FEBS 29247 FEBS Letters 579 (2005) 1089–1100

Self-organizing-map-based molecular signature representing the development of hepatocellular carcinoma

Norio Iizuka[a,b], Masaaki Oka[a,*], Hisafumi Yamada-Okabe[c], Naohide Mori[a], Takao Tamesa[a], Toshimasa Okada[a], Norikazu Takemoto[a], Kazuhiko Sakamoto[a], Kenji Hamada[c], Hideo Ishitsuka[c,1], Takanobu Miyamoto[d], Shunji Uchimura[d], Yoshihiko Hamamoto[d]

[a] *Departments of Surgery II, Yamaguchi University School of Medicine, 1-1-1 Minami-Kogushi, Ube, Yamaguchi 755-8505, Japan*
[b] *Departments of Bioregulatory Function, Yamaguchi University School of Medicine, 1-1-1 Minami-Kogushi, Ube, Yamaguchi 755-8505, Japan*
[c] *Pharmaceutical Research Department 4, Kamakaura Research Laboratories, Chugai Pharmaceutical Co. Ltd., 200 Kajiwara, Kamakura, Kanagawa 247-8530, Japan*
[d] *Department of Computer Science and Systems Engineering, Faculty of Engineering, Yamaguchi University, 2-16-1 Tokiwadai, Ube, Yamaguchi 755-8611, Japan*

Received 8 August 2004; revised 9 October 2004; accepted 21 October 2004

Available online 18 January 2005

Edited by Lukas Huber

Abstract **Using high-density oligonucleotide array, we comprehensively analyzed expression levels of 12 600 genes in 50 hepatocellular carcinoma (HCC) samples with positive hepatitis C virus (HCV) serology (well (G1), moderately (G2), and poorly (G3) differentiated tumors) and 11 non-tumorous livers (L1 and L0) with and without HCV infection. We searched for discriminatory genes of transition (L0 vs. L1, L1 vs. G1, G1 vs. G2, G2 vs. G3) with a supervised learning method, and then arranged the samples by self-organizing map (SOM) with the discriminatory gene sets. The SOM arranged the five clusters on a unique sigmoidal curve in the order L0, L1, G1, G2, and G3. The sample arrangement reproduced development-related features of HCC such as p53 abnormality. Strikingly, G2 tumors without venous invasion were located closer to the G1 cluster, and most G2 tumors with venous invasion were located closer to the G3 cluster ($P = 0.001$ by Fisher's exact test). Our present profiling data will serve as a framework to understand the relation between the development and dedifferentiation of HCC.**

Keywords: DNA microarray; Dedifferentiation; HCV; HCC; Hepatocellular carcinoma

1. Introduction

Hepatocellular carcinoma (HCC) is one of the most common cancers and represents a major international health problem because the incidence is increasing in many countries [1–3]. Hepatitis B and C virus (HBV and HCV) infections are major risk factors for HCC [3–5]. Many factors [6–10] such as HBx protein, HCV core protein, *AXIN1*, *p53*, and allele loss on chromosome 16 are associated with the pathogenesis of HCC. Despite intense research efforts in this field, it remains unclear how HCC develops and how many genes are involved in the course of development.

It has been suggested that the majority of small, early stage HCCs are well differentiated and that tumor advancement, which is characterized by pathologic features, occurs via dedifferentiation [10–13]. Dedifferentiation also occurs in transgenic mice carrying the HCV core gene [7]. An example of dedifferentiation is nodule-in-nodule-type HCC (progressed HCC within early HCC) [10,11,13]. However, it remains unclear whether dedifferentiation is associated with the development of all HCC cases. A limitation to testing this is the inability to collect longitudinal samples from a single patient during the development of HCC. Molecular profiling of a population in which each member is at a different stage of differentiation might enable us to understand the development of this disease. To test this and to elucidate the molecular basis of HCC, we investigated the levels of expression of 12 600 genes in HCV antibody-positive HCCs of three grades of differentiation, non-tumorous livers without HCV infection, and HCV-infected non-tumorous livers by high-density oligonucleotide array. With the use of a supervised learning method [14–17], we profiled genes whose expression differed significantly between classes at a different stage. We arranged all samples according to a self-organizing map (SOM) [18,19] with the most discriminatory gene sets and investigated the development of HCC on the basis of dedifferentiation.

2. Materials and methods

2.1. Samples

We analyzed levels of expression of 12 600 genes in 76 HCC samples (Supplementary Table) from 76 patients who underwent surgical treatment for HCC at Yamaguchi University Hospital between May 1997 and August 2000. Written informed consent was obtained from all patients before surgery. The study protocol was approved by the Institutional Review Board for the Use of Human Subjects at the Yamaguchi University School of Medicine. Among the 76 patients, 50 were seropositive for HCV antibody (HCVAb) and seronegative for hepatitis B virus surface antigen (HBsAg) (Table 1). The remaining 26 were seronegative for HCVAb. A histopathologic diagnosis of HCC was made in all cases after surgery by two experienced pathologists. In the 50 patients with HCV-related HCC, histopathologic examination based on TNM classification of the International Union Against

* Corresponding author.
E-mail address: 2geka-1@po.cc.yamaguchi-u.ac.jp (M. Oka).

[1] Present address: Roche Diagnostics K.K., 6-1, Shiba 2-choume, Minato-ku, Tokyo 105-0014, Japan.

0014-5793/$30.00
doi:10.1016/j.febslet.2004.10.113

Table 1
Clinocopathologic characteristics per study group of 50 HCV-positive HCCs

Factors	Well (G1)[a]	Moderately (G2)[a]	Poorly (G3)[a]	*P*-value
Sex				$P = 0.8007$
Male	4	24	6	
Female	3	11	2	
Age (year)[b]	65.3 ± 7.0	65.4 ± 7.1	67.2 ± 9.5	$P = 0.9612$ (G1 vs. G2) $P = 0.6595$ (G1 vs. G3) $P = 0.5406$ (G2 vs. G3)
Primary lesion				$P = 0.0568$
Single tumor	6	15	2	
Multiple tumors	1	20	6	
Capsule formation				$P = 0.3339$
Present	4	29	6	
Absent	3	6	2	
Tumor size (cm)[b]	2.0 ± 0.8	5.0 ± 3.2	6.0 ± 7.0	$P = 0.0007$ (G1 vs. G2) $P = 0.0279$ (G1 vs. G3) $P = 0.6397$ (G2 vs. G3)
Stage[a]				$P = 0.0656$
I	6	10	2	
II	1	17	3	
IIIA/IV	0	8	3	
Microscopic venous invasion[a]				$P = 0.0381$
(−)	7	21	3	
(+)	0	14	5	
Alpha-feto protein (ng/ml)				$P = 0.1504$
<or =100	6	24	3	
>100	1	11	5	
Non-tumorous liver				$P = 0.7569$
Normal or chronic hepatitis	2	15	2	
Liver cirrhosis	5	20	6	

Fisher's exact test, Student's *t* test and Mann–Whitney's *U* test were used to elucidate differences in backgrounds between each group.
[a]Tumor differentiation, stage, and microscopic venous invasion were determined on the basis of TNM classification of UICC (Ref. [20]). G1–G3 tumors are equal to types I–III of Edmondson and Steiner classification, respectively.
[b]Mean ± S.D.

Cancer [20] revealed that seven had well differentiated HCC (G1), 35 had moderately differentiated HCC (G2), and the remaining eight had poorly differentiated HCC (G3). The degree of tumor differentiation in all samples was reproducible between the pathologists. Clinicopathologic factors were also determined according to TNM classification. Fisher's exact test, Student's *t* test, and the Mann–Whitney *U* test were used to evaluate differences in clinicopathologic characteristics among G1, G2, and G3 tumors. A value of $P < 0.05$ was considered to be statistically significant.

As controls, six non-tumorous liver samples (L0) from six patients who underwent hepatic resection for benign or metastatic liver tumors (1 focal nodular hyperplasia, 2 hemanigiomas, and 3 metastatic liver cancers derived from 2 colon cancers and 1 gastric cancer) and who had histologically normal livers were used. All control subjects were seronegative for both HBsAg and HCVAb. We also had five HCV-infected non-tumorous liver samples from five HCC patients. Among the five liver samples (L1), 2 were histopathologically diagnosed as chronic hepatitis and 3 as liver cirrhosis. Informed consent in writing was obtained from all of these patients before surgery.

