EUROPEAN JOURNAL OF CANCER 45 (2009) 1239–1247

available at www.sciencedirect.com

journal homepage: www.ejconline.com

Common genetic variations of the cytochrome P450 1A1 gene and risk of hepatocellular carcinoma in a Chinese population

Rui Li[a], *Yin Yao Shugart*[b,c], *Weiping Zhou*[d], *Yu An*[a], *Yuan Yang*[d,e], *Yun Zhou*[e], *Beibei Zhang*[e], *Daru Lu*[a], *Hongyang Wang*[e,*], *Ji Qian*[a,*], *Li Jin*[a]

[a]*State Key Laboratory of Genetic Engineering and MOE Key Laboratory of Contemporary Anthropology, School of Life Sciences, Fudan University, 220 Handan Rd., Shanghai 200433, PR China*
[b]*Department of Epidemiology, Johns Hopkins Bloomberg School of Public Health, Baltimore, MD, United States*
[c]*Department of Biostatistics, Johns Hopkins Bloomberg School of Public Health, Baltimore, MD, United States*
[d]*The Third Department of Hepatic Surgery, Eastern Hepatobiliary Surgery Hospital, Shanghai, PR China*
[e]*International Cooperation Laboratory on Signal Transduction, Eastern Hepatobiliary Surgery Institute, Shanghai, PR China*

ARTICLE INFO

Article history:
Received 27 August 2008
Received in revised form 29 October 2008
Accepted 7 November 2008
Available online 26 December 2008

Keywords:
CYP1A1
Genetic polymorphism
Haplotype
Hepatocellular carcinoma
Hepatitis

ABSTRACT

Cytochrome P450 1A1 is a major enzyme in the bioactivation of exogenous procarcinogens of hepatocellular carcinoma (HCC). However, the contribution of common genetic variants in *CYP1A1* to the HCC risk in Chinese populations has not been thoroughly investigated. In this study, we examined the association between HCC and four selected tagging single nucleotide polymorphisms (SNPs) of *CYP1A1*, and the risk of *CYP1A1* haplotypes/diplotypes in 1006 pathologically confirmed HCC patients and 1015 cancer-free controls, from a Han Chinese population. Haplotypes/diplotypes were constructed from observed genotypes using the Haplo.Stats program. Relative risk was estimated by using multivariable logistic regression method. To summarise, we detected an increased HCC risk in rs4646421 variant carriers (OR 1.30, 95% CI 1.05–1.61) and rs2198843 variant carriers (OR 1.33, 95% CI 1.05-1.69), and a reduced risk of HCC (OR 0.70. 95% CI 0.52–0.94) associated with homozygote carriers of rs4886605 variant. These association signals were also observed in non-smokers with rs4646421 (OR 1.56, 95% CI 1.16–2.08) and rs4886605 (OR 0.61, 95% CI 0.40–0.91). Compared to the most common *CYP1A1* haplotype CCAG, the haplotype TTGC conferred an increased risk of HCC (OR 1.26, 95% CI 1.04–1.52). Similarly, the TTGC/TTGC diplotype conferred an increased risk of HCC compared with diplotype CCAG/CCAG (OR 2.06, 95% CI 1.23–3.45, $P = 0.006$). Interestingly, the diplotype TTAC/CCAG also conferred an increased risk of HCC (OR 1.76, 95% CI 1.22–2.54, $P = 0.003$). Our results suggested that common genetic variants in *CYP1A1* may modulate the risk of developing HCC in the study population, particularly in non-smokers. However, our findings need to be validated in at least one independent study of Han Chinese population.

* *Corresponding authors:* Tel./fax: +86 21 65642426 (J. Qian).
E-mail addresses: hywangk@vip.sina.com (H. Wang), jiqian@fudan.edu.cn (J. Qian).
0959-8049/$ - see front matter
doi:10.1016/j.ejca.2008.11.007

1. Introduction

Hepatocellular carcinoma (HCC) is the sixth most prevalent malignant tumour worldwide and ranks the third as a cancer killer, causing more than half million deaths annually.[1,2] Although HCC has been generally considered more endemic in China and sub-Saharan Africa, recent epidemiology studies revealed that low survival rate of individuals with liver cancer has made this malignancy the second most lethal cancer in America.[3,4] The incidence of liver cancer is lower in Europe (consistently below 10/100,000) than areas of high incidence. Age-adjusted incidence rates of liver cancer in men were estimated to be 9.8/100,000 in Southern Europe, 5.8/100,000 in Eastern and Western Europe and 2.6/100,000 in Northern Europe.[1,2] The distribution among women follows a similar pattern. Chronic hepatitis B and C viral infections have been well characterised to play a major role in the HCC aetiology. Given that not all the individuals infected with HBV/HCV develop HCC, other risk factors including environment and genes may be involved in the multistage process of this complex disease.[5]

Many polycyclic aromatic hydrocarbons (PAHs) exhibit carcinogenic effects through interaction of their reactive metabolites and DNA, thereby causing DNA damage. PAHs may also induce hepatocellular carcinogenesis, because human liver participates in the metabolism of these exogenous chemical compounds derived from diet, smoking and drugs. It has been shown that, compared to the tissues obtained from non-HCC controls, higher levels of PAH-DNA adducts were found in adjacent non-tumour tissues of HCC cases.[6] For example, Wu and colleagues reported the association between PAH exposure and risk of HCC by comparing the levels of PAH-albumin adducts in blood samples.[7] Moreover, PAH exposure modulated the expression levels of several genes in the liver.[8] The genes involved in the process of hepatocarcinogenesis are highly correlated to the levels of PAH-DNA adducts.[8]

