Analysis of loss of heterozygosity on chromosome 4q in hepatocellular carcinoma using high-throughput SNP array

HUAKUN ZHANG[1,3*], HAIQING MA[1*], QIJING WANG[1], MINSHAN CHEN[2], DESHENG WENG[1], HUI WANG[1], JUN ZHOU[1], YONGQIANG LI[1], JIANCONG SUN[1], YIBING CHEN[1], XIAOTING LIANG[1], JINGJING ZHAO[1], KE PAN[1], HUIYUN WANG[1] and JIANCHUAN XIA[1]

[1]State Key Laboratory of Oncology in Southern China and Department of Experimental Research, and
[2]Department of Hepatobilary Oncology, Sun Yat-sen University Cancer Center, Guangzhou, Guangdong 510060;
[3]Central Laboratory of Shenzhen Maternity and Child Healthcare Hospital, Shenzhen, Guangdong 518033, P.R. China

Received September 16, 2009; Accepted October 30, 2009

DOI: 10.3892/or_00000654

Abstract. To identify tumour suppressor genes (TSGs) associated with hepatocellular carcinoma (HCC) on chromosome 4q using a high-throughput single nucleotide polymorphism (SNP) array, we first scanned for loss of heterozygosity (LOH) of 40 SNPs on chromosome 4q and discovered 2 hot regions: 4q24-26 and 4q34.3-35. We then further scanned for LOH of 338 SNPs in genes around 4q34.3-35 and discovered 3 genes with the most frequent LOH: nei endonuclease VIII-like 3 (NEIL3), interferon regulatory factor 2 (IRF2) and inhibitor of growth family member 2 (ING2). A review of the literature indicates only ING2 might be a TSG associated with HCC.

Introduction

Primary liver cancer, which consists predominantly of hepatocellular carcinoma (HCC), has risen to become the fifth most common malignancy worldwide and the third leading cause of cancer mortality (1). Previous studies have demonstrated that accumulation of the activation of oncogenes and/or inactivation of tumour suppressor genes (TSGs) is involved in the carcinogenesis of human cancer (2-4). However, the inactivation of TSGs was especially reported to play a more critical role in HCC pathogenesis (5). Recently, loss of heterozygosity (LOH) as an indirect procedure to identify TSGs has been widely used. Frequent allele deletions are observed on chromosomes 1p, 4q, 5q, 8p, 8q, 9p, 10q, 11p, 13q, 14q, 16q, and 17p in HCC (6-13).

Whole-genome scans and chromosome analyses have been performed, and the chromosome 4q carries deletions in HCC more frequently than in other tumours (14-22). In early studies, we investigated LOH on 22 autosomes with 382 sets of microsatellite markers (MS) in 65 cases of HCC. The results showed that the frequency of LOH on 4q was 48.1% and that TSGs associated with HCC might be on 4q.

In the present study, we first scanned chromosome 4q with a high-throughput SNP array to analyse the LOH and found the region with the most frequent LOH, then further scanned the SNPs around the region to search for the TSGs associated with HCC.

Materials and methods

Patients and tissue specimens. All matched primary hepatocellular carcinoma tissue and adjacent cancer-free tissue specimens (n=69) were obtained from those patients who underwent surgical resection of their diseases in the Sun Yat-sen University Cancer Center between 2005 and 2007. The patients who received any preoperative treatment such as chemotherapy and radiotherapy were excluded. The 69 patients included 60 males and 9 females with a median age of 52 years (range, 21-75 years). The fresh tissues were immediately immersed in RNAlater (Ambion, Inc., USA) after surgical resection, stored at 4˚C overnight to allow thorough penetration of the tissues, then frozen at -80˚C until RNA and DNA extraction. Both cancer and corresponding adjacent cancer-free tissues not less than 2 cm away from the liver were sampled from cancer patients and verified by pathological examination.

DNA extraction. Both cancer and corresponding cancer-free liver tissues were digested with proteinase K (1 mg/ml). After the tissues were homogenised, genomic DNA was extracted via the phenol-chloroform method (23).

SNP array analysis. In the system adopted in this study, SNP-containing sequences are first amplified in a single tube

Correspondence to: Dr Jianchuan Xia and Dr Huiyun Wang, State Key Laboratory of Oncology in Southern China and Department of Experimental Research, 651 Dongfeng Road East, Yuexiu District, Guangzhou, Guangdong 510060, P.R. China
E-mail: xiajch@mail.sysu.edu.cn

*Contributed equally

Key words: hepatocellular carcinoma, loss of heterozygosity, single nucleotide polymorphism, tumour suppressor gene

to a detectable amount by multiplex PCR and then used as templates to generate single-stranded DNA (ssDNA), which is then hybridised to the probes on a microarray. The probes are designed in such a way that their 3' ends are next to the polymorphic sites in the hybridising ssDNA. In this way, the probes can be labelled with the commonly used single-base extension method (24-26) during which single dideoxyribonucleotides (ddNTPs) are added to the probe in an allele-specific way dependent on the hybridising allelic sequence(s). When the corresponding ddNTPs are labelled with different fluorescent chromophores (cyanine dyes, either Cy3 or Cy5, in our system), the allelic state of the SNPs can be determined by analysing the amount of incorporated fluorescence. In addition, SNP genotypes may be determined independently by using the two DNA strands as separate templates so that results from such dual-probe analysis can be compared to ensure a high degree of accuracy.

SNP selection. A computer program written for SNP selection was used as previously described (27,28). SNPs in human chromosome 4q were selected from the dbSNP database (ftp://ftp.ncbi.nih.gov/snp/human/chr_rpts/) maintained by National Center for Biotechnology Information (NCBI). To ensure that the selected SNPs were real and suitable for the multiplex system, a series of criteria was used for selection: i) only transition SNPs (A/T to G/C changes or vice versa) could be selected in order to use a two-colour fluorescent labelling system (using the cyanine dyes Cy3 and Cy5) for genotype determination; ii) SNPs selected on chromosome 4q should be evenly distributed as far as possible, separated probably by 350 kb; iii) uppercase sequences should be better in the 150-bp region around SNPs; iv) SNPs could not be flanked by a significant number of short tandem repeats, palindromes or other SNPs (i.e., SNPs should be separated by >150 bp), which might significantly affect the specificity of amplification; and v) heterozygosity should be >0.4.

In addition, to select suitable SNPs in the interior of genes around the region with the most frequent LOH on chromosome 4q, the corresponding criteria were as follows: i) SNPs should be located inside or within 3000 bp away from the gene; ii) SNPs selected within one gene should be evenly distributed as far as possible; iii) SNPs with frequency of heterozygosity >0.1 must be chosen; the higher the heterozygosity, the better. Furthermore, SNPs could not be flanked by a significant number of short tandem repeats, palindromes or other SNPs (i.e., SNPs should be separated by >150 bp), which might significantly affect the specificity of amplification; iv) only transition SNPs (A/T to G/C changes or vice versa) could be selected; and v) 5-10 SNPs must be selected inside one gene, however, only 5 SNPs would be enough for genes with frequency of heterozygosity >0.2.

Candidate SNP sequences were then submitted to the NCBI and UCSC websites for BLAST (http://www.ncbi.nlm.nih.gov/BLAST) and BLAT (http://www.genome.ucsc.edu/cgi-bin/hgBlat?db=hg8) searches to eliminate possible false SNPs caused by repetitive sequences.