2.2. DNA microarray analysis

We divided the resected specimens into two groups immediately after surgery; one was immediately frozen in liquid nitrogen for RNA extraction, and the other was fixed in 10% formaldehyde solution and embedded in paraffin. Samples from the latter group were used to confirm histopathologically that tissues used for RNA extraction were not necrotic. Extraction of RNA was performed as described previously [14–17]. In all samples, the quality of extracted RNA was confirmed by the appearance of characteristic 28S and 18S rRNA fragments on agarose gels. Synthesis of cDNA and cRNA and oligonucleotide microarray screening were performed as described previously [14–17]. In the present study, huU95A DNA Chips® (Affymetrix, Santa Clara, CA) [21] containing 12 600 genes were used.

2.3. Gene selection

We first selected genes with expression levels >40 (arbitrary units by Affymetrix) in all 50 HCC samples and 11 non-tumorous liver samples. This filtering resulted in the identification of 3559 genes. We used the Fisher ratio [14–17] to evaluate the potentials of the selected genes to discriminate L0 vs. L1, L1 vs. G1, G1 vs. G2, and G2 vs. G3. For each transition, the 3559 genes were ranked in order of decreasing magnitude of the Fisher ratio. To determine the number of genes to be considered, a random permutation test was performed as described previously [14–17]. From the distribution of the Fisher ratios based on randomized data, all genes that passed the random permutation test ($P < 0.005$) were selected. We found that the expression levels of 152 genes with Fisher ratios >4.90 between L0 and L1 were statistically significant. Likewise, we identified significant discriminatory genes, the top 191 genes with the Fisher ratios >4.08 between L1 and G1, the top 54 genes with the Fisher ratios >1.52 between G1 and G2, and the top 40 genes with the Fisher ratios >1.34 between G2 and G3. To estimate the percentage of genes identified by chance in the above gene selections, the false discovery rate (FDR) was calculated as described previously [22]. Permutations were defined as balanced, and the Fisher ratio for each gene was computed for each of the balanced permutations. To verify the results provided by the random permuta-

tion test, the number of falsely significant genes corresponding to each balanced permutation was computed by counting the number of genes that exceeded the threshold for the random permutation test. The estimated number of falsely significant genes was the average of the number of genes termed significant from all balanced permutations. The FDR was determined as the ratio of the estimated number of falsely significant genes to the number of genes identified by the random permutation test.

2.4. Minimum distance classifier

To compare classes, we designed the minimum distance classifier with the top 40 genes selected in each transition. First, the level of expression of each gene was normalized to have zero mean and unit variance with training samples from two classes. To classify a sample, we measured the Euclidean distance from the sample to each of the mean vectors and assigned the sample to the grade of the nearest mean vector.

2.5. Self-organizing map

A SOM is a neural network algorithm that is used widely for clustering and is an efficient tool for the visualization of multi-dimensional data [18,19]. We used MATLAB R13 (The MathWorks, Inc. Natick, MA) with the SOM toolbox 2.0 (http://www.cis.hut.fi/projects/somtoolbox/) to arrange all samples.

3. Results

3.1. Clinicopathologic characteristics of HCV-related HCCs according to differentiation grade

The tumor size in G2 and G3 was significantly larger than that in G1 (Table 1) ($P = 0.0007$ and $P = 0.028$, respectively, by Mann–Whitney U test). None of the seven G1 tumors involved vessels. G2 and G3 tumors involved vessels more frequently than did G1 tumors ($P = 0.038$ by Fisher's exact test). From G1 to G3, tumor stage tended to be more advanced ($P = 0.066$ by Fisher's exact test). Thus, HCCs of three grades of differentiation showed features characteristic of cancer development (tumor size, vessel involvement, and tumor stage) [11,13]. We hypothesized that HCC develops sequentially from L0 to L1 to G1 to G2 to G3, and we searched for discriminatory genes for the four transitions (L0 to L1, L1 to G1, G1 to G2, G2 to G3).

3.2. Genes selected in each transition of hypothesized development

Our gene selection procedure identified 152 genes with expression levels that differed significantly between L0 and L1. Among these genes, expression of 67 was upregulated and that of 85 was downregulated in L1 compared to L0 (Fig. 1(a), Table 2, and supplemental tables). Likewise, we identified 191 genes with expression levels that differed significantly between L1 and G1. Among these genes, expression of 95 was upregulated and that of 96 was downregulated in G1 compared to L1 (Fig. 1(b), Table 3, and supplemental tables). Among the 54 genes differentially expressed between G1 and G2 tumors, expression of 36 genes was increased and that of 18 genes was decreased in G2 compared to G1 (Fig. 1(c), Table 4, and supplemental tables). Among the 40 genes that were differentially expressed between G2 and G3 tumors, expression of

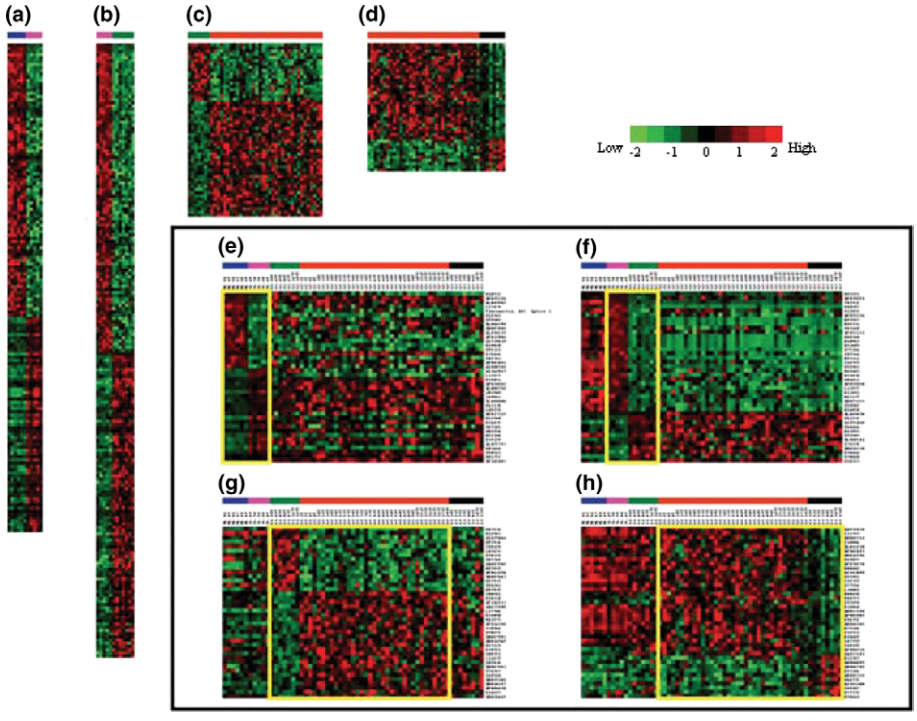

Fig. 1. Genes that discriminate transitions in the development of HCC expression of 152 genes with significantly altered expression during transition from L0 (blue) to L1 (pink) (a), 191 genes with significantly altered expression during transition from L1 to G1 (green) (b), 54 genes with significantly altered expression during transition from G1 to G2 (orange) (c) and 40 genes with significantly altered expression during transition from G2 to G3 (black) (d) are shown in color. Panels e, f, g, and h illustrate expression of the 40 most discriminatory genes for each transition in all samples. Expression of the 40 genes with significantly altered expression in the transitions from L0 to L1 (e), L1 to G1 (f), G1 to G2 (g), and G2 to G3 (h). The 40 selected genes for each transition discriminate pre-transition and post-transition grade samples. Genes are shown in decreasing order of the Fisher ratio (see Section 2) and are listed as an accession number. Accession number of each gene was obtained from PubMed (Internet address: http://www3.ncbi.nlm.nih.gov/PubMed/) or the Institute for Genomic Research databases (Internet address: http://www.tigr.org/tdb/hgi/searching/reports.html).

