Central genetic alterations common to all HCV-positive, HBV-positive and non-B, non-C hepatocellular carcinoma: A new approach to identify novel tumor markers

YUKINORI KUROKAWA[1,2], KIMI HONMA[3,4], ICHIRO TAKEMASA[2], SHOJI NAKAMORI[2], HIROKO KITA-MATSUO[1], MASAAKI MOTOORI[1,2], HIROAKI NAGANO[2], KEIZO DONO[2], TAKAHIRO OCHIYA[3], MORITO MONDEN[2] and KIKUYA KATO[1]

[1]Taisho Laboratory of Functional Genomics, Nara Institute of Science and Technology, Nara 630-0101; [2]Department of Surgery and Clinical Oncology, Graduate School of Medicine, Osaka University, Osaka 565-0871; [3]National Cancer Center Research Institute, Tokyo 104-0045; [4]Koken Bioscience Institute, Tokyo 115-0051, Japan

Received September 15, 2005; Accepted November 4, 2005

Abstract. Hepatocellular carcinoma (HCC) is a common malignancy, but the prognosis remains poor due to the lack of sensitive diagnostic markers. To gain insight into the central molecular features common to all types of HCC, and to identify novel diagnostic markers or therapeutic targets for HCC, we performed a gene expression profiling analysis using a high throughput RT-PCR system. After examining the mRNA expression of 3,072 genes in 204 (119 tumor and 85 non-tumor) liver samples, we identified differential gene expression between the HCV group (n=80), HBV group (n=19) and non-B, non-C group (n=20) with a principal component analysis and a correlation spectrum analysis. After selection of genes differentially expressed between tumor and non-tumor tissues ($p<0.01$) within each HCC group, a total of 51 differentially expressed genes (23 upregulated and 28 downregulated genes) were found to be common to the three HCC groups. Gene Ontology grouping analysis revealed that genes with functions related to cell proliferation or differentiation and genes encoding extracellular proteins were found to be significantly enriched in these 51 common genes. Using an Atelocollagen-based cell transfection array for functional analysis of eight upregulated genes, five (*CANX, FAM34A, PVRL2, LAMR1*, and *GBA*) significantly inhibited cellular apoptosis by two independent assays. In conclusion, we identified 51 differentially expressed genes, common to all HCC types. Among these genes, there was a high incidence of anti-apoptotic activity. This combination approach with the advanced statistical methods and the bioinformatical analysis may be useful for finding novel molecular targets for diagnosis and therapy.

Correspondence to: Dr Kikuya Kato, Research Institute, Osaka Medical Center for Cancer and Cardiovascular Diseases, 1-3-2 Nakamichi, Higashinari-ku, Osaka 537-8511, Japan
E-mail: katou-ki@mc.pref.osaka.jp

Abbreviations: HCC, hepatocellular carcinoma; HCV, hepatitis C virus; HBV, hepatitis B virus; ATAC-PCR, adaptor-tagged competitive PCR; T, tumor; NT, non-tumor; PCA, principal component analysis; CR, correlation ratio; FDR, false discovery rate; GO, gene ontology

Key words: DNA microarray, PCR-array, gene expression profiling, apoptosis

Introduction

Hepatocellular carcinoma (HCC) contributes significantly to death from malignancy throughout the world, and the incidence of HCC is rising sharply, most likely due to the spread of the hepatitis C virus (HCV), particularly in Japan, Spain and Italy (1). Hepatitis B virus (HBV) also represents a major risk factor for HCC in East Asia where the prevalence of this infection is high (1). Although recent progress in diagnostic procedures and surgical techniques have resulted in considerably improved morbidity and mortality rates, the prognosis of HCC remains poor (2). Alpha-fetoprotein (AFP) is the only widely used molecular marker for the diagnosis of HCC, but it fails to identify early stages of HCC (3). Additional biochemical markers are sorely needed to accurately detect HCC during its early stages. Since HCC is one of the many tumors whose origins differ based on the milieu of epidemiological, clinical, and pathological risk factors under which it develops, the search for such markers is complicated by the high degree of hetero-geneity of HCC (4).

DNA microarray studies examining thousands of genes have recently clarified the genetic characteristics of HCV- or HBV-infected HCC to some degree (5-7), and we characterized the gene expression profiles of non-B, non-C HCC (8). In this report, we wished to identify molecular features common to the HCV group, the HBV group and the non-B, non-C group. The identified molecules represent genes intrinsic to HCC with little heterogeneity, and these are promising candidates for novel molecular targets for both diagnostic and therapeutic interventions for HCC (9,10).

We performed gene expression profiling using adaptor-tagged competitive PCR (ATAC-PCR) technology: a PCR-based array system (11). To avoid inflated results due to small sample size, we prepared more than 200 liver samples for this study. Strict statistical analysis allowed us to identify 51 genes common to all etiological groups, which revealed extremely distinct expression patterns between tumorous and non-tumorous livers. Bioinformatical Gene Ontology analysis was employed to examine the annotated functions and cellular characteristics of these common genes. Additionally, we analyzed the relationship of identified genes to apoptosis with hepatoma cell lines using an Atelocollagen-based cell transfection array (12-14).

Materials and methods

Tissues and patients. We obtained 204 liver samples (119 HCC, 85 non-tumorous liver) from 119 patients, with informed written consent, who underwent hepatic resection for HCC at Osaka University Hospital between January 1997 and September 2001. Tumor (T) and non-tumor (NT) tissue samples were enucleated separately from either the tumorous or non-tumorous part of the resected tissue. Serologically, of the 119 patients, 80 were HCV antibody-positive, and 16 were HBs antigen-positive. No patients were double-positive for HCV antibody and HBs antigen. Although three patients were double-negative for HCV antibody and HBs antigen, the presence of HBV mRNA was confirmed using the previously reported RT-PCR method (15). These patients were considered to have HBV-positive HCC for further analysis. For the remaining 20 non-B, non-C patients who were double-negative for HCV antibody and HBs antigen, the absence of viral genomes was confirmed by RT-PCR as reported to rule out possible cryptic HCV or HBV infection (15,16). In total, 80 cases were in the HCV group, 19 in the HBV group, and 20 in the non-B, non-C group. All aspects of our study protocol were approved by the Ethics Committee of Osaka University Medical School.

