

Transcriptome analysis of hepatitis B virus-associated small hepatocellular carcinoma by serial analysis of gene expression

JIN-WOOK KIM¹, SANG HYUB LEE¹, YOUNG SOO PARK¹, SOOK-HYANG JEONG¹,
NAYOUNG KIM¹, DONG HO LEE¹ and HYO-SUK LEE²

¹Department of Medicine, Seoul National University Bundang Hospital, Seongnam;

²Department of Internal Medicine, Seoul National University College of Medicine, Seoul, Korea

Received February 12, 2009; Accepted April 20, 2009

DOI: 10.3892/ijo_00000321

Abstract. Serial analysis of gene expression (SAGE) is a sequence-based tool that enables quantitative study of comprehensive gene expression profiles. In this study, we generated a SAGE library of hepatitis B virus (HBV)-associated small hepatocellular carcinoma (HCC) and compared it with libraries of normal liver and other organs in order to identify genes which are specifically expressed in HBV-associated small HCC. A total of 18,107 SAGE tags were obtained from the HCC SAGE library. Among the 7,405 unique tags, 240 were significantly overexpressed in HCC compared to normal liver. Seventeen genes were unequivocally matched to SAGE tags which were up-regulated 15-fold or higher in HCC compared to normal liver: four genes (fatty acid desaturase 2, α -L-2-fucosidase, testis-specific protein Y-encoded-like 2 and Gon-4-like) were significantly overexpressed in the HCC SAGE library compared to any other SAGE libraries of human normal tissues. The significance of these genes with respect to carcinogenesis and early diagnosis of HCC needs to be elucidated in further studies.

Introduction

Hepatocellular carcinoma (HCC) is the third most common cause of death from cancer worldwide, and has a very poor prognosis (1). Curative treatment modalities, such as surgical resection or percutaneous ablation, are precluded in many patients with HCC due to advanced tumor stage at the time of presentation. Early diagnosis of HCC by surveillance program can reduce cancer-related mortality in patients

infected with hepatitis B virus (HBV) (2), a major risk factor for HCC. Therefore, diagnosis of small HCC is crucial in order to improve overall prognosis of HCC. α -fetoprotein (AFP) is currently the most widely used serum marker of HCC for screening purposes. However, the sensitivity of AFP is around 60% at best (3,4), so that AFP alone is not sufficient for the early detection of HCC. Other serum markers such as des- γ carboxyprothrombin and lens culinaris agglutinin-reactive AFP have been reported to be promising, but the clinical usefulness of these markers needs further validation (5). Thus, new effective diagnostic markers are urgently needed for the optimal screening of HCC.

Transcriptome analysis is a powerful tool in cancer research for the identification of novel diagnostic markers, prognostic indicators and therapeutic targets. Techniques of transcriptome analysis, represented by cDNA and oligonucleotide microarrays, have been extensively applied to HCC research with a special focus on carcinogenesis (6), molecular classification (7-10) and identification of diagnostic markers (11,12). Microarrays, however, cannot identify novel transcripts which are not included in the pre-defined set of target genes. Another limitation of microarray is that data from other researchers cannot be compared directly. Large scale sequencing of expressed sequence tags (ESTs) can detect novel transcripts, and numerical comparisons of ESTs are, in theory, possible between two unrelated cDNA libraries. Because generation of comprehensive library needs a huge number of sequencing, however, there have been only a few studies of large scale EST libraries from HCC tissues (13-15). As each library was generated from only a few patients, heterogeneity of gene expression profiles makes it difficult to interpret the variable results from these studies.

Serial analysis of gene expression (SAGE) is a sequencing-based technique which allows genome-wide analysis of gene expression in a quantitative manner without prior knowledge of the gene sequences (16). Furthermore, the results from any SAGE experiment can be directly compared with SAGE libraries of other normal and cancerous tissues deposited at the public SAGE databases (17,18).

In this study, we constructed a SAGE library from HBV-associated small HCC and compared the gene expression profile of HCC with normal liver and extrahepatic normal organs to identify specifically expressed genes in HBV-associated small HCC.

Correspondence to: Dr Jin-Wook Kim, Department of Medicine, Seoul National University Bundang Hospital, 300 Gumi-dong, Bundang-gu, Seongnam 463-802, Korea
E-mail: kimjw@snubh.org

Key words: hepatitis B virus, hepatocellular carcinoma, serial analysis of gene expression, transcriptome

Table I. Clinical features of HCC patients enrolled for construction of HCC SAGE library.

Age	Gender	Underlying liver disease	Tumor size (mm)	AFP (ng/ml)	Edmondson grade
70	M	HBV-associated liver cirrhosis	25	5	II
67	F	HBV-associated liver cirrhosis	30	13	I
70	F	HBV-associated liver cirrhosis	15	7400	II
66	M	HBV-associated liver cirrhosis	20	153	I
67	M	HBV-associated liver cirrhosis	25	25	III

Table II. Clinical features of HCC patients enrolled for real-time RT-PCR assay.

Patients (Age/Gender)	Underlying liver disease	Tumor size (mm)	AFP (ng/ml)	Edmondson grade
70/M ^a	HBV-associated liver cirrhosis	25	5	II
67/F ^a	HBV-associated liver cirrhosis	30	13	I
52/M	HBV-associated chronic hepatitis	25	4	III
42/M	HBV-associated chronic hepatitis	22	1.8	II
63/M	HBV-associated liver cirrhosis	21	256	III
56/M	HBV-associated liver cirrhosis	12	7	II

^aSame patients enrolled in Table I.

Materials and methods

Patients and tissue samples. Nine Korean patients with HBV-associated small HCC were enrolled in this study. The SAGE library was constructed by using pooled samples from five patients (Table I) and real-time RT-PCR was performed in six patients (Table II, including two patients in Table I). Informed consent was obtained from each patient according to the regulations of the Institutional Review Board of Seoul National University Bundang Hospital. All tissue samples were obtained in pairs (HCC and adjacent non-neoplastic liver tissue) by percutaneous needle biopsy during radio-frequency ablation of small HCC (<3 cm). After harvest, the tissue specimens were immediately submerged in RNAlater solution (Ambion, Austin, TX, USA), stored at 4°C overnight and then at -70°C until RNA extraction.