Table 2
Top-40 discriminatory genes in L0 and L1

Fisher ratio	GB number	Description	Symbol	Locus	Function
Eighteen genes downregulated in L1 in comparison with L0					
50.45	M18533	Dystrophin	DMD	Xp21.2	Cytoskeleton
23.02	AF035316	Homolog to tubulin beta chain		6p24.3	Unknown
20.65	AL049942	Zinc finger protein 337	ZNF337	20p11.1	Unknown
18.34	L27479	Friedreich ataxia region gene X123	X123	9q13-q21	Unknown
16.63		Fibronectin (Alt. Splice 1)			Extracellular matrix
16.13	U19765	Zinc finger protein 9	ZNF9	3q21	Transcription/retroviral nucleic acid binding protein
14.91	X55503	Metallothionein IV	MTIV	16q13	Detoxification
13.71	AL046394	Poly(rC) binding protein 3	PCBP3	21q22.3	RNA-binding protein/ post-transcriptional control
12.56	AB007886	KIAA0426 gene product	KIAA0426	6p22.2-p21.3	Unknown
12.41	AL050139	Hypothetical protein FLJ13910	FLJ13910	2p11.1	Unknown
12.37	AF012086	RAN binding protein 2-like 1	RANBP2L1	2q12.3	Signal transduction/ small GTP-binding protein
11.66	AI539439	S100 calcium binding protein A2	S100A2	1q21	Extracellular stimuli and cellular responses
11.24	M19828	Apolipoprotein B	APOB	2p24-p23	Lipid metabolism
10.59	U92315	Sulfotransferase family, cytosolic, 2B, member 1	SULT2B1	19q13.3	Steroid metabolism
10.53	D76444	Zinc finger protein 103 homolog (mouse)	ZFP103	2p11.2	Central nervous system development
10.50	X02761	Fibronectin 1	FN1	2q34	Extracellular matrix/cell adhesion and motility
10.20	AF001891	Zinc finger protein-like 1	ZFPL1	11q13	Unknown
9.74	AI400326	EST		2	UniGene Cluster Hs.356456
Twenty-two genes upregulated in L1 in comparison with L0					
40.49	AI362017	Cystatin C	CST3	20p11.21	Cysteine protease inhibitor
21.66	L13977	Prolylcarboxypeptidase (angiotensinase C)	PRCP	11q14	Metabolism/lysosome-related protein
20.59	D32053	Lysyl-tRNA synthetase	KARS	16q23-q24	Protein biosynthesis
13.70	AF038962	Voltage-dependent anion channel 3	VDAC3	8p11.2	Transport of adenine nucleotides
11.90	AL008726	Protective protein for beta-galactosidase (cathepsin A)	PPGB	20q13.1	Lysosomal protein/ enzyme activator
11.71	J03909	Interferon, gamma-inducible protein 30	IFI30	19p13.1	Lysosomal thiol reductase/IFN-inducible
11.32	Z69043	Signal sequence receptor, delta	SSR4	Xq28	Translocatation of newly synthesized polypeptides
11.17	AL080080	Thioredoxin-related transmembrane protein	TXNDC	14q21.3	Redox reaction
11.15	M63138	Cathepsin D	CTSD	11p15.5	Lysosomal aspartyl protease/proteolysis
11.12	L09159	Ras homolog gene family, member A	ARHA	3p21.3	Oncogenesis/actin cytoskeleton reorganization
10.99	AF017115	Cytochrome c oxidase subunit IV isoform 1	COX4I1	16q22-qter	Energy pathway
10.76	M13560	CD74 antigen	CD74	5q32	Immune response
10.22	M36035	Benzodiazapine receptor	BZRP	22q13.31	Flow of cholesterol into mitochondria
10.08	U47101	Nitrogen fixation cluster-like	NIFU	12q24.1	Unknown
9.70	U81554	Calcium/calmodulin-dependent protein kinase II gamma	CAMK2G	10q22	Signal transduction
9.59	M21186	Cytochrome b-245, alpha polypeptide	CYBA	16q24	Energy generation
9.47	D32129	Major histocompatibility complex, class I, A	HLA-A	6p21.3	Immune response
9.44	AL022723	Major histocompatibility complex, class I, F	HLA-F	6p21.3	Immune response
9.41	M83664	Major histocompatibility complex, class II, DP beta 1	HLA-DPB1	6p21.3	Immune response
9.16	U50523	Actin related protein 2/3 complex, subunit 2	ARPC2	13q12-q13	Cell motility and cytoskeleton
9.02	M81757	Ribosomal protein S19	RPS19	19q13.2	Oncogenesis/RNA-binding protein
8.89	AF102803	Catenin (cadherin-associated protein), alpha 1	CTNNA1	5q31	Cell adhesion

GB number of each gene was obtained from PubMed (http://www3.nibi.nlm.nih.gov/PubMed/).
Symbol used is based on the data from LocusLink (http://www.nibi.nlm.nih.gov/LocusLink/).
Downregulated genes; Fold change of L1 vs. L0 < 1.
Upregulated genes; Fold change of L1 vs. L0 > 1.

Table 3
Top-40 discriminatory genes in L1 and G1

Fisher ratio	GB number	Description	Symbol	Locus	Function
Twenty-eight genes downregulated in G1 in comparison with L1					
26.84	M93221	Mannose receptor, C type 1	MRC1	10p13	Phagocytosis and pinocytosis
26.08	AF079221	BCL2/adenovirus E1B 19 kD interacting protein 3-like	BNIP3L	8p21	Tumor suppressor/induction of apoptosis
21.46	V01512	v-fos FBJ murine osteosarcoma viral oncogene homolog	FOS	14q24.3	Oncogenesis/transcription
21.45	D88587	Ficolin 3 (Hakata antigen)	FCN3	1p35.3	Extracellular space
20.15	U12022	Calmodulin 1	CALM1	14q24-q31	Signal transduction/calcium-binding protein
19.73	AF055376	v-maf musculoaponeurotic fibrosarcoma oncogene homolog	MAF	16q22-q23	Oncogenesis/transcription
19.19	R93527	Metallothionein 1H	MT1H	16q13	Detoxification
18.19	R92331	Metallothionein 1E	MT1E	16q13	Detoxification
17.65	U83460	Solute carrier family 31, member1	SLC31A1	9q31-q32	Copper ion transport
17.30	AF052113	RAB14, member RAS oncogene family	RAB14	9q32-q34.11	Ras superfamily member of GTP-binding proteins
15.26	H68340	RNA helicase-related protein	RNAHP	17q22	Alteration of RNA secondary structure
14.96	M10943	Metallothionein 1F	MT1F	16q13	Detoxification
14.18	M13485	Metallothionein 1B	MT1B	16q13	Detoxification
13.34	U75744	Deoxyribonuclease I-like 3	DNASE1L3	3p21.1-3p14.3	DNA metabolism
12.65	X02544	Orosomucoid 1	ORM1	9q31-q32	Immune response/acute-phase response
11.95	M93311	Metallothionein 3	MT3	16q13	Detoxification
11.58	Z24725	Mitogen inducible 2	MIG2	14q22.1	Cell cycle and cell proliferation
11.52	U22961	Unknown			Unknown
11.45	M62403	Insulin-like growth factor binding protein 4	IGFBP4	17q12-q21.1	Signal transduction/cell proliferation
11.01	M35878	Insulin-like growth factor binding protein 3	IGFBP3	7p13-p12	Signal transduction/cell proliferation
10.80	U84011	Amylo-1, 6-glucosidase, 4-alpha-glucanotransferase	AGL	1p21	Glycogen degradation
10.74	AF055030	PHD zinc finger protein XAP135, isoform b	XAP135	6q27	Unknown
10.29	L13977	Prolylcarboxypeptidase (angiotensinase C)	PRCP	11q14	Metabolism/lysosome-related protein
10.02	D13891	Inhibitor of DNA binding 2	ID2	2p25	Negative regulator of cell differentiation
9.95	M63175	Autocrine motility factor receptor	AMFR	16q21	Signal transduction/cell motility
9.94	AB023157	KIAA0940 protein	KIAA0940	10q23.33	Unknown
9.76	U20982	Insulin-like growth factor binding protein 4	IGFBP4	17q12-q21.1	Signal transduction/cell proliferation
9.09	M14058	Complement component 1, r subcomponent	C1R	12p13	Immune response
Twelve genes upregulated in G1 in comparison with L1					
30.42	AL049650	Small nuclear ribonucleoprotein polypeptides B and B1	SNRPB	20p13	RNA processing/modification/RNA splicing
20.95	U61232	Tubulin-specific chaperone e	TBCE	1q42.3	Microtubule/cochaperonin
11.95	AI991040	DR1-associated protein 1	DRAP1	11q13.3	Transcription
10.96	U64444	Ubiquitin fusion degradation 1-like	UFD1L	22q11.21	Proteolysis
10.71	D63997	Golgi autoantigen, golgin subfamily a, 3	GOLGA3	12q24.33	Stabilization of Golgi structure
10.60	X55503	Metallothionein IV	MT4	16q13	Detoxification
10.23	AL080181	Immunoglobulin superfamily, member 4	IGSF4	11q23.2	It possess low similarity to viral receptor
10.01	X76228	aTX1 antioxidant protein 1 homolog (yeast)	ATP6V1E	22q11.1	Proton transport
9.77	AB018330	Calcium/calmodulin-dependent protein kinase 2, beta	CAMKK2	12q24.2	Signal transduction/calcium-binding protein
9.41	D76444	Zinc finger protein 103 homolog (mouse)	ZFP103	2p11.2	Central nervous system development
9.31	U70660	ATX1 antioxidant protein 1 homolog (yeast)	ATOX1	5q32	Copper homeostasis and ion transport
9.10	U10323	Interleukin enhancer binding factor 2, 45 kD	ILF2	1q21.1	Transcription

GB number of each gene was obtained from PubMed (http://www3.nibi.nlm.nih.gov/PubMed/).
Symbol used is based on the data from LocusLink (http://www.nibi.nlm.nih.gov/LocusLink/).
Downregulated genes; Fold change of G1 vs. L1 < 1.
Upregulated genes; Fold change of G1 vs. L1 > 1.