PCR-based array system. To select genes expressed in liver tissue, we constructed three cDNA libraries: one from a mixture of HCC and non-tumorous liver, one from normal liver, and one from metastatic liver cancer, as described (17). We designed PCR primers for ATAC-PCR reactions for a total of 2,774 genes from these EST collections. In total, we prepared 3,072 primers for ATAC-PCR; this total includes an additional 298 genes established in the literature. The specificity of this gene selection provides an advantage over more universal gene sets, such as those selected from the UniGene database, which include genes not expressed in liver tissue. The ATAC-PCR experimental procedure was performed as described (18,19). The complete list of genes and detailed protocols for the ATAC-PCR experimental procedure are available on our website (http://genome.mc.pref.osaka.jp/).

Statistical analysis of PCR-based array data. The relative expression level of each gene was calculated by calibrating against a standard mixture of more than 20 liver tissues including HCC and non-tumor samples, as described (8,19,20). Following conversion to a logarithmic scale (base 2), the data matrix was normalized to a mean of 0 by standardizing each sample. Principal component analysis (PCA) was performed using GeneMaths 2.0 software.

The correlation ratio (CR) of gene i is defined as follows (21):

$$(CR_i)^2 \equiv \frac{\sum_{c=1}^{C} n_c((\sum_{j=J_c} x_{i,j})/n_c - \bar{x}_i)^2}{\sum_{j=1}^{M}(x_{i,j} - \bar{x}_i)^2}$$

where n_c is the number of genes in a particular class, J_c; $x_{i,j}$ is the expression level of gene i with sample j; and x_i is the average of the expression levels of gene i. In drawing a *CR* curve, genes were sorted by *CR* value order, and the ratios of the original total dataset were compared with those of the permutated data.

Permutation testing, which involves 50,000 randomly permutating class labels to determine gene-class correlations, was used to determine statistical significance as described (22,23). To estimate the false discovery rate (FDR), which is the percentage of genes erroneously identified, we used software downloaded from http://faculty.washington.edu/~jstorey/qvalue (24). Gene Ontology (GO) grouping was performed according to the annotated function in biological process and cellular component with a Generic Gene Ontology Term Finder program downloaded from http://helix.princeton.edu/cgi-bin/GOTermFinder/GOTermFinder (25). This software can find significant GO terms shared among a list of selected genes, helping us to discover what these genes may have in common (26). Each gene annotation was designated by the Gene Ontology Consortium at the European Bioinformatics Institute.

Atelocollagen-based cell transfection array. HepG2 cells were grown in DMEM (Sigma, St. Louis, MO) supplemented with 10% FBS (Gibco-BRL), at 37˚C in a 5% CO_2 humidified atmosphere. Atelocollagen-based cell transfection array, which enables reverse transfection of cells by Atelocollagen-mediated gene transfer, was used for functional analysis of candidate genes (12-14). The full-length cDNAs were cloned downstream of the CMV promoter of the expression vector. The expression vector without an insert was used as the control. The plasmid vectors and Atelocollagen complexes were prepared according to the described method (13). An equal volume of Atelocollagen (0.016% in PBS at pH 7.4) and plasmid DNA solution (5 μg/ml) was combined and mixed for 20 min at 4˚C. The final concentration of Atelocollagen was 0.008%. The Atelocollagen/plasmid DNA complexes were then arrayed and precoated on 96-well plates (plasmid DNA 0.25 μg/50 μl/well). The cultured HepG2 cells were plated into the complex-prefixed 96-well plate and the effects of overexpressed genes were then observed.

In vitro growth and apoptosis assay. For growth assay, cells were seeded at a density of $5x10^3$ cells per well in 100 μl of culture medium on Atelocollagen-based cell transfection array. Cell proliferation was assayed for 4 days with TetraColor One, a cell-proliferation assay reagent (Seikagaku Co., Tokyo, Japan), according to the recommended method; cells were then measured for absorbency in the well at 450 nm with a reference wavelength of 650 nm. For apoptosis assay, cells were seeded at a density of $1x10^4$ cells per well in 100 μl of

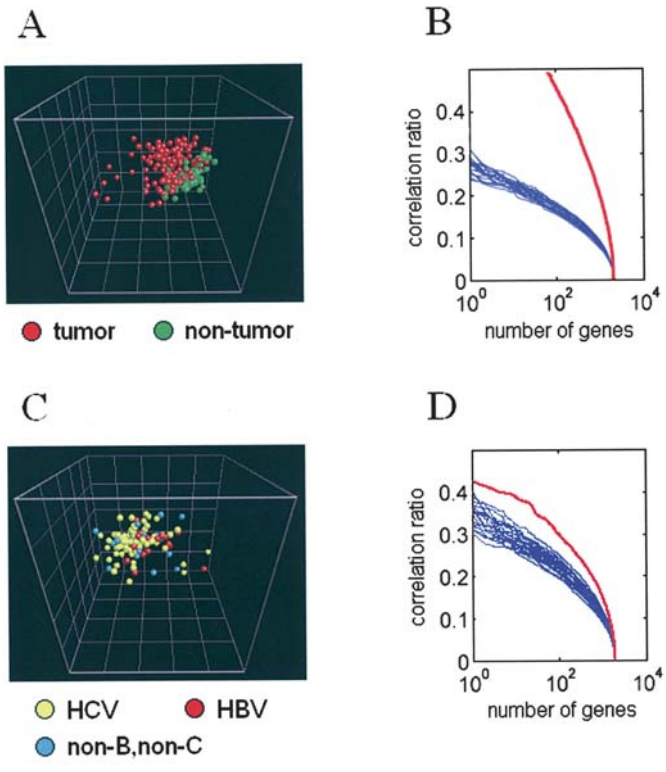

Figure 1. Expression profiling analyses of 204 liver samples (A and B, tumor vs. non-tumor) and 119 HCC samples (C and D, HCV group vs. HBV group vs. non-B, non-C group). (A and C) Principal component analyses. (B and D) Correlation spectrum analyses. Vertical axis represents the correlation ratio of the differences between tumor and non-tumor tissues (B) and between the three HCC groups (D). Horizontal axis, genes sorted by correlation ratios. Red, original data; blue, twenty trials of randomized data.

