

ONCOGENOMICS

Localization of a susceptibility locus for hepatocellular carcinoma to chromosome 4q in a hepatitis B hyperendemic area

W-L Shih^{1,7}, M-W Yu^{1,7}, P-J Chen^{2,7}, S-H Yeh^{2,7}, M-T Lo¹, H-C Chang³, Y-F Liaw⁴, S-M Lin⁴, C-J Liu², S-D Lee⁵, C-L Lin⁶, CK Hsiao¹, S-Y Yang¹ and C-J Chen¹

¹Graduate Institute of Epidemiology, College of Public Health, National Taiwan University, Taipei, Taiwan; ²Hepatitis Research Center, National Taiwan University Hospital, Taipei, Taiwan; ³Department of Internal Medicine, Shin Kong Wu Ho-Su Memorial Hospital, Taipei, Taiwan; ⁴Liver Research Unit, Chang-Gung Memorial Hospital, Chang-Gung University, Taoyuan, Taiwan;

⁵Department of Medicine, Veterans General Hospital and School of Medicine, National Yang-Ming University, Taipei, Taiwan and

⁶Division of Gastroenterology, Department of Internal Medicine, Taipei Municipal Jen-Ai Hospital, Taipei, Taiwan

Chromosome 4q is one of the most common regions with a high frequency of allelic loss in hepatocellular carcinoma (HCC). To identify the HCC-susceptibility locus on chromosome 4q, we have performed linkage and family-based association analyses on Chinese families with HCC from Taiwan, where hepatitis B is hyperendemic. Using 77 microsatellite markers spanning chromosome 4q on 52 multiplex families, we found suggestive evidence of linkage to 4q22.3–28.1 with a maximum two-point heterogeneity LOD (HLOD) score of 2.55 at marker D4S3240 on chromosome 4q25. Multipoint analyses with microsatellite markers in the region 4q22.3–28.1 resulted in a maximum HLOD score of 3.12 and a maximum nonparametric linkage (NPL) Z score of 1.98 (pointwise $P=0.0080$; region-wide empirical $P=0.021$) for D4S3240. The evidence for linkage to D4S3240 was seen mostly in a subset of 28 families lacking affected parents, which showed multipoint HLOD and NPL scores of 3.25 and 2.79 (pointwise $P=0.0028$; region-wide empirical $P=0.008$), respectively. Family-based association analyses of the 77 microsatellite markers in 191 families (53 multiplex plus 138 singleton families) using the pedigree disequilibrium test provide further support for observed linkage. Additional genotyping in the 52 multiplex families informative for linkage analyses was performed for 29 single-nucleotide polymorphisms around D4S3240. A common haplotype (at markers rs7442180 and rs221330) positioned ~873 kb away from D4S3240 was associated with HCC, with $P=0.0074$.

Oncogene (2006) 25, 3219–3224. doi:10.1038/sj.onc.1209345; published online 16 January 2006

Keywords: Chromosome 4q; family-based association analysis; hepatocellular carcinoma; linkage analysis; microsatellite markers; single-nucleotide polymorphism

Introduction

Hepatocellular carcinoma (HCC) is a highly fatal malignant neoplasm, with the majority of cases occurring in chronic hepatitis B virus (HBV) carriers (Befeler and Di Bisceglie, 2002; Wands, 2004). The HCC risk is 20-fold higher for individuals who are seropositive for HBV surface antigen (HBsAg) than for those who are HBsAg-seronegative (Yu and Chen, 1994). In high incidence areas, familial aggregation of HCC is closely associated with perinatal transmission of HBV (Beasley, 1982). However, epidemiological data suggest that a genetic component may also be involved in the familial clustering of HCC. For example, HBsAg carriers with a first-degree-affected relative have a 2.4-fold increase in risk for developing HCC compared with HBsAg carriers without a family history of HCC (Yu *et al.*, 2000). Segregation analyses revealed that clustering of HCC in Chinese families could be best explained by autosomal recessive inheritance of a susceptibility gene that may act in conjunction with HBV to increase risk for HCC (Shen *et al.*, 1991; Cai *et al.*, 2003). The underlying molecular defect(s) in these familial clusters of HCC has not been elucidated.

