Integrative Transcriptome Analysis Reveals Common Molecular Subclasses of Human Hepatocellular Carcinoma

Yujin Hoshida,[1,2] Sebastian M.B. Nijman,[1,5] Masahiro Kobayashi,[6] Jennifer A. Chan,[1,7] Jean-Philippe Brunet,[1] Derek Y. Chiang,[1] Augusto Villanueva,[8] Philippa Newell,[10] Kenji Ikeda,[6] Masaji Hashimoto,[6] Goro Watanabe,[6] Stacey Gabriel,[1] Scott L. Friedman,[10] Hiromitsu Kumada,[6] Josep M. Llovet,[8,9,10] and Todd R. Golub[1,2,3,4]

[1]Broad Institute of Massachusetts Institute of Technology and Harvard University, Cambridge, Massachusetts; [2]Pediatric Oncology, Dana-Farber Cancer Institute; [3]Children's Hospital Boston, Harvard Medical School; [4]Howard Hughes Medical Institute, Boston, Massachusetts; [5]Center for Molecular Medicine of the Austrian Academy of Sciences, Vienna, Austria; [6]Toranomon Hospital, Tokyo, Japan; [7]University of Calgary, Calgary, Alberta, Canada; [8]Barcelona-Clínic Liver Cancer Group, Liver Unit, CIBERehd, Hospital Clínic, IDIBAPS; [9]Institució Catalana de Recerca i Estudis Avançats, Barcelona, Spain; and [10]Liver Cancer Program, Mount Sinai School of Medicine, New York, New York

Abstract

Hepatocellular carcinoma (HCC) is a highly heterogeneous disease, and prior attempts to develop genomic-based classification for HCC have yielded highly divergent results, indicating difficulty in identifying unified molecular anatomy. We performed a meta-analysis of gene expression profiles in data sets from eight independent patient cohorts across the world. In addition, aiming to establish the real world applicability of a classification system, we profiled 118 formalin-fixed, paraffin-embedded tissues from an additional patient cohort. A total of 603 patients were analyzed, representing the major etiologies of HCC (hepatitis B and C) collected from Western and Eastern countries. We observed three robust HCC subclasses (termed S1, S2, and S3), each correlated with clinical parameters such as tumor size, extent of cellular differentiation, and serum α-fetoprotein levels. An analysis of the components of the signatures indicated that S1 reflected aberrant activation of the WNT signaling pathway, S2 was characterized by proliferation as well as MYC and AKT activation, and S3 was associated with hepatocyte differentiation. Functional studies indicated that the WNT pathway activation signature characteristic of S1 tumors was not simply the result of *β-catenin* mutation but rather was the result of transforming growth factor-β activation, thus representing a new mechanism of WNT pathway activation in HCC. These experiments establish the first consensus classification framework for HCC based on gene expression profiles and highlight the power of integrating multiple data sets to define a robust molecular taxonomy of the disease. [Cancer Res 2009;69(18):7385–92]

Introduction

Hepatocellular carcinoma (HCC) affects approximately half a million patients worldwide and is the most rapidly increasing cause of cancer death in the United States owing to the lack of effective treatment options for advanced disease (1). Numerous lines of clinical and histopathologic evidence suggest that HCC is a heterogeneous disease, but a coherent molecular explanation for this heterogeneity has yet to be reported. Genomic approaches to the classification of HCC therefore hold promise for a molecular taxonomy of the disease.

Mutations in the WNT signaling pathway have been found to be common in HCC, but other DNA level classification approaches have proven challenging. This relates to the enormous complexity of the genomic alterations observed in HCC, likely attributable to the accumulation of chromosomal rearrangements resulting from decades of chronic viral hepatitis and cirrhosis. This complexity makes it difficult to identify the causal genetic events promoting HCC development and progression (2, 3). An alternate approach to HCC classification has been to study tumors at the level of their gene expression profiles. Although several such profiling efforts have been reported (4–11), a cohesive view of expression-based subclasses of HCC has yet to emerge. In part, this is because each of the reported studies analyzed different patient populations (most of them small) on a different microarray platform, with a different primary biological or clinical question in mind. Perhaps not surprisingly then, each study reported a somewhat different view of the heterogeneity of HCC, and it has been therefore impossible to see whether there exists a common biological thread that links these disparate studies.

We believe that any biologically or clinically meaningful classification system should be informative across multiple patient populations and should be independent of any particular microarray platform. In the present study, we therefore set out to define molecular subclasses of HCC that existed across all available HCC data sets, including eight previously reported studies and one new one reported here, totaling 603 patients. We report that indeed there exist three distinct molecular subclasses of HCC that are present in all nine data sets examined, regardless of technical differences between the microarray platforms used to generate the profiles. We show that these subclasses are correlated with histologic, molecular, and clinical features of HCC, and we highlight the important role of transforming growth factor-β (TGF-β) signaling in one of the HCC subclasses. These findings thus create a solid foundation for HCC classification on which to build informed clinical trials for patients with HCC and also suggest new opportunities for therapeutic intervention.

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

Requests for reprints: Todd R. Golub, Broad Institute of Massachusetts Institute of Technology and Harvard University, 7 Cambridge Center, Cambridge, MA 02142. Phone: 617-324-0700; Fax: 617-258-0903; E-mail: golub@broad.mit.edu.

