Association of human aryl hydrocarbon receptor gene polymorphisms with risk of lung cancer among cigarette smokers in a Chinese population

Dan Chen[a,*], Tian Tian[c,*], Haifeng Wang[b], Hongliang Liu[a], Zhibin Hu[c], Yi Wang[a], Yanhong Liu[a,d], Hongxia Ma[c], Weiwei Fan[a], Ruifen Miao[c], Weiwei Sun[b], Yi Wang[b], Ji Qian[a], Li Jin[a,b], Qingyi Wei[d], Hongbing Shen[c], Wei Huang[b] and Daru Lu[a]

Background and objective **Most of the carcinogenic effects of polycyclic aromatic hydrocarbons present in tobacco smoke are mediated by the aryl hydrocarbon receptor (AHR), a ligand-dependent transcription factor that regulates tobacco-induced expression of carcinogen metabolic enzymes. We hypothesized that genetic variations in *AHR* might confer individual susceptibility to lung cancer.**

Methods **Eight selected single-nucleotide polymorphisms in *AHR* were genotyped using the Illumina SNP genotyping BeadLab platform in a case–control study of 500 lung cancer patients and 517 cancer-free controls in a Chinese population.**

Results **We found that significantly increased lung cancer risk was associated with heterozygous genotypes of rs2158041 (adjusted odds ratio = 1.53 and 95% confidence interval = 1.17–1.99 for GA, compared with the GG genotype) and rs7811989 (adjusted odds ratio = 1.48 and 95% confidence interval = 1.13–1.93 for GA, compared with the GG genotype), although these two single-nucleotide polymorphisms were in linkage disequilibrium. Furthermore, haplotype analysis revealed significant differences in haplotype distributions of *AHR* between cases and controls (Global *P* = 1.38e-5). We also observed statistically significant interaction between the polymorphism rs2066853 (p.Arg554Lys) and cumulative cigarette smoking as a discrete or continuous variable (*P* = 0.033 and 0.019, respectively), and the Lys/Lys genotype conferred an increased risk of lung cancer in the heavy smokers (adjusted odds ratio = 3.36 and 95% confidence interval = 1.07–10.55).**

Conclusion **These findings suggest that *AHR* polymorphisms and potential gene–smoking interaction may be involved in the etiology of lung cancer. Further large prospective studies with ethnically diverse populations and functional studies are warranted to validate these findings. *Pharmacogenetics and Genomics* 19:25–34**

Pharmacogenetics and Genomics 2009, **19**:25–34

Keywords: *AHR*, genetic susceptibility, haplotype, lung cancer, molecular epidemiology, polymorphism, tobacco carcinogen

[a]State Key Laboratory of Genetic Engineering and Key Laboratory of Contemporary Anthropology, School of Life Sciences, Fudan University, [b]Rui Jin Hospital, School of Medicine, Shanghai Jiaotong University and Chinese National Human Genome Center, Shanghai, [c]Department of Epidemiology and Biostatistics, Cancer Research Center of Nanjing Medical University, Nanjing, China and [d]Department of Epidemiology, The University of Texas M.D. Anderson Cancer Center, Houston, Texas, USA

Correspondence to Daru Lu, State Key Laboratory of Genetic Engineering and Key Laboratory of Contemporary Anthropology, School of Life Sciences, Fudan University, Handan Road 220, Shanghai 200433, China
Tel/fax: +86 21 65642799; e-mail: drlu@fudan.edu.cn
or
Wei Huang, Rui Jin Hospital, School of Medicine, Shanghai Jiaotong University and Chinese National Human Genome Center, Shanghai 200025, China
Tel: +86 21 5080 1795; fax: +86 21 5080 1922;
e-mail: huangwei@chgc.sh.cn

*Dan Chen and Tian Tian have contributed equally to this study

Received 14 March 2008 **accepted** 28 August 2008

Introduction

Tobacco smoking is a primary etiologic factor for lung cancer, and about 10–20% of long-term smokers develop lung cancer in their lifetime [1,2]. It is well established that susceptibility to lung cancer may in part be attributable to interindividual variation in metabolic activation or detoxification of tobacco carcinogens [3–5], suggesting the importance of genetic determinants in lung cancer etiology, including the gene–environment interaction between genetic polymorphisms and environmental factors, such as smoking.

