Resampling Based on Geographic Patterns of Hepatitis Virus Infection Reveals a Common Gene Signature for Early Intrahepatic Recurrence of Hepatocellular Carcinoma

SHUNJI UCHIMURA[1], NORIO IIZUKA[2,3], TAKAO TAMESA[2], TAKANOBU MIYAMOTO[1], YOSHIHIKO HAMAMOTO[1] and MASAAKI OKA[2*]

[1]Department of Computer Science and Systems Engineering, Faculty of Engineering, Yamaguchi University, Yamaguchi; Departments of [2]Surgery II and [3]Complementary Medicine, Yamaguchi University Graduate School of Medicine, Ube, Yamaguchi 755-8505, Japan

Abstract. *The majority of hepatocellular carcinomas (HCCs) correlated with infection by either hepatitis B or C virus (HBV or HCV) showing various geographic distributions, making it impossible to identify common gene signatures responsible for HCC recurrence. In this study we performed in silico resampling analysis of DNA microarray data that can reproduce virtually the geographic distribution pattern of HBV and HCV in 6 representative geographic regions. With the use of the Fisher ratio, genes associated with early intrahepatic recurrence of HCC within 1 year of surgery were identified in the 6 geographic virtual cohorts, each consisting of 1,000 virtual samples. The top 100 genes among each virtual cohort were compared. Many human leukocyte antigen (HLA) family genes were common among the 6 geographic virtual cohorts, suggesting that this gene family represents the pathway most responsible for early intrahepatic recurrence of HCC worldwide. This resampling approach is useful for identifying common pathways involved in various aspects of HCC.*

Significant progress has been made in the surgical treatment of hepatocellular carcinoma (HCC) over the last decade and the 5-year survival in well selected patients with resectable HCC is around 50% (1-3). Nevertheless, HCC is still one of the most fatal types of cancer because of its high rate of recurrence (1-3). Thus, unnecessary treatment of patients who can be cured by surgery alone must be avoided and patients must be provided with more personalized therapeutic options to improve the poor outcome. To address these two problems, it is critical to understand the molecular basis of HCC recurrence better.

Several gene-profiling studies have identified molecular signatures associated with HCC metastatic potential (4-7). However, few common genes were identified in these studies (4-7). This may be due to the use of differing microarrays and gene selection procedures (*i.e.* algorithms) or to the use of HCCs of different viral origins, *i.e.* hepatitis B or C virus (HBV or HCV), which are major causes of human HCC (8-10). These two viruses have different worldwide distribution patterns (11). Given that there are significant differences in gene expression patterns between HBV- and HCV-infected HCCs (12), sample bias may be caused by different infection patterns of HBV and HCV (13). To resolve these issues, it is critical to identify common genes or pathways in the recurrence of HCC, taking into consideration these infection patterns.

To consider sample variation and to address the small sample size of DNA microarray data, Michiels *et al*. developed an elegant gene selection approach involving resampling (14). This approach is able to create various virtual samples in selecting feature genes; however, it does not consider sample variation due to geographic distribution of more than one agent, for example, HBV and HCV. To refine this resampling method, a geographic bias-based resampling approach was developed. With the use of virtual cohorts, we examined how geographic bias of the frequency of HBV and HCV affected identification of early intrahepatic recurrence (IHR)-related genes in HCC and which gene families or pathways were most common.

Materials and Methods

The expression levels of approximately 6,000 genes in human HCC samples have been analyzed using high-density oligonucleotide array (HuGeneFL Array, Affymetrix, Inc., Santa Clara, CA, USA) (5, 12, 15-17). In the present study, DNA microarray data (http://surgery2.med. yamaguchi-u.ac.jp/research/DNAchip/) from

Correspondence to: M. Oka, Department of Surgery II, Yamaguchi University Graduate School of Medicine, 1-1-1 Minami-Kogushi, Ube, Yamaguchi 755-8505, Japan. Fax: +81836222262, e-mail: 2geka-1@po.cc.yamaguchi-u.ac.jp

Key Words: DNA microarray, bioinformatics, resampling, hepatocellular carcinoma.