doi:10.1158/0008-5472.CAN-09-1089

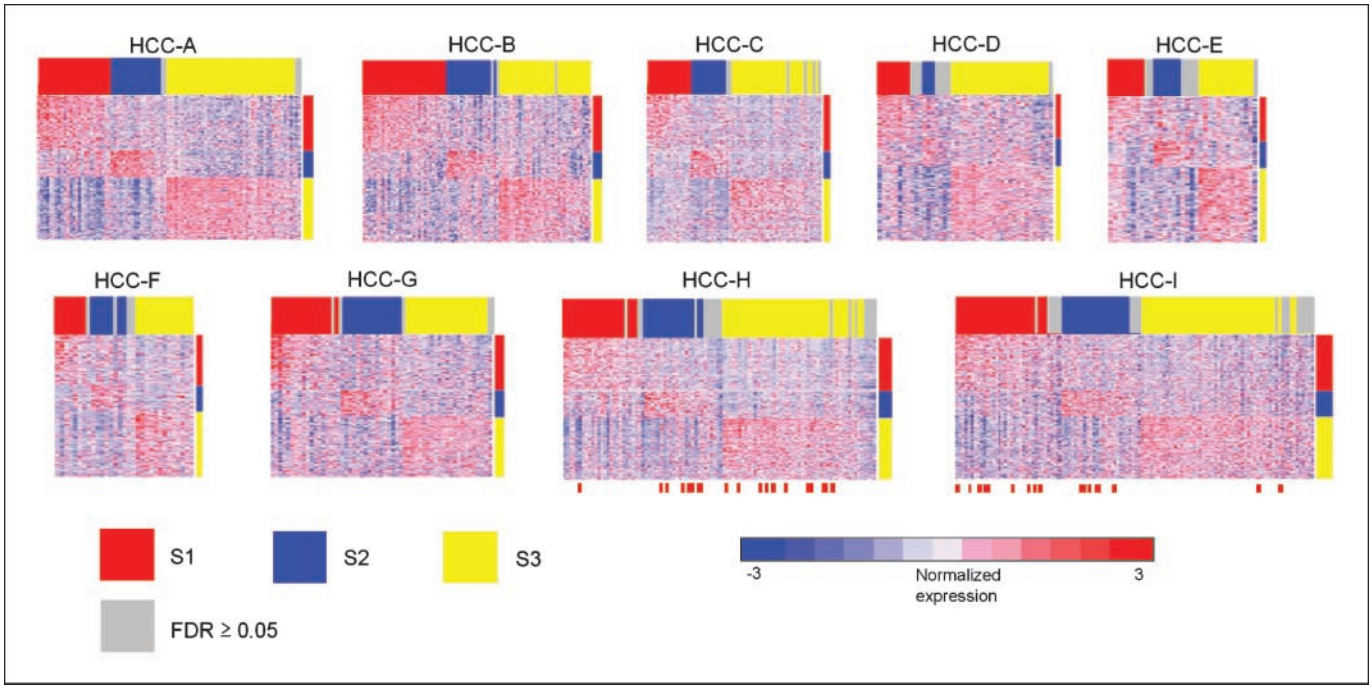

Figure 1. HCC subclasses predicted in nine independent data sets. Predicted subclasses are shown in red (S1), blue (S2), and yellow (S3) with expression pattern of the HCC subclass signature. The proportion of the cases with confident prediction (FDR < 0.05) in HCC-A, HCC-B, HCC-C, HCC-D, HCC-E, HCC-F, HCC-G, HCC-H, and HCC-I were 96%, 96%, 90%, 81%, 79%, 87%, 94%, 83%, and 83%, respectively. Red bars attached to HCC-H and HCC-I indicate positive *β-catenin* mutations and nuclear staining of p53, respectively.

Materials and Methods

Microarray Data Sets and Statistical Analysis

Identification of common HCC subclasses. To define and validate a gene expression–based model of common molecular subclasses of HCC, we collected publicly available gene expression data sets from eight independent cohorts profiled on a wide variety of microarray platforms (HCC-A, HCC-B, HCC-C, HCC-D, HCC-E, HCC-F, HCC-G, and HCC-H; see Supplementary Tables S1 and S2 for details; refs. 4–11). Between the training data sets (HCC-A, HCC-B, and HCC-C) chosen as larger data sets covering major etiologies of HCC to avoid overfitting a model to any particular cohort or microarray platform, corresponding subgroups of the samples were defined by subclass mapping (SubMap) method (12) based on subclasses that identified three unsupervised clustering methods: hierarchical clustering, k-means clustering, and nonnegative matrix factorization, which finds clusters of samples after collapsing the data set into representative "meta-genes" (13).

For each subclass defined by SubMap, meta-analysis marker genes were selected as overexpressed genes compared with the rest of the subclasses (HCC subclass signature) in all the three clustering methods to avoid defining gene expression–based subclasses that were unique to a particular clustering algorithm. Prediction of the subclass was performed for each sample using nearest template prediction method (14, 15) to accommodate the diverse microarray platforms (see Supplementary Data for details).

Molecular annotation of HCC subclasses. Functional characterization of the HCC subclasses was performed using gene set enrichment analysis (GSEA; ref. 16). Two categories of gene sets in Molecular Signature Database (MSigDB)[11] were used: target gene sets regulated by experimental perturbations (377 gene sets) and literature-based manually curated molecular pathway gene sets (150 gene sets).

Clinical data analysis. The hazard of tumor recurrence was calculated to estimate the pattern of HCC recurrence over time after the surgery as previously reported (17, 18). Continuous and proportional data were tested by Wilcoxon rank sum test and Fisher's exact test, respectively. All clinical data analyses were performed using the R statistical package version 2.4.0.[12]

Microarrays for Fixed Tissues

We generated a ninth data set of formalin-fixed, paraffin-embedded (FFPE) tumors (HCC-I), reasoning that any meaningful classification system should be applicable to routinely collected fixed (as opposed to frozen) tissues. We analyzed FFPE tissue blocks from 118 HCC patients who consecutively underwent surgical resection during 1990 to 2001 at Toranomon Hospital. Ethical approval for use of the FFPE tissues, obtained and archived as part of routine clinical care, was acquired from the institutional review board granted on condition that all samples be made anonymous. Total RNA was extracted from macrodissected 10-μm tissue slices (three to four slices for each sample) using High Pure RNA Paraffin kit (Roche). Expression of transcriptionally informative 6,000 genes, selected to capture global state of transcriptome (14), was profiled using DNA-mediated annealing, selection, extension, and ligation (DASL) assay (Illumina; see Supplementary Data; ref. 19).