In this regard, the enzymes converting PAHs into their biologically reactive forms may play a critical role in the determination of an individual's susceptibility to HCC. The CYP1A1 (*cytochrome P450, family1, subfamily A, polypeptide1*) enzyme is essential for the catalysis of the first step in PAHs metabolism and therefore has been considered a primary candidate in HCC susceptibility. The *CYP1A1* (MIM 108330) gene is composed of seven exons and encompasses 5.8 kb on 15q22-24.[9] Two genetic variants of *CYP1A1* (rs4646903 and rs1048943) have been reported to be associated with an increased risk of HCC in smokers,[10] but the association was not replicated in subsequent studies.[11] These inconsistent findings justify the need for additional studies of larger sample sizes to further evaluate the role of the *CYP1A1* variants in HCC development. Although the variants investigated in the previous studies were indicated to have putative effects on the enzyme activity, their involvement may not be sufficient to account for a significant proportion of the differences in the levels of the CYP1A1 activity among individuals.[12] In light of these findings, we utilised the tag single nucleotide polymorphisms (tagSNPs) dependent on the linkage disequilibrium (LD) pattern of the gene to provide a better understanding of the relationship between genetic variation in *CYP1A1* and the risk of HCC. In the present study, therefore, we conducted a large case-control study to extend our investigation of genetic risk factors for HCC in a Han Chinese population, and evaluated the effect of common genetic polymorphisms (both within the *CYP1A1* and its surrounding gene region) on the risk of developing HCC.

2. Materials and methods

2.1. Study population

This study included 1006 cases and 1015 controls that were unrelated ethnic Han Chinese from Shanghai and its surrounding provinces including Zhejiang, Jiangsu and Anhui in the eastern China. Patients were consecutively recruited from Eastern Hepatobiliary Surgery Hospital between January 2003 and December 2005. Ninety percent of the recruited HCC patients were incident cases. The remaining 106 (10%) cases had been diagnosed to have HCC before 2003 and revisited the hospital during the period of the study. The diagnosis of HCC was confirmed by a pathological examination or α-fetoprotein elevation (>400 ng/ml) combined with positive imaging (Magnetic resonance imaging, MRI and/or computerised tomography, CT).[13] The cases that were 25 years old or younger were excluded because these HCC cases may have developed from hepatitis B virus infection through mother-infant transmission. In order to reduce the confounding effect of HBV infection in research of genetic susceptibility to HCC, controls were randomly selected from the individuals who attended hepatitis examination in the hospital from the same metropolitan region during the period of case collection. The selection criterion for control subjects included no personal history of any kind of cancer at the time of ascertainment, and these subjects were frequency-matched to HCC cases by age (±5 years). HBV infection was considered if the seropositivity of hepatitis B surface antigen (HBsAg) was detected. At the end of the recruitment, 52.2% individuals were HBsAg positive in the controls. For all the cases and controls, we detected the serum HBV markers, antibodies to hepatitis C virus, alanine transaminase and aspartate transaminase.

Once cases and controls subjects gave their written informed consent to participate in the study, they were interviewed by using a standard questionnaire including demographic characteristics and the history of smoking, drinking and infectious disease. At the end of the 30-minute interview, a sample of approximately 5ml venous blood was collected from each subject. Demographical data (e.g. age, sex and ethnicity) and lifetime consumption of tobacco smoking and alcohol drinking were also collected. The subjects who smoked more than one cigarette per day for more than one year were classified as smokers. Others were defined as non-smokers. The average number of cigarettes smoked per day and the total number of years of smoking were used to calculate cumulative smoking dose as 'pack-years' ([cigarettes per day/20] × years smoked). Subjects were considered as alcohol drinkers, if they drank at least once per week. They were asked about the frequency of their consumption of wine, beer and spirits, and the consumption was quantified according to the average ethanol content of wine (12% in volume),

beer (5%) and spirits (40%). This study protocol was approved by the Ethics Committee for Human Subject Research at Fudan University.

2.2. SNP selection and laboratory methods

A few previous studies examined the relationship between HCC risk and some putatively functional polymorphisms in *CYP1A1*. One goal of this study was to investigate the contribution of common SNPs of *CYP1A1* in the Han Chinese population to the development of HCC. To balance the efficiency of association study and the genotyping cost, a linkage disequilibrium-based approach was implemented to select a minimum number of tagSNPs to cover genetic variation in the gene.[14] Using the online dataset (release 19/Phase II Oct 05) provided by the International HapMap Project (http://www.hapmap.org), we identified a tagSNP set having pairwise $r^2 > 0.8$ with other unselected SNPs. Because few SNPs were discovered in the gene region of 5.8 kb, we defined a broader region including 10 kb upstream and downstream sequence and identified three tagSNPs rs4886605, rs4646421 and rs2198843 with a minor allele frequency (MAF) greater than 0.05 for *CYP1A1* (15q22-24, NC_000015.8). Each selected SNP was located in either the non-coding, or the 5′ or 3′ untranslated regions. We also included a non-synonymous SNP rs1048943 located in exon 7 which was reported to be associated with the risk of lung cancer[15] and oral cancer.[16]

Whole blood samples collected from case and control subjects were processed for the extraction of genomic DNA using the conventional phenol/chloroform extraction method. Isolated DNA was diluted to 5 ng/ul and distributed into 384-well plates. TaqMan SNP Genotyping Assay (Applied Biosystems, Foster City, CA) was performed to detect genotypes for the selected SNPs following the manufacturers's instructions. The unlabelled PCR primers and TaqMan MGB probes (FAM and VIC dye-labelled) were directly supplied by the Assay-by-Design service from Applied Biosystems. The details of sequences and reaction conditions are available upon request (https://products.appliedbiosystems.com/ab/en/US/adirect/ab). The genotypes were determined by scanning plates on the ABI PRISM 7900HT Sequence Detection System with endpoint reading. The call rate across all the assays was greater than 96%. Four duplicate samples and four no-template controls (NTCs) were also included in each plate for quality control, and the genotyping achieved 100% reproducibility.

