

Transcriptome Classification of HCC Is Related to Gene Alterations and to New Therapeutic Targets

Sandrine Boyault,^{1,2*} David S. Rickman,^{3*} Aurélien de Reyniès,^{3*} Charles Balabaud,^{4,5} Sandra Rebouissou,^{1,2} Emmanuelle Jeannot,^{1,2} Aurélie Hérault,^{1,2} Jean Saric,⁶ Jacques Belghiti,^{7,8} Dominique Franco,^{8,9} Paulette Bioulac-Sage,^{4,9} Pierre Laurent-Puig,¹⁰ and Jessica Zucman-Rossi^{1,2}

Hepatocellular carcinomas (HCCs) are a heterogeneous group of tumors that differ in risk factors and genetic alterations. We further investigated transcriptome-genotype-phenotype correlations in HCC. Global transcriptome analyses were performed on 57 HCCs and 3 hepatocellular adenomas and validated by quantitative RT-PCR using 63 additional HCCs. We determined loss of heterozygosity, gene mutations, promoter methylation of *CDH1* and *CDKN2A*, and HBV DNA copy number for each tumor. Unsupervised transcriptome analysis identified 6 robust subgroups of HCC (G1-G6) associated with clinical and genetic characteristics. G1 tumors were associated with low copy number of HBV and overexpression of genes expressed in fetal liver and controlled by parental imprinting. G2 included HCCs infected with a high copy number of HBV and mutations in *PIK3CA* and *TP53*. In these first groups, we detected specific activation of the AKT pathway. G3 tumors were typified by mutation of *TP53* and overexpression of genes controlling the cell cycle. G4 was a heterogeneous subgroup of tumors including *TCF1*-mutated hepatocellular adenomas and carcinomas. G5 and G6 were strongly related to β -catenin mutations that lead to Wnt pathway activation; in particular, G6 tumors were characterized by satellite nodules, higher activation of the Wnt pathway, and E-cadherin underexpression. **Conclusion:** These results have furthered our understanding of the genetic diversity of human HCC and have provided specific identifiers for classifying tumors. In addition, our classification has potential therapeutic implications because 50% of the tumors were related to WNT or AKT pathway activation, which potentially could be targeted by specific inhibiting therapies. (HEPATOLOGY 2007;45:42-52.)

From ¹INSERM U674, CEPH—Fondation Jean Dausset, Paris France; ²Université Paris 7, IUH, Saint-Louis, Paris, France; ³Bioinformatic Department, Ligue Nationale Contre le Cancer (LNCC), Paris, France; ⁴INSERM E0362, GREC, Université Bordeaux 2, CHU Bordeaux, France; ⁵Department of Hepatology, Centre Hospitalier Universitaire de Bordeaux, Bordeaux, France; ⁶Department of Surgery, Centre Hospitalier Universitaire de Bordeaux, Bordeaux, France; ⁸AP-HP, Surgery Department, Hôpital Beaujon, Clichy, France; ⁹AP-HP, Surgery Department, Hôpital Antoine Béchère, Clamart, France; ¹⁰INSERM U775, Université Paris 5, Paris, France.

Received July 13, 2006; accepted August 24, 2006.

Supported by the Ligue Nationale Contre le Cancer (including a fellowship to S.R.); Fondation de France (including a fellowship to S.B.); and Association pour la Recherche sur le Cancer (fellowship to E.J.).

*These authors contributed equally to this work.

Address reprint requests to: Jessica Zucman-Rossi, INSERM U674, CEPH—Fondation Jean Dausset, IUH-Saint-Louis, 27 rue Juliette Dodu, 75010 Paris, France. E-mail: zucman@cepb.fr; fax: (33) 1-53-72-51-58.

Copyright © 2006 by the American Association for the Study of Liver Diseases. Published online in Wiley InterScience (www.interscience.wiley.com).

DOI 10.1002/hep.21467

Potential conflict of interest: Nothing to report.

Supplementary material for this article can be found on the HEPATOLOGY website (<http://interscience.wiley.com/jpages/0270-9139/suppmat/index.html>).

Hepatocellular carcinoma (HCC) is one of the most frequently occurring solid tumors worldwide and is the third-leading cause of death from cancer.¹ Cirrhosis of any origin and dysplastic regenerative nodules have long been considered the likely precursors of HCC because of their frequent association with HCC occurrence.^{2,3} As in other solid tumors, a large number of genetic alterations accumulate during the carcinogenetic process. Some of these genetic alterations are specific to HCC etiological risk factors, particularly HBV infection, which can induce chromosome instability or insertional mutagenesis.⁴ Among the genetic alterations unrelated to HCC risk factors, microsatellite allelotypes and comparative genomic hybridization (CGH) studies have demonstrated recurrent chromosome aberrations.⁵ Altogether, the principal carcinogenetic pathways known to be deregulated in HCC are inactivation of TP53,⁶ Wnt/wingless activation mainly through *CTNNB1* mutations activating β -catenin- and *AXIN1*-inactivating mutations,⁷⁻⁹ retinoblastoma inactivation through *RBI* and *CDKN2A* promoter methylation and rare gene mu-

Table 1. Associations Between Transcriptomic Groups and Clinical, Pathological, and Genetic Variables

	Associated group	Affymetrix hybridizations (57 HCCs)	QRT-PCR validation set (63 HCCs)	Complete set (120 HCCs)
<i>Clinical characteristics</i>				
AFP > 100 IU/mL	G1	.01	.006	< 10 ⁻⁴
Female	G1	.06	.05	< 10 ⁻²
HBV low	G1	< 10 ⁻⁴	.04	10 ⁻⁵
HBV high	G2	.05	.07	< 10 ⁻²
Age < 60 years	G1 and G2	.04	.09	< 10 ⁻²
African origin	G1	< 10 ⁻²	.3	< 10 ⁻²
Hemochromatosis	G2	1	< 10 ⁻³	.04
<i>Pathological characteristics</i>				
Satellite nodules	G6	< 10 ⁻²	.3	< 10 ⁻²
Genetic alterations				
FAL > 0.128	G1, G2, and G3	< 10 ⁻²	< 10 ⁻³	< 10 ⁻⁵
CTNNB1 mutation	G5 and G6	< 10 ⁻¹⁰	< 10 ⁻⁵	< 10 ⁻¹⁵
TP53 mutation	G2 and G3	.03	.001	< 10 ⁻³
AXIN1 mutation	G1 and G2	.1	< 10 ⁻²	< 10 ⁻²
PIK3CA mutations	G2	.01	—	.01
CDH1 methylation	G5 and G6	.01	< 10 ⁻²	< 10 ⁻²
LOH 17p	G2 and G3	.005	.02	< 10 ⁻³
LOH 16p	G1, G2, and G3	.005	.05	.0005
LOH 16q	G1	.05	.04	< 10 ⁻²
LOH 4q	G1, G2, and G3	.002	.002	3.10 ⁻⁶
LOH 5q	G3	.02	.02	< 10 ⁻³
LOH 13q	G2	.02	< 10 ⁻²	< 10 ⁻³
LOH 21q	G3	< 10 ⁻²	.03	< 10 ⁻³
LOH 22q	G3	< 10 ⁻²	.02	< 10 ⁻³

Shown are *P* values obtained from Fisher exact tests based on the given genetic or clinical variable.

tations,^{10,11} insulin growth factor activation through *IGF2* overexpression, and *IGF2R*-inactivating mutations.^{12,13}

We previously showed that genetic alterations are indeed closely associated with clinical characteristics of HCC that define 2 mechanisms of hepatocarcinogenesis.¹⁴ The first type of HCC was associated with not only a high level of chromosome instability and frequent *TP53* and *AXIN1* mutations but also was closely linked to HBV infection and a poor prognosis. Conversely, the second subgroup of HCC tumors was chromosome-stable, having a high incidence of activating β -catenin alterations, and was not associated with viral infection. To further investigate genotype–phenotype correlations in HCC and identify pathways and/or biological processes deregulated in such heterogeneous tumors, we performed a comprehensive analysis at the clinical, genetic, and transcriptomic level in 57 HCCs and 3 adenomas, and we validated the results using quantitative RT-PCR in 63 new independent HCCs.