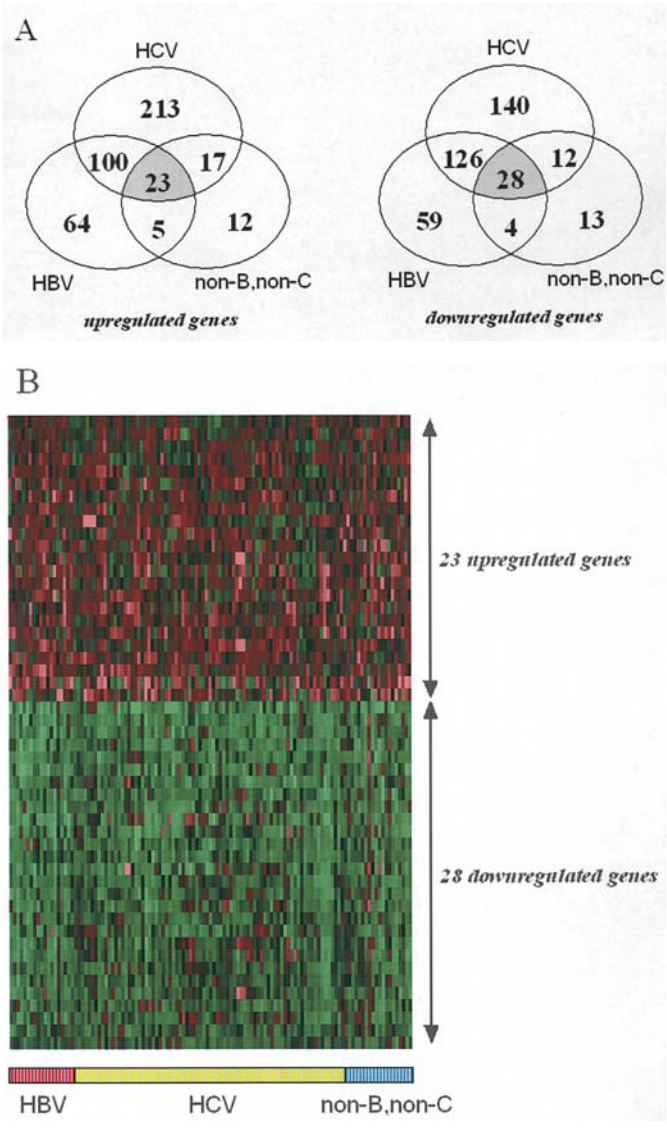

Figure 2. (A) The number of differentially expressed genes between tumor and non-tumor tissues with division of three HCC groups. Left, upregulated genes; right, downregulated genes. (B) Expression pattern of 51 genes commonly appearing in all three HCC groups.

culture medium on Atelocollagen-based cell transfection array, and, after three days, docetaxel, an anticancer agent, was added at a concentration of 10 nM for induction of apoptosis. For detection of apoptotic cells, Caspase-3/7 assay was performed using an Apo-ONE™ Homogeneous Caspase-3/7 assay kit (Promega). Condensation and fragmentation of cell nuclei was also evaluated by fluorescence microscopy following Hoechst DNA staining.

Results

Expression features of HCC. After measuring the expression levels of 3,072 genes in 204 liver samples, we selected 1,812 genes with few missing values for further analysis. We first applied PCA, a statistical method for reducing the number of data dimensions, to present the relationship between tumor (T) and non-tumor (NT) tissues. Upon displaying the expression patterns of all 1,812 genes in three-dimensional space, we observed that T and NT were located separately, indicating distinct gene expression patterns (Fig. 1A). To statistically evaluate the correlation between the gene expression patterns of these two groups, we performed a correlation spectrum analysis. The correlation ratio (*CR*) is a statistical indicator of correlation of clinicopathological parameters. The *CR* value of the original dataset was much higher than that of any permutated dataset for any number of genes (Fig. 1B). Second, to estimate the differences in expression profiles among the 119 HCC samples from the three hepatitis groups (HCV associated, HBV associated and non-B, non-C associated), PCA and correlation spectrum analyses were performed using T:NT ratios. Although the expression difference between hepatitis virus groups was not clear after PCA (Fig. 1C), the *CR* curve of the original dataset was consistently higher than that of any permutated dataset (Fig. 1D). Thus, not only are gene expression patterns different between tumor and non-tumor liver tissues, but tumors arising in patients infected with HCV, HBV, or neither virus exhibit expression differences as well.

Genes differentially expressed between tumor and non-tumor tissues. To select genes differentially expressed between T and NT tissues, we performed random permutation testing. Because expression pattern heterogeneity was confirmed between the three hepatitis virus groups, permutation tests were performed after separating each HCC group. Using $p<0.01$ as a measure

Table I. Differentially expressed genes common to all types of HCC.