In HCCs, frequent loss of heterozygosity (LOH) has been reported on at least 11 different chromosomal arms, including 1p, 1q, 4q, 5q, 6q, 8p, 9p, 13q, 16p, 16q, and 17p (Thorgeirsson and Grisham, 2002). LOH at microsatellite markers on chromosome 4q was observed in more than 40% of sporadic HCCs, and high rates of LOH on this region were associated with HBV infection (Yeh *et al.*, 1996; Piao *et al.*, 1998; reviewed in Buendia, 2000; Okabe *et al.*, 2000; Laurent-Puig *et al.*, 2001; Bluteau *et al.*, 2002; reviewed in Thorgeirsson and Grisham, 2002). Allelic loss at 4q is of great interest because it is also prevalent in cirrhotic nodules, which have long been considered to be a premalignant lesion of HCC (Yeh *et al.*, 2001), and chromosome 4q contains genes encoding growth factors (e.g. epidermal growth factor) or genes expressed predominantly in the liver (e.g. albumin, α -fetoprotein, alcohol dehydrogenase, UDP-glucuronyl-transferase). Identification of the

Correspondence: Professor M-W Yu, Graduate Institute of Epidemiology, College of Public Health, National Taiwan University, No. 1 Jen-Ai Road, Sec.1, Rm. 1550, Taipei 100, Taiwan.
E-mail: mingwhei@ha.mc.ntu.edu.tw

⁷These authors contributed equally to this work.

Received 30 August 2005; revised 15 November 2005; accepted 24 November 2005; published online 16 January 2006

presumed tumor-suppressor gene on 4q is currently an active area of research.

To determine whether chromosome 4q harbors a susceptibility gene for familial HCC, we have performed linkage and family-based association analyses. Study families were from Taiwan, where hepatitis B and HCC are hyperendemic.

Results

Linkage analyses

Of the 53 multiplex HCC families included in this study, one was a mother–father–child trio (in which the child and mother were affected with HCC) with no linkage information. Thus, we performed linkage analyses on 52 multiplex families. We genotyped 77 microsatellite polymorphisms distributed throughout the entire chromosome 4q region at an average spacing of 1.81 cM. Two-point linkage analyses revealed suggestive evidence for linkage to the region 4q22.3–28.1. Four markers within this region had heterogeneity LOD scores (HLODs) ≥ 2 . Marker D4S3240, on 4q25, yielded a maximum HLOD score of 2.55 at $\theta=0$, with the proportion (α) of families linked estimated at 55%. The peak two-point nonparametric linkage (NPL) Z score (1.94) occurred at D4S2989, which is 3.49 cM from D4S3240 (where the second-highest NPL score occurred, 1.66, $P=0.02156$) (Figure 1).

We further conducted multipoint analyses with 16 markers in the region 4q22.3–28.1. The HLOD score of marker D4S3240 increased to 3.12 ($\alpha=53\%$) from the multipoint analysis. Similarly, the NPL Z score also increased to 1.98 (pointwise $P=0.0080$) for D4S3240 (Figure 2). This NPL score corresponds to a region-wide empirical P -value of 0.021, based on simulations of 10 000 replicates each consisting of 16 markers with the same marker information in our study. We also evaluated the degree of allele sharing because of identity by descent (IBD) among the affected sibling pairs. The

multipoint estimate of the mean proportion of alleles sharing IBD was 0.60 for marker D4S3240, with a calculated standard error of 0.0524.

In all, 25 families (48.1%) had NPL Z scores >0 at markers within the 4q25 region and 27 families (51.9%) had NPL scores ≤ 0 . We then sought to characterize families showing linkage to 4q25 (families with positive NPL scores). The number of affected members, the mean within-family age at HCC diagnosis, and the occurrence of non-HCC malignancy could not distinguish families showing linkage from the nonlinked families. However, a parent with HCC was found in 18 of 27 (66.7%) nonlinked families but in only five of 25 (20.0%) families showing linkage ($P=0.0003$); that is, the families with positive NPL scores tended to exhibit horizontal transmission, a major characteristic of recessive traits.

