Carcinogenesis vol.30 no.12 pp.2031–2036, 2009
doi:10.1093/carcin/bgp239
Advance Access publication October 5, 2009

A Case–control and a family-based association study revealing an association between *CYP2E1* polymorphisms and nasopharyngeal carcinoma risk in Cantonese

Wei-Hua Jia[1,2,3,*,†], Qing-Hua Pan[1,2,†], Hai-De Qin[2], Ya-Fei Xu[1,2], Guo-Ping Shen[1,2], Lina Chen[4], Li-Zhen Chen[1,2], Qi-Sheng Feng[1,2], Ming-Huang Hong[1], Yi-Xin Zeng[1,2] and Yin Yao Shugart[5]

[1]State Key Laboratory of Oncology in Southern China, Sun Yat-sen University Cancer Center, Guangzhou 510060, People's Republic of China, [2]Department of Experimental Research, Sun Yat-sen University Cancer Center, Guangzhou 510060, People's Republic of China, [3]Department of Statistics and Epidemiology, School of Public Health, Sun Yat-sen University, Guangzhou 510060, People's Republic of China, [4]Department of Social Medicine, University of Bristol, White Ladies Road BS8 1TH, UK and [5]Department of Epidemiology, Johns Hopkins Bloomberg School of Public Health, Baltimore, MD 21205, USA

[*]To whom correspondence should be addressed. Tel: +86 20 87343195; Fax: +86 20 87343392;
Email: jiaweih@mail.sysu.edu.cn

Nasopharyngeal carcinoma (NPC) is rare in most parts of the world but is more prevalent in Southern China, especially in Guangdong. The cytochrome P450 2E1 (CYP2E1) has been recognized as one of the critically important enzymes involved in oxidizing carcinogens and is probably to be associated with NPC carcinogenesis. To systematically investigate the association between genetic variants in *CYP2E1* and NPC risk in Cantonese, two independent studies, a family-based association study and a case–control study, were conducted using the haplotype-tagging single-nucleotide polymorphism approach. A total of 2499 individuals from 546 nuclear families were initially genotyped for the family-based association study. Single-nucleotide polymorphisms (SNPs) *rs9418990*, *rs915908*, *rs8192780*, *rs1536826*, *rs3827688* and one haplotype *h2* (*CGTGTTAA*) were revealed to be significantly associated with the NPC phenotype ($P = 0.045$–0.003 and $P = 0.003$, respectively). To follow up the initial study, a case–control study including 755 cases and 755 controls was conducted. Similar results were observed in the case–control study in individuals <46 years of age and had a history of cigarette smoking, with odds ratios (ORs) of specific genotypes ranging from 1.88 to 2.99 corresponding to SNP *rs9418990*, *rs3813865*, *rs915906*, *rs2249695*, *rs8192780*, *rs1536826*, *rs3827688* and of haplotypes *h2* with OR = 1.65 ($P = 0.026$), *h5* (*CCCGTTAA*) with OR = 2.58 ($P = 0.007$). The values of false-positive report probability were <0.015 for six SNPs, suggesting that the reported associations are less probably to be false. This study provides robust evidence for associations between genetic variants of *CYP2E1* and NPC risk.

Introduction

Nasopharyngeal carcinoma (NPC) is rare in most parts of the world, with an incidence rate of <1/100 000 person-years in most populations. However, NPC occurs much more frequently in southern China and Southeast Asia, with incidence rates ranging from 15 to 50/100 000 (1). Numerous studies have shown that NPC is a multifactorial disease and that risk factors include genetic components, Epstein-Barr virus (EBV) infection, environmental factors and probably interactions among these factors (2–4). Among these factors, EBV has been consistently identified as an important risk factor for NPC, and a dose-response relationship between EBV antibody titers and NPC risk has been previously demonstrated (5). Additionally, epidemiological studies have suggested that other environmental factors are related to NPC susceptibility, including salted fish consumption, cigarette smoking, alcohol consumption and occupational exposure to wood and formaldehyde. Of these, salted fish consumption and cigarette smoking are environmental risk factors for NPC that are consistently reported in the literature (6–8).

In terms of genetic influences, the roles that multiple genetic factors play in NPC carcinogenesis have been investigated over the last few decades (9–12). Hildesheim *et al.* (13) reported associations between human leukocyte antigen class I, II alleles and haplotypes and risk of NPC. In a later study, Goldsmith *et al.* (14) carried out a meta-analysis of 13 human leukocyte antigen type association studies and reported positive associations between human leukocyte antigen alleles *A2*, *B14* and *B46* and NPC. Other studies have reported that the polymorphisms in the genes encoding glutathione *S*-transferase M1 (15), the polymeric immunoglobulin receptor (16), Toll-like receptors (17), the DNA repair enzymes *XRCC1* and *hOGG1* (18,19), as well as polymorphisms in several other genes (20,21) were associated with the risk of developing NPC. In our previous study, which included 2252 Cantonese families with at least one member who was diagnosed with NPC, we reported that NPC tended to aggregate in families (22). In 2002, Feng *et al.* (23) conducted a genome-wide linkage analysis with microsatellite markers in 32 Cantonese families who had been identified as at high risk for developing NPC and mapped a NPC susceptibility locus to chromosome 4p15.1–q12. This linkage region was further narrowed to a 8.29 cM region on chromosome 4p11–4p14 using high-density microsatellites and single-nucleotide polymorphism (SNP) markers (24). In 2004, Xiong *et al.* (25) reported a susceptibility locus for familial NPC to chromosome 3p21. Together, these studies suggest that multiple genetic factors may contribute to the development of NPC. However, the underlying mechanism remains to be elucidated.

