Focal Gains of VEGFA and Molecular Classification of Hepatocellular Carcinoma

Derek Y. Chiang,1,4 Augusto Villanueva,6 Yujin Hoshida,2,4 Judit Peix,6 Philippa Newell, 6 Beatriz Minguez,3 Amanda C. LeBlanc,7,8 Diana J. Donovan,6 Swan N. Thung,10 Manel Solé,10 Victoria Tovar,6 Clara Alsinaet,10 Alex H. Ramos,14 Jordi Barretina,14 Sasan Roayaie,9 Myron Schwartz,9 Samuel Waxman,7 Jordi Bruix,10 Vincenzo Mazzaferrro,11 Azra H. Ligon,3 Vesna Najfeld,7,8 Scott L. Friedman,6 William R. Sellers,5 Matthew Meyerson, and Josep M. Llovet 10

1Department of Medical Oncology and Center for Cancer Genome Discovery and 2Department of Pediatric Oncology, Dana-Farber Cancer Institute, Department of Pathology, Brigham and Women's Hospital, Boston, Massachusetts; 3Cancer Program, Broad Institute of Harvard and MIT; 4Novartis Institutes for Biomedical Research, Cambridge, Massachusetts; 5Mount Sinai Liver Cancer Program, Division of Liver Diseases, Division of Hematology/Oncology, Department of Medicine, 6Department of Pathology, and Division of Surgical Oncology, Department of Surgery, Mount Sinai School of Medicine, New York, New York; 7Liver Unit, BCLC Group, IICC Translational Research Lab and Department of Pathology, IDIBAPS, CIBERehd, Hospital Clinic, Barcelona, Spain; and 8Gastrointestinal Surgery and Liver Transplantation Unit, National Cancer Institute, Milan, Italy

Abstract

Hepatocellular carcinomas represent the third leading cause of cancer-related deaths worldwide. The vast majority of cases arise in the context of chronic liver injury due to hepatitis B virus or hepatitis C virus infection. To identify genetic mechanisms of hepatocarcinogenesis, we characterized copy number alterations and gene expression profiles from the same set of tumors associated with hepatitis C virus. Most tumors harbored 1q gain, 8q gain, or 8p loss, with occasional alterations in 13 additional chromosome arms. In addition to amplifications at 11q13 in 6 of 103 tumors, 4 tumors harbored focal gains at 6p21 incorporating vascular endothelial growth factor A (VEGFA). Fluorescence in situ hybridization on an independent validation set of 210 tumors found 6p21 high-level gains in 14 tumors, as well as 2 tumors with 6p21 amplifications. Strikingly, this locus overlapped with copy gains in 4 of 371 lung adenocarcinomas. Overexpression of VEGFA via 6p21 gain in hepatocellular carcinomas suggested a novel, non–cell-autonomous mechanism of oncogene activation. Hierarchical clustering of gene expression among 91 of these tumors identified five classes, including “CTNNB1”, “proliferation”, “IFN-related”, a novel class defined by polysomy of chromosome 7, and an unannotated class. These class labels were further supported by molecular data; mutations in CTNNB1 were enriched in the “CTNNB1” class, whereas insulin-like growth factor 1 receptor and RPS6 phosphorylation were enriched in the “proliferation” class. The enrichment of signaling pathway alterations in gene expression classes provides insights on hepatocellular carcinoma pathogenesis. Furthermore, the prevalence of VEGFA high-level amplifications in multiple tumor types suggests indications for clinical trials of antiangiogenic therapies. [Cancer Res 2008;68(16):6779–88]

Introduction

Hepatocellular carcinoma is the third leading cause of cancer deaths worldwide. These tumors arise in the context of chronic liver injury, inflammation, and hepatocyte proliferation provoked by several etiologies. Hepatitis C virus is the most prevalent etiology in the United States and Europe, whereas hepatitis B virus is endemic in Asia and Africa (1, 2). The critical genomic hits at preneoplastic stages are not well defined, although aberrant methylation, genomic instability, and TERT activation have been reported. In overt hepatocellular carcinoma, deregulated oncogenes (MET, CTNNB1, MYC, and CCND1) and tumor suppressor genes (TP53, PTEN, CDKN2A, and CDH1) due to genetic or epigenetic alterations have been characterized.

Gene expression studies have converged on major molecular classes of hepatocellular carcinoma by analyzing patients at different stages and with different etiologic factors (3–8). One class is associated with high proliferation, chromosomal instability, and activation of insulin-like growth factor (IGF) or Akt/mammalian target of rapamycin signaling. Another class features activation of the Wnt signaling pathway associated with a high prevalence of mutations in CTNNB1, which is the gene with the second highest frequency of known mutations in hepatocellular carcinomas, after TP53 (7, 9). Mutations or deletions in CTNNB1 exon 3 prevent the ubiquitination of its protein product, β-catenin, leading to its nuclear translocation and trans-activation of target genes (10). A third class is associated with IFN signaling due to leukocyte infiltration (4). Intriguingly, an interleukin gene expression signature from the adjacent liver tissue has also been linked to higher rates of venous invasion and metastasis, pointing to the importance of tumor microenvironment in the genesis and intrahepatic dissemination of hepatocellular carcinoma (11).