Up/down[a]	UniGene ID	Symbol	Gene name
Up	Hs.497674	*FAM34A*	Family with sequence similarity 34, member A (FAM34A)
Up	Hs.465224	*NARS*	Asparaginyl-tRNA synthetase
Up	Hs.333579	*HSPC152*	Hypothetical protein HSPC152
Up	Hs.308709	*GRP58*	Glucose regulated protein, 58 kDa
Up	Hs.511984	*GBA*	Glucosidase, ß
Up	Hs.326371	*PVRL2*	Poliovirus receptor-related 2
Up	Hs.433427	*RPS17*	Ribosomal protein S17
Up	Hs.374553	*LAMR1*	Laminin receptor 1
Up	Hs.304694	*GNB1*	Guanine nucleotide binding protein, ß polypeptide 1
Up	Hs.502756	*AHNAK*	AHNAK nucleoprotein (desmoyokin)
Up	Hs.405913	*GRCC10*	Likely ortholog of mouse gene rich cluster, C10 gene
Up	Hs.387905	*SPTAN1*	Spectrin, α, non-erythrocytic 1
Up	Hs.128065	*CTSC*	Cathepsin C
Up	Hs.284295	*NSE1*	Non-SMC (structural maintenance of chromosomes) element 1 protein
Up	Hs.523744	*COP1*	Constitutive photomorphogenic protein
Up	Hs.318567	*NDRG1*	N-myc downstream regulated gene 1
Up	Hs.310769	*HSPA5*	Heat shock 70 kDa protein 5
Up	Hs.111779	*SPARC*	Secreted protein, acidic, cysteine-rich
Up	Hs.521056	*ATP5J2*	ATP synthase, H^+ transporting, mitochondrial F0 complex, subunit f, isoform 2
Up	Hs.532399	*KIAA0663*	KIAA0663 gene product
Up	Hs.375108	*CD24*	CD24 antigen (small cell lung carcinoma cluster 4 antigen)
Up	Hs.188882	*NUDT3*	Nudix (nucleoside diphosphate linked moiety X)-type motif 3
Up	Hs.155560	*CANX*	Calnexin
Down	Hs.8109	*SMYD3*	SET and MYND domain containing 3
Down	Hs.380135	*FABP1*	Fatty acid binding protein 1, liver
Down	Hs.282557	*CP*	Ceruloplasmin (ferroxidase)
Down	Hs.74561	*A2M*	α-2-macroglobulin
Down	Hs.440409	*IGFBP3*	Insulin-like growth factor binding protein 3
Down	Hs.405156	*PPAP2B*	Phosphatidic acid phosphatase type 2B
Down	Hs.293636	*SHMT1*	Serine hydroxymethyltransferase 1
Down	Hs.82101	*PHLDA1*	Pleckstrin homology-like domain, family A, member 1
Down	Hs.93194	*APOA1*	Apolipoprotein A-I
Down	Hs.405946	*SOCS2*	Suppressor of cytokine signaling 2
Down	Hs.292477	*ETS2*	V-ets erythroblastosis virus E26 oncogene homolog 2
Down	Hs.418241	*MT2A*	Metallothionein 2A
Down	Hs.387367	*CYP39A1*	Cytochrome P450, family 39, subfamily A, polypeptide 1
Down	Hs.8821	*HAMP*	Hepcidin antimicrobial peptide
Down	Hs.418127	*CYP2C9*	Cytochrome P450, family 2, subfamily C, polypeptide 9
Down	Hs.117367	*SLC22A1*	Solute carrier family 22, member 1
Down	Hs.143436	*PLG*	Plasminogen
Down	Hs.81131	*GAMT*	Guanidinoacetate N-methyltransferase
Down	Hs.75431	*FGG*	Fibrinogen, γ polypeptide
Down	Hs.170953	-	EST
Down	Hs.356123	*KRT8*	Keratin 8
Down	Hs.458414	*IFITM1*	Interferon induced transmembrane protein 1
Down	Hs.524224	*C1R*	Complement component 1, r subcomponent
Down	Hs.300774	*FGB*	Fibrinogen, B ß polypeptide
Down	Hs.351593	*FGA*	Fibrinogen, A α polypeptide
Down	Hs.418497	*GC*	Group-specific component (vitamin D binding protein)
Down	Hs.534293	*SERPINA3*	Serine proteinase inhibitor, clade A, member 3
Down	Hs.435610	*WAC*	WW domain-containing adapter with a coiled-coil region

[a]Up- or down-regulation are defined as expression in HCC tissue compared to non-tumor tissue.

Table II. Gene Ontology categories enriched in biological process (A) and cellular component (B) for 51 common genes.

A. Biological process Gene Ontology term	p-value[a]	% FDR[b]	Genes annotated to the term
Copper ion homeostasis	0.0043	0.00	*CP, MT2A*
Blood coagulation	0.0050	0.00	*FGG, PLG, FGB, FGA*
Regulation of blood pressure	0.0062	0.00	*FGG, FGB, FGA*
Hemostasis	0.0067	0.00	*FGG, PLG, FGB, FGA*
Circulation	0.0087	0.00	*FGG, APOA1, FGB, FGA*
Regulation of body fluids	0.0094	0.00	*FGG, PLG, FGB, FGA*
Organismal physiological process	0.0152	0.25	*CTSC, PLG, HAMP, SERPINA3, CYP39A1, IFITM1, CD24, FGG, APOA1, FGB, SPARC, FGA, C1R*
Regulation of cell proliferation	0.0232	0.22	*FGG, PLG, FGB, FGA, IFITM1*
Lipid metabolism	0.0315	0.40	*FAM34A, SERPINA3, APOA1, PPAP2B, CYP39A1, FABP1, GBA*
Regulation of cellular process	0.0340	0.36	*FGG, PLG, NUDT3, SOCS2, FGB, IGFBP3, FGA, IFITM1*
Cell growth and/or maintenance	0.0352	0.33	*PLG, NUDT3, GC, SOCS2, SLC22A1, CP, PPAP2B, KRT8, IFITM1, FABP1, MT2A, FGG, APOA1, GRP58, FGB, A2M, IGFBP3, FGA, ETS2, GBA*
Positive regulation of cell differentiation	0.0388	0.31	*SOCS2, IGFBP3*

B. Cellular component Gene Ontology term	p-value[a]	% FDR[b]	Genes annotated to the term
Fibrinogen complex	0.0000	0.00	*FGG, FGB, FGA*
Extracellular region	0.0003	0.00	*HAMP, GC, SERPINA3, CP, FGG, APOA1, A2M, FGB, SPARC, IGFBP3, FGA, C1R*
Extracellular space	0.0089	0.00	*FGG, GC, FGB, CP, FGA*
Cytosolic small ribosomal subunit	0.0107	0.00	*RPS17, LAMR1*
Eukaryotic 48S initiation complex	0.0107	0.00	*RPS17, LAMR1*
Eukaryotic 43S preinitiation complex	0.0440	0.33	*RPS17, LAMR1*

[a]Corrected p-value and [b]false discovery rate (FDR) were calculated by a Generic Gene Ontology Term Finder program.

of significance, we selected 659, 409 and 114 significant genes in the HCV group, the HBV group and the non-B, non-C group, respectively. Considering the multiplicity of gene selection, we also evaluated the differentially expressed genes by FDR. The FDRs in the three HCC groups were 1.0%, 1.9% and 12.1%, respectively. After three random permutation tests, 23 upregulated and 28 downregulated genes were identified as common to all three groups (Fig. 2A, Table I).

Genes commonly identified among the three HCC groups. The expression patterns of 51 genes were similarly changed among the three HCC groups (Fig. 2B). The Gene Ontology annotations for these common genes were analyzed with the Generic GO Term Finder program with a p-value threshold of 0.05 (Table II). Twelve terms in the biological processes field and six terms in the cellular components field revealed significant p-values and low FDRs of <1%. As expected, many genes involved in cell proliferation/growth or cell differentiation were identified. Furthermore, the 51 common genes were also enriched for cellular components of the extracellular region/space. The identified genes are potential candidates for novel tumor markers or drug targets of HCC.