The cytochrome P450 family of enzymes is known to be involved in the metabolism of numerous xenobiotics. Some of these enzymes catalyze the first step in the metabolism of procarcinogens and thereby often create reactive metabolic intermediates that are capable of forming DNA adducts and leading to genetic mutations (26). Among the cytochrome P450 superfamily of enzymes, cytochrome P450 2E1 (CYP2E1) has been indicated as one of the critically important enzymes involved in oxidizing carcinogens such as nitroaromatic compounds, polycyclic aromatic hydrocarbons and arylamines, which are found in high concentrations in salted fish and tobacco.

Several *CYP2E1* genetic variants have been reported to be risk factors for NPC in the Taiwanese population (18,27,28). Hildesheim *et al.* analyzed *CYP2E1* RsaI (rs2031920) and DraI (rs6413432) polymorphisms in 50 NPC patients and 50 healthy controls collected from Taiwan and found that individuals who possessed the homozygous variant form of *CYP2E1* (*T/T* for RsaI or *A/A* for DraI) were at a 7.7-fold [95% confidence interval (CI) 0.87–68] or a 5-fold (95% CI 0.95–16) higher risk of developing NPC compared with those who carried the homozygous wild-type form (*C/C* for RsaI or *T/T* for DraI), respectively (27). Later, the authors expanded their study to include 364 NPC patients and 320 healthy controls to validate their initial findings. They reported that the carriers of the *T/T* genotype for RsaI had a 2.6-fold increased risk of developing NPC (95% CI 1.2–5.7), whereas those with the *A/A* genotype for DraI had a 1.9-fold increased risk of developing NPC (95% CI 0.98–3.7) (28). On the other hand,

Abbreviations: CI, confidence interval; CYP2E1, cytochrome P450 2E1; EBV, Epstein-Barr virus; FBAT, family-based association test; FPRP, false-positive report probability; HWE, Hardy–Weinberg equilibrium; htSNP, haplotype-tagging single-nucleotide polymorphism; LD, linkage disequilibrium; NPC, nasopharyngeal carcinoma; OR, odds ratio; SNP, single-nucleotide polymorphism; SYSUCC, Sun Yat-Sen University Cancer Center.

[†]These authors contributed equally to this work.

Kongruttanachok *et al.* analyzed the *CYP2E1* RsaI polymorphism among individuals of Thai origin and Chinese origin in Thailand. They did not detect any association between this polymorphism and NPC risk (29). It is possible that the differences in ethnic groups and the sample size may have led to incongruent results.

Recently, we conducted a family-based association study in a high-risk area in Guangdong, China. To comprehensively assess the role of genetic variants in *CYP2E1* for NPC susceptibility, we used the haplotype-tagging single-nucleotide polymorphism (htSNP) approach and included 11 *CYP2E1* htSNPs that were thought to represent the majority of genetic variants in the genomic region of the *CYP2E1* gene that we were studying. We then performed an independent case–control study (including 755 cases and 755 controls) to verify the results of our previous study. Moreover, information on cigarette smoking and salted fish consumption were collected and analyzed in the case–control study. Similar results we found in these two independent studies indicated that the association between genetic variants in *CYP2E1* and NPC risk may be robust.

Materials and methods

Study participants

In the family-based association study, all patients with pathologically confirmed NPC who were seen between December 2001 and December 2004 at Sun Yat-Sen University Cancer Center (SYSUCC), Guangzhou, China, the largest center for cancer treatment in Southern China, were recruited for participation in this study. Each NPC patient who consented to participate in the study was regarded as the proband of his or her family. All probands were administered a structured questionnaire including basic demographic information, including birth place, sex, age, etc. We also recruited their family members to participate in this study. Clinical data, including age at NPC diagnosis, histopathological diagnosis, as well as several other clinical parameters, were collected from their medical records after written informed consent was obtained. We also collected peripheral blood samples from each individual. By the end of December 2004, a total of 443 NPC pedigrees, which included 546 unrelated nuclear families, were enrolled in our study.

To validate the results of the family-based association study, an independent case–control study was performed. Cantonese patients living in Guangdong Province who had pathologically confirmed NPC and were seen at SYSUCC between October 2005 and October 2007 were recruited to participate in this study. Clinical data were collected from their medical records. During the same period, healthy controls were recruited from the physical examination centers of several large comprehensive hospitals in Guangdong and were frequency matched to the cases by age (±5 years), sex, geographic region and ethnicity. Using face-to-face interviews, trained SYSUCC staff interviewers collected data on demographics, dietary habits, cigarette smoking history, etc. To evaluate food intake, subjects were asked to choose from five intake-frequency categories including 'never', 'sometimes', 'monthly', 'weekly' and 'daily'. While analyzing the data, we merged the never and sometimes responses to create a negative reference group and combined the other three categories to form an exposure group. Individuals who had smoked at least 100 cigarettes in their lifetime were defined as 'smokers'. All study subjects had signed informed consent agreements before epidemiological data and blood samples were collected by trained SYSUCC staff interviewers.

