

Genes Involved in Viral Carcinogenesis and Tumor Initiation in Hepatitis C Virus–Induced Hepatocellular Carcinoma

Valeria R Mas,^{1,2} Daniel G Maluf,¹ Kellie J Archer,^{3,4} Kenneth Yanek,¹ Xiangrong Kong,³ Laura Kulik,⁵ Chris E Freise,⁶ Kim M Olthoff,⁷ Rafik M Ghobrial,⁸ Paula McIver,⁹ and Robert Fisher¹

¹Division of Transplantation, Department of Surgery, ²Departments of Pathology and ³Biostatistics, and ⁴Massey Cancer Center, Virginia Commonwealth University, Richmond, Virginia; ⁵Department of Medicine, Northwestern University, Chicago, Illinois; ⁶Department of Surgery, University of California, San Francisco, California; ⁷Department of Surgery, University of Pennsylvania, Philadelphia, Pennsylvania; ⁸Department of Surgery, University of California, Los Angeles, California; ⁹Department of Surgery, University of North Carolina, Chapel Hill, North Carolina, United States of America

The role of chronic hepatitis C virus (HCV) in the pathogenesis of HCV-associated hepatocellular carcinoma (HCC) remains controversial. To understand the transition from benign to malignant, we studied the gene expression patterns in liver tissues at different stages, including normal, cirrhosis, and different HCC stages. We studied 108 liver tissue samples obtained from 88 distinct patients (41 HCV-cirrhotic tissues, 17 HCV-cirrhotic tissues from patients with HCC, and 47 HCV-HCC tissues). Differentially expressed genes (DEG) were studied by use of high-density oligonucleotide arrays. Among probe sets identified as differentially expressed via the F test, all pairwise comparisons were performed. Cirrhotic tissues with and without concomitant HCC were further evaluated, and a classifier was used to predict whether the tissue type was associated with HCC. Differential expression profiles were analyzed using Interaction Networks and Functional Analysis. Characteristic gene signatures were identified when normal tissue was compared with cirrhosis, cirrhosis with early HCC, and normal with HCC. Pathway analysis classified the cellular and biological functions of the DEG as related to cellular growth and proliferation, cell death and inflammatory disease in cirrhosis; cell death, cell cycle, DNA replication, and immune response in early HCCs; and cell death, cell growth and proliferation, cell cycle, and DNA repair in advanced HCCs. Characteristic gene signatures were identified at different stages of HCV-HCC progression. A set of genes were identified to predict whether the cirrhotic tissue was associated with HCC.

© 2009 The Feinstein Institute for Medical Research, www.feinsteininstitute.org

Online address: <http://www.molmed.org>

doi: 10.2119/molmed.2008.00110

INTRODUCTION

According to the most recently available worldwide estimates, liver cancer is the sixth leading cancer type, with 626,162 cases estimated in 2002, and is the third leading cause of cancer death, with an estimated 598,321 deaths that same year (<http://www-dep.iarc.fr/>) (1). The incidence of hepatocellular carcinoma (HCC) in the US is increasing because of the increased prevalence of hepatitis C virus (HCV) infection (2–4).

Surgical resection and liver transplantation are still the only potentially curative treatments for HCC (5–7). Because 80% of HCC patients in the US have cirrhosis, optimum care requires the complex analysis of cancer stage to predict recurrence, and the determination of liver reserve to predict suitability of resection versus total hepatic replacement to prevent death from liver failure (5,6,8,9).

A vast amount of information regarding genetic markers and genomic aberrations,

as well as gene expression, is being accumulated for the study of HCC (9–15). The major risk factors for HCC development are now well defined, and some of the multiple steps involved in hepatocarcinogenesis have been elucidated in recent years (16,17). Although an association between HCV infection and later development of HCC has been established, currently the actual mechanisms that lead to malignant transformation are largely unknown.

Molecular genetic analyses have shown that genomic changes accumulate during the development and progression of HCC (18–23). The molecular pathogenesis of HCC is very complex (18–25). Furthermore, different risk factors (viral infections, alcohol consumption, aflatoxin, hemochromatosis, and nonalcoholic steatohepatitis) are implicated in

Address correspondence and reprint requests to Valeria Mas, Associate Professor of Surgery and Pathology, Division of Transplant, Department of Surgery, West Hospital ninth floor, 1200 East Broad Street, PO Box 980057, Richmond, VA, 23298-0057. Phone: (804) 828-2364; E-mail: vmass@mcvh.vcu.edu.

Submitted December 10, 2008; Accepted for publication December 11, 2008; Epub (www.molmed.org) ahead of print December 15, 2008.

HCC development. Each of these scenarios involves different genetic and epigenetic alterations, chromosome aberrations, gene mutations, and altered molecular pathways (18–25). Moreover, the study of the step-by-step oncogenic process is particularly difficult, owing to the fact that samples from preneoplastic lesions can be prospectively collected almost only in transplantation centers by using the whole explanted liver. Despite these complexities, DNA microarrays have been used recently to profile global changes in gene expression in liver samples obtained from patients with HCC to identify subgroups of HCC that differ according to etiological factors (26), rate of recurrence (27), and intrahepatic metastasis (28), as well as novel molecular markers for HCC diagnosis (26–28). However, most of these studies identified genes that are related to limited aspects of tumor pathogenesis. In addition, the majority of these studies looked at the wide variety of HCC tumors. In this study we studied the genes involved in viral tumorigenesis and tumor initiation in HCV-induced HCC.

MATERIALS AND METHODS

Samples

Samples were obtained from patients awaiting and undergoing liver transplantation at Virginia Commonwealth University; University of North Carolina; University of California, Los Angeles; University of Pennsylvania; and Northwestern University as part of the Adult to Adult Living Donor Liver Transplant Cohort Study (A2ALL; NIDDK 1U01DK62531). Laboratory techniques were centralized in the Molecular Transplant Research section of the Division of Transplant at the Virginia Commonwealth University. The research protocol was approved by the respective institutional review boards, and informed consent was obtained for all study participants. In addition, normal livers and tumor samples were also obtained through the Liver Tissue Cell Distribution System (NO1-DK-7-0004/

HHSN26700700004C). Patient and sample information was obtained through the Data Coordinator Center at the University of Michigan (A2ALL: Adult to Adult Living Donor Liver Transplant Cohort Study 1U01DK62531).

Sample Characteristics

This study was restricted to patients with HCV infection. The characteristics of the samples are described in Supplemental Table 1. The study included 108 liver tissue samples obtained from 88 distinct patients, 41 HCV-cirrhotic tissues from patients without HCC, 17 cirrhotic tissues from patients with HCC, and 47 HCV-HCC tissues; 13 patients with HCC-cirrhosis provided both tumor and cirrhotic tissue, and 3 patients contributed with duplicate cirrhotic or tumor tissue for array processing. Also, 19 normal livers were included. Liver function and histopathology for these liver donors were shown to be normal. All 19 patients from whom normal liver samples were obtained were seronegative for HCV Ab.