Patients and Methods

Tumor Samples and Clinical Data. A series of 120 HCCs and 3 hepatocellular adenomas were collected from 123 patients surgically treated in 3 French surgical departments from 1992 to 1999. A first group of tumors

(57 HCCs and 3 hepatocellular adenomas) was used for genome-wide transcriptome microarray analysis, and a second group of 63 independent HCCs was used for validation by quantitative RT-PCR analysis. All these tumors were clinically characterized as previously described¹⁴ (Supplementary Table 1). Briefly, the 123 patients had a 4:1 male/female sex ratio, 79% were born in France, and their mean age was 60 years. Thirty percent, 25%, 33%, and 5% of the patients had the HCC risk factors of hepatitis B virus, hepatitis C virus, alcohol abuse, and hemochromatosis, respectively. In 68% of the patients, their HCCs developed from cirrhotic liver or severe chronic hepatitis disease. Histological grade of tumor differentiation was assigned according to the Edmondson and Steiner grading system: 7% were grade I, 49% were grade II, 39% were grade III, and 4% were grade IV. Tumor size ranged from 3 to 190 mm (median = 60 mm). The preoperative α -fetoprotein serum level was available for 103 patients, which showed that 37 patients had a level of more than 100 IU/mL. Macroscopic and/or microscopic vascular invasion was recorded for 38% of the patients. Satellite tumors defined by nodules found less than 1 cm from the main tumor were recorded for 42%. For the Affymetrix analysis, 5 pools of 3 nontumorous liver tissues matching the analyzed tumors were used. They were alcoholic cirrhosis (pool 1), alcoholic

noncirrhotic liver (pool 2), HBV noncirrhotic liver (pool 3), HCV cirrhosis (pool 4), and HBV cirrhosis (pool 5). In the QRT-PCR experiments, we analyzed the RNA of these 15 nontumorous tissue samples and of 6 additional normal nontumorous liver samples. We used 19 human fetal liver samples at different stages of development (ranging from 11 to 29 weeks of pregnancy) to test for genes potentially expressed during fetal life using QRT-PCR. The study was approved by the local Ethics Committee (CCPPRB Paris Saint-Louis), and informed consent was obtained in accordance with French legislation.

Gene Mutations, Chromosome Imbalance, Quantification of HBV Genome and DNA Methylation. For each tumor, we screened for gene mutations in *TP53*, *CTNNB1* coding for β -catenin, *AXIN1*, *PIK3CA*, *EGFR*, *KRAS*, *NRAS*, and *HRAS* using direct sequencing. Primers and protocols are available on request, and all mutations are described in Supplementary Table 2. Genomewide allelic loss was assessed as previously described.¹⁴ For all samples related to HBV infection either by serological results or viral DNA amplification, *HBS* and *HBX* copies of DNA were quantified in tumor and nontumor DNA using the Syber Green (Applied Biosystems) method (for further details, see Supplementary Materials and Methods). DNA methylation at the *CDH1* and *CDKN2A* promoter was assessed using bisulfite DNA and methylation-specific amplification as previously described.^{15,16}

Gene Expression Analysis. Microarray analyses of 57 HCCs, 3 hepatocellular adenomas, and 5 pools of nontumorous samples were performed using 5 μ g of total RNA and 20 μ g of cRNA (GeneChip Fluidics Station 400) of HG-U133A Affymetrix GeneChipTM arrays per hybridization and analyzed following the manufacturer's protocols (all methodological details of the gene expression analysis are given in Supplementary Materials and Methods).

Quantitative RT-PCR Data Processing. Quantitative RT-PCR of 109 HCCs (46 from the "microarray" series of 57 HCCs and 63 new samples) and 21 nontumor liver samples was performed in duplicate, using TaqMan[®] Low Density Array and the ABI PRISM[®] 7900HT System (Applied Biosystems). Briefly, expression of a gene was normalized to internal control ribosomal 18S relative to the mean expression of the corresponding gene in nontumorous samples normalized to internal control ribosomal 18S.

Western Blot and Immunohistochemistry. Frozen tissue was homogenized in RIPA Lysis buffer (Santa Cruz), and protein concentration was determined by a BCA protein assay kit (Pierce). Immunoblot analysis was

performed using the antibodies E-cadherin (SC-7870, 1:500; Santa Cruz), AKT (#9272, 1:2000; Cell Signaling), phospho-AKT ser473 (#9271, 1:200; Cell Signaling), phospho-GSK3 ser9 (#9336, 1:500; Cell Signaling), IGF1R β (#3027, 1:500; Cell Signaling), peroxidase-conjugated secondary antibody (1:2000; Santa Cruz), and enhanced chemoluminescence (ECL; Pierce). Immunostaining was performed using monoclonal anti- β -catenin as previously described.¹⁷

Results

Unsupervised Transcriptome Analysis Defines Clusters of Tumors Closely Associated with Clinical Data and Genetic Alterations

In the whole series of 123 tumors (120 HCCs and 3 hepatocellular adenomas) we identified mutations in the *CTNNB1* (encoding β -catenin), *TP53*, *AXIN1*, *TCF1*, *PIK3CA*, and *KRAS* genes in 34, 31, 13, 5, 2, and 1 tumors, respectively (Supplementary Table 2). No mutations were found in *NRAS*, *HRAS*, and *EGFR*. Hypermethylation of the *CDKN2A* and *CDH1* promoter was identified in 35% and 16% of the tumors, respectively.

Fifty-seven HCCs, 3 hepatocellular adenomas, and 5 samples of pooled nontumorous tissues were analyzed using Affymetrix HG-U133A GeneChipTM arrays. On the basis of the results of a hierarchical clustering analysis using the 6,712 probe sets whose expression varied the most across samples, we classified tumors into 2 major groups with each further divided into 3 subgroups (G1-G6; Fig. 1). This classification was found to be extremely robust—the mean reproducibility score was found to be at least 90% for all major groups and subgroups—and was conserved across different gene lists and clustering methods.

We observed a high level of association between this HCC classification and genetic alterations and clinical factors (Table 1). The tumors in groups G1-G3 were associated with high chromosomal instability, yielding significantly higher fractional allelic loss (FAL), compared to the tumors in groups G4-G6 ($P < 10^{-2}$). Among the frequent somatic gene mutations observed, the G5-G6 subgroups were highly associated with *CTNNB1* mutations ($P < 10^{-10}$) and the G2-G3 subgroups with *TP53* mutations ($P = 0.03$). The rare *TCF1* and *PIK3CA* mutations were associated with the G4 subgroup ($P = 0.01$) and the G2 subgroup ($P = 0.01$), respectively. Hypermethylation of E-cadherin (*CDH1*) and P16 (*CDKN2A*) gene promoter were most frequently observed in the G5-G6 and G3 subgroups, respectively. In addition, HCC related to HBV infection was found in groups G1-G3; more specifically, tumors with a low number of HBV DNA copies were related to the G1 subgroup ($P < 10^{-4}$).