Despite all these efforts, only two studies have been able to correlate homogeneous biological tumor patterns with survival outcomes (5, 8), although confirmatory-independent validations are not yet available. In parallel, genome-wide chromosomal aberration studies (12–16) and methylation studies (17) have begun to decipher the molecular heterogeneity of hepatocellular carcinoma.

The increasing worldwide incidence of hepatocellular carcinomas due to chronic hepatitis C infection underscores the need to...
discover biomarkers for early detection and tumor progression, as well as gene targets for primary and adjuvant therapies. We conducted extensive genomic and transcriptomic characterization of tumors from hepatitis C virus–positive patients with early disease, as defined by a clinical staging system. We show that high-level copy number gain of 6p21 represents one mechanism for increasing vascular endothelial growth factor A (VEGFA) expression. In addition, we define five molecular classes that integrate copy number and gene expression alterations with signaling pathway assessment and describe a new molecular class defined by increased dosage of chromosome 7 and overexpression of oncogenes on this region.

Materials and Methods

Primary hepatocellular carcinoma specimens. Fresh frozen tissue specimens measuring 1 cm³ were obtained at the time of surgical resection or liver transplantation from 100 anonymous patients from the HCC Genomic Consortium (Barcelona Hospital Clinic, Milan National Cancer Institute, and Mount Sinai School of Medicine) according to institutional review board–approved protocols. Two nodules were obtained from each of three transplantation patients, for a total of 103 tumor samples. For 94 of these samples, uninvolved cirrhotic tissue was obtained from the same patient at time of surgery. Clinicopathologic annotated data were analyzed in 82 samples obtained from surgical resection (Supplementary Table S1). Hepatitis C virus infection was confirmed by serology in all cases, and other etiologies of hepatocellular carcinoma were ruled out. Tissue samples were treated with RNAlater (Qiagen) or liquid nitrogen and stored at −80 °C until DNA and RNA extraction, as previously reported (18). Genomic DNA was extracted with the ChargeSwitch gDNA mini kit (Invitrogen) and further purified with phenol-chloroform extraction. An independent set of 210 samples and 8 cirrhotic liver controls already reported (18) were preprocessed with GISTIC software, which includes steps for batch correction, data normalization, copy number estimation versus the average of the five closest normal samples, and copy number segmentation (refer to supplement of ref. 21). Copy number alterations at each probe set i are evaluated by the score $G_i = \frac{1}{\sigma_i}$, where $a_i$ represents the average amplitude of gain or loss with an overall frequency $f_i$ in the data set. We used copy number cutoffs of 2.2 to detect broad regions of copy gain and 1.8 for loss because these cutoffs exceeded the variation observed in adjacent cirrhotic tissues obtained from the same patients. We obtained similar results by inferring copy numbers with the GEMCA algorithm (Supplementary Fig. S5; ref. 22). To detect high-level copy number gains of VEGFA by fluorescence in situ hybridization (FISH) of tissue microarrays (Supplementary Table S1).

Single-nucleotide polymorphism arrays. Genotypes and hybridization intensities for more than 238,000 single-nucleotide polymorphisms (SNP) were measured with the Sty1 chip of the Affymetrix 500K Human Mapping Array set (Affymetrix). Data were analyzed with the GenePattern software package (19) according to methods previously described (20). Intensity (CEL) files were normalized with invariant-set normalization and modeled using the PM-MM difference modeling method with the SNPFileCreator module (19). Copy numbers were inferred by comparing signal intensities between each tumor and a selection of the five most similar normals selected from nontumoral cirrhotic tissue, as well as a database of multiple tumor types (21). The similarity between tumor and normal profiles was measured as the Euclidean distance between the log 2–transformed signal profiles (21).

Statistical analysis of copy number alterations. Copy number data were preprocessed with GISTIC software, which includes steps for batch correction, data normalization, copy number estimation versus the average of the five closest normal samples, and copy number segmentation (refer to supplement of ref. 21). Copy number alterations at each probe set i are evaluated by the score $G_i = \frac{1}{\sigma_i}$, where $a_i$ represents the average amplitude of gain or loss with an overall frequency $f_i$ in the data set. We used copy number cutoffs of 2.2 to detect broad regions of copy gain and 1.8 for loss because these cutoffs exceeded the variation observed in adjacent cirrhotic tissues obtained from the same patients. We obtained similar results by inferring copy numbers with the GEMCA algorithm (Supplementary Fig. S5; ref. 22). To detect high-level gains of VEGFA, we used a 1−Pearson correlation distance metric after mean-centering of genes (25).

Selection of class-specific marker genes. For each of the five gene expression classes obtained from consensus hierarchical clustering, we performed supervised analysis to identify genes that were significantly up-regulated or down-regulated among tumors. Significance Analysis of Microarrays with 500 permutations of class labels evaluated the significance of gene expression changes that were significantly overexpressed or underexpressed among the tumors in a single class. Up to 200 overexpressed transcripts and up to 200 underexpressed transcripts with a ≥2-fold change and a false discovery rate (FDR) $q < 0.01$ were selected as marker genes. The list of marker genes for each class is provided in Supplementary Tables S6 to S10.

