

## ARTICLE

# Evidence for association with hepatocellular carcinoma at the *PAPSS1* locus on chromosome 4q25 in a family-based study

Wei-Liang Shih<sup>1</sup>, Ming-Whei Yu<sup>\*1</sup>, Pei-Jer Chen<sup>2</sup>, Tai-Wei Wu<sup>1</sup>, Chih-Lin Lin<sup>3</sup>, Chun-Jen Liu<sup>4</sup>, Shi-Ming Lin<sup>5</sup>, Dar-In Tai<sup>5</sup>, Shou-Dong Lee<sup>6</sup> and Yun-Fan Liaw<sup>5</sup>

<sup>1</sup>Graduate Institute of Epidemiology, College of Public Health, National Taiwan University, Taipei, Taiwan; <sup>2</sup>Hepatitis Research Center, National Taiwan University Hospital, Taipei, Taiwan; <sup>3</sup>Department of Gastroenterology, Ren-Ai Branch, Taipei City Hospital, Taipei, Taiwan; <sup>4</sup>Department of Internal Medicine, National Taiwan University Hospital, Taipei, Taiwan; <sup>5</sup>Liver Research Unit, Chang Gung Memorial Hospital, Chang Gung University College of Medicine, Taipei, Taiwan; <sup>6</sup>Division of Gastroenterology, Department of Medicine, Taipei Veterans General Hospital, Taipei, Taiwan

A region on chromosome 4q25 has recently been highlighted as linked to hepatocellular carcinoma (HCC). In this study, we performed a family-based association analysis with 67 single-nucleotide polymorphisms (SNPs) to map this linkage region in 240 families with HCC, 212 (88.3%) of which were ascertained through hepatitis B virus surface antigen (HBsAg)-positive index cases. Individual SNP analysis with correction for multiple testing identified 10 SNPs in two correlated haplotype blocks, located in or around the 3'-phosphoadenosine 5'-phosphosulfate synthetase-1 (*PAPSS1*) gene (all *P*-values: <0.0075). Our linkage data and GIST (Genotype identity-by-descent sharing test) indicate that 6 of these 10 SNPs contributed to the linkage signal. The haplotype block of the strongest association with HCC extended from the intron 5 to the 3'-flanking region of *PAPSS1*; multiple consecutive three-SNP haplotypes in this region were significant. The most significant haplotype showed odd ratios of 3.41 (95% confidence interval (CI) = 1.36–8.53) for homozygous individuals in a case-affected sibling analysis. This haplotype also revealed an association with elevated serum  $\alpha$ -fetoprotein and with poor survival in familial cases and an independent series of HBsAg-positive cases with small tumor present at the time of hospital admission. These results implicate *PAPSS1* as a candidate HCC-susceptibility gene in hepatitis B carriers.

*European Journal of Human Genetics* (2009) 17, 1250–1259; doi:10.1038/ejhg.2009.48; published online 1 April 2009

**Keywords:** chromosome 4q; hepatocellular carcinoma; haplotype; single-nucleotide polymorphisms; hepatitis B; *PAPSS1* gene

\*Correspondence: Professor M-W Yu, Graduate Institute of Epidemiology, College of Public Health, National Taiwan University, Room 522 No.17, Xuzhou Road Zhongzheng District, Taipei City 10055, Taiwan.

Tel: +886 2 332 280 31; Fax: +886 2 235 119 55;

E-mail: yumw@ntu.edu.tw

Received 22 December 2008; revised 17 February 2009; accepted 18 February 2009; published online 1 April 2009

## Introduction

More than half of the hepatocellular carcinoma (HCC) cases worldwide occur in Southeast Asia and sub-Saharan Africa, where hepatitis B virus (HBV) infection is hyper-endemic. The HCC risk is extremely high, with a relative risk of ~20, for chronic HBV carriers when compared with that for non-carriers.<sup>1</sup> However, it was found that less than one-quarter of chronic HBV carriers were affected with HCC in their lifetime.<sup>2</sup> The heterogeneity among chronic

HBV carriers in the variable rates of progression to HCC is the result of many factors, including viral factors, such as viral variants and replication activity, environmental conditions, and host factors.<sup>3–11</sup>

In Taiwan where hepatitis B is hyperendemic, ~15% of patients with HCC have first-degree relatives with a history of HCC. A familial tendency is one of the risk factors for HCC.<sup>2,12</sup> Several candidate gene-association studies on HBV-related HCC have been carried out, but have not found any genetic difference that can account for the underlying mechanisms of heritable causes for the familial clustering of HCC.<sup>7,8,11–16</sup> Using linkage analysis with a dense map of microsatellite markers, we recently identified a HCC-susceptibility locus in the vicinity of the marker D4S3240 on chromosome 4q25.<sup>17</sup> Although this linkage region has been identified as a region of frequent allelic loss in HCC,<sup>18</sup> only a polymorphism in the epidermal growth factor gene on 4q25 has been shown to predispose individuals to HCC among cirrhotic patients who were mostly affected by factors other than HBV.<sup>19</sup>

In this study, we used family-based data with single-nucleotide polymorphisms (SNPs) to search for evidence of association within the 4q25 linkage region. We have provided evidence for an association of SNPs with haplotypes in the 3'-phosphoadenosine 5'-phosphosulfate synthetase-1 (*PAPSS1*) gene with HCC and have also shown that this association explains our earlier reported linkage peak.<sup>17</sup> Using death registry data and another set of unrelated cases, we further showed that *PAPSS1* SNPs were associated with serum  $\alpha$ -fetoprotein (AFP) and HCC survival.

## Materials and methods

The research ethics committee at the College of Public Health, National Taiwan University approved this study, and all subjects provided written informed consent.

### Patients

Patients with HCC were consecutively recruited from the Chang Gung Memorial Hospital, Taipei Veterans General Hospital, and from National Taiwan University Hospital, the three major teaching hospitals that has provided primary to tertiary medical care in Taiwan, since 1997.<sup>2</sup> The diagnosis of HCC was based on either histological or cytological findings, or on elevated serum AFP levels ( $\geq 100$  ng/ml) combined with at least one positive liver image on arteriography, sonography, and/or computed tomography. By 2004, 3204 patients who represented 90% of those who were originally contacted had been recruited with questionnaire interview and from DNA samples extracted from peripheral white blood cells and/or buccal swabs. The annual follow-up for date of death was obtained through data linkage with computer files of the national death registry system.