Sample Collection and Pathological Data

Liver tissue samples were collected in liquid nitrogen or RNA later solution (Ambion, Austin, TX, USA), and stored at -80°C until use. Explanted livers were sliced at intervals of 4–5 mm, and all suspicious nodules for HCC processed for light microscopy. Sections from the tumor/tumors were also fixed in formalin and processed for light microscopy. The size, multiplicity, grade, infiltration into adjacent liver, encapsulation, and vascular invasion (micrometer versus millimeter portal or hepatic vein) were determined for all tumors. Histopathological classification of HCC was performed according to the Edmondson grading system; clinical stages were determined according to the modified TNM classification of the American Tumor Study Group mandated by the United Network for Organ Sharing (29). In the HCC cases, only samples with at least 85% of tumor were further studied.

The cirrhotic tissues from explanted livers were classified using Knodell score and Ishak grade (30).

RNA isolation, cDNA Synthesis, and *in vitro* Transcription for Labeled cRNA Probe

The sample preparation protocol followed the Affymetrix GeneChip® Expression Analysis Manual (Santa Clara, CA, USA). To ensure processing of samples with a high percentage of tumor and minimal to no necrosis, 5- μm frozen section slides were made and stained with hematoxylin and eosin for pathologist evaluation to determine tumor cell percentage. After the evaluation, normal/necrotic tissue was macrodissected. Sections were prepared and total RNA was extracted by use of TRIzol (Life Technologies, Rockville, MD, USA).

Briefly, total RNA was reverse transcribed with T7-polydT primer and converted into double-stranded cDNA (One-Cycle Target Labeling and Control Reagents, Affymetrix), with templates being used for an *in vitro* transcription reaction to yield biotin-labeled antisense cRNA. The labeled cRNA was chemically fragmented and made into the hybridization cocktail according to the Affymetrix GeneChip protocol, which was then hybridized to HG-U133A (n = 9) and HG-U133A 2.0 (n = 134) GeneChips. The array image was generated by the high-resolution GeneChip® Scanner 3000 by Affymetrix®. The data were analyzed by use of Affymetrix Microarray Suite software, version 5.0 and Bioconductor packages (31) available in the R programming environment (32). Data is available from Gene Expression Omnibus (GEO), accession # GSE14323. Information about quality control (QC) procedures performed are provided in QC Supplemental Data.

Data Analysis

To mitigate any effect due to GeneChip type, prior to obtaining probe set expression summaries, the 9 HG-U133A GeneChips and 115 HG-U133A 2.0 GeneChips were independently read into the R programming environment (32)

using the *affy* Bioconductor package (31,33). Thereafter, probe-level data from the two GeneChip types were merged by probe sequence using the *matchprobes* package in R (34). Subsequently, the robust multiarray average method was used to obtain probe-set expression summaries (35,36).

Because of the lack of independence among the 124 GeneChips present owing to contribution of more than one sample by some patients, for each probe set, a mixed-effects analysis of variance model was fit, taking into consideration probe set expression as the response, tissue type (normal liver, HCV cirrhosis without HCC, HCV cirrhosis with HCC, HCV-HCC) as the fixed effect of interest, and including subject as a random-effect term. A group means parameterization of tissue type was included as the fixed-effect term to facilitate extraction of linear contrasts of interest. To set the P value threshold at which probe sets would be declared significantly differentially expressed, we first examined the P values from the F test for all Affymetrix control probe sets. Because the expectation is that control probe sets are not differentially expressed, we selected the minimum P value among control probe sets as our threshold for declaring a probe set significant (37). All probe sets with a P value from the F test less than this threshold were retained.

Thereafter, among probe sets identified as differentially expressed via the F test, all six pairwise comparisons were performed. The *limma* package was used to fit the mixed-effects models by using restricted maximum likelihood procedure, and an empirical Bayes method was applied to moderate probe set standard errors by borrowing information across the entire set of probe sets (38).

In addition to assessing significance across the three diagnostic groups, we performed a Jockheere-Terpstra test for trend to identify whether gene expression was significantly increasing or decreasing as the tissue progressed from normal to cirrhosis to early HCV-HCC to advanced HCV-HCC (early HCV-HCC was defined

as TNM stage T1N0M0 or T2N0M0, and advanced HCV-HCC was defined as T3N0M0 or T4N0M0 or T4N0M1). Prior to performing the hypothesis tests, the gene expression matrix was filtered to include only independent samples, so for patients contributing both HCV-HCC and HCV cirrhotic tissues, only the HCV-HCC sample was used.

For those with replicate samples from the same tissue, the observed correlation among replicates was high ($P = 0.944$, 0.962 , and 0.923), so that the average probe-set intensities for two replicates was used in place of either individual CEL file. The *SAGx* package in the R programming environment was used in performing the nonparametric test for trend.

It was also of interest to compare the cirrhotic tissues with and without concomitant HCC. The 58 CEL files representing the cirrhotic tissues were read into the R programming environment and normalized together by use of quantile normalization, and RMA expression summaries were obtained. There were two patients who contributed replicate cirrhotic samples, therefore subject was included as a random effect term in the linear model. A group means parameterized linear model was fit, and an empirical Bayes method was used to adjust probe-set variance by borrowing information across the entire set of probe sets. P values from the modulated t -statistics were then obtained. It is known that the Affymetrix control probe sets should not be differentially expressed, thus the minimum P value among the control probe sets was selected as the threshold for identifying differentially expressed probe sets.

Interaction Networks and Functional Analysis

Gene ontology and gene interaction analyses were executed using Ingenuity Pathways Analysis tools 3.0 (IPA) (<http://www.ingenuity.com>). The gene lists containing Entrez GeneIDs as clone identifiers, as well as fold-change values from corresponding supervised analyses, were mapped to their corresponding

gene object in the Ingenuity Pathways Knowledge Base (IPKB). These so-called focus genes were then used in the network generation algorithm, based on the list of molecular interactions in IPKB. Significance for the enrichment of the genes in a network with particular biologic functions was determined by the right-tailed Fisher's exact test, using a list of all the genes on the array as a reference set.

Validation of Microarray Results

We carried out a quantitative reverse transcriptase–real-time polymerase chain reaction (QPCR) for EDF1 (endothelial differentiation-related factor 1), tissue factor pathway inhibitor (TFPI), cadherin 4 (CDH4), and tumor necrosis factor α -induced protein 3 (TNFAIP3) mRNAs from the same RNA samples that were subjected to microarray study. Total RNA was subjected to reverse transcription using TaqMan® Reverse Transcription Reagents (Applied Biosystems, Foster City, CA, USA) according to the manufacturer's protocol. Real-time PCR reactions then were carried out in an ABI Prism 7700 Sequence Detection System, using TaqMan® Gene Expression Assays (Applied Biosystems) (EDF1: Hs00610152_m1, TFPI: Hs00196731_m1, CDH4: Hs00899698_m1, and TNFAIP3: Hs00234713_m1). Data were analyzed according to the comparative cycle threshold (Ct) method and normalized by two different housekeeping genes, as recently recommended for studies including HCC tissues (39,40), including β 2-microglobulin and TATA box binding protein expression in each sample. Pearson's correlation coefficient (r) was calculated to examine the relation between microarray and real-time PCR results. $P < 0.05$ were considered significant.