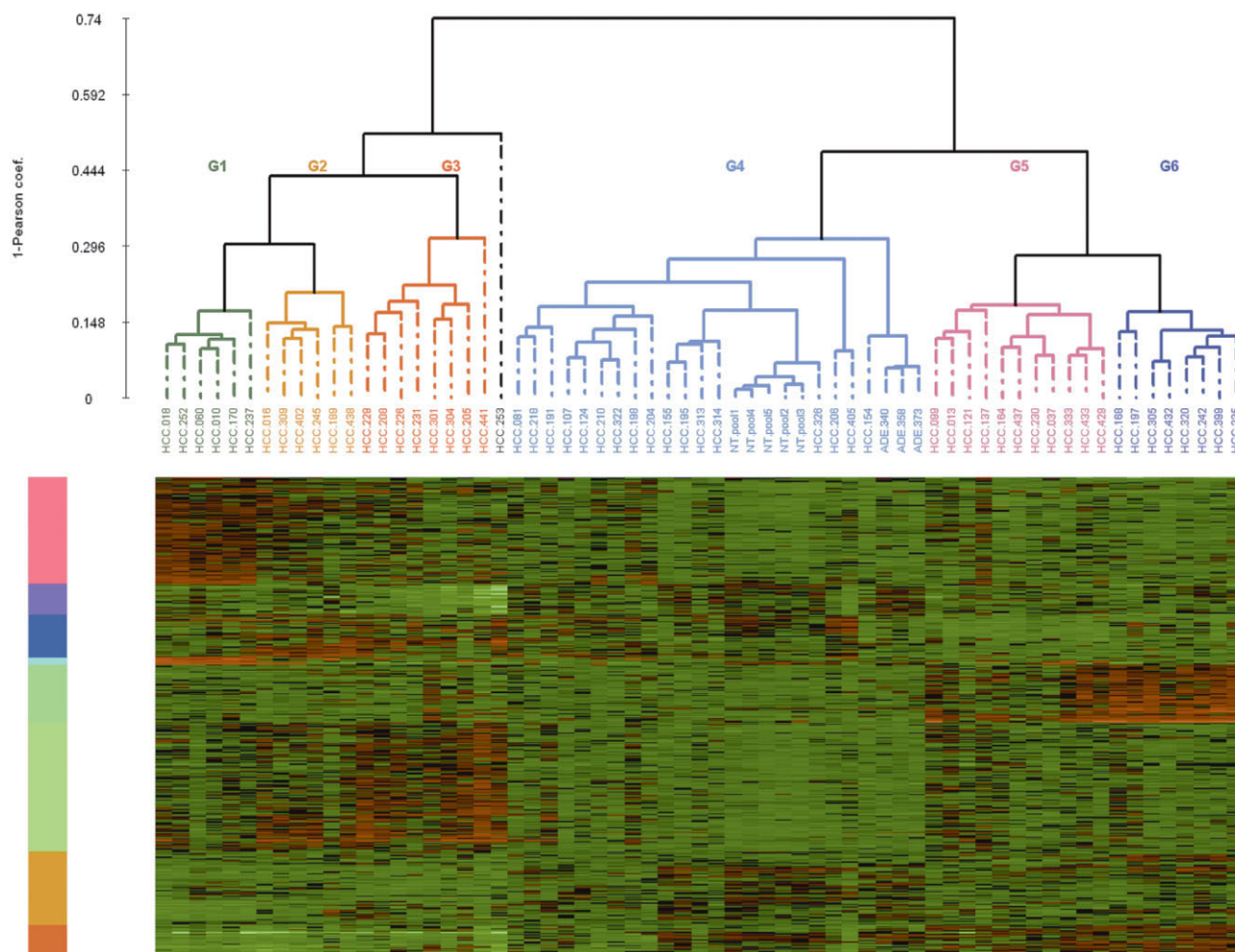


Fig. 1. Unsupervised hierarchical clustering. The dendrogram in the upper panel was obtained on the basis of the expression profile of the 6,712 probe sets whose expression varied the most across samples from 57 HCC tumors, 3 hepatocellular adenomas, and 5 pools of nontumor tissues using Ward's linkage and 1-Pearson correlation coefficient. The different HCC subgroups (G1-G6) are indicated in color. The heat map, in the lower panel, shows the expression of 493 probe sets corresponding to a subselection of HCC subgroup-specific genes (supplementary Table 4). For each probe set, data are median-centered (black), with the lowest and highest intensity values in green and red, respectively.

We validated these associations on the basis of the classification of a second set of 63 independent tumors using quantitative RT-PCR.

Identification of a Predictor of the 6-Group Classification

To more easily classify tumors, we sought to define a gene predictor that would use quantitative RT-PCR data (for details, see Supplementary Materials and Methods and Supplementary Table 3). We identified a combination of 16 genes (Fig. 2) that correctly classified 45 of 46 HCCs previously characterized in the Affymetrix experiment. We then used this signature to partition the validation series of 63 independent HCC samples into 6 subgroups. As observed in the first set of tumors analyzed in the Affymetrix experiment, significant associations were found between genetic alterations and the different

predicted subgroups (Table 1). These associations were confirmed using the complete series of 120 HCC tumors classified with either Affymetrix and QRT-PCR ($n = 46$), Affymetrix only ($n = 11$), or QRT-PCR only ($n = 63$); see Table 1 and Figure 3. We further searched to identify deregulated pathways in each groups of tumors.

Identification of Key Signaling Pathways Implicated in HCC Subgroups

HCC groups G1-G3 (associated with a high rate of chromosomal instability) were enriched with overexpression of cell-cycle/proliferation/DNA metabolism genes ($P < 0.01$; see Supplementary Tables 4 and 5).

Parentally Imprinting Genes Were Overexpressed in HCC Subgroup G1. Among the high number of genes specifically overexpressed in the G1 subgroup, which was associated with HBV infection with a low

A

Predicted class ($sample_i$) = $\arg \min_{k=1..6} (\text{Distance}(sample_i, class_k))$

where : $\text{Distance}(sample_i, class_k) = \sum_{t=1..16} \frac{(\Delta Ct(sample_i, gene_t) - \mu(class_k, gene_t))^2}{\sigma(gene_t)}$