Table I. *Geographic patterns of each virus-type HCC and assigned sample in virtual cohort.*

Actual distribution

Geographic area	HBs Ag (+)	HCV Ab (+)	Negative for both (non B, non C)
Europe	6.2%-20.6%	6.1%-79.2%	11.5%-39.3%
Greece	59.5%	12.3%	24.9%
United States	23.6%	9.8%	47.1%
Asia	58.1%-68.0%	0.5%-15.3%	10.0%-29.0%
Japan	10%	82.9%	5.4%
Africa	44.6%	16.9%	29.9%
Our institue (Japan, Yamaguchi)	11/60 (18.3%)	40/60 (66.7%)	9/60 (15%)

Proportion of 16 samples assigned to virtual cohort out of our 60 HCC samples

Geographic area	HBs Ag (+)	HCV Ab (+)	Negative for both (non B, non C)
Europe	3 (18.7%)	8 (50%)	5 (31.3%)
Greece	10 (62.5%)	2 (12.5%)	4 (25%)
United States	5 (31.3%)	2 (12.5%)	9 (56.2%)
Asia	10 (62.4%)	1 (6.3%)	5 (31.3%)
Japan	2 (12.5%)	13 (81.2%)	1 (6.3%)
Africa	8 (50%)	3 (18.7%)	5 (31.3%)

The above distribution data were obtained from Iizuka *et al.* (12). HBs Ag, hepatitis B surface antigen; HCV Ab, hepatitis C virus antibody.

60 HCC samples was used to predict early IHR, according to Brazma *et al*. (18). Among the 60 HCC samples, early IHR was found in 20, which included 4 HBV-infected, 13 HCV-infected and 3 samples negative for both. No recurrence within 1 year of surgery was found in the remaining 40 HCCs, which included 7 HBV-infected, 27 HCV-infected and 6 samples negative for both.

From a pool of approximately 6,000 genes, all genes with mean average differences (Ads, arbitrary units from Affymetrix) of greater than 20 were selected from the 60 HCCs. This filtering step yielded 4,493 genes. The geographic distribution pattern data (11) for HCCs positive for either HBV or HCV, or negative for both were used (Table I). Artificial resampling cohorts were created *in silico* for 6 geographic regions (*i.e.* Europe, Greece, United States, Asia, Japan and Africa) for which data have been presented by Bosch *et al*. (11). The following procedure was repeated 1,000 times: (i) 16 HCCs were randomly selected from the 60 HCCs according to the proportion of HCC positive for HBV, HCC positive for HCV and HCC negative for both of each region as shown in Table I (*e.g.* in European-type virtual cohort, 3 from the 11 HCCs positive for HBV, 8 from the 40 HCCs positive for HCV and 5 from the 9 HCCs negative for both were respectively selected); (ii) The 4,493 filtered genes were ranked in order of decreasing magnitude of Fisher ratios among the 16 samples, as described elsewhere (19); (iii) Genes showing statistically significant ($p<0.01$) differences in expression between the early IHR and no recurrence groups were selected from the 4,493 genes with the use of a random permutation test of the Fisher ratio (15-17, 19). Finally, the top 100 genes that appeared frequently throughout the 1,000 trials were identified.

Results

The actual distribution of HCCs with distinct hepatitis virus infections and the number of samples assigned to each group (HCC positive for HBV or HCV, or HCC negative for both) in each virtual cohort are shown in Table I. We selected the top 100 genes that appeared most frequently in the 1,000 trials in each virtual cohort (Supplementary Tables). The frequency of the top 100 early IHR-related genes in the European-type virtual cohort ranged from 94 to 431 per 1,000. The frequencies of the top 100 genes in the Greek-, United States-, Asian-, Japanese- and African-type virtual cohorts were 52-504, 144-779, 48-642, 100-423 and 71-419 per 1000, respectively (supplementary Tables). No genes were listed in more than half of the 1,000 trials in the virtual cohorts of European-, Japanese- and African-type HCCs. This result differed from that of a previous study (14), which used an appearance rate of greater than 50% in virtual cohorts; we have previously reported on this difference (20). Therefore, in the present study, we focused on genes that were ranked in the top 100 in order of ability to distinguish between the early IHR and no recurrence groups.