Primer design. To select sequence frames for primers, a computer program designed by Wang *et al* (27,28) was used. The candidate sequence frames were first selected based on a user-defined melting temperature range (55-75˚C in this application) within a user-defined sequence range surrounding the polymorphic sites (150 bp in the present study). All primers were designed at the same time to avoid interaction during amplification in the same tube. The typical length of the primers was 20-25 bp. To minimise and avoid primer-primer interaction, further selection was performed on qualified frames based on the following criteria: i) fewer than 4 consecutively complementary bases between the 3'-ends of any frames; ii) fewer than 8 but 1 consecutively complementary bases between the 3'-ends of any frames; iii) fewer than 10 consecutively complementary bases between the 3'-ends of any frame and anywhere in all the others; iv) fewer than 12 but 1 consecutively complementary base between the 3'-ends of any frame and anywhere in all others; v) less than 75% complementary bases anywhere between any 2 frames; and vi) fewer than 13 complementary bases between the 3'-end of any frame and any amplicon sequence.

In addition, every pair of primers was tested individually by PCR in order to guarantee the right length and specificity of the product. For primers generating non-specific products during detection, repeated design and detection was needed until the product met the requirements. Otherwise, primers were excluded.

Furthermore, these primers were designed in such a way that their 3'-ends would anneal next to the polymorphic sites, and they could therefore also be used as probes (primer-probes) on the microarray for genotyping. For each SNP, two such primer-probes were designed in opposite directions so that they could be used to generate ssDNA in different directions.

Multiplex PCR and ssDNA preparation. The procedures for multiplex amplification were performed following the method of Wang *et al* (27,28) with minor modifications. In brief, first multiplex PCR was performed in 25 μl of PCR mix containing 2.5 μl of 10X PCR buffer (50 mM KCl, 100 mM Tris-HCl at pH 8.3, 1.5 mM $MgCl_2$, and 100 μg/ml gelatin), 0.5 μl of 10 nmol/l dNTPs, primer mix (200 μM each) for all SNPs in the multiplex group, 6 units of HotStar *Taq* DNA polymerase (Qiagen), and 200 ng of DNA template. The samples were first heated to 94˚C for 15 min to activate the *Taq* DNA polymerase, followed by 40 PCR cycles. Each PCR cycle consisted of 40 sec at 94˚C for denaturation and 2 min at 54˚C followed by 5 min of ramping from 54 to 70˚C for annealing and extension. A final extension step was carried out at 72˚C for 3 min at the end of the 40th cycle. ssDNA was generated in both directions using the same conditions for multiplex PCR with the following modifications: i) product from the multiplex PCR (1.0 μl) was used as template; ii) only one primer (one of the primer-probes) for each SNP was used; and iii) 45 PCR cycles were used.

Genotype determination by microarray. Preparation of microarray slides. Gold Seal Micro slides (Becton-Dickinson) were soaked in 30% bleach with shaking for 6 h followed by rinsing 5 times with deionised H_2O and 3 times with MilliQ H_2O. The slides were then sonicated in 15% Fisher brand Versa-Clean Liquid Concentrate with heat on for 2 h and

Table I. Frequency of LOH in chromosome 4q of 69HCC patients.

SNP locus	Informative	No. of LOH	Frequency of LOH (%)	Location	Genotype
rs7682030	20	1	5	4q12	C/T→C
rs722709	5	0	0	4q12	-
rs184732	8	1	12.5	4q13.1	C/T→C
rs1384725	11	1	9.1	4q13.1	C/T→T
rs12649371	15	2	13.3	4q13.2	C/T→C
rs7656234	16	4	25	4q13.3	C/T→C
rs2609200	22	2	9.1	4q13.3	C/T→T
rs10028815	17	2	11.8	4q21.21	C/T→C
rs755527	7	0	0	4q21.21	-
rs1867544	16	1	6.2	4q21.23	C/T→T
rs6531983	12	2	16.7	4q22.1	C/T→C
rs12642565	15	2	13.3	4q22.1	C/T→T
rs1368726	9	1	11.1	4q22.2	C/T→C
rs1568102	17	2	11.8	4q22.3	C/T→C
rs867381	5	0	0	4q23	-
rs3774933	11	1	9.1	4q24	C/T→C
rs2298732	19	7	36.8	4q24	C/T→C
rs3181187	6	1	16.7	4q25	C/T→C
rs965012	25	8	32.0	4q25	C/T→C
rs10013652	11	5	45.4	4q26	C/T→T
rs2292493	13	2	15.4	4q27	C/T→C
rs7677254	13	5	38.4	4q28.1	C/T→T
rs12645636	18	1	5.6	4q28.2	C/T→C
rs10049646	14	2	14.3	4q28.3	C/T→C
rs1386363	3	0	0	4q28.3	-
rs11099441	12	1	8.3	4q28.3	C/T→T
rs3805315	15	1	6.7	4q31.1	C/T→T
rs10033674	8	3	37.5	4q31.21	C/T→T
rs7435720	28	3	10.7	4q31.22	C/T→C
rs4574387	6	0	0	4q31.23	-
rs3733390	9	1	11.1	4q32.1	C/T→T
rs10000610	8	1	12.5	4q32.1	C/T→T
rs359512	13	2	15.4	4q32.2	C/T→C
rs6536943	16	3	18.8	4q32.3	C/T→C
rs1876442	23	8	34.8	4q33	C/T→T
rs4695942	15	4	26.6	4q34.1	C/T→C
rs1567475	17	6	35.3	4q34.3	C/T→T
rs1514479	13	5	38.5	4q35.1	C/T→C
rs2249916	17	8	47.0	4q35.2	C/T→T
rs10027026	18	2	11.1	4q35.2	C/T→C

then rinsed with shaking in deionised H_2O 10 times and in MilliQ H_2O 5 times. Slides were dried by centrifugation at 1000 rpm for 5 min. The slides were then baked at 160˚C in a vacuum oven for 4-6 h.

Microarray preparation. One volume of probe was mixed with four volumes of microarray printing solution EZ'nBrite (distributed by GenBase Biosciences Corp.), for a final concentration of 40 μM for each probe in each well of 384-well plates. Probes were then spotted onto the washed glass slides using a microarray spotter, OmniGrid Accent (GeneMachines), under 50-55% humidity at 22-25˚C. Two matrixes were made in one slide.

Hybridisation. Hybridisation was done with 2X hybridisation solution (5X Denhart's solution, 0.5% SDS, 5X SSC, 20 μl of ssDNA/1000 microarray spots) in a Hybridisation Chamber (Corning) at 56˚C for 2.5-4 h. Chambers were immersed

Table II. Distribution of selected SNPs in genes around 4q34.3-4q35.2.