2.3. Statistical methods

Differences in demographic characteristics (age and sex), smoking status, pack-years of smoking, drinking status, alcohol intake per day, HBsAg status and genotype frequency distribution were compared between cases and controls using χ^2 tests for categorical variables or Student's *t* tests for continuous measures. Departure from Hardy-Weinberg equilibrium (HWE) was assessed by Pearson's χ^2 test among controls for each polymorphism in *CYP1A1*. Unconditional logistic regression was conducted to estimate the relative risk of each SNP adjusted for age, sex, pack-years, alcohol intake per day and HBsAg status. Odds ratios (ORs) and their 95% confidence intervals (CIs) were calculated. The most common genotype in controls was considered the referent. To evaluate the effect of the genotype containing SNP variant, we ran analyses under both codominant and dominant models. Further, we stratified the data by smoking status, drinking status and infection of hepatitis B virus. All tests were two-sided with a significance level of *P* value < 0.05. To account for the issue of multiple testing of SNPs, we used SNPSpD[17] to correct the significance threshold taking into account LD between SNPs. Unlike popular Bonferroni correction which may result in a power reduction when LD exists between SNPs, SNPSpD generates the experiment-wide significance threshold required to keep the Type I error rate at <5%, on the basis of the spectral decomposition (SpD) of matrics of pairwise LD between SNPs, which is a measurement of SNP correlation. Moreover, this method is more simplistic than the permutation tests that are typically computationally intensive. The haplotype analysis was conducted using modules implemented in Haplo.Stats program[18] (http://mayoresearch.mayo.edu/mayo/research/biostat/splusfunctions.cfm). The expectation-maximisation algorithm in haplo.em was used to estimate population haplotype frequencies. Common haplotypes with frequencies greater than 0.01 were compared between cases and controls and in stratification analysis of smoking status. The permutation tests were used to control multiple testing errors with 10,000 times simulation. Adjusted ORs and 95% CIs were also generated for each haplotype compared to the most common haplotype by haplo.glm algorithm. The most probable pair of haplotypes (diplotypes) was constructed by Haplo.Stats program and validated by PHASE 2.1.[19] Adjusted ORs and 95% CIs were calculated for individual diplotypes with frequencies greater than 0.02 compared to the diplotype constituted with the most common haplotype pair.

3. Results

3.1. Population characteristics

Table 1 provides demographic data and distribution of risk factors for both cases and controls. The controls were slightly younger than cases and had a higher proportion of females. The cases and controls had no statistical difference in smoking status, while there were more heavy smokers in cases (>16 pack-years) than in controls. Although there were a few more drinkers in controls than in cases, the difference in drinking status between cases and controls was not significant, with more light drinkers (⩽15 g/day) but less heavy drinkers (>15 g/day) in controls (29.4% and 13.7%, respectively) than in cases (20.3% and 20.3%, respectively). The positive rate for HBsAg was 87.5% in cases and 52.2% in controls.

3.2. Single SNP association

We obtained genotyping results of four selected SNPs of *CYP1A1* in cases and controls with success rate greater than 96%. The allele frequencies in our study were in line with those of Han Chinese in HapMap project (Table S1), and genotype frequencies were consistent with Hardy-Weinberg proportions in all the control subjects. By comparing the overall genotype frequency distribution, no statistically significant differences were tested. However, we identified significant

Table 1 – Distribution comparison of selected characteristics of HCC cases and controls.

Characteristics	Controls (n = 1015) no. (%)	Cases (n = 1006) no. (%)	P value
Age (mean ± SD), y	50.1 ± 13.1	51.4 ± 10.6	0.02
Age group			<0.001
⩽40	251 (24.7)	162 (16.1)	
40–⩽50	266 (26.2)	317 (31.5)	
50–⩽60	273 (26.9)	340 (33.8)	
>60	225 (22.2)	187 (18.6)	
Gender			<0.001
Male	817 (80.5)	880 (87.5)	
Female	198 (19.5)	126 (12.5)	
Smoking status			0.23
Never	538 (53.0)	561 (55.8)	
Ever	477 (47.0)	445 (44.2)	
Pack-years smoked[a,b]			
0	538 (53.2)	561 (61.9)	<0.001
⩽16	296 (29.3)	142 (15.7)	
>16	177 (17.5)	203 (22.4)	
Drinking status			0.28
Never	577 (56.9)	596 (59.4)	
Ever	437 (43.1)	408 (40.6)	
Alcohol intake (g/day)			
0	577 (56.9)	596 (59.4)	<0.001
⩽15	298 (29.4)	204 (20.3)	
>15	139 (13.7)	204 (20.3)	
HBsAg			<0.001
Negative	485 (47.8)	126 (12.5)	
Positive	530 (52.2)	880 (87.5)	

a Pack-year information is missing for 108 cases, so the percentage of non-smoker is 61.9%.
b Pack-year and alcohol intake are categorised on the bases of the mean number in controls.

associations for rs4646421 and rs2198843 after adjusting for age, sex, pack-years of smoking, amount of alcohol intake per day and HBsAg status, with both rs4646421 and rs2198843 variant carriers having an approximately 30% higher risk of HCC in comparison with their homozygote carriers of common alleles (95% CI for OR were 1.05–1.61 and 1.05–1.69, respectively, Table 2). Associated variant genotypes of rs4886605 had a reduced HCC risk (OR 0.70, 95% CI

Table 2 – Genotype frequencies of CYP1A1 among cases and controls, and risk of HCC.