### Family sample

Relatives aged 25–75 years were recruited through the index cases. All of the participating relatives were interviewed in person, using a structured questionnaire, and underwent a clinical assessment at the time of enrollment. Relatives with HCC were ascertained through an annual follow-up examination and data linkage with computer files of the national cancer and death registry system. The clinical evaluation that was performed at study entry and

**Table 1** Characteristics of the 240 study families with HCC

Characteristic	Result
Number (%) of families ascertained through HBsAg-positive index cases	212 (88.3)
Number of multiplex families	74
Number of case-parent trios	166
Number of discordant sibships	145
Number of affected individuals	183
Number of unaffected individuals	433
Total number of affected individuals	328
Affected individuals genotyped <sup>a</sup>	295
Mean $\pm$ SD of age at diagnosis in years	43.7 $\pm$ 11.3
Number (%) of males/females	252 (85.4)/43 (14.6)
Number (%) of subjects who were HBsAg-positive	258 (87.5)
Number (%) of subjects who were anti-HCV-positive	30 (10.9)
Number (%) of subjects with an alcohol intake $\geq 80$ g/day	35 (12.1)
Total number of unaffected individuals genotyped <sup>b</sup>	708
Number (%) of males/females	293 (41.4)/415 (58.6)
Number (%) of subjects who were HBsAg positive	275 (39.6)
Number (%) of subjects who were anti-HCV positive	24 (3.5)
Number (%) of subjects with an alcohol intake $\geq 80$ g/day	16 (2.4)

HCC, hepatocellular carcinoma.

<sup>a</sup>Twenty individuals had missing anti-HCV status and six individuals had missing quantity of alcohol intake.

<sup>b</sup>Thirteen individuals had missing HBsAg status, 15 individuals had missing anti-HCV status, and 28 individuals had missing quantity of alcohol intake.

during follow-up comprised abdominal ultrasonography, conventional liver biochemical tests, serological tests for HBV surface antigen (HBsAg) and antibodies against hepatitis C virus (anti-HCV) (only at the time of entry), anthropometry, and six tests of cardiovascular disease risk factors.

Families that met the following criteria were included in this study: (1) at least one blood relative with HCC available for genotyping (multiplex families); (2) both parents available for genotyping (trios); or (3) one parent and at least two unaffected siblings available for genotyping. The data set comprised 1003 genotyped individuals (295 affected and 708 unaffected individuals) from 240 families (74 multiplex and 166 singleton families) (Table 1).

### Independent case series

To determine whether genetic variants in the critical region were also associated with patient subgroups and outcome, an additional set of unrelated HBsAg-positive cases under the age of 75 years were selected from the database of 3204 HCC patients for the purpose of survival analysis. This series of cases comprised the first 912 patients who met the inclusion criteria in the database list. Our analysis was restricted to HBsAg carriers who represented 87.5% of the affected individuals in the family sample; hence, we could compare the disease-specific survival between patients unselected for family history and those from families with HCC.

### Selection of SNPs and genotyping

SNPs were identified from public databases (<http://www.hapmap.org/> and <http://snp.ims.u-tokyo.ac.jp/>). First, we genotyped 43 SNPs, with an average inter-marker distance of 146.63 kb across a 6.16-Mb region, which were 1-heterogeneity LOD (logarithm of odds) score down from the linkage peak (at marker D4S3240) of 4q25 in the family sample.<sup>17</sup> When associations with HCC were identified for two SNPs located between 108.69 and 108.79 Mb, the additional 24 SNPs in the surrounding region were then tested for high-resolution association mapping. In an independent case series, we selected a subset of SNPs, which can capture all other untyped SNPs in the critical region.

SNPs were genotyped using the TaqMan technology implemented on an ABI PRISM 7900HT Sequence Detection System (Applied Biosystems, Foster City, CA, USA) according to the manufacturer's protocol. Genotypes of missing relatives were inferred from available genotyping information of more than three offsprings and a spouse. All genotypes were checked for inconsistent Mendelian inheritance using the PedCheck software,<sup>20</sup> and all inconsistencies were eliminated by either retyping or by removal from the analysis.

### Sequencing

To identify any missense SNPs that might be of functional significance in the implicated region, we re-sequenced

**Table 2** Allele tests for associations between SNPs in a 127.45-kb interval at 4q25 and HCC

SNP number	dbSNP number	Distance to next marker (bp)	Context	Allele	Trios			Frequency (%)		PDT		Permutated <sup>b</sup> P-value	FBAT Z	<sup>a</sup> P-value	GIST <sup>a</sup> P-value
					T	NT	NT	Affected	Unaffected	Discordant siblings	<sup>a</sup> P-value				
16	rs1514728	0	Intergenic	C	44.2	54.9	42.2	50.0	0.0025	0.0020	0.0413	0.0793			
17	rs3936010	13,272	Intergenic	C	11.5	15.0	13.9	14.0	0.6644	0.6233	0.3328	0.2165			
18	rs1355990	4508	Intergenic	C	46.2	55.1	42.8	50.5	0.0038	0.0035	0.0391	0.0975			
19	rs932100	14,218	Intergenic	G	44.9	35.2	46.8	38.7	0.0046	0.0037	0.0544	0.0082			
20	rs4451005	7187	Intergenic	A	43.5	32.8	42.9	35.3	0.0025	0.0013	0.0265	0.0551			
21	rs2726165	9245	Intergenic	T	11.3	15.1	10.9	10.2	0.9555	1.0000	0.1533	0.9901			
22	rs9569	12,319	3'-UTR	T	42.7	33.0	45.8	36.7	0.0015	0.0006	0.0316	0.0396			
23	rs2726670	1050	Intron 11	A	53.1	50.0	58.1	48.9	0.0063	0.0045	0.2521	0.2240			
24	rs2726673	4897	Intron 11	C	15.0	14.5	12.5	11.1	0.4504	0.4246	0.455	0.9893			
25	rs4956112	11,321	Intron 11	A	43.9	32.2	45.5	36.1	0.0008	0.0006	0.0250	0.0306			
26	rs2726659	12,612	Intron 10	C	12.5	13.0	11.5	10.6	0.6724	0.7139	0.3731	0.9893			
27	rs11934943	394	Intron 10	A	44.1	33.2	44.5	36.9	0.0025	0.0009	0.0216	0.0147			
28	rs2271590	9496	Intron 8	T	43.9	32.2	43.3	36.1	0.0032	0.0022	0.0175	0.0077			
29	rs973213	5102	Intron 6	T	14.0	13.5	12.2	10.8	0.4458	0.4194	0.7200	0.9893			
30	rs4956024	12,458	Intron 5	T	46.4	31.5	46.4	37.1	0.0004	0.0002	2.990	0.0081			
31	rs3805346	9371	Intron 5	C	12.9	15.1	11.0	10.5	0.9554	1.0000	0.3671	0.9901			