We found 7 genes commonly listed among the 6 geographic virtual cohorts, 3 (42.9%) of which were HLA family genes (*HLA-DRB1, HLA-DRA* and *HLA-DPB1*). Expression levels of all 7 genes were lower (0.67- to 0.41-fold) in HCC with early IHR than in HCC without recurrence (Table II). We then searched for common genes among N (N=2-5) virtual cohorts (Tables III-VI). Total numbers of common genes were 54 in the 7 combinations of 2 virtual cohorts (Table III), 50 in the 5 combinations of 3 virtual cohorts (Table IV), 26 in the 5 combinations of 4 virtual cohorts (Table V) and 17 in the 3 combinations of 5 virtual cohorts (Table VI). Two (3.7%), 5 (10.0%), 3 (11.5%) and 3 (17.6%) of the genes common to the 2 to 5 virtual cohorts, respectively, were HLA family genes. The relative appearance of HLA family genes increased with the number of virtual cohorts examined ($p=0.02$ by Fisher's exact test, Table VII). Notably, expression levels of 148 out of a total of 154 common genes (96.1%) were lower in HCC with early IHR than in HCC without recurrence.

Discussion

DNA microarray analysis allows for the elucidation of simultaneous regulation of thousands of genes that participate in cancer biology (21). Recent advances have yielded many microarray-based systems for outcome prediction (5-7, 22, 23). However, issues remain with respect to the clinical use of this

Table II. *Genes common to 6 virtual cohorts.*

Gene accession number	Symbol (or description)	Function	Level in recurrence group vs. non-recurrence group	Fold change (recurrence/non-recurrence)*					
				Europe	Greece	US	Asia	Japan	Africa
Y10032	SGK	Signal transduction	Reduced	0.50±0.16	0.55±0.10	0.48±0.07	0.55±0.09	0.53±0.20	0.54±0.12
M33600	**HLA-DRB1**	Immune system	Reduced	0.46±0.15	0.61±0.11	0.41±0.12	0.60±0.10	0.53±0.16	0.57±0.13
X00274	**HLA-DRA**	Immune system	Reduced	0.49±0.13	0.58±0.10	0.41±0.10	0.56±0.09	0.60±0.15	0.55±0.12
X16663	HCLS1	Signal transduction	Reduced	0.55±0.13	0.67±0.09	0.53±0.08	0.66±0.08	0.60±0.15	0.64±0.11
AB000409	MKNK1	Signal transduction	Reduced	0.49±0.14	0.58±0.09	0.46±0.10	0.56±0.07	0.54±0.16	0.55±0.11
M17733	TMSB4X	Immune system	Reduced	0.67±0.13	0.62±0.07	0.64±0.09	0.62±0.07	0.68±0.15	0.64±0.10
X03100	**HLA-DPB1**	Immune system	Reduced	0.48±0.10	0.56±0.08	0.43±0.06	0.54±0.07	0.56±0.12	0.53±0.09

Bold, HLA family genes. *Values are mean ± SD of fold change in individual 6-type virtual cohorts (Europe, Greece, US, Asia, Japan and Africa) (n=1000).

Table III. *Genes common to 5 virtual cohorts.*

Gene accession number	Symbol (or description)	Function	Level in recurrence group *vs*. non-recurrence group
[Japan, Africa, Asia, Europe, Greece]			
HG2755-HT2862	PLS3	Cell motility and extracellular matrix	Reduced
U89336	**HLA class III region containing 9 genes**	Unknown	Reduced
U15642	E2F5	Transcription	Reduced
[Japan, Africa, Asia, Europe, US]			
M34996	**HLA-DQA**	Immune system	Reduced
[Africa, Asia, Europe, Greece, US]			
D87465	KIAA0275 gene	Unknown	Reduced
L41870	RB1	Cell growth and apoptosis	Reduced
M27891	CST3	Signal transduction	Reduced
M69066	MSN	Cell motility and extracellular matrix	Reduced
M91029	AMPD2	Metabolism	Reduced
U48405	GPR68	Signal transduction	Reduced
U60115	FHL1	Cell growth and apoptosis	Reduced
U97105	N2A3 mRNA	Similar to Collapsin response mediator protein 2	Reduced
X76538	MPV17	Unknown	Reduced
X83416	PRNP	Prion protein	Reduced
U33838	NF-kappa-B p65delta3 mRNA	Transcription	Reduced
M36284	GYPC	Integral membrane glycoprotein	Reduced
X03068	**HLA-DQB1**	Immune system	Reduced

Bold, HLA family genes.

technology. In particular, sample size is too small to select remarkable genes and to evaluate predictive performance (24). To address the former issue, Michiels *et al*. developed a unique gene selection approach involving resampling, which randomly creates sample variation and is useful for narrowing down the list of candidate genes (14). This elegant approach is useful when considering that single-pass gene selection is insufficient for the selection of robust remarkable genes for prediction (20). Human HCC has a complex etiology involving two hepatitis viruses with different worldwide distribution patterns (11). Thus, it is likely that the mathematical approach of Michiels *et al*. (14) has limitations when applied to HCC cohorts. We developed a resampling approach that is suitable for analysis of HCC and identified common pathways in early IHR of HCC in virtual cohorts created *in silico* according to actual geographic distribution patterns of hepatitis viruses.