All supplementary materials are available online at www.molmed.org.

RESULTS

Samples

Liver tissue samples were classified for the analysis as normal livers ($n = 19$),

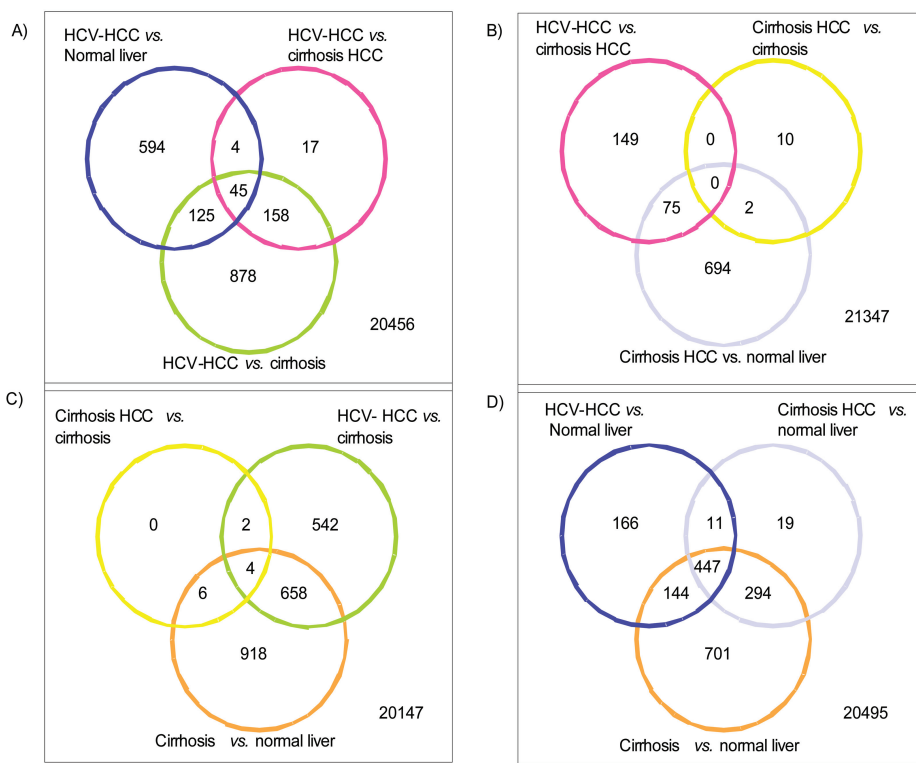


Figure 1. Venn diagram illustrating the number of significant genes for all the pairwise comparisons among the following groups: normal livers, HCV cirrhotic tissues, HCV cirrhotic tissues from patients with HCC (non-tumor samples), and HCV-HCC. Blue circle: comparison analysis between HCV-HCC versus normal liver; pink circle: comparison analysis between HCV-HCC versus cirrhotic HCC; green circle: comparison analysis between HCV-HCC versus cirrhosis; yellow circle: comparison analysis between cirrhosis HCC versus cirrhosis; light blue circle: comparison analysis between cirrhosis HCC versus normal liver; orange circle: comparison analysis between cirrhosis versus normal liver.

HCV cirrhosis (n = 41), early HCC (T1N0M0 and T2N0M0, n = 26), advanced HCC (T3N0M0 and T4N0M0, n = 13), and very advanced HCC (T4N0M1, n = 7). For some of the analysis, HCV-HCC samples were associated in the same group (HCV-HCC samples).

Identifying Differentially Expressed Genes among Different Tissue Groups

The minimum *P* value among control probe sets, which was used as our threshold, was 1.037E-10. From this analysis, 2262 probe sets were found to be significantly differentially expressed among the four diagnostic groups (normal liver, HCV cirrhosis without HCC, HCV cirrhosis with HCC, HCV-HCC) (Figure 1).

Figure 1A illustrates the number of significant probe sets for the three pairwise comparisons, including HCV-HCC versus normal liver (n = 768 with 594 probe sets being unique for this comparison), HCV-HCC versus HCV cirrhosis from patients with HCC (n = 224, with 17 probe sets being unique for this comparison), and HCV-HCC versus HCV cirrhosis without HCC (n = 1206 with 878 probe sets being unique for this comparison). From the analysis of the 17 unique probe sets (n = 16 genes) differentially expressed in HCV-HCC samples when compared with HCV cirrhotic tissues from patients with HCC (Supplemental Table 2), genes involved in regulation of transcription (*TAF10*) and DNA repair

(*APEX2*) were upregulated in HCV-HCC samples, whereas coagulation factors (*F5*, *TFPI*) and apoptosis genes (*BAG5*) were downregulated.

We identified 878 probe sets (n = 867 genes) that were differentially expressed in HCV-HCC samples compared with HCV cirrhotic tissues from patients without HCC (after elimination of overlapping probe sets). From the analysis of the resulting 878 probe sets (Figure 1A) we identified genes related to cell division (*CDC2L6*, *CDKN1C*, *CDCA4*), cell adhesion (*CDH22*, *CDH4*, *CDH6*), apoptosis (*TNFAIP3*, *TNFRSF21*, *MCL1*, *BNIP1*, *BAK1*) being differentially expressed in HCV-HCC tissue samples compared with HCV cirrhotic tissues.

Figure 1B illustrates the number of significant probe sets for the three pairwise comparisons including HCV-HCC versus HCV cirrhosis from patients with HCC (n = 224 probe sets), HCV cirrhosis from patients with HCC versus HCV cirrhosis (n = 12 probe sets), and HCV cirrhosis from patients without HCC versus normal liver (n = 771 probe sets). Also, overlapping of probe sets among the different analyses is shown. In this analysis, only 10 probe sets were identified as differentially expressed when HCV cirrhosis tissues from patients with and without HCC were compared. *TNFSF12*, *CD97*, and *TMEM109* were the more relevant genes in this list. The protein encoded by the *TNFSF12* gene is a cytokine that belongs to the tumor necrosis factor ligand family. This protein is a ligand for the FN14/TWEAKR receptor. This cytokine can induce apoptosis via multiple pathways of cell death in a cell type-specific manner. This cytokine is also found to promote proliferation and migration of endothelial cells, and thus acts as a regulator of angiogenesis. However, when this analysis was evaluated with HCV-HCC versus cirrhotic tissues from patients without HCC and tissues from patients without HCC versus normal liver, these probe sets (n = 10) were shared with the other comparisons, leaving no unique probe sets dif-

ferentially expressed when HCV cirrhotic tissues from patients with and without HCC were compared (Figure 1C).