Exposure to polycyclic aromatic hydrocarbons (PAHs), a major class of potent tobacco procarcinogens, results in a wide range of toxic and carcinogenic responses in humans [6]. Most of the biological effects of PAHs are mediated by the aryl hydrocarbon receptor (AHR), a ubiquitously expressed ligand-activated transcriptional

1744-6872 © 2009 Wolters Kluwer Health | Lippincott Williams & Wilkins DOI: 10.1097/FPC.0b013e328316d8d8

factor, which upon ligand binding activates the transcription of genes coding multiple phase I and phase II enzymes [7]. Phase I enzymes, such as CYP1A1 and CYP1B1, and phase II enzymes, such as glutathione S-transferases, in turn, participate in the metabolism of PAHs [8–10]. Each of these enzymes is expressed in human lung and is inducible through the activation of the AHR [11]. Therefore, AHR plays a critical role in the regulation of carcinogen metabolism in the lung. Benzo[a] pyrene (BaP), one of the most important members of PAHs in cigarette smoke, is a major potent carcinogen implicated in the etiology of lung cancer through the induction of BaP diol epoxidation-DNA adducts through the AHR pathway [12,13]. There is direct evidence that BaP carcinogenicity is lost in the AHR-null mice [14]; besides, resveratrol, a natural competitive inhibitor of the AHR, could protect the lung from DNA damage and apoptosis caused by BaP, suggesting that the AHR is involved in the carcinogenesis initiated by BaP, especially in the lung [15]. In laboratory animals, genetic variations in *AHR* can dramatically alter its affinity for ligands or transcriptional regulation, leading to substantial differences in sensitivity to biochemical and toxic effects of PAHs and structurally related halogenated aromatic hydrocarbons, such as 2,3,7,8-tetrachlorodibenzo-*p*-dioxin [16–18].

Human lung cancer is a well-known example of environmentally induced carcinogenesis, which occurs through a complex multistage process as a result of the combination of carcinogen exposure and genetic susceptibility [19]. Human beings vary widely in their sensitivity to toxic effects of xenobiotic chemicals, and tobacco-induced expression of carcinogen metabolic enzymes in the lung also has a wide interindividual variation [20,21]. It has been known for several decades that phenotypic diversity exists among humans in the responses mediated by the AHR [22–24]. As the AHR is a central factor that regulates enzyme induction in response to tobacco carcinogens, we hypothesized that genetic variations in *AHR* may confer individual susceptibility to lung cancer. To test this hypothesis, we genotyped for eight *AHR* common single-nucleotide polymorphisms (SNPs) including one nonsynonymous SNP (p.Arg554Lys) in exon 10 and evaluated potential gene–smoking interaction in lung cancer risk in a case–control study of 500 incident lung cancer cases and 517 age and sex frequency matched cancer-free controls in a Chinese population.

Materials and methods

Study population

The study population and patients' characteristics have been described elsewhere [25]. In brief, this study was composed of 500 lung cancer patients and 517 cancer-free controls. All the participants were genetically unrelated ethnic Han Chinese and were from Nanjing City and surrounding regions in southeastern China. Patients newly diagnosed with histopathologically confirmed lung cancer were consecutively recruited between July 2002 and December 2004 at the Cancer Hospital of Jiangsu province (Nanjing) and the First Affiliated Hospital of Nanjing Medical University, Nanjing, China with a response rate of 90.5% without the restrictions of age, sex, and histology. The exclusion criteria included self-reported cancer history and previous radiotherapy and chemotherapy for unknown disease conditions. Cancer-free controls were randomly selected from 10 500 individuals who participated in a community-based screening program for noninfectious diseases conducted in Jiangsu province during the same period when the patients were recruited, with a response rate of 83.8%. The exclusion criteria for the control participants were the same as those for the case patients. All the control participants were frequency matched to the patients by age ($\pm$ 5 years), sex, and residential area (urban or rural areas). Each participant was scheduled for an interview after a written informed consent was obtained, and a structured questionnaire was administered by interviewers to collect information on demographic data and environmental exposure history including tobacco smoking. Smoking status and cumulative smoking dose were used to evaluate the smoking exposure. For the smoking status, those who had smoked less than one cigarette per day in less than 1 year in their lifetime were defined as nonsmokers; those smokers who had quit for longer than 1 year were considered as former smokers; otherwise they were considered as current smokers. In addition, pack-years smoked [(cigarettes per day/20) $\times$ years smoked] were calculated to indicate the cumulative smoking dose, which took one's total consumption of cigarette smoking into account. Family history of cancer was defined as any self-reported cancer in first-degree relatives (parents, siblings, or children). After the interview, an approximately 5-ml venous blood sample was collected from each participant. The study was approved by the institutional review boards of Fudan University.