	Gene	Region (bp)	Cyto	SNP
1	NEIL3	178,465K-178,524K	4q34.3	8
2	AGA	178,586K-178,603,600	4q32-q33	4
3	LOC285500	178,600K-178,838K	4q34.3	12
4	LOC285501	178,884K-179,153K	4q34.3	12
5	LOC643040	179,292,080-179,296,060	4q34.3	0
6	LOC391719	179,441,840-179,448,430	4q34.3	11
7	LOC643179	181,728,710-181,734,900	4q34.3	5
8	LOC643195	182,118,190-182,125,290	4q34.3	4
9	LOC132386	182,677,810-182,683,810	4q35.1	5
10	LOC643296	183,130K-183,246K	4q35.1	7
11	MGC45800	183,296,800-183,305,610	4q35.1	4
12	ODZ3	183,399K-183,962K	4q35.1	19
13	DCTD	184,045,200-184,078,600	4q35.1	9
14	LOC643437	184,192,900-184,240K	4q35.1	6
15	FLJ30277	184,252,170-184,260,350	4q35.1	2
16	BOMB	184,349K-184,477K	4q35.1	10
17	CLDN22	184,474,370-184,480,370	4q35.1	0
18	LOC643482	184,477,900-184,495,300	4q35.1	4
19	LOC643492	184,493,860-184,498,770	4q35.1	1
20	CARF	184,599,780-184,609,040	4q35.1	4
21	LOC389246	184,638,500-184,645,580	4q35.1	0
22	ING2	184,660,200-184,672,200	4q35.1	2
23	RWDD4A	184,793,800-184,820,300	4q35.1	8
24	FLJ12716	184,817,400-184,871,700	4q35.1	11
25	STOX2	185063503-185175869	4q35.1	10
26	ENPP6	185,244K-185,379K	4q35.1	11
27	LOC391722	185,454,300-185,478,500	4q35.1	3
28	MGC24125	185,495,900-185,515,100	4q35.1	6
29	IRF2	185,542,900-185,635,700	4q34.1-q35.1	11
30	CASP3	185,782,800-185,810,600	4q34	6
31	FLJ33167	185,807,900-185,856,100	4q35.1	4
32	MLF1IP	185,852,200-185,892,300	4q35.1	7
33	ACSL1	185,913,700-185,984,200	4q34-q35	9
34	LOC643036	185,956,450-185,957,188	4q35.1	0
35	HELT	186,174,080-186,179,220	4q35.1	3
36	SLC25A4	186,298,400-186,308,400	4q35	3
37	KIAA1430	186,309,900-186,367,500	4q35.1	6
38	SNX25	186,368K-186,522K	4q35.1	10
39	LRP2BP	186,522K-186,537,100	4q35.1	6
40	ANKRD37	186,551,830-186,561,380	4q35.1	3
41	FLJ11200	186,551,400-186,587,100	4q35.1	6
42	KM-HN-1	186,600,300-186,632,900	4q35.1	8
43	PDLIM3	186,656,900-186,696,700	4q35	8
44	SORBS2	186,741K-187,118K	4q35.1	9
45	TLR3	187,224,300-187,246,200	4q35	6
46	DKFZP564J102	187,300K-187,333,300	4q35.2	5
47	CYP4V2	187,346,700-187,374,600	4q35.2	8
48	KLKB1	187,382,700-187,419,600	4q34-q35	8
49	F11	187,421,300-187,449,900	4q35	7
50	LOC644042	187,483,800-187,532,500	4q35.2	6
51	MTNR1A	187,688,800-187,716,500	4q35.1	5
52	FAT	187,746K-187,882K	4q35	9

Table II. Continued.

	Gene	Region (bp)	Cyto	SNP
53	MRPS36P2	188,055,890-188,061,990	4q35.1	2
54	LOC644282	188,599,080-188,605,440	4q35.2	2
55	LOC644317	188,885,100-188,893K	4q35.2	1
56	LOC644325	188,901,280-188,907,660	4q35.2	5
57	LOC389249	189,125,350-189,132,430	4q35.2	1
58	ZFP42	189,150,900-189,166,200	4q35.2	2
59	FLJ25801	189,246,400-189,266,400	4q35.2	7
60	FLJ36180	189,297,590-189,305,640	4q35.2	3
61	LOC152663	189,504,550-189,511,400	4q35.2	2
62	LOC644491	189,544,100-189,560K	4q35.2	4
63	LOC401164	189,556K-189,763K	4q35.2	7
64	LOC644514	189,696,500-189,790K	4q35.2	6
65	LOC285442	189,838,100-189,900,200	4q35.2	8
Total				381

in ice water for ~30 sec before opening. Slides were washed at 56˚C with 1X SSC and 0.1% SDS for 10 min, twice with 0.2X SSC for 30 sec, and twice with 0.2X SSC for 30 sec.

Labelling probes by single-base extension. Probes were labelled in 25 μl of labelling solution containing 1/7 volume of Sequenase buffer (supplied by the vendor), 0.5 units/μl Sequenase (Amersham Pharmacia Biosciences), Cy3-ddATP and Cy5-ddGTP (PE Biosystems) for AG probes, and Cy3-ddUTP and Cy5-ddCTP for CT probes (750 nM each). The reaction was incubated at 70˚C for 10 min (not allowing it to dry). Chambers were immersed in ice water for ~30 sec before opening. Slides were washed under the conditions specified for the washing after hybridisation, as described above.

Data analysis. Microarrays were scanned with GenePix 4000B (Axon Instruments). The resulting images were analysed with GenePix Pro (Axon Instruments) software. Genotypes were determined using the computer program AccuTyping.

Statistical analysis. The SPSS 15.0 software (SPSS Inc., Chicago, IL, USA) was used for all statistical analyses and a p-value <0.05 was considered significant.

Results

Analysis of LOH on chromosome 4q. In this study, 42 SNPs were first selected on chromosome 4q from the dbSNP database (ftp://ftp.ncbi.nih.gov/snp/human/chr_rpts/) maintained by NCBI. PCR primers and SNP probes of these SNPs were designed by the software (27,28). Every pair of primers was further selected by pretesting amplification in order to guarantee only one specific product. Finally, 40 pairs of primers were successfully determined. Multiplex amplification was performed as described in Materials and methods. Cancer tissue and corresponding cancer-free tissue were hybridised and labelled in two different matrixes of the same chip under the same reaction conditions.

In this part of the results, the LOH frequency of 40 SNPs on chromosome 4q in HCC varied. In all 69 HCC samples, at least 1 patient was found to have one mutation of 35 out of the total 40 SNPs, with the exceptions of rs722709, rs1386363, rs755527, rs4574387 and rs867381 (Table I, Fig. 2). In addition, 51 of 69 patients were found to have LOH on more than one SNP. The highest frequency of LOH was found on rs2249916 (47.0%), which was located at 4q35.2; the second highest frequency of LOH happened on rs10013652 (45.4%) of 4q26. There were 9 SNPs with frequencies of LOH >30%, which were mainly clustered in two regions. rs2298732 (36.8%), rs965012 (32.0%) and rs10013652 (45.4%) were located at 4q24-26; rs1567475 (35.3%), rs1514479 (38.5%) and rs2249916 (47.0%) were located at 4q34.3-35. In addition, rs7677254 (38.4%), rs10033674 (37.5%) and rs1876442 (34.8%) were located in 4q28.1, 4q31.2 and 4q33, respectively.

The correlation between SNPs with high frequency of LOH and clinical parameters was examined using the χ^2 test. This test found that, among SNPs with frequency of LOH >30%, LOH of rs1567475 (4q34.3) and rs1514479 (4q35.1) was shown to be correlated with the histological grade of HCC ($p<0.05$, i.e., HCC with worse differentiation showed higher frequency of LOH at these sites) but independent from other clinical parameters such as age, gender, HBV infection, serum AFP level, tumour size and hepatic cirrhosis ($p>0.05$). LOH of rs7656234 (4q13.3) was found to be correlated with HBV infection ($p<0.05$), in agreement with previous studies (29) and suggested the existence of some HBV infection-associated genes in this region. However, LOH of rs7656234 was independent from age, gender, serum AFP level, tumour size, hepatic cirrhosis and histological grade ($p>0.05$).