As segregation analyses of familial HCC also supported an autosomal recessive mode of inheritance for HCC-susceptibility genes (Shen *et al.*, 1991; Cai *et al.*, 2003), we divided families into two subsets based on consistency with a recessive mode of inheritance, using the apparent presence or absence of an affected

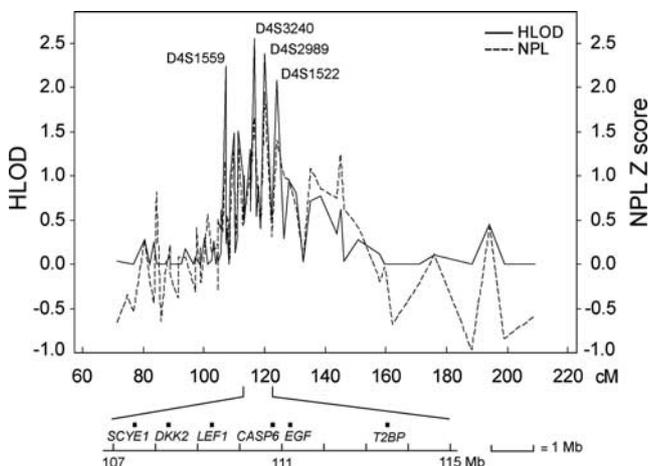


Figure 1 Two-point HLOD and NPL Z scores by distance from 4pter.

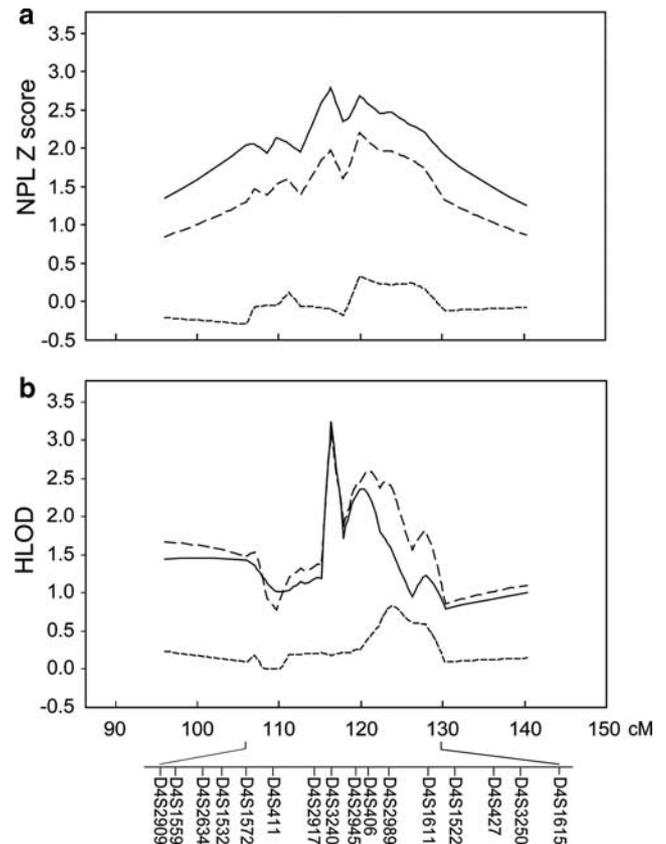


Figure 2 Multipoint linkage analysis of 4q22.3–28.1 markers. Dashed line = plot of multipoint scores in all 52 multiplex families informative for linkage analysis; solid line = plot of multipoint scores in 28 multiplex families lacking affected parents; dotted line = plot of multipoint scores in 24 multiplex families with affected parents. (a) Multipoint NPL Z scores. (b) Multipoint HLOD scores calculated under the assumption of an autosomal recessive model.

Table 1 Linkage results for familial HCC and 4q22.3–28.1 markers in 28 multiplex families lacking affected parents