Cross-dataset prediction of gene expression classes. Affymetrix U133A array data were obtained from EBI ArrayExpress, with accession no. E-TABM-36 (7). These U133A array probe sets were mapped to 12,145 AceView transcripts based on exact sequence matches (25). NCI-Operon oligonucleotide microarray data were obtained from Gene Expression Omnibus, with accession nos. GSE1988 and GSE4024 (5, 6). The Prediction Analysis of Microarrays R package was used to train shrunken centroid classifiers from previously published gene expression data and class labels (see Supplementary Methods; ref. 22).

Quantitative real-time PCR. Total RNA was extracted from three to four tissue sections, each 10 μm thick, using the TRIzol LS reagent (Invitrogen). cDNA synthesis and PCR conditions were conducted as previously described (23). Expression levels were measured with TaqMan probes Hs00173626-m1 for VEGFA and Hs03023943-g1 for ACTB obtained from TaqMan Gene Expression Assays (Applied Biosystems). Reactions were set up as triplicates for each gene, and the median threshold cycle ($C_i$) value from the triplicates was used. We calculated Δ$C_i$ values between VEGFA and ACTB for each tumor, and the median threshold cycle ($C_i$) for five uninvolved, noncancerous cirrhotic controls. The Mann-Whitney test was used to evaluate differences in expression levels between groups.

Affymetrix U133 Plus 2.0 expression data. Gene expression assessment was done in 91 tumor samples: 69 new samples obtained from surgical resection extracted as described previously and additional 22 tumor samples and 8 cirrhotic liver controls already reported (18). Affymetrix array intensity (CEL) files were processed with the RMA algorithm (24). To compare expression levels across microarray platforms, we remapped probe sequences on each platform to AceView transcripts (25, 26). Array intensity data (CEL files) have been deposited in the National Center for Biotechnology Information Gene Expression Omnibus (GEO) and are accessible through Gene Expression Omnibus Series accession no. GSE9829.

Unsupervised classification of gene expression data. We adopted a robust hierarchical clustering approach for unsupervised classification of Affymetrix U133Plus 2.0 data (7). We averaged the results from modifying the three clustering parameters: transcript lists, linkage method, and number of clusters (see Supplementary Methods). The reproducibility of hierarchical clustering was assessed with the GenePattern ConsensusClustering module, version 3, using a 1−Pearson correlation distance metric after mean-centering of genes (27).

Genotypes and hybridization intensities for more than 238,000 single-nucleotide polymorphisms (SNP) were measured with the Sty1 chip of the Affymetrix 500K Human Mapping Array set (Affymetrix). Data were analyzed with the GenePattern software package (19) according to methods previously described (20). Intensity (CEL) files were normalized with invariant-set normalization and modeled using the PM-MM difference modeling method with the SNPFileCreator module (19). Copy numbers were inferred by comparing signal intensities between each tumor and a selection of the five most similar normals selected from nontumoral cirrhotic tissue, as well as a database of multiple tumor types (21). The similarity between tumor and normal profiles was measured as the Euclidean distance between the log 2–transformed signal profiles (21).

Statistical analysis of copy number alterations. Copy number data were preprocessed with GISTIC software, which includes steps for batch correction, data normalization, copy number estimation versus the average of the five closest normal samples, and copy number segmentation (refer to supplement of ref. 21). Copy number alterations at each probe set i are evaluated by the score $G_i = \frac{1}{\sigma_i}$, where $a_i$ represents the average amplitude of gain or loss with an overall frequency $f_i$ in the data set. We used copy number cutoffs of 2.2 to detect broad regions of copy gain and 1.8 for loss because these cutoffs exceeded the variation observed in adjacent cirrhotic tissues obtained from the same patients. We obtained similar results by inferring copy numbers with the GEMCA algorithm (Supplementary Fig. S5; ref. 22). To detect high-level gains of VEGFA, we used a 1−Pearson correlation distance metric after mean-centering of genes (25).

Selection of class-specific marker genes. For each of the five gene expression classes obtained from consensus hierarchical clustering, we performed supervised analysis to identify genes that were significantly up-regulated or down-regulated among tumors. Significance Analysis of Microarrays with 500 permutations of class labels evaluated the significance of gene expression changes that were significantly overexpressed or underexpressed among the tumors in a single class. Up to 200 overexpressed transcripts and up to 200 underexpressed transcripts with a ≥2-fold change and a false discovery rate (FDR) $q < 0.01$ were selected as marker genes. The list of marker genes for each class is provided in Supplementary Tables S6 to S10.