Construction of SAGE library. To construct the HCC SAGE library, I-SAGE kit (Invitrogen, Carlsbad, CA, USA) was used according to the manufacturer's instructions. Briefly, total RNA was obtained from each of five HCC specimens by using TRIzol (Invitrogen) according to the manufacturer's instructions. RNA (1 µg) from five samples was pooled together and mixed with oligo (dT) magnetic beads to select the poly A⁺ mRNA. First- and second-strand cDNA was synthesized using SuperScript II reverse transcriptase and *E. coli* DNA polymerase, respectively. The double-stranded cDNA was digested with *Nla*III, an anchoring enzyme which generates CATG sticking ends. The cDNA was ligated with two adaptors, cleaved with *Bsm*FI and subjected to a Klenow reaction to fill in the 5' overhangs. Two adaptor-bound tags were ligated together by T4 DNA ligase to form 100-bp ditags. These ditags were PCR-amplified, gel-purified and digested with *Nla*III to get 26-bp ditags. After gel purification, the

26-bp ditags were ligated to give concatemers, cloned into *Sph*I site of pZerO-1 plasmid. Plasmid DNA was purified using Montage plasmid miniprep96 kit (Millipore) and sequenced using M13F primer.

SAGE data analysis. SAGE library of normal liver was downloaded from SAGEmap (17), a public SAGE data resource run by National Center for Biotechnology Information (NCBI) (ftp://ftp.ncbi.nih.gov/pub/sage/obsolete/seq/SAGE_normal_liver.zip) for comparison with HCC SAGE library constructed in this study. SAGE libraries of normal tissues other than liver were also downloaded from the SAGEmap for comparison. SAGE Genie, another public repository of SAGE data maintained by NCBI (18), was also used to study the expression profiles of particular genes in normal human tissues (<http://cgap.nci.nih.gov/SAGE/AnatomicViewer>). SAGE tags were extracted from the SAGE library sequence files containing SAGE ditag concatemers by using eSAGE software (ver 1.2) (19). The reference database for tag-to-gene matching was downloaded from SAGEmap (ftp://ftp.ncbi.nih.gov/pub/sage/mappings/SAGEmap_Hs_NlaIII_10_best.gz, updated on September 2, 2008). The result of tag-to-gene mapping was validated by using the SAGE Genie web-base tool (<http://cgap.nci.nih.gov/SAGE/AnatomicViewer>). If there were more than one gene that matched a unique tag, a gene with tag matching closest to the 3' end was selected (20). Tags that were not found more than once in any SAGE library and did not map to the UniGene database were filtered out as sequencing errors.

For each tag, the statistical significance of difference in tag frequency between two SAGE libraries was analyzed by the method developed by Audic and Claverie (21). A web based significance test by Poisson statistics (<http://genome.dfci.harvard.edu/sager>) was also used to validate the results.

Table III. Primers for real-time RT-PCR.

Gene	Accession no.	Forward primer	Reverse primer
GSTA1	BC110891	GCTGAAGGCCCTGAAAACC	GGCTGCCAGGCTGTAGAAAC
FUCA2	BC003060	GCCCGCGTGGTTTGAC	GCTGGGCACGGAACAACT
GPC3	BC035972	CCTTGCAGAACTGGCCTATGA	GGAGTTGCCTGCTGACTGTTT
LCN2	NM_005564	GCCTCCCTGAAAACACATC	TGCACTCAGCCGTCGATACA
LGALS4	BC003661	GCCTGCCCCACCATGGAA	TTGCAGCCTCCCCGAAATATG
18SrRNA	X03205	GCAATTATCCCCATGAACGA	CAAAGGGCAGGGACTTAATCAA

For genes which are expressed at least 15 times higher in the HCC SAGE library compared to the normal liver SAGE library, gene ontology (cellular process and localization) data was obtained from Gene Ontology Annotation Database (<http://www.ebi.ac.uk/GOA/>).

Real-time quantitative RT-PCR. To validate the results obtained from the SAGE method, the mRNA levels of selected genes which were differentially expressed between normal liver and HCC SAGE libraries were quantified by real-time RT-PCR. Six paired samples (HCC and adjacent non-neoplastic liver tissue) obtained from patients (Table II) were used for the isolation of total RNA by using TRIzol. The RNA was converted into cDNA using the Thermoscript RT system with random hexamers (Invitrogen). Primer pairs for each gene were designed by using Primer Express software (version 2.0, Applied Biosystems, Foster City, CA, USA) (Table III). Real-time PCR was performed using the Applied Biosystems Prism 7500 Sequence Detection System. The 12 μ l PCR mixture included 6 μ l SYBR-Green PCR master mix (Applied Biosystems), primer pair (900 nM) and 100 ng template. Semi-log amplification curves were evaluated by the comparative quantification method and the gene expression levels were normalized to human 18s ribosomal RNA (22).

Expression of HCC-related genes in normal organs. For genes which are expressed at least fifteen times higher in the HCC SAGE library compared to the normal liver SAGE library, the expression in extrahepatic normal organs (brain, thyroid, lung, heart, breast, stomach, colon, pancreas, peritoneum, kidney, prostate, placenta, muscle, skin, bone marrow, lymph node, WBC, breast endothelial cells and liver sinusoidal endothelial cells) were assessed by SAGE Anatomic Viewer (<http://cgap.nci.nih.gov/SAGE/AnatomicViewer>). The tag frequencies in each SAGE library of normal organs were normalized (tag frequencies per 200,000 total tag counts) for comparison and the level of significance was analyzed as described above.