Marker	Cytogenetic location	Heterozygosity	Distance in cM from 4pter	Two-point		Multipoint parametric linkage analysis		Multipoint nonparametric linkage analysis		
				HLOD	(α^a , θ^b)	HLOD	(α)	Z	Pointwise P	Region-wide empirical P
D4S2909	4q22.3	0.44	106.07	0.94	(0.90, 0.06)	1.43	(0.46)	2.05	0.0200	0.0570
D4S1559	4q22.3	0.46	107.05	1.71	(1.00, 0.06)	1.36	(0.43)	2.06	0.0193	0.0564
D4S2634	4q23	0.82	108.59	0.92	(0.40, 0.00)	1.13	(0.38)	1.94	0.0278	0.0696
D4S1532	4q23	0.64	109.76	1.83	(0.70, 0.00)	1.02	(0.36)	2.14	0.0176	0.0475
D4S1572	4q24	0.81	111.23	1.12	(1.00, 0.15)	1.04	(0.37)	2.07	0.0193	0.0554
D4S411	4q24	0.72	112.71	0.91	(0.45, 0.02)	1.15	(0.38)	1.95	0.0240	0.0684
D4S2917	4q25	0.39	115.22	0.23	(0.40, 0.00)	1.19	(0.41)	2.59	0.0046	0.0141
D4S3240	4q25	0.74	116.41	3.21	(0.70, 0.00)	3.25	(0.63)	2.79	0.0028	0.0080
D4S2945	4q25	0.82	117.91	1.10	(1.00, 0.15)	1.71	(0.46)	2.35	0.0095	0.0302
D4S406	4q25	0.70	118.61	0.86	(0.45, 0.00)	2.07	(0.53)	2.40	0.0088	0.0267
D4S2989	4q25	0.90	119.90	2.72	(0.65, 0.00)	2.37	(0.56)	2.69	0.0038	0.0102
D4S1611	4q26	0.56	122.31	0.69	(0.40, 0.00)	1.81	(0.49)	2.46	0.0073	0.0223
D4S1522	4q26	0.65	123.89	1.55	(0.65, 0.00)	1.53	(0.46)	2.47	0.0073	0.0215
D4S427	4q27	0.78	126.30	0.11	(1.00, 0.28)	0.96	(0.34)	2.30	0.0099	0.0336
D4S3250	4q27	0.78	127.90	0.44	(0.60, 0.13)	1.24	(0.41)	2.21	0.0147	0.0404
D4S1615	4q28.1	0.57	130.31	1.08	(1.00, 0.12)	0.79	(0.32)	1.91	0.0285	0.0743

^aProportion of families estimated to be linked. ^bMaximum likelihood estimate of the recombination fraction.

parent as a surrogate stratification criterion. Following this stratification, we found stronger evidence of linkage to the 4q25 region among 28 families lacking affected parents. Two-point analyses yielded a maximum HLOD score of 3.21 for D4S3240 in the 28 families. Multipoint analyses with 16 markers in the region 4q22.3–28.1 gave a maximum HLOD score of 3.25 ($\alpha=63\%$) and a maximum NPL Z score of 2.79 (pointwise $P=0.0028$) at the same location (Table 1 and Figure 2). Simulation to correct for multiple testing indicated that the NPL score corresponded to a region-wide empirical P -value of 0.008. In contrast, the 24 families with affected parents exhibited no significant evidence of disease linkage to 4q25 (Figure 2).

For the 33 sibling pairs from the 28 families lacking affected parents, the multipoint estimate of the mean proportion of alleles sharing IBD was 0.64 for marker D4S3240, with a calculated standard error of 0.0579. Test of mean allele sharing showed evidence of linkage demonstrated by significantly increased allele sharing ($P=0.0243$).

Family-based association analyses

Having found evidence of linkage, we examined each allele of each marker for evidence of association using the pedigree disequilibrium test (PDT) in all 191 families (53 multiplex plus 138 singleton families). When no correction was made for multiple testing, seven of the 77 microsatellite markers have alleles with significant PDT results. Four of the seven markers identified cluster in the region 4q22.1–27 (Figure 3a). Therefore, 29 single-nucleotide polymorphisms (SNPs) spanning ~1.0 Mb within the 1-HLOD drop interval from linkage study were further genotyped for multiplex families. SNP rs221330 achieved significance levels of $P<0.05$