Table IV. *Genes common to 4 virtual cohorts.*

Gene accession number	Symbol (or description)	Function	Level in recurrence group *vs.* non-recurrence group
[Japan, Africa, Asia, Greece]			
D83407	DSCR1L1	Nucleotide binding	Reduced
[Japan, Africa, Europe, US]			
U19713	AIF1	Signal transduction	Reduced
HG3576-HT3779	**MHC Class Ii Beta W52**	Immune system	Reduced
K02405	**HLA-DQB1**	Immune system	Reduced
[Africa, Asia, Europe, Greece]			
M93425	PTPN12	Signal transduction	Reduced
X15880	COL6A1	Cell motility and extracellular matrix	Reduced
X51345	JUNB	Transcription	Reduced
[Africa, Asia, Europe, US]			
M58285	HEM1	Cell motility and extracellular matrix	Reduced
Z11697	CD83	Immune system	Reduced
X04729	SERPINE1	Cell motility and extracellular matrix	Reduced
D32129	**HLA-A**	Immune system	Reduced
[Africa, Asia, Greece, US]			
J02906	CYP2F1	Metabolism	Reduced
J04456	LGALS1	Signal transduction	Reduced
L10343	PI3	Immune system	Reduced
M31013	MYH9	Cell motility and extracellular matrix	Reduced
M63167	AKT1	Signal transduction	Reduced
M76378	CSRP1	Zinc ion binding	Increased
U06452	MLANA	Immune system	Reduced
U20350	CX3CR1	Immune system	Reduced
U49973	TIGD1	Nucleotide binding	Reduced
U58970	TOMM34	Protein folding	Reduced
U90918	hypothetical protein dJ462O23.2	Unknown	Reduced
X59350	CD22	Immune system	Reduced
X69398	CD47	Immune system	Reduced
X69433	IDH2	Metabolism	Increased
U88902	*Homo sapiens* clone g10.34 integrase	Unknown	Increased

Bold, HLA family genes.

Our results show that many HLA family genes are involved in early IHR of HCC among the virtual cohorts. In particular, 3 HLA genes (*HLA-DRB1, HLA-DRA* and *HLA-DPB1*) were commonly listed in all of the 6 geographic virtual cohorts, reflecting individual virus patterns (11) of 6 geographic regions. The relative appearance of HLA genes also increased with the number of virtual cohorts tested. Thus, HLA genes are the only family of genes that can survive in virtual sample variation created by etiologic hepatitis virus infection patterns. Recent DNA microarray studies have shown reduced expression levels of HLA genes associated with high metastatic potential and/or poor prognosis in hematopoietic tumors (25) and also in several adenocarcinomas (26) and in HCC (5, 7, 16), suggesting a common function of HLA molecules in tumor progression. These reports support our present finding of lower expression levels of many HLA family genes in HCC with early IHR.

Our recent study showed that reduced protein levels of HLA-DR, an HLA antigen, by tumor cells is closely associated with early IHR of HCC in a manner independent of other clinicopathological factors (16). Although the precise mechanism remains unclear, the immunological pathway represented by the reduced expression of HLA family genes plays a central role in IHR within 1 year of surgery. This concept warrants further investigation. For example, it has been suggested that differences in autoimmunity of distinct haplotypes of HLA-DQ1 and -DR3 alleles may contribute to HCC development in patients with HBV or HCV infection (27). Thus, the distinct quality of HLA family genes, as well as the distinct quantity, should be taken into consideration. In addition, present study focused on IHR within 1 year of surgery; it is necessary to elucidate pathways critical for IHR within 2 years or more of surgery or for recurrence in organs other than the liver.