SNP[a]	Genotype	Controls no. (%)	Cases no. (%)	P[b]	Logistic regression	
					OR[c] (95% CI[d])	P value
rs4886605	TT	293 (29.9)	324 (33.1)	0.16	1.00 (Reference)	
	TC	489 (49.9)	486 (49.6)		0.91 (0.72–1.15)	0.42
	CC	197 (20.1)	169 (17.3)		0.70 (0.52–0.94)	0.018
	TC + CC	686 (70.1)	655 (66.9)		0.88(0.71–1.09)	0.24
rs4646421	CC	357 (35.7)	312 (32.4)	0.29	1.00 (Reference)	
	TC	491 (49.1)	502 (52.1)		1.31 (1.05–1.64)	0.018
	TT	153 (15.3)	149 (15.5)		1.25 (0.92–1.72)	0.16
	TC + TT	644 (64.3)	651 (67.6)		1.30 (1.05–1.61)	0.017
rs1048943	AA	598 (59.8)	560 (57.7)	0.19	1.00 (Reference)	
	AG	357 (35.7)	349 (36.0)		1.04 (0.84–1.29)	0.71
	GG	45 (4.5)	61 (6.3)		1.54 (0.98–2.43)	0.06
	AG + GG	402 (40.2)	410 (42.3)		1.09 (0.89–1.34)	0.41
rs2198843	GG	259 (26.4)	220 (22.9)	0.18	1.00 (Reference)	
	GC	502 (51.1)	506 (52.6)		1.33 (1.03–1.70)	0.03
	CC	221 (22.5)	236 (24.5)		1.38 (1.03–1.86)	0.03
	GC + CC	723 (73.6)	742 (77.1)		1.33 (1.05–1.69)	0.017

a Chromosome positions are 72813041, 72803245, 72800038 and 72788283 for rs4886605, rs4646421, rs1048943 and rs2198843, respectively, from NCBI dbSNP database.
b Two-sided χ^2 test.
c Adjusted for age, gender, number of pack-years smoked, gram of alcohol intake per day and HBsAg status (positive or negative).
d OR, odds ratio; CI, confidence interval.

0.52–0.94) compared with TT genotypes. The association test on SNP rs1048943 achieved a borderline significant *P* value of 0.06 (GG versus AA, OR 1.54, 95% CI 0.98–2.43). After correcting for multiple comparisons by using SNPSpD (a type of Bonferroni correction method which takes the LD between SNPs into account), the associations observed in overall samples on SNP rs4646421, rs2198843 and rs4886605 remained significant (threshold significant *P* value was set at 0.019 by the program SNPSpD). Many kinds of PAHs are contained in cigarette smoking. To explore the putative effects of cigarette smoking on association of *CYP1A1* polymorphisms with HCC risk, we further examined the association stratified by smoking status. We found that these relative risks were still evident for rs4646421 (OR 1.56, 95% CI 1.16–2.08 for TC + TT versus CC), rs2198843 (OR 1.41, 95% CI 1.02–1.95 for GC + CC versus GG) and rs4886605 (OR 0.61, 95% CI 0.40–0.91 for CC versus TT) in non-smokers, although the association on rs2198843 may not be significantly rendered by multiple testing adjustment (Table 3). On the other hand, rs1048943 mutant homozygotes appeared to be associated with an increased risk of HCC (OR = 2.44, 95% CI: 1.22–4.88) in smokers, while no significant associations were detected on other polymorphisms. Association analysis of *CYP1A1* genotypes and risk of HCC were also stratified on risk factors including drinking status and HBV infection. Similarly, we found that associations were significant for rs4886605 (OR 0.57, 95% CI 0.35–0.84 for CC versus TT) and rs4646421 (OR 1.41, 95% CI 1.06–1.88 for TC + TT versus CC) in non-drinkers after multiple testing correction, and there was a marginal association for rs1048943 in drinkers (Supplemental Table S2). On the other hand, we observed no significant heterogeneity of effects for rs4886605 and rs4646421 genotypes in HBsAg negative subjects and HBsAg positive subjects (data not shown).

3.3. *Haplotype and diplotype association analysis*

LD analysis further revealed that all selected SNPs of *CYP1A1* were in strong LD with pairwise $D' > 0.85$, and formed a haplotype block in our study population. Haplotypes were inferred by the expectation-maximisation algorithm implemented in haplo.em, and six major haplotypes (CCAG, TTAC, TCAC, CTAC, TCAG and TTGC) that had allele frequency > 0.01 were included in the analysis as presented in Table 4. The most common haplotype CCAG containing no risk allele at each SNP locus was significantly more frequent in controls (43.3%) than in cases (39.7%) ($P = 0.004$, $P_{sim} = 0.004$) showed

Table 3 – Association of CYP1A1 genotypes with HCC risk stratified by smoking status.