PDT, pedigree disequilibrium test; T, transmission; NT, non-transmission; FBAT, family-based association test; GIST, genotype identity-by-descent sharing test

<sup>a</sup>Nominal P-values.

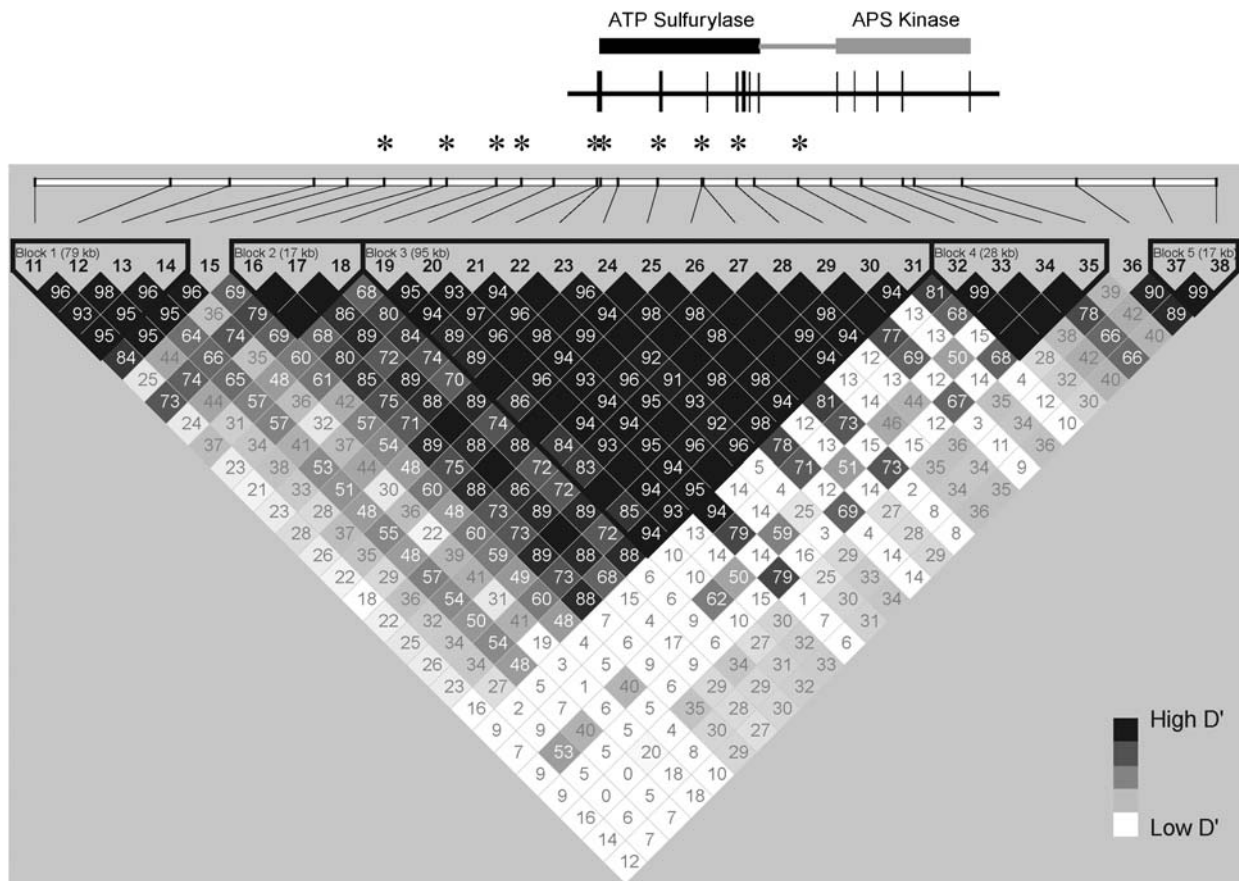
<sup>b</sup>Derived from 10 000 permutations.

each exon from 6 to 12 of the *PAPSS1* gene in DNAs from 48 unrelated individuals (including 40 HBsAg-positive cases and 8 unaffected HBsAg carriers randomly selected from our study subjects). This sample size for analysis was deliberately chosen (96 chromosomes) to identify the common variants with a frequency of >5% and a probability of 95%. The coding regions and at least 60bp of each flanking intronic sequence were amplified using polymerase chain reaction and sequenced using conventional dye-primer sequencing on the ABI3730 DNA Analyzer System (Applied Biosystems).

**Statistical methods**

A Hardy–Weinberg disequilibrium test was performed for each SNP in 270 unrelated founders, using the SAS software with 10 000 Monte Carlo simulations to obtain the exact *P*-value estimates. The Haploview version 4.1 software was used to estimate pairwise linkage disequilibrium (measured as Lewontin’s standardized disequilibrium coefficient, *D'*). Haplotype blocks were defined using the default confidence interval (CI) algorithm.<sup>21</sup> We performed the

pedigree disequilibrium test (PDT) as implemented in the PDTPHASE program within the UNPHASED (version 2.404) suite of programs to examine associations between SNPs or haplotypes and HCC.<sup>22</sup> The reported *P*-values were based on the PDT-sum statistic. In addition to PDT, we also included the family-based association test (FBAT) and the haplotype-based association test using the FBAT software version 1.7.2.<sup>23</sup> The complementary use of the additional method will reduce the chance for false positive findings. We calculated both nominal and empirical *P*-values that were derived from 10 000 simulating replicates. The Benjamini-corrected false discovery rate was used to correct for multiple testing.<sup>24</sup> After adjustment for multiple testing and with a false discovery rate of ≤0.05, the cutoff for a significant association was *P* = 0.0075 for single-locus analyses. The GIST (Genotype identity-by-descent sharing test) was performed using affected-sibship data to test whether our earlier described evidence for linkage on 4q25 could be explained by any of the tested SNPs.<sup>25</sup> We used conditional logistic regression on sibship data with family as the stratification variable to estimate the odds