Microarrays for Cell Lines

Total RNA was isolated using Trizol reagent (Invitrogen) according to the manufacturer's instruction. Microarray experiment was performed using HT_HG-U133A High-Throughput Arrays (Affymetrix). The raw data were normalized using BioConductor affy package.[13]

All microarray data sets are available through National Center for Biotechnology Information Gene Expression Omnibus[14] with accession numbers of GSE10186 (DASL), GSE10393 (U133A), and GPL5474 (transcriptionally informative gene panel for DASL) and our Web site.[15]

[11] http://www.broad.mit.edu/gsea/msigdb/

[12] http://www.r-project.org
[13] http://www.bioconductor.org/
[14] http://www.ncbi.nlm.nih.gov/geo/
[15] http://www.broad.mit.edu/cancer/pub/HCC

Table 1. Clinical phenotypes associated with HCC subclasses

Variable	S1	S2	S3	*P*
Tumor size (cm)*	3.0 [2.0,4.5]	4.5 [2.5,7.0]	2.5 [1.8,4.3]	0.003
Tumor differentiation*				
Well	8 (16%)	4 (10%)	37 (44%)	
Moderate	27 (53%)	23 (59%)	45 (53%)	<0.001
Poor	16 (31%)	12 (31%)	3 (4%)	
AFP (ng/mL)†	50 [14,332]	171 [27,1,251]	13 [5,43]	<0.001
Hepatitis B virus infection‡	39 (38%)	27 (36%)	39 (25%)	0.05
Hepatitis C virus infection‡	55 (53%)	44 (58%)	109 (69%)	0.03

NOTE: Median [25%,75%]. Wilcoxon rank sum test for continuous data, and Fisher's exact test for categorical data.
*HCC-F, HCC-H, and HCC-I: S1, $n = 55$; S2, $n = 46$; S3, $n = 96$.
†HCC-H and HCC-I: S1, $n = 48$; S2, $n = 39$; S3, $n = 83$.
‡HCC-B, HCC-C, HCC-F, HCC-H, and HCC-I: S1, $n = 103$; S2, $n = 76$; S3, $n = 158$.

Immunostaining

Immunohistochemical staining was performed on 10-μm FFPE sections using antibodies for β-catenin (Becton Dickinson), phospho-AKT (Cell Signaling), and p53 (Immunotech) followed by detection using the Envision+ DAB system (Dako). The stains were evaluated by a pathologist blinded to the results of the gene expression profiling, and the results were scored in a binary system. For immunofluorescence staining, cells grown on multiwell chamber slides were fixed by 4% paraformaldehyde and stained for β-catenin (see Supplementary Data for details).

Cell Culture

Human HCC cell lines, Huh-7 (Riken Bioresource Center) and SNU-387 (American Type Culture Collection), were grown in DMEM supplemented with 10% heat-inactivated fetal bovine serum at 37°C in a 5% CO_2 atmosphere. No *β-catenin* mutation has been found in these cell lines.

Western Blotting

Cell lysates were separated on NuPAGE 4% to 12% gels (Invitrogen) and transferred to polyvinylidene difluoride membranes (Bio-Rad) and blotted for α-fetoprotein (AFP; Santa Cruz Biotechnology), β-catenin, phospho-SMAD3 (Cell Signaling), and proliferating cell nuclear antigen (Santa Cruz Biotechnology) antibodies.

β-Catenin Knockdown

Cells were infected with the indicated short hairpin RNA (shRNA) vectors (construct 1, TRCN0000003843; construct 2, TRCN0000003844; The RNAi Consortium[16]) and puromycin selected. Ninety-six hours after infection, cells were counted and seeded in triplicate (20,000/six well). After 10 d, cells were fixed in paraformaldehyde and stained with crystal violet. Dye was extracted with 10% acetic acid and absorbance was determined at 600 nm.

Luciferase Assay

Cells were transfected using Lipofectamine 2000 (Invitrogen) with either TOP-flash or FOP-flash constructs and stimulated with 100 pmol/L TGF-β (R&D Systems) for 48 h. Luciferase activity was measured using Dual-Glo kit (Promega). All transfections were performed in triplicate and measurements were normalized to a SV40 promoter-driven *Renilla* luciferase construct.

Results

Three common molecular subclasses of HCC. The SubMap method identified three robust subclasses in each of the three initial data sets analyzed. We refer to these subclasses as S1, S2, and S3 (Fig. 1), and the complete list of genes that correlated with each of the subclasses is available in Supplementary Table S3. As with any unsupervised clustering–based definition of cancer subclasses, it is essential to establish the validity of the newfound classification system. In the following sections, we describe three independent validations of the three-class structure of HCC. First, we show that the three subclasses are detected with statistical significance in each of the six remaining HCC data sets (totaling 371 patients). Second, we show that the subclasses are associated with clinical parameters. Third, we show that the subclasses are associated with biological mechanism known to be operative in the pathogenesis of HCC.

Statistical validation of subclasses across nine HCC cohorts. As a statistical measure of the validity of the three subclasses, we determined the confidence with which HCC samples could be classified into one of the three subclasses using the HCC subclass signature–based classifier, including 619 genes. As expected, subclass labels were assigned with high confidence [false discovery rate (FDR) < 0.05] to 94% of the training samples (HCC-A, HCC-B, and HCC-C), which were used to define the subclasses in the first place. More importantly, high confidence subclass labels were assigned to 84% of the 371 samples in the validation set (HCC-D, HCC-E, HCC-F, HCC-G, HCC-H, and HCC-I). In contrast, a classifier based on the same number of genes chosen randomly yielded high confidence predictions in <1% of the samples. In addition, our classification system was superior to those reported previously (10, 11, 20, 21) when those classifiers were tested across all of the validation data sets (high confidence predictions using reported signatures were 27–75%; Supplementary Fig. S1). These results, taken together, indicate the statistical significance of our three subclasses and point to the limitation of defining subclasses based on only a single cohort, where overfitting often leads to failure of the classifier to validate on new samples, particularly when profiled on a different microarray platform.

Clinical relevance of HCC subclasses. Having established the statistical validity of the HCC subclasses, we next asked whether any of the subclasses were associated with clinical parameters to add the validity of the subclasses. Clinical data were available for 197 patients, as summarized in Table 1.

Our first observation was that tumors in subclass S2 were larger than the others, whereas tumors in S3 were smaller compared with the rest ($P = 0.003$). In addition, subclass S3 included the majority of well-differentiated tumors (37 of 49; $P < 0.001$), whereas there was no