	Genotype	Non-smoker				Smoker			
		Controls no. (%)	Cases no. (%)	OR[a] (95% CI)	*P* value	Controls no. (%)	Cases no. (%)	OR[a] (95% CI)	*P* value
rs4886605	TT	150 (29.1)	166 (30.9)	1.00 (Reference)		143 (30.9)	158 (35.7)	1.00 (Reference)	
	TC	252 (48.8)	286 (53.3)	1.00 (0.73–1.38)	0.99	237 (51.2)	200 (45.2)	0.79 (0.56–1.13)	0.19
	CC	114 (22.1)	85 (15.8)	0.61 (0.40–0.91)	0.016	83 (17.9)	84 (19.0)	0.78 (0.49–1.22)	0.27
	TC + CC	366 (70.9)	370 (68.3)	0.87 (0.64–1.17)	0.35	320 (69.1)	285 (65.2)	0.87 (0.61–1.23)	0.42
rs4646421	CC	203 (38.2)	165 (31.0)	1.00 (Reference)		154 (32.8)	147 (34.1)	1.00 (Reference)	
	TC	244 (46.0)	292 (54.9)	1.64 (1.21–2.23)	0.001	247 (52.6)	210 (48.7)	0.98 (0.69–1.38)	0.89
	TT	84 (15.8)	75 (14.1)	1.19 (0.78–1.82)	0.42	69 (14.7)	74 (17.2)	1.49 (0.92–2.39)	0.10
	TC + TT	328 (61.8)	371 (69.2)	1.56 (1.16–2.08)	0.003	316 (67.2)	284 (65.9)	1.09 (0.77–1.54)	0.63
rs1048943	AA	320 (60.2)	313 (58.4)	1.00 (Reference)		278 (59.4)	247 (56.9)	1.00 (Reference)	
	AG	185 (34.8)	192 (35.8)	1.05 (0.78–1.41)	0.63	172 (36.8)	157 (36.2)	1.08 (0.78–1.51)	0.64
	GG	27 (5.1)	31 (5.8)	1.30 (0.69–2.44)	0.40	18 (3.8)	30 (6.9)	2.44 (1.22–4.88)	0.01
	AG + GG	212 (39.8)	226 (41.8)	1.11 (0.84–1.47)	0.46	190 (40.6)	184 (42.9)	1.17 (0.84–1.63)	0.35
rs2198843	GG	140 (27.2)	117 (22.1)	1.00 (Reference)		119 (25.5)	103 (23.8)	1.00 (Reference)	
	GC	264 (51.3)	293 (55.3)	1.38 (0.98–1.93)	0.06	238 (51.0)	213 (49.3)	1.23 (0.84–1.80)	0.30
	CC	111 (21.6)	120 (22.6)	1.35 (0.90–2.03)	0.14	110 (23.6)	116 (26.9)	1.51 (0.97–2.36)	0.07
	GC + CC	375 (72.8)	416 (77.9)	1.41 (1.02–1.95)	0.04	348 (74.5)	326 (76.2)	1.29 (0.88–1.89)	0.19

a Adjusted for age, gender, number of pack-years smoked (optionally), gram of alcohol intake per day and HBsAg status (positive or negative).

Table 4 – Associations between haplotypes of CYP1A1 and risk of HCC.

Haplotype[a]	Control no. (%)	Case no. (%)	*P*	P_{sim}[b]	OR[c] (95% CI)	*P* value
CCAG	878 (43.3)	799 (39.7)	0.004	0.004	1.00 (reference)	
TTAC	313 (15.4)	299 (14.9)	0.59	0.59	1.20 (0.96–1.49)	0.12
TCAC	169 (8.3)	171 (8.5)	0.85	0.85	1.15 (0.89–1.50)	0.29
CTAC	31 (1.5)	40 (2.0)	0.15	0.15	1.76 (0.99–3.12)	0.05
TCAG	173 (8.5)	195 (9.7)	0.30	0.20	1.31 (0.99–1.71)	0.05
TTGC	454 (22.4)	491 (24.4)	0.15	0.15	1.26 (1.04–1.52)	0.02

a Order of polymorphism-rs4886605, rs4646421, rs1048943, rs2198843.
b Generated by 10,000 times permutation.
c Adjusted for age, gender, number of pack-years smoked, gram of alcohol intake per day and HBsAg status (positive or negative).

Table 5 – Associations between diplotypes of CYP1A1 and HCC.

Diplotype[a,b]	Control no. (%)	Case no. (%)	OR (95% CI)[c]	P value
CCAG/CCAG	182 (17.9)	147 (14.6)	1.00 (reference)	
TTAC/TTGC	69 (6.8)	60 (6.1)	1.23(0.76–1.98)	0.41
TTAC/TCAG	30 (3.0)	29 (2.9)	1.31(0.69–2.49)	0.41
TTAC/TCAC	24 (2.4)	32 (3.2)	1.90(0.99–3.63)	0.05
TTAC/CCAG	136 (13.4)	143 (14.2)	1.76(1.22–2.54)	0.003
TTGC/TTGC	43 (4.2)	58 (5.8)	2.06(1.23–3.45)	0.006
TTGC/TCAG	37 (3.6)	48 (4.8)	1.77(1.02–3.05)	0.04
TTGC/TCAC	42 (4.1)	43 (4.3)	1.52(0.89–2.60)	0.13
TTGC/CCAG	209 (20.6)	204 (20.3)	1.41(1.01–1.96)	0.04
TCAG/CCAG	83 (8.2)	70 (7.0)	1.19(0.76–1.84)	0.45
TCAC/CCAG	62 (6.1)	58 (5.8)	1.33(0.82–2.17)	0.25
Other diplotypes[d]	98 (9.7)	113 (11.2)	1.59(1.07–2.36)	0.02

a Order of polymorphism-rs4886605, rs4646421, rs1048943, rs2198843.
b In cases where more than one diplotype is inferred for the individual, the diplotype with the greatest probability was selected.
c Adjusted for age, gender, number of pack-years smoked, gram of alcohol intake per day and HBsAg status (positive or negative).
d All other diplotypes had frequency <2% in either cases or controls.

in score test of haplotype distribution. Using the regression model in haplo.glm to evaluate other haplotype effects on HCC compared to CCAG with age, gender, pack-years of smoking, alcohol intake per day and HBsAg carrying status as covariates, an overall association was observed on the haplotype TTGC (OR 1.26, 95% CI 1.04–1.52) (Table 4). And two other haplotypes CTAC (OR 1.76, 95% 0.99–3.12) and TCAG (OR 1.31, 95% CI 0.99–1.71) showed borderline significant association ($P = 0.05$) compared to CCAG. Consistent with the single SNP analysis results, haplotype associations restricted to non-smokers revealed significant results (Supplemental Table S3). CCAG frequency difference remains significant between cases and controls ($P = 0.009$, $P_{sim} = 0.009$). By comparison, positive association on haplotype TTGC (OR 1.26, 95% CI 0.97–1.63) showed only marginal significance ($P = 0.08$). Furthermore, associations were observed for the haplotypes TCAG (OR 1.59, 95% CI: 1.10–2.31) and TTAC (OR = 1.39, 95% CI: 1.02–1.88).