HepG2 in vitro assays for growth and apoptosis. We performed *in vitro* assays for growth and apoptosis to better understand the role of some of the commonly identified genes in HCC. Eight genes (*CANX, FAM34A, PVRL2, ATP5J2, LAMR1, GBA, AHNAK*, and *RPS17*) were selected from the 23 common upregulated genes for further analysis. These genes were cloned into expression vectors for use in an Atelocollagen-based cell transfection array of HepG2 cells. Although growth assays were performed for these eight genes,

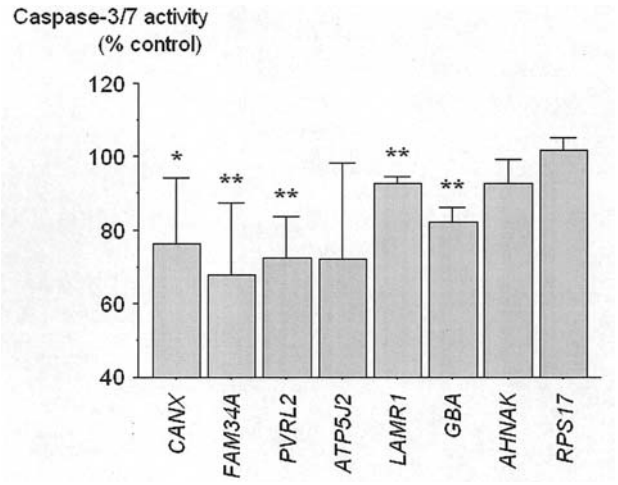

Figure 3. Caspase-3/7 activities of HepG2 cells measured using an Apo-ONE™ Homogeneous Caspase-3/7 assay kit. Vertical axis exhibits the percentage of Caspase-3/7 activity compared with control vector transfected cells. Data are expressed as the mean ± SD. Statistical difference in activity between each transfected cell and control vector cell was estimated by t-test. $^{*}p<0.05$, $^{**}p<0.01$.

no gene induced a significant change in cell proliferation (data not shown). To study the effects of these genes on apoptosis, we carried out Caspase-3/7 assays on transfected cells. Five (*CANX, FAM34A, PVRL2, LAMR1*, and *GBA*) of eight genes significantly inhibited apoptosis in transfected tumor cells (Fig. 3). We validated these results using a Hoechst staining assay, an independent confirmation of apoptosis. In all five genes, the number of cells with apoptotic nuclear fragmentation was clearly reduced in transfected cells compared with control vector transfected cells (Fig. 4). Thus, expression of these genes may play an important role in the inhibition of apoptosis in hepatoma cells.

Discussion

HCC typically develops following chronic liver inflammation caused by hepatitis viruses such as HCV and HBV (27). Some individuals who develop HCC are neither infected with HCV nor HBV, however, and these non-B, non-C cases can arise from various pathogenesis such as aflatoxin B1, alcoholic hepatitis, or nonalcoholic steatohepatitis (NASH) (28-30). Although structural alterations in many cancer-related genes are found in HCC (31), the high number of genes involved suggests that different etiological factors may affect different gene subsets within hepatocytes. Thus, distinct, but related, genetic pathways may be altered during hepatocarcinogenesis, possibly due to different initiators and promoters. Multiple studies linking hepatitis viruses and chemical carcinogens to hepatocarcinogenesis have provided clues for understanding this molecular system (32,33), but it is still unclear whether a single subset of genes is commonly modulated across all subsets of HCC. Genetic alterations common to all types of HCC would help elucidate a molecular mechanism underlying all human hepatocarcinogenesis independent of pathogenetic pathways. We wished to identify the genetic changes common to any types of HCC, which may lead to the discovery of novel diagnostic markers and therapeutic targets.