We were also able to identify 149 unique probe sets statistically differentially expressed in HCV-HCC samples compared with HCV cirrhotic tissues from patients with HCC (Figure 1B). The downregulated genes in HCV-HCC samples were principally associated with apoptosis. Also, coagulation-factor genes were downregulated in HCV-HCC. The expression of oncogenes such as *FYN* and *RAS* was also affected in HCV-HCC samples compared with HCV cirrhosis from patients with HCC.

Figure 1C illustrates the number of significant probe sets for the three pairwise comparisons, including HCV cirrhosis from patients with HCC versus HCV cirrhosis from patients without HCC (n = 12 probe sets), HCV-HCC versus HCV cirrhosis (n = 1206 probe sets), HCV cirrhosis from patients with HCC versus normal liver (n = 771 probe sets). From this analysis, we were able to identify the unique statistically different probe sets (n = 542) in HCV-HCC compared with HCV cirrhosis tissue from patients without HCC. This analysis indicated that genes related to immune response, apoptosis, and growth factors were upregulated in HCV cirrhosis when compared with HCV-HCC. Cell-division and cell-cycle genes were upregulated in HCV-HCC samples. Core analysis was performed to interpret the data set in the context of biological processes, pathways, and molecular networks. We identified 21 networks with a score ≥ 15 (including at least 13 molecules). Two of the top-scored networks were merged (Figure 2). The top molecular and cellular functions in the networks were classified as: cell morphology, development, and cancer (score = 41); and gene expression, viral function, and cancer (score = 32).

Finally, Figure 1D illustrates the number of significant probe sets for the three pairwise comparisons, including HCV-HCC versus normal liver (n = 768 probe sets); HCV cirrhosis with HCC versus

normal liver (n = 771 probe sets), and HCV cirrhosis without HCC versus normal liver (N = 1586 probe sets).

Gene Signatures Involved in HCV-Induced HCC

In testing whether gene expression significantly increased or decreased across the tissue types from normal (n = 19) to HCV cirrhosis (n = 43) to early HCV-HCC (n = 26) to advanced HCV-HCC (n = 20), there were 1997 probe sets (1598 genes) with a significant decreasing trend and 56 probe sets (50 genes) with a significant increasing trend in expression values.

From this analysis we observed that an important number of the downregulated genes from normal to advanced HCV-HCC were genes associated with normal liver function. These included, for example, coagulation factors such as *TFPI*, *C8A*, *C1RL*, and *F9*, among others. Also, genes related to immune response (that is, *IL18R*, *IL33*, *IL4R*, and *ILF3*, among others) and interferon-inducible genes and interferon receptors (*IFITM2*, *ISG20L2*, *IFNAR1*, *IFNAR2*) presented a negative trend among tissues. Many genes of the RAS family (that is, *RASB27A*, *RAB14*, *RAB1A*, *KRAS*, and *RAB2A*, among others) as well as growth factors and receptors were present in the negative trend list. From the analysis of the more important molecular and cellular functions associated with these genes, gene expression ($P = 1.4E-07$ – $9.0E-03$), cell death ($P = 2.7E-06$ – $9.9E-03$), RNA post transcriptional modifications ($P = 5.6E-06$ – $2.1E-03$), molecular transport ($P = 8.1E-6$ – $7.4E-03$), and protein trafficking ($P = 8.0E-6$ – $7.4E-03$) were the more relevant. Acute phase-response signaling ($P = 2E-10$) and JAK/STAT signaling ($6.7E-05$) represented the more important canonical pathways.

From the analysis of the probe sets, with a positive trend from normal liver to HCV cirrhosis to HCV-HCC, we observed that the list included MHC class-I receptor activity (that is, *HLA-G*, *-F*, and *-B*), DNA damage checkpoint (that

is, *CHECK1*), cell division (that is, *BIRC5*, *KNTC1*), and ubiquitin cycle (that is, *FBXO41*, *FBXL7*, *UBD*) genes.

The associated network functions: cancer, cell cycle, cell death (score = 44); and inflammatory disease, cancer, cellular growth and proliferation (score = 32) were the top-scored networks. The top canonical pathways included cell cycle ($P = 1.3E-05$), p53 signaling ($P = 4.2E-03$), antigen presentation pathway ($P = 5.2E-03$), and protein ubiquitination pathway ($P = 1.4E-2$).

Differential Gene Expression Patterns between HCV-Cirrhotic Liver Tissue from Patients with and without HCC

When we examined only 58 HCV cirrhotic tissues, the comparison between cirrhotic tissues with (n = 17) and without (n = 41) HCC yielded differentially expressed 863 probe sets. The 17 cirrhotic tissues were collected from explanted livers to ensure the absence of HCC. From this analysis we observed that genes related to apoptosis (*BCL7*, *TNFRSF1A*, *TNF5F13*, *TRAADD*, *LITAF*) and angiogenesis (*TNFAIP2*, *TNFSF12*) were downregulated in HCV cirrhotic tissues from patients with HCC, whereas genes related to immune response (*IL1R*, *IL9R*, *IL8R*) and response to stress (*TP53I11*, *TP53TG3*) were downregulated compared with HCV cirrhotic tissues from patients without HCC. To redefine the significance of the altered genes, we next investigated the biological interactions using the IPA tool and found the genes to map to genetic networks with functional relationships. Five networks with high scores (>15) were found associated to HCV cirrhosis without HCC. These networks had between 10–17 genes affected, being related to cellular death; cell-to-cell signaling; cancer; and DNA replication, recombination, and repair. We also performed gene ontology analysis using the IPA tool. Twenty functions were identified as having high scores. The functions with the highest P values were related to cell cycle ($P = 1.6E-06$ – $1.1E-02$ [25 molecules]) and cell death