**Figure 1** Schematic representation of the *PAPSS1* gene showing locations of the ATP-sulfurylase and adenosine 5'-phosphosulfate (APS) kinase domains and the haplotype blocks. Values for *D'* (original magnification, ×100) are shown. The five-color scheme (white to black) represents the increasing strength of linkage disequilibrium. SNPs significantly associated with HCC after multiple testing adjustments are indicated with an asterisk (\*).

ratio and 95% CI for the association between implicated haplotypes and HCC.

Survival was calculated from the date of hospital admission to the last reported search for death entries (31 December 2005) in the national death certification system. Patients whose causes of death were not due to HCC or cirrhosis were censored in the survival analysis. The Kaplan–Meier method was used to generate survival curves, and the Wilcoxon test was used to compare the survival curves between groups. Hazard ratios were calculated using the Cox regression model with adjustment for putative risk factors and clinical features. The analyses were performed using the SAS version 9.1 software (SAS Institute, Cary, NC, USA). All statistical tests were two-sided.

**Results**

**Characteristics of study family and SNPs**

Of the 240 study families, 212 (88.3%) were ascertained through HBsAg-positive index cases. HCC occurred more frequently in males than in females. The mean age (±SD) at diagnosis of HCC was 43.7 (±11.3) years. A majority of the subjects were HBsAg carriers (87.5% affected and 39.6% unaffected individuals). The positive rate of the antibodies against hepatitis C virus was 10.9% in the affected subjects and 3.5% in the unaffected (Table 1). None of the 67 genotyped SNPs showed evidence of deviation from the Hardy–Weinberg equilibrium. The minor allele frequency for these SNPs was between 0.067 and 0.50 (Supplementary Table 1).

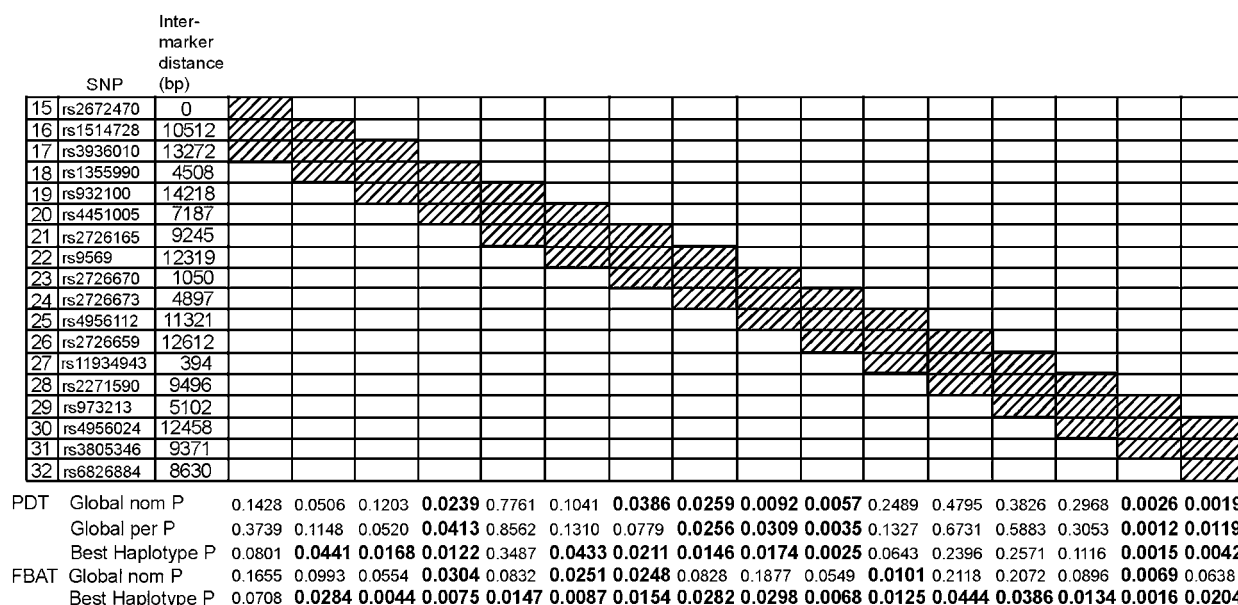
**Family-based single-locus association analysis**

The initial assessment of the 43 SNPs by PDT revealed evidence of association for 5 SNPs (rs1514728, *P* = 0.0025; rs2271590, *P* = 0.0032; rs12512339, *P* = 0.0350; rs2704108, *P* = 0.0371; and rs2739206, *P* = 0.0421), of which the 2 most significant SNPs resided in the region flanking or within the *PAPSS1* gene. This prompted us to analyze an additional 24 SNPs in a 300 kb interval (108.63–108.93 Mb) with the *PAPSS1* gene as the principal target. We then identified 8 other SNPs with nominal *P*-values from 0.0004 to 0.0063 (permuted *P* = 0.0002–0.0045). The 10 significant SNPs in the region surrounding *PAPSS1* remained significant after correction for multiple testing. Although the level of significance was lower in the FBAT analysis compared with that in the PDT results, the same pattern of association was observed (Table 2).

With the exception of rs9569, which is in the 3'-untranslated region of *PAPSS1*, 9 of the 10 HCC-associated SNPs were in the intronic or intergenic region without any known function. We applied specialized algorithms (including miRBase,<sup>26</sup> microRNA.org,<sup>27</sup> PupaSuite,<sup>28</sup> and miRNAmap<sup>29</sup>) with the default parameters included in the software to investigate whether rs9569 resides in a microRNA target site. The results indicated that microRNAs, hsa-mir-520d-5p and hsa-mir-421, can bind to a site containing rs9569.