Table V. *Genes common to 3 virtual cohorts.*

Gene accession number	Symbol (or description)	Function	Level in recurrence group *vs.* non-recurrence group
[Japan, Africa, Europe]			
U00672	IL10RA	Immune system	Reduced
S68271	CREM	Transcription	Reduced
[Japan, Europe, US]			
D67029	SEC14L1	Molecule transport	Reduced
M22995	RAP1A	Signal transduction	Reduced
M59465	TNFAIP3	Immune system	Reduced
M93221	MRC1	Immune system	Reduced
U03105	PNRC1	Protein binding	Reduced
U07802	ZFP36L2	Transcription	Reduced
U15085	**HLA-DMB**	Immune system	Reduced
M97935	STAT1	Signal transduction	Reduced
M13560	**HLADG**	Immune system	Reduced
[Africa, Asia, Greece]			
AB002380	ARHGEF12	Signal transduction	Reduced
AF003743	KCNQ1	Molecule transport	Increased
D13644	USP6NL	Cell motility and extracellular matrix	Reduced
J02902	PPP2R1A	Cell motility and extracellular matrix	Reduced
L13689	PCGF4	Protein modification	Reduced
L20316	GCGR	Signal transduction	Increased
L49169	FOSB	Transcription	Reduced
M14565	CYP11A1	Metabolism	Reduced
M34668	PTPRA	Signal transduction	Increased
M80563	S100A4	Metastasis	Reduced
M84349	CD59	Immune system	Reduced
U07000	BCR	Signal transduction	Reduced
U07681	IDH3A	Metabolism	Reduced
U45285	TCIRG1	Immune system	Reduced
U51096	CDX2	Transcription	Reduced
U65932	ECM1	Cell motility and extracellular matrix	Reduced
U66075	GATA	Transcription	Reduced
U77664	RPP38	Nucleotide processing	Reduced
U89896	CSNK1G2	Signal transduction	Reduced
U92015	*Human clone 143789 defective mariner transposon Hsmar2 mRNA*		Reduced
X14787	THBS1	Metastasis	Reduced
X57025	IGF1	Cell growth and apoptosis	Reduced
X75342	SHB	Signal transduction	Reduced
X84908	PHKB	Metabolism	Reduced
Y07829	TRIM10	Protein modification	increased
U33632	KCNK1	Molecule transport	Reduced
U38291	MAP1A	Cell motility and extracellular matrix	Reduced
U60061	FEZ2	Unknown	Reduced
X15422	MBL2	Immune system	Reduced
HG2175-HT2245	MYH10	Cell motility and extracellular matrix	Reduced
HG2987-HT3136	VIP	Hormone	Reduced
HG3342-HT3519	ID1	Transcription	Reduced
X57579	INHBA	Cell growth and apoptosis	Reduced
D17427	DSC3	Cell motility and extracellular matrix	Reduced
[Africa, Asia, US]			
M64082	FMO1	Molecule transport	Reduced
[Africa, Europe, US]			
HG1872-HT1907	**Major Histocompatibility Complex, Dg**	Immune system	Reduced
L33075	IQGAP1	Signal transduction	Reduced
HG3597-HT3800	**MHC (HLA) class I-type DNA**	Immune system	Reduced
HG2915-HT3059	**HLA-E**	Immune system	Reduced

Bold, HLA family genes.