For the purpose of both studies, we collected 5–10 ml venous blood specimens from study participants and genomic DNA was then extracted from the lymphocytes using the QIAamp DNA Blood Midi Kit (QIAGEN, Germany) following the manufacturer's protocol. Both studies were reviewed and approved by the Human Ethics Approval Committee of SYSUCC.

Genotyping and detection of virus capsid antigen/early antigen–immunoglobulin A antibody

Han Chinese SNP information for the candidate genes *CYP2E1* and *CYP1A1*, including variants located at 10–20 kb upstream and downstream of each gene, was obtained from the Single Nucleotide Polymorphism Database from National Center for Biotechnology Information and the International HapMap Project database. Using the selection criteria of $r^2 > 0.8$ and minor allele frequency $\geq 5\%$ in the Han Chinese population, we selected htSNPs to capture the common haplotypes (with frequency $\geq 5\%$) associated with each fragment. A total of 11 htSNPs were selected and genotyped using standardized TaqMan SNP Genotyping Assays (Applied Biosystems, Foster City, CA). TaqMan assays include unlabeled polymerase chain reaction primers and TaqMan minor groove binder probes that are separately labeled with 6-carboxy-fluorescein and victoria green-fluorescent protein. All polymerase chain reactions were performed in 96-well plates with either a GeneAmp polymerase chain reaction System 9700 thermocycler or a 7500 Real-Time System (both from Applied Biosystems).

The measurements of virus capsid antigen– and early antigen–immunoglobulin A antibody titers were conducted when the peripheral blood samples were collected. The sera were tested for IgA class antibody against virus capsid antigen and early antigen using an immunoenzymatic method in accordance with the kit instructions (Guangdong Zhongshan Biotechnology company, Zhong Shan City, China), which has been a routine method for NPC diagnosis in high-risk areas in China for decades. Serial dilutions of quality control sera (1:5, 1:10, 1:20, 1:40 and 1:80) were applied to each assay for evaluation of intra-set variability. To minimize the experimental error rate, all of the tests related to this study were conducted in the same lab and by the same technicians, and kits were also from the same company.

Statistical analysis

We analyzed the genotype data for the family-based association study using the Haploview software (version 4.0) (30) to conduct the Hardy–Weinberg equilibrium (HWE) test, standard transmission disequilibrium test and haplotype analysis. Haploview was also used to calculate Lewontin's D' value and the r^2 correlation coefficient (31). The default settings were used in these analyses, which invoke a one-sided upper 95% confidence limit of $D' > 0.98$ and a lower 95% confidence limit of > 0.7 to define SNP pairs in strong linkage disequilibrium (LD). Haplotypes were then constructed and their frequencies were estimated using an accelerated expectation–maximization algorithm in Haploview (30,32). We partitioned LD blocks using the method proposed by Gabriel *et al.* (31). Tests evaluating the association between single-loci, haplotypes and NPC were performed with the family-based association test software package (FBAT, version 1.7.3) (33). The genotypes of individuals with Mendelian errors were excluded from further analysis.

For the case–control study, the chi-squared and Fisher's exact tests were used to assess the differences in the distribution of alleles and genotypes between cases and controls. To control for confounding factors, an unconditional logistic regression analysis was conducted and adjusted odds ratios (ORs) and 95% CIs were calculated. The analyses were performed using the Stata statistical software package (version 10.0) (Stata Corporation, College Station, TX). For power calculation, the QUANTO program (Version 1.2) was used (34,35). The values of false-positive report probability (FPRP) were assessed by the use of method described by Wacholder *et al.* (36).

Results

In the family-based association study, a total of 443 NPC pedigrees, were enrolled. In the case–control study, 755 cases and 755 controls were used, and no significant differences were noted between cases and controls with regard to age, gender or cigarette smoking history. Cases had significantly higher frequencies of salted fish consumption than controls during both childhood and adulthood. Baseline demographic, clinical and lifestyle characteristics of individuals participating were shown in Table I. EBV infection is strongly associated with NPC development. In this study, 96.3% of the case patients versus 17% of control subjects were shown positive of virus capsid antigen–immunoglobulin A, 77.4% versus 0.5% of early antigen–immunoglobulin A, the correlation coefficient was 0.80 and 0.79, respectively.

The results of the allele frequency and HWE tests are presented in Table II. All htSNPs were located either in the intronic region or in the 5′ or 3′-flanking region of the *CYP2E1* gene. The minor allele frequencies of the SNPs we tested ranged from 0.198 to 0.463. The genotypic frequencies in the study population did not violate the principles of HWE for any loci (all $P > 0.05$) except rs2987800 in the family-based dataset; the genotypic distributions in rs2987800 significantly deviated from HWE ($P_{H-W} = 0.0002$). We therefore excluded this SNP from the analysis.