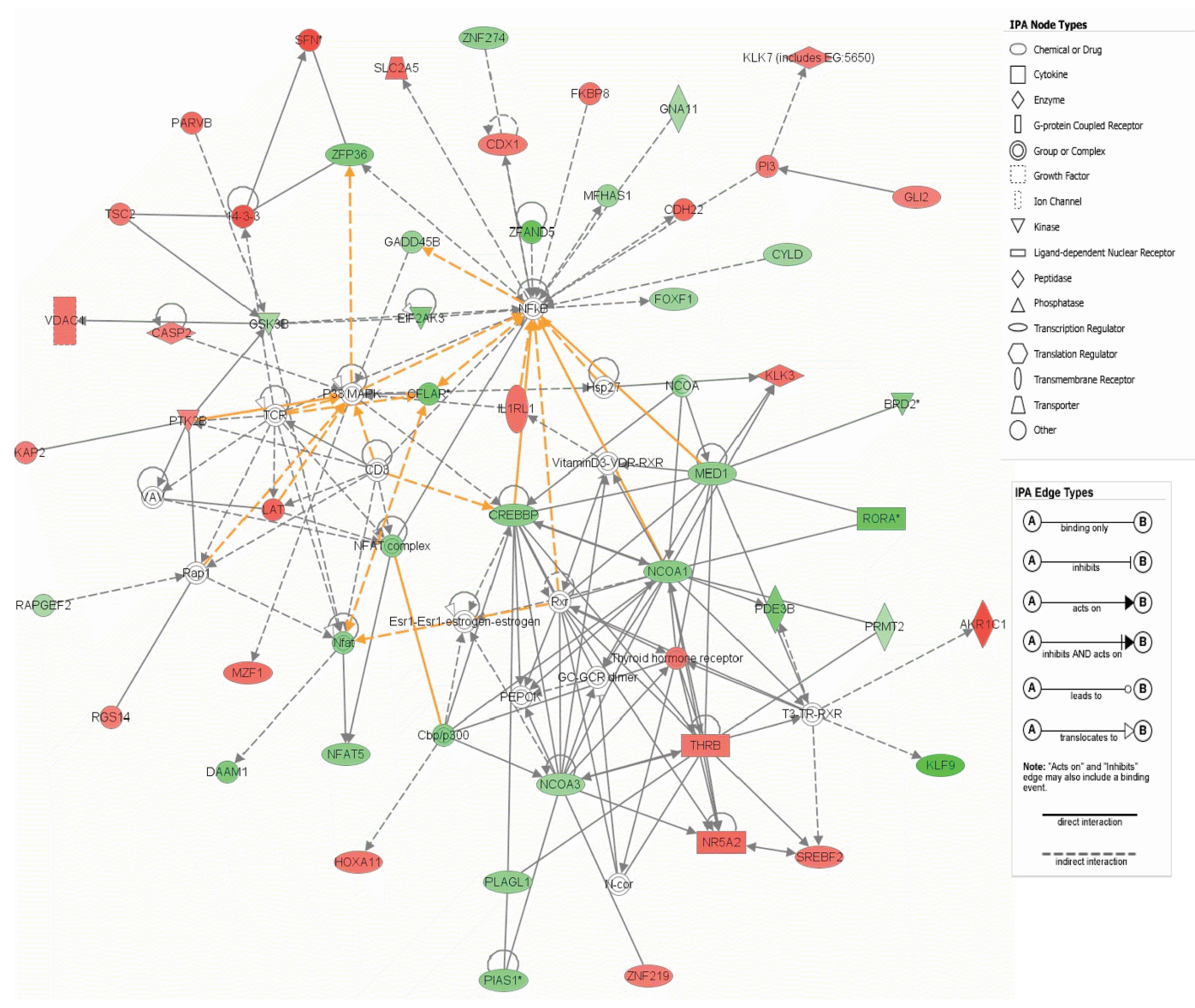


Figure 2. The 2 top-scoring networks of interactions among the differentially expressed genes between HCV-HCC versus HCV-cirrhotic tissues. Core analysis was performed to interpret the data set in the context of biological processes, pathways, and molecular networks for the 542 probe sets differentially expressed only in the HCV-HCC versus HCV cirrhosis comparison. The top molecular and cellular functions in the networks were classified as cell morphology, development, and cancer (score = 41) and gene expression, viral function, and cancer (score = 32). Interconnections of significant functional networks in HCV cirrhosis are indicated. Spatial location of nodes into their corresponding subcellular locations is also indicated. The meaning of the node shapes and interaction edges is also indicated. Image, ©2000–2008 Ingenuity Systems, Inc. All rights reserved.

($P = 3.0E-05$ – $1.1E-02$ [12 molecules]). The top canonical pathways associated with HCV cirrhosis included p53 signaling ($P = 0.0012$), acute-phase response signaling ($P = 0.0016$), xenobiotic metabolism signaling ($P = 0.008$), IL-6 signaling ($P = 0.01$), and NFR2-mediated oxidative stress response ($P = 0.013$).

Classifier to Predict whether Tissue Type was Associated with HCC

We sought to derive a classifier to predict whether the HCV cirrhotic tissue was from a patient without HCC versus cirrhotic tissue with HCC. The 58 CEL files representing the cirrhotic tissues were read into the R programming envi-

ronment and normalized together using quantile normalization, and RMA expression summaries were obtained. Prior to deriving a classifier, all Affymetrix control-probe sets were removed. Afterward, the random forest algorithm was used to predict tissue type using the 22,215 RMA probe-set expression sum-

Table 1. Fifteen probe sets consistently identified among the random forest classifiers as being important both with respect to the mean decrease in accuracy and the mean decrease in the Gini index.

AffyID	Symbol	Gene name	HCV-cirrhosis from patients with HCC	HCV-cirrhosis only	Fold change
218062_x_at	<i>CDC42EP4</i>	<i>CDC42 effector protein (Rho GTPase binding) 4</i>	7.73	8.34	1.52
201362_at ^a	<i>IVNS1ABP</i>	<i>Influenza virus NS1A-binding protein</i>	6.83	7.27	1.36
221270_s_at	<i>QTRT1</i>	<i>Queuine tRNA-ribosyltransferase 1</i> (alias: <i>tRNA-guanine transglycosylase</i>)	7.08	7.89	1.75
221217_s_at	<i>A2BP1</i>	<i>Ataxin 2-binding protein 1</i>	5.5	5.85	1.27
209059_s_at	<i>EDF1</i>	<i>Endothelial differentiation-related factor 1</i>	9.57	9.99	1.34
211241_at	<i>ANXA2P3</i>	<i>Annexin A2 pseudogene 3</i>	8.23	8.7	1.38
218059_at ^a	<i>ZNF706</i>	<i>Zinc finger protein 706</i>	5.57	5.91	1.26
201332_s_at	<i>STAT6</i>	<i>Signal transducer and activator of transcription 6, interleukin-4 induced</i>	7.9	8.31	1.32
213183_s_at	<i>CDKN1C</i>	<i>Cyclin-dependent kinase inhibitor 1C (p57, Kip2)</i>	4.71	5.21	1.42
208891_at	<i>DUSP6</i>	<i>Dual specificity phosphatase 6</i>	7.58	8.31	1.66
200827_at	<i>PLOD1</i>	<i>Procollagen-lysine 1, 2-oxoglutarate 5-dioxygenase 1</i>	9.38	9.99	1.53
213042_s_at	<i>ATP2A3</i>	<i>ATPase, Ca++ transporting, ubiquitous</i>	7.47	7.99	1.43
45749_at	<i>FAM65A</i>	<i>Family with sequence similarity 65, member A</i>	6.06	6.53	1.39
204164_at	<i>SIPA1</i>	<i>Signal-induced proliferation-associated gene 1</i>	5.23	6.03	1.73
203315_at	<i>NCK2</i>	<i>NCK adaptor protein 2</i>	6.29	6.62	1.26

^a201362_at and 218059_at were the probe sets in the final multivariate logistic regression model.

maries as covariates; owing to memory limitations, three random forests were separately derived by using roughly 1/3 of the probe sets. Thereafter, for the three independent random forests, all probe sets with a Gini importance measure exceeding the $\mu_{\text{Gini}} + 3\text{SE}_{\text{Gini}}$ were retained, and a subsequent random forest predicting tissue type using only these important probe sets was derived. All random forests consisted of 5000 trees. This entire process was repeated three times to examine the stability of the probe sets with the highest variable importance values. Because the random forest uses bootstrap samples in deriving each classification tree, there is a natural test set, which consists of those observations not in the bootstrap sample, to provide an unbiased estimate of classification error. The random forest had an unbiased error rate of 8.93% estimated using the observations not in the bootstrap re-samples (Supplemental Figure 1).