[16] www.broad.mit.edu/genome_bio/trc/rnai.html

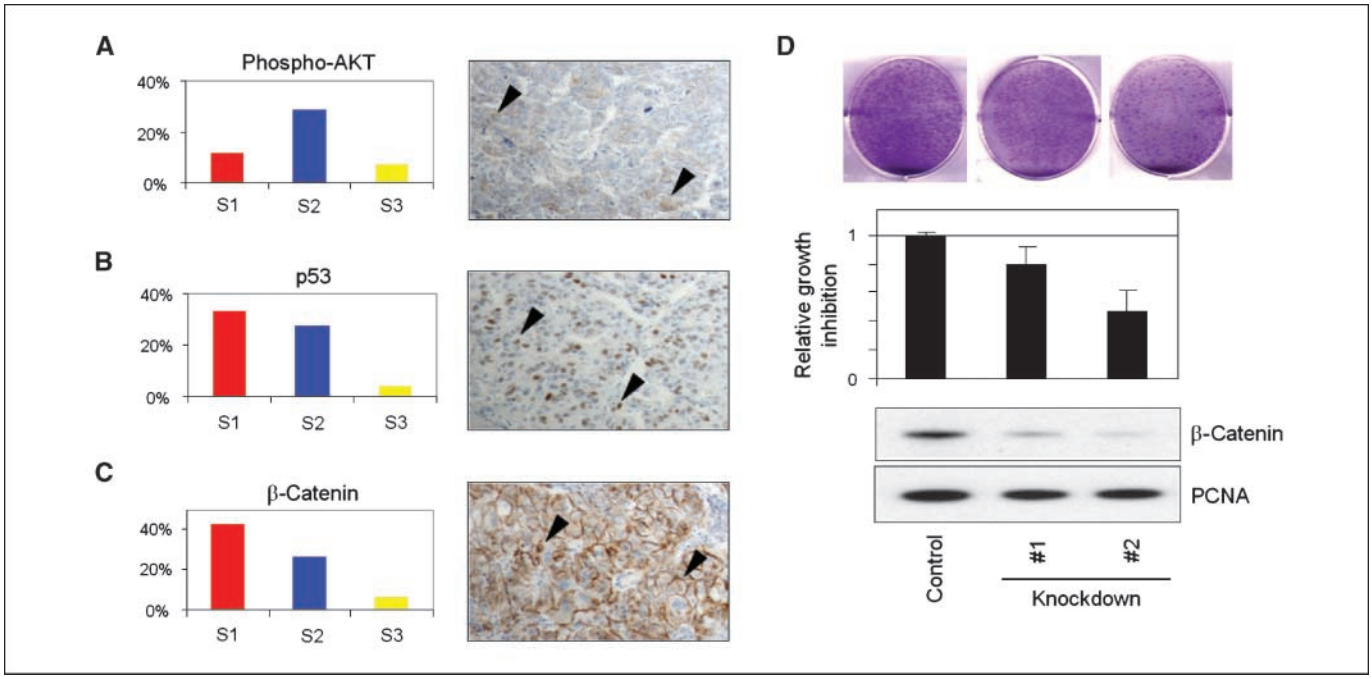

Figure 2. Molecular pathways associated with HCC subclasses. Immunohistochemical analysis of phospho-AKT (*A*), p53 (*B*), and β-catenin (*C*) proteins in HCC-I. *Left,* proportions of the cases with positive staining in each HCC subclass; *right,* representative positive staining (*arrowheads*). Magnification, ×20. *D,* growth inhibition of SNU-387 cells (predicted to be subclass S1) by knocking down β-catenin protein using two different shRNA constructs.

histologic difference between S1 and S2 ($P = 0.73$). We also examined the serum levels of the one clinically used serum biomarker of HCC—AFP. Serum AFP was the highest in S2 (0.001), further supporting the notion that our subclasses are clinically relevant.

Next, we sought to determine whether the HCC subclasses were associated with clinical outcome following surgical resection. We were careful to analyze the two major patterns of HCC recurrence: early recurrence, which is related to residual dissemination of primary tumor cells within the liver, and late recurrence, which is attributable to new primary tumors arising in a hypercarcinogenic state of a cirrhotic liver (17, 18, 22). "Early" recurrence is associated with malignant characteristics of the primary tumor itself, and we reported that it has less effect on patient survival in earlier-stage HCC, in which "late" recurrence is the major determinant of survival (14, 23). We found that subclass S1 was associated with a significantly greater risk of earlier recurrence ($P = 0.03$ within 1 year; Supplementary Fig. S2). This association remained to be significant even after correcting for tumor size in multivariate Cox regression modeling (Supplementary Table S4). Consistent with this observation, S1 tumors exhibited more vascular invasion and satellite lesions (both known risk factors for early recurrence; Supplementary Table S5). These results may suggest that the S1 subclass is associated with a more invasive/disseminative phenotype. Interestingly, we found that a recently reported signature of poor survival defined in patients with more advanced HCC, where early recurrence is the major determinant of survival (4), was associated with S1 and S2, whereas the good survival signature defined in that study was enriched in S3 tumors (Supplementary Table S6), lending further credence to our subclass model. Importantly, our HCC subclasses were not associated with late recurrence, consistent with our recent study indicating that late recurrence is determined not by the characteristics of the tumor itself but rather by the biological state of the surrounding liver at risk (14). Furthermore, consistent with our previous observation, the HCC subclasses showed no association with survival ($P = 0.12$) in our data set (HCC-H) consisting mostly of early-stage tumors.

Molecular pathways associated with HCC subclasses. We next asked whether we could ascribe biological meaning to our validated HCC subclasses. The GSEA results indicated that indeed our HCC subclasses were associated with distinct biological processes, several of which have been implicated in HCC pathogenesis (Supplementary Tables S7 and S8). For example, S2 were tumors associated with a relative suppression of IFN target genes (7), of interest because of the use of IFN as an experimental chemopreventive strategy for HCC. S2 tumors were also enriched in MYC target genes, suggesting that MYC activation is a feature of S2 tumors. Consistent with this observation, we found that a recently reported mouse model of HCC based on MYC overexpression (24) exhibited the S2 subclass signature (Supplementary Fig. S3). S2 tumors were also strongly enriched in a signature of EpCAM positivity (Supplementary Table S6; ref. 25), and in addition, we found that S2 tumors overexpressed *AFP* (consistent with S2 patients having elevated serum AFP levels; Table 1). Lastly, S2 tumors were enriched in a signature of AKT activation (10), and validation experiment indicated a trend toward elevated phosphorylation of AKT as determined by immunohistochemistry ($P = 0.07$; Fig. 2*A*). An AFP-AKT association has been previously observed (10, 26), and we see here that this association is being driven primarily by S2 tumors. The mechanism of AKT activation in these tumors is not known but likely reflects upstream signaling of the phosphatidylinositol 3-kinase (PI3K) pathway (27). As PI3K inhibitors are now entering clinical development, it may be of value to examine their role in S2 tumors in particular.