Cross-dataset prediction of gene expression classes. Affymetrix U133A array data were obtained from EBI ArrayExpress, with accession no. E-TABM-36 (7). These U133A array probe sets were mapped to 12,145 AceView transcripts based on exact sequence matches (25). NCI-Operon oligonucleotide microarray data were obtained from Gene Expression Omnibus, with accession nos. GSE1988 and GSE4024 (5, 6). The Prediction Analysis of Microarrays R package was used to train shrunken centroid classifiers from previously published gene expression data and class labels (see Supplementary Methods; ref. 22).
were incubated with anti–phosphorylated (Ser 473) Akt antibody, quenched with 3% hydrogen peroxide. After washing with PBS, samples were rehydrated in a graded series of ethanol solutions. Antigens were unmasked by microwave heating the samples in 10 mmol/L sodium citrate buffer (pH 6.0) for 5 min (×3). Endogenous peroxidase from the tissue was quenched with 3% hydrogen peroxide. After washing with PBS, samples were incubated with anti–phosphorylated (Ser240/244) RPS6 (1:200; Cell Signaling Technology) at pH 6.0 for 5 min (×3). Endogenous peroxidase from the tissue was quenched with 3% hydrogen peroxide. After washing with PBS, samples were incubated with anti–phosphorylated (Ser134) Grb2 antibody (1:50; immunohistochemistry specific, Cell Signaling Technology), anti–phosphorylated (Tyr1316) IGF-I receptor (IGF-IR; 1:100; Dr. Rubini, University of Ferrara, Ferrara, Italy), anti-CTNNB1 (1:750; Abcam), or anti–phosphorylated (Ser172) Akt antibody (1:50; immunohistochemistry specific, Cell Signaling Technology), anti–phosphorylated (Tyr118) IGF-I receptor (IGF-IR; 1:100; Dr. Rubini, University of Ferrara, Ferrara, Italy), anti-CTNNB1 (1:750; Abcam), or anti–phosphorylated (Ser240/244) RPS6 (1:200; Cell Signaling Technology) at 4°C overnight. 3,3'-Diaminobenzidine was used as a detection system (EnVision+ System-HRP, DAKO). Immunoreactivity was independently graded by two liver pathologists (S.T. and M.S.) and both agreed in the final staining score. The variables measured were as follows: (a) intensity (0, absent; 1, weak; 2, moderate; 3, strong); (b) distribution (1, very focal; 2, focal; 3, diffuse); and (c) localization of the staining (membranous, cytoplasmic, or nuclear). Samples were defined positive for Akt or IGF-IR phosphorylation when intensity was ≥2, regardless of distribution. For RPS6 phosphorylation, only those samples with intensity and distribution of ≥2 were considered positive. Tumors with CTNNB1 activation were defined by the presence of nuclear staining or by >5% of cytoplasmic staining.

Statistical associations. Each of the five gene expression classes was assessed for enrichment with molecular characteristics. Fisher’s exact test was used to assess the enrichment of binary variables such as immunostaining, mutation status, and chromosomal gain or loss. The Mann-Whitney was used to test for associations with continuous variables, such as α-fetoprotein (AFP) levels.

Analysis of clinical outcomes. The probability curves of recurrence and early recurrence were calculated according to Kaplan-Meier and compared by Mantel-Cox test. Median inferred copy numbers for chromosomal arms for each sample were used as covariates for univariate association with recurrence. Median copy numbers of chromosomal arms were calculated for each tumor sample, and cutoffs of 2.2 for copy gains and 1.8 for copy loss were used. For continuous variables, the cutoff level was their median value. Significant variables (P < 0.05) were included in a stepwise Cox proportional hazard regression analysis of recurrence and early recurrence. The calculations were done with the SPSS package (SPSS 15.0).

Results

Copy number alterations in clinically early, HCV-related hepatocellular carcinomas. We measured copy number alterations at over 238,000 genomic loci with Affymetrix Sty mapping arrays among 103 primary tumors. According to the Barcelona Clinic Liver Cancer staging system, 80% of these patients had very early or early stage disease (stage 0 or stage A), representing asymptomatic tumors with either a single nodule or up to three nodules <3 cm in diameter (Supplementary Table S1; ref. 29). To find evidence for driver alterations in tumor genomes, we evaluated the frequency and magnitude of copy number gains and losses with the GISTIC algorithm (21). Broad gains and losses of chromosome arms were the most prevalent changes among these tumors (Fig. 1A). The most significant chromosomal gains affected...
Figure 2. Overlapping high-level gains at 6p21 in hepatocellular carcinomas and lung adenocarcinomas. A, copy number alterations among four hepatocellular carcinomas (light blue), four lung adenocarcinomas (light green), four matched cirrhotic tissues from adjacent liver (dark blue), and three matched normals for the lung tumors (dark green). B, inferred copy numbers at individual SNP probe sets. Copy numbers are shown in red for tumor samples or in gray for the matched normal from the same patient. Vertical dashed lines delineate the boundaries of the minimal overlapping region among the eight tumors. C, annotated genes found in the minimal overlapping region at 6p21. Inclusion of the SM-11XV lung tumor narrows the overlapping region to the boundaries of the shaded box. The genomic location of BAC probe RP11-710L16 used for FISH is indicated. D, confirmation of VEGFA copy gains by FISH. Red signals, BAC probe RP-710L16 centered on VEGFA; green signals, BAC probe CEP-6 for the centromere of chromosome 6. Representative images are displayed for the three of four hepatocellular carcinomas with available tissue blocks, and detailed counts of probe signals are provided in Supplementary Table S4.
chromosome arms 1q (72%) and 8q (66%), which are early events in hepatocarcinogenesis (Fig. 1B; ref. 30). Focal gains centered on 6p21 (4%) and 11q13 (6%) were the next most significant alterations due to their high levels of copy increase. Significant copy number gains were also observed for chromosomes 5p (33%), 7 (30%), 17q (19%), and 20 (18%). Notably, gains of chromosomes 5p and 7 significantly co-occurred among these tumors ($P < 10^{-4}$, Fisher’s exact), which may indicate cooperation among multiple genes with elevated dosage on both chromosomes. Significant chromosomal losses affected chromosome arms 8p (60%), 17p (39%), 6q (27%), 4q (19%), 13q (19%), 16q (13%), and 10q (11%; Fig. 1C). Aside from 1q gain, 8q gain, and 8p loss in the majority of tumors, substantially lower frequencies of gains and losses on other chromosome arms suggested the presence of multiple tumor subtypes that select for alterations in distinct signaling pathways. We did not detect chromosomal gains or losses in the 94 uninvolved, noncancerous cirrhotic tissue samples (Supplementary Fig. S4D).