As an exploratory effort, we reconstructed diplotypes for each individual to test the association with HCC risk. We designated diplotype CCAG/CCAG as the referent. Consistent with the haplotype analysis, a significantly increased risk was detected for TTGC/TTGC diplotype carriers (OR 2.06, 95% CI 1.23–3.45, $P = 0.006$, Table 5). Two other diplotypes containing one TTGC haplotype (TTGC/TCAG and TTGC/CCAG) also conferred risk effect with adjusted ORs ranging from 1.41 to 1.77. The most significant risk effect was observed for haplotype TTAC/CCAG with an adjusted OR of 1.76 (95% CI 1.22–2.54, $P = 0.003$).

4. Discussion

Well recognised to be an environmentally related disease, variation in *CYP1A1* is a relevant genetic factor to predispose HCC risk, for CYP1A1 is the most active enzyme in converting procarcinogens into active compounds, which leads to DNA adduct formation.[20,21] However, previous studies investigated only a limited number of putatively functional polymorphisms in relatively small studies. Therefore, we evaluated the effects of four common *CYP1A1* SNPs on HCC in this large-scale case-control study. The associations between SNPs rs4886605, rs4646421 and rs2198843 and overall risk of HCC remained significant after multiple testing correction, and the effects of rs4886605 and rs4646421 were still pronounced in non-smokers with a 40% reduction (CC versus TT) and a 1.56-fold increase (TC + TT versus CC) in HCC risk, respectively. In haplotype/diplotype analysis, the haplotype consisted of risk allele in all loci (TTGC) and its haplotype pair exhibited an increased risk of HCC with adjusted OR of 1.26 and 2.06, respectively. Corresponding to positive results on rs4886605 and rs4646421, both the haplotypes TTGC and TTAC had T allele of rs4886605 at the first locus and T allele of rs4646421 at the second locus. Compared to TT genotype, the rs4886605 CC homozygotes showed a protective effect of HCC, and the rs4886605 C allele appears most frequently combined with wild type allele of the other three loci in haplotype CCAG (43.3% in control and 39.7% in case). The consistency in single-locus and haplotype/diplotype analyses suggest that inherited variation in *CYP1A1* may have a modest impact on HCC susceptibility.

Our result suggested a positive association with rs4886605 variant in 5′ flanking region. Many xenobiotic responsive elements (XREs) are present in this region of the *CYP1A1* gene. Chemical procarcinogens undergo metabolic activation by elevating CYP1A1 enzyme expression via ligand-activated aryl hydrocarbon receptor (AHR) requiring interaction with XREs. Such elevation of CYP1A1 mRNA expression was reported in mice livers and promoted liver tumourigenesis.[22] It is possible that the variation in this sequence may affect the affinity of AHR for some nearby XREs depending on changed topological structure or some other complex mechanism that is observed as an impact of rs4886605 on HCC risk. Moreover, genetic variants in the 5′ flanking region of *CYP1A1* may be important as *CYP1A1* shares the 5′ flanking region with *CYP1A2*, which is another crucial enzyme involved in the activation of chemical procarcinogens in the liver. The common region sequence has a bidirectional regulatory activity of the *CYP1A1/1A2* gene cluster, a phenomenon proposed to be abundant in the human genome.[23] Recently, two novel regulatory regions in the shared intergenic spacer were identified

to modulate *CYP1A1* and *CYP1A2* transcriptional activation simultaneously.[24] The modified HCC risk associated with rs4886605 may be related to regulated activities of both *CYP1A1* and *CYP1A2* genes. Although there was no previous report on association of rs4886605 in *CYP1A1* with HCC risk, one haplotype containing upstream SNPs of the *CYP1A2* gene was found to be associated with HCC risk in a recent study,[25] which indicates biological plausibility of sequence variation in common region. Because little is known about the direct relationship between rs4886605 and the activity of the enzyme, further research on the functions of SNPs in this region is needed.

We also provided consistent and relatively strong evidence of a significantly increased HCC risk associated with SNP rs4646421 located in intron 1 region of *CYP1A1*. Without sustained induction caused by exposure to procarcinogens, the CYP1A1 enzyme has a low level of transcription in normal conditions. Consequently, the existence of negative regulatory element has been speculated. Intron 1 sequence was found to have a repressive effect on *CYP1A1* transcription after AHR activation, and its internal sequence is highly homologous to a transcription inhibitor binding site.[26] According to functional prediction using FASTSNP,[27] rs4646421 variant may cause impaired inhibition of the enzyme by altering the sequence specificity adequate for the upstream stimulatory factor (USF) recognition, which may repress *CYP1A1* transcription, resulting in high inducibility of CYP1A1 and carcinogenesis of HCC. However, the finding with rs4646421 is intriguing, because the risk appears to be independent of allelic dose. Although the biological basis for a higher risk among carriers of heterozygotes remains unknown, it is possible that the observed risk associated with rs4646421 heterozygotes may reflect the effect of some potential susceptibility loci in linkage disequilibrium with this genotype. One recent study reported that a cryptic intron excision in exon 2 can form an enzymatically active novel spliced variant of *CYP1A1* (*CYP1A1v*) in epithelial ovarian cancer cells.[28] The alternative splicing may contribute to ovarian cancer initiation and progression, and it may be a widespread event in cancers.[28] It is likely that causal variants within or in proximity of exon 2 may have the potential to form a more active enzyme that facilitates tumour formation and progression, as suggested by the observed association between HCC and rs4646421, although this hypothesis needs to be confirmed in future studies.

Our approach to the SNP selection in this study was to select common SNPs that can capture genetic variation of the gene. The inclusion of SNP rs1048943 was based on its putative function. The association between rs1048943 and the risk of HCC in overall samples was only borderline significant but was more significant in a subgroup of smokers carrying the variant homozygous genotype. In a previous study, an increased risk of HCC was reported with rs1048943 (Ile462Val) in smokers, while no statistically significant difference was observed in overall study population.[10] In a subsequent study of hepatitis C infection, the rs1048943 variant allele was overrepresented in HCV patients with liver disease compared with carriers but was not associated with HCC risk.[11] Our results supported the previous finding that rs1048943 may modulate HCC risk in smokers.