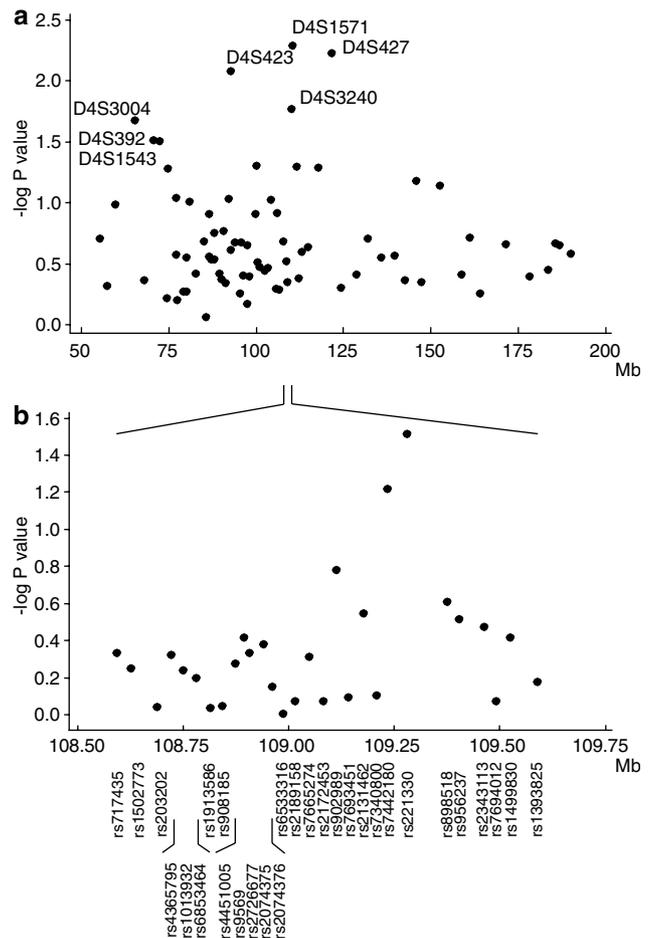


Figure 3 P -values for PDT analysis of chromosome 4q markers. (a) Single-marker allelic associations for 77 microsatellite markers spanning chromosome 4q. The smallest P -value for common alleles (frequency >0.10) at each marker locus was reported. (b) Single-marker allelic associations for 29 SNPs around the linkage peak.

Table 2 Summary of haplotypes with significant PDT results

	Frequency in all subjects from multiplex families ^a	Frequency among trios (n = 36)		Frequency in discordant siblings		P-value	
		Parental trans- missions ^a	Parental non- transmissions ^a	Affected (n = 52)	Unaffected (n = 86)	Individual haplotype	Global ^b
Haplotypes for rs7442180 and rs221330							
A-A	0.249	0.361	0.222	0.308	0.250	0.0750	0.023
A-G	0.077	0.069	0.042	0.077	0.064	0.8738	
C-A	0.066	0.097	0.069	0.096	0.058	0.1255	
C-G	0.607	0.472	0.667	0.519	0.628	0.0074	
Haplotypes for rs221330 and rs898518							
A-A	0.228	0.333	0.183	0.298	0.244	0.0083	0.0517
A-C	0.089	0.100	0.083	0.115	0.076	0.8002	
G-A	0.124	0.150	0.150	0.058	0.087	0.5290	
G-C	0.559	0.417	0.583	0.529	0.593	0.0892	

^aOwing to rounding, percentages do not always total 100. ^bP-values of significance were computed using the χ^2 distribution with three df.

(unadjusted for multiple testing), whereas its neighboring SNP rs7442180 showed marginally significant association (Figure 3b). Next, we analysed two-marker haplotypes for rs221330 and its flanking 5 SNPs. All estimates of pairwise D' between adjacent markers of the six SNPs were >0.7 . A common haplotype (at SNPs rs7442180 and rs221330; frequency 60.7% in this sample) positioned ~ 873 kb away from marker D4S3240 was associated with HCC, with $P=0.0074$ (Table 2).

Discussion

The etiology of HCC appears to be complex and multifactorial (reviewed in Chen and Chen, 2002). Mapping susceptibility loci for HCC is difficult due to the presence of phenocopies, genetic heterogeneity, age-dependent penetrance, and a variety of heterogeneous environmental risk factors. The high case-fatality rate of HCC further makes the study of affected families particularly challenging, because it is difficult to collect DNA samples from an affected family member. In spite of these difficulties, our results have provided evidence for a gene conferring susceptibility to HCC on chromosome 4q.