**SNPs potentially explaining the observed linkage**

According to the GIST analysis under the additive weighting scheme, the genotypes observed for 6 of these



**Figure 2** Haplotype analyses for all three SNP sliding windows over blocks 2 and 3. Global nom P, global nominal P-value; Global per P, permutation-based global haplotype P-value. P-value of 0.05 or less. Significant P-values are shown in boldface. PDT, pedigree disequilibrium test; FBAT, family-based association test.

10 implicated SNPs located in or around *PAPSS1* could be responsible for the linkage of HCC to marker D4S3240 on 4q25 in our sample (Table 2). Investigating the same issue using a dominant or recessive model yielded similar results (data not shown).

### Linkage disequilibrium structure

The region associated with HCC spanned 127.45 kb, extending from the intron 5 to the 3'-flanking region of *PAPSS1*. Two highly correlated haplotype blocks (multi-allelic  $D' = 0.72$  between blocks 2 and 3) were identified across this region (Figure 1).

The fractions of all common (minor allele frequency  $\geq 0.05$ ) SNPs (Han Chinese HapMap release #23a) that were captured by the tested SNPs were 0.95 and 0.86, respectively, for blocks 2 and 3. The haplotype block structure surrounding *PAPSS1* in our sample is comparable with the structure defined using the HapMap SNPs. To confirm the critical region for HCC-susceptibility locus, we then searched the HapMap data to identify additional SNPs that were in  $r^2 \geq 0.8$  with any of the SNPs in blocks 2 and 3. Although an exhaustive search with a 1.25-Mb window surrounding *PAPSS1* was carried out, none of the SNPs outside the critical region met this criterion (data are available from the author on request).

### Family-based haplotype association analysis

To prioritize regions of interest, we also performed sliding window haplotype analysis of varying sizes from 2 to 4 SNPs per window on SNPs 15–32, which covered the associated region. The same pattern was observed using sliding windows of 2–4 SNPs. Multiple haplotypes showed evidence for association with HCC. The haplotype at SNPs 29–30–31 revealed the minimum *P*-value across all windows (global permutation  $P = 0.0012$ , with the corrected *P*-value approaching significance (cutoff  $P$ -value = 0.0010) after multiple testing adjustment for 48 tests) and high significance for both PDT and FBAT (Figure 2).

### Effect of *PAPSS1* SNPs

To estimate the magnitude of all potential effects on the risk of HCC associated with *PAPSS1* genetic variants, we performed a conditional logistic regression on 145 sibships discordant for HCC (183 affected and 433 unaffected individuals) stratified by family. The matched odds ratios of HCC were calculated for haplotypes of SNPs 29–30–31, the marker combination leading to the lowest global *P*-value in the PDT/FBAT analyses. The odds ratios for heterozygotes and homozygotes of the C–T–T haplotype were 1.55 (95% CI = 0.81–2.98) and 3.41 (95% CI = 1.36–8.53), respectively, as compared with that for individuals without this haplotype.

We selected 10 (including 6 significant SNPs identified by the family sample) SNPs, which captured a 100%

**Table 3** Odds ratios (ORs) of haplotypes of SNPs 29–30–31 in the *PAPSS1* gene with the risk of HCC and elevated serum AFP levels (ng/mL)<sup>a</sup>

Haplotype <sup>b</sup>	Allele frequency (%)		Case-unaffected sibling analysis				Analysis in unrelated case series					
	Controls	Cases	OR <sup>c</sup>	P-value	OR <sup>c</sup>	Two copies (95% CI)	OR <sup>d</sup>	One copy (95% CI)	P-value	OR <sup>d</sup>	Two copies (95% CI)	P-value
CCT	53.6	45.5	0.35	0.0030	0.32	(0.13–0.82)	1.02	(0.72–1.44)	0.9297	0.82	(0.56–1.20)	0.3014
CCT	36.5	43.1	1.55	0.1840	3.41	(1.36–8.53)	1.48	(1.10–2.01)	0.0109	1.18	(0.78–1.79)	0.4285
TCC	9.6	10.5	0.64	0.3222	1.04	(0.06–17.01)	0.96	(0.67–1.38)	0.8330	0.75	(0.22–2.56)	0.6477

CI, confidence interval; AFP,  $\alpha$ -fetoprotein.

<sup>a</sup>Haplotype consisting of marker rs973213, rs4956024, and rs3805346.

<sup>b</sup>Haplotypes with frequencies  $> 0.01$  are listed.

<sup>c</sup>ORs were calculated with 'no copies' as the reference group from conditional logistic regression (stratified according family) adjusting age (continuous variable) and sex.

<sup>d</sup>ORs were calculated with 'no copies' as the reference group from unconditional logistic regression adjusting age and sex.

genetic variation of the 16 SNPs in the critical region (blocks 2 and 3) for survival analysis in the independent series of 912 unrelated cases. Several SNPs showed associations with increased AFP levels (SNP17,  $P=0.0459$ ; SNP19,  $P=0.0152$ ; SNP27,  $P=0.0261$ ; and SNP30,  $P=0.0145$ ) or with worse survival (SNP17,  $P=0.0173$ ; SNP19,  $P=0.0183$ ; SNP22,  $P=0.0325$ ; and SNP30,  $P=0.0076$ ) in cases with small tumor ( $\leq 2$  cm) present at hospital admission. Consistent with the family study, we found SNP30 to be the most associated SNP. The C–T–T haplotype at SNPs 29–30–31 was also associated with an elevation of the serum AFP levels (Table 3). None of the haplotypes were associated with tumor size or number of lesions (data not shown).

Characteristics and clinical features of the cases included in the survival analysis are shown in Table 4. The median age at diagnosis of HCC was lesser in non-familial cases from singleton families than in familial cases and the independent set of unrelated cases unselected for family history. Among unrelated cases with small tumor ( $\leq 2$  cm) present at hospital admission, those with haplotype-1

C–T–T of SNPs 29–30–31 had an increased hazard ratio of 1.74 (95% CI=1.12–2.71) for death when compared with cases without haplotype-1. For familial cases with small tumor, the hazard ratio of haplotype-1 for death was 4.25 (95% CI=1.01–17.86) (Figure 3). We did not find an association between survival and haplotype-1 in non-familial cases or in cases with tumor  $> 2$  cm.