B

μ	class 1 (G1)	class 2 (G2)	class 3 (G3)	class 4 (G4)	class 5 (G5)	class 6 (G6)	σ
gene 1 (RAB1A)	-16.04	-16.39	-16.29	-17.15	-17.33	-16.95	0.23
gene 2 (PAP)	-27.02	-28.75	-23.48	-27.87	-19.23	-11.33	16.63
gene 3 (NRAS)	-17.41	-16.92	-16.25	-17.31	-16.96	-17.26	0.27
gene 4 (RAMP3)	-23.12	-23.54	-25.34	-22.36	-23.09	-23.06	1.23
gene 5 (MERTK)	-18.43	-18.72	-21.24	-18.29	-17.03	-16.16	7.23
gene 6 (PIR)	-19.81	-18.44	-16.73	-18.28	-17.09	-17.25	0.48
gene 7 (EPHA1)	-16.51	-16.68	-19.89	-17.04	-18.70	-21.98	1.57
gene 8 (LAMA3)	-20.44	-20.58	-20.19	-21.99	-18.77	-16.85	2.55
gene 9 (GUS2)	-17.45	-14.82	-18.18	-14.78	-17.99	-16.06	3.88
gene 10 (HNF1)	-17.16	-16.92	-15.91	-17.88	-17.72	-17.93	0.54
gene 11 (PAK2)	-16.56	-17.86	-16.99	-18.14	-17.92	-17.97	0.58
gene 12 (AFP)	-12.36	-16.68	-26.80	-27.28	-25.97	-23.47	14.80
gene 13 (CYP2C9)	-16.99	-18.27	-16.26	-16.23	-13.27	-14.44	5.47
gene 14 (CDH2)	-14.76	-15.20	-18.91	-15.60	-15.48	-17.32	10.59
gene 15 (HAMP)	-20.19	-19.53	-21.32	-18.51	-25.06	-26.10	13.08
gene 16 (SAE1)	-17.10	-17.37	-16.79	-18.22	-17.72	-18.16	0.31

Fig. 2. Predictor of HCC classification using quantitative RT-PCR. (A) Formula used for class-membership prediction. To classify a new sample requires the ΔCt values (control: 18S) of the 16 genes used in the formula. Having calculated the distance between the given sample and the centroid representation of each class, we associated the new sample to the closest class. (B) Parameters for each gene and for each class used in the formula.

number of viral DNA copies, *AXIN1* mutations, younger age, high serum level of AFP, and frequent origin from Africa (Fig. 3, Table 1, and Supplementary Table 4), we found genes encoding proteins expressed during development, myosin heavy chain IIb (*MYH4*) and the transcription factor *SOX9*, and the parentally imprinted genes insulin-like growth factor 2 (*IGF2*), paternally expressed genes 3 and 10 (*PEG3* and *PEG10*), alpha-fetoprotein (*AFP*), and sarcoglycan epsilon (*SGCE*). The differential expression of all these genes was validated using QRT-PCR on 109 HCC tumors (Fig. 4A). The parentally imprinted genes were highly overexpressed in normal fetal liver. *H19* mRNA was also overexpressed, not only in the G1 samples but also in fetal samples, correlating with the level of *IGF2* mRNA expression in these 2 groups ($R^2 = 0.4$ and 0.6 , respectively). Interestingly, in the unsupervised classification, the magnitude of expression of AFP mRNA was identified as a discriminating marker of the G1 tumors and was observed to be correlated with a high AFP serum level (Table 1 and Fig. 3).

HCC Tumors in Subgroup G2 Included the Rare *PIK3CA* Mutated Cases. Subgroup G2 tumors were related to HBV infection with a high number of viral DNA copies, frequent local and vascular invasion, and *TP53* mutations (Table 1 and Fig. 3). An association with hemochromatosis-related tumors was only observed in the validation set, possibly because the number of tested patients overall was small ($n = 6$, Table 1). Interestingly, in the whole series we identified mutations in the *PIK3CA* gene in only 2 tumors that belonged to the G2 group. These 2 samples were closely associated in the unsupervised clustering analysis (HCC189 and HCC438; Fig. 1). We identified 38 genes specifically overexpressed in the *PIK3CA*-mutated samples when compared with the other

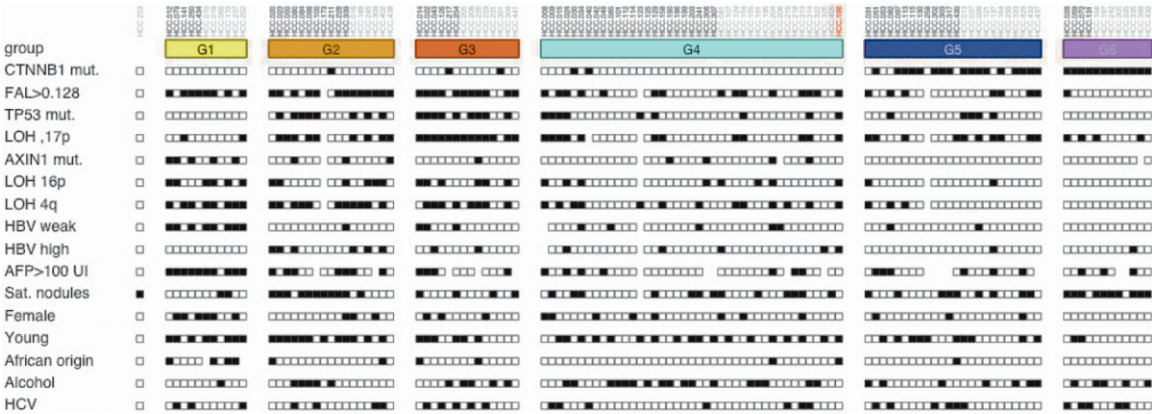


Fig. 3. Clinical and genetic annotations for the classification of all 120 HCCs. Samples (in columns) are ordered by group (G1-G6). Sample names are in gray (56 Affymetrix-classified samples), red (1 HCC misclassified by QRT-PCR), or black (independent series of 63 HCCs). Features that are positive and negative are indicated in black and white boxes, respectively (FAL, fractional allelic loss, in black tumors deleting more than 5 chromosome arms; Mut, mutation; sat. nodules, satellite nodules less than 1 cm from the principal tumor; AFP, alpha-fetoprotein).