Fifteen probe sets were consistently identified among the random forest classifiers as being important both with respect to the mean decrease in accuracy and the mean decrease in the Gini index (Table 1). A pairwise scatterplot for

these 15 probe sets revealed that all probe sets were correlated, with the cirrhotic tissues with HCC having lower expression values (red) than the cirrhotic tissues without HCC (blue) (Supplemental Figure 2). Owing to the correlation among these 15 probe sets, a multivariable logistic regression model was derived using a forward variable selection strategy to obtain a more parsimonious set of genes predictive of tissue of origin. First, all univariable logistic regression models were fit, and that model with the smallest log-likelihood was selected as the most important probe set. Thereafter, all possible two-variable models containing this probe set and one other were fit, and that probe set having the most significant decrease in the log-likelihood was retained. This process was repeated until there was no significant decrease in the log-likelihood. The probe sets in the final multivariate logistic regression model were 201362_at and 218059_at (Table 1). Moreover, other bivariable models using these 15 probe sets will yield similar performance, so that there is a multiplicity of models that could be used to predict tissue source. After the

final model was determined, the Hosmer and Lemeshow goodness-of-fit test was performed, which indicated no departure from goodness-of-fit ($P = 0.4238$). The model was then used in cal-

Table 2. A. Cross-tabulation of true and predicted class where the predicted class was determined from the multivariable logistic regression model using $\hat{p}_i \geq 0.5$.

Classified	True class		Total
	No HCC	HCC	
No HCC	40	2	42
HCC	0	14	14
Total	40	16	56

B. Sensitivity, specificity, positive predicted value, negative predicted value, and accuracy of the predicted class was determined from the multivariable logistic regression model using $\hat{p}_i \geq 0.5$.

Using $\hat{p}_i \geq 0.5$	%
Sensitivity	100.0%
Specificity	87.5%
Positive predictive value	95.2%
Negative predictive value	100.0%
Accuracy	96.4%

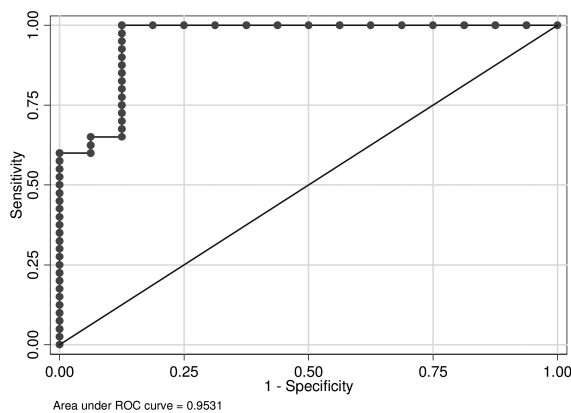


Figure 3. Receiver operating characteristic (ROC) curve.

culating the predicted probabilities \hat{p}_i for patients $i = 1, \dots, 56$. If $\hat{p}_i \geq 0.5$ the patient was predicted to have HCC, and no HCC otherwise. These classifications cross-tabulated against the true class (Table 2A) were then used in estimating sensitivity, specificity, positive predicted value, negative predicted value, and accuracy (Table 2B). The receiver-operator curve was also plotted (Figure 3).

Gene Expression Profiles among Early, Advanced, and Very Advanced HCV-HCC

After applying the Jonckheere-Terpstra test for trend to assess which probe sets have a significant increasing trend and again to assess which probe sets have a significant decreasing trend, the q -values were estimated as a means of assessing the false discovery rate. The increasing trend analysis yielded 189 probe sets, and the decreasing trend analysis yielded only 3 probe sets. The *angiopoietin 1* gene was affected by the positive trend. Angiopoietins are proteins with important roles in vascular development and angiogenesis. All angiopoietins bind with similar affinity to an endothelial-cell-specific tyrosine-protein kinase receptor. The protein encoded by this gene is a secreted glycoprotein that activates the receptor by inducing its tyrosine phosphorylation and plays a critical role in mediating reciprocal interactions between the endothelium and surrounding matrix and mesenchyme. Genes involved

in DNA repair were also present in the positive trend results (*RAD51*, *RAD17*, *PML*, *FANCI*).

Gene Expression Quantitation Using QPCR

Expression levels of the *EDF1*, *TFPI*, *CDH4*, and *TNFAIP3* mRNAs were further confirmed using QPCR. These genes were identified as significant differen-

tially expressed between different tissues types and/or as a part of classifier to predict whether the tissue type was associated with HCC. The results from the microarray were reproduced by QPCR ($r = 0.71$, $P < 0.001$; $r = 0.78$, $P = 0.002$; $r = 0.68$, $P < 0.001$; and $r = 0.77$, $P < 0.001$, respectively) (Figure 4).

DISCUSSION

HCC due to HCV may be an indirect result of enhanced hepatocyte turnover that occurs in an effort to replace infected cells that have been immunologically attacked. Chronic inflammation, immune-mediated hepatocellular destruction, and liver regeneration underlie cirrhosis, and are thought to play central roles in primary carcinogenesis (18–23). Viral functions may also play a more direct role in mediating oncogenesis.

In the present study, we aimed to identify the genes implicated in the origin and progression of HCV-HCC. We studied the molecular biology of HCV-HCC including the study of HCV-

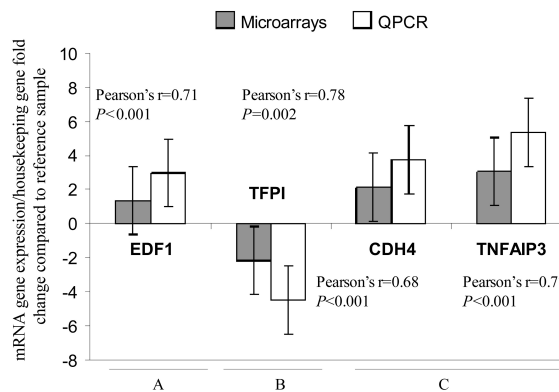


Figure 4. Validation of microarray results obtained using QPCR to analyze expression levels of *EDF1*, *TFPI*, *CDH4*, and *TNFAIP3* genes. Total RNA from the same RNA samples that were subjected to microarray study were used in the QPCR validation reactions. Values were expressed as mRNA gene expression level (for each specific gene)/housekeeping gene fold-change expression compared with the reference sample. The gene expression of the study genes was identified as significant differentially expressed when comparing different group of tissues samples. Specifically, *EDF1* was identified as differentially expressed when HCV cirrhotic tissues from patients with HCC (nontumor samples) ($n = 17$) were compared with HCV cirrhotic tissues from patients without HCC ($n = 41$) (A). *TFPI* was differentially expressed in HCV-HCC samples ($n = 47$) when compared with HCV cirrhotic tissues from patients with HCC ($n = 17$) (B). *CDH4* and *TNFAIP3* were identified as differentially expressed when HCV-HCC samples ($n = 47$) were compared with HCV cirrhotic tissues ($n = 41$) (C).

cirrhotic liver, HCV-cirrhotic livers from patients with HCC, and HCC at different stages.