### Sequence analysis

No missense variants were detected by re-sequencing all coding regions of exons 6–12 of *PAPSS1* among 48 individuals, 40 of whom were cases. Five non-coding/silent variants were found, including three earlier reported SNPs and two insertion events (Supplementary Table 2).

### Discussion

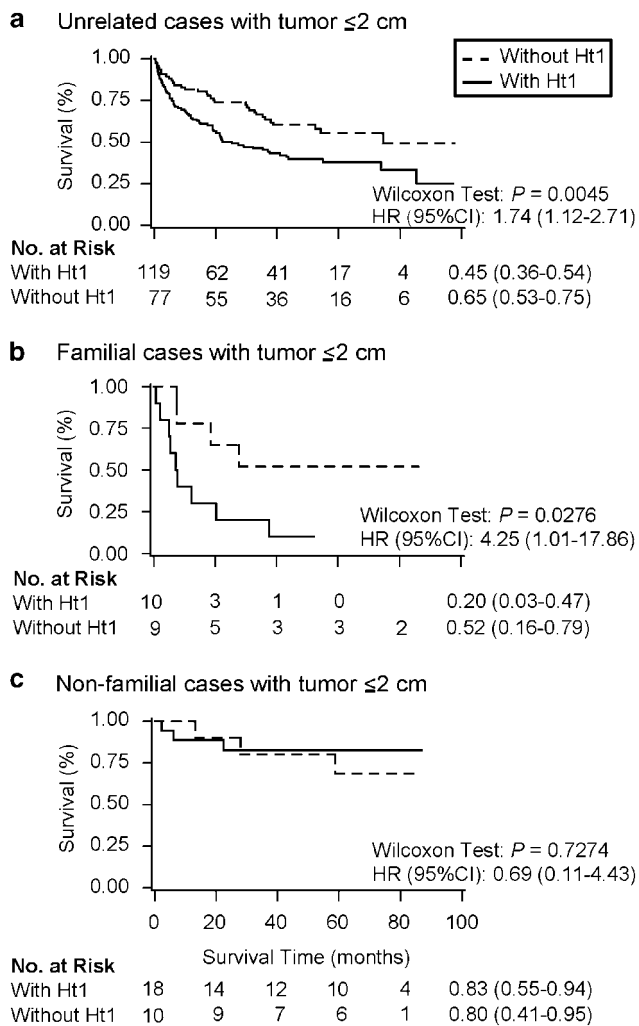
In this study, we detected an association of the *PAPSS1* gene with HCC in a family sample. This gene is located  $\sim 1.15$  Mb away from the linkage peak at D4S3240 in the

**Table 4** Characteristics of HCC cases in the survival analysis<sup>a</sup>

Characteristic	Unrelated cases (n = 912) n (%)	Familial cases (n = 85) n (%)	Non-familial cases (n = 142) n (%)
<i>At hospital admission</i>			
Age at diagnosis, median (range) in years	52.0 (18.0–69.0)	49.0 (20.0–69.0)	38.0 (16.0–61.0)
Male/Female	775 (85.0)/137 (15.0)	76 (89.4)/9 (10.6)	125 (88.0)/17 (12.0)
First-degree family history of HCC	143 (15.7)	85 (100.0)	0 (0.0)
HBsAg-positive	912 (100.0)	76 (89.4)	124 (87.3)
Anti-HCV-positive	84 (9.5)	11 (13.3)	10 (7.8)
Missing	26	2	14
Daily alcohol intake $\geq 80$ g	105 (11.5)	9 (10.6)	19 (13.6)
Missing	2	0	2
<i>Tumor size (cm)</i>			
$\leq 2$	217 (23.8)	20 (23.5)	29 (20.4)
2.1–5	297 (32.6)	29 (34.1)	43 (30.3)
$> 5$	398 (43.6)	36 (42.4)	70 (49.3)
<i>No. of lesions</i>			
1	516 (57.1)	51 (62.2)	71 (51.4)
2–3	101 (11.2)	8 (9.8)	16 (11.6)
$\geq 4$	286 (31.7)	23 (28.0)	51 (37.0)
Missing	9	3	4
<i>Serum AFP levels (ng/ml)</i>			
$\leq 100$	195 (21.5)	29 (35.4)	48 (34.5)
101–400	191 (21.1)	18 (21.9)	17 (12.2)
$> 400$	520 (57.4)	35 (42.7)	74 (53.3)
Missing	6	3	3
<i>Follow-up</i>			
Range of follow-up time (months)	0.1–99.4	0.3–97.9	0.1–89.1
<i>Deceased</i>			
HCC	597 (87.2)	56 (87.5)	80 (83.3)
Liver cirrhosis	55 (8.0)	3 (4.7)	5 (5.2)
Other causes	33 (4.8)	5 (7.8)	11 (11.5)

HCC, hepatocellular carcinoma; AFP,  $\alpha$ -fetoprotein.

<sup>a</sup>Unrelated cases are an independent series of cases. Familial (in multiplex families) and non-familial (in singleton families) cases are from the family sample. Cases with data on tumor size were included for analysis.



**Figure 3** Haplotype 1 (Ht1, C–T–T at SNPs 29–30–31) and HCC survival among cases with small tumor ( $\leq 2$  cm) present at the time of hospital admission. Unrelated cases (a) are an independent series of cases. Familial (in multiplex families) (b) and non-familial (in singleton families) (c) cases are from the family sample. The 3-year survival (95% confidence interval (CI)) for each group of cases, Wilcoxon  $P$ -value and hazard ratios (HRs) of death for cases with Ht1 *versus* those without Ht1 are presented. HRs were adjusted for age (continuous variable), sex, cigarette smoking (yes *versus* no), alcohol abuse ( $\geq 80$  *versus*  $< 80$  g/day), anti-HCV status, serum  $\alpha$ -fetoprotein levels ( $> 400$  *versus*  $\leq 400$  ng/ml), and number of lesions (single *versus* multiple).

4q25 region. Several points suggest that our results may reflect a genuine association between *PAPSS1* and HCC. First, the association was identified using family-based association analysis, which was robust against false positive associations from cryptic population stratification.<sup>30</sup> Second, multiple SNPs, which belonged to two correlated haplotype blocks, were associated with HCC, and the  $P$ -value associated with 10 SNPs was significant on correction for multiple testing. Third, we showed that the observed associations with HCC for multiple *PAPSS1* SNPs at least partially accounted for the linkage signal on 4q25.