The results of the single-locus analysis that we performed using FBAT are shown in Table III. The results of the chi-squared test examining transmitted and untransmitted alleles suggested that minor alleles of 5 of 10 SNPs were significantly associated with NPC risk under either a dominant or recessive model, with P values ranging from 0.045 to 0.003. Another SNP, rs915906, had a marginally significant P value ($P = 0.055$). Haplotype construction was conducted and one block including seven SNPs was constructed using Haploview. Five common haplotypes (each with a frequency of ≥ 0.05) were constructed using the expectation–maximization algorithm implemented in haplotype FBAT. Haplotype '*CGTGTTA*' was found to be

significantly associated with the risk of developing NPC ($P = 0.004$). Further, Using 5000 permutation test runs, we also detected a significant association ($P = 0.003$) between the haplotype and NPC risk (data not shown). The variant rs3827688, which gave the strongest signal of association in univariate analysis, was not in the main LD block as the other SNPs. To examine the potential influence of this SNP on the risk haplotype of the main block, we included this variant in haplotype construction and frequency estimation. The haplotype '*CGTGTTAA*' (h2), which contained the rs3827688 risk allele *A*, remained to be significantly associated with the NPC phenotype ($Z = 2.851$, $P = 0.003$). The frequency of this haplotype was estimated to be 15.9% in the study subjects (Table V). These results provided evidence that a susceptibility locus may exist and be harbored in this region.

To more precisely assess the association between *CYP2E1* genetic variants, environmental factors and NPC risk, we conducted an independent case-control study (Table IV). We genotyped eight SNPs in the family-based association study. While we included the entire case–control set in the analysis, we did not obtain significant results in the single SNP analysis (except for rs3813865: OR = 1.25, $P = 0.050$) or the haplotype analysis. However, after stratifying the data by age and smoking history, we found that among individuals who had a positive smoking history and who were <46 years old, seven of eight SNPs were observed to have significantly higher frequencies in the cases than in the controls. Patients possessing genotypes containing risk alleles had a higher risk of developing NPC (OR = 1.88 to 2.99; $P = 0.0140$–0.0001). After we stratified the data, power calculation showed that the power was >0.96 for six significant SNPs, but *rs3827688* had power of 0.79. Moreover, two haplotypes (h2 and h5) were associated with a significantly increased risk of NPC (OR 1.65, $P = 0.026$ and OR 2.58, $P = 0.007$, respectively). These results were consistent with the findings from the family-based association study, in which we also found that the h2 haplotype was associated with NPC risk. We then calculated the value of FPRP using an appropriate method. Our calculation showed that six SNPs with significant findings showed FPRP values <0.015, one SNP *rs3827688* was 0.176. For the haplotype of *h2* and *h5*, FPRP was 0.220 and 0.129, respectively.

Table I. Characteristics of study participants in case–control study

Variable	Case ($n = 755$)	Control ($n = 755$)	*P*
Sex (%)			
Male	557 (74)	556 (74)	0.953
Female	198 (26)	199 (26)	
Age (%), mean ± SD (years)	46.6 ± 10.74	46.7 ± 11.0	0.975
<30	33 (4)	30 (4)	
30–39	173 (23)	166 (22)	
40–49	237 (31)	250 (33)	
50–59	219 (29)	216 (29)	
≥60	92 (12)	93 (12)	
Smoking history (%)			
Never	344 (45.6)	363 (48.1)	0.431
Ever	401 (53.1)	390 (51.7)	
Salted fish consumption (%)			
Adulthood			
Never	587	691	<0.0001
Ever	162	63	
Childhood			
Never	356	561	<0.0001
Ever	393	190	
VCA–IgA (%)			
Negative (<1:10)	28 (3.7)	627 (83)	<0.0001
1:10–1:40	78 (10.3)	119 (15.8)	
1:80–1:320	430 (57.0)	8 (1.1)	
≥1:640	208 (27.5)	1 (0.1)	
EA–IgA (%)			
Negative	171 (22.6)	751 (99.5)	<0.0001
1:10–1:40	369 (48.9)	4 (0.5)	
≥1:80	204 (27.0)	0 (0)	

EA–IgA, early antigen–immunoglobulin A; VCA–IgA, virus capsid antigen–immunoglobulin A.

Discussion

Several previous studies have demonstrated an association between NPC risk and consumption of nitrosamine-containing foods, cigarette smoking and organic solvent exposure (7,13,37–39). Yi *et al.* (40) demonstrated that NPC patients living in high-risk areas in southern China had a higher potential to form nitrosamines endogenously than those living in low-risk areas. Later, Yu *et al.* reported that feeding Wistar rats Chinese salted fish containing nitrosamines could induce the formation of malignant nasal cavity tumors (8). *CYP2E1* is involved in the metabolic activation of numerous procarcinogens, such as nitrosamines, halogenated hydrocarbons, polycyclic aromatic hydrocarbons and arylamines, as well as other procarcinogens (41,42). Therefore, genetic variation in this carcinogen-metabolic gene may lead to increased nasopharyngeal carcinoma susceptibility. However, strong and consistent links between *CYP2E1* genetic variants and NPC risk had not been established in Cantonese.