Selection of single-nucleotide polymorphisms of aryl hydrocarbon receptor

The human *AHR* gene encompasses 47.146 kb and is located on chromosome 7p15. As this study was initiated in January 2005 when the SNP density of phase I HapMap SNP database was not adequate (14 SNPs were included, eight of which had a minor allele frequency $\geq$ 5%), we chose SNPs from both the HapMap and dbSNP databases. In the genetic model of *AHR* (contig accession NT_007819.16), it had 12 exons and 11 introns with 65 SNPs listed in the dbSNP Build 121. To capture the majority of common variants across this gene, an algorithm to score the SNPs was developed and a set of SNPs were selected based on their final scores. The detailed procedure has been described elsewhere [26]. In brief, for a SNP, the following criteria were considered: (i) interdistance between two adjacent SNPs; (ii) heterozygosity; (iii) functional relevance; (iv) compatibility

with the genotyping platform. As a result, eight SNPs (rs713150, rs6951212, rs3802083, rs1476080, rs2158041, rs7811989, rs2074113, and rs2066853) were chosen for genotyping in this study, of which one was a nonsynonymous variant rs2066853 (p.Arg554Lys) in exon 10 and seven were intronic variants.

Genotyping assays

We genotyped the SNPs of rs713150, rs6951212, rs3802083, rs1476080, rs2158041, rs7811989, and rs2074113 by using the Illumina SNP genotyping BeadLab platform (Illumina Corp, Foster City, California, USA) at the Chinese National Human Genome Centre and one additional SNP rs2066853 (p.Arg554Lys) located in exon 10 by using the fluorogenic 5′ nuclease Taqman assay (Applied Biosystems, Foster City, California, USA) at Fudan University. The information on assay conditions and the primers and probes is available upon request. For the Illumina high-throughput genotyping platform, more detailed description of quality control method and each of the steps are available in our previously published paper [25] and the HapMap website (*http://www.hapmap.org/downloads/protocols_overiview.html*). As 37 genomic DNA samples of case patients and 18 of control participants have run out in the former genotyping assay, there were 463 samples of patients and 499 samples of controls left for the Taqman assay of rs2066853. Two no-template controls and two duplicated samples in each 384-well format were used for quality control procedure. Four hundred and fifty (97.2%) of 463 patients and 489 (98.0%) of 499 controls were successfully genotyped in this locus. The consistency rate observed in these duplicated samples was 100%.

Statistical analyses

For each polymorphism, deviation of the genotype frequencies from those expected under Hardy–Weinberg equilibrium was assessed in the controls by a χ^2 test. Genotype frequencies in patients and controls were compared using a χ^2 test. Odds ratios (ORs) and 95% confidence intervals (CIs) were calculated by unconditional logistic regression analysis with adjustment for age, sex, pack-years of smoking, and family history of cancer. Stratification analyses were also performed by variables of interest, such as age, sex, smoking status (nonsmokers, former smokers, and current smokers), trichotomized pack-years of smoking [nonsmokers (0 pack-years), light smokers ($\leq$ 30 pack-years), and heavy smokers ($>$ 30 pack-years)], family history of cancer, and histological types.

The pairwise linkage disequilibrium (LD) among the SNPs was examined using Lewontin's standardized coefficient D' and LD coefficient r^2 [27], and haplotype blocks were defined by the method of Gabriel *et al.* [28] in the publicly available Haploview software (*http://www.broad.mit.edu/personal/jcbarret/haplo/*) [29] with default settings [the CI for a strong LD was minimal for (upper 0.98 and low 0.7) and maximal for a strong recombination of 0.9, and a fraction of strong LD in informative comparisons was at least 0.95]. The haplotypes were inferred from all genotype data, and haplotype (with a frequency of higher than 0.01) analyses were conducted by using Haplo.stats (*http://www.mayo.edu/stagen*), which is a score test based on the generalized linear model framework [30]. This program allows adjustment for possible confounding variables and provides both global and haplotype-specific tests for associations. To reduce the potential of spurious findings due to multiple testing, we conducted permutation testing to guide interpretation of nominally statistically significant associations in the single-locus, haplotype and stratification analyses. Case–control status was randomly permuted 100 000 times.