Table 4
Top-40 discriminatory genes in G1 and G2

Fisher ratio	GB number	Description	Symbol	Locus	Function
Fifteen genes downregulated in G2 in comparison with G1					
2.89	M87434	2′-5′-oligoadenylate synthetase 2	OAS2	12q24.2	Antiviral response protein/IFN-inducible
2.63	M12963	Class I alcohol dehydrogenase alpha subunit	ADH1A	4q21-q23	Detoxification
2.51	AI625844	Hypothetical protein FLJ20378			Unknown
2.43	M97936	Signal transducer and activator of transcription 1	STAT1	2q32.2	Transcription/IFN-signaling pathway
2.12	Z99129	Heat shock transcription factor 2	HSF2	6q22.33	Transcription
2.08	L07633	Proteasome activator subunit1	PSME1	14q11.2	Proteolysis and peptidolysis/IFN-inducible
2.06	D50312	Potassium inwardly-rectifying channel subfamily J, member8	KCNJ8	12p11.23	Potassium transport
2.02	U07364	Proteasome activator subunit1	PSME1	14q11.2	Proteolysis and peptidolysis/IFN-inducible
2	AA883502	Ubiquitin-conjugating enzyme E2L6	UBE2L6	11q12	Proteolysis and peptidolysis
1.85	M97935	Signal transducer and activator of transcription 1	STAT1	2q32.2	Transcription/IFN-signaling pathway
1.83	AF061258	LIM protein	LIM	4q22	Signal transduction
1.74	AB007447	FLN 29 gene product	FLN29	12q	Signal transduction
1.72	M97935	Signal transducer and activator of transcription 1	STAT1	2q32.2	Transcription/IFN-signaling pathway
1.7	W28281	GABA(A) receptor-associated protein like 1	GABARAPL1	12p13.1	Microtubule associated protein
1.66	M97935	Signal transducer and activator of transcription 1	STAT1	2q32.2	Transcription/IFN-signaling pathway
Twenty-five genes upregulated in G2 in comparison with G1					
4.41	Y00281	Ribophorin I	RPNI	3q21.3-q25.2	Protein modification/ RNA binding
3.25	D28118	Zinc finger protein 161	ZNF161	17q23.3	Transcription
2.83	AF104913	Eukaryotic protein synthesis initiation facto 4 gamma	EIF4G1	3q27-qter	Translation
2.27	AA675900	Formin binding protein 3	FNBP3	2q23.3	Proteolysis and peptidolysis
2.27	L27706	Chaperonin containing TCP1, subunit 6A (zeta 1)	CCT6A	7p14.1	Chaperone/protein folding
2.15	D32050	Alanyl-tRNA synthetase	AARS	16q22	tRNA processing/ protein synthesis
2.1	M63573	Peptidylprolyl isomerase B	PPIB	15q21-q22	Chaperone/immune response
2.09	AF014398	Inositol(myo)-1(or 4)-monophosphatase 2	IMPA2	18p11.2	Signal transduction
2.08	X70944	Splicing factor proline/glutamine rich	SFPQ	1p34.2	mRNA splicing/ mRNA processing
2.03	U70671	Ataxin 2 related protein	A2LP	7	Unknown
1.89	AA447263	Golgi reassembly stacking protein 2, 55 kDa	GORASP2	2p24.3-q21.3	Golgi stacking
1.87	AB014569	KIAA0669 gene product	KIAA0669	3	Unknown
1.85	M23115	ATPase, Ca++ transporting, cardiac muscle, slow twitch 2	ATP2A2	12q23-q24.1	Small molecule transport
1.83	D38521	Proteasome activator 200 kDa	PA200	2p16.2	Proteolysis and peptidolysis
1.82	X00351	Actin, beta	ACTB	7p15-p12	Cytoskeleton
1.75	L11672	Zinc finger protein 91	ZNF91	19p13.1-p12	Transcription
1.75	X82834	Golgi autoantigen, golgin subfamily a, 4	GOLGA4	3p22-p21.3	Vesicle transport
1.74	AB007963	KIAA0494 gene product	KIAA0494	1pter-p22.1	Unknown
1.74	U76247	Seven in absentia homolog 1 (Drosophila)	SIAH1	16q12	Proteolysis and peptidolysis/ apoptosis
1.73	X68560	Sp3 transcription factor	SP3	2q31	Transcription
1.73	AB015344	Ubiquilin 2	UBQLN2	Xp11.23-p11.1	Ubiquitination
1.73	AB018327	Activity-dependent neuroprotector	ADNP	20q13.13-q13.2	Unknown
1.7	AF004430	Tumor protein D52-like 2	TPD52L2	20q13.2-q13.3	Cell proliferation
1.67	D14697	Farnesyl diphosphate synthase	FDPS	1q21.2	Cholesterol biosynthesis
1.67	AB028449	Dicer1, Dcr-1 homolog (Drosophila)	DICER1	14q32.2	RNA helicase

GB number of each gene was obtained from PubMed (http://www3.nibi.nlm.nih.gov/PubMed/).
Symbol used is based on the data from LocusLink (http://www.nibi.nlm.nih.gov/LocusLink/).
Downregulated genes; Fold change of G2 vs. G1 < 1.
Upregulated genes; Fold change of G2 vs. G1 > 1.