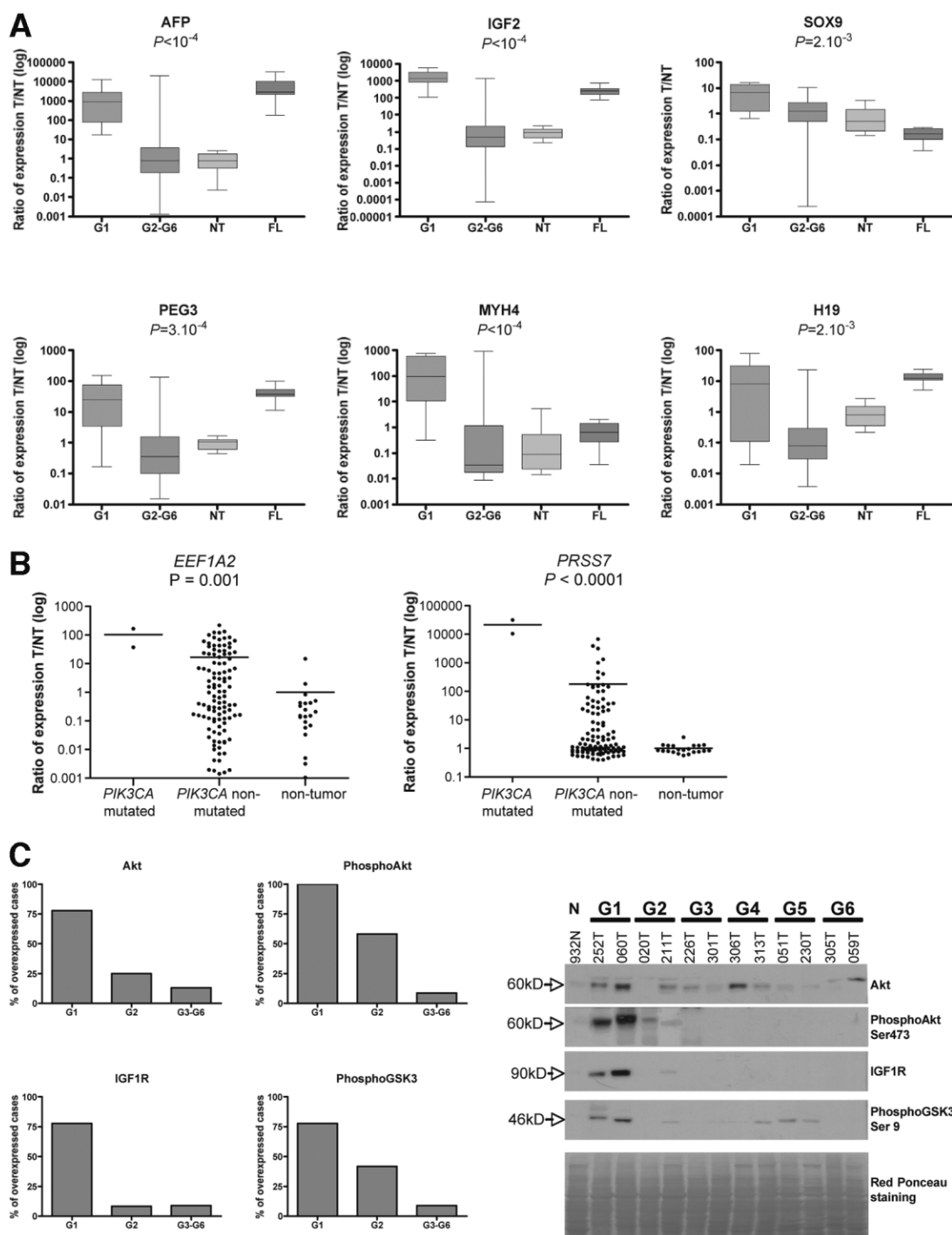


Fig. 4. Validation of selected genes and proteins specific for HCC subgroups G1 and G2. (A) Validation of *AFP*, *IGF2*, *SOX9*, *PEG3*, *MYH4*, and *H19* gene expression using QRT-PCR. Box plots (from 25th percentile to the 75th percentile with a line at the median) show the range of relative expression (tumor versus the mean of 21 nontumors (T/NT)) in the predicted subgroups G1 ($n = 10$) and G2 to G6 ($n = 99$), 21 nontumor samples (NT), and 19 fetal liver samples (FL). The whiskers extend above and below the box to show the first and ninth deciles. P values from the Kruskal-Wallis test are indicated. (B) Validation of *EEF1A2* and *PRSS7* genes overexpressed in *PIK3CA*-mutated tumors compared to 107 nonmutated HCCs. P values from a t test are shown. (C) Protein expression analysis of Akt, phosphoAkt (Ser473), PhosphoGSK3 β (Ser9), and IGF1R was performed using Western blotting in 44 HCCs distributed in all transcriptomic subgroups. The graphs show the percentage of tumors that overexpress these proteins in the G1, G2, and G3-G6 subgroups compared to nontumor tissues. The panel on the right illustrates protein expression profiles of the different HCC subgroups.

tumors in groups G1-G3. Using QRT-PCR, we validated overexpression of 2 of these genes, the genes coding for protein elongation factor *EEF1A2* and enterokinase *PRSS7*, specifically overexpressed in *PIK3CA*-mutated tumors ($P = 0.001$, Fig. 4B).

IGF2 overexpression in G1 and *PIK3CA* mutations in G2 were predicted to activate the AKT pathway. To test this hypothesis, we searched for possible activation of the AKT pathway using Western blotting in a series of 44 HCCs representing each of the transcriptomic groups. We found specific overexpression of Insulin-like growth factor 1 receptor (IGF1R) in 7 of 9 G1 tumors in contrast to IGF1R overexpression found in only 9% of the tumors in the other groups ($P < 10^{-4}$; Fig. 4C). We also identified frequent overexpression of AKT in groups G1 (75%) and G2 (25%), compared to only 13% found in the other groups. Moreover, phosphorylated AKT proteins were overexpressed in all tumors of G1 and in 58% of the G2 tumors in contrast to only 8% in the other groups, demonstrating frequent activation of AKT in the tumors in groups G1 and G2 ($P < 10^{-3}$; Fig. 4C). Accordingly, GSK3 β , a target of activated AKT, was found to be phosphorylated in 78% of G1 tumors, in 42% of G2 tumors, but in only 9% of the tumors in the other groups ($P < 10^{-3}$; Fig. 4C).

HCCs in Subgroup G3 Overexpressed Genes Encoding Proteins of the Nuclear Pore and That Control the Cell Cycle. HCC subgroup G3 mainly included tumors with *TP53* mutation but without HBV infection. Significant associations were found with chromosome losses at 17p, 5q, 21q, and 22q (Fig. 3 and Table 1), and we frequently observed a high incidence of *CDKN2A* promoter hypermethylation (data not shown). Examples of key genes specifically differentially expressed in the G3 group are provided in Supplementary Figure 1. We observed overexpression of a large number of genes implicated in cell-cycle control ($P < 0.0003$), including those encoding proteins that control cell-cycle checkpoints (*CDC6*, *MAD2L1*, *BUB1*, *TTK*, *SMC1L1*), cyclins (*CCNA2*, *CCNE2*), and DNA replication checkpoints (*MCM2*, *MCM3*, *MCM6*, *ASK*). Interestingly, genes encoding proteins implicated in nucleus import/export were also overexpressed, including 6 proteins of the nuclear pore (*KPNB1*, *RANBP5*, *XPO1*, *IPO7*, *NUP155*, *NUP107*).

G4 Was a Heterogeneous Group of HCC Tumors. The G4 subgroup of samples comprised the 5 sample pools of nontumorous liver tissues, which clustered tightly together within a large, heterogeneous group containing 20 tumors (Fig. 1). We also observed that 4 tumors, all with a *TCF1* mutation, were closely associated in a small cluster of 5 tumors including the 3 hepatocellular

adenomas and 2 well-differentiated HCCs. These results suggest a possible continuum of early transcriptome alterations between nontumorous, benign, and HCC tissues. We did not find other recurrent or unifying genetic and/or clinical characteristics in the remaining small clusters of G4 HCCs (Fig. 1).