Two regions (4q24-26 and 4q34.3-4q35) with a high frequency of LOH on chromosome 4q were revealed by LOH analysis of 40 SNPs, suggesting some hepatocarcinogenesis-associated tumour suppressor genes in these regions, and

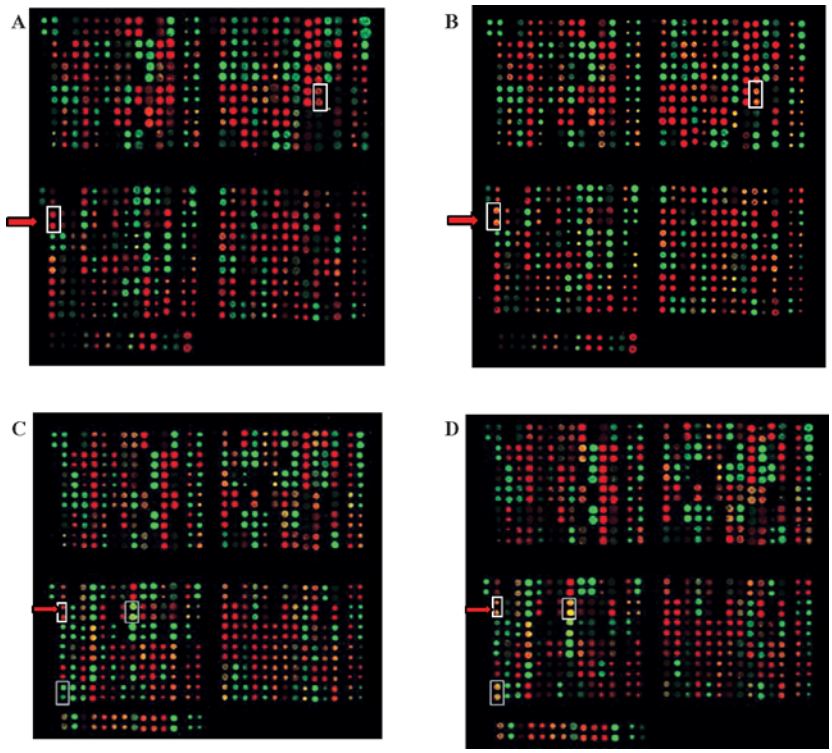

Figure 1. Two representative SNP hybridization results. (A) Cancer tissue of sample 12; (B) corresponding cancer-free tissue of sample 12; (C) cancer tissue of sample 18; (D) corresponding cancer-free tissue of sample 18. Each probe was printed twice and shown as neighbouring spots. Spots in red and green, homozygous; yellow, heterozygous; and dark, low signal, but not necessarily no signal, or too low of a signal for genotype calls. White panel without arrow indicates LOH. White panel with red arrow indicates LOH of rs6830958, which is homozygous in cancer tissue but heterozygous in matched cancer-free tissue.

4q34.3-4q35 showed higher frequency of LOH than 4q24-26. With regard to 4q34.3-4q35, there are many reports on deletions in this region in HCC, but no hepatocarcinogenesis-associated tumour suppressor genes have been found. There have not even been any reports on screening hepatocarcinogenesis-associated tumour suppressor genes on a large scale. Therefore, in this study, we first carried out LOH analysis of genes around 4q34.3-4q35 using SNP arrays, in order to discover HCC-associated TSGs.

Identification of tumour suppressor genes around 4q34.3-4q35 in HCC. To further search for the target TSGs in the identified region, we first focused on the region around 4q34.3-4q35.2. Objective genes were selected in the 12-Mb region. There were 65 genes according to the NCBI genome database in 2005. Finally, 381 SNPs (Table II) were selected inside or near 63 genes. No suitable SNPs were found in the other 2 genes according to the series of criteria given in Materials and methods. In total, 338 suitable pairs of primer-probes were selected after software design and detection.

Similarly, cancer tissue and corresponding cancer-free tissue were hybridised and labelled in two different matrixes of the same chip under the same reaction conditions (Fig. 1).

These results showed that LOH frequency of 338 SNPs in genes of the 4q34-35.2 region in HCC was different. In all 69 HCC samples, there were 65 (94.2%) samples with LOH on more than one SNP. The highest frequency of LOH was found on rs2048076 (46.7%), which was located inside NEIL3; the second highest frequency of LOH was on rs6830958 (41.2%) inside ING2. High frequency of LOH >40% was found on rs10000856, which was located within IRF2. There were 19 SNPs with frequency of LOH >20% (Table III, Fig. 3). However, the rest of the SNPs had a frequency of LOH <20% or no LOH.

Statistical analysis of the relationship between SNPs with high frequency of LOH and clinical parameters suggested that, among SNPs with frequency of LOH >20%, LOH was independent from age, gender, HBV infection, serum AFP level, tumour size, hepatic cirrhosis and histological grade ($p>0.05$).

Discussion

The long arm of chromosome 4 (4q) is a hot region in tumour studies. Recently, LOH of 4q has been detected in oesophageal carcinoma, glioma, cervical carcinoma, carcinoma of gallbladder and squamous cancer in head and neck (14-19). Interestingly, there are many reports on higher frequencies of LOH on 4q in hepatocellular carcinoma (20-22), and 4q has become a hot region for discovery of TSGs associated with HCC. So far, traditional microsatellite detection has still been the primary method for detecting LOH on 4q in HCC. Even when LOH analysis with SNP array has been used, it has been performed as a general scan of the whole genome of HCC (30). There is no report using SNP arrays in LOH analysis on 4q.

In the present study, we first selected 42 SNPs on 4q, which covered 4q and were distributed as evenly as possible