Table 5
Top-40 discriminatory genes in G2 and G3

Fisher ratio	GB number	Description	Symbol	Locus	Function
Thirty genes downregulated in G3 in comparison with G2					
2.36	AA976838	Apolipoprotein C-I	APOC1	19q13.2	Lipid metabolism
2.20	Z11793	Selenoprotein P, plasma, 1	SEPP1	5q31	Antioxidant activity
1.86	AB002311	PDZ domain containing guanine nucleotide exchange factor 1	PDZ-GEF1	4q32.1	Ras/Rap1A-associating signal transduction
1.80	Y18004	Sex comb on midleg-like 2 (Drosophila)	SCML2	Xp22	Transcription/embryogenesis and morphogenesis
1.76	AL031230	Aldehyde dehydrogenase 5 family, member A1	ALDH5A1	6p22	Electron transporter/amino butyrate catabolism
1.71	AF002697	BCL2/adenovirus E1B 19 kD interacting protein 3	BNIP3	14q11.2-q12	Apoptosis
1.65	AB014596	F-box and WD-40 domain protein 1B	FBXW1B	5q35.1	Ubiquitination
1.64	U49897	Phenylalanine hydroxylase	PAH	12q22-q24.2	Amino acid biosynthesis
1.62	AF070570	Homo sapiens clone 24473 mRNA sequence		4	Unknown
1.59	M80482	Paired basic amino acid cleaving system 4	PACE4	15q26	Cell–cell signalling/proteolysis
1.59	AI263099	FLJ31305 fis or clone LIVER1000104		16	Similar to Rattus norvegicus kidney-specific protein mRNA
1.57	U22961	Unknown			Unknown
1.57	Z24725	Mitogen inducible 2	MIG2	14q22.1	Cell cycle control
1.53	U77594	Retinoic acid receptor responder (tazarotene induced) 2	RARRES2	7q35	Retinoic acid receptor/ retinoic acid-inducible
1.49	L34081	Bile acid Coenzyme A: amino acid N-acyltransferase	BAAT	9q22.3	Liver enzyme for glycine and bile acid metabolisms
1.49	M88458	KDEL endoplasmic reticulum protein retention receptor 2	KDELR2	7p22.2	Intracellular protein traffic
1.48	U68723	Checkpoint suppressor 1	CHES1	14q24.3-q31	Transcription/cell cycle
1.48	X92098	Coated vesicle membrane protein	RNP24	12q24.31	Intracellular protein traffic
1.44	D10040	Fatty-acid-Coenzyme A ligase, long-chain 2	FACL2	4q34-q35	Fatty acid metabolism
1.43	AB023194	KIAA0977 protein	KIAA0977	2q24.3	Unknown
1.42	AF001903	L-3-hydroxyacyl-Coenzyme A dehydrogenase, short chain	HADHSC	4q22-q26	Mitochondrial enzyme/ energy generation
1.40	X96752	L-3-hydroxyacyl-Coenzyme A dehydrogenase, short chain	HADHSC	4q22-q26	Mitochondrial enzyme/ energy generation
1.40	AB006202	Succinate dehydrogenase complex, subunit D	SDHD	11q23	Mitochondrial protein/ electron transporter
1.37	M75106	Carboxypeptidase B2	CPB2	13q14.11	Proteolysis and peptidolysis
1.37	Y12711	Rogesterone receptor membrane component 1	PGRMC1	Xq22-q24	Liver-rich protein that binds to progesterone
1.36	D14662	Anti-oxidant protein 2	AOP2	1q23.3	Antioxidant activity/non-selenium glutathione peroxidase
1.36	S87759	Protein phosphatase 1A	PPM1A	14q23.1	Cellular stress responses
1.36	Z48199	Syndecan 1	SDC1	2p24.1	Cell adhesion and metastasis
1.35	AF088219	Chemokine (C–C motif) ligand 14	CCL14	17q11.2	Cell proliferation
1.35	AA453183	EST			Unknown
Ten genes upregulated in G3 in comparison with G2					
2.80	D31767	DAZ associated protein 2	DAZAP2	2q33-q34	RNA-binding protein
2.57	AB000095	Serine protease inhibitor, Kunitz type 1	SPINT1	15q13.3	Inhibitor specific for HGF activator
2.40	AB006782	Galectin 9	LGALS9	17q11.1	Cell adhesion and metastasis
2.18	M21186	Cytochrome b-245, alpha polypeptide	CYBA	16q24	Energy generation
1.96	AB002312	Bromodomain adjacent to zinc finger domain 2A	BAZ2A	12q24.3-qter	DNA-binding protein
1.84	U44772	Palmitoyl-protein thioesterase 1	PPT1	1p32	Neuronal maturation
1.77	AI541308	S100 calcium binding protein A13	S100A13	1q21	Extracellular stimuli and cellular responses
1.53	Z49107	Galectin 9	LGALS9	17q11.1	Cell adhesion and metastasis
1.36	U77735	Pim-2 oncogene	PIM2	Xp11.23	Cell proliferation
1.34	M38449	Transforming growth factor, beta 1	TGFB1	19q13.2	Cell growth and adhesion

GB number of each gene was obtained from PubMed (http://www3.nibi.nlm.nih.gov/PubMed/).
Symbol used is based on the data from LocusLink (http://www.nibi.nlm.nih.gov/LocusLink/).
Downregulated genes; Fold change of G3 vs. G2 < 1.
Upregulated genes; Fold change of G3 vs. G2 > 1.

10 genes was increased and that of 30 was decreased in G3 tumors in comparison to G2 tumors (Fig. 1(d) and Table 5). Interestingly, there was almost no overlap among these discriminatory genes, with exception of 17 (0.39%) of the total 437 genes (see supplementary table). FDRs, the percentage of genes identified by chance for L0 vs. L1, L1 vs. G1, G1 vs. G2, and G2 vs. G3, were 0%, 0%, 0.24%, and 0.29%, respectively.

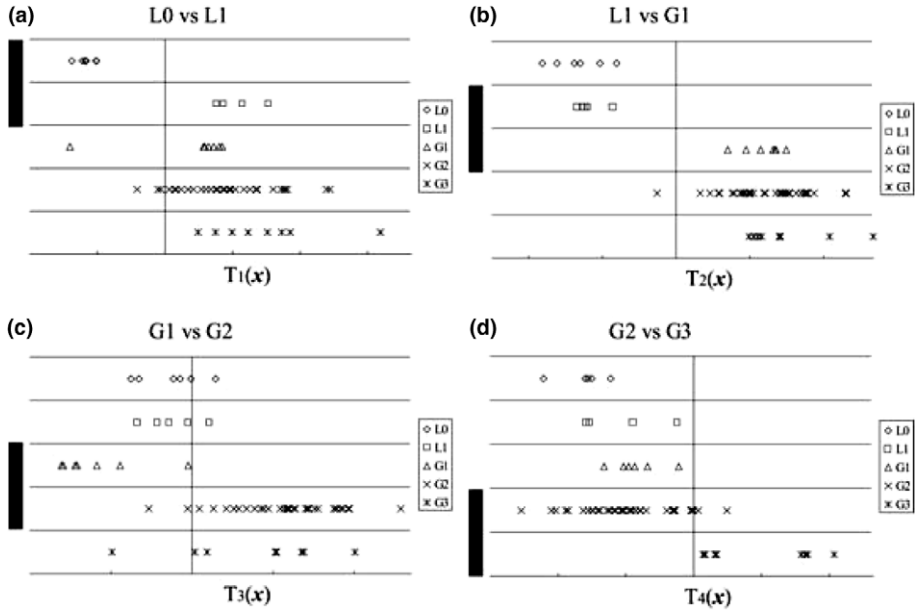

Fig. 2. Validation of the 40 selected genes for each transition in distinguishing differentiation grade In each transition, the minimum distance classifier was constructed with the samples from two consecutive differentiation grades, which are indicated by a black bar, and was applied to samples in the remaining differentiation grades. The resulting classifiers classified pre- and post-differentiation grades of the samples with 92% (46/50) (a), 98% (48/49) (b), 84% (16/19) (c), and 100% (18/18) (d) accuracies, respectively.

3.3. Significance of selected genes

Assessment of the 61 samples with the 40 genes with the greatest differential expression at the L0 to L1, L1 to G1, G1 to G2, and G2 to G3 transitions clearly discriminated samples staged before and after each transition (Fig. 1(e)–(h)). Thus, the 40 most discriminatory genes for each transition discriminated pre-transition samples from post-transition samples. We further examined the discriminative power of the 40 genes for each transition by the minimum distance classifier. In each transition, the minimum distance classifier was constructed with samples from two consecutive differentiation grades (indicated by the black bar in Fig. 2), and it was applied to samples in the remaining differentiation grades. The classifier designed locally with L0 and L1 samples, whereas it classified correctly

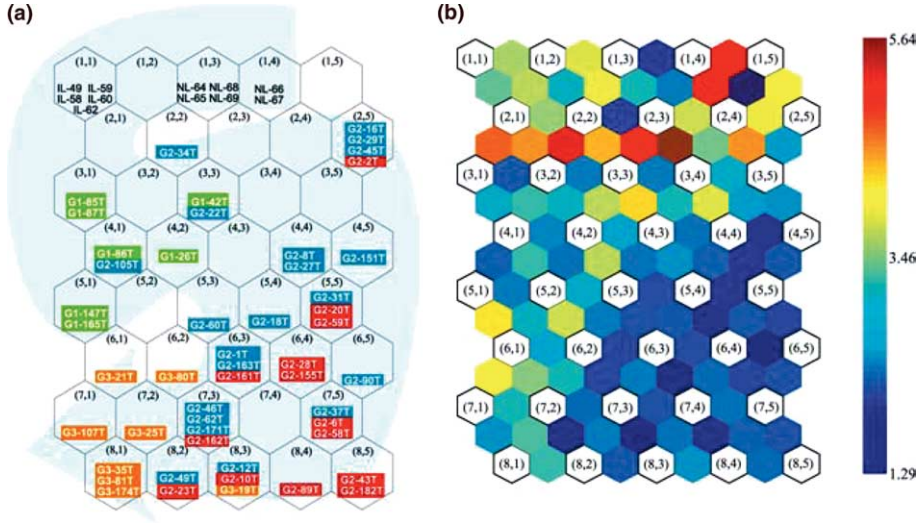

Fig. 3. Visualization of sample arrangement by SOM (a) Clusters of samples. Each cell in the SOM grid corresponds to one cluster. The vectors of neighboring cells are typically near each other. (*m*, *n*), index of the cell located at the *m*th row and *n*th column. NL-XX, samples from non-tumorous livers without HCV infection (L0); IL-XX, samples from HCV-infected non-tumorous livers (L1); G1-XXT, samples from well differentiated tumors (G1); G2-XXT, samples from moderately differentiated tumors (G2); G3-XXT, samples from poorly differentiated tumors (G3). The samples are clearly arranged on a sigmoidal curve in the order L0, L1, G1, G2, and G3. Note that G2 tumors without venous invasion (blue) are located close to G1 tumors and that G2 tumors with venous invasion (red) are located close to G3 tumors. (b) Distance between neighboring clusters. (*m*, *n*), index of the cell located at the *m*th row and *n*th column. The color of the cells indicates the distance between the neighboring clusters, and red indicates a long distance. Red cells in the upper area clearly show that non-tumorous liver and tumor samples are relatively far apart on all of the 40 most discriminatory genes.