Results

Comparison of SAGE tag profiles between HCC and normal liver. Our HCC SAGE library consisted of 1,395 cloned sequences, from which 18,107 tags were extracted. As for the normal liver SAGE library, 66,588 tags were extracted from 2,122 sequences deposited at the SAGEmap. Overall, 19,968 different tags were identified from both libraries: 7,405 and 15,578 different tags from the HCC and normal liver SAGE library, respectively. Among these tags, >90%

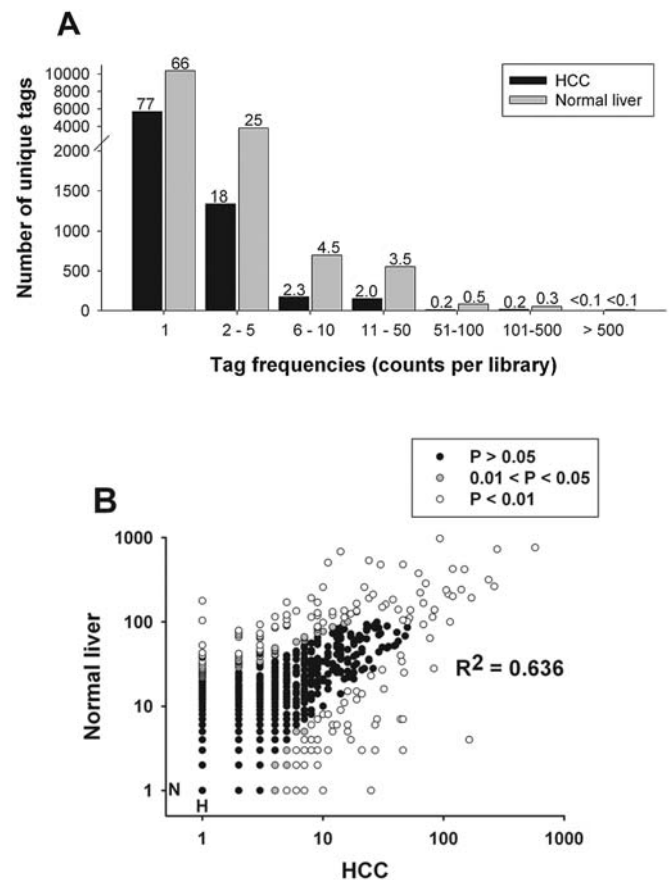


Figure 1. SAGE libraries of HCC and normal liver. (A) Distribution of SAGE tag frequencies in HCC and normal liver. The total tag number of the HCC SAGE library is 18,107 with 7,405 different tags, and that of the normal liver SAGE library is 66,588 with 15,578 different tags. Each bar represents the number of unique tags with tag frequencies specified on the horizontal axis. Figures on the top of each bar represent percentages in the corresponding library. (B) Comparison of SAGE tag frequencies between HCC and normal liver. Each circle represents a unique tag among a total of 19,968 different tags. The horizontal and the vertical axes represent frequencies of each SAGE tag in HCC and normal liver, respectively. Gray circles indicate differently expressed tags with P-values between 0.01 and 0.05 (172 tags; 97 were overexpressed and 75 were underexpressed in HCC) and white circles indicate differently expressed tags with P-values <0.01 (244 tags; 143 were overexpressed and 101 were underexpressed in HCC). H (3,990 tags) and N (9,456 tags) indicate SAGE tags that occur only once in HCC and normal liver, respectively.

occurred five times or less in each library. Tags that appeared only once comprised 77.1 and 66.5% of the HCC and normal liver SAGE library, respectively (Fig. 1A).

Plotting of the frequencies of each tag showed correlation between the HCC and normal liver SAGE libraries ($R^2 = 0.636$,

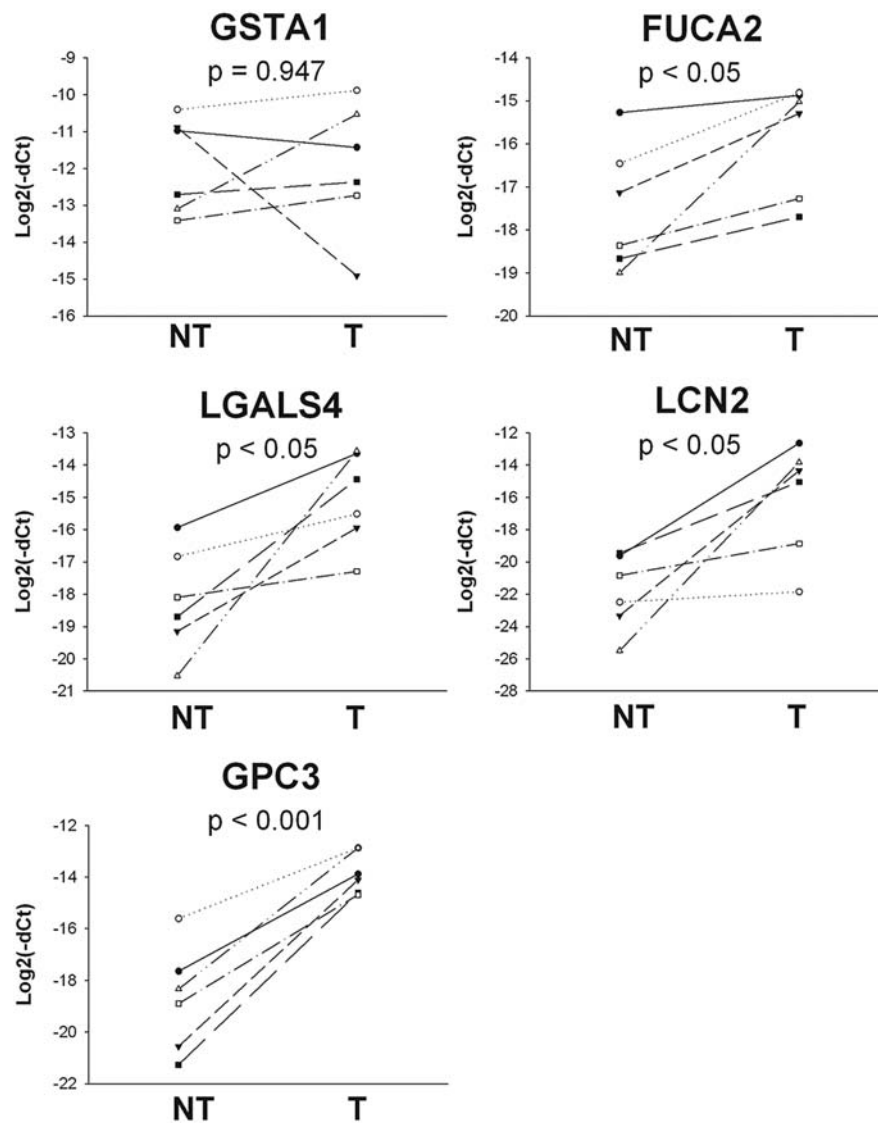


Figure 2. Validation of differentially expressed genes between HCC and chronic liver disease by real-time quantitative RT-PCR. mRNA levels of five genes and one EST were compared between HCC (T) and adjacent non-neoplastic liver tissue (NT) from six patients. Filled (\bullet) and empty (\circ) circles denote samples that were also used for HCC SAGE library construction. Filled (\blacktriangle) and empty (\triangle) triangles denote patients with chronic hepatitis B as non-neoplastic control; the other four patients showed liver cirrhosis in the background.