For elucidating the genes and molecular pathways involved in the HCV-induced HCC progression, different analyses were performed. Among probe sets identified as differentially expressed via the F test, all possible six pairwise comparisons were performed. This analysis allowed us to identify the unique probe sets characterizing every tissue type/condition when overlapping genes from another comparison were eliminated. As expected, the gene expression patterns were found to vary significantly among the HCC and normal liver samples. In concordance with previous publications (41,42), genes associated with cell proliferation and mitosis were found to have increased expression in HCC samples. In contrast, most genes that were expressed at lower levels in HCC than in normal liver tissues comprised genes specifically expressed in differentiated hepatocytes. These observations suggest that accelerated cell proliferation accomplished by the upregulation of proliferation and mitotic-associated genes are generally involved in hepatocarcinogenesis. However, the downregulation of liver-specific genes is possibly associated with dedifferentiation of cancer cells during tumor progression.

Cell adhesion, cell division, and apoptosis were the more important cellular and molecular functions when HCV-HCC samples were compared with HCV-cirrhotic tissues. A lower number of differentially expressed genes were observed when HCV-HCC samples were compared with HCV-cirrhotic tissues from patients with HCC (nontumor samples). Cirrhosis is a recognized, established risk factor for HCC, and this finding might represent a molecular compromise already present in these tissues even when nonhistological evidence of tumor was present.

To examine the molecular mechanisms during the process of liver carcinogenesis, it is considered reasonable to investi-

gate the noncancerous portion of the liver tissue, because multicentric occurrence of HCC is mainly associated with underlying chronic liver damage rather than adverse tumor factors. For this reason, we evaluated HCV-cirrhotic tissues from patients with HCC as an independent sample group. When examining only 58 cirrhotic tissues, the comparison between cirrhotic tissues with and without HCC yielded 863 probe sets differentially expressed. The top canonical pathways associated with HCV cirrhosis in patients with HCC included p53 signaling, acute-phase response signaling, xenobiotic metabolism signaling, IL-6 signaling, and NFR2-mediated oxidative stress response. These results might have important impact in the early and accurate diagnosis of HCV-HCC in cirrhotic patients.

It was of interest to derive a classifier to predict whether the studied HCV cirrhotic tissue was from a patient without HCC (from explanted livers) versus cirrhotic tissue with HCC. It can be of great importance to identify markers that can indicate the presence of HCC in cirrhotic tissues even when nontumor cells are present in the study sample. Fifteen probe sets were consistently identified among the random forest classifiers. Moreover, in cross-tabulation of true and predicted class for which the predicted class was determined, very good values were obtained for sensitivity, specificity, positive predicted value, negative predicted value, and accuracy. These findings might be important in diagnosis of HCC in HCV-cirrhotic patients.

In testing whether gene expression significantly increased or decreased across the tissue types from normal liver to HCV-cirrhosis to early HCV-HCC to advanced HCV-HCC, trend analysis was performed. A high number of genes related with normal liver function were present in the negative-trend gene list. Acute phase response signaling and JAK/STAT signaling represented the more important downregulated canonical pathways. Cell cycle, p53 signaling,

antigen presentation pathway, and protein ubiquitination pathway were the top canonical pathway in positive-trend analysis.

In conclusion, we identified gene signatures that distinguish the pathological stages of HCC and potential molecular markers for early HCC diagnosis in high risk cirrhotic HCV patients. These findings provide a comprehensive molecular portrait of genomic changes in progressive HCV-related HCC.

ACKNOWLEDGMENTS

This work has been presented in part at the American Transplant Congress meeting, Toronto, 2008.

This project was supported by a National Institute of Diabetes and Digestive and Kidney Diseases (NIDDK) grant, RO1DK069859.

The following individuals were instrumental in the planning, conduct and/or care of patients enrolled in this study at each of the Adult to Adult Living Donor Liver Transplant Cohort Study (A2ALL; NIDDK 1U01DK62531) participating institutions as follows:

Northwestern University, Chicago, IL (DK62467): PI: MMI Abecassis, MD, MBA; Co-PI: A Blei, MD; Study Coordinator: P Al-Saden, RN, CTCC

University of Pennsylvania Health System, Philadelphia, PA (DK62494): PI: A Shaked, MD, PhD; Co-PI: KM Olthoff, MD; Study Coordinators: M Kaminski, PA-C; M Shaw, RN, BBA

University of California Los Angeles, Los Angeles, CA (DK62496): PI: RM Ghobrial, MD, PhD; Co-PI: RW Busuttill, MD, PhD; Study Coordinator: L Artinian, RN, MN

University of California San Francisco, San Francisco, CA (DK62444): PI: CE Freise, MD, FACS; Co-PI: NA Terrault, MD; Study Coordinator: D MacLeod, RN

University of Michigan Medical Center, Ann Arbor, MI (DK62498): PI: RM Merion, MD; DCC Staff: ASF Lok, MD; AO Ojo, MD, PhD; BW Gillespie, PhD; DR Armstrong, RN, MS; M Hill-Callahan, BS, LSW; T Howell, BS; L Tong, MS; TH Shearon, MS; KA Wisniewski, MPH; M Lowe, BS

University of North Carolina, Chapel Hill, NC (DK62505); PI: JH Fair, MD; Study Coordinator: CA Nielsen, MA Medical College of Virginia Hospitals, Virginia Commonwealth University, Richmond, VA (DK62531); PI: RA Fisher, MD, FACS; Co-PI: ML Shiffman, MD; Study Coordinators: E Fenick, RN; A Ashworth, RN

DISCLOSURE

We declare that the authors have no competing interests as defined by *Molecular Medicine*, or other interests that might be perceived to influence the results and discussion reported in this paper.