Focal regions of copy gain or loss provide greater resolution to map oncogenes or tumor suppressors by narrowing the overlapping genomic region among multiple tumors. However, these focal lesions were rare events among clinically early hepatocellular carcinomas. Only 15% of tumors harbored at least one high-level gain, which we define as a genomic region smaller than 12 Mb with inferred copy number >3.8 (Supplementary Table S2). We found 9 candidate regions of homozygous deletion, all of which

<table>
<thead>
<tr>
<th>Copy number alteration</th>
<th>n (%)</th>
<th>Mean VEGFA probe signals</th>
<th>Mean CEP6 probe signals</th>
<th>VEGFA/CEP6 ratios</th>
</tr>
</thead>
<tbody>
<tr>
<td>VEGFA amplifications</td>
<td>2 (1.0)</td>
<td>6.5–14.9</td>
<td>2.1–2.1</td>
<td>4.1–7.2</td>
</tr>
<tr>
<td>VEGFA high-level gains</td>
<td>14 (6.7)</td>
<td>4.0–6.5</td>
<td>2.1–5.9</td>
<td>1.1–1.8</td>
</tr>
<tr>
<td>Chromosome 6 gains</td>
<td>58 (27.6)</td>
<td>2.3–3.9</td>
<td>2.0–3.7</td>
<td>1.0–1.8</td>
</tr>
<tr>
<td>Normal chromosome 6</td>
<td>136 (64.8)</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

| Table 1. FISH analyses of 6p21 copy number alterations in 210 hepatocellular carcinomas on tissue microarrays |

Figure 3. High-level gains at 6p21 lead to VEGFA overexpression. A, Significance Analysis of Microarrays identified overexpression of two genes among the four tumors with 6p21 amplifications. Each point represents the observed $t$ statistic, plotted against its expected score from permutation tests. The most significant gene is VEGFA, followed by TMEM63B (purple). Dashed lines, confidence interval for the random, null distribution. VEGFA (B) or TMEM63B (C) is overexpressed in tumors with high-level gains of 6p21. Each of the 91 points represents the corresponding copy number and expression level for a single tumor. Horizontal axis, median inferred copy number for 68 SNP probes in the minimal overlapping region at 6p21. Vertical axis, log 2 ratio between the expression level of each tumor, compared with the median of 10 normal liver samples on the Affymetrix U133 Plus 2 array. D, higher VEGFA expression is associated with increased copy numbers in an independent cohort. Each of the 45 points represents a single tumor chosen from the panel of 210 tumors evaluated by FISH. Horizontal axis, average number of VEGFA probes per nucleus. Vertical axis, negative $\Delta \Delta C_t$ values of VEGFA normalized to ACTB, compared with the median of five uninvolved, cirrhotic tissue samples. Red, tumors with $>4$ copies of VEGFA ($n = 10$); black, tumors with gains of chromosome 6 ($n = 19$); gray, diploid tumors ($n = 16$). The average negative $\Delta \Delta C_t$ value for each of these three classes is plotted with a dashed blue line.
Figure 4. Integrated classification of genomic and molecular alterations in hepatocellular carcinomas. A, consensus hierarchical clustering of gene expression in 91 tumors. Colors in the sample dendrogram indicate the five gene expression classes determined by consensus hierarchical clustering among 32 different parameter set combinations. Rows display mean-centered expression levels for 1,242 marker genes associated with each of the five gene expression classes.

B, overlap of consensus classes with previously published classes. Class labels were assigned based on the distance to the closest shrunken centroid classifier for previous classifications (5, 7). C, copy number alterations among the gene expression classes. Gains and losses for the indicated chromosomes are displayed according to the color scheme in Fig. 2. Mutation status (CTNNB1 exon 3 or TP53 exons 5–8) and immunohistochemical staining are indicated in grayscale: present (black), absent (gray), or missing (white) data.
occurred in only one tumor (Supplementary Table S3). Two of these regions affected single exons of CSMD1 or MSR1, whereas two regions deleted exons of GPR143, SHROOM2, WWC3, MAOA, and MAOB.

To pinpoint genomic regions that may harbor key oncogenes, we repeated GISTIC analysis to find significantly recurrent high-level gains. The most frequent amplifications occurred in six tumors at 11q13, with a minimal common region from 68.83 to 69.30 Mb (FDR $q < 10^{-11}$; Supplementary Fig. S1). This region included only four known genes: CCND1, ORAOV1, FGF19, and FGF4. FISH with a BAC probe to these four genes confirmed that all three of these tumors with available tissue blocks harbored amplifications. In 50% to 90% of tumor nuclei, more than 10 signals corresponding to BAC RP11-710L16 were detected (Supplementary Fig. S1).