The potential gene–environment interaction between the SNPs and cigarette smoking was evaluated with the genotype–smoking joint effect and interaction models that considered cumulative smoking dose as both discrete [nonsmokers, light smokers ($\leq$ 30 pack-years), and heavy smokers ($>$ 30 pack-years)] and continuous (square root of pack-years) variables. The genotype–smoking interaction model included the following covariates: interaction term between genotypes and cumulative smoking dose (a discrete or continuous variable), genotypes, cumulative smoking dose (a discrete or continuous variable), age, sex, and family history. All the statistical analyses were performed with the SPSS15.0 software, with two-sided tests and a significant level of 0.05, unless indicated otherwise.

Results

Characteristics of the study population

The distributions of selected characteristics between lung cancer patients and controls have also been described elsewhere [25]. In brief, our frequency matching on age and sex was adequate ($P = 0.661$ for age and $P = 1.000$ for sex). The patients were, however, more likely to be smokers than the controls (77.0 vs. 51.8%), and more patients (35.4%) smoked more than 30 pack-years than did the controls (18.8%), and this difference was statistically significant ($P < 0.0001$). Furthermore, 24.4% of the lung cancer patients reported a family history of cancer in their first-degree relatives, which was significantly higher than that of the controls (16.8%), and this difference accounted for a significantly 60% increased lung cancer risk (OR = 1.60 and 95% CI = 1.17–2.17). Of the 500 lung cancer patients, 229 (45.8%) were adenocarcinoma, 141 (28.2%) squamous cell carcinoma, 34 (6.8%) small cell lung cancer, and 96 (19.2%) large cell, mixed cell carcinomas or undifferentiated carcinoma.

Individual single-nucleotide polymorphism association analysis

The SNP IDs, locations, and frequencies are given in Table 1 (information was updated according to the present dbSNP Build 129). All *AHR* genotype distributions in the controls were consistent with those expected from the Hardy–Weinberg equilibrium.

In the single-locus analyses, no significant differences in SNP allele distributions were found between patients and controls (Table 1). We, however, observed statistically significant differences between case patients and control participants in genotype distributions for two SNPs ($P = 0.0038$ for rs2158041 and $P = 0.008$ for rs7811989) and the significance remained after applying 100 000-time permutation tests ($P = 0.0193$ for rs2158041 and $P = 0.0407$ for rs7811989).

Further logistic regression analyses revealed that in the genotypic effect model with 2 d.f., significant risks were associated with the variant genotypes of rs2158041 G/A (adjusted OR = 1.53 and 95% CI = 1.17–1.99 for GA, $P = 0.002$, compared with the GG genotype) and rs7811989 G/A (adjusted OR = 1.48 and 95% CI = 1.13–1.93, $P = 0.004$ for GA, compared with the GG genotype) (Table 2).

Furthermore, we performed stratified analyses by selected variables and histological types on the SNPs of rs2158041G/A, rs7811989G/A, and rs2066853G/A. As shown in Table 3, compared with the wild-type genotype, the risk effects were almost similar in magnitude and direction to the point estimates obtained from fitted ORs of rs2158041 GA genotype among the subgroups stratified by age, sex, smoking status, pack-years of smoking, family history of cancer, and histological types, although the significance was mostly diminished in some subgroups because of reduced sample sizes. The risk effects of the rs7811989 GA genotype had the same trend as that of the rs2158041 GA genotype among all subgroups (data not shown). As for rs2066853, adjusted ORs of the AA genotype versus the GG genotype increased significantly as the number of pack-years increased, and the elevated risk was statistically significant among the heavy smokers (adjusted OR = 3.36 and 95% CI = 1.07–10.55, $P = 0.038$). This was, however, no longer significant after the multiple-testing correction ($P > 0.05$).

Haplotype block structure and haplotype analysis

Figure 1 shows plots of the pairwise LD (D') values for the eight SNPs and LD structure of *AHR*. The LD plot indicated that, seven SNPs (SNP 2–8) were in strong LD with each other (all $D' > 0.7$) and therefore formed a haplotype block, and the other SNP rs713150 was located in the upstream of the block. In addition, the following pairs of SNPs were in strong LD: rs6951212 and rs3802083 ($D' = 0.996$, $r^2 = 0.946$); rs6951212 and rs1476080 ($D' = 0.996$, $r^2 = 0.939$); rs7811989 and rs2158041 ($D' = 1.0$, $r^2 = 0.984$); and rs2074113 and rs2066863 ($D' = 0.990$, $r^2 = 0.958$) (supplementary Table 1).