REFERENCES

- Parkin DM, Bray F, Ferlay J, Pisani P. (2001) Estimating the world cancer burden: Globocan 2000. *Int. J. Cancer* 94:153–6.
- Tagger A, et al. (1999) Case-control study on hepatitis C virus (HCV) as a risk factor for hepatocellular carcinoma: the role of HCV genotypes and the synergism with hepatitis B virus and alcohol. Brescia HCC Study. *Int. J. Cancer* 81:695–9.
- Tsukuma H, et al. (1993) Risk factors for hepatocellular carcinoma patients with chronic liver disease. *N. Engl. J. Med.* 328:1797–801.
- El-Serag HB, Mason AC. (1999) Rising incidence of hepatocellular carcinoma in the United States. *N. Engl. J. Med.* 340:745–50.
- Marsh JW, Dvorchik I. (2003) Liver organ allocation for hepatocellular carcinoma: are we sure? *Liver Transpl.* 9:693–6.
- Fisher RA, et al. (2007) Is hepatic transplantation justified for primary liver cancer? *J Surg. Oncol.* 95:674–9.
- Barnett CC Jr, Curley SA. (2001) Ablative techniques for hepatocellular carcinoma. *Semin. Oncol.* 28:487–96.
- Figueras J, et al. (2000) Resection or transplantation for hepatocellular carcinoma in cirrhotic patients: outcomes based on indicated treatment strategy. *J. Am. Coll. Surg.* 190:580–7.
- Iwatsuki S, et al. (2000) Liver transplantation for hepatocellular carcinoma: a proposal of a prognostic scoring system. *J. Am. Coll. Surg.* 191:389–94.
- Marsh JW, et al. (2003) Genotyping of hepatocellular carcinoma in liver transplant recipients adds predictive power for determining recurrence-free survival. *Liver Transpl.* 7:664–71.
- Nam SW, et al. (2005) Molecular changes from dysplastic nodule to hepatocellular carcinoma through gene expression profiling. *Hepatology* 42:809–18.
- Iizuka N, et al. (2003) Oligonucleotide microarray for prediction of early intrahepatic recurrence of hepatocellular carcinoma after curative resection. *Lancet* 361:923–9.
- Mas VR, et al. (2007) Genes associated with progression and recurrence of hepatocellular carcinoma in hepatitis C patients waiting and undergoing liver transplantation: preliminary results. *Transplantation* 83:973–81.
- Mas VR, et al. (2006) Differentially expressed genes between early and advanced hepatocellular carcinoma (HCC) as a potential tool for selecting liver transplant recipients. *Mol. Med.* 12:97–104.
- Moinzadeh P, Breuhahn K, Stutzer H, Schirmacher P. (2005) Chromosome alterations in human hepatocellular carcinomas correlate with aetiology and histological grade - results of an explorative CGH meta-analysis. *Br. J. Cancer* 92:935–41.
- Schafer DF, Sorrell MF. (1999) Hepatocellular carcinoma. *Lancet* 353:1253–7.
- El-Serag HB. (2002) Hepatocellular carcinoma and hepatitis C in the United States. *Hepatology* 36: S74–83.
- Guan XY, et al. (2000) Recurrent chromosome alterations in hepatocellular carcinoma detected by comparative genomic hybridization. *Genes Chromosomes Cancer* 29:110–6.
- Marchio A, et al. (2001) Chromosomal abnormalities in liver cell dysplasia detected by comparative genomic hybridization. *Mol. Pathol.* 54:270–4.
- Wong N, et al. (2001) Hypomethylation of chromosome 1 heterochromatin DNA correlates with q-arm copy gain in human hepatocellular carcinoma. *Am. J. Pathol.* 159:465–71.
- Nishida N, et al. (2003) Chromosomal instability and human hepatocarcinogenesis. *Histol. Histopathol.* 18:897–909.
- Buendia MA. (2002) Genetic alterations in hepatoblastoma and hepatocellular carcinoma: common and distinctive aspects. *Med. Pediatr. Oncol.* 39:530–5.
- Taniguchi K, et al. (2002) Mutational spectrum of beta-catenin, AXIN1, and AXIN2 in hepatocellular carcinomas and hepatoblastomas. *Oncogene* 21:4863–71.
- Al-Sukhun S, Hussain M. (2003) Current understanding of the biology of advanced bladder cancer. *Cancer* 97:2064–75.
- Theodorescu D (2003). Molecular pathogenesis of urothelial bladder cancer. *Histol. Histopathol.* 18:259–74.
- Iizuka N, et al. (2002) Comparison of gene expression profiles between hepatitis B virus- and hepatitis C virus-infected hepatocellular carcinoma by oligonucleotide microarray data on the basis of a supervised learning method. *Cancer Res.* 62:3939–44.
- Okabe H, et al. (2001) Genome-wide analysis of gene expression in human hepatocellular carcinomas using cDNA microarray: identification of genes involved in viral carcinogenesis and tumor progression. *Cancer Res.* 61:2129–37.
- Shirota Y, Kaneko S, Honda M, Kawai HF, Kobayashi K. (2001) Identification of differentially expressed genes in hepatocellular carcinoma with cDNA microarrays. *Hepatology* 33:832–840.
- American Liver Tumor Study Group. (1998) A randomized prospective multi-institutional trial of orthotopic liver transplantation or partial hepatic resection with or without adjuvant chemotherapy for hepatocellular carcinoma. Investigators Booklet and protocol 1998. United Network for Organ Sharing. Policy 3.6.4.4.1998.
- Knodel RG, et al. (1981) Formulation and application of a numerical scoring system for assessing histological activity in asymptomatic chronic active hepatitis. *Hepatology* 1:431–5.
- Gentleman RC, et al. (2004) Bioconductor: open software development for computational biology and bioinformatics. *Genome Biol.* 5:R80.
- R Development Core Team. R: a language and environment for statistical computing. Vienna, Austria: R Foundation for Statistical Computing; 2007.
- Gautier L, Cope L, Bolstad BM, Irizarry RA. (2004) affy-analysis of Affymetrix GeneChip data at the probe level. *Bioinformatics* 20:307–15.
- Huber W, Gentleman R. (2004) matchprobes: a Bioconductor package for the sequence-matching of microarray probe elements. *Bioinformatics* 20:1651–2.
- Irizarry RA, et al. (2003) Exploration, normalization, and summaries of high density oligonucleotide array probe level data. *Biostatistics* 4:249–64.
- Irizarry RA, et al. (2003) Summaries of Affymetrix GeneChip probe level data. *Nucleic Acids Res.* 31:e15.
- Smyth GK. (2005) Limma: linear models for microarray data. In: Gentleman R, Carey VJ, Dudoit S, Irizarry R, Huber W, editors. *Bioinformatics and computational biology solutions using R and Bioconductor*. New York: Springer; p. 397–420.
- Llovet JM, Burroughs A, Bruix J. (2003) Hepatocellular carcinoma. *Lancet* 362:1907–17.
- Gao Q, et al. (2008) Selection of reference genes for real-time PCR in human hepatocellular carcinoma tissues. *J. Cancer Res. Clin. Oncol.* 134:979–86.
- Waxman S, Wurmbach E. (2007) De-regulation of common housekeeping genes in hepatocellular carcinoma. *BMC Genomics.* 8:243.
- Chen X, et al. (2002) Gene expression patterns in human liver cancers. *Mol. Biol. Cell* 13:1929–39.
- Delpuech O, et al. (2002). Identification, using cDNA macroarray analysis, of distinct gene expression profiles associated with pathological and virological features of hepatocellular carcinoma. *Oncogene* 21:2926–37.