In addition, the observed association between *PAPSS1* SNPs and HCC is unlikely because of long-range linkage disequilibrium with neighboring genes as the structure of haplotype blocks in the critical region, where multiple SNPs were found to be associated with HCC, are comparable with that of the blocks defined by HapMap in Han Chinese population; and we did not detect any  $r^2 \geq 0.8$  between alleles of SNPs in the critical region and those in other genes by using a 1.25-Mb window surrounding *PAPSS1*.

We also found an association for a *PAPSS1* haplotype with high serum AFP levels and with poor HCC survival in an independent set of HBsAg-positive cases with small tumor present at hospital admission. Homozygosity for this haplotype increased the risk of developing HCC threefold, estimated by using the case-sibling matched design. Interestingly, familial cases with small tumor also had worse survival associated with harboring the same haplotype. It is unclear why no such association was observed in non-familial cases from singleton families. However, these cases were substantially younger than the other two groups of cases. Their survival also seemed to be longest among cases.

The *PAPSS1* gene contains 12 exons, including the adenosine 5'-triphosphate (ATP) sulfurylase and adenosine 5'-phosphosulfate kinase domains. *PAPSS1* catalyzes the synthesis of PAPS through reactions of ATP sulfurylase and adenosine 5'-phosphosulfate kinase. There are two *PAPSS* isoforms, *PAPSS1* and *PAPSS2*. *PAPSS1* is expressed in virtually all human tissues.<sup>31,32</sup> PAPS is a co-substrate for all sulfation reactions, which is a key step in the metabolism of a broad range of compounds, of which some have been implicated in cancer; these include bile acids, phenolic xenobiotics, endogenous and exogenous estrogens, and certain classes of environmental carcinogens, such as hydroxylated aromatic amines.<sup>33–35</sup> In addition, sulfation of extracellular matrix is of crucial importance for a multitude of processes, such as extracellular signaling and adhesion of leukocytes to endothelial cells at the site of inflammation.<sup>34,36</sup> It was recently shown that *PAPSS1* influenced retroviral replication by affecting a step during provirus establishment.<sup>37</sup>

It is noted that multiple SNPs associated with HCC clustered within the *PAPSS1* ATP sulfurylase domain that included exons 6–12 and 3'-untranslated region. Two rare missense variants in this domain have been reported in other races; one of which can change *in vitro* substrate kinetic properties.<sup>32</sup> We did not find any missense variant in the ATP sulfurylase domain. The lack of coding SNPs perhaps suggests that regulatory SNPs that influence gene expression might be involved, although we may not have re-sequenced enough cases to identify the functional, causal variants. In this study, we identified a significant association with HCC for the 3'-untranslated region SNP rs9569, which resides in putative microRNA binding site. Given the functional nature of rs9569, it may be the most suspected causal variant for HCC that lies in our implicated



region. However, the haplotypes that are significantly associated with HCC almost completely span the critical region. We have not been able to implicate a single causative variant, because of the tight linkage disequilibrium in the region. By performing a sliding window haplotype analysis, we found that the most significant association was at SNPs 29–30–31, in which only SNP30 showed significant association with HCC. Therefore, the association with C–T–T haplotype might be derived from SNP30. Taken together, the haplotype-based analysis supports the association between single SNP and HCC.

In conclusion, we have shown that polymorphisms within the ATP sulfurylase domain of the *PAPSS1* gene are associated with HCC in a family sample enriched with familial or early-onset cases. We have also shown that an allelic combination in a haplotype significantly increases HCC susceptibility, and is associated with a poor prognosis of HCC. These associations were seen mostly in HBV carriers. Considering the recent discovery that *PAPSS1* plays a role in retroviral replication,<sup>37</sup> *PAPSS1* polymorphisms may modify the response to viral infection. Perhaps it is plausible to suggest that a similar mechanism involving *PAPSS1* may occur in HBV infection. Our study is the first to suggest the importance of *PAPSS1* in HCC. Future genetic and functional studies are necessary to further validate our findings.

#### Acknowledgements

This work was supported by National Research Program for Genomic Medicine (Grant nos. NSC94-3112-B-002-017, NSC95-3112-B-002-001, NSC96-3112-B-002-001, and NSC95-2314-B-002-244-MY3) from the National Science Council, Taiwan. W-L Shih was the recipient of an investigator award of the Liver Disease Prevention and Treatment Research Foundation. The authors acknowledge the technical supports provided by the Sequencing Core Facility of the National Yang-Ming University Research Center. The authors declare no conflict of interests.

#### Conflict of interest

The authors declare no conflict of interest.