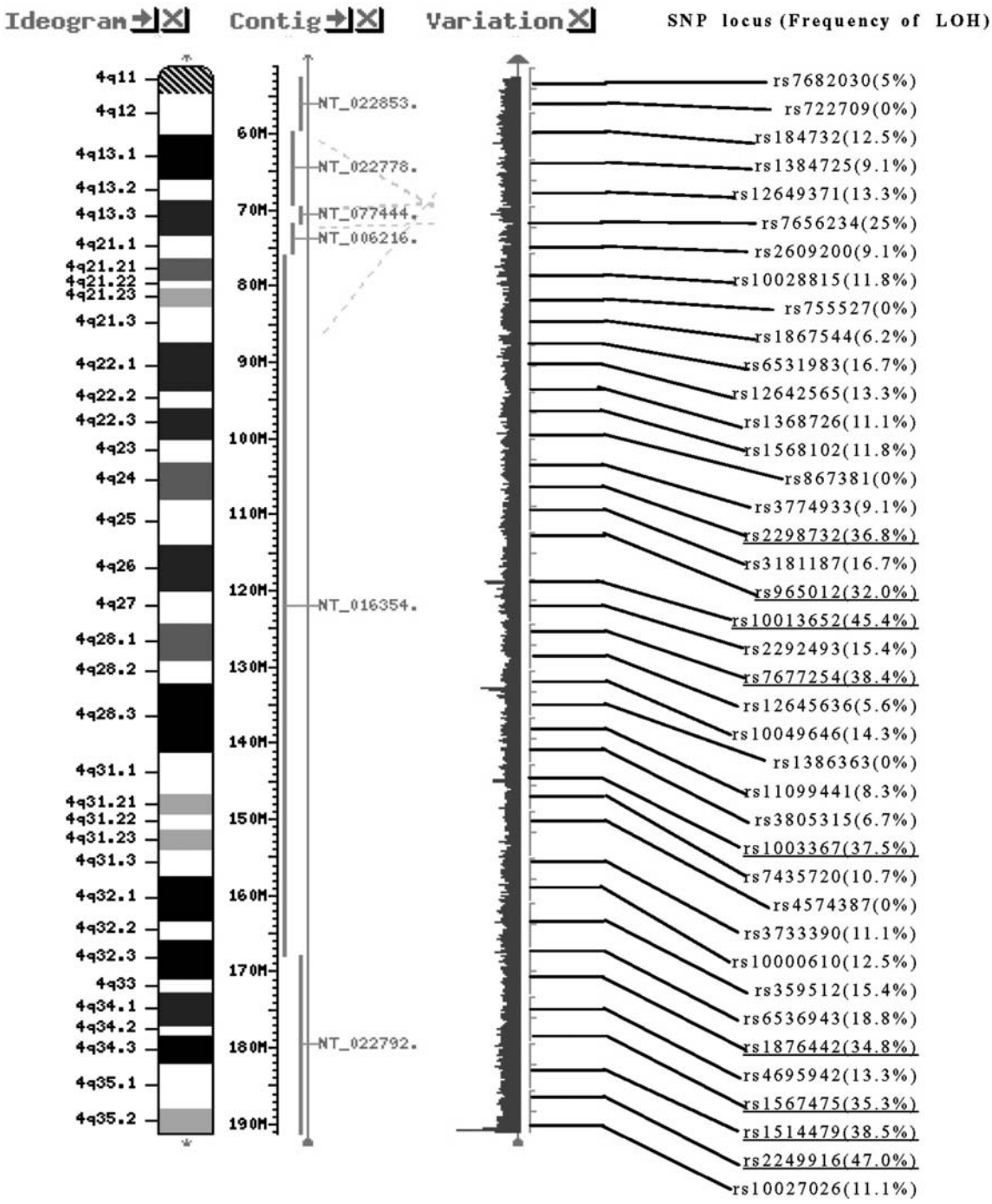

Figure 2. Map of selected 40 SNPs on chromosome 4q and their frequency of LOH. The 9 SNPs with frequency of LOH >30% are marked with a transverse line.

with a separating distance of 350 kb. The fluorescence intensity in cancer tissue and cancer-free tissue in the SNP array was compared to discover SNPs with LOH and to choose SNPs or regions associated with clinical parameters.

In all 69 HCC samples, 51 samples were found to have more than one SNP with LOH. The total frequency of LOH was 73.9%. The highest frequency of LOH was found on rs2249916 (47.0%) located at 4q35.2; the second highest frequency of LOH occurred on rs10013652 (45.4%) of 4q26. There were nine SNPs with frequency of LOH >30%, which were mainly distributed in two regions. rs2298732 (36.8%), rs965012 (32.0%) and rs10013652 (45.4%) were located at 4q24-26; rs1567475 (35.3%), rs1514479 (38.5%) and rs2249916 (47.0%) were located at 4q34.3-35.

In our study, among SNPs with frequency of LOH >30%, rs1567475 (4q34.3) and rs1514479 (4q35.1) were shown to have LOH correlated with histological grade of HCC ($p<0.05$), and poorly differentiated HCCs showed high frequency of LOH at these two sites.

Several similar studies have been reported. Bando *et al* (31) examined 96 primary HCCs for their patterns of allelic loss at 39 microsatellite marker loci distributed along chromosome arm 4q and identified two distinct commonly deleted regions: 4q21-22 and 4q35. Furthermore, allelic loss at 4q35

Table III. Distribution of SNPs with high frequency of LOH in 4q34-35.2.

SNP locus	Informative	No. of LOH	Frequency of LOH (%)	Gene	Genotype
rs2048076	15	7	46.7	NEIL3	C/T→C
rs4690567	16	4	25	LOC285501	C/T→T
rs998326	5	1	20	LOC132386	C/T→T
rs6814485	9	2	22.2	ODZ3	C/T→C
rs6830958	17	7	41.2	ING2	C/T→C
rs4862229	4	1	25	RWDD4A	C/T→C
rs7657090	18	5	27.8	FLJ12716	C/T→C
rs13103966	22	5	22.7	ENPP6	C/T→T
rs1049216	7	2	28.6	CASP3	C/T→C
rs6817412	23	5	21.7	IRF2	C/T→C
rs10000856	10	4	40	IRF2	C/T→T
rs3108230	9	2	22.2	SNX25	C/T→C
rs2120414	7	2	28.6	KM-HN-1	C/T→C
rs9312338	13	3	23.1	SORBS2	C/T→T
rs4862632	11	3	27.3	TLR3	C/T→C
rs1877320	5	1	20	CYP4V2	C/T→T
rs11132390	16	4	25	LOC644042	C/T→T
rs4446379	7	2	28.6	LOC644325	C/T→C
rs12502770	12	3	25	FLJ25801	C/T→T

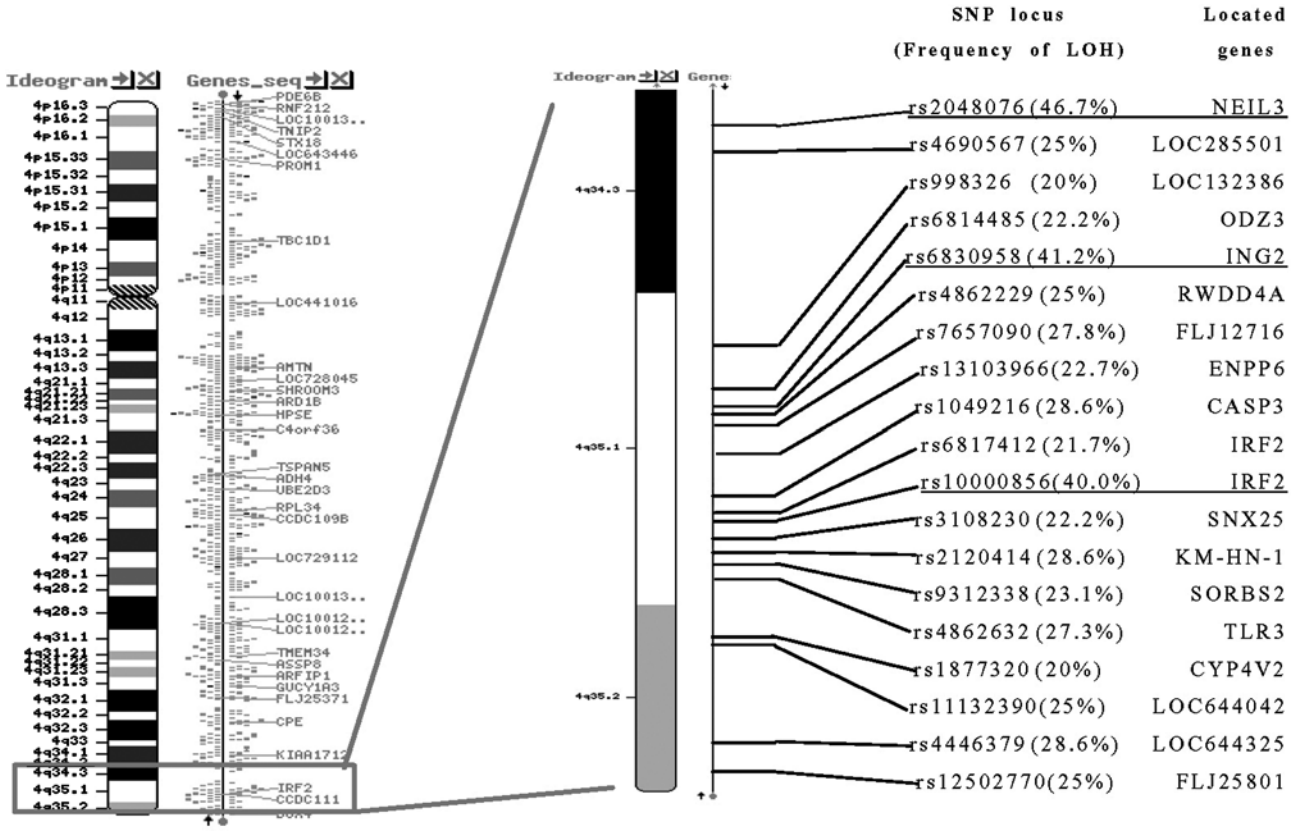

Figure 3. Map of SNPs with high frequency of LOH and their located genes in 4q34-35.2. The 3 SNPs with frequency of LOH >30% are marked with a transverse line.