Table VI. *Genes common to 2 virtual cohorts.*

Gene accession number	Symbol (or description)	Function	Level in recurrence group *vs.* non-recurrence group
[Japan, Europe]			
D10522	MARCKS	Cell motility and extracellular matrix	Reduced
D87434	KIAA0247	Unknown	Reduced
HG1155-HT4822	CSF1	Immune system	Reduced
HG3521-HT3715	RAP1B	Signal transduction	Reduced
J04162	FCGR3B	Immune system	Reduced
L22075	GNA13	Signal transduction	Reduced
L36033	CXCL12	Immune system	Reduced
M64925	MPP1	Signal transduction	Reduced
M68941	PTPN4	Signal transduction	Reduced
M87503	ISGF3G	Immune system	Reduced
U26710	CBLB	Signal transduction	Reduced
U79294	PPAP2B	Cell motility and extracellular matrix	Reduced
X74496	PREP	Protein modification	Reduced
X75042	REL	Transcription	Reduced
X82200	TRIM22	Transcription	Reduced
HG162-HT3165	Tyrosine Kinase, Receptor Axl	Signal transduction	Reduced
L08895	MEF2C	Transcription	Reduced
HG2090-HT2152	M130 antigen extracellular variant	Immune system	Reduced
U72936	ATRX	Nucleotide binding	Reduced
HG688-HT688	**HLA-DRB2**	Immune system	Reduced
[Africa, Asia]			
K01383	MT1A	Metabolism	Reduced
X57351	IFITM2	Immune system	Reduced
M23323	CD3E	Immune system	Reduced
L08904			Reduced
[Africa, Europe]			
L09717	LAMP2	Metastasis	Reduced
M87434	OAS2	Immune system	Reduced
[Africa, Greece]			
U90878	PDLIM1	Cell motility and extracellular matrix	Reduced
[Africa, US]			
M59807	IL32	Immune system	Reduced
X77588	ARD1A	Protein modification	Reduced
[Asia, Greece]			
D87742	KIAA0268	Metabolism	Reduced
M10612	APOC2	Metabolism	Reduced
M12529	APOE	Metabolism	Reduced
M14764	NGFR	Cell growth and apoptosis	Reduced
U35246	VPS45A	Immune system	Reduced
U66083	MAGEA9	Unknown	Reduced
U83410	CUL2	Cell growth and apoptosis	Reduced
X60655	EVX1	Transcription	Reduced
D29805	B4GALT1	Metabolism	Increased
L10338	SCN1B	Molecule transport	Reduced
L11244	C4BPB	Immune system	Increased
[Europe, US]			
D28915	IFI44	Immune system	Reduced
K03430	C1QB	Immune system	Reduced
L20971	PDE4B	Signal transduction	Reduced
M37033	CD53	Immune system	Reduced
M37766	CD48	Immune system	Reduced
X04011	CYBB	Immune system	Reduced
X95404	CFL1	Cell motility and extracellular matrix	Reduced

Table VI. *continued*

Table VI. *continued*

Gene accession number	Symbol (or description)	Function	Level in recurrence group *vs.* non-recurrence group
X96586	NSMAF	Signal transduction	Reduced
L20688	ARHGDIB	Signal transduction	Reduced
M31724	PTPN1	Signal transduction	Reduced
L06797	CXCR4	Immune system	Reduced
J03077	PSAP	Metabolism	Reduced
S54005	TMSB10	Cell motility and extracellular matrix	Reduced
HG2917-HT3061	**HLA-E**	Immune system	Reduced

Bold, HLA family genes.

Table VII. *Characteristics of genes listed.*

Number of virtual cohorts	6	5	4	3	2	total
Number of genes down-regulated in recurrence group *vs*. non-recurrence group (%)	7/7 (100%)	17/17 (100%)	26/26 (100%)	46/50 (92.0%)	52/54 (96.2%)	148/154 (96.1%)
Number of HLA (MHC) family genes included (%)	3/7 (42.9%)	3/17 (17.6%)	3/26 (11.5%)	5/50 (10.0%)	2/54 (3.7%)	$p=0.02$ (Fisher's exact test).

In the present study, expression levels of 148 out of a total of 154 genes (96.1%) were lower in HCC with early IHR than in HCC without recurrence. HCC with early IHR is more advanced than HCC without. Reduced expression of these genes is likely to play a role in the progression of HCC. Down-regulation of genes may account more for early IHR of HCC than up-regulation. These concepts are supported by recent DNA microarray studies in which levels of many genes were reduced in HCC tissues compared to those in non-tumorous liver tissues (28, 29).

Investigators have used single-pass gene selection methods to identify molecular signatures related to HCC metastatic potential (4, 6, 7). If single-pass analysis is performed, the genes identified may depend largely on sample variation. Repeating the sampling procedure and analyzing the appearance rates of genes may minimize the variation in genes identified. Thus, we can successfully identify a common pathway (*i.e.* reduced expression of HLA genes) related to early IHR of HCC regardless of sample variation. We believe that the gene selection based on our resampling method has an advantage over single-pass gene selection. Gene selection in virtual cohorts created considering HBV and HCV infection patterns will serve as an attractive tool to identify common pathways responsible for several aspects of HCC with complex etiology and also to overcome issues of DNA microarray analysis (*i.e.* small sample size of high-density data) (30).

Acknowledgements

This work was supported in part by Grants-in-Aid from the Ministry of Education, Culture, Sports, Science and Technology of Japan (No.16500113) and New Energy and Industrial Technology Development Organization of Japan (Project ID 03A02018a).