Fig. 1B). Among the 19,968 tags, 416 were expressed differentially between the two libraries ($P < 0.05$): 240 and 176 tags were significantly up- and down-regulated in the HCC SAGE library, respectively.

Genes up-regulated in the HCC SAGE library compared to normal liver. Among the 240 tags which were significantly up-regulated in HCC, 56 were expressed 15 times or higher in HCC than in normal liver (Table IV). Twenty tags were matched to 19 unique UniGene clusters: 17 genes (tag for IGHV3-23 matched twice) and two hypothetical proteins (LOC115110 and LOC100129060). When these matched genes were classified according to their ontology annotations (biological process), genes associated with metabolic process was the most frequent [Aldo-keto reductase family 1, member B10 (AKR1B10), glutathione S-transferase A1 (GSTA1), fucosidase, α -L-2 (FUCA2), fatty acid desaturase 2 (FADS2), lysozyme (LYZ), ST3 β -galactoside α -2,3-sialyltransferase 3 (ST3GAL3)], followed by genes affecting cell proliferation

[TSPY-like 2 (TSPYL2), Ras homolog gene family, member G (RHOG), glypican-3 (GPC3)]. Other genes were classified as transcriptional regulation [Gon-4-like (GON4L)], transport (Lipocalin-2 [LCN2]), immune response [immunoglobulin heavy constant μ (IGHV3-23), immunoglobulin λ joining 3 (IGLV3-21), immunoglobulin κ light chain VJ region (IGKV@), CD74], cytoskeleton organization [thymosin β 10 (TMSB10)] and cell adhesion [lectin, galactoside-binding, soluble, 4 (LGALS4)]. Thirty transcripts were matched to known ESTs (cDNA clones and transcribed locus) without unequivocal match to any known gene. These ESTs were from various human tissues, either benign or malignant. Three ESTs are worthy of note, because they were originally cloned from HCC (13) (AV684577, AV687841, AV687243). The remaining six tags had no match to existing SAGE tags or UniGene clusters (labeled as no match in Table IV).

Quantitative validation of HCC-specific gene expression from SAGE data (Fig. 2). The expression of five genes (GSTA1,

Table IV. Genes expressed 15 times or higher in HCC SAGE library compared with normal liver.

Tag sequence	HCC/normal liver ratio	UniGene cluster no.	Accession no.	Symbol	Title	Location	Cellular compartment
GAAATAAAGC	149.9	Hs.510635	BC006402	IGHV3-23	Immunoglobulin heavy constant mu cDNA clone	14q32.33	Membrane fraction
CGAATAAGAG	106.6		BF219475		cDNA clone		
TGAAAAGTGG	91.9		BI826151		cDNA clone		
GGGGAATCG	69.9	Hs.446574	BC016025	TMSB10	Thymosin β 10	2p11.2	Cytoskeleton
AAGGGAGCAC	56.4	Hs.449585	BC012159	IGLV3-21	Immunoglobulin lambda joining 3 cDNA clone	22q11.2	Membrane fraction
TACCTAATG	51.5		BF896098		Transcribed locus		
TTGCCCCAGG	47.8	Hs.574052	BX645969		Aldo-keto reductase family 1, member B10	7q33	Cytoplasm
GCTTGAATAA	44.1	Hs.116724	NM_020299	AKR1B10	No match		
TGTCATAGGG	44.1				cDNA clone		
GCTTTTCAA	40.5		BM682785		cDNA clone		
CGTGTGAAC	36.8		AV684577		cDNA clone		
CTCATACACC	36.8		BG461525		cDNA clone		
GACCTAATTG	36.8		AA193101		cDNA clone		
GAGACCAACCG	36.8	Hs.136164	BC024270	TSPYL2	TSPY-like 2	Xp11.2	Nucleus
TCCATCGTCC	36.8		CV328243		cDNA clone		
GAGGCCAAGA	33.1	Hs.446309	NM_145740	GSTA1	Glutathione S-transferase A1	6p12.1	Cytoplasm
GGGCCCCAGG	29.4	Hs.709907	AL359943	LOC115110	Hypothetical protein LOC115110	1p36.32	
CGGCCCCAGG	25.7	Hs.591332	NM_032020	FUCA2	Fucosidase, α -L-2	6q24	Extracellular region
GGTCTGGCCT	25.7	Hs.502745	NM_004265	FADS2	Fatty acid desaturase 2	11q12-q13.1	Endoplasmic reticulum
TAGGGAGCAC	25.7				No match		
TGACCACGGG	25.7		AA627944		cDNA clone		
TGACCCACAG	25.7		BE963226		cDNA clone		
TTGTGGTTAA	25.7		BF949459		cDNA clone		
GTTACACATTA	24.2	Hs.436568	BC024272	CD74	CD74 molecule, MHC class II invariant chain	5q32	Membrane
GGAATAAGAG	23.1	Hs.595979	H47090		Transcribed locus		
ACCATCGTCC	22.1		BE878792		cDNA clone		
AGAAGACGTT	22.1		AA421948		cDNA clone		
CAGGTGTCT	22.1		BE831108		cDNA clone		
CGTGGAGAGC	22.1		BF930209		cDNA clone		
CGTGTTGAGA	22.1				No match		
GATGCCCCCC	22.1		BQ581978		cDNA clone		
GCAGAGTAGC	22.1	Hs.656361	AK098734	GON4L	Gon-4-like	1q22	Nucleus
GCCACCGTCC	22.1		AA503888		cDNA clone		
GGCTGGTCTG	22.1		MN_052871		cDNA clone		
TGCCCTCAGG	22.1	Hs.204238	NM_005564	LCN2	Lipocalin-2	9q34	Extracellular region
TGTCCCCAGG	22.1		BQ345334		cDNA clone		
TTGTCTTTGG	22.1		AV687841		cDNA clone		
TTTGTGGTTA	22.1		BE162345		cDNA clone		
AAATAAAGCA	18.4				No match		
AGGGTCCCCG	18.4	Hs.708980	Z69026	IGKV@	Immunoglobulin κ light chain VJ region	2p12	Membrane fraction

Table IV. Continued.