Overlapping high-level gains at 6p21 occur in multiple tumor types. The second most frequent high-level gains pinpointed an overlapping region that had not previously been described in hepatocellular carcinomas (FDR $q < 10^{-11}$). Four tumors harbored high-level gains of 6p21, ranging in size from 2.4 to 22 Mb (Fig. 2A and B). Remarkably, a recent study of 371 lung adenocarcinomas identified this locus as the 17th most frequent region of copy gains, yet this frequency barely passed statistical significance (FDR $q < 0.08$; ref. 20). Assuming that the same oncogene targets are under selection in multiple tumor types (31), we pooled regions of high-level gains from four hepatocellular carcinomas and four lung adenocarcinomas. The minimal region of overlap, between positions 43.792 and 44.110 Mb, encompassed only two genes, VEGFA and MGCG45491 (Fig. 2C).

High-level gains of 6p21 were confirmed in two FISH studies with a BAC probe spanning the MRPS18A and VEGFA genes (Fig. 2C). First, FISH analysis on thin sections from all three hepatocellular carcinomas with available tissue blocks validated copy gains at 6p21, with an average of >5.5 VEGFA probe signals per tumor (Fig. 2D; Supplementary Fig. S2; Supplementary Table S4). Second, we also performed FISH analysis on an independent tissue microarray collection of 210 hepatocellular carcinomas (Table 1; Supplementary Table S1; Supplementary Table S4). We defined three cytogenetic categories based on the average number of VEGFA probe signals observed per nucleus. Gains of chromosome 6 were found in 58 of 210 (27.6%) tumors, with ≥2.3 probe signals. An additional 14 (6.7%) tumors also had high-level gains of VEGFA, as defined by ≥4 VEGFA probe signals. Two tumors showed VEGFA amplifications (1.0%), with ≥8 VEGFA probe signals. VEGFA is a likely oncogene target of high-level gains at 6p21. Genomic regions with increased copy number are often retained in tumor cells due to selective advantages conferred by the increased expression of one or more target genes residing in the region. To identify candidate oncogene targets in the overlapping regions of copy gains at 6p21 and 11q13, we evaluated the association between copy numbers at these loci and the expression of each gene. A genome-wide, permutation-based $t$ test identified only two known genes that were significantly overexpressed among the four tumors with high-level gains at 6p21: VEGFA and TMEM63B (FDR $q < 0.001$; Fig. 3A). VEGFA was the most significantly overexpressed transcript, with an average 2.8-fold higher expression among the four tumors with 6p21 high-level gains (Fig. 3B). TMEM63B encodes a transmembrane protein with unknown function, and the four tumors with 6p21 high-level gains had 1.9-fold higher average expression (Fig. 3C). Thus, elevated levels of TMEM63B may represent a marker indicating VEGFA overexpression by increased copy dosage, rather than VEGFA overexpression due to trans-regulation by hypoxia-inducible factor 1A or other regulators.

Notably, VEGFA expression was also associated with copy numbers measured by FISH among 45 tumors in the tissue microarray cohort. Ten tumors with >4 copies of VEGFA had an average of 2.8-fold higher expression than 16 tumors that were diploid at this locus ($P = 0.0007$, Mann-Whitney; Fig. 3D). Thus, we have validated that high-level gains of VEGFA are associated with higher VEGFA expression in two independent cohorts.

For the six tumors with 11q13 amplifications, we found four genes to be significantly overexpressed: ORAOV1, TPCN2, CCND1, and MRPL21 (FDR $q < 0.001$). Both ORAOV1 and CCND1 reside in the minimal common region of amplification. CCND1 is a confirmed oncogene in hepatocarcinogenesis (32), whereas the log expression levels of ORAOV1 showed a strikingly linear correlation with the median copy number of 11q13 (Supplementary Fig. S1D).

Gene expression classes are associated with signaling pathway alterations. To characterize the molecular heterogeneity of hepatocellular carcinomas, we measured gene expression profiles in 91 of the 103 tumors with oligonucleotide microarrays (Affymetrix U133 Plus 2 arrays). We obtained five gene expression classes from unsupervised classification with consensus hierarchical clustering, which considered 32 different parameter combinations (Fig. 4A; ref. 7). To validate these five gene expression classes with previously published studies, we trained shrunken centroid classifiers on two additional gene expression data sets (6, 7). Two classes showed extremely good concordance with the labels predicted by these external classifiers (Fig. 4B). Furthermore, signaling pathway annotations could be assigned for three of these five classes by the enrichment of marker genes, immunohistochemistry, and mutations (Fig. 4C; Supplementary Tables S5–S10).