GSEA also identified differential activation of p53 and p21 target gene sets, with these genes being more abundantly expressed in S3 tumors compared with S1 and S2, consistent with our observation that S3 tumors tend to be lower grade (Table 1). To further validate this result, we performed

immunohistochemistry for p53, wherein nuclear accumulation of p53 protein is well known to reflect inactivating p53 mutation (28). As predicted by the GSEA analysis, S1 and S2 tumors exhibited significantly greater nuclear p53 staining compared with S3 (P = 0.001; Fig. 2*B*). The more well-differentiated nature of S3 tumors was also reflected in the S3 gene expression profile, with S3 tumors exhibiting relatively higher levels of expression of hepatocyte function-related genes involved in glycogen/lipid/alcohol metabolism (*APO/ALDH/ADH* family genes), detoxification (*CYP* family genes), coagulation, and oxygen radical scavenging (*CAT, SOD1*; Supplementary Tables S3, S7, and S8).

WNT pathway activation in S1. The WNT signaling pathway is perhaps the best characterized oncogenic pathway in HCC, with pathway activation occurring through *β-catenin* mutation (specifically via mutation in exon 3 in up to 44% of cases) and less frequently in *AXIN1* (<10% of cases; refs. 2, 3). We addressed WNT status with regard to HCC subclass in two ways. First, we performed GSEA analysis using an experimentally defined WNT activation signature. We found strong enrichment of the WNT signature in subclass S1 compared with S2 or S3 (FDR = 0.03; Supplementary Table S7), suggesting preferential WNT activation in S1 tumors. We validated this result via immunohistochemistry for β-catenin (the principal downstream effector of WNT in HCC) and found that S1 tumors indeed had higher levels of cytoplasmic β-catenin protein expression compared with the other subclasses (0.001), again indicating preferential activation of the canonical WNT pathway in S1 (Fig. 2*C*; ref. 29). In addition, we found that shRNAs targeting *β-catenin* resulted in growth inhibition when introduced into the SNU-387 cell line (predicted to be subclass S1), thereby further supporting the hypothesis that WNT activation is functionally important in S1 tumors (Fig. 2*D*).

Mechanisms of WNT activation in S1 tumors. Having determined that S1 tumors exhibit preferential activation of the WNT/β-catenin pathway, we next addressed potential mechanisms for this activity. We first asked whether the S1 tumors were associated with *β-catenin* mutation in HCC-H data set, for which we previously sequenced exon 3 of *β-catenin* gene (11). Surprisingly, *β-catenin* mutations were preferentially found in S3 tumors, consistent with previously reported "CTNNB1" class representing a subset of S3 (11, 30). This result is also consistent with recent evidence indicating that *β-catenin* mutations do not regulate the canonical WNT target genes (e.g., *cyclin D1* and *MYC*) that characterize our S1-associated WNT activation signature (31). These results further suggest that the WNT pathway is activated in S1 tumors by a mechanism other than *β-catenin* mutation.

To explore alternate explanations for WNT pathway activation, we again turned to GSEA, asking whether other gene sets (signatures) enriched in S1 tumors might provide insight into WNT activation in these tumors. Strikingly, we observed strong overexpression of TGF-β target gene sets (i.e., genes expressed as a result of experimental activation of TGF-β) in S1 tumors (Supplementary Table S7). We similarly observed enrichment of a gene set associated with epithelial-to-mesenchymal transition, a phenomenon implicated in tumor invasion and metastasis (32) and known to be regulated by TGF-β signaling in HCC (Supplementary Table S7; refs. 33, 34). Furthermore, a previously reported TGF-β activation signature associated with an invasive phenotype (35) showed strong enrichment in S1 (Supplementary Table S6). We observed no genomic copy number change associated with S1 subclass in *TGFB1* locus, suggesting that chromosomal aberration is not the causative mechanism of the activation (Supplementary Fig. S4). These results indicate that TGF-β and WNT signaling co-occur in the same HCC subclass (subclass S1), and suggest the hypothesis that TGF-β might in some way lead to WNT activity that defines the S1 molecular phenotype.

TGF-β-WNT interactions. We next explored the hypothesis that TGF-β regulates WNT pathway activity in HCC cells. First, we treated the HCC cell line Huh-7 with intact WNT pathway components (classified as subclass S2 and with no activation of S1 and WNT activation signature) with TGF-β and monitored the genome-wide expression consequence. As predicted, TGF-β stimulation induced expression of WNT target genes (FDR < 0.001; Fig. 3*A*) and induced the expression of genes characteristic of subclass S1 (FDR = 0.04; Fig. 3*B*; Supplementary Data) characterized by WNT/

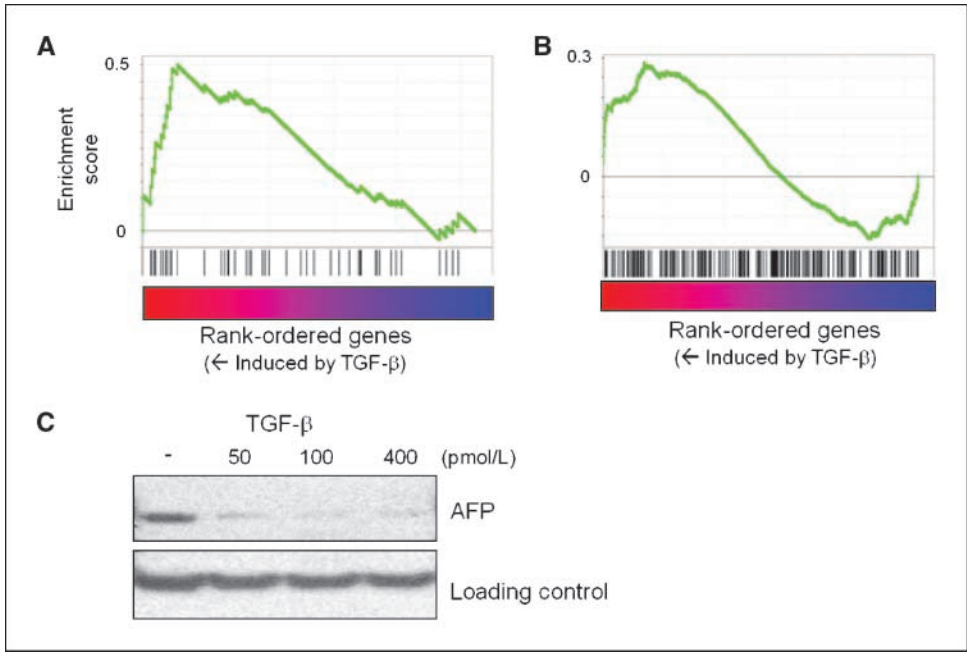

Figure 3. Interaction between WNT pathway and TGF-β. *A,* up-regulation of an experimentally defined WNT target gene set, "KENNY_WNT_UP" (FDR < 0.001), by TGF-β. Genes were rank ordered based on differential expression between TGF-β–treated and untreated Huh-7 cells (predicted to be subclass S2). A database of target gene sets for experimental perturbations (377 gene sets) was assessed by GSEA. *B,* up-regulation of the S1 signature by TGF-β treatment. Genes were rank ordered based on differential expression between treated and untreated Huh-7 cells, and induction of the subclass signature was evaluated by GSEA (FDR = 0.04). *C,* suppression of AFP protein expression by TGF-β treatment. Loading control is nonspecific for AFP antibody to show that equal amounts of protein were loaded.