We conducted two independent studies to investigate the association between *CYP2E1* genetic variants and NPC risk. To our knowledge, this is the first comprehensive genetic association study examining the relationship between *CYP2E1* genetic variants and

Table II. The MAFs and the results of HWE tests of 11 htSNPs

SNP ID	Position[a]	Genotype[b]	Gene region	Family-based association study		Case–control study	
				MAF	P_{H-W}[c]	MAF	P_{H-W}[c]
rs9418984	135172346	G/C	5′ flanking	0.463	0.574	—	—
rs10857733	135182793	G/A	5′ flanking	0.198	0.165	—	—
rs9418990	135187956	T/C	5′ flanking	0.441	0.688	0.436	0.495
rs3813865	135189234	G/C	5′ flanking	0.250	0.973	0.231	0.736
rs915906	135193728	T/C	Intron 2	0.265	0.316	0.246	0.743
rs915908	135196949	G/A	Intron 5	0.203	0.215	0.173	0.933
rs2249695	135202158	C/T	Intron 8	0.435	0.325	0.417	0.354
rs8192780	135204115	G/T	3′ flanking	0.444	0.834	0.436	0.186
rs1536826	135207229	C/A	3′ flanking	0.442	0.781	0.436	0.221
rs3827688	135219248	G/A	3′ flanking	0.321	0.737	0.256	0.779
rs2987800	135231917	G/T	3′ flanking	0.248	0.0002	—	—

MAF, minor allele frequency.
[a]The chromosome position listed here is taken from the Single Nucleotide Polymorphism Database from National Center for Biotechnology Information build 129.
[b]First allele is major allele and the second is minor allele.
[c]P_{H-W} represents the *P*-value of HWE tests.

NPC risk using a family-based association study design and a large sample case–control study design in Cantonese. We assessed the association between genetic polymorphisms within *CYP2E1* and NPC risk in the 546 Cantonese nuclear families and were able to identify several genetic variants in the *CYP2E1* gene that conferred an elevated risk of developing NPC. Consequently, we conducted an independent case–control study to validate these results. We found similar results in the single SNP and haplotype analyses we performed in this case–control study only when we limited our analysis to include only patients ≤46 years of age who smoked cigarettes. These results were inconsistent with the findings from a previous Taiwanese study, which found that the increased risk conferred by a given CYP2E1 variant (RsaI digestion, c2 allele) was limited to nonsmokers (28). Notably, we did not observe differences in NPC risk after we stratified the data by other risk factors, such as salted fish intake, salted vegetable consumption, etc. (Data not shown).

We would like to point out that the significantly associated variant rs3827688, resides ∼5.5 kb from the 3′-terminal region of *CYP2E1* gene. We speculate that this variant may contribute to NPC risk via affecting the transcriptional regulation of *CYP2E1* gene expression or the posttranscriptional modification of *CYP2E1* messenger RNA, in a manner similar to the way the RsaI polymorphism of the gene exerts its effect (43–45). However, it is also plausible that rs3827688 may serve as a genetic marker that is in strong LD with other yet-to-be-defined genetic variants, which may themselves affect the function of *CYP2E1*. Therefore, it will be meaningful to analyze the DNA sequence surrounding this SNP to discovery of the true causal genetic variants that directly alter *CYP2E1* function and confer an increased risk of NPC. Therefore, further functional studies are needed to better understand the association between *CYP2E1* genetic variants and NPC risk.

Our study has several unique strengths. First, we systematically studied the association between *CYP2E1* genetic polymorphisms and NPC risk using the htSNP approach and successfully identified disease-associated variants in a homogeneous population. Using an htSNP approach, a subset of SNPs was selected to represent the majority of genetic variants in a defined genomic region, which is

Table III. Single-locus association study examining the association between a specific *CYP2E1* genetic variant and NPC risk in the family-based association study

SNP ID	Minor allele	Dominant model		Recessive model	
		$\chi 2$	$P_{Aa/aa}$[a]	$\chi 2$	P_{aa}[a]
rs9418990	C	8.184	**0.017**[b]	4.017	0.134
rs3813865	C	3.546	0.170	2.708	0.258
rs915906	C	1.321	0.517	5.796	**0.055**[b]
rs915908	A	1.305	0.521	6.534	**0.038**[b]
rs2249695	T	4.331	0.115	0.556	0.757
rs8192780	T	6.211	**0.045**[b]	2.433	0.296
rs1536826	A	6.371	**0.041**[b]	3.727	0.155
rs3827688	A	11.588	**0.003**[b]	4.584	0.101

$\chi 2$, chi-squared test.
[a]Analyzed using FBAT software, adjusted by age of onset.
[b]The statistical significantly *P* values or boundary *P* values.

Table IV. Association between *CYP2E1* gene and NPC risk in the case–control study