In this study, both parametric and NPL analyses indicate that a susceptibility locus on 4q22.3–28.1 may account for a significant subset of familial HCC. The 1-HLOD drop interval from the multipoint analysis with 16 markers within the region 4q22.3–28.1 among the 28 families lacking affected parents extends from D4S2917 (115.22 cM) to D4S2989 (119.90 cM) in the region 4q25. Several studies in HCC have revealed high frequency of allelic loss at 4q25 and/or its neighboring regions (Yeh *et al.*, 1996; Piao *et al.*, 1998; reviewed in Buendia, 2000; Okabe *et al.*, 2000; Bluteau *et al.*, 2002). This region was also identified as a commonly deleted region in other virus-related carcinomas including cervix and the head

and neck (Mitra *et al.*, 1994; Pershouse *et al.*, 1997; Wang *et al.*, 1999).

To provide further support for the findings from linkage analyses, PDT was then performed. Several microsatellite markers within or closely to the 1-HLOD drop interval yielded significant PDT results, including marker D4S3240, which is located at the center of the linkage peak. Prompted by this observation, we sought additional confirmation of our results by examining 29 SNPs in the vicinity of marker D4S3240. A common haplotype consisting of rs7442180 and rs221330, which positioned ~ 873 kb away from marker D4S3240, was associated with HCC, with $P=0.0074$.

Although the PDT results nicely mirror the findings from linkage study, the reported P -values should be viewed with caution because we have presented raw P -values uncorrected for multiple testing. Several procedures have been suggested for correcting multiple testing, but there is as yet little consensus as to an 'ideal statistical framework' for reporting P -values in analysis of SNP data because the correlation structure between neighboring SNPs complicates the problem (Nyholt, 2004; Wacholder *et al.*, 2004; Roeder *et al.*, 2005).

The haplotype (at SNPs rs7442180 and rs221330) showing the strongest association with HCC from linkage disequilibrium mapping by use of the PDT is located ~ 47 kb from the *LEF1* gene. This gene regulates the Wnt/ β -catenin signaling pathway, which has been implicated in hepatocarcinogenesis (reviewed in Buendia, 2000). Using the National Center for Biotechnology information Mapviewer, we identified other five potential candidate genes residing in or proximal to the 4.68 cM linkage interval on 4q25. Of these, *DKK2* also has a role in the Wnt/ β -catenin signaling pathway; *CASP6* is involved in cellular apoptosis; *EGF* encodes the epidermal growth factor and *SCYE1* encodes a cytokine; *T2BP* is associated with the oligomerization and polyubiquitination of tumor necrosis factor receptor-associated factor 6, which is important for the regulation of inflammatory process (Ea *et al.*, 2004).

To our knowledge, this is the first report to describe a genetic basis for familial HCC. Since more than 90% of the affected individuals were HBsAg carriers in this study, chronic HBV infection is also an important risk factor, even in families with linkage to a high-penetrance locus. Identification of the relevant gene predisposing to HCC on chromosome 4q will provide further insight into both the problems of HCC etiology and viral oncogenesis in general. Although we have also performed linkage disequilibrium analyses for fine mapping by use of a dense set of SNPs spanning ~1.0 Mb and the PDT results are intriguing, the 29 SNPs genotyped in the 52 multiplex families are unlikely to provide sufficient coverage of the linked region to conclusively exclude genes by association mapping. Further fine-mapping studies with additional SNPs in a more extended region are needed.

Materials and methods

Families and genotyping

Since 1997, a family-based study designed to search for environmental or genetic factors that may contribute to the familial clustering of HCC has recruited patients with HCC from three major medical centers in Taiwan (Yu *et al.*, 2000). Diagnosis of HCC was confirmed by liver biopsy or the combination of increased α -fetoprotein (≥ 400 ng/ml) plus typical features on angiography, sonography, or computed tomography. In 2001, we had 2448 patients with questionnaire interview and DNA samples. For this study, a total of 902 individuals (257 affected individuals and 645 unaffected individuals) from 191 families, identified through family-history interview among these patients, were available. This sample included 51 multiplex nuclear families with at least two affected members, two extended families containing affected paternal uncle-nephew pairs and/or affected parent-offspring and sib pairs, as well as 138 singleton families consisted of 67 complete trios (patients and both parents) and 71 case-parent pairs with unaffected siblings. Totally, 34 sibships containing at least two affected siblings were used for sib-pair analyses. In all, 30 sibships had two affected siblings and four sibships had three. The affected siblings defined 42 affected sibling pairs when all possible distinct-affected sibling pairs were used from the triplet sibships. The mean age at onset in affected individuals was 44.9 ± 12.3 (standard deviation) years. In total, 91% of the affected individuals and 36.1% of the unaffected individuals tested for HBsAg were seropositive, whereas only 9.5% of the affected individuals and 4.7% of the unaffected individuals were seropositive for antibodies against hepatitis C virus. All participants in this study gave informed consent.