was more frequent in poorly or moderately differentiated tumours than in well-differentiated tumours. In addition, they also found allelic loss at 4q35 to be more frequent in tumours larger than 2 cm in size, and in tumours that arose from liver

cirrhosis as opposed to HCCs arising from chronic hepatitis. Zhang *et al* (32) detected LOH on chromosome 4q using PCR-simple sequence length polymorphism (PCR-SSLP) on 10 high-polymorphic microsatellite markers in 56 HCC and found that LOH on D4S426 (4q35), D4S1615 (4q31.1) and D4S1652 (4q34-qter) was more frequent in poorly or moderately differentiated HCC than in well-differentiated HCC.

However, there were also some opposing results. For example, Bluteau *et al* (33) systematically analysed 85 micro-satellite markers spanning chromosome 4q in a series of 154 well-characterised primary liver tumours and defined three minimal common regions of deletion (MCRDs) of 15, 9 and 8 Mb in the 4q22, 4q34 and 4q35 regions, respectively. However, a search for relationships between the specific regions of deletion and clinical parameters showed a significant association of loss of the 4q34-35 region with alcohol intake (p=0.005) and with high grade of differentiation (p=0.02). The study results were inconsistent with those of Bando *et al* (31) and Zhang *et al* (32).

Furthermore, in our study, LOH of rs7656234 (4q13.3) was correlated with HBV infection (p<0.05), which was in agreement with previous studies (29), and suggests the possiblity of some HBV infection-associated genes in this region.

Similar results have been reported in many studies. Yeh *et al* (34) investigated allelic loss on chromosome 4q using 13 sets of microsatellite polymorphic markers in 42 HCC samples and showed 77% allelic loss on chromosome 4q with the common region mapped to 4q12-23. In addition, the allelic loss of chromosome 4q was significantly associated with hepatocellular carcinoma having elevated serum AFP. They then detected 149 HCC samples using 49 microsatellite markers and identified a common region with allelic loss between D4S1534 and D4S1572. In these HCCs, 4q allelic loss was associated with hepatitis B virus infection status and the elevation of serum AFP. Marchio *et al* (35) compared the comparative genomic hybridisation (CGH) analysis of 34 HCCs resected on non-cirrhotic livers from patients serologically negative for both hepatitis B (HBV) and C (HCV) viruses with the CGH analysis of 50 HCCs selected on the basis of their positivity for HBV infection. They found a significant decrease (40% on average) of losses on 4q, 16q and 17p in non-viral HCC samples, suggesting that these abnormalities are tightly associated with HBV infection.

On the contrary, Zhang *et al* (36) examined loss of heterozygosity at 34 loci on 23 chromosomes in 35 surgically resected human hepatocellular carcinomas and found that LOH on chromosomes 4 seemed to be important in the development of human hepatocellular carcinoma irrespective of the presence of hepatitis B virus infection.

Furthermore, there have been studies on other deletion regions of chromosome 4q. Hammond *et al* (37) examined the frequency of genetic alterations in 28 HCC cases using short tandem repeat (STR)-microsatellites and restriction fragment length polymorphism (RFLP) analyses, and found the 4q28 region to have one of the highest frequency of alterations. Moreover, the deletion frequency of 4q28 was higher in males than in females, indicating important TSGs associated with gender within this region. In the present study, we also identified a high frequency of LOH of rs7677254 at 4q28.1, suggesting the region 4q28 has a common allelic loss. However, we did not find any correlation between the LOH of rs7677254 and clinical parameters such as age, gender, HBV infection, serum AFP level, tumour size, hepatic cirrhosis and pathological grade (p>0.05). Further research is still needed to determine the relationship between LOH of 4q28 and gender.

Based on our study and many other previous reports, it seems clear that high frequency of LOH is common at 4q24-26 and 4q34.3-4q35. A search of 4q24-26 in the NIH genome database found 138 genes in this region. However, very few of those genes are related to cell growth. Only caspase-6 and RAC1P5 seem to be more important for tumour growth. In contrast, 4q34.3-4q35, especially 4q35.2 with the highest frequency of LOH, seems to have important genes involved in cell growth, and cell cycle regulation.

So far, hepatocarcinogenesis-associated TSGs have not been reported at 4q34.3-4q35, and there have not even been any studies on screening HCC-associated TSGs on a large scale within this region. In the present study, to further identify any HCC-associated TSGs, we first scanned all 65 genes within 4q34.3-4q35 using SNP arrays developed by us.

Out of the first selected 381 SNPs inside or near these 65 genes, only 338 pairs of primer-probes were suitable following software design and detection. The hybridisation results and analysis showed that the frequency of LOH in 338 SNPs was quite different. In all 69 HCC samples, there were 65 (94.2%) samples with LOH in more than one SNP, and 19 SNPs showed frequency of LOH >20%. The first three SNPs with the highest frequency of LOH were rs2048076 (46.7%), rs6830958 (41.2%) and rs10000856 (40%), which were located inside NEIL3, ING2 and IRF2, respectively.

NEIL3 (nei endonuclease VIII-like 3), first named as FPG2, was discovered by Bandaru *et al* (38) in 2002. It is one of the members of the Fpg/Nei family primarily found in the bacterial kingdom, showing homology with DNA glycosylase. Oxidative base damage leads to alteration of genomic information and is implicated as a cause of aging and carcinogenesis. DNA glycosylases initiates base excision repair (BER) by hydrolysing the N-glycosidic bond and releasing the damaged base (39). NEIL3, as a kind of DNA glycosylase, may be involved in carcinogenesis. In this study, NEIL3 was found to have high frequency of LOH, suggesting a relationship between NEIL3 genetic abnormalities and hepatocarcinogenesis. However, existing studies could not answer whether NEIL3 is an HCC-associated TSG. Further studies are still needed.

Interferon regulatory factor 2 (IRF2) is another gene with higher frequency of LOH in our study. The protein encoded by IRF2 is among the members of the interferon regulatory factor family, which have thus far been shown to be transcriptional mediators in many biological processes, including innate and adaptive immune responses, cell growth regulation and apoptosis, and hematopoietic development (40-43). IRF2 was initially found to antagonise IRF1 in terms of transcriptional activity (44). In contrast to the anti-oncogenic activity of IRF1, IRF2 shows oncogenic potential: NIH3T3 cells with overexpressed IRF2 became transformed and were more

tumourigenic in nude mice, indicating IRF2 as a potential oncoprotein (45). One putative mechanism of the oncogenic activity of IRF2 is that IRF2 antagonises the antiproliferative action of IRF1 by competing for binding sites in the promoters of several growth-suppressing genes (46-48). Consistently, several studies showed that the IRF2 expression level increased in clinical samples from oesophageal squamous cell cancer (49) and breast cancer (50), whereas the IRF1 expression level decreased in these cancers. In this study, IRF2 showed higher frequency of LOH. However, there has not been any report on the function of IRF2 as a TSG in HCC.