Table 4 summarizes the associations between frequencies of the haplotypes and risk of lung cancer. Global score test showed statistically significant differences in the *AHR* haplotype distributions between case patients and control participants (Global $P = 1.38e\text{-}5$). After adjusting for confounding factors, the risk of lung cancer was significantly increased among individuals carrying the haplotype 'TAAGGAG' (adjusted OR = 12.85 and 95% CI = 2.98–55.53, $P = 0.0007$), compared with those carrying the most common haplotype 'CGCGGCG'. These results remained significant after the correction of 100 000-time permutation tests for multiple testing. Furthermore, one highly protective haplotype 'TAAGGCG' with a 60% reduction in risk of developing lung cancer was identified among the heavy smokers (adjusted OR = 0.40 and 95% CI = 0.18–0.89, $P = 0.025$).

Table 1 **Information about eight genotyped SNPs of the *AHR* gene**

Gene name (OMIM no.[a]) locus and build	No.	NCBI SNP ID	Genomic position[b]	Genic position	Base change	MAF			P[d]	P value for HWE[e]	Genotyping rate (%)
						In database[c]	Cases	Controls			
AHR	1	rs713150	16829115	Intron_1	C>G	0.381	0.355	0.341	0.520	0.21	100
OMIM: 600253	2	rs6951212	16833956	Intron_1	T>C	0.419	0.415	0.449	0.125	1.00	100
7p15	3	rs3802083	16837865	Intron_1	A>G	0.420	0.400	0.437	0.090	0.88	100
Build 129	4	rs1476080	16846831	Intron_2	A>C	0.400	0.396	0.435	0.073	1.00	100
	5	rs2158041	16857378	Intron_4	G>A	0.256	0.259	0.227	0.095	0.33	100
	6	rs7811989	16860321	Intron_6	G>A	0.256	0.257	0.224	0.085	0.51	100
	7	rs2074113	16862729	Intron_7	C>A	**0.333**	0.334	0.318	0.447	1.00	100
	8	rs2066853	16868068	Exon_10	G>A	0.367	0.334	0.323	0.601	0.77	92.33

[a]OMIM, Online Mendelian Inheritance in Man (*http://www.ncbi.nlm.nih.gov/sites/entrez?db=OMIM*).
[b]Single-nucleotide polymorphism (SNP) position in NCBI dbSNP database (*http://www.ncbi.nlm.nih.gov/SNP/snp_ref.cgi?locusId=196&chooseRs=all*).
[c]MAF, minor allele frequency, from both HapMap and dbSNP databases, the MAF in bold is from the dbSNP database.
[d]*P* value for difference in allele distributions between cases and controls.
[e]HWE, Hardy–Weinberg equilibrium, HWE *P* value in the control group.

Table 2 Genotype frequencies of selected SNPs of *AHR* among patients and controls and their associations with lung cancer risk

		Cases	Controls		Logistic regression	
SNP	Genotype	No.[a](%)	No.[a](%)	*P* (2 d.f.)[b]	OR (95% CI)[c]	*P*[d]
rs713150	CC	201 (40.2)	231 (44.7)	0.133	1.00 (Reference)	
	CG	243 (48.6)	219 (42.4)		1.28 (0.97–1.67)	0.079
	GG	56 (11.2)	67 (13.0)		0.99 (0.65–1.50)	0.960
rs6951212	TT	174 (34.8)	157 (30.4)	0.288	1.00 (Reference)	
	TC	237 (47.4)	256 (49.5)		0.84 (0.63–1.12)	0.240
	CC	89 (17.8)	104 (20.1)		0.75 (0.52–1.08)	0.118
rs3802083	AA	182 (36.4)	165 (31.9)	0.240	1.00 (Reference)	
	AG	236 (47.2)	252 (48.7)		0.85 (0.64–1.13)	0.266
	GG	82 (16.4)	100 (19.3)		0.70 (0.48–1.02)	0.061
rs1476080	AA	184 (36.8)	165 (31.9)	0.199	1.00 (Reference)	
	AC	236 (47.2)	254 (49.1)		0.84 (0.63–1.11)	0.216
	CC	80 (16.0)	98 (19.0)		0.70 (0.48–1.02)	0.062
rs2158041	GG	262 (52.4)	313 (60.5)	**0.0038**	1.00 (Reference)	
	GA	217 (43.4)	173 (33.5)		**1.53 (1.17–1.99)**	**0.002**
	AA	21 (4.2)	31 (6.0)		0.81 (0.45–1.46)	0.476
rs7811989	GG	264 (52.8)	314 (60.7)	**0.008**	1.00 (Reference)	
	GA	215 (43.0)	174 (33.7)		**1.48 (1.13–1.93)**	**0.004**
	AA	21 (4.2)	29 (5.6)		0.88 (0.48–1.61)	0.679
rs2074113	CC	219 (43.8)	240 (46.4)	0.703	1.00 (Reference)	
	CA	228 (45.6)	225 (43.5)		1.11 (0.85–1.45)	0.454
	AA	53 (10.6)	52 (10.1)		1.14 (0.74–1.76)	0.564
rs2066853	GG	197 (43.8)	222 (45.4)	0.870	1.00 (Reference)	
	GA	205 (45.6)	218 (44.6)		1.08 (0.82–1.42)	0.603
	AA	48 (10.7)	49 (10.0)		1.15 (0.73–1.81)	0.554