TGF-β activity while suppressing expression of AFP protein, one of the top markers for S2 (Fig. 3*C*).

Second, we asked whether TGF-β could regulate the activity of a T-cell factor-lymphoid enhancer factor (TCF-LEF) reporter, further reflecting WNT/β-catenin activity. TGF-β stimulation of Huh-7 cells resulted in activation of a wild-type (TOP-flash) but not mutant (FOP-flash) TCF-LEF luciferase reporter (Fig. 4*A*). Interestingly, the superactivation of TCF-LEF activity was also observed in the presence of cotransfected mutant β-catenin. These results validate the hypothesis that TGF-β enhances WNT activity in HCC, consistent with the subclass S1 molecular profile.

We next explored the mechanism by which TGF-β augments WNT/β-catenin activity. A simple explanation would be that TGF-β induced expression of β-catenin RNA or protein levels, but we found no evidence for this (Fig. 4*B*). Strikingly, however, TGF-β treatment resulted in a marked change in β-catenin subcellular localization. Specifically, TGF-β treatment induced a shift from membranous β-catenin staining to a cytoplasmic distribution with focal perinuclear aggregation (Fig. 4*C*). This suggests that TGF-β enhances WNT signaling by modulating the intracellular pool of free β-catenin.

Taken together, these results validate the observation that TGF-β and WNT activity together typify the S1 subclass of HCC, and further suggest that TGF-β augments WNT activity via alteration of the subcellular localization of β-catenin, consistent with the cross-talk between these pathways observed in other biological contexts (34, 36). This implies that therapeutic cotargeting TGF-β and β-catenin in S1 tumors might be explored as a strategy for the treatment of S1 subclass HCC.

Discussion

Advances in genome technologies are now supporting a breadth of cancer genome characterization studies, including those focusing on HCC. Along with this proliferation of studies has come, however, a certain confusion in the field—different studies often report different results relating to the same set of underlying questions. For example, ~10 articles on the gene expression–based classification of HCC have been published in recent years, but a consensus molecular taxonomy of the disease has yet to emerge. This might lead some to believe that either expression technologies are insufficiently stable or HCC is so hopelessly heterogeneous and complex that regular, reproducible patterns in the data are nonexistent. We report here that in fact a highly reproducible molecular architecture of HCC is identifiable and is detected across all available HCC data sets.

Our analysis of nine HCC data sets totaling 603 patients indicated that there exist three major subclasses of HCC, which we refer to as subclasses S1, S2, and S3 (Fig. 5). Importantly, although the proportion of each subclass varied slightly from study to study, the subclasses were identifiable regardless of the geographic location of the study patients (Asia versus Europe versus United States) or the technology platform used (cDNA versus oligonucleotide arrays, and frozen versus FFPE tissues). Notably, the new data set generated in the present study used FFPE tissues, thereby showing that the three-class structure is readily detectable in specimens collected and stored in the routine clinical setting. This is relevant because the future deployment of diagnostic tests aimed at cancer classification should ideally be applicable to the standard FFPE specimens that are obtained in clinical practice.

Several biological insights can be made from the observed three-class structure of HCC. Class S1 is particularly notable for the prominence of a WNT activation gene expression signature. This is notable because such WNT activation is not simply explained by the presence of activated *β-catenin* mutations, suggesting that additional mechanisms of WNT activation seem to be at play,

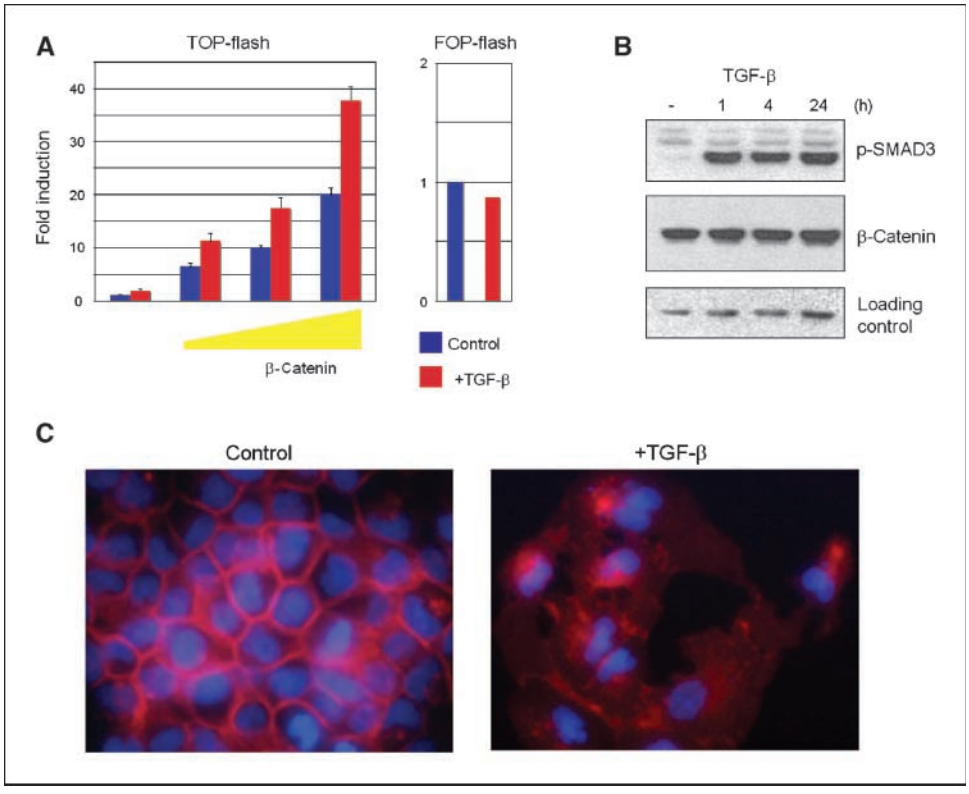

Figure 4. Activation of WNT pathway by TGF-β. *A,* Huh-7 cells were transfected with the indicated reporter constructs and increasing amounts of mutant β-catenin (2, 5, and 10 ng of plasmid). *B,* TGF-β pathway activation was confirmed by phosphorylation of SMAD3. Abundance of β-catenin protein was not changed by TGF-β treatment (100 pmol/L, 48 h). Loading control is nonspecific for phospho-SMAD3 antibody to show that equal amounts of protein were loaded. *C,* Huh-7 cells were stimulated as above and stained for β-catenin. Cellular distribution of β-catenin changed from predominantly membranous to cytoplasmic and perinuclear, and clustered cells spread out with more elongated and flattened morphology.