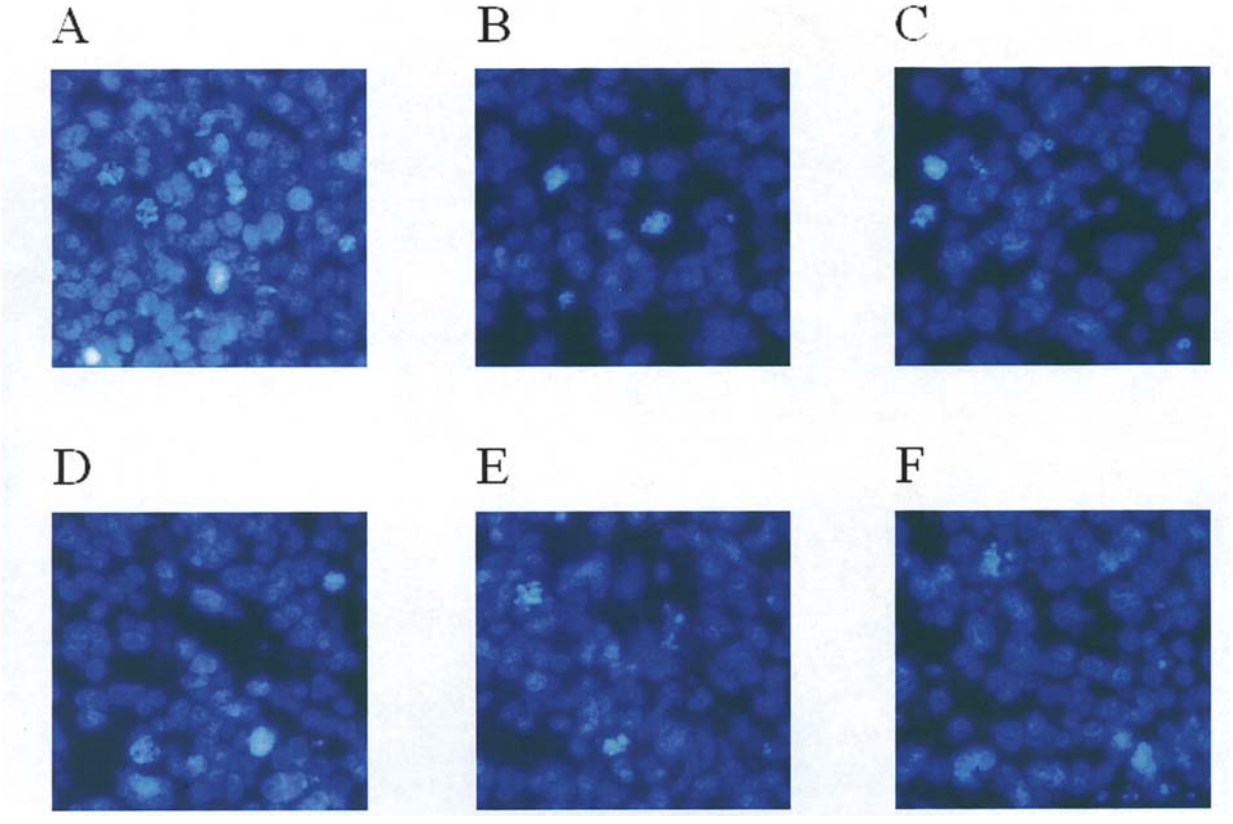

Figure 4. Hoechst staining of HepG2 cells. (A) Control vector transfected cells, (B) *CANX* transfected cells, (C) *FAM34A* transfected cells, (D) *PVRL2* transfected cells, (E) *LAMR1* transfected cells, and (F) *GBA* transfected cells.

Gene expression profiling is a powerful molecular technique wherein the expression levels of an entire mRNA population of a tissue are characterized. It allows for the increased understanding of global molecular abnormalities in various malignancies, and HCC has been previously studied using this approach (5-7). Previous studies identified differences in gene expression profiles underlying different causal viruses; however, few studies examined common gene expression profiles between HBV and HCV. Here, we performed gene expression profiling analysis with more than 200 liver samples using a PCR-based array system instead of DNA microarrays. Although DNA microarrays have contributed to the literature of gene expression profiling, they detect only a fraction of the changes in gene expression which can be detected by RT-PCR (34). We have therefore prepared a PCR-based technology which is a high throughput quantitative PCR method based on ATAC-PCR (11). The aforementioned benefits and the strength of this system for cancer research, established in previous work on not only HCC but also other cancers, makes this technique a powerful method of obtaining a better understanding of the molecular characteristics of cancer (35-37).

Two advanced statistical methods revealed distinct gene expression patterns between tumor and non-tumor tissues. We performed random permutation tests to select the differentially expressed genes between HCC and non-tumor samples by isolating causative hepatitis virus groups, because there was heterogeneity between the HCV, HBV, and non-B, non-C groups. We selected 659, 409 and 114 significant genes in the HCV, HBV and non-B, non-C groups, respectively. From these genes, we obtained 51 genes that were common to these three types of HCC. Comparing these common genes with the previous report (38), over half of the 51 differentially expressed genes identified in this study were also identified among the 1,648 differentially expressed genes.

Among these 51 genes, several cancer-related genes previously reported to be associated with HCC were identified. For example, a well known HCC-related gene, *insulin-like growth factor binding protein 3* (*IGFBP3*), was included. *IGFBP3* plays a key role in regulating cell proliferation and apoptosis, and it also plays an important inhibitory role in the development and/or growth of HCC (39,40). *GRP58* is a chaperone in the endoplasmic reticulum lumen, which was previously identified as one of several accessory proteins in the S100 cytosol fraction of human hepatoma Hep3B cells (41). *GRP58* was also reported to be associated with hepatocellular carcinoma by a cDNA library analysis using human liver cancer vascular endothelial cells (42). The overexpression of *CD24*, a sialoglycoprotein anchored to the cell surface by a glycosyl phosphatidylinositol linkage, and *SPARC*, a glycoprotein involved in extracellular matrix remodeling, were observed with other previous studies using human HCC (43,44). In contrast, *α-2-macroglobulin* (*A2M*), a protease inhibitor and cytokine transporter, or *plasminogen* (*PLG*), the zymogen in circulating blood from which plasmin is formed, are significantly underexpressed in human HCC and hepatoma cells (45-47). *Apolipoprotein A-I* (*APOA1*), the major apoprotein of high density lipoprotein (HDL) in the plasma, is a candidate for an HCC biomarker, because its expression level is significantly lower in serum samples from HCC patients compared to healthy controls (48). Furthermore, using the Gene Ontology program, genes with functions related to cell proliferation were found to be significantly enriched in the list of 51 common genes. Such genes are candidates for novel diagnostic markers because most traditional protein markers are associated with cell proliferation (49). We also identified several genes that encode extracellular proteins, which have a greater likelihood of acting as serological markers.

Additionally, we performed *in vitro* assays for growth and apoptosis to analyze the functions of some selected genes. Although no genes significantly altered cell growth, five of the eight examined genes inhibited apoptosis in two independent *in vitro* assays. Among these five genes, we identified *Glucosylceramidase* (*GBA*), a gene encoding a lysosomal membrane protein that cleaves the ß-glucosidic linkage of glycosylceramide, an intermediate in glycolipid metabolism. Although many studies have reported that mutations in this gene cause Gaucher disease, a lysosomal storage disease characterized by the accumulation of glucocerebrosides (50), the correlation between the *GBA* gene and HCC has not yet been identified. This gene also inhibited apoptosis in hepatoma cells. In fact, although an increased risk of malignancy in patients with Gaucher disease has been observed, the coexistence of Gaucher disease and hepatocellular carcinoma is extremely rare (51). In patients with Gaucher disease, the *GBA* mutation may induce not only abnormal storage of glucocerebrosides but also inhibition of hepatocarcinogenesis through the promotion of apoptosis. The regulation of apoptosis in hepatoma cells is poorly understood, and further studies are needed to identify the role of *GBA* and other genes in this process.

We used over 200 separate liver samples in this large scale gene expression profiling study to generate several important findings. A total of 51 genes were identified as differentially expressed between tumor and non-tumor tissues regardless of the etiology of HCC. It is thought that these genes may play significant roles in the development of cancer independent of hepatitis viruses. Such genes are potential targets for the rational development of new cancer drugs and for the early detection of all types of HCC.