Tag sequence	HCC/normal liver ratio	UniGene cluster no.	Accession no.	Symbol	Title	Location	Cellular compartment
ATTGAGCAG	18.4		BC011481		cDNA clone		
CCGGAAGGA	18.4		CF121551		cDNA clone		
CTTCTGGGA	18.4	Hs.501728	NM_001665	RHOG	Ras homolog gene family, member G	11p15.5	Plasma membrane
GAAAGCAATAA	18.4	Hs.597915	NM_006279	ST3GAL3	ST3 β -galactoside α -2,3-sialyltransferase 3	1p34.1	Golgi apparatus
GATCACTGCT	18.4	Hs.332649	AV687243		Transcribed locus		
GATTTCTTTG	18.4	Hs.713537	NM_004484	GPC3	Glypican-3	Xq26.1	Extracellular region/plasma membrane
GGAAAACAGA	18.4	Hs.5302	NM_006149	LGALS4	Lectin, galactoside-binding, soluble, 4 (Galectin-4)	19q13.2	Cytoplasm/plasma membrane
GTAAGAAGTGG	18.4				No match		
TCCTGGGTTC	18.4		F00468		cDNA clone		
TGCACATTTT	18.4		AL589357		cDNA clone		
TGCCCTACAA	18.4		AV759594		cDNA clone		
TTCACATTAG	18.4				No match		
ATGTAAAAAA	16.5	Hs.706744	BC004147	LYZ	Lysozyme	12q15	Extracellular region
CTCCCCCAAA	16.5	Hs.510635	BC032249	IGHV3-23	Immunoglobulin heavy constant μ	14q32.33	Membrane fraction
TGAATAAGAG	16.5	Hs.689477	AK127934		Hypothetical LOC100129060	3p21.31	
AGAAAAGTGG	15.6		CD518154		cDNA clone		

FUCA2, LGALS4, LCN2, GPC3) were quantified by real-time PCR in paired tumor/non-tumor samples from six patients with HBV-related small HCC: two patients contributed to the SAGE library construction and the other four patients were enrolled after the construction of the HCC SAGE library. Underlying liver disease was liver cirrhosis in four patients and chronic hepatitis in two patients (Table II). The mRNA levels of FUCA2, LGALS4, LCN2, GPC3 were significantly higher in HCC compared to adjacent non-neoplastic liver tissues. The expression of GSTA1 was, however, variable in HCC: expression was up-regulated in four, down-regulated in two, and overall expression levels were not different between HCC and adjacent liver.

Expression of HCC-related genes in SAGE libraries of extrahepatic normal tissues (Fig. 3). Seventeen HCC-related genes identified by comparing the SAGE libraries of HCC and normal liver were further studied by comparing SAGE libraries between HCC and extrahepatic normal organs to elucidate the expression patterns of these genes in the human body. Four genes were expressed significantly higher in HCC library than in any other libraries of normal extrahepatic tissues ($P<0.05$): fatty acid desaturase 2 (FADS2), fucosidase, α -L-2 (FUCA2), testis-specific protein Y-encoded-like 2 (TSPYL2) and Gon-4-like (GON4L). The other 13 genes were significantly expressed in one or more libraries of normal organs.

Discussion

As SAGE technique produces genome-wide transcript data in a digital format, simple statistical analysis can detect a set of differently expressed genes from any independent libraries. In search of potential diagnostic markers of HBV-associated small HCC, we compared our HCC SAGE library not only with normal liver but also with SAGE libraries of other normal organs to confirm the specificity of the identified genes in the whole human body. To our knowledge, this is the first study on the comparison of HCC transcriptome across the major organs. Two previous studies on the transcriptome analysis of HCV-related HCC by SAGE method (23,24) did not compare their HCC SAGE library with other normal tissues.

The distribution of tag frequencies (Fig. 1A) and the correlation coefficient between our HCC SAGE library and normal liver SAGE library ($R^2 = 0.636$, Fig. 1B) were quite similar to the previous study (24). However, our list of up-regulated genes in HCC (>5 -fold, data not shown) shares only a couple of genes with the previous studies on HCC SAGE [LGALS4 (23), vitronectin (23), MHC class IC (24), GPC3 (24), diubiquitin (24)]. Although SAGE is believed to be a highly reproducible method (25), we expect inter-individual variability which is not an inherent problem of SAGE (25-28). In addition, further variability can be introduced to the HCC transcriptome with respect to viral etiology (29), tumor differentiation, size of tumor (30) and contamination of non-tumor cells. Although both of the previous HCC SAGE data were built from Japanese patients infected with hepatitis C virus, the catalogs of top six up-regulated transcripts shared no common genes between the two studies (23,24). In order to overcome the biological and/or technical variability and to obtain adequate statistical power, sample size needs to

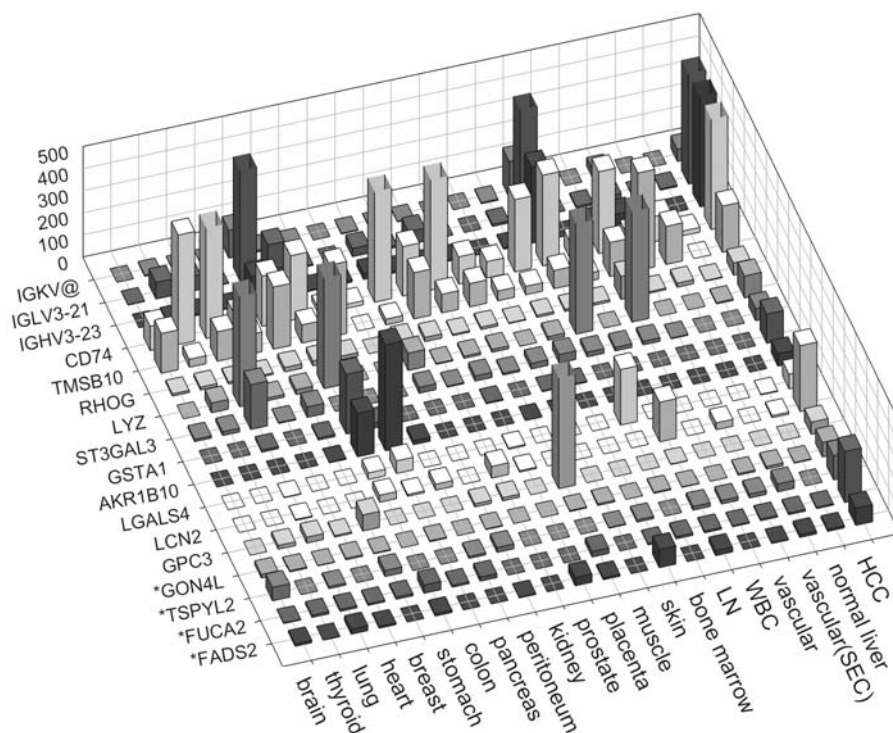


Figure 3. Expression of HCC-associated tags in normal tissues. Expression frequencies of tags which were 15 times or more abundant in HCC than in normal liver SAGE library were obtained from extrahepatic SAGE libraries and normalized to 200,000 total tags per library. Four tags (*) were expressed significantly higher in HCC library than any other SAGE libraries. Vertical axis represents tag frequencies (trimmed at 500).