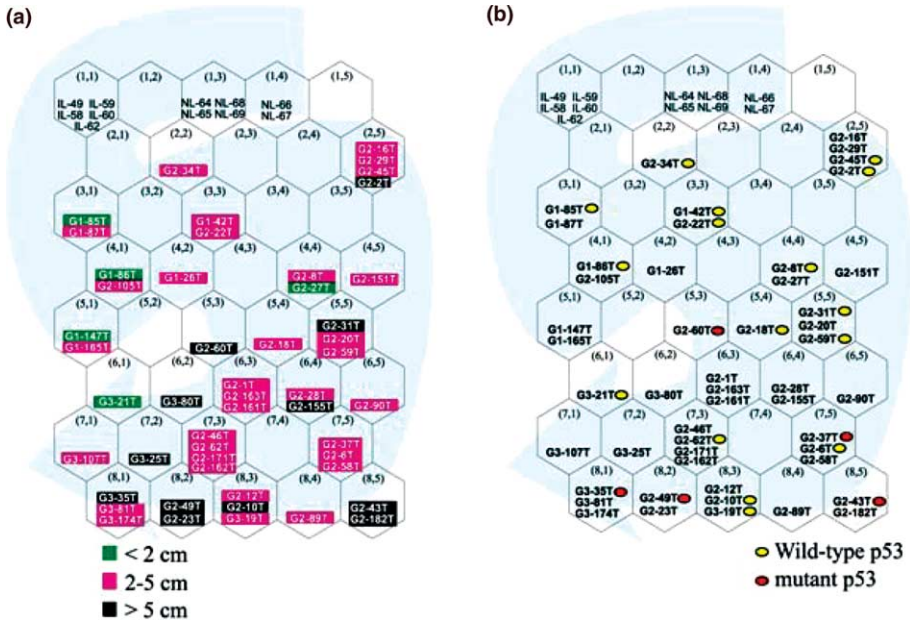

Fig. 4. Tumor size and p53 status on sigmoidal orbit of HCC development. (a) Tumor size. This shows the reproducible relation between tumor size and differentiation grade in most samples. Note that there are aberrant cases with small tumor size (G2-27T and G3-21T) during development. (b) P53 status. When p53 data [17] from 22 HCCs were applied to the sigmoidal curve, most HCCs with wild-type p53 were located within or close to G1 cluster on the sigmoidal curve. In contrast, four of five HCCs with p53 mutation were located at the most distant point from the three clusters L0, L1, and G1, verifying our hypothesis that this disease develops sequentially from L0 to L1 to G1 to G2 to G3. NL-XX, samples from non-tumorous livers without HCV infection (L0); IL-XX, samples from HCV-infected non-tumorous livers (L1); G1-XXT, samples from well differentiated tumors (G1); G2-XXT, samples from moderately differentiated tumors (G2); G3-XXT, samples from poorly differentiated tumors (G3).

46 of 50 (92%) remaining samples in G1, G2, and G3 (Fig. 2(a)). The classifiers designed with L1 and G1 samples, G1 and G2 samples, and G2 and G3 samples classified the remaining samples with 98% (48/49), 84% (16/19), and 100% (18/18) accuracies, respectively (Fig. 2(b)–(d)). Thus, the 40 selective genes represent molecular signatures for each transition during hypothesized development (Fig. 2).

3.4. Arrangement of all samples by SOM

We analyzed our data by SOM [18] with the use of all 40 selective genes. The SOM correctly arranged the five clusters on a sigmoidal curve in the order L0, L1, G1, G2, and G3 (Fig. 3(a) and (b)). Strikingly, 11 (52%) of 21 G2 tumors without venous invasion were located closer to the G1 cluster, and 13 (92%) of 14 G2 tumors with venous invasion were located closer to the G3 cluster ($P = 0.001$ by Fisher's exact test) (Fig. 3(a)). Thus, the SOM classified G2 tumors into two subtypes: tumors with venous invasion and those without. Given the finding that tumor sizes in G2 and G3 were significantly larger than that in G1 (Table 1), we assigned tumor size to the samples on the sigmoidal curve (Fig. 4(a)). This reproduced the relation between tumor size and differentiation grade in most cases; however, tumor size was not always consistent with dedifferentiation. Rather, even if they belonged to G2 or G3 HCCs, small tumors (G2-27T and G3-21T) were located closer to the G1 cluster (Fig. 4(a)). Thus, our sample rearrangement detected aberrant samples.

We next applied p53 abnormality data [17] in 22 of the HCCs to the sigmoidal curve (Fig. 4(b)). Many HCCs with wild-type p53 were located within or close to G1 cluster. In contrast, four of five HCCs with p53 mutation were located at the most distant point from the three clusters L0, L1, and G1. This result was consistent with a previous report that p53 abnormality is frequent in HCC at late stage [9]. Collectively, these results verify our hypothesis that this disease develops sequentially from L0 to L1 to G1 to G2 to G3.

3.5. Validation of microarray data by quantitative RT-PCR

To validate the microarray data, we randomly selected 4 discriminatory genes for each transition and analyzed them by real-time quantitative RT-PCR. The expression patterns of *CD74*, *IGFBP3*, *STAT1*, and *TGFB1* by microarray were reproduced by real-time quantitative RT-PCR (supplementary figures S1 and S2).

4. Discussion

Since the introduction of DNA microarray technology [23,24], the patterns of gene expression associated with HCC have been described [14–17,25–32]. Among these studies, four identified genes related to dedifferentiation of HCC [26–29], and one identified molecular markers specific for HCV-related HCC [30]. More recently, we have used molecular profiling to identify a new class of HCC according to metastatic potentials [16]. However, even this state-of-the-art technology does not address the molecular basis underlying the development of HCC.

According to histologic features at biopsy, Poynard et al. [33] investigated liver fibrosis progression after HCV infection. Their elegant work provided insights into the mechanism of HCV-related liver fibrosis, which is closely related to hepatocarcinogenesis. Unfortunately, they did not investigate

features of developing HCC. Comparing HCC at biopsy and at autopsy from the same patients, Sugihara et al. [12] showed that 9 of 12 tumors, which had been well differentiated at biopsy, developed into moderately differentiated tumors at autopsy. With the use of resected HCC specimens, Nakashima et al. [11] proposed that well differentiated HCCs become progressively less differentiated as they enlarge. Nodule-in-nodule-type HCC may provide a model for the development of HCC. However, these phenomena do not necessarily explain all aspects of the development of HCC. Thus, a new framework is necessary to better understand the actual development of HCC. We profiled gene expression patterns in a population in which each member was at a distinct stage of differentiation and characterized molecular differences between distinct classes. In this viewpoint, there may be a limitation of our present study in understanding the actual development. Nevertheless, the arrangement of samples by SOM was consistent with classical development-related parameters and supports the hypothesis that this disease develops sequentially from L0 to L1 to G1 to G2 to G3.

SOMs with gene expression data can calculate the distance between samples [18]. Our result is intriguing: G2 tumors without venous invasion were located closer to the G1 cluster, and G2 tumors with venous invasion were located closer to the G3 cluster. This result is consistent with the increased invasiveness observed during the progression of many malignant tumors [34]. Genetic abnormality of p53 is a feature of HCC at the late stage [9]. In support of this, we found that most HCCs with *p53* abnormality were markedly advanced in the sigmoidal curve. Moreover, the sigmoidal orbit showed that HCCs become progressively less differentiated as they enlarge. This result reproduces the development of HCC reported by Nakashima et al. [11]. Interestingly, two G2 and G3 HCCs (G2-27T and G3-21T), whose sizes were <2 cm in diameter, were located closer to the G1 cluster (Fig. 4(a)). Thus, these two may be aberrant cases that skipped the normal development scenario.