Acknowledgements

This work was supported by Grant-in-Aid for the Development of Innovative Technology from the Ministry of Education, Culture, Sports, Science and Technology, Japan. The authors thank Ms. Noriko Ueno, Ms. Chiyuri Maruyama, Ms. Keiko Ikenaga, Ms. Mihoko Yoshino and Ms. Satoko Kinjo for their expert technical assistance.

References

1. Montalto G, Cervello M, Giannitrapani L, Dantona F, Terranova A and Castagnetta LA: Epidemiology, risk factors, and natural history of hepatocellular carcinoma. Ann NY Acad Sci 963: 13-20, 2002.
2. Poon RT, Fan ST, Lo CM, Ng IO, Liu CL, Lam CM and Wong J: Improving survival results after resection of hepatocellular carcinoma: a prospective study of 377 patients over 10 years. Ann Surg 234: 63-70, 2001.
3. Nguyen MH and Keeffe EB: Screening for hepatocellular carcinoma. J Clin Gastroenterol 35: S86-S91, 2002.

4. Thorgeirsson SS and Grisham JW: Molecular pathogenesis of human hepatocellular carcinoma. Nat Genet 31: 339-346, 2002.
5. Okabe H, Satoh S, Kato T, Kitahara O, Yanagawa R, Yamaoka Y, Tsunoda T, Furukawa Y and Nakamura Y: 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: 2129-2137, 2001.
6. Delpuech O, Trabut JB, Carnot F, Feuillard J, Brechot C and Kremsdorf D: Identification, using cDNA macroarray analysis, of distinct gene expression profiles associated with pathological and virological features of hepatocellular carcinoma. Oncogene 21: 2926-2937, 2002.
7. Iizuka N, Oka M, Yamada-Okabe H, *et al*: 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: 3939-3944, 2002.
8. Kurokawa Y, Matoba R, Takemasa I, *et al*: Molecular features of non-B, non-C hepatocellular carcinoma: a PCR-array gene expression profiling study. J Hepatol 39: 1004-1012, 2003.
9. Smith MW, Yue ZN, Geiss GK, Sadovnikova NY, Carter VS, Boix L, Lazaro CA, Rosenberg GB, Bumgarner RE, Fausto N, Bruix J and Katze MG: Identification of novel tumor markers in hepatitis C virus-associated hepatocellular carcinoma. Cancer Res 63: 859-864, 2003.
10. Petricoin EF III, Hackett JL, Lesko LJ, Puri RK, Gutman SI, Chumakov K, Woodcock J, Feigal DW Jr, Zoon KC and Sistare FD: Medical applications of microarray technologies: a regulatory science perspective. Nat Genet 32: 474-479, 2002.
11. Kato K: Adaptor-tagged competitive PCR: a novel method for measuring relative gene expression. Nucleic Acids Res 25: 4694-4696, 1997.
12. Ochiya T, Takahama Y, Nagahara S, Sumita Y, Hisada A, Itoh H, Nagai Y and Terada M: New delivery system for plasmid DNA *in vivo* using atelocollagen as a carrier material: the Minipellet. Nat Med 5: 707-710, 1999.
13. Honma K, Ochiya T, Nagahara S, Sano A, Yamamoto H, Hirai K, Aso Y and Terada M: Atelocollagen-based gene transfer in cells allows high-throughput screening of gene functions. Biochem Biophys Res Commun 289: 1075-1081, 2001.
14. Honma K, Miyata T and Ochiya T: The role of atelocollagen-based cell transfection array in high-throughput screening of gene functions and in drug discovery. Curr Drug Discovery Technologies 1: 287-294, 2004.
15. Yotsuyanagi H, Shintani Y, Moriya K, Fujie H, Tsutsumi T, Kato T, Nishioka K, Takayama T, Makuuchi M, Iino S, Kimura S and Koike K: Virologic analysis of non-B, non-C hepatocellular carcinoma in Japan: frequent involvement of hepatitis B virus. J Infect Dis 181: 1920-1928, 2000.
16. Schneeberger PM, Keur I, van der Vliet W, van Hoek K, Boswijk H, van Loon AM, van Dijk WC, Kauffmann RH, Quint W and van Doorn LJ: Hepatitis C virus infections in dialysis centers in the Netherlands: a national survey by serological and molecular methods. J Clin Microbiol 36: 1711-1715, 1998.
17. Matoba R, Kato K, Saito S, Kurooka C, Maruyama C, Sakakibara Y and Matsubara K: Gene expression in mouse cerebellum during its development. Gene 241: 125-131, 2000.
18. Matoba R, Kato K, Kurooka C, Maruyama C, Sakakibara Y and Matsubara K: Correlation between gene functions and developmental expression patterns in the mouse cerebellum. Eur J Neurosci 12: 1357-1371, 2000.
19. Kurokawa Y, Matoba R, Nakamori S, Takemasa I, Nagano H, Dono K, Umeshita K, Sakon M, Monden M and Kato K: PCR-array gene expression profiling of hepatocellular carcinoma. J Exp Clin Cancer Res 23: 135-141, 2004.
20. Kurokawa Y, Matoba R, Takemasa I, *et al*: Molecular-based prediction of early recurrence in hepatocellular carcinoma. J Hepatol 41: 284-291, 2004.
21. Muro S, Takemasa I, Oba S, *et al*: Identification of expressed genes linked to malignancy of human colorectal carcinoma by parametric clustering of quantitative expression data. Genome Biol 4: R21, 2003.
22. Golub TR, Slonim DK, Tamayo P, Huard C, Gaasenbeek M, Mesirov JP, Coller H, Loh ML, Downing JR, Caligiuri MA, Bloomfield CD and Lander ES: Molecular classification of cancer: class discovery and class prediction by gene expression monitoring. Science 286: 531-537, 1999.
23. Pomeroy SL, Tamayo P, Gaasenbeek M, *et al*: Prediction of central nervous system embryonal tumour outcome based on gene expression. Nature 415: 436-442, 2002.
24. Storey JD and Tibshirani R: Statistical significance for genome-wide studies. Proc Natl Acad Sci USA 100: 9440-9445, 2003.
25. Boyle EI, Weng S, Gollub J, Jin H, Botstein D, Cherry JM and Sherlock G: GO::TermFinder - open source software for accessing Gene Ontology information and finding significantly enriched Gene Ontology terms associated with a list of genes. Bioinformatics 20: 3710-3715, 2004.
26. Patil MA, Chua MS, Pan KH, Lin R, Lih CJ, Cheung ST, Ho C, Li R, Fan ST, Cohen SN, Chen X and So S: An integrated data analysis approach to characterize genes highly expressed in hepatocellular carcinoma. Oncogene 24: 3737-3747, 2005.
27. Kasai Y, Takeda S and Takagi H: Pathogenesis of hepatocellular carcinoma: a review from the viewpoint of molecular analysis. Semin Surg Oncol 12: 155-159, 1996.
28. Wogan GN: Aflatoxins as risk factors for hepatocellular carcinoma in humans. Cancer Res 52: 2114s-2118s, 1992.
29. Mohamed AE, Kew MC and Groeneveld HT: Alcohol consumption as a risk factor for hepatocellular carcinoma in urban Southern African blacks. Int J Cancer 51: 537-541, 1992.
30. Bugianesi E, Leone N, Vanni E, Marchesini G, Brunello F, Carucci P, Musso A, De Paolis P, Capussotti L, Salizzoni M and Rizzetto M: Expanding the natural history of nonalcoholic steatohepatitis: from cryptogenic cirrhosis to hepatocellular carcinoma. Gastroenterology 123: 134-140, 2002.
31. Murakami Y, Hayashi K, Hirohashi S and Sekiya T: Aberrations of the tumor suppressor p53 and retinoblastoma genes in human hepatocellular carcinomas. Cancer Res 51: 5520-5525, 1991.
32. Andrisani OM and Barnabas S: The transcriptional function of the hepatitis B virus X protein and its role in hepatocarcinogenesis (Review). Int J Oncol 15: 373-379, 1999.
33. Bressac B, Kew M, Wands J and Ozturk M: Selective G to T mutations of p53 gene in hepatocellular carcinoma from Southern Africa. Nature 350: 429-431, 1991.
34. Holland MJ: Transcript abundance in yeast varies over six orders of magnitude. J Biol Chem 277: 14363-14366, 2002.
35. Iwao K, Matoba R, Ueno N, Ando A, Miyoshi Y, Matsubara K, Noguchi S and Kato K: Molecular classification of primary breast tumors possessing distinct prognostic properties. Hum Mol Genet 11: 199-206, 2002.
36. Kurokawa Y, Matoba R, Nagano H, Sakon M, Takemasa I, Nakamori S, Dono K, Umeshita K, Ueno N, Ishii S, Kato K and Monden M: Molecular prediction of response to 5-fluorouracil and interferon-alpha combination chemotherapy in advanced hepatocellular carcinoma. Clin Cancer Res 10: 6029-6038, 2004.
37. Iwao-Koizumi K, Matoba R, Ueno N, Kim SJ, Ando A, Miyoshi Y, Maeda E, Noguchi S and Kato K: Prediction of docetaxel response in human breast cancer by gene expression profiling. J Clin Oncol 23: 422-431, 2005.
38. Chen X, Cheung ST, So S, *et al*: Gene expression patterns in human liver cancers. Mol Biol Cell 13: 1929-1939, 2002.
39. Huynh H, Chow PK, Ooi LL and Soo KC: A possible role for insulin-like growth factor-binding protein-3 autocrine/paracrine loops in controlling hepatocellular carcinoma cell proliferation. Cell Growth Differ 13: 115-122, 2002.
40. Hanafusa T, Yumoto Y, Nouso K, *et al*: Reduced expression of insulin-like growth factor binding protein-3 and its promoter hypermethylation in human hepatocellular carcinoma. Cancer Lett 176: 149-158, 2002.
41. Ndubuisi MI, Guo GG, Fried VA, Etlinger JD and Sehgal PB: Cellular physiology of STAT3: where's the cytoplasmic monomer? J Biol Chem 274: 25499-25509, 1999.
42. Zhong X, Ran YL, Lou JN, Hu D, Yu L, Zhang YS, Zhou Z and Yang ZH: Construction of human liver cancer vascular endothelium cDNA expression library and screening of the endothelium-associated antigen genes. World J Gastroenterol 10: 1402-1408, 2004.
43. Huang LR and Hsu HC: Cloning and expression of CD24 gene in human hepatocellular carcinoma: a potential early tumor marker gene correlates with p53 mutation and tumor differentiation. Cancer Res 55: 4717-4721, 1995.
44. Le Bail B, Faouzi S, Boussarie L, Guirouilh J, Blanc JF, Carles J, Bioulac-Sage P, Balabaud C and Rosenbaum J: Osteonectin/SPARC is overexpressed in human hepatocellular carcinoma. J Pathol 189: 46-52, 1999.