The inhibitor of growth family member 2 (ING2), cloned and mapped to human chromosome 4q35 (51,52), is an important member of the ING family, which encodes a series of proteins that are important cofactors of p53 associated with cell cycle progression, apoptosis, and DNA repair (51,53). A number of studies indicated that ING2 possessed tumour-suppressive functions, such as induction of growth arrest, senescence, apoptosis and enhancement DNA repair (52,54-57). Recently, there have been several reports on the role of ING2 as a candidate TSG in human cutaneous melanomas and lung cancer (58,59). In our study, we found it to have a high frequency of LOH in HCC, indicating its possible function as a TSG in HCC. In fact, in a previous study (60), we investigated the expression pattern of ING2 in primary HCC, further elaborated its relationship with clinicopathological features, and evaluated its prognostic value for prediction of post-resectional survival in HCC. The findings suggested that ING2 down-regulation frequently occurred in HCC and that its significantly decreased expression in HCC may lead to an unfavourable prognosis. The study results suggest that ING2 may be a candidate tumour suppressor gene for HCC and that ING2 may have diagnostic and therapeutic potential for patients with HCC. However, elucidation of the molecular mechanism used by the target TSG ING2 to perform its role in liver carcinogenesis and the possibility that AFP is one of the downstream effectors regulated by ING2 is worthy of future study.

In conclusion, the SNP array used in this study was a practicable and reliable technique for scanning SNPs with high frequency of LOH. In this method, primer-probes were designed according to different hot spots following the preparation and analysis of microarrays. Experimental flexibility was thus greatly improved compared with commercial arrays. In addition, as a high-throughput experimental technique, SNP arrays can be used to analyse hundreds of loci, critically boosting experimental efficiency. For large sample size analyses, considerable expenditures are needed to purchase commercial arrays. However, lower expenditures are sufficient for the SNP arrays that we spotted ourselves, and larger sample sizes lead to even lower expenses. The data from this study not only validate the use of SNP arrays for research itself but also emphasises the potential of such high-throughput approaches for use in the clinical setting.

References

1. El-Serag HB and Rudolph KL: Hepatocellular carcinoma: epidemiology and molecular carcinogenesis. Gastroenterology 132: 2557-2576, 2007.
2. Vogelstein B, Fearon ER, Hamilton SR, *et al*: Genetic alterations during colorectal-tumor development. N Engl J Med 319: 525-532, 1988.
3. Sheu JC, Huang GT, Shih LN, *et al*: Hepatitis C and B viruses in hepatitis B surface antigen-negative hepatocellular carcinoma. Gastroenterology 103: 1322-1327, 1992.
4. Weissenbach J, Gyapay G, Dib C, *et al*: A second-generation linkage map of the human genome. Nature 359: 794-801, 1992.
5. Ozturk M: Biology of hepatocellular carcinoma. In: Gastrointestinal Cancer. Biology, Diagnosis, and Therapy. Rustgi AK (ed.) Lippincott-Raven, Philadelphia, pp520-521, 1995.
6. Yeh SH, Chen PJ, Chen HL, Lai MY, Wang CC and Chen DS: Frequent genetic alterations at the distal region of chromosome 1p in human hepatocellular carcinomas. Cancer Res 54: 4188-4192, 1994.
7. Kuroki T, Fujiwara Y, Tsuchiya E, *et al*: Accumulation of genetic changes during development and progression of hepatocellular carcinoma: loss of heterozygosity of chromosome arm 1p occurs at an early stage of hepatocarcinogenesis. Genes Chromosomes Cancer 13: 163-167, 1995.
8. Takahashi K, Kudo J, Ishibashi H, Niho Y, *et al*: Loss of heterozygosity in hepatocellular carcinoma. Nippon Rinsho 51: 370-374, 1993.
9. Fujimoto Y, Hampton LL, Wirth PJ, Wang NJ, Xie JP and Thorgeirsson SS: Alterations of tumor suppressor genes and allelic losses in human hepatocellular carcinomas in China. Cancer Res 54: 281-285, 1994.
10. Yumoto Y, Hanafusa T, Hada H, *et al*: Loss of heterozygosity and analysis of mutation of p53 in hepatocellular carcinoma. J Gastroenterol Hepatol 10: 179-185, 1995.
11. Nagai H, Pineau P, Tiollais P, Buendia MA and Dejean A: Comprehensive allelotyping of human hepatocellular carcinoma. Oncogene 14: 2927-2933, 1997.
12. Piao Z, Park C, Park JH, Kim H, *et al*: Allelotype analysis of hepatocellular carcinoma. Int J Cancer 75: 29-33, 1998.
13. Sheu JC, Lin YW, Chou HC, *et al*: Loss of heterozygosity and microsatellite instability in hepatocellular carcinoma in Taiwan. Br J Cancer 80: 468-76, 1999.
14. Sterian A, Kan T, Berki AT, *et al*: Mutational and LOH analyses of the chromosome 4q region in esophageal adenocarcinoma. Oncology 70: 168-172, 2006.
15. Hu N, Wang C, Hu Y, *et al*: Genome-wide loss of heterozygosity and copy number alteration in esophageal squamous cell carcinoma using the Affymetrix GeneChip Mapping 10 K array. BMC Genomics 7: 299, 2006.
16. Wong KK, Tsang YT, Chang YM, *et al*: Genome-wide allelic imbalance analysis of pediatric gliomas by single nucleotide polymorphic allele array. Cancer Res 66: 11172-11178, 2006.
17. Backsch C, Rudolph B, Kuhne-Heid R, *et al*: A region on human chromosome 4 (q35.1→qter) induces senescence in cell hybrids and is involved in cervical carcinogenesis. Genes Chromosomes Cancer 43: 260-272, 2005.
18. Arakawa A, Fujii H, Matsumoto T, Hirai S and Suda K: Loss of heterozygosity in clonal evolution with genetic progression and divergence in spindle cell carcinoma of the gallbladder. Hum Pathol 35: 418-423, 2004.
19. Ye H, Pungpravat N, Huang BL, *et al*: Genomic assessments of the frequent loss of heterozygosity region on 8p21.3-p22 in head and neck squamous cell carcinoma. Cancer Genet Cytogenet 176: 100-106, 2007.
20. Zhang SH, Cong WM, Xian ZH and Wu MC: Clinicopathological significance of loss of heterozygosity and microsatellite instability in hepatocellular carcinoma in China. World J Gastroenterol 11: 3034-3039, 2005.
21. Li X, Ding M and Lai B: Frequent loss of heterozygosity on chromosome 4 in human hepatocellular carcinoma. Zhonghua Yi Xue Za Zhi 81: 37-40, 2001.
22. Parkin DM, Pisani P and Ferlay J: Estimates of the worldwide incidence of eighteen major cancers in 1985. Int J Cancer 54: 594-606, 1993.
23. Maniatis TF and Sambrook J: Molecular cloning: a laboratory manual. Cold Spring Harbor Laboratory, Cold Spring Harbor, NY, 1989.
24. Pastinen T, Raitio M, Lindroos K, Tainola P, Peltonen L and Syvanen AC: A system for specific, high-throughput genotyping by allele-specific primer extension on microarrays. Genome Res 10: 1031-1042, 2000.