The high prevalence of CTNNB1 mutations indicated alterations to canonical Wnt signaling in a subset of hepatocellular carcinomas. The CTNNB1-associated class of 24 tumors was signifi-

### Table 2. Variables with independent predictive value for clinical outcomes in multivariate analysis

<table>
<thead>
<tr>
<th></th>
<th>Early recurrence</th>
<th>Overall recurrence</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Hazard ratio (95% CI)</td>
<td>$P$</td>
</tr>
<tr>
<td>Barcelona-Clinic Liver Cancer staging classification</td>
<td>8.8 (4.3–17.6)</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>Chromosome 7 median copy number ≥2.2</td>
<td>5.8 (2.4–13.8)</td>
<td>&lt;0.001</td>
</tr>
</tbody>
</table>

Abbreviation: 95% CI, 95% confidence interval.
canty enriched for CTNNB1 exon 3 mutations (62%; $P \leq 6 \times 10^{-4}$, Fisher's exact) and nuclear localization (71%; $P < 3 \times 10^{-3}$, Fisher's exact; Supplementary Table S5). Further evidence for CTNNB1 activation among tumors in this class came from the significant overexpression of several liver-specific marker genes including GLUL, LGRB, TXB3, and REG3A (Supplementary Table S6; refs. 33–35). This class was also associated with tumor diameters $>3$ cm ($P = 0.002$, Fisher's exact).

Another molecular subtype of hepatocellular carcinomas featured increased proliferation, high levels of serum AFP, and chromosomal instability (5, 7). We found a corresponding proliferation class of 23 tumors that overexpressed AFP (median serum level, 472 ng/mL; $P = 0.001$, Mann-Whitney), along with several genes corresponding to a proliferation gene expression signature (Supplementary Tables S5 and S7; ref. 36). Because these tumors were enriched for IGF-IR phosphorylation (48%; $P < 1 \times 10^{-3}$, Fisher's exact), RPS6 phosphorylation (83%; $P = 0.003$, Fisher's exact), and Akt phosphorylation (55%; $P = 0.01$, Fisher's exact), it is likely that tyrosine kinase activation drives the proliferation of these tumors. Reduced frequencies of CTNNB1 exon 3 mutations among these tumors (13%; $P = 0.02$, Fisher's exact) confirmed the model that tyrosine kinase activation and CTNNB1 activation represent distinct routes for tumor progression (7, 37). Tumors in this class also had higher frequencies of 4q loss (38%; $P = 0.03$, Fisher's exact) and 13q loss (33%; $P = 0.04$, Fisher's exact), as well as lower frequencies of 6q loss (12%; $P = 0.04$, Fisher's exact). A significant correlation with macrovascular invasion was observed for tumors in this proliferation class ($P = 0.007$, Fisher's exact).

A third class of 18 tumors also harbored significantly lower rates of CTNNB1 exon 3 mutation (6%; $P = 0.004$, Fisher's exact). Lower expression of IGF-II, as well as CTNNB1 target genes, suggested that these tumors represent a distinct class. Strikingly, 4 of the 28 significantly overexpressed marker genes corresponded to IFN-stimulated genes: STAT1, ISG15, IFI6, and IFI27 (Supplementary Table S8; ref. 38). STAT1 is a transcription factor that mediates the response to hepatitis C viral infection, whereas elevated expressions of ISG15, IFI6, and IFI27 are predictive markers of Hepatitis C virus patients who failed to respond to pegylated IFN and ribavirin therapy (39). Notably, tumors about this IFN-related class were more likely to be $<3$ cm in diameter ($P = 0.005$, Fisher’s exact).

**Polysony of chromosome 7 defines a novel gene expression class.** We discovered a new class of 9 tumors that was significantly associated with a lack of gains of chromosome 8q (78%; $P = 0.007$, Fisher's exact), as well as polysony of chromosome 7 above a median copy number of 2.7 (89%; $P < 10^{-4}$, Fisher's exact). Increased dosage led to significantly higher expression levels of multiple genes on chromosome 7 (Supplementary Table S9), as confirmed by gene set enrichment analysis based on cytobands (FDR $q < 0.01$). The most significantly overexpressed genes on chromosome 7 included COBIL, CLDN15, MAD1L1, FOLPD2, and EPHA1; although oncogene targets of these high-level gains remain to be identified. Due to the small number of tumors in this class, we lacked the power to detect significant clinicopathologic correlations.

**Prognostic significance of genomic alterations.** We evaluated associations between copy number alterations and clinical outcomes among 82 patients who underwent surgical resection. Recurrence was associated with chromosome 7 gain and 13q loss by univariate analysis (Supplementary Table S11). On multivariate analysis, chromosome 7 gain was retained as an independent predictor of recurrence [hazard ratio, 3.4 (95% confidence interval, 1.7–6.8); $P < 0.001$], along with BCLC staging [hazard ratio, 7.1 (95% confidence interval, 3.8–13.5); $P < 0.001$]. Notably, gains of chromosome 7 were associated with a significantly higher risk of recurrence within 2 years (early recurrence) after surgical resection (Table 2; Supplementary Fig. S3).