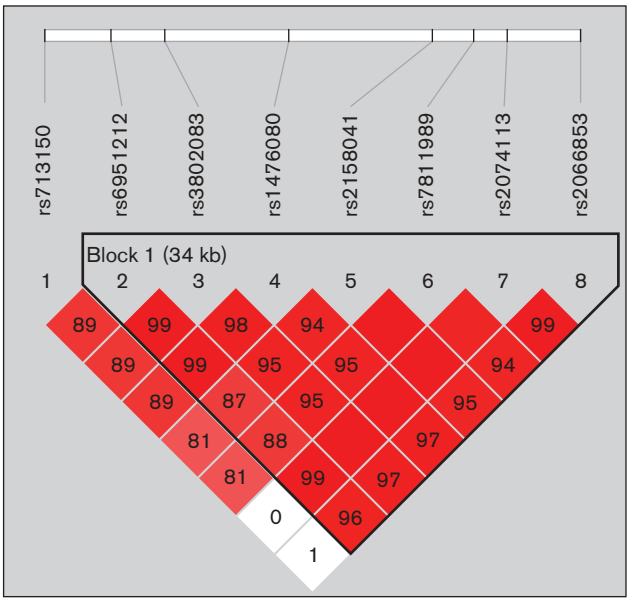
Bold characters indicate corresponding *P* values are less than 0.05.
CI, confidence interval; OR, odds ratio; SNP, single-nucleotide polymorphism.
[a]Numbers of cases/controls.
[b]Genotype frequencies in cases and controls were compared using two-sided χ^2 test with 2 d.f.
[c]Adjusted for age, sex, pack-years of smoking, and family history of cancer.
[d]*P* values from unconditional logistic regression analyses.

Table 3 Stratified analyses of associations between the *AHR* rs2158041 and rs2066853 genotypes and lung cancer risk by selected variables

	rs2158041			rs2066853		
Variables	Cases/controls (500/517)	OR (95% CI)[a]		Cases/controls (450/489)	OR (95% CI)[a]	
	No.[b]	GA vs. GG	AA vs. GG	No.[b]	GA vs. GG	AA vs. GG
Age (year)						
≤ 60	262/304	1.75 (1.22–2.50)	0.91 (0.43–1.92)	237/287	0.97 (0.67–1.41)	0.97 (0.53–1.77)
>60	238/213	1.28 (0.86–1.91)	0.66 (0.25–1.77)	213/202	1.23 (0.81–1.87)	1.44 (0.71–2.91)
Sex						
Male	386/400	1.44 (1.06–1.95)	0.69 (0.35–1.37)	349/380	1.06 (0.77–1.46)	1.30 (0.77–2.19)
Female	114/117	1.82 (1.05–3.14)	1.30 (0.39–4.35)	101/109	1.13 (0.63–2.00)	0.77 (0.30–1.99)
Smoking status[c]						
Nonsmokers	165/249	1.55 (1.02–2.35)	1.59 (0.58–4.37)	147/234	1.30 (0.83–2.02)	0.67 (0.31–1.45)
Former smokers	108/52	1.53 (0.76–3.07)	0.63 (0.12–3.31)	97/51	0.95 (0.47–1.95)	1.64 (0.48–5.63)
Current smokers	227/216	1.44 (0.97–2.15)	0.55 (0.25–1.24)	206/204	1.01 (0.67–1.53)	1.63 (0.81–3.26)
Pack-years of smoking[d]						
Nonsmokers	165/249	1.55 (1.02–2.35)	1.57 (0.57–4.32)	147/234	1.30 (0.83–2.02)	0.67 (0.31–1.44)
Light smokers	158/171	1.77 (1.11–2.81)	0.73 (0.30–1.82)	145/165	0.84 (0.52–1.36)	1.12 (0.52–2.39)
Heavy smokers	177/97	1.22 (0.72–2.07)	0.40 (0.12–1.32)	158/90	1.12 (0.65–1.92)	**3.36 (1.07–10.55)**[f]
Family history of cancer						
No	378/430	1.55 (1.15–2.09)	0.88 (0.46–1.68)	349/405	1.11 (0.81–1.52)	1.06 (0.64–1.74)
Yes	122/87	1.40 (0.77–2.54)	0.50 (0.11–2.26)	101/84	0.95 (0.51–1.77)	1.83 (0.57–5.90)
Histological types						
Adenocarcinoma	229/517	1.63 (1.17–2.28)	0.93 (0.45–1.94)	202/489	1.22 (0.86–1.73)	0.71 (0.36–1.39)
Squamous cell	141/517	1.46 (0.96–2.22)	0.60 (0.22–1.65)	130/489	0.93 (0.59–1.45)	1.48 (0.72–3.02)
Small cell	34/517	1.28 (0.62–2.64)	0.46 (0.06–3.62)	31/489	1.15 (0.53–2.49)	1.16 (0.31–4.31)
Other carcinomas[e]	96/517	1.47 (0.93–2.32)	0.77 (0.26–2.29)	87/489	0.88 (0.53–1.47)	2.20 (1.13–4.28)