25. Pastinen T, Kurg A, Metspalu A, Peltonen L and Syvanen AC: Minisequencing: a specific tool for DNA analysis and diagnostics on oligonucleotide arrays. Genome Res 7: 606-614, 1997.
26. Shumaker JM, Metspalu A and Caskey CT: Mutation detection by solid phase primer extension. Hum Mutat 7: 346-354, 1996.
27. Wang HY, Luo M, Tereshchenko IV, *et al*: A genotyping system capable of simultaneously analyzing >1000 single nucleotide polymorphisms in a haploid genome. Genome Res 15: 276-283, 2005.
28. Hu G, Wang HY, Greenawalt DM, *et al*: AccuTyping: new algorithms for automated analysis of data from high-throughput genotyping with oligonucleotide microarrays. Nucleic Acids Res 34: e116, 2006.
29. Li S, Wang H and Zhang C: Genome-wide loss of heterozygosity analyses in primary hepatocellular carcinoma. Zhonghua Yi Xue Za Zhi 80: 577-581, 2000.
30. Midorikawa Y, Yamamoto S, Ishikawa S, *et al*: Molecular karyotyping of human hepatocellular carcinoma using single-nucleotide polymorphism arrays. Oncogene 25: 5581-5590, 2006.
31. Bando K, Nagai H, Matsumoto S, *et al*: Identification of a 1-cM region of common deletion on 4q35 associated with progression of hepatocellular carcinoma. Genes Chromosomes Cancer 25: 284-289, 1999.
32. Zhang SH, Cong WM, Xian ZH and Wu MC: Features of microsatellite alterations on chromosome 4 in hepatocellular carcinoma. Zhonghua Gan Zang Bing Za Zhi 12: 223-226, 2004.
33. Bluteau O, Beaudoin JC, Pasturaud P, *et al*: Specific association between alcohol intake, high grade of differentiation and 4q34-q35 deletions in hepatocellular carcinomas identified by high resolution allelotyping. Oncogene 21: 1225-1232, 2002.
34. Yeh SH, Lin MW, Lu SF, *et al*: Allelic loss of chromosome 4q21 approximately 23 associates with hepatitis B virus-related hepatocarcinogenesis and elevated alpha-fetoprotein. Hepatology 40: 847-854, 2004.
35. Marchio A, Pineau P, Meddeb M, *et al*: Distinct chromosomal abnormality pattern in primary liver cancer of non-B, non-C patients. Oncogene 19: 3733-3738, 2000.
36. Zhang WD, Hirohashi S, Tsuda H, *et al*: Frequent loss of heterozygosity on chromosomes 16 and 4 in human hepatocellular carcinoma. Jpn J Cancer Res 81: 108-111, 1990.
37. Hammond C, Jeffers L, Carr BI and Simon D: Multiple genetic alterations, 4q28, a new suppressor region, and potential gender differences in human hepatocellular carcinoma. Hepatology 29: 1479-1485, 1999.
38. Bandaru V, Sunkara S, Wallace SS and Bond JP: A novel human DNA glycosylase that removes oxidative DNA damage and is homologous to *Escherichia coli* endonuclease VIII. DNA Repair (Amst) 1: 517-529, 2002.
39. Takao M, Oohata Y, Kitadokoro K, *et al*: Human Nei-like protein NEIL3 has AP lyase activity specific for single-stranded DNA and confers oxidative stress resistance in *Escherichia coli* mutant. Genes Cells 14: 261-270, 2009.
40. Takaoka A, Tamura T and Taniguchi T: Interferon regulatory factor family of transcription factors and regulation of oncogenesis. Cancer Sci 99: 467-478, 2008.
41. Barnes B, Lubyova B and Pitha PM: On the role of IRF in host defense. J Interferon Cytokine Res 22: 59-71, 2002.
42. Mamane Y, Heylbroeck C, Genin P, *et al*: Interferon regulatory factors: the next generation. Gene 237: 1-14, 1999.
43. Taniguchi T, Ogasawara K, Takaoka A and Tanaka N: IRF family of transcription factors as regulators of host defense. Annu Rev Immunol 19: 623-655, 2001.
44. Harada H, Fujita T, Miyamoto M, *et al*: Structurally similar but functionally distinct factors, IRF-1 and IRF-2, bind to the same regulatory elements of IFN and IFN-inducible genes. Cell 58: 729-739, 1989.
45. Harada H, Kitagawa M, Tanaka N, *et al*: Anti-oncogenic and oncogenic potentials of interferon regulatory factors-1 and -2. Science 259: 971-974, 1993.
46. Benech P, Vigneron M, Peretz D, Revel M and Chebath J: Interferon-responsive regulatory elements in the promoter of the human 2',5-oligo(A) synthetase gene. Mol Cell Biol 7: 4498-4504, 1987.
47. Tan RS, Taniguchi T and Harada H: Identification of the lysyl oxidase gene as target of the antioncogenic transcription factor, IRF-1, and its possible role in tumor suppression. Cancer Res 56: 2417-2421, 1996.
48. Tanaka H and Samuel CE: Mechanism of interferon action: structure of the mouse PKR gene encoding the interferon-inducible RNA-dependent protein kinase. Proc Natl Acad Sci USA 91: 7995-7999, 1994.
49. Wang Y, Liu DP, Chen PP, Koeffler HP, Tong XJ and Xie D: Involvement of IFN regulatory factor (IRF)-1 and IRF-2 in the formation and progression of human esophageal cancers. Cancer Res 67: 2535-2543, 2007.
50. Doherty GM, Boucher L, Sorenson K and Lowney J: Interferon regulatory factor expression in human breast cancer. Ann Surg 233: 623-629, 2001.
51. Shimada Y, Saito A, Suzuki M, Takahashi E and Horie M: Cloning of a novel gene (ING1L) homologous to ING1, a candidate tumor suppressor. Cytogenet Cell Genet 83: 232-235, 1998.
52. Nagashima M, Shiseki M, Miura K, *et al*: DNA damage-inducible gene p33ING2 negatively regulates cell proliferation through acetylation of p53. Proc Natl Acad Sci USA 98: 9671-9676, 2001.
53. Doyon Y, Cayrou C, Ullah M, *et al*: ING tumor suppressor proteins are critical regulators of chromatin acetylation required for genome expression and perpetuation. Mol Cell 21: 51-64, 2006.
54. Chin MY, Ng KC and Li G: The novel tumor suppressor p33ING2 enhances UVB-induced apoptosis in human melanoma cells. Exp Cell Res 304: 531-543, 2005.
55. Gozani O, Karuman P, Jones DR, *et al*: The PHD finger of the chromatin-associated protein ING2 functions as a nuclear phosphoinositide receptor. Cell 114: 99-111, 2003.
56. Pedeux R, Sengupta S, Shen JC, *et al*: ING2 regulates the onset of replicative senescence by induction of p300-dependent p53 acetylation. Mol Cell Biol 25: 6639-6648, 2005.
57. Wang J, Chin MY and Li G: The novel tumor suppressor p33ING2 enhances nucleotide excision repair via inducement of histone H4 acetylation and chromatin relaxation. Cancer Res 66: 1906-1911, 2006.
58. Lu F, Dai DL, Martinka M, Ho V and Li G: Nuclear ING2 expression is reduced in human cutaneous melanomas. Br J Cancer 95: 80-86, 2006.
59. Okano T, Gemma A, Hosoya Y, *et al*: Alterations in novel candidate tumor suppressor genes, ING1 and ING2 in human lung cancer. Oncol Rep 15: 545-549, 2006.
60. Zhang HK, Pan K, Wang H, *et al*: Decreased expression of ING2 gene and its clinicopathological significance in hepatocellular carcinoma. Cancer Lett 261: 183-192, 2008.