The positive association of rs4886605 and rs4646421 was primarily observed in non-smokers as revealed in stratification analysis. Similar results were observed in studies of other cancers in which the risk of non-smokers was modulated by *CYP1A1* genetic variations.[29–31] These findings together support the notion that 'a low dose effect' is more distinct among individuals exposed to lower levels of environmental carcinogens,[32,33] which is a feature of genetic susceptibility to cancer. Another explanation is that the effect of *CYP1A1* polymorphism on HCC risk may be related to some unidentified, low-level carcinogens in the environment. Although cigarette smoking contains multiple PAHs, it is not the only source of the procarcinogens, which are widespread pollutants found in air and diet in addition to smoking. A similar significant association observed in non-drinkers may reflect a correlated pattern between smoking and drinking habits in Han Chinese population.

This study has a few strengths. The first one is the selection of common SNPs in the upstream and non-coding regions of *CYP1A1*, where increasing studies focused on to explore possible regulatory function. Second, the present study also benefited from relatively larger sample size than the previously published studies. The adequate number of cases and controls made the genotype frequency more stable, especially for SNPs with low MAF, and made significant results in certain subgroup to be reliable. Using these genotypes, *CYP1A1* haplotypes were reconstructed with high certainty (the probability of the most likely haplotype pair is >95% for 92.7% subjects). However, several limitations exist in the current study. One potential concern is population stratification. However, the SNP frequencies we observed were similar to that previously reported for Han Chinese in the HapMap, and the residences of controls were comparable to those of cases; it is unlikely that our findings were significantly influenced by this type of bias. Furthermore, there is evidence to show that population stratification may not be obvious in the Han Chinese population of East China, in which we ascertained our sampling. For example, F_{st} between northern Han Chinese population and HapMap Han Chinese population samples (CHB) estimated from 19,934 SNPs was 0.0007, while the F_{st} between a southern Han Chinese population and CHB was only slightly higher (0.0008).[34] According to the previous study on genomic control, for case-control studies where F_{st} is on the order of 0.01,[35] the population structure is relatively subtle. Therefore, we expected that the issue of population structure in this study is not serious to affect our results. Because controls were selected from hepatitis examination we could not rule out the possible selection bias. However, the inadequacy may not have a significant impact on risk evaluation of genetic variation for the same effects were observed in both HBsAg negative and HBsAg positive individuals. Finally, the risk effects of genetic factor we reported were only modest, and may not be significant after correction for multiple testing. More studies with larger samples will be required to confirm the results.

In summary, our findings support the hypothesis that genetic variants of *CYP1A1* are associated with HCC risk, especially in subjects exposed to low level of environmental carcinogens such as smoking. In light of our results, more epidemiologic and functional studies are warranted to further

explore the role of CYP1A1 in HCC carcinogenesis other than metabolising PAHs in tobacco.

Conflict of interest statement

None declared.

Acknowledgements

This work was supported in part by the China National Key Basic Research Program Grants 2002CB512902 and National '211' Environmental Genomics Grant. This work was completed when Dr. Yin Yao Shugart moved to NIMH/NIH. The views expressed in this article do not necessarily represent the views of the NIMH, NIH, HHS or the United States Government.

Appendix A. Supplementary material

Supplementary data associated with this article can be found, in the online version, at doi:10.1016/j.ejca.2008.11.007.

