a new independent prognostic marker of patient survival. Am J Pathol 2002;161:1215 – 21.

37. Marrero JA, Lok AS. Newer markers for hepatocellular carcinoma. Gastroenterology 2004;127:S113 – 9.

38. Safe S, Abdelrahim M. Sp transcription factor family and its role in cancer. Eur J Cancer 2005;41:2438 – 48.

39. Abdelrahim M, Smith R III, Burghardt R, Safe S. Role of Sp proteins in regulation of vascular endothelial growth factor expression and proliferation of pancreatic cancer cells. Cancer Res 2004;64:6740 – 9.

40. Zannetti A, Del Vecchio S, Carriero MV, et al. Coor-dinate up-regulation of Sp1 DNA-binding activity and urokinase receptor expression in breast carcinoma. Cancer Res 2000;60:1546 – 51.

41. Yao JC, Wang L, Wei D, et al. Association between expression of transcription factor Sp1 and increased vascular endothelial growth factor expression, advanced stage, and poor survival in patients with resected gastric cancer. Clin Cancer Res 2004;10: 4109 – 17.

42. Peters JM, Cheung C, Gonzalez FJ. Peroxisome proliferator-activated receptor-α and liver cancer: where do we stand? J Mol Med 2005;83:774 – 85.

43. Gonzalez FJ. The peroxisome proliferator-activated receptor α (PPARα): role in hepatocarcinogenesis. Mol Cell Endocrinol 2002;193:71 – 9.

44. Gervois P, Torra IP, Chinetti G, et al. A truncated human peroxisome proliferator-activated receptor α splice variant with dominant negative activity. Mol Endocrinol 1999;13:1535 – 49.

45. Grabacka M, Plonka PM, Urbanska K, Reiss K. Peroxisome proliferator-activated receptor α activation decreases metastatic potential of melanoma cells *in vitro* via down-regulation of Akt. Clin Cancer Res 2006;12:3028 – 36.

46. Borlak J, Meier T, Halter R, Spanel R, Spanel-Borowski K. Epidermal growth factor-induced hepatocellular carcinoma: gene expression profiles in precursor lesions, early stage and solitary tumours. Oncogene 2005;24:1809 – 19.

47. Leng J, Han C, Demetris AJ, Michalopoulos GK, Wu T. Cyclooxygenase-2 promotes hepatocellular carcinoma cell growth through Akt activation: evidence for Akt inhibition in celecoxib-induced apoptosis. Hepatology 2003;38:756 – 68.

48. Gold LI. The role for transforming growth factor-β (TGF-β) in human cancer. Crit Rev Oncog 1999;10: 303 – 60.