References

1 Bruix J, Sherman M, Llovet JM, Beaugrand M, Lencioni R, Burroughs AK, Christensen E, Pagliaro L, Colombo M and Rodes J: Clinical management of hepatocellular carcinoma. Conclusions of the Barcelona-2000 EASL conference. European Association for the Study of the Liver. J Hepatol *35*: 421-430, 2001.

2 Poon RT, Fan ST, Ng IO, Lo CM, Liu CL and Wong J: Different risk factors and prognosis for early and late intrahepatic recurrence after resection of hepatocellular carcinoma. Cancer *89*: 500-507, 2000.

3 Blum HE: Treatment of hepatocellular carcinoma. Best Pract Res Clin Gastroenterol *19*: 129-145, 2005.

4 Cheung ST, Chen X, Guan XY, Wong SY, Tai LS, Ng IO, So S and Fan ST: Identify metastasis-associated genes in hepatocellular carcinoma through clonality delineation for multinodular tumor. Cancer Res *62*: 4711-4721, 2002.

5 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: Oligonucleotide microarray for prediction of early intrahepatic recurrence of hepatocellular carcinoma after curative resection. Lancet *361*: 923-929, 2003.

6 Lee JS, Chu IS, Heo J, Calvisi DF, Sun Z, Roskams T, Durnez A, Demetris AJ and Thorgeirsson SS: Classification and prediction of survival in hepatocellular carcinoma by gene expression profiling. Hepatology *40*: 667-676, 2004.

7 Kurokawa Y, Matoba R, Takemasa I, Nagano H, Dono K, Nakamori S, Umeshita K, Sakon M, Ueno N, Oba S, Ishii S, Kato K and Monden M: Molecular-based prediction of early recurrence in hepatocellular carcinoma. J Hepatol *41*: 284-291, 2004.

8 Brechot C: Pathogenesis of hepatitis B virus-related hepatocellular carcinoma: Old and new paradigms. Gastroenterology *127*: S56-S61, 2004.

9 Lai CL, Ratziu V, Yuen MF and Poynard T: Viral hepatitis B. Lancet *362*: 2089-2094, 2003.

10 Poynard T, Yuen MF, Ratziu V and Lai CL: Viral hepatitis C. Lancet *362*: 2095-2100, 2003.

11 Bosch FX, Ribes J, Diaz M and Cleries R: Primary liver cancer: worldwide incidence and trends. Gastroenterology *127*: S5-S16, 2004.

12 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: 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, 2002.

13 Iizuka N, Hamamoto Y and Oka M: Predicting individual outcomes in hepatocellular carcinoma. Lancet *364*: 1837-1839, 2004.

14 Michiels S, Koscielny S and Hill C: Prediction of cancer outcome with microarrays: a multiple random validation strategy. Lancet 365: 488-492, 2005.

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: Differential gene expression in distinct virologic types of hepatocellular carcinoma: Association with liver cirrhosis. Oncogene *22*: 3007-3014, 2003.

16 Matoba K, Iizuka N, Gondo T, Ishihara T, Yamada-Okabe H, Tamesa T, Takemoto N, Hashimoto K, Sakamoto K, Miyamoto T, Uchimura S, Hamamoto Y and Oka M: Tumor HLA-DR expression linked to early intrahepatic recurrence of hepatocellular carcinoma. Int J Cancer *115*: 231-240, 2005.

17 Takemoto N, Iizuka N, Yamada-Okabe H, Hamada K, Tamesa T, Okada T, Hashimoto K, Sakamoto K, Takashima M, Miyamoto T, Uchimura S, Hamamoto Y and Oka M: Sex-based molecular profiling of hepatitis C virus-related hepatocellular carcinoma. Int J Oncol *26*: 673-678, 2005.

18 Brazma A, Hingamp P, Quackenbush J, Sherlock G, Spellman P, Stoeckert C, Aach J, Ansorge W, Ball CA, Causton HC, Gaasterland T, Glenisson P, Holstege FC, Kim IF, Markowitz V, Matese JC, Parkinson H, Robinson A, Sarkans U, Schulze-Kremer S, Stewart J, Taylor R, Vilo J and Vingron M: Minimum information about a microarray experiment (MIAME)-toward standards for microarray data. Nat Genet *29*: 365-371, 2001.