Activation of the Wnt Pathway in G5 and G6 with Underexpression of Cell Adhesion Proteins in G6 Is Related to Satellite Tumor Nodules. G5 and G6 HCCs were highly related to β -catenin activation (70% and 100% of the tumors contained *CTNNB1* mutations, respectively). No chromosome deletions specific to G5 and/or G6 were identified. In a search for possible β -catenin-targeted genes, we found 83 genes to be significantly overexpressed in G5 and/or G6 HCCs (Supplementary Table 4). All but 2 of these genes were also found to be significantly differentially expressed in a supervised test comparing β -catenin-mutated and -nonmutated tumors. In addition to overexpression of *GPR49*, *GLUL*, and *PAP/HIP*, 3 well-known β -catenin target genes in the liver,¹⁸⁻²⁰ overexpression of 6 putative β -catenin target genes potentially implicated in tumorigenesis was confirmed using QRT-PCR (Fig. 5A). We observed that the expression of all these putative β -catenin-targeted genes was more significant in G6 than in G5, even after exclusion of the samples without *CTNNB1* mutation. We also found greater overexpression of β -catenin in G6 tumors than in G5 tumors combined with a loss of signal at the plasma membrane and a strong localization in cytoplasm and nucleus (Fig. 5B). Consistent with this observation, we found overexpression of *LEF1*, a transcription factor that interacts with β -catenin to activate Wnt-responsive target genes, in G6 tumors. In addition, we found underexpression of *CDH1* (encoding E-cadherin) in the G6 subgroup (Fig. 5A), which may have accounted for the local invasion of these HCCs as shown by the quasi-constant presence of satellite nodules around the principal tumor (Fig. 3 and Table 1). We showed that the level of *CDH1* mRNA down-regulation was highly related to the down-regulated expression of the E-cadherin protein in G6, consistent with the high level of promoter methylation of *CDH1* in these tumors (Fig. 5C). Interestingly, in G5-related tumors, we observed enrichment of underexpressed genes involved in the response to biotic stimuli and immune response such as *ARHG-DIB*, *HLA-DPA1/B1*, *IFI16*, *IFI44*, *PTGER4*, *STAT1*, and *CLECSF2* ($P < 0.003$). Finally, combining transcriptome and immunohistochemical data, we found no evidence of significant activation of the β -catenin pathway in HCC subgroups G1-G4.

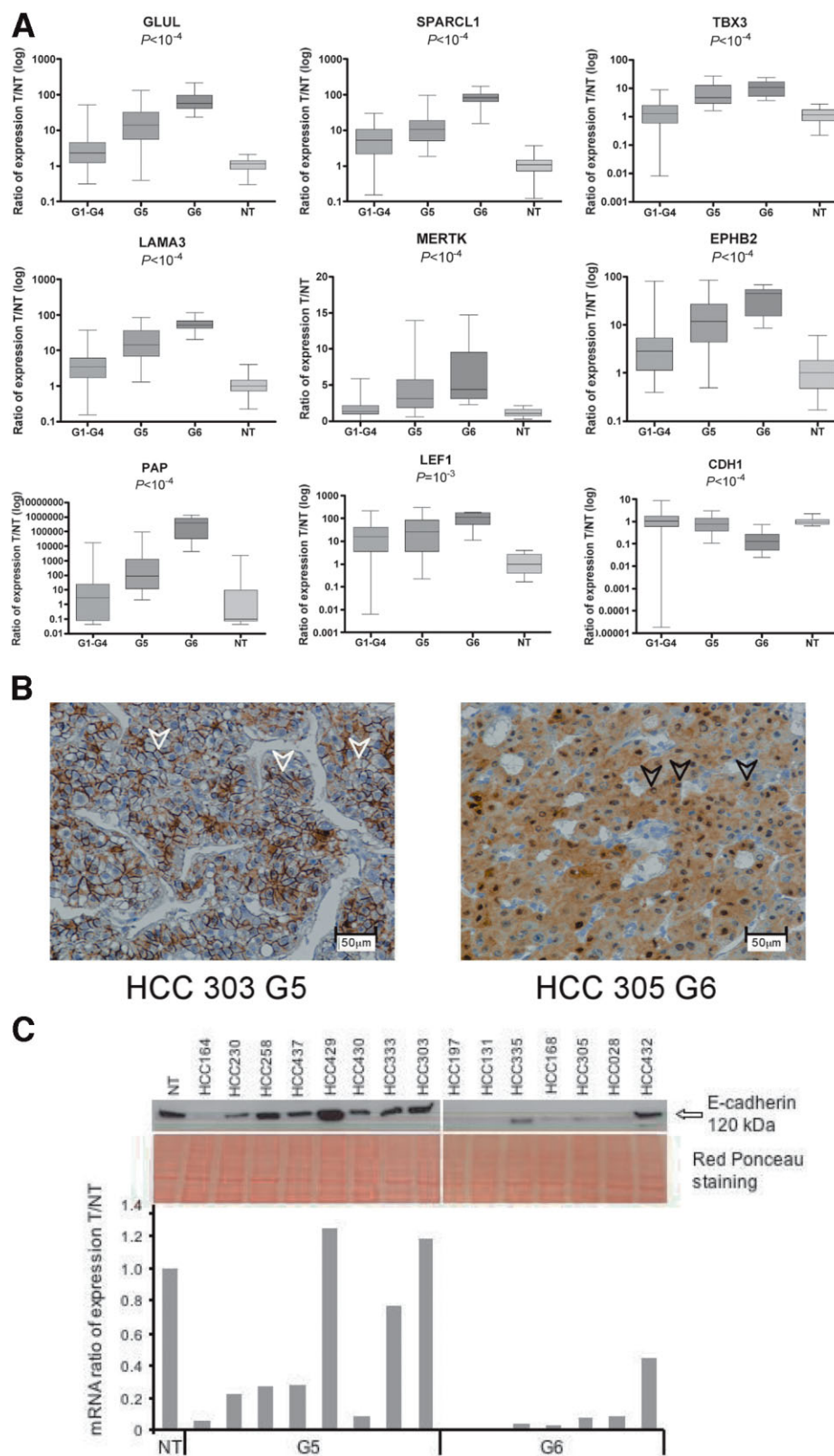


Fig. 5. Characterization of HCC tumors leading to G5 and G6 subgroups. (A) Validation of genes using QRT-PCR. Relative expression (tumor versus mean of nontumors [T/NT]) for *GLUL*, *SPARCL1*, *TBX3*, *LAMA3*, *MERTK*, *EPHB2*, *PAP*, *LEF1*, and *CDH1* was analyzed in tumors related to G1-G4 ($n = 75$), G5 ($n = 23$), G6 ($n = 11$), and nontumor samples ($n = 21$). P values from the Kruskal-Wallis test are indicated. (B) β -Catenin immunostaining in representative cases of HCC mutated for β -catenin and leading to G5 and G6. In case HCC303 (G5), note the small number of stained nuclei and the intense staining of the plasma membrane (white arrows). In case HCC305 (G6), the cytoplasm and nuclei of hepatocytes are intensely stained (black arrows) without a signal at the plasma membrane. (C) Protein expression of E-cadherin using Western blotting (upper panels) and mRNA expression using QRT-PCR (lower panel) in HCCs of groups G5 and G6.