**Discussion**

To date, this study represents the largest characterization effort of paired copy number and gene expression alterations in clinically early hepatocellular carcinomas with hepatitis C virus etiology. Combined annotations from marker gene lists, mutation analyses, and immunohistochemistry could associate three of five gene expression classes with previously described signaling pathway alterations (4, 5, 7). First, a CTNNB1-activated class was enriched for CTNNB1 exon 3 mutations and immunohistochemical staining. Second, a proliferation class was enriched for IGF-IR and RPS6 phosphorylation, thus pointing to potential targets for combined therapeutic strategies. Small-molecule inhibitors and monoclonal antibodies against IGF-IR [such as A12 (Imclone)] and mammalian target of rapamycin [such as rapamycin or everolimus (Novartis)] are still at preclinical or early clinical stage of drug development for hepatocellular carcinoma (40). Third, an IFN-related class that overexpressed several IFN-stimulated genes was associated with smaller tumor size. Notably, tumors with up-regulated IFN response in another study also had higher leukocyte infiltration and tumor apoptosis (4).

A novel gene expression class was defined by polysony of chromosome 7 and the concomitant overexpression of multiple genes along this chromosome. Intriguingly, most of these tumors lacked 8q gains, which are the second most frequent chromosomal alterations in hepatocellular carcinomas and include the known oncogenes MYC, PTK2, and COPIS (12, 41). This observation suggests that unknown target oncogenes on chromosome 7 may contribute to a distinct mechanism of tumor progression. Indeed, we found that even low-level gains of chromosome 7 were independent predictors of recurrence within 2 years of surgical resection. Although EGFR and MET are frequently cited oncogene candidates on chromosome 7, we only observed a modest overexpression of EGFR (1.39-fold versus normals) and MET (1.82-fold versus normals) among the nine tumors in this class. Surprisingly, we did not observe amplifications of MET among these hepatitis C virus–positive tumors, although previous studies have reported a few instances of focal MET amplification, mostly in tumors arising in the context of hepatitis B virus infection (12, 13, 31). Thus, this discrepancy in MET amplification or overexpression could be explained by differences in viral etiologies or accompanying cirrhosis.

Due to the limited number of patients and events in each class, the size of this cohort was underpowered to detect robust differences in overall survival, and only correlation with metastatic potential was captured. Nevertheless, we speculate that these gene expression classes may represent distinct combinations of signaling pathway alterations and would thus show different responses to molecular targeted therapies. Thus, biomarkers for hepatocellular carcinoma molecular classes should be assayed in future clinical trials to provide evidence of treatment response in a genetically defined subset of patients. In addition, animal models that recapitulate these molecular classes should be instrumental for testing novel antitumoral agents in preclinical experiments. For instance, a double transgenic MYC/TKGA mouse model of hepatocellular carcinoma has similar gene expression patterns.
with human tumors in the proliferation class (42). Additional mouse models that recapitulate the remaining classes need to be characterized.

We propose that high-level gains of VEGFA may represent a genetic dependency of tumors arising in multiple tissue types, and that these high-level gains may predict response to a variety of inhibitors against VEGFA or its receptors. Whereas chromosome 6p gains have previously been reported in hepatocellular carcinoma (30, 43), the higher resolution of SNP arrays in this study has localized a minimal common region of copy gains that included VEGFA. We also found that three of the four tumors with focal gains of 6p21 were assigned to an unannotated gene expression class.

Although the prognostic significance of elevated transcript and serum VEGFA levels in early hepatocellular carcinomas has been recognized (44, 45), this study is the first report of focal copy gains as a mechanism for elevated VEGFA expression. Intriguingly, KDR/VEGFR-2 and FLT1/VEGFR-3 are among the putative targets of angiogenesis, as a mechanism for elevated VEGFA expression. Intriguingly, KDR/VEGFR-2 and FLT1/VEGFR-3 are among the putative targets of VEGFA that these high-level gains may predict response to a variety of genetic dependency of tumors arising in multiple tissue types, and that these high-level gains may predict response to a variety of inhibitors against VEGFA or its receptors. Whereas chromosome 6p gains have previously been reported in hepatocellular carcinoma (30, 43), the higher resolution of SNP arrays in this study has localized a minimal common region of copy gains that included VEGFA. We also found that three of the four tumors with focal gains of 6p21 were assigned to an unannotated gene expression class.


41. Okamoto H, Yasui K, Zhao C, Arii S, Inazawa A. PTK2 and EIF3S3 genes may be amplification targets at 8q23–24 and are associated with large hepatocellular carcinomas. Hepatology 2005;41:1242–9.


Focal Gains of VEGFA and Molecular Classification of Hepatocellular Carcinoma

Derek Y. Chiang, Augusto Villanueva, Yujin Hoshida, et al.


Updated version
Access the most recent version of this article at:
http://cancerres.aacrjournals.org/content/68/16/6779

Supplementary Material
Access the most recent supplemental material at:
http://cancerres.aacrjournals.org/content/suppl/2008/08/12/68.16.6779.DC1

Cited articles
This article cites 49 articles, 14 of which you can access for free at:
http://cancerres.aacrjournals.org/content/68/16/6779.full.html#ref-list-1

Citing articles
This article has been cited by 38 HighWire-hosted articles. Access the articles at:
/content/68/16/6779.full.html#related-urls

E-mail alerts
Sign up to receive free email-alerts related to this article or journal.

Reprints and Subscriptions
To order reprints of this article or to subscribe to the journal, contact the AACR Publications Department at pubs@aacr.org.

Permissions
To request permission to re-use all or part of this article, contact the AACR Publications Department at permissions@aacr.org.