All 191 families were genotyped for 73 microsatellite markers with fluorescently labeled primers and DNA sequencers (model 377 or 3100; Applied Biosystems). After preliminary analysis suggested the possible location of the HCC-susceptibility locus, additional four microsatellite markers and 29 SNPs spanning ~1.0 Mb around the linkage peak were genotyped for the 52 multiplex families used for linkage analyses. Genotyping of SNPs was provided by the National Genotyping Center at Academia Sinica (<http://ngc.sinica.edu.tw>), where the genotypes were determined using a MassARRAY (SEQUENOM, Inc., San Diego, CA, USA). All genotypes were checked for inconsistent Mendelian

inheritance using the PedCheck software (O'Connell and Weeks, 1998). Inconsistencies were eliminated by either retyping or by removal from the analysis.

Data analyses

Genetic map distances were taken from the Rutgers Linkage-Physical Maps (Kong *et al.*, 2004). Marker allele frequencies were estimated from pedigree founders. Autosomal recessive inheritance was assumed for parametric linkage analysis, with a disease-allele frequency of 0.13 and 12 age-sex specific liability classes. For males, the age-dependent penetrances were 0.0012 (0–29 years), 0.0047 (30–39 years), 0.0144 (40–49 years), 0.9999 (50–59 years), 0.9999 (60–69 years), and 0.9999 (≥ 70 years). We assumed no phenocopies before age 50 years and the phenocopy rates for the age group 50–59, 60–69 and ≥ 70 years were 0.064, 0.064, and 0.1, respectively. For female subjects, the penetrances assigned to each age group were 0.0004 (0–29 years), 0.0016 (30–39 years), 0.0048 (40–49 years), 0.01 (50–59 years), 0.9999 (60–69 years), and 0.9999 (≥ 70 years). We assumed no phenocopies before age 60 years and the phenocopy rates were 0.064 and 0.08, respectively, for the age group 60–69 and ≥ 70 years.

Since locus heterogeneity confounds the discovery of susceptibility genes, we calculated LOD scores with the assumption of heterogeneity (HLOD) (Ott, 1986). Two-point HLOD scores were computed using the TABLE 1.8 software (<http://www.mds.qmw.ac.uk/statgen/dcurtis/software.html>). Multipoint HLOD scores and NPL Z scores were computed with GENEHUNTER version 2.1 (Kruglyak *et al.*, 1996). To examine the false-positive rate, empirical *P*-values were calculated for the NPL scores via simulation. The program MERLIN version 1.0.0 (Abecasis *et al.*, 2002) was used to generate 10 000 replicates of families identical to those in our sample. Markers with similar allele sizes and frequencies were also generated under the assumption of no linkage. Linkage analyses were then performed on these unlinked replicates, and region-wide empirical *P*-values were calculated as the proportion of replicates showing an equal or more extreme NPL score at any point within the studied chromosomal region. In affected sib-pair analyses, the mean proportion of alleles sharing IBD was analysed by use of the SIBPAL program in the program package SAGE version 4.5. We used the affected-sib-pair mean test (Blackwelder and Elston, 1985) to test for linkage.

The program GENEHUNTER was also used to construct haplotype for family data. Linkage disequilibrium between pairs of markers was estimated using the PowerMarker software program (Liu and Muse, 2004). We performed PDT (Martin *et al.*, 2000) for family-based association analyses. The average PDT statistic, for which the contribution of large families to the end results does not exceed that of small families, was used to examine associations between markers and HCC. To protect against misleading results due to infrequent alleles or haplotypes at a moderate number of study families, we aggregated all alleles or haplotypes with frequencies ≤ 0.10 before PDT analysis.