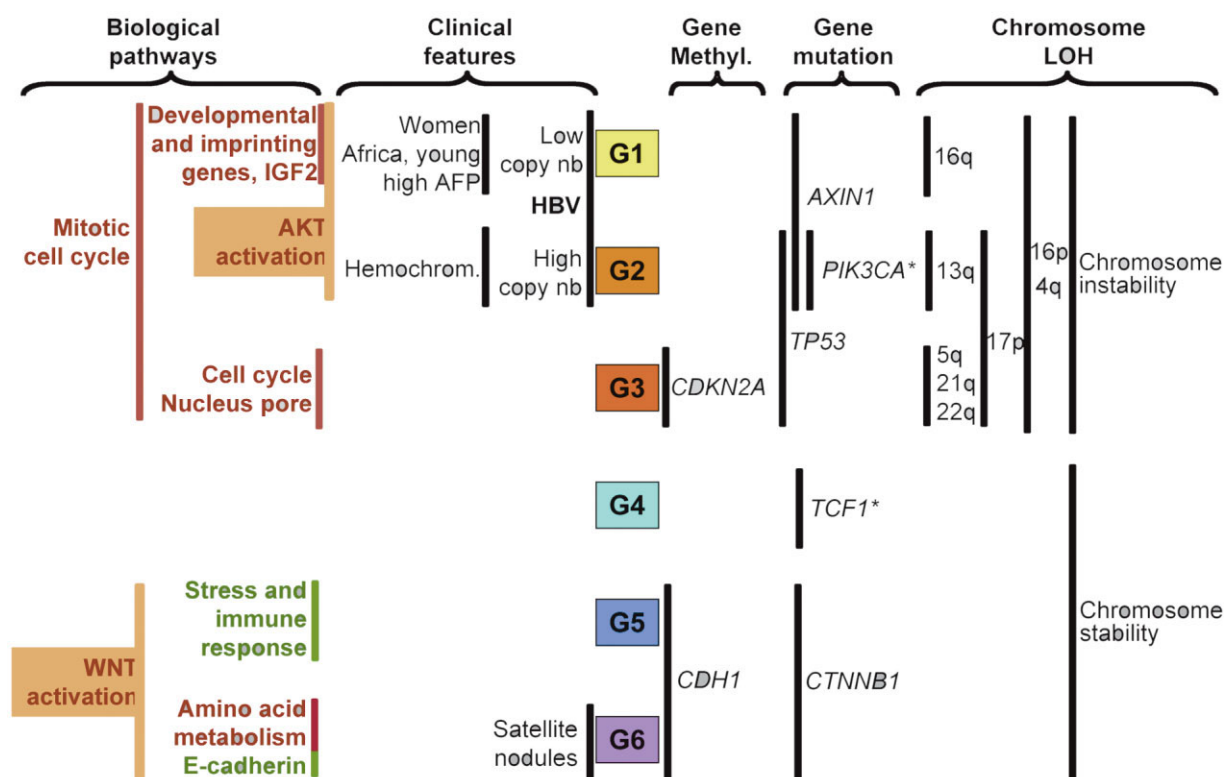


Fig. 6. Schematization of the different HCC subgroups defined by transcriptome analysis with their related clinical and genetic pathways. G1 to G6 are the subgroups of HCCs defined by transcriptome analysis. Vertical lines indicate significantly associated features (see Table 1, Fig. 3, and supplementary Table 5). Red and green primarily indicate over- and underexpressed genes, respectively, in that particular functional category. LOH, loss of heterozygosity; Hemochrom, hemochromatosis; AFP, alpha-fetoprotein; HBV, hepatitis B virus; *rare feature.

Discussion

To elucidate the molecular diversity of HCC tumors without any a priori assumptions, we used an unsupervised transcriptome-wide approach to classify a large number of tumors that have been extensively clinically and genetically annotated. Using this approach, we obtained a robust classification of HCC that yielded 6 main subgroups reflecting the large spectrum of these tumors.^{1,2,21} This classification was achieved by QRT-PCR using 16 diagnostic genes and was applied to an independent set of 63 tumors, for a total of 120 HCCs studied. Our classification is in full agreement with previously published transcriptomic analyses of HCC,²²⁻²⁴ which identified 2 tumor groups linked to chromosome stability/instability (corresponding to meta-groups 1-3 and 4-6). However, our study went further by exquisitely refining this classification. Indeed, the strength of our study resides in our inclusion of a vast amount of medical data: (1) our series of tumors of patients surgically treated in France who had the main risk factors for HCC (HBV infection, HCV infection, alcohol abuse, and hemochromatosis), and (2) an exhaustive number of clinical, histopathological, and genetic annotations. From the results

of our study, we have concluded that the primary clinical determinant of class membership is HBV infection, and the other main determinants are genetic and epigenetic alterations, including chromosome instability ($FAL > 0.128$), *CTNNB1* and *TP53* mutations, and parental imprinting (Fig. 6).

The natural history of HCC has taught us that HBV-related tumors (defining the G1 and G2 subgroups) are strikingly molecularly distinct from other HCCs. Indeed, tumors related to HCV infection and alcohol abuse were interspersed across subgroups G3-G6. Our transcriptomic classification has enabled the identification of new entities of tumors. Subgroup G1 included HBV-related tumors from younger patients (relative to the other HBV HCCs), frequently from Africa, with an equal sex ratio, a low number of viral DNA copies, frequent *AXIN1* mutations, absence of *TP53* mutation, and overexpression of genes normally imprinted (Fig. 6). These results suggest that HBV infection early in life^{25,26} leads to a specific type of HCC that has immature features with abnormal parental gene imprinting selection, possibly through the persistence of fetal hepatocytes or alternatively through partial dedifferentiation of adult hepatocytes. These G1 tumors

are related to the high-risk populations found in epidemiological studies.^{4,27} Subgroup G6 had a 100% incidence of *CTNNB1* mutation, a high level of Wnt pathway activation (higher than in G5), and inactivation of E-cadherin consistent with the high invasive potential of these tumors, as E-cadherin inactivation is known to participate in the cell invasion process.²⁸ Interestingly, *AXIN1*-mutated tumors segregated to different groups that were not characterized by harboring a *CTNNB1* mutation, suggesting that β -catenin and Axin1 may operate in distinct ways to promote carcinogenesis. In human HCCs, the inactivating mutations of the *AXIN1* gene seem to differ strongly from the gain of function of β -catenin, as confirmed by *in vitro* experiments.¹⁷

Apart from these large subgroups of tumors defined by frequent genetic alterations, our analysis also defined homogeneous subgroups of tumors related to rare genetic alterations like *TCF1* or *PIK3CA* mutations.^{29,30} However, mutations specific to other small homogeneous subgroups of tumors have yet to be identified.

Thorgeirsson and collaborators defined a hepatoblast-like subtype of HCC that exhibited the most severe prognostic characteristics.³¹ Interestingly, our G1 tumor group shared the same characteristics as this hepatoblast-like type of HCC. However, in our series of HCC, the most severe prognosis was related to G3 tumors that did not overexpress fetal genes (Supplementary Fig. 2). Previous reports also identified the 2 main transcriptomic group classifications as a relevant survival predictor.²² We found better survival for subgroups G4-G6 than subgroups G1-G3; however, this difference did not reach statistical significance. These discrepancies may be related to the different distribution of risk factors in each tumor series. Furthermore, unsupervised transcriptomic classification does not appear to be a robust survival determinant, emphasizing the need to carefully consider the wide spectrum of epidemiological risk factors and genetic diversity of HCC when attempting to define "universal" HCC survival predictors.

In conclusion, our global transcriptomic analysis has established a robust classification reflecting the natural diversity of human HCC and the structural gene alterations and epigenetic de-regulations accumulated during tumor progression. The high diversity of HCC tumors has clinical implications, and our classification will help to further identify patients who will benefit from targeted therapies.

Acknowledgment: We warmly thank Jacqueline Godet, François Sigaux, and Philippe Dessen, managers of the Carte d'Identité des Tumeurs (CIT) program, founded by the Ligue Nationale contre le Cancer; Daniela

Geromin, Christelle Thibault, Patricia Legoix, Damien Gerald, and Olivier Bluteau for their experimental help; Fabien Petel for his help in the submission of the data to EBI; and Philippe Bois for critically reading the manuscript. We thank the technicians from the CEPH, Fondation Jean Dausset, for their help in sequencing and all the clinicians who referred patients.