Another striking finding of our study is that the 40 most discriminatory genes for each transition clearly divided all samples into pre-transition and post-transition stages (Fig. 1(e)–(h)). This finding was verified mathematically (Fig. 2). In addition, the finding that there was almost no overlap among the discriminatory genes indicates that their biological functions are specific for each aspect of development from L0 to L1 to G1 to G2 to G3. This finding is unique among microarray studies [26–29] that have identified many genes related to the dedifferentiation of HCC. Such an altered level of gene expression in specific stages may support the multi-step transformation theory that was proposed initially by Vogelstein et al. [35]. In comparison with those studies [26–29], there were a few genes that overlapped with genes identified in the present study. This discrepancy is not surprising and may be attributable in part to differences in sample background and type of microarray and algorithm used. The genes identified in this study represent pathways that are common to each aspect of development of HCV antibody-positive HCC. For example, levels of expression of many immune response-related genes, including MHC class I family members (*HLA-A*, *-C*, *-E*, and *-F*), MHC class II family (*HLA DPB1* and *HLA-DRA*), *CD74*, *NK4*, and *IFI30*, were increased in L1 compared to L0. This is reasonable considering that HCV infection causes chronic inflammation [3,4]. Additionally, high levels of these MHC class I or II genes were found in HCV-related cirrhotic liver [36].

In the present study, expression of many oncogenesis-related genes (*BNIP3L*, *FOS*, *MAF*, *IGFBP3*, and *IGFBP4*) was downregulated from L1 to G1. Insulin-like growth factor binding protein 3 (IGFBP3) induces apoptosis of some types of cancer cells, and IGFBP4 acts as an inhibitor of IGF-induced cell proliferation. A previous microarray study showed that levels of *IGFBP3* and *IGFBP4* transcripts were decreased in HCC compared to those in non-tumorous liver [27]. Chuma et al. [29] showed that *IGFBP3* levels were reduced in the early component of nodule-in-nodule-type HCC compared to those in non-tumorous liver. Thus, the IGF pathway may play an important role in well differentiated HCC arising from HCV infection. We also found upregulation of *ATOX1* in G1 HCC. On the basis of the result of a previous report [30], this finding is likely to be specific to HCV-related HCC.

We found that in dedifferentiation of G1 to G2, the most striking event was downregulation of expression of IFN-related genes (*OAS2*, *STAT1*, *PSME1*, *ISGF3G*, and *PSMB9*). This result is intriguing considering our previous observation that IFN-related genes are involved in the pathogenesis of HCV-related HCC and not HBV-associated HCC [14]. Upregulation of *STAT1* expression in HCC cell lines was observed during differentiation induced by sodium butyrate [37]. Interestingly, it was shown that HCV Nonstructural 5A (NS5A) protein attenuated inducible expressions of IFN-related genes including *STAT1* [38]. These reports suggest that these genes listed here play a specific role in the pathogenesis of HCV-related HCC. Downregulation of expression of these IFN-related genes also reflects decreased immune response. This concept is supported by a previous study [27] showing that several immune response-related genes were repressed in G2 and G3 HCCs compared with G1 HCC.

We observed upregulation of *SPINT1*, *LGALS9*, and *TGFB1* and downregulation of *SDC1* in G3 compared with G2. LGALS9 is member of the lectin family, which is involved in cell adhesion, cell growth regulation, immunomodulation, apoptosis, and metastasis. Several galectins are thought to be involved in cancer cell adhesion [39]. It has been reported that TGFB1 triggers invasiveness of HCC cells via α3β1 integrin [40]. Matsumoto et al. [41] reported rdecreased expression of SDC1 in HCC with high metastatic potential. Given the present finding that venous invasion is found in G2 HCC, altered levels of these metastasis-related genes would increase further the metastatic potential of G3 HCC and would provide additional molecular targets for HCC treatment.

Our present study focused on HCC with positive HCV serology. The identified discriminatory genes were consistent with the molecular patterns [30,36] of HCV-infected liver disease. However, we could not exclude the possibility of an occult HBV infection because of a lack of the data for hepatitis B virus core antibody and HBV-DNA in our cohort. This dilemma prompted us to examine how the discriminatory genes are involved in HCC with negative HCV serology. We found that the SOM with the discriminatory gene set failed to arrange correctly the samples of 26 HCCs with negative HCV serology (supplementary figure S3). This striking finding suggests that changes in our identified discriminatory genes are specific for HCC with positive HCV serology.

Because the prognosis of HCC is extremely poor even when curative surgery is performed [2,3,16], the greatest impact on this disease will be prevention. A primary strategy of prevention of transfusion-related HCV infection has almost been achieved, and a current focus is to prevent the development of HCC in HCV infection. Currently, oilgonucleotide arrays representing the whole known genes (about 38 000 genes) are available [42]. In this regard, our profiling data may be less informative; however, these provide additional biomarkers and molecular targets for the prevention, diagnosis, and treatment of this disease. The sigmoidal orbit constructed here may provide a framework to explain the development of HCC with positive HCV serology.

Acknowledgements: We thank Prof. T Ishihara and Prof. K Sasaki for pathologic diagnosis.

Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version at doi:10.1016/j.febslet.2004.10.113.