45. Paradis V, Bieche I, Dargere D, Laurendeau I, Laurent C, Bioulac Sage P, Degott C, Belghiti J, Vidaud M and Bedossa P: Molecular profiling of hepatocellular carcinomas (HCC) using a large-scale real-time RT-PCR approach: determination of a molecular diagnostic index. Am J Pathol 163: 733-741, 2003.
46. Baffet G, Braciak TA, Fletcher RG, Gauldie J, Fey GH and Northemann W: Autocrine activity of interleukin 6 secreted by hepatocarcinoma cell lines. Mol Biol Med 8: 141-156, 1991.
47. Steel LF, Shumpert D, Trotter M, Seeholzer SH, Evans AA, London WT, Dwek R and Block TM: A strategy for the comparative analysis of serum proteomes for the discovery of biomarkers for hepatocellular carcinoma. Proteomics 3: 601-609, 2003.
48. Kinoshita M and Miyata M: Underexpression of mRNA in human hepatocellular carcinoma focusing on eight loci. Hepatology 36: 433-438, 2002.
49. Hui AM, Makuuchi M and Li X: Cell cycle regulators and human hepatocarcinogenesis. Hepatogastroenterology 45: 1635-1642, 1998.
50. Tsuji S, Choudary PV, Martin BM, Stubblefield BK, Mayor JA, Barranger JA and Ginns EI: A mutation in the human glucocerebrosidase gene in neuronopathic Gaucher's disease. N Engl J Med 316: 570-575, 1987.
51. Erjavec Z, Hollak CEM and Vries EGE: Hepatocellular carcinoma in a patient with Gaucher disease on enzyme supplementation therapy. Ann Oncol 10: 243, 1999.