SNP	Genotype	Total				Smokers aged <46 years			
		Case	Control	OR (95% CI)[a]	*P*	Case	Control	OR (95% CI)[a]	*P*
rs9418990	TT	234	252			41	65		
	TC+CC	519	503	1.20 (0.95–1.52)	0.125	133	91	2.95 (1.69–5.17)	0.0002
rs3813865	GG	428	467			86	104		
	GC+CC	326	287	1.25 (1.00–1.56)	0.050	89	52	2.35 (1.40–3.92)	0.0011
rs915906	TT	416	444			83	103		
	TC+CC	339	310	1.16 (0.93–1.44)	0.196	92	53	2.45 (1.47–4.08)	0.0006
rs915908	GG	509	524			124	104		
	GA+AA	246	230	1.18 (0.93–1.49)	0.167	51	52	0.93 (0.55–1.57)	0.7860
rs2249695	CC	258	263			48	67		
	CT+TT	495	490	1.12 (0.89–1.41)	0.334	127	89	2.59 (1.51–4.45)	0.0006
rs8192780	GG	241	252			43	67		
	GT+TT	513	502	1.16 (0.92–1.46)	0.224	132	88	2.99 (1.72–5.21)	0.0001
rs1536826	CC	241	251			43	66		
	CA+AA	513	504	1.15 (0.91–1.45)	0.243	132	90	2.94 (1.69–5.13)	0.0001
rs3827688	GG	422	410			92	101		
	GA+AA	328	340	1.00 (0.80–1.25)	0.982	80	54	1.88 (1.13–3.13)	0.0140

[a]ORs were adjusted for sex, age, smoking and Cantonese salted fish consumption in childhood and adulthood.

Table V. Association between *CYP2E1* haplotype and NPC risk in the family-based association study and the case–control study

Hap[a]	Block	Family-based association study (Total 546 nuclear families)				Case–control study Smokers aged <46 years				
		Haplotype frequency[b]	Fam[c]	Z	*P*	Haplotype frequency[b]	Case	Control	OR (95% CI)	*P*
h1	TGTGCGCG	0.329	88	1.316	0.188	0.403	120	149	Reference	—
h2	CGTGTTAA	0.159	63	2.851	**0.003**[d]	0.173	65	49	1.65 (1.03–2.63)	**0.026**[d]
h3	TGTACGCG	0.168	88	−1.483	0.138	0.166	50	61	1.02 (0.64–1.63)	0.938
h4	CCCGTTAG	0.159	77	−0.213	0.831	0.150	51	48	1.32 (0.81–2.15)	0.239
h5	CCCGTTAA	0.077	41	−0.141	0.887	0.062	27	13	2.58 (1.22–5.68)	**0.007**[d]

[a]The code of each haplotype.
[b]The frequency of each haplotype.
[c]The number of the nuclear families carrying corresponding haplotype.
[d]The statistical significantly *P* values.

considered to be a powerful method to identify the association between genetic loci and disease states. A given set of htSNPs in-and-of themselves may or may not be functionally related to a given disease state but may be capable of indirectly identifying the true causal variants of that disease state.

We included a total of 2499 subjects within 546 nuclear families in the family-based association study. This study population appeared to be fairly homogeneous with respect to social and cultural norms. Furthermore, FBAT is a generalized approach, which is not susceptible to biases due to population admixture/stratification, misspecification of the trait distribution or selection based on trait (33,46,47). However, although we made an effort to collect information regarding tobacco use and dietary history, as well as exposure to other environmental factors, there was nonetheless a severe amount of missing data on these parameters. Therefore, we were not able to adjust for environmental exposures as confounders or assess the effects of gene–environment interactions implicated in NPC susceptibility in the family-based study. However, we followed up the results revealed by the family-based association study using an independent case–control study, in which we included environmental measures and supported the findings obtained from the family-based design.

Several important polymorphisms in *CYP2E1* (e.g. m1(RsaI), a C:T mutation in the 5′-flanking region of the gene) appear to be involved in the transcriptional regulation of *CYP2E1* (43). Individuals with this genetic variant form have higher hepatic *CYP2E1* messenger RNA and protein levels and have a greater ability to metabolize acetaminophen (44,48). CYP2E1 is also expressed in nasopharyngeal tissue (49,50). Hou *et al.* (49) reported that the expression of *CYP2E1* was regulated by doxycycline in a dose-dependent manner when the catalytic activity of CYP2E1 was assayed in the NIH 3T3/rtTA cell line. CYP2E1 may play an important role in the development of NPC that is induced by certain indirect carcinogens. However, there is not enough direct evidence to link the genetic variants and functional mechanism of this carcinogen-metabolizing gene to NPC carcinogenesis, therefore, an in-depth biological study aimed at further clarifying this linkage is needed to further elucidate the genetic factors that are directly involved in NPC carcinogenesis.

To summarize, we consistently identified disease-associated *CYP2E1* variants in a homogeneous Cantonese population. To our knowledge, this is by far the largest association study performed in the Cantonese population that has focused on a carcinogen-metabolizing gene and provided a robust evidence for an association between the presence of specific genetic variants and cancer risk. In-depth investigations of the specific variants linked to *CYP2E1* function are needed to further elucidate the relationship between these genetic variants and NPC carcinogenesis.

Electronic-database information

URLs for data presented herein are as follows: dbSNP home page, http://www.ncbi.nlm.nih.gov/SNP/; HapMap home page, http://www.hapmap.org/; Haploview, http://www.broad.mit.edu/mpg/haploview/index.php available in the public domain, provided by the Massachusetts Institute of Technology, Cambridge, MA and FBAT, http://www.biostat.harvard.edu/∼fbat/fbat.htm available in the public domain, provided by Harvard Medical School, Boston, MA.

Funding

National Natural Science Foundation of China (30671798, 30471487); National Science and Technology Support Program of China (2006BAI02A11); National Major Basic Research Program of China (863:2006AA02A404); National Institute of Health (R03CA113240-01).