REFERENCES

1. Bosch FX, Ribes J, Diaz M, Cleries R. Primary liver cancer: worldwide incidence and trends. *Gastroenterology* 2004;**127**(5 Suppl 1):S5–S16.
2. Parkin DM, Bray F, Ferlay J, Pisani P. Global cancer statistics, 2002. *CA Cancer J Clin* 2005;**55**(2):74–108.
3. Hertl M, Cosimi AB. Liver transplantation for malignancy. *Oncologist* 2005;**10**(4):269–81.
4. Farazi PA, DePinho RA. Hepatocellular carcinoma pathogenesis: from genes to environment. *Nat Rev Cancer* 2006;**6**(9):674–87.
5. Chen CJ, Yu MW, Liaw YF. Epidemiological characteristics and risk factors of hepatocellular carcinoma. *J Gastroenterol Hepatol* 1997;**12**(9–10):S294–308.
6. Chen SY, Wang LY, Lunn RM, et al. Polycyclic aromatic hydrocarbon-DNA adducts in liver tissues of hepatocellular carcinoma patients and controls. *Int J Cancer* 2002;**99**(1):14–21.
7. Wu HC, Wang Q, Wang LW, et al. Polycyclic aromatic hydrocarbon- and aflatoxin-albumin adducts, hepatitis B virus infection and hepatocellular carcinoma in Taiwan. *Cancer Lett* 2007;**252**(1):104–14.
8. Staal YC, van Herwijnen MH, Pushparajah DS, et al. Modulation of gene expression and DNA-adduct formation in precision-cut liver slices exposed to polycyclic aromatic hydrocarbons of different carcinogenic potency. *Mutagenesis* 2007;**22**(1):55–62.
9. Kawajiri K, Watanabe J, Gotoh O, Tagashira Y, Sogawa K, Fujii-Kuriyama Y. Structure and drug inducibility of the human cytochrome P-450c gene. *Eur J Biochem* 1986;**159**(2): 219–25.
10. Yu MW, Chiu YH, Yang SY, et al. Cytochrome P450 1A1 genetic polymorphisms and risk of hepatocellular carcinoma among chronic hepatitis B carriers. *Brit J Cancer* 1999;**80**(3–4):598–603.
11. Silvestri L, Sonzogni L, De Silvestri A, et al. CYP enzyme polymorphisms and susceptibility to HCV-related chronic liver disease and liver cancer. *Int J Cancer* 2003;**104**(3):310–7.
12. Smart J, Daly AK. Variation in induced CYP1A1 levels: relationship to CYP1A1, Ah receptor and GSTM1 polymorphisms. *Pharmacogenetics* 2000;**10**(1):11–24.
13. Ryder SD. Guidelines for the diagnosis and treatment of hepatocellular carcinoma (HCC) in adults. *Gut* 2003;**52 Suppl 3**:iii1–iii8.
14. Carlson CS, Eberle MA, Rieder MJ, Yi Q, Kruglyak L, Nickerson DA. Selecting a maximally informative set of single-nucleotide polymorphisms for association analyses using linkage disequilibrium. *Am J Hum Genet* 2004;**74**(1):106–20.
15. Song N, Tan W, Xing DY, Lin DX. CYP 1A1 polymorphism and risk of lung cancer in relation to tobacco smoking: a case-control study in China. *Carcinogenesis* 2001;**22**(1):11–6.
16. Park JY, Muscat JE, Ren Q, et al. CYP1A1 and GSTM1 polymorphisms and oral cancer risk. *Cancer Epidemiol Biomarkers Prev* 1997;**6**(10):791–7.
17. Nyholt DR. A simple correction for multiple testing for single-nucleotide polymorphisms in linkage disequilibrium with each other. *Am J Hum Genet* 2004;**74**(4):765–9.
18. Schaid DJ, Rowland CM, Tines DE, Jacobson RM, Poland GA. Score tests for association between traits and haplotypes when linkage phase is ambiguous. *Am J Hum Genet* 2002;**70**(2):425–34.
19. Stephens M, Donnelly P. A comparison of bayesian methods for haplotype reconstruction from population genotype data. *Am J Hum Genet* 2003;**73**(5):1162–9.
20. Stiborova M, Martinek V, Rydlova H, Koblas T, Hodek P. Expression of cytochrome P450 1A1 and its contribution to oxidation of a potential human carcinogen 1-phenylazo-2-naphthol (Sudan I) in human livers. *Cancer Lett* 2005;**220**(2):145–54.
21. Roberts-Thomson SJ, McManus ME, Tukey RH, Gonzalez FF, Holder GM. The catalytic activity of four expressed human cytochrome P450s towards benzo[a]pyrene and the isomers of its proximate carcinogen. *Biochem Biophys Res Commun* 1993;**192**(3):1373–9.
22. Moennikes O, Loeppen S, Buchmann A, et al. A constitutively active dioxin/aryl hydrocarbon receptor promotes hepatocarcinogenesis in mice. *Cancer Res* 2004;**64**(14): 4707–10.
23. Trinklein ND, Aldred SF, Hartman SJ, Schroeder DI, Otillar RP, Myers RM. An abundance of bidirectional promoters in the human genome. *Genome Res* 2004;**14**(1):62–6.
24. Ueda R, Iketaki H, Nagata K, et al. A common regulatory region functions bidirectionally in transcriptional activation of the human CYP1A1 and CYP1A2 genes. *Mol Pharmacol* 2006;**69**(6):1924–30.
25. Chen X, Wang H, Xie W, et al. Association of CYP1A2 genetic polymorphisms with hepatocellular carcinoma susceptibility: a case-control study in a high-risk region of China. *Pharmacogenet Genomics* 2006;**16**(3):219–27.
26. Sakata Y, Yoshioka W, Tohyama C, Ohsako S. Internal genomic sequence of human CYP1A1 gene is involved in superinduction of dioxin-induced CYP1A1 transcription by cycloheximide. *Biochem Biophys Res Commun* 2007;**355**(3):687–92.
27. Yuan HY, Chiou JJ, Tseng WH, et al. FASTSNP: an always up-to-date and extendable service for SNP function analysis and prioritization. *Nucleic Acids Res* 2006;**34**(Web Server issue): W635–41.
28. Leung YK, Lau KM, Mobley J, Jiang Z, Ho SM. Overexpression of cytochrome P450 1A1 and its novel spliced variant in ovarian cancer cells: alternative subcellular enzyme compartmentation may contribute to carcinogenesis. *Cancer Res* 2005;**65**(9):3726–34.
29. Taioli E, Ford J, Trachman J, Li Y, Demopoulos R, Garte S. Lung cancer risk and CYP1A1 genotype in African Americans. *Carcinogenesis* 1998;**19**(5):813–7.

30. Ishibe N, Wiencke JK, Zuo ZF, McMillan A, Spitz M, Kelsey KT. Susceptibility to lung cancer in light smokers associated with CYP1A1 polymorphisms in Mexican- and African–Americans. *Cancer Epidemiol Biomarkers Prev* 1997;**6**(12):1075–80.
31. Vineis P, Kogevinas M, Simonato L, Brennan P, Boffetta P. Levelling-off of the risk of lung and bladder cancer in heavy smokers: an analysis based on multicentric case-control studies and a metabolic interpretation. *Mutat Res* 2000;**463**(1):103–10.
32. Vineis P, Bartsch H, Caporaso N, et al. Genetically based N-acetyltransferase metabolic polymorphism and low-level environmental exposure to carcinogens. *Nature* 1994;**369**(6476):154–6.
33. Vineis P, Alavanja M, Garte S. Dose-response relationship in tobacco-related cancers of bladder and lung: a biochemical interpretation. *Int J Cancer* 2004;**108**(1):2–7.
34. Xu S, Jin L. A genome-wide analysis of admixture in Uyghurs and a high-density admixture map for disease-gene discovery. *Am J Hum Genet* 2008;**83**(3):322–36.
35. Pritchard JK, Donnelly P. Case-control studies of association in structured or admixed populations. *Theor Popul Biol* 2001;**60**(3):227–37.