Bold characters indicate corresponding *P* values are less than 0.05.
CI, confidence interval; OR, odds ratio.
[a]Adjusted for age, sex, pack-years of smoking, and family history of cancer, accordingly.
[b]Numbers of cases/controls.
[c]Those who had smoked less than one cigarette per day and in less than 1 year in their lifetime were defined as nonsmokers; those smokers who had quit for longer than 1 year were considered former smokers; otherwise they were considered current smokers.
[d]Light smokers (≤ 30 pack-years), and heavy smokers (>30 pack-years).
[e]Large cell, mixed cell carcinomas or undifferentiated carcinoma.
[f]*P*=0.038.

Fig. 1



Graphical representation of the SNP locations and LD structure of *AHR* using 8 genotyped SNPs in 517 Southeast Han Chinese controls. The exact SNP positions are listed in Table 1. One haplotype block (colored) was defined by the Haploview program using the approach given in Gabriel *et al.* [28] with default settings (confidence interval minima for strong LD: upper 0.98, low 0.7; upper confidence interval maximum for strong recombination, 0.9; fraction of strong LD in informative comparisons must be at least 0.95). The rs number (top; from left to right) corresponds to the SNP name and the numbers in squares are D' values ($|D'| \times 100$). The measure of LD (D') among all possible pairs of SNPs is shown graphically according to the shade of red, whereas white represents very low D' and dark red represents very high D'. *AHR*, aryl hydrocarbon receptor gene; LD, linkage disequilibrium; SNP, single-nucleotide polymorphism.

Table 4 Association between *AHR* common haplotypes and lung cancer risk in overall population and subpopulation stratified by pack-years of smoking

	Overall							Nonsmokers	Light smokers	Heavy smokers
	Frequencies				Logistic regression					
Haplotype[a]	Case	Control	*P*	*P-sim*[b]	OR (95% CI)[c]	*P*	Global score test	OR (95% CI)[b]	OR (95% CI)[b]	OR (95% CI)[b]
CGCGGCG	0.373	0.422	**0.031**	**0.031**	1.00 (Reference)	–	Global stat = 30.147	1.00 (Reference)	1.00 (Reference)	1.00 (Reference)
TAAGGAA	0.296	0.315	0.438	0.437	1.08 (0.87–1.35)	0.470	d.f. = 5, ***P = 1.38e-5***	1.03 (0.74–1.43)	0.99 (0.68–1.44)	1.35 (0.89–2.05)
TAAAACG	0.231	0.211	0.288	0.289	1.24 (0.97–1.58)	0.094	***P-sim*[b] = *1.00e-5***	1.42 (0.97–2.07)	1.19 (0.79–1.80)	1.04 (0.65–1.67)
TAAGGAG	0.031	0.002	**1.335e-5**	**0.000**	**12.85 (2.98–55.53)**	**0.0007**	**13.98 (4.71–41.49)**		6.74 (1.15–39.52)	**NA[d]**
TAAGGCG	0.011	0.016	0.287	0.297	0.87 (0.39–1.96)	0.740	**0.98 (0.48–2.03)**		0.74 (0.28–1.97)	**0.40 (0.18–0.89)[e]**

Bold characters indicate corresponding *P* values are less than 0.05.
CI, confidence interval; OR, odds ratio; SNP, single-nucleotide polymorphism.
[a]Polymorphic bases were in 5′-3′ order as listed in Table 1. Loci chosen for block: SNP 2–8.
[b]Generated by permutation test with 100 000 times simulation.
[c]Adjusted for age, sex, pack-years of smoking, and family history of cancer.
[d]NA, not available because of the rarity of haplotype.
[e]$P = 0.025$.