be increased (31,32). HCC SAGE library from single patient may not be as representative as microarray data obtained from hundreds of HCC samples. However, construction of multiple SAGE libraries is, although less demanding in workload than cDNA library sequencing, much more difficult compared to microarray study. Another realistic option is pooling of samples before generating a SAGE library. For microarray experiments in which only one or two datasets are available, pooling is recommended because it can reduce the effect of individual variability via biological averaging (33). Pooling is also desirable when the study is interested in identifying biomarkers rather than in making an individual inference about prognosis and diagnosis. It has not been studied how pooling affects SAGE libraries, but it seems reasonable to assume that pooling can reduce the effect of individual variation and can help making generalization of the SAGE results. Our HCC SAGE library was constructed from pooled samples of five small HCC patients and may have the advantage of averaging effect.

The total tag number of our HCC SAGE library was smaller than that of normal liver. However, it was reported that the relative tag frequencies were not influenced by total tag size when the total tag counts were >10,000 and individual tag frequencies were >20 (34). Libraries of our size can represent highly expressed genes because biological variability seems to be greater than binomial sampling variability (28). Aurora kinase A, a reference gene of HCC library (35), was detected in our HCC SAGE library but not in the normal liver SAGE library (data not shown), indicating that our HCC SAGE library contains enough tags for comparison with normal liver.

Identification of novel transcripts is another potential advantage of the SAGE technique. Among the top 56 SAGE

tags from Table IV, 18 and 30 tags were matched to characterized genes and ESTs, respectively, whereas six tags were not matched to the most current SAGE Genie database. Percentage of mismatched SAGE tags was reported to be 7 to 20% (36,37). Sequencing errors might contribute to the unmatched tags in our study, but this is not very likely because of relatively high tag frequencies in our library. Alternatively, they may represent truly novel transcripts (38). The significance of the 30 ESTs is also to be defined in a further study. The fact that three of these ESTs were originally identified in cDNA libraries generated from HCC tissues (13) suggests that our ESTs may be truly associated with HBV-related small HCC.

As we compared the HCC SAGE data to the SAGE library of normal liver, it can be argued that up-regulated genes from our results may not necessarily be overexpressed in HCC compared to cirrhotic background which is present in the majority of HCC patients. Our real-time quantitative RT-PCR, however, used liver cirrhosis tissues as controls and confirmed that five of the six genes identified from our SAGE library analysis (except for GSTA1) were still overexpressed in small HBV-related HCC tissues compared to background cirrhotic tissues from same patients. GSTA is a member of detoxification enzymes in normal hepatocytes, and yet the expression of GSTA1 in liver cirrhosis and HCC has not been studied extensively. One study of GSTA expression also reported similar levels between HBV-associated HCC and non-neoplastic liver tissues (39). The Lack of GSTA1 expression in the normal liver SAGE library may be explained by genetic polymorphism (40) and larger sample number would be needed to determine if there is real difference in GSTA expression between HCC and liver cirrhosis.

Among the 17 genes that were expressed 15 times more in HCC compared to normal liver, 13 genes (GSTA1, LCN2, LGALS4, AKR1B10, GPC3, ST3GAL3, LYZ, RHOG, TMSB10, CD74, IGHV3-23, IGLV3-2 and IGKV@) were significantly expressed in one or more extrahepatic normal organs. Although these genes may contribute to carcinogenesis and/or growth of HCC, application of these genes in the early diagnosis of HCC may be limited by non-specific extrahepatic expression.

It is an elusive and yet quite interesting finding that immunoglobulin genes were top ranking among the 15 highly expressed genes in HCC. A couple of microarray-based studies reported overexpression of immunoglobulin genes in HCC (12,13). Up-regulation of IGHV3-23 tag was also observed in SAGE libraries of other cancers such as lung, pancreas and ovary (data not shown). mRNA from tumor-infiltrating lymphocytes may have contaminated the HCC SAGE library, but this hypothesis needs further validation.

Four genes (FADS2, FUCA2, TSPYL2 and GON4L) were expressed significantly higher in HCC SAGE library than in any other SAGE libraries of normal human body. FADS2 exists as a membrane-bound form in human (41). FADS2 (Δ -6 desaturase) introduces double bonds to linoleic acid to generate arachidonic acid, an important molecule in cell signaling. The role of FADS2 in HCC has not been studied yet, but there is a possibility that FADS2-mediated arachidonic acid may be involved in early carcinogenesis (42). Serum activity of α -L-fucosidase was previously reported as a potential marker of HCC (43,44). New kinetic assay of α -L-fucosidase showed promising results with sensitivity and specificity better than AFP (45). Further studies are needed to elucidate the role of α -L-fucosidase in the screening of small HCC. TSPYL2 is homologous to testis-specific protein Y-encoded (TSPY), a cancer/testis antigen which was recently reported to be expressed in 35% of HCC tissues (46). Also known as cutaneous T-cell lymphoma (CTCL)-associated tumor antigen se20-4 (47) and differentially expressed nucleolar TGF- β 1 target protein (48), TSPYL2 is elevated in the cell lines of CTCL (47), melanoma (47), leukemia (47) and non-small cell lung cancer (48). Although localized to the nucleus, TSPYL2 was detected in the serum of CTCL patients (47). It would be worth exploring the usefulness of TSPYL2 for a serum marker of HCC. GON4L is a putative evolutionarily conserved transcription factor (49), but the functional significance of GON4L in cancer is not known.