#### References

- Parkin DM: International variation. *Oncogene* 2004; **23**: 6329–6340.
- Yu MW, Chang HC, Liaw YF *et al*: Familial risk of hepatocellular carcinoma among chronic hepatitis B carriers and their relatives. *J Natl Cancer Inst* 2000; **92**: 1159–1164.
- Yu MW, Yeh SH, Chen PJ *et al*: Hepatitis B virus genotype and DNA level and hepatocellular carcinoma: a prospective study in men. *J Natl Cancer Inst* 2005; **97**: 265–272.
- Wu CF, Yu MW, Lin CL *et al*: Long-term tracking of hepatitis B viral load and the relationship with risk for hepatocellular carcinoma in men. *Carcinogenesis* 2008; **29**: 106–112.
- Chou YC, Yu MW, Wu CF *et al*: Temporal relationship between hepatitis B virus enhancer II/basal core promoter sequence variation and risk of hepatocellular carcinoma. *Gut* 2008; **57**: 91–97.
- Jee SH, Ohrr H, Sull JW *et al*: Cigarette smoking, alcohol drinking, hepatitis B, and risk for hepatocellular carcinoma in Korea. *J Natl Cancer Inst* 2004; **96**: 1851–1856.
- El-Serag HB, Rudolph L: Hepatocellular carcinoma: epidemiology and molecular carcinogenesis. *Gastroenterology* 2007; **132**: 2557–2576.
- Yu MW, Yang YC, Yang SY *et al*: Hormonal markers and hepatitis B virus-related hepatocellular carcinoma risk: a nested case-control study among men. *J Natl Cancer Inst* 2001; **93**: 1644–1651.
- Yu MW, Chang HC, Chang SC *et al*: Role of reproductive factors in hepatocellular carcinoma: impact on hepatitis B- and C-related risk. *Hepatology* 2003; **38**: 1393–1400.
- Yu MW, Horng IS, Hsu KH *et al*: Plasma selenium levels and risk of hepatocellular carcinoma among men with chronic hepatitis virus infection. *Am J Epidemiol* 1999; **150**: 367–374.
- Yu MW, Chiu YH, Chiang YC *et al*: Plasma carotenoids, glutathione S-transferase M1 and T1 genetic polymorphisms, and risk of hepatocellular carcinoma: independent and interactive effects. *Am J Epidemiol* 1999; **149**: 621–629.
- Yu MW, Yang SY, Chiu YH *et al*: A p53 genetic polymorphism as a modulator of hepatocellular carcinoma risk in relation to chronic liver disease, familial tendency, and cigarette smoking in hepatitis B carriers. *Hepatology* 1999; **29**: 697–702.
- Yoon VJ, Chang HY, Ahn SH *et al*: MDM2 and p53 polymorphisms are associated with the development of hepatocellular carcinoma in patients with chronic hepatitis B virus infection. *Carcinogenesis* 2008; **29**: 1192–1196.
- Yu MW, Pai CI, Yang SY *et al*: Role of N-acetyltransferase polymorphisms in hepatitis B related hepatocellular carcinoma: impact of smoking on risk. *Gut* 2000; **47**: 703–709.
- Yu MW, Yang SY, Pan IJ *et al*: Polymorphisms in XRCC1 and glutathione S-transferase genes and hepatitis B-related hepatocellular carcinoma. *J Natl Cancer Inst* 2003; **95**: 1485–1488.
- Chen CC, Yang SY, Liu CJ *et al*: Association of cytokine and DNA repair gene polymorphisms with hepatitis B-related hepatocellular carcinoma. *Int J Epidemiol* 2005; **34**: 1310–1318.
- Shih WL, Yu MW, Chen PJ *et al*: Localization of a susceptibility locus for hepatocellular carcinoma to chromosome 4q in a hepatitis B hyperendemic area. *Oncogene* 2006; **25**: 3219–3224.
- Buendia MA: Genetics of hepatocellular carcinoma. *Semin Cancer Biol* 2000; **10**: 185–200.
- Tanabe KK, Lemoine A, Finkelstein DM *et al*: Epidermal growth factor gene functional polymorphism and the risk of hepatocellular carcinoma in patients with cirrhosis. *JAMA* 2008; **299**: 53–60.
- O'Connell JR, Weeks DE: PedCheck: a program for identification of genotype incompatibilities in linkage analysis. *Am J Hum Genet* 1998; **63**: 259–266.
- Barrett JC, Fry B, Maller J *et al*: Haploview: analysis and visualization of LD and haplotype maps. *Bioinformatics* 2005; **21**: 263–265.
- Dudbridge F: Pedigree disequilibrium tests for multilocus haplotypes. *Genet Epidemiol* 2003; **25**: 115–121.
- Laird NM, Horvath S, Xu X: Implementing a unified approach to family-based tests of association. *Genet Epidemiol* 2000; **19** (Suppl. 1): 36–42.
- Benjamini Y, Hochberg Y: Controlling the false discovery rate: a practical and powerful approach to multiple testing. *J R Stat Soc* 1995; **57**: 289–300.
- Li C, Scott LJ, Boehnke M: Assessing whether an allele can account in part for a linkage signal: the Genotype-IBD Sharing Test (GIST). *Am J Hum Genet* 2004; **74**: 418–431.
- Griffiths-Jones S, Grocock RJ, van Dongen S *et al*: miRBase: microRNA sequences, targets and gene nomenclature. *Nucleic Acids Res* 2006; **34**: D140–D144.
- Betel D, Wilson M, Gabow A *et al*: The microRNA.org resource: targets and expression. *Nucleic Acids Res* 2008; **36**: D149–D153.
- Conde L, Vaquerizas JM, Dopazo H *et al*: PupaSuite: finding functional SNPs for large-scale genotyping purposes. *Nucleic Acids Res* 2006; **34**: W621–W625.

- 29 Hsu PW, Huang HD, Hsu SD *et al*: miRNAMap: genomic maps of microRNA genes and their target genes in mammalian genomes. *Nucleic Acids Res* 2006; **34**: D135–D139.
- 30 Reich DE, Goldstein DB: Detecting association in a case-control study while correcting for population stratification. *Genet Epidemiol* 2001; **20**: 4–16.
- 31 Xu ZH, Otterness DM, Freimuth RR *et al*: Human 3'-phosphoadenosine 5'-phosphosulfate synthetase 1 (PAPSS1) and PAPSS2: gene cloning, characterization and chromosomal localization. *Biochem Biophys Res Commun* 2000; **268**: 437–444.
- 32 Xu ZH, Thomae BA, Eckloff BW *et al*: Pharmacogenetics of human 3'-phosphoadenosine 5'-phosphosulfate synthetase 1 (PAPSS1): gene resequencing, sequence variation, and functional genomics. *Biochem Pharmacol* 2003; **65**: 1787–1796.
- 33 Banoglu E: Current status of the cytosolic sulfotransferases in the metabolic activation of promutagens and procarcinogens. *Curr Drug Metab* 2000; **1**: 1–30.
- 34 Strott CA: Sulfonation and molecular action. *Endocr Rev* 2002; **23**: 703–732.
- 35 Allali-Hassani A, Pan PW, Dombrowski L *et al*: Structural and chemical profiling of the human cytosolic sulfotransferases. *PLoS Biol* 2007; **5**: e97, doi:10.1371/journal.pbio.0050097.
- 36 Muramatsu T: Essential roles of carbohydrate signals in development, immune response and tissue functions, as revealed by gene targeting. *J Biochem* 2000; **127**: 171–176.
- 37 Bruce JW, Ahlquist P, Young JA: The host cell sulfonation pathway contributes to retroviral infection at a step coincident with provirus establishment. *PLoS Pathog* 2008; **4**: e1000207.

Supplementary Information accompanies the paper on European Journal of Human Genetics website (<http://www.nature.com/ejhg>)