19 Iizuka N, Oka M, Yamada-Okabe H, Mori N, Tamesa T, Okada T, Takemoto N, Sakamoto K, Hamada K, Ishitsuka H, Miyamoto T, Uchimura S and Hamamoto Y: Self-organizing-map-based molecular signature representing the development of hepatocellular carcinoma. FEBS Lett *579*: 1089-1100, 2005.

20 Iizuka N, Hamamoto Y and Oka M: Prediction of cancer outcome with microarrays. Lancet *365*: 1683-1684, 2005.

21 Schena M, Shalon D, Davis RW and Brown PO: Quantitative monitoring of gene expression patterns with a complementary DNA microarray. Science *270*: 467-470, 1995.

22 van 't Veer LJ, Dai H, van de Vijver MJ, He YD, Hart AA, Mao M, Peterse HL, van der Kooy K, Marton MJ, Witteveen AT, Schreiber GJ, Kerkhoven RM, Roberts C, Linsley PS, Bernards R and Friend SH: Gene expression profiling predicts clinical outcome of breast cancer. Nature *415*: 530-536, 2002.

23 Rosenwald A, Wright G, Chan WC, Connors JM, Campo E, Fisher RI, Gascoyne RD, Muller-Hermelink HK, Smeland EB, Giltnane JM, Hurt EM, Zhao H, Averett L, Yang L, Wilson WH, Jaffe ES, Simon R, Klausner RD, Powell J, Duffey PL, Longo DL, Greiner TC, Weisenburger DD, Sanger WG, Dave BJ, Lynch JC, Vose J, Armitage JO, Montserrat E, López-Guillermo A, Grogan TM, Miller TP, LeBlanc M, Ott G, Kvaloy S, Delabie J, Holte H, Krajci P, Stokke T and Staudt LM: Lymphoma/Leukemia Molecular Profiling Project: The use of molecular profiling to predict survival after chemotherapy for diffuse large B cell lymphoma. N Engl J Med *346*: 1937-1947, 2002.

24 Ntzani EE and Ioannidis JP: Predictive ability of DNA microarrays for cancer outcomes and correlates: an empirical assessment. Lancet *362*: 1439-1444, 2003.

25 Rimsza LM, Roberts RA, Miller TP, Unger JM, LeBlanc M, Braziel RM, Weisenberger DD, Chan WC, Muller-Hermelink HK, Jaffe ES, Gascoyne RD, Campo E, Fuchs DA, Spier CM, Fisher RI, Delabie J, Rosenwald A, Staudt LM and Grogan TM: Loss of MHC class II gene and protein expression in diffuse large B-cell lymphoma is related to decreased tumor immunosurveillance and poor patient survival regardless of other prognostic factors: a follow-up study from the Leukemia and Lymphoma Molecular Profiling Project. Blood *103*: 4251-4258, 2004.

26 Ramaswamy S, Ross KN, Lander ES and Golub TR: A molecular signature of metastasis in primary solid tumors. Nat Genet *33*: 49-54, 2003.

27 Pellegris G, Ravagnani F, Notti P, Fissi S and Lombardo C: B and C hepatitis viruses, HLA-DQ1 and -DR3 alleles and autoimmunity in patients with hepatocellular carcinoma. J Hepatol *36*: 521-526, 2002.

28 Coulouarn C, Derambure C, Lefebvre G, Daveau R, Hiron M, Scotte M, Francois A, Daveau M and Salier JP: Global gene repression in hepatocellular carcinoma and fetal liver, and suppression of dudulin-2 mRNA as a possible marker for the cirrhosis-to-tumor transition. J Hepatol *42*: 860-869, 2005.

29 Kim BY, Lee JG., Park S, Ahn JY, Ju YJ, Chung JH, Han CJ, Jeong SH, Yeom YI, Kim S, Lee YS, Kim CM, Eom EM, Lee DH, Choi KY, Cho MH, Suh KS, Choi DW and Lee KH: Feature genes of hepatitis B virus-positive hepatocellular carcinoma, established by its molecular discrimination approach using prediction analysis of microarray. Biochim Biophys Acta *1739*: 50-61, 2004.

30 Liu ET and Karuturi KR: Microarrays and clinical investigations. N Engl J Med *350*: 1595-1597, 2004.

Received March 28, 2007
Revised June 20, 2007
Accepted July 3, 2007