GPC3 is an emerging marker of HCC: early studies showed that >70% of HCC express mRNA of GPC3 (50,51) and 40-53% of HCC patients were positive for serum GPC3 (51-53). Test for serum GPC3 may improve the sensitivity of HCC screening because serum GPC3 did not correlate with serum AFP (52). Our SAGE and real-time PCR results also confirmed that GPC3 is up-regulated in small HCC, while there was no significant difference of AFP expression between SAGE libraries of HCC and normal liver (data not shown). Although the expression of GPC3 is relatively liver-specific, it is expressed in placenta (54) and our data also verified this finding (Fig. 3). GPC3 is still very promising, however, as a histochemical marker for HCC that can be used to distinguish HCC from benign hepatocellular mass lesions (55,56).

In summary, our SAGE library has identified differentially expressed genes in HBV-associated small HCC. The significance of these genes in the carcinogenesis and diagnosis of HBV-related small HCC is to be elucidated in further studies.

Acknowledgements

This study was supported by grant 03-2004-007 from the SNUBH Research Fund to Dr J.W. Kim.

References

1. Parkin DM, Bray F, Ferlay J and Pisani P: Global cancer statistics, 2002. *CA Cancer J Clin* 55: 74-108, 2005.
2. Zhang BH, Yang BH and Tang ZY: Randomized controlled trial of screening for hepatocellular carcinoma. *J Cancer Res Clin Oncol* 130: 417-422, 2004.
3. Daniele B, Bencivenga A, Megna AS and Tinessa V: Alpha-fetoprotein and ultrasonography screening for hepatocellular carcinoma. *Gastroenterology* 127: S108-S112, 2004.
4. Bruix J and Sherman M: Management of hepatocellular carcinoma. *Hepatology* 42: 1208-1236, 2005.
5. Marrero JA and Lok AS: Newer markers for hepatocellular carcinoma. *Gastroenterology* 127: S113-S119, 2004.
6. Nam SW, Park JY, Ramasamy A, *et al*: Molecular changes from dysplastic nodule to hepatocellular carcinoma through gene expression profiling. *Hepatology* 42: 809-818, 2005.
7. Iizuka N, Oka M, Yamada-Okabe H, *et al*: Oligonucleotide microarray for prediction of early intrahepatic recurrence of hepatocellular carcinoma after curative resection. *Lancet* 361: 923-929, 2003.
8. Lee JS, Chu IS, Heo J, *et al*: Classification and prediction of survival in hepatocellular carcinoma by gene expression profiling. *Hepatology* 40: 667-676, 2004.
9. Matoba K, Iizuka N, Gondo T, *et al*: Tumor HLA-DR expression linked to early intrahepatic recurrence of hepatocellular carcinoma. *Int J Cancer* 115: 231-240, 2005.
10. Boyault S, Rickman DS, de Reynies A, *et al*: Transcriptome classification of HCC is related to gene alterations and to new therapeutic targets. *Hepatology* 45: 42-52, 2007.
11. Smith MW, Yue ZN, Geiss GK, *et al*: Identification of novel tumor markers in hepatitis C virus-associated hepatocellular carcinoma. *Cancer Res* 63: 859-864, 2003.
12. Patil MA, Chua MS, Pan KH, *et al*: An integrated data analysis approach to characterize genes highly expressed in hepatocellular carcinoma. *Oncogene* 24: 3737-3747, 2005.
13. Xu XR, Huang J, Xu ZG, *et al*: Insight into hepatocellular carcinogenesis at transcriptome level by comparing gene expression profiles of hepatocellular carcinoma with those of corresponding noncancerous liver. *Proc Natl Acad Sci USA* 98: 15089-15094, 2001.
14. Tsai CC, Huang KW, Chen HF, *et al*: Gene expression analysis of human hepatocellular carcinoma by using full-length cDNA library. *J Biomed Sci* 13: 241-249, 2006.
15. Yoon SY, Kim JM, Oh JH, *et al*: Gene expression profiling of human HBV- and/or HCV-associated hepatocellular carcinoma cells using expressed sequence tags. *Int J Oncol* 29: 315-327, 2006.
16. Velculescu VE, Zhang L, Vogelstein B and Kinzler KW: Serial analysis of gene expression. *Science* 270: 484-487, 1995.
17. Lash AE, Tolstoshev CM, Wagner L, *et al*: SAGEmap: a public gene expression resource. *Genome Res* 10: 1051-1060, 2000.
18. Boon K, Osorio EC, Greenhut SF, *et al*: An anatomy of normal and malignant gene expression. *Proc Natl Acad Sci USA* 99: 11287-11292, 2002.
19. Margulies EH and Innis JW: eSAGE: managing and analysing data generated with serial analysis of gene expression (SAGE). *Bioinformatics* 16: 650-651, 2000.
20. Pleasance ED, Marra MA and Jones SJ: Assessment of SAGE in transcript identification. *Genome Res* 13: 1203-1215, 2003.
21. Audic S and Claverie JM: The significance of digital gene expression profiles. *Genome Res* 7: 986-995, 1997.