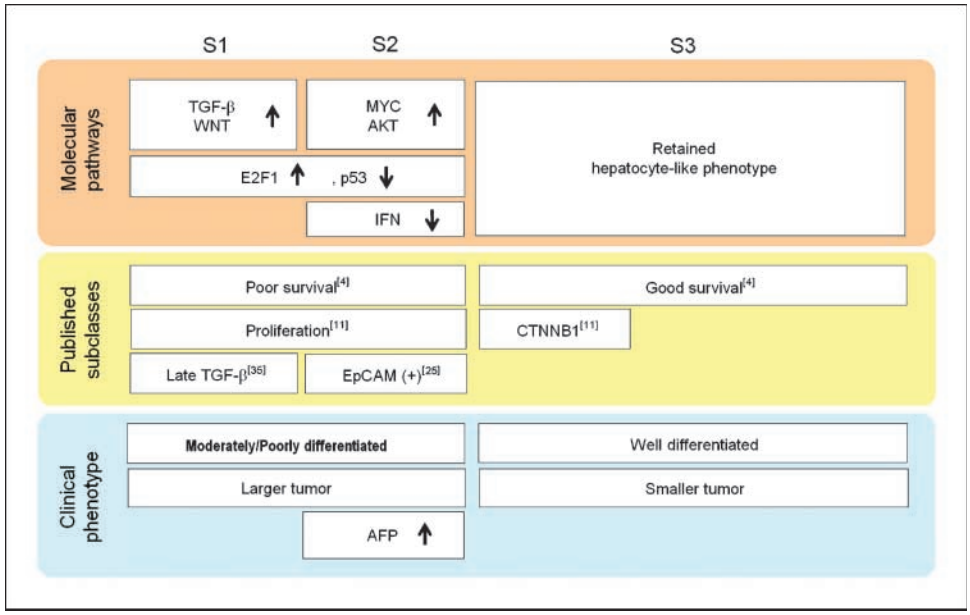

Figure 5. Schematic summary of the characteristics of HCC subclasses.

including TGF-β activation. This may be particularly important in the setting of clinical trials testing β-catenin inhibitors in HCC. Our data suggest that such inhibitors may be worth exploring in HCC beyond those patients harboring *β-catenin* mutation. Although additional mechanistic studies are clearly required, our data support the existence of an interaction between WNT activation and TGF-β activation in S1 tumors, an interaction that has been recently proposed in HCC (34).

Class S2 tumors were notable for their high level of expression of *AFP*, associated with elevated plasma levels of AFP protein compared with non-S2 patients. S2 tumors also tended to be enriched in MYC tumors harboring a MYC activation signature. This is of relevance because it suggests that genetically engineered mouse models of HCC based on MYC activation may be used to interrogate biological basis of the S2 subclass of human HCC. In addition, the finding of an AKT activation signature in S2 tumors suggests that AKT or PI3K inhibitors might be particularly worthy of exploration in this subclass. Further studies are required to establish the mechanism by which AKT is activated in these tumors.

S3 tumors were notable for their relative histologic evidence of differentiation, and the S3 gene expression program was accordingly suggestive of a molecular program of differentiated hepatocyte function. It is tempting to speculate that these tumors might be particularly well suited to differentiation therapy with agents such as retinoids, as has been previously suggested (37). Whether S3 tumors have distinct mechanisms of transformation or rather simply allow for more complete cellular differentiation remains to be determined. The preserved p53 function in S3 suggests that the abrogation of p53 is associated with stepwise malignant transformation of well-differentiated tumors rather than initiation of carcinogenesis. The less frequent *β-catenin* mutations in S1 and S2 may suggest that these tumors arose through different carcinogenic mechanisms compared with S3.

Clearly, much remains to be learned about the biological basis of our observed HCC subclasses. But the fact that they are observed in all studies of HCC examined to date suggests that they represent a reproducible classification framework for the disease. We therefore propose that it will be important to know the subclass of HCC patients entering clinical trials for the treatment of HCC because the response to targeted agents (e.g., β-catenin and PI3K) is likely to be different across the subsets (38). Early observations of differential sensitivity of these distinct tumor types may help guide the design of future clinical trials aimed at targeting agents to distinct patient populations.

Disclosure of Potential Conflicts of Interest

No potential conflicts of interest were disclosed.

Acknowledgments

Received 3/25/09; revised 6/9/09; accepted 6/13/09; published OnlineFirst 9/1/09.

Grant support: NIH/National Cancer Institute grant 5U54 CA112962-03 (T.R. Golub), NIH/National Institutes of Diabetes, Digestive and Kidney Diseases grant 1R01DK076986 (J.M. Llovet), NIH (Spain) grant I+D Program SAF-2007-61898 (J.M. Llovet), and Samuel Waxman Cancer Research Foundation. Y. Hoshida is supported by Charles A. King Trust fellowship. S.M.B. Nijman was supported by the Netherlands Organisation for Scientific Research (Rubicon) and Dutch Cancer Society fellowships.

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.

We thank Menno Creyghton for reagents and helpful suggestion; So Young Kim, Ron Firestein, William Hahn, and David Root for the shRNA constructs; Megan Hanna for technical help; Weijia Zhang for critical reading of the manuscript; and Jadwiga Grabarek and Mariko Kobayashi for general support.