References

[1] El-Serag, H.B. and Mason, A.C. (1999) Rising incidence of hepatocellular carcinoma in the United States. N. Engl. J. Med. 340, 745–750.
[2] Schafer, D.F. and Sorrell, M.F. (1999) Hepatocellular carcinoma. Lancet 353, 1253–1257.
[3] Llovet, J.M., Burroughs, A. and Bruix, J. (2003) Hepatocellular carcinoma. Lancet 362, 1907–1917.
[4] De Mitri, M.S., Pisi, E., Poussin, K., Paterlini, P., Bréchot, C., Baccarini, P., D'Errico, A., Grigiani, W., Alberti, A., Pontisso, P., Simon, N. and Beaugrand, M. (1995) HCV-associated liver cancer without cirrhosis. Lancet 345, 413–415.
[5] Thorgeirsson, S.S. and Grisham, J.W. (2002) Molecular pathogenesis of human hepatocellular carcinoma. Nat. Genet. 31, 339–346.
[6] Kim, C.M., Koike, K., Saito, I., Miyamura, T. and Jay, G. (1991) HBx gene of hepatitis B virus induces liver cancer in transgenic mice. Nature 351, 317–320.
[7] Moriya, K., Fujie, H., Shintani, Y., Yotsuyanagi, H., Tsutsumi, T., Ishibashi, K., Matsuura, Y., Kimura, S., Miyamura, T. and Koike, K. (1998) The core protein of hepatitis C virus induces hepatocellular carcinoma in transgenic mice. Nat. Med. 4, 1065–1067.
[8] Satoh, S., Daigo, Y., Furukawa, Y., Kato, T., Miwa, N., Nishiwaki, T., Kawasoe, T., Ishiguro, H., Fujita, M., Tokino, T., Sasaki, Y., Imaoka, S., Murata, M., Shimano, T., Yamaoka, Y. and Nakamura, Y. (2000) AXIN1 mutations in hepatocellular carcinomas, and growth suppression in cancer cells by virus-mediated transfer of AXIN1. Nat. Genet. 24, 245–250.
[9] Oda, T., Tsuda, H., Scarpa, A., Sakamoto, M. and Hirohashi, S. (1992) p53 gene mutation spectrum in hepatocellular carcinoma. Cancer Res. 52, 6358–6364.
[10] Tsuda, H., Zhang, W.D., Shimosato, Y., Yokota, J., Terada, M., Sugimura, T., Miyamura, T. and Hirohashi, S. (1990) Allele loss on chromosome 16 associated with progression of human hepatocellular carcinoma. Proc. Natl. Acad. Sci. USA 87, 6791–6794.
[11] Nakashima, O., Sugihara, S., Kage, M. and Kojiro, M. (1995) Pathomorphologic characteristics of small hepatocellular carcinoma: a special reference to small hepatocellular carcinoma with indistinct margins. Hepatology 22, 101–105.
[12] Sugihara, S., Nakashima, O., Kojiro, M., Majima, Y., Tanaka, M. and Tanikawa, K. (1992) The morphologic transition in hepatocellular carcinoma. A comparison of the individual histologic features disclosed by ultrasound-guided fine-needle biopsy with those of autopsy. Cancer 70, 1488–1492.
[13] Kojiro, M. (2002) Pathological evolution of early hepatocellular carcinoma. Oncology 62, 43–47.
[14] Iizuka, N., Oka, M., Yamada-Okabe, H., Mori, N., Tamesa, T., Okada, T., Takemoto, N., Tangoku, A., Hamada, K., Nakayama, H., Miyamoto, T., Uchimura, S. and Hamamoto, Y. (2002) Comparison of gene expression profiles between hepatitis B virus- and hepatitis C virus-infected hepatocellular carcinoma by oligonucleotide microarray data based on a supervised learning method. Cancer Res. 62, 3939–3944.
[15] Iizuka, N., Oka, M., Yamada-Okabe, H., Mori, N., Tamesa, T., Okada, T., Takemoto, N., Hashimoto, K., Tangoku, A., Hamada, K., Nakayama, H., Miyamoto, T., Uchimura, S. and Hamamoto, Y. (2003) Differential gene expression in distinct virologic types of hepatocellular carcinoma: association with liver cirrhosis. Oncogene 22, 3007–3014.
[16] Iizuka, N., Oka, M., Yamada-Okabe, H., Nishida, M., Maeda, Y., Mori, N., Takao, T., Tamesa, T., Tangoku, A., Tabuchi, H., Hamada, K., Nakayama, H., Ishitsuka, H., Miyamoto, T., Hirabayashi, A., Uchimura, S. and Hamamoto, Y. (2003) Oligonucleotide microarray for prediction of early intrahepatic recurrence of hepatocellular carcinoma after curative resection. Lancet 361, 923–929.
[17] Okada, T., Iizuka, N., Yamada-Okabe, H., Mori, N., Tamesa, T., Takemoto, N., Tangoku, A., Hamada, K., Nakayama, H., Miyamoto, T., Uchimura, S., Hamamoto, Y. and Oka, M. (2003) Gene expression profile linked to *p53* status in hepatitis C virus-related hepatocellular carcinoma. FEBS Lett. 555, 583–590.
[18] Tamayo, P., Slonim, D., Mesirov, J., Zhu, Q., Kitareewan, S., Dmitrovsky, E., Lander, E.S. and Golub, T.R. (1999) Interpreting patterns of gene expression with self-organizing maps: methods and application to hematopoietic differentiation. Proc. Natl. Acad. Sci. USA 96, 2907–2912.
[19] Toronen, P., Kolehmainen, M., Wong, G. and Castren, E. (1999) Analysis of gene expression data using self-organizing maps. FEBS Lett. 451, 142–146.
[20] Sobin, L.H. and Wittekind, C. (2002) TNM classification of Malignant Tumours, sixth ed, UICC, Wiley-Liss, 81–83.
[21] Iizuka, N., Oka, M., Yamamoto, K., Tangoku, A., Miyamoto, K., Miyamoto, T., Uchimura, S., Hamamoto, Y. and Okita, K. (2003) Identification of common or distinct genes related to antitumor activities of a medicinal herb and its major component by oligonucleotide microarray. Int. J. Cancer 107, 666–672.
[22] Tusher, V.G., Tibshirani, R. and Chu, G. (2001) Significance analysis of microarrays applied to the ionizing radiation response. Proc. Natl. Acad. Sci. USA 98, 5116–5121.
[23] Schena, M., Shalon, D., Davis, R.W. and Brown, P.O. (1995) Quantitative monitoring of gene expression patterns with a complementary DNA microarray. Science 270, 467–470.
[24] DeRisi, J., Penland, L., Brown, P.O., Bittner, M.L., Meltzer, P.S., Ray, M., Chen, Y., Su, Y.A. and Trent, J.M. (1996) Use of a cDNA microarray to analyse gene expression patterns in human cancer. Nat. Genet. 14, 457–460.
[25] Kim, J.W. and Wang, X.W. (2003) Gene expression profiling of preneoplastic liver disease and liver cancer: a new era for improved early detection and treatment of these deadly diseases?. Carcinogenesis 24, 363–369.
[26] Shirota, Y., Kaneko, S., Honda, M., Kawai, H.F. and Kobayashi, K. (2001) Identification of differentially expressed genes in hepatocellular carcinoma with cDNA microarrays. Hepatology 33, 832–840.
[27] Okabe, H., Satoh, S., Kato, T., Kitahara, O., Yanagawa, R., Yamaoka, Y., Tsunoda, T., Furukawa, Y. and Nakamura, Y. (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, 2129–2137.
[28] Midorikawa, Y., Tsutsumi, S., Taniguchi, H., Ishii, M., Kobune, Y., Kodama, T., Makuuchi, M. and Aburatani, H. (2002) Identification of genes associated with dedifferentiation of hepatocellular carcinoma with expression profiling analysis. Jpn. J. Cancer Res. 93, 636–643.

[29] Chuma, M., Sakamoto, M., Yamazaki, K., Ohta, T., Ohki, M., Asaka, M. and Hirohashi, S. (2003) Expression profiling in multistage hepatocarcinogenesis: identification of HSP70 as a molecular marker of early hepatocellular carcinoma. Hepatology 37, 198–207.
[30] Smith, M.W., Yue, Z.N., Geiss, G.K., Sadovnikova, N.Y., Carter, V.S., Boix, L., Lazaro, C.A., Rosenberg, G.B., Bumgarner, R.E., Fausto, N., Bruix, J. and Katze, M.G. (2003) Identification of novel tumor markers in hepatitis C virus-associated hepatocellular carcinoma. Cancer Res. 63, 664–859.
[31] Chen, X., Cheung, S.T., So, S., Fan, S.T., Barry, C., Higgins, J., Lai, K.M., Ji, J., Dudoit, S., Ng, I.O., Van De Rijn, M., Botstein, D. and Brown, P.O. (2002) Gene expression patterns in human liver cancers. Mol. Biol. Cell 13, 1929–1939.
[32] Delpuech, O., Trabut, J.B., Carnot, F., Feuillard, J., Brechot, C. and Kremsdorf, D. (2002) dentification, using cDNA macroarray analysis, of distinct gene expression profiles associated with pathological and virological features of hepatocellular carcinoma. Oncogene 21, 2926–2937.
[33] Poynard, T., Bedossa, P. and Opolon, P. (1997) atural history of liver fibrosis progression in patients with chronic hepatitis C. The OBSVIRC, METAVIR, CLINIVIR, and DOSVIRC groups. Lancet 349, 825–832.
[34] Kohn, E.C. and Liotta, L.A. (1995) Molecular insights into cancer invasion: strategies for prevention and intervention. Cancer Res. 55, 1856–1862.
[35] Vogelstein, B., Fearon, E.R., Hamilton, S.R., Kern, S.E., Preisinger, A.C., Leppert, M., Nakamura, Y., White, R., Smits, A.M. and Bos, J.L. (1988) Genetic alterations during colorectal-tumor development. N. Engl. J. Med. 319, 525–532.
[36] Smith, M.W., Yue, Z.N., Korth, M.J., Do, H.A., Boix, L., Fausto, N., Bruix, J., Carithers Jr., R.L. and Katze, M.G. (2003) Hepatitis C virus and liver disease: global transcriptional profiling and identification of potential markers. Hepatology 38, 1458–1467.
[37] Hung, W.C. and Chuang, L.Y. (1999) Sodium butyrate enhances STAT 1 expression in PLC/PRF/5 hepatoma cells and augments their responsiveness to interferon-alpha. Br. J. Cancer 80, 705–710.
[38] Geiss, G.K., Carter, V.S., He, Y., Kwieciszewski, B.K., Holzman, T., Korth, M.J., Lazaro, C.A., Fausto, N., Bumgarner, R.E. and Katze, M.G. (2003) Gene expression profiling of the cellular transcriptional network regulated by alpha/beta interferon and its partial attenuation by the hepatitis C virus nonstructural 5A protein. J. Virol. 77, 6367–6375.
[39] Ohannesian, D.W., Lotan, D., Thomas, P., Jessup, J.M., Fukuda, M., Gabius, H.J. and Lotan, R. (1995) Carcino-embryonic antigen and other glycoconjugates act as ligands for galectin-3 in human colon carcinoma cells. Cancer Res. 55, 2191–2199.
[40] Giannelli, G., Fransvea, E., Marinosci, F., Bergamini, C., Colucci, S., Schiraldi, O. and Antonaci, S. (2002) Transforming growth factor-beta1 triggers hepatocellular carcinoma invasiveness via alpha3beta1 integrin. Am. J. Pathol. 16, 183–193.
[41] Matsumoto, A., Ono, M., Fujimoto, Y., Gallo, R.L., Bernfield, M. and Kohgo, Y. (1997) Reduced expression of syndecan-1 in human hepatocellular carcinoma with high metastatic potential. Int. J. Cancer 74, 482–491.
[42] Llovet, J.M. and Wurmbach, E. (2004) Gene expression profiles in hepatocellular carcinoma: not yet there. J. Hepatol. 41, 336–339.