22. Fernandez-Velasco M, Goren N, Benito G, Blanco-Rivero J, Bosca L and Delgado C: Regional distribution of hyperpolarization-activated current (I_f) and hyperpolarization-activated cyclic nucleotide-gated channel mRNA expression in ventricular cells from control and hypertrophied rat hearts. *J Physiol* 553: 395-405, 2003.
23. Kondoh N, Wakatsuki T, Ryo A, *et al*: Identification and characterization of genes associated with human hepatocellular carcinogenesis. *Cancer Res* 59: 4990-4996, 1999.
24. Yamashita T, Kaneko S, Hashimoto S, *et al*: Serial analysis of gene expression in chronic hepatitis C and hepatocellular carcinoma. *Biochem Biophys Res Commun* 282: 647-654, 2001.
25. Blackshaw S, Kuo WP, Park PJ, *et al*: MicroSAGE is highly representative and reproducible but reveals major differences in gene expression among samples obtained from similar tissues. *Genome Biol* 4: R17, 2003.
26. Baggerly KA, Deng L, Morris JS and Aldaz CM: Differential expression in SAGE: accounting for normal between-library variation. *Bioinformatics* 19: 1477-1483, 2003.
27. Wong LY, Hafeman A, Boyd VL, *et al*: Assessing gene expression variation in normal human tissues using GeneTag, a novel, global, sensitive profiling method. *J Biotechnol* 101: 199-217, 2003.
28. Vencio RZ, Brentani H, Patrao DF and Pereira CA: Bayesian model accounting for within-class biological variability in Serial Analysis of Gene Expression (SAGE). *BMC Bioinformatics* 5: 119, 2004.
29. 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.
30. Mas VR, Maluf DG, Archer KJ, Yanek K, Williams B and Fisher RA: Differentially expressed genes between early and advanced hepatocellular carcinoma (HCC) as a potential tool for selecting liver transplant recipients. *Mol Med* 12: 97-104, 2006.
31. Wei C, Li J and Bumgarner RE: Sample size for detecting differentially expressed genes in microarray experiments. *BMC Genomics* 5: 87, 2004.
32. Shih JH, Michalowska AM, Dobbin K, Ye Y, Qiu TH and Green JE: Effects of pooling mRNA in microarray class comparisons. *Bioinformatics* 20: 3318-3325, 2004.
33. Kendzierski C, Irizarry RA, Chen KS, Haag JD and Gould MN: On the utility of pooling biological samples in microarray experiments. *Proc Natl Acad Sci USA* 102: 4252-4257, 2005.
34. Yamamoto M, Wakatsuki T, Hada A and Ryo A: Use of serial analysis of gene expression (SAGE) technology. *J Immunol Methods* 250: 45-66, 2001.
35. Su WH, Chao CC, Yeh SH, Chen DS, Chen PJ and Jou YS: OncoDB.HCC: an integrated oncogenomic database of hepatocellular carcinoma revealed aberrant cancer target genes and loci. *Nucleic Acids Res* 35: D727-D731, 2007.
36. Lu J, Lal A, Merriman B, Nelson S and Riggins G: A comparison of gene expression profiles produced by SAGE, long SAGE, and oligonucleotide chips. *Genomics* 84: 631-636, 2004.
37. Li YJ, Xu P, Qin X, *et al*: A comparative analysis of the information content in long and short SAGE libraries. *BMC Bioinformatics* 7: 504, 2006.
38. Chen J, Sun M, Lee S, Zhou G, Rowley JD and Wang SM: Identifying novel transcripts and novel genes in the human genome by using novel SAGE tags. *Proc Natl Acad Sci USA* 99: 12257-12262, 2002.
39. Zhou T, Evans AA, London WT, *et al*: Glutathione S-transferase expression in hepatitis B virus-associated human hepatocellular carcinogenesis. *Cancer Res* 57: 2749-2753, 1997.
40. Coles BF and Kadlubar FF: Human alpha class glutathione S-transferases: genetic polymorphism, expression, and susceptibility to disease. *Methods Enzymol* 401: 9-42, 2005.
41. Sperling P, Ternes P, Zank TK and Heinz E: The evolution of desaturases. *Prostaglandins Leukot Essent Fatty Acids* 68: 73-95, 2003.
42. Hansen-Petrik MB, McEntee MF, Johnson BT, *et al*: Selective inhibition of Delta-6 desaturase impedes intestinal tumorigenesis. *Cancer Lett* 175: 157-163, 2002.
43. Takahashi H, Saibara T, Iwamura S, *et al*: Serum alpha-L-fucosidase activity and tumor size in hepatocellular carcinoma. *Hepatology* 19: 1414-1417, 1994.
44. Giardina MG, Matarazzo M, Morante R, *et al*: Serum alpha-L-fucosidase activity and early detection of hepatocellular carcinoma: a prospective study of patients with cirrhosis. *Cancer* 83: 2468-2474, 1998.
45. Wang JJ and Cao EH: Rapid kinetic rate assay of the serum alpha-L-fucosidase in patients with hepatocellular carcinoma by using a novel substrate. *Clin Chim Acta* 347: 103-109, 2004.
46. Yin YH, Li YY, Qiao H, *et al*: TSPY is a cancer testis antigen expressed in human hepatocellular carcinoma. *Br J Cancer* 93: 458-463, 2005.
47. Eichmuller S, Usener D, Dummer R, Stein A, Thiel D and Schadendorf D: Serological detection of cutaneous T-cell lymphoma-associated antigens. *Proc Natl Acad Sci USA* 98: 629-634, 2001.
48. Ozbun LL, You L, Kiang S, Angdisen J, Martinez A and Jakowlew SB: Identification of differentially expressed nucleolar TGF-beta1 target (DENTT) in human lung cancer cells that is a new member of the TSPY/SET/NAP-1 superfamily. *Genomics* 73: 179-193, 2001.
49. Kuryshv VY, Vorobyov E, Zink D, *et al*: An anthropoid-specific segmental duplication on human chromosome 1q22. *Genomics* 88: 143-151, 2006.
50. Hsu HC, Cheng W and Lai PL: Cloning and expression of a developmentally regulated transcript MXR7 in hepatocellular carcinoma: biological significance and temporospatial distribution. *Cancer Res* 57: 5179-5184, 1997.
51. Zhu ZW, Friess H, Wang L, *et al*: Enhanced glypican-3 expression differentiates the majority of hepatocellular carcinomas from benign hepatic disorders. *Gut* 48: 558-564, 2001.
52. Capurro M, Wanless IR, Sherman M, *et al*: Glypican-3: a novel serum and histochemical marker for hepatocellular carcinoma. *Gastroenterology* 125: 89-97, 2003.
53. Nakatsura T, Yoshitake Y, Senju S, *et al*: Glypican-3, overexpressed specifically in human hepatocellular carcinoma, is a novel tumor marker. *Biochem Biophys Res Commun* 306: 16-25, 2003.
54. Khan S, Blackburn M, Mao DL, Huber R, Schlessinger D and Fant M: Glypican-3 (GPC3) expression in human placenta: localization to the differentiated syncytiotrophoblast. *Histol Histopathol* 16: 71-78, 2001.
55. Coston WM, Loera S, Lau SK, *et al*: Distinction of hepatocellular carcinoma from benign hepatic mimickers using Glypican-3 and CD34 immunohistochemistry. *Am J Surg Pathol* 32: 433-444, 2008.
56. Wang HL, Anatelli F, Zhai QJ, Adley B, Chuang ST and Yang XJ: Glypican-3 as a useful diagnostic marker that distinguishes hepatocellular carcinoma from benign hepatocellular mass lesions. *Arch Pathol Lab Med* 132: 1723-1728, 2008.