Acknowledgements

The authors would like to thank all patients and their families who participated in this study.

Conflict of Interest Statement: None declared.

References

1. Yu,M.C. *et al*. (2002) Epidemiology of nasopharyngeal carcinoma. *Semin. Cancer Biol.*, **12**, 421–429.
2. Young,L.S. *et al*. (2004) Epstein-Barr virus: 40 years on. *Nat. Rev. Cancer*, **4**, 757–768.
3. Zeng,Y.X. *et al*. (2002) Familial nasopharyngeal carcinoma. *Semin. Cancer Biol.*, **12**, 443–450.
4. Jia,W.H. *et al*. (2005) Complex segregation analysis of nasopharyngeal carcinoma in Guangdong, China: evidence for a multifactorial mode of inheritance (complex segregation analysis of NPC in China). *Eur. J. Hum. Genet.*, **13**, 248–252.
5. Raab-Traub,N. (2002) Epstein-Barr virus in the pathogenesis of NPC. *Semin. Cancer Biol.*, **12**, 431–441.
6. Yu,M.C. *et al*. (1987) Intake of Cantonese-style salted fish as a cause of nasopharyngeal carcinoma. *IARC Sci. Publ.*, **84**, 547–549.
7. Chow,W.H. *et al*. (1993) Tobacco use and nasopharyngeal carcinoma in a cohort of US veterans. *Int. J. Cancer*, **55**, 538–540.
8. Yu,M.C. *et al*. (1989) Induction of malignant nasal cavity tumours in Wistar rats fed Chinese salted fish. *Br. J. Cancer*, **60**, 198–201.
9. Ward,M.H. *et al*. (2000) Dietary exposure to nitrite and nitrosamines and risk of nasopharyngeal carcinoma in Taiwan. *Int. J. Cancer*, **86**, 603–609.
10. Yuan,J.M. *et al*. (2000) Non-dietary risk factors for nasopharyngeal carcinoma in Shanghai, China. *Int. J. Cancer*, **85**, 364–369.
11. IARC. (2004) Tobacco smoke and involuntary smoking. *IARC Monogr. Eval. Carcinog. Risks Hum.*, **83**, 1–1438.
12. Hildesheim,A. *et al*. (2001) Occupational exposure to wood, formaldehyde, and solvents and risk of nasopharyngeal carcinoma. *Cancer Epidemiol. Biomarkers Prev.*, **10**, 1145–1153.
13. Hildesheim,A. *et al*. (2002) Association of HLA class I and II alleles and extended haplotypes with nasopharyngeal carcinoma in Taiwan. *J. Natl Cancer Inst.*, **94**, 1780–1789.
14. Goldsmith,D.B. *et al*. (2002) HLA associations with nasopharyngeal carcinoma in Southern Chinese: a meta-analysis. *Clin. Otolaryngol. Allied Sci.*, **27**, 61–67.
15. Nazar-Stewart,V. *et al*. (1999) Glutathione S-transferase M1 and susceptibility to nasopharyngeal carcinoma. *Cancer Epidemiol. Biomarkers Prev.*, **8**, 547–551.
16. Hirunsatit,R. *et al*. (2003) Polymeric immunoglobulin receptor polymorphisms and risk of nasopharyngeal cancer. *BMC Genet.*, **4**, 3.
17. Zhou,X.X. *et al*. (2006) Sequence variants in toll-like receptor 10 are associated with nasopharyngeal carcinoma risk. *Cancer Epidemiol. Biomarkers Prev.*, **15**, 862–6.
18. Cho,E.Y. *et al*. (2003) Nasopharyngeal carcinoma and genetic polymorphisms of DNA repair enzymes XRCC1 and hOGG1. *Cancer Epidemiol. Biomarkers Prev.*, **12**, 1100–1104.
19. Cao,Y. *et al*. (2006) Polymorphisms of XRCC1 genes and risk of nasopharyngeal carcinoma in the Cantonese population. *BMC Cancer*, **6**, 167.
20. Duh,F.M. *et al*. (2004) Characterization of a new SNP c767A/T (Arg222Trp) in the candidate TSG FUS2 on human chromosome 3p21.3: prevalence in Asian populations and analysis of association with nasopharyngeal cancer. *Mol. Cell. Probes*, **18**, 39–44.
21. Ren,W. *et al*. (2005) A functional single nucleotide polymorphism site detected in nasopharyngeal carcinoma-associated transforming gene Tx. *Cancer Genet. Cytogenet.*, **157**, 49–52.
22. Jia,W.H. *et al*. (2004) Familial risk and clustering of nasopharyngeal carcinoma in Guangdong, China. *Cancer*, **101**, 363–369.
23. Feng,B.J. *et al*. (2002) Genome-wide scan for familial nasopharyngeal carcinoma reveals evidence of linkage to chromosome 4. *Nat. Genet.*, **31**, 395–399.
24. Chen,H.K. *et al*. (2003) The susceptibility gene for familial nasopharyngeal carcinoma is mapped on chromosome 4p11-p14 by haplotype analysis. *Chin. Sci. Bull.*, **48**, 2327–2330.
25. Xiong,W. *et al*. (2004) A susceptibility locus at chromosome 3p21 linked to familial nasopharyngeal carcinoma. *Cancer Res.*, **64**, 1972–1974.
26. Gonzalez,F.J. *et al*. (1994) Role of human cytochromes P450 in the metabolic activation of chemical carcinogens and toxins. *Drug Metab. Rev.*, **26**, 165–183.
27. Hildesheim,A. *et al*. (1995) Cytochrome P4502E1 genetic polymorphisms and risk of nasopharyngeal carcinoma: results from a case-control study conducted in Taiwan. *Cancer Epidemiol. Biomarkers Prev.*, **4**, 607–610.
28. Hildesheim,A. *et al*. (1997) CYP2E1 genetic polymorphisms and risk of nasopharyngeal carcinoma in Taiwan. *J. Natl Cancer Inst.*, **89**, 1207–1212.
29. Kongruttanachok,N. *et al*. (2001) Cytochrome P450 2E1 polymorphism and nasopharyngeal carcinoma development in Thailand: a correlative study. *BMC Cancer*, **1**, 4.