References

1. El-Serag HB, Rudolph KL. Hepatocellular carcinoma: epidemiology and molecular carcinogenesis. Gastroenterology 2007;132:2557–76.
2. Farazi PA, DePinho RA. Hepatocellular carcinoma pathogenesis: from genes to environment. Nat Rev Cancer 2006;6:674–87.
3. Villanueva A, Newell P, Chiang DY, Friedman SL, Llovet JM. Genomics and signaling pathways in hepatocellular carcinoma. Semin Liver Dis 2007;27:55–76.
4. 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.
5. Chen X, Cheung ST, So S, et al. Gene expression patterns in human liver cancers. Mol Biol Cell 2002;13:1929–39.
6. 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.
7. Breuhahn K, Vreden S, Haddad R, et al. Molecular profiling of human hepatocellular carcinoma defines mutually exclusive interferon regulation and insulin-like growth factor II overexpression. Cancer Res 2004;64:6058–64.
8. 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.
9. Midorikawa Y, Tsutsumi S, Nishimura K, et al. Distinct chromosomal bias of gene expression signatures in the progression of hepatocellular carcinoma. Cancer Res 2004;64:7263–70.
10. Boyault S, Rickman DS, de Reynies A, et al. Transcriptome classification of HCC is related to gene alterations and to new therapeutic targets. Hepatology 2007;45:42–52.
11. Chiang DY, Villanueva A, Hoshida Y, et al. Focal gains of VEGFA and molecular classification of hepatocellular carcinoma. Cancer Res 2008;68:6779–88.
12. Hoshida Y, Brunet JP, Tamayo P, Golub TR, Mesirov JP. Subclass mapping: identifying common subtypes in independent disease data sets. PLoS ONE 2007;2:e1195.
13. Brunet JP, Tamayo P, Golub TR, Mesirov JP. Metagenes and molecular pattern discovery using matrix factorization. Proc Natl Acad Sci U S A 2004;101:4164–9.
14. Hoshida Y, Villanueva A, Kobayashi M, et al. Gene expression in fixed tissues and outcome in hepatocellular carcinoma. N Engl J Med 2008;359:1995–2004.
15. Xu L, Shen SS, Hoshida Y, et al. Gene expression changes in an animal melanoma model correlate with aggressiveness of human melanoma metastases. Mol Cancer Res 2008;6:760–9.
16. Subramanian A, Tamayo P, Mootha VK, et al. Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. Proc Natl Acad Sci U S A 2005;102:15545–50.
17. Imamura H, Matsuyama Y, Tanaka E, et al. Risk factors contributing to early and late phase intrahepatic recurrence of hepatocellular carcinoma after hepatectomy. J Hepatol 2003;38:200–7.
18. Mazzaferro V, Romito R, Schiavo M, et al. Prevention of hepatocellular carcinoma recurrence with α-interferon after liver resection in HCV cirrhosis. Hepatology 2006;44:1543–54.
19. Fan JB, Yeakley JM, Bibikova M, et al. A versatile assay for high-throughput gene expression profiling on universal array matrices. Genome Res 2004;14:878–85.
20. Kaposi-Novak P, Lee JS, Gomez-Quiroz L, Coulouarn C, Factor VM, Thorgeirsson SS. Met-regulated expression signature defines a subset of human hepatocellular carcinomas with poor prognosis and aggressive phenotype. J Clin Invest 2006;116:1582–95.
21. 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.
22. Bruix J, Sherman M. Management of hepatocellular carcinoma. Hepatology 2005;42:1208–36.
23. Hoshida Y, Villanueva A, Llovet JM. Molecular profiling to predict hepatocellular carcinoma outcome. Expert Rev Gastroenterol Hepatol 2009;3:101–3.
24. Lee JS, Chu IS, Mikaelyan A, et al. Application of comparative functional genomics to identify best-fit mouse models to study human cancer. Nat Genet 2004;36:1306–11.
25. Yamashita T, Forgues M, Wang W, et al. EpCAM and α-fetoprotein expression defines novel prognostic subtypes of hepatocellular carcinoma. Cancer Res 2008;68:1451–61.
26. Sahin F, Kannangai R, Adegbola O, Wang J, Su G, Torbenson M. mTOR and P70 S6 kinase expression in primary liver neoplasms. Clin Cancer Res 2004;10:8421–5.
27. Villanueva A, Chiang DY, Newell P, et al. Pivotal role of mTOR signaling in hepatocellular carcinoma. Gastroenterology 2008;135:1972–83.
28. Vousden KH, Lane DP. p53 in health and disease. Nat Rev Mol Cell Biol 2007;8:275–83.
29. Miller JR, Moon RT. Signal transduction through β-catenin and specification of cell fate during embryogenesis. Genes Dev 1996;10:2527–39.
30. Thorgeirsson SS, Lee JS, Grisham JW. Functional genomics of hepatocellular carcinoma. Hepatology 2006;43:S145–50.
31. Zucman-Rossi J, Benhamouche S, Godard C, et al. Differential effects of inactivated Axin1 and activated β-catenin mutations in human hepatocellular carcinomas. Oncogene 2007;26:774–80.
32. Zavadil J, Bottinger EP. TGF-β and epithelial-to-mesenchymal transitions. Oncogene 2005;24:5764–74.
33. Giannelli G, Bergamini C, Fransvea E, Sgarra C, Antonaci S. Laminin-5 with transforming growth factor-β1 induces epithelial to mesenchymal transition in hepatocellular carcinoma. Gastroenterology 2005;129:1375–83.
34. Fischer AN, Fuchs E, Mikula M, Huber H, Beug H, Mikulits W. PDGF essentially links TGF-β signaling to nuclear β-catenin accumulation in hepatocellular carcinoma progression. Oncogene 2007;26:3395–405.
35. Coulouarn C, Factor VM, Thorgeirsson SS. Transforming growth factor-β gene expression signature in mouse hepatocytes predicts clinical outcome in human cancer. Hepatology 2008;47:2059–67.
36. Jian H, Shen X, Liu I, Semenov M, He X, Wang XF. SMAD3-dependent nuclear translocation of β-catenin is required for TGF-β1-induced proliferation of bone marrow-derived adult human mesenchymal stem cells. Genes Dev 2006;20:666–74.
37. Muto Y, Moriwaki H, Ninomiya M, et al. Prevention of second primary tumors by an acyclic retinoid, polyprenoic acid, in patients with hepatocellular carcinoma. Hepatoma Prevention Study Group. N Engl J Med 1996;334:1561–7.
38. Llovet JM, Bruix J. Molecular targeted therapies in hepatocellular carcinoma. Hepatology 2008;48:1312–27.