Acknowledgements

This work was supported by Grants NSC 92-2320-B-002-031 (to M-W Yu) and NSC 92-3112-B002-007 (to P-J Chen) from the National Science Council and DOH92-TD-1054 (National Research Program for Genomic Medicine) (to M-W Yu) from the Department of Health, Executive Yuan, Taiwan.

References

- Abecasis GR, Cherny SS, Cookson WO, Cardon LR. (2002). *Nat Genet* **30**: 97–101.
- Beasley RP. (1982). *Hepatology* **2**(Suppl): 21–26.
- Befeler AS, Di Bisceglie AM. (2002). *Gastroenterology* **122**: 1609–1619.
- Blackwelder WC, Elston RC. (1985). *Genet Epidemiol* **2**: 85–97.
- Bluteau O, Beaudoin JC, Pasturaud P, Belghiti J, Franco D, Bioulac-Sage P *et al* (2002). *Oncogene* **21**: 1225–1232.
- Buendia MA. (2000). *Semin Cancer Biol* **10**: 185–200.
- Cai RL, Meng W, Lu HY, Lin WY, Jiang F, Shen FM. (2003). *World J Gastroenterol* **9**: 2428–2432.
- Chen CJ, Chen DS. (2002). *Hepatology* **36**: 1046–1049.
- Ea CK, Sun L, Inoue J, Chen ZJ. (2004). *Proc Natl Acad Sci USA* **101**: 15318–15323.
- Kong X, Murphy K, Raj T, He C, White PS, Matisse TC. (2004). *Am J Hum Genet* **75**: 1143–1148.
- Kruglyak L, Daly MJ, Reeve-Daly MP, Lander ES. (1996). *Am J Hum Genet* **58**: 1347–1363.
- Laurent-Puig P, Legoix P, Bluteau O, Belghiti J, Franco D, Binot F *et al* (2001). *Gastroenterology* **120**: 1763–1773.
- Liu K, Muse S. (2004). <http://www.powermarker.net>.
- Martin ER, Monks SA, Warren LL, Kaplan N. (2000). *Am J Hum Genet* **67**: 146–154.
- Mitra AB, Murty VV, Li RG, Pratap M, Luthra UK, Chaganti RS. (1994). *Cancer Res* **54**: 4481–4487.
- Nyholt DR. (2004). *Am J Hum Genet* **74**: 765–769.
- O'Connell JR, Weeks DE. (1998). *Am J Hum Genet* **63**: 259–266.
- Okabe H, Ikai I, Matsuo K, Satoh S, Momoi H, Kamikawa T *et al* (2000). *Hepatology* **31**: 1073–1079.
- Ott J. (1986). *Genet Epidemiol* **1**(Suppl): 251–257.
- Pershouse MA, El-Naggar AK, Hurr K, Lin H, Yung WA, Steck PA. (1997). *Oncogene* **14**: 369–373.
- Piao Z, Park C, Park JH, Kim H. (1998). *Int J Cancer* **79**: 356–360.
- Roeder K, Bacanu SA, Sonpar V, Zhang X, Devlin B. (2005). *Genet Epidemiol* **28**: 207–219.
- Shen FM, Lee MK, Gong HM, Cai XQ, King MC. (1991). *Am J Hum Genet* **49**: 88–93.
- Thorgeirsson SS, Grisham JW. (2002). *Nat Genet* **31**: 339–346.
- Wacholder S, Chanock S, Garcia-Closas M, El ghormli L, Rothman N. (2004). *J Natl Cancer Inst* **96**: 434–442.
- Wands JR. (2004). *N Engl J Med* **351**: 1567–1570.
- Wang XL, Uzawa K, Imai FL, Tanzawa H. (1999). *Oncogene* **18**: 823–825.
- Yeh SH, Chen PJ, Lai MY, Chen DS. (1996). *Gastroenterology* **110**: 184–192.
- Yeh SH, Chen PJ, Shau WY, Chen YW, Lee PH, Chen JT *et al* (2001). *Gastroenterology* **121**: 699–709.
- Yu MW, Chang HC, Liaw YF, Lin SM, Lee SD, Liu CJ *et al* (2000). *J Natl Cancer Inst* **92**: 1159–1164.
- Yu MW, Chen CJ. (1994). *Crit Rev Oncol Hematol* **17**: 71–91.