References

1. Bosch FX, Ribes J, Borrás J. Epidemiology of primary liver cancer. *Semin Liver Dis* 1999;19:271-285.
2. Edmondson HA, Peters RL. Tumors of the liver: pathologic features. *Semin Roentgenol* 1983;18:75-83.
3. Thorgeirsson SS, Grisham JW. Molecular pathogenesis of human hepatocellular carcinoma. *Nat Genet* 2002;31:339-346.
4. Brechot C. Pathogenesis of hepatitis B virus-related hepatocellular carcinoma: old and new paradigms. *Gastroenterology* 2004;127:S56-S61.
5. Laurent-Puig P, Zucman-Rossi J. Genetics of hepatocellular tumors. *Oncogene* 2006;25:3778-3786.
6. Bressac B, Galvin KM, Liang TJ, Isselbacher KJ, Wands JR, Ozturk M. Abnormal structure and expression of p53 gene in human hepatocellular carcinoma. *Proc Natl Acad Sci U S A* 1990;87:1973-1977.
7. de La Coste A, Romagnolo B, Billuart P, Renard CA, Buendia MA, Soubrane O, et al. Somatic mutations of the beta-catenin gene are frequent in mouse and human hepatocellular carcinomas. *Proc Natl Acad Sci U S A* 1998;95:8847-8851.
8. Miyoshi Y, Iwao K, Nagasawa Y, Aihara T, Sasaki Y, Imaoka S, et al. Activation of the beta-catenin gene in primary hepatocellular carcinomas by somatic alterations involving exon 3. *Cancer Res* 1998;58:2524-2527.
9. Satoh S, Daigo Y, Furukawa Y, Kato T, Miwa N, Nishiwaki T, et al. AXIN1 mutations in hepatocellular carcinomas, and growth suppression in cancer cells by virus-mediated transfer of AXIN1. *Nat Genet* 2000;24:245-250.
10. Roncalli M, Bianchi P, Bruni B, Laghi L, Destro A, Di Gioia S, et al. Methylation framework of cell cycle gene inhibitors in cirrhosis and associated hepatocellular carcinoma. *HEPATOLOGY* 2002;36:427-432.
11. Azechi H, Nishida N, Fukuda Y, Nishimura T, Minata M, Katsuma H, et al. Disruption of the p16/cyclin D1/retinoblastoma protein pathway in the majority of human hepatocellular carcinomas. *Oncology* 2001;60:346-354.
12. De Souza AT, Hankins GR, Washington MK, Orton TC, Jirtle RL. M6P/IGF2R gene is mutated in human hepatocellular carcinomas with loss of heterozygosity. *Nat Genet* 1995;11:447-449.
13. Su TS, Liu WY, Han SH, Jansen M, Yang-Fen TL, P'Eng FK, et al. Transcripts of the insulin-like growth factors I and II in human hepatoma. *Cancer Res* 1989;49:1773-1777.
14. Laurent-Puig P, Legoix P, Bluteau O, Belghiti J, Franco D, Binot F, et al. Genetic alterations associated with hepatocellular carcinomas define distinct pathways of hepatocarcinogenesis. *Gastroenterology* 2001;120:1763-1773.
15. Lee S, Lee HJ, Kim JH, Lee HS, Jang JJ, Kang GH. Aberrant CpG island hypermethylation along multistep hepatocarcinogenesis. *Am J Pathol* 2003;163:1371-1378.
16. Zochbauer-Muller S, Fong KM, Virmani AK, Geradts J, Gazdar AF, Minna JD. Aberrant promoter methylation of multiple genes in non-small cell lung cancers. *Cancer Res* 2001;61:249-255.
17. Zucman-Rossi J, Benhamouche S, Godard C, Boyault S, Grimber G, Balabaud C, et al. Differential effects of inactivated Axin1 and activated β -catenin mutations in human hepatocellular carcinomas. *Oncogene* 2006, September 11 [Epub ahead of print].
18. Cadoret A, Ovejero C, Terris B, Souil E, Levy L, Lamers WH, et al. New targets of beta-catenin signaling in the liver are involved in the glutamine metabolism. *Oncogene* 2002;21:8293-8301.
19. Yamamoto Y, Sakamoto M, Fujii G, Tsuiji H, Kenetaka K, Asaka M, et al. Overexpression of orphan G-protein-coupled receptor, Gpr49, in human hepatocellular carcinomas with beta-catenin mutations. *HEPATOLOGY* 2003;37:528-533.

20. Cavard C, Terris B, Grimber G, Christa L, Audard V, Radenen-Bussiere B, et al. Overexpression of regenerating islet-derived 1 alpha and 3 alpha genes in human primary liver tumors with beta-catenin mutations. *Oncogene* 2006;25:599-608.
21. Bosch FX, Ribes J, Diaz M, Cleries R. Primary liver cancer: worldwide incidence and trends. *Gastroenterology* 2004;127:S5-S16.
22. Lee JS, Chu IS, Heo J, Calvisi DF, Sun Z, Roskams T, et al. Classification and prediction of survival in hepatocellular carcinoma by gene expression profiling. *HEPATOLOGY* 2004;40:667-676.
23. Breuhahn K, Vreden S, Haddad R, Beckebaum S, Stippel D, Flemming P, et al. Molecular profiling of human hepatocellular carcinoma defines mutually exclusive interferon regulation and insulin-like growth factor II overexpression. *Cancer Res* 2004;64:6058-6064.
24. Chen X, Cheung ST, So S, Fan ST, Barry C, Higgins J, et al. Gene expression patterns in human liver cancers. *Mol Biol Cell* 2002;13:1929-1939.
25. Davis LG, Weber DJ, Lemon SM. Horizontal transmission of hepatitis B virus. *Lancet* 1989;1:889-893.
26. Wong VC, Lee AK, Ip HM. Transmission of hepatitis B antigens from symptom free carrier mothers to the fetus and the infant. *Br J Obstet Gynaecol* 1980;87:958-965.
27. Yu MC, Yuan JM. Environmental factors and risk for hepatocellular carcinoma. *Gastroenterology* 2004;127:S72-S78.
28. Behrens J, Mareel MM, Van Roy FM, Birchmeier W. Dissecting tumor cell invasion: epithelial cells acquire invasive properties after the loss of uvomorulin-mediated cell-cell adhesion. *J Cell Biol* 1989;108:2435-2447.
29. Bluteau O, Jeannot E, Bioulac-Sage P, Marques JM, Blanc JF, Bui H, et al. Bi-allelic inactivation of TCF1 in hepatic adenomas. *Nat Genet* 2002;32:312-315.
30. Lee JW, Soung YH, Kim SY, Lee HW, Park WS, Nam SW, et al. PIK3CA gene is frequently mutated in breast carcinomas and hepatocellular carcinomas. *Oncogene* 2005;24:1477-1480.
31. Lee JS, Heo J, Libbrecht L, Chu IS, Kaposi-Novak P, Calvisi DF, et al. A novel prognostic subtype of human hepatocellular carcinoma derived from hepatic progenitor cells. *Nat Med* 2006;12:410-416.