30. Barrett,J.C. *et al*. (2005) Haploview: analysis and visualization of LD and haplotype maps. *Bioinformatics*, **21**, 263–265.
31. Gabriel,S.B. *et al*. (2002) The structure of haplotype blocks in the human genome. *Science*, **296**, 2225–2229.
32. Niu,T. *et al*. (2002) Bayesian haplotype inference for multiple linked single-nucleotide polymorphisms. *Am. J. Hum. Genet.*, **70**, 157–169.
33. Laird,N.M. *et al*. (2000) Implementing a unified approach to family-based tests of association. *Genet. Epidemiol.*, **19** (suppl. 1), S36–S42.
34. Gauderman,W. *et al*. (2006) *QUANTO 1.1: A Computer Program for Power and Sample Size Calculations for Genetic-Epidemiology Studies*. http://hydra.usc.edu/gxe (22 October 2009, date last accessed).
35. Gauderman,W.J. (2002) Sample size calculations for matched case-control studies of gene-environment interaction. *Stat. Med.*, **21**, 35–50.
36. Wacholder,S. *et al*. (2004) Assessing the probability that a positive report is false: an approach for molecular epidemiology studies. *J. Natl Cancer Inst.*, **96**, 434–442.
37. Yu,M.C. *et al*. (1988) Preserved foods and nasopharyngeal carcinoma: a case-control study in Guangxi, China. *Cancer Res.*, **48**, 1954–1959.
38. Yu,M.C. (1991) Nasopharyngeal carcinoma: epidemiology and dietary factors. *IARC Sci. Publ.*, **105**, 39–47.
39. Ning,J.P. *et al*. (1990) Consumption of salted fish and other risk factors for nasopharyngeal carcinoma (NPC) in Tianjin, a low-risk region for NPC in the People's Republic of China. *J. Natl Cancer Inst.*, **82**, 291–296.
40. Yi,Z. *et al*. (1993) Urinary excretion of nitrosamino acids and nitrate by inhabitants of high- and low-risk areas for nasopharyngeal carcinoma in southern China. *Cancer Epidemiol. Biomarkers Prev.*, **2**, 195–200.
41. Ioannides,C. *et al*. (2004) Cytochromes P450 in the bioactivation of chemicals. *Curr. Top. Med. Chem.*, **4**, 1767–1788.
42. Raucy,J.L. *et al*. (1993) Bioactivation of halogenated hydrocarbons by cytochrome P4502E1. *Crit. Rev. Toxicol.*, **23**, 1–20.
43. Watanabe,J. *et al*. (1994) Different regulation and expression of the human CYP2E1 gene due to the RsaI polymorphism in the 5′-flanking region. *J. Biochem.*, **116**, 321–326.
44. Tsutsumi,M. *et al*. (1994) Hepatic messenger RNA contents of cytochrome P4502E1 in patients with different P4502E1 genotypes. *Alcohol Alcohol. Suppl.*, **29**, 29–32.
45. Hayashi,S. *et al*. (1991) Genetic polymorphisms in the 5′-flanking region change transcriptional regulation of the human cytochrome P450IIE1 gene. *J. Biochem.*, **110**, 559–565.
46. Spielman,R.S. *et al*. (1996) The TDT and other family-based tests for linkage disequilibrium and association. *Am. J. Hum. Genet.*, **59**, 983–989.
47. Rabinowitz,D. *et al*. (2000) A unified approach to adjusting association tests for population admixture with arbitrary pedigree structure and arbitrary missing marker information. *Hum. Hered.*, **50**, 211–223.
48. Dai,Y. *et al*. (1995) Cytotoxicity of acetaminophen in human cytochrome P4502E1-transfected HepG2 cells. *J. Pharmacol. Exp. Ther.*, **273**, 1497–1505.
49. Hou,D.F. *et al*. (2007) Expression of CYP2E1 in human nasopharynx and its metabolic effect *in vitro*. *Mol. Cell. Biochem.*, **298**, 93–100.
50. Jiang,J.H. *et al*. (2004) Expression of cytochrome P450 enzymes in human nasopharyngeal carcinoma and non-cancerous nasopharynx tissue. *Ai Zheng*, **23**, 672–677.

Received July 17, 2009; revised September 7, 2009;
accepted September 28, 2009