Gene–smoking interaction analysis

As shown in Table 5, the genotype–smoking joint effect analyses revealed that the adjusted ORs of the 'GA' versus 'GG' and 'AA' versus 'GG' genotypes increased significantly as the number of pack-years increased, suggesting that the effect of the rs2066853 appeared to be strongly modified by cumulative cigarette smoking. Notably, the magnitude and statistical significance for the 'AA' versus 'GG' genotype were larger and stronger, respectively, than that for the 'GA' versus 'GG 'comparison, suggesting a gene–dose relationship for the number of the variant A allele. A statistically significant multiplicative interaction between rs2066853 genotypes and cumulative smoking dose was observed when the genotype–smoking interaction model considered cumulative smoking dose as both discrete [nonsmokers, light smokers ($\leq$ 30 pack-years), and heavy smokers ($>$ 30 pack-years)] and continuous variables (square root of pack-years) with $P = 0.033$ and 0.019, respectively. However, no interaction between the genotypes and cigarette exposure was found for other polymorphisms.

Table 5 Interaction analyses of rs2066853 genotypes and cumulative smoking dose

Genotype	Genotype–smoking joint effects			Genotype–smoking interaction model	
	Nonsmokers OR (95% CI)[a]	Light smokers OR (95% CI)[a]	Heavy smokers OR (95% CI)[a]	P^b for interaction	P^c for interaction
GG (Arg/Arg)	1.00 (Reference)	2.33 (1.40–3.90)	3.69 (2.12–6.43)	**0.033**	**0.019**
GA (Arg/Lys)	1.30 (0.83–2.02)	1.97 (1.17–3.31)	4.13 (2.38–7.17)		
AA (Lys/Lys)	0.67 (0.31–1.44)	2.61 (1.19–5.73)	12.41 (3.94–39.08)		

[a]Nonsmokers carrying the wild-type genotype GG were used as reference group with adjustment for age, sex, and family history of cancer.
[b]The logistic regression model included the following covariates: interaction term between genotypes and three categories of pack-years (nonsmokers, light smokers, and heavy smokers), genotypes, three categories of pack-years, age, sex, and family history.
[c]The logistic regression model included the following covariates: interaction term between genotypes and square root of pack-years (as a continuous variable), genotypes, square root of pack-years, age, sex, and family history.

Discussion

In this lung cancer case–control study in a Chinese population, we found, for the first time, that the variant genotypes of two (rs2158041 and rs7811989) of eight selected SNPs in *AHR* were significantly associated with lung cancer risk, although these two SNPs were in strong LD. Furthermore, there was significant difference in the haplotype distributions of *AHR* between case patients and control participants. All the above positive associations derived from single-locus and haplotype analyses remained significant after correction by permutation tests. However, considering the ambiguity in haplotype inference and the missing data (especially rs2066853), the validity of information derived from the rare haplotypes warrants further studies. We also, for the first time, presented statistical evidence of significant interaction between the rs2066853 SNP and cumulative cigarette smoking in the risk of lung cancer. These findings support our hypothesis that *AHR* polymorphisms and the potential gene–smoking interaction may contribute to susceptibility to lung cancer. Larger studies with ethnically diverse populations are warranted to confirm our findings.

AHR is ubiquitously distributed in cells and tissues of humans and has a long evolutionary history [31,32]. The high degree of conservation in the primary structure of the human AHR implies that this protein has important biological functions [33]. In addition to regulating responses to environmental contaminants, a recent study provided further evidence that the AHR pathway may also play an important role in cigarette smoke-mediated cyclooxygenase-2 and prostaglandin production in human lung fibroblasts and thus may contribute to tobacco-associated inflammation and lung diseases, including lung cancer [34]. Moreover, AHR modulates cellular signaling pathways critical to cell-cycle regulation, differentiation, mitogen-activated protein kinase cascades, immediate-early gene induction, and functions of the RB protein [35]. It is important to note that AHR is associated with increased oxidative stress and tumor promotion independent of the CYP1 activity [36,37]. In the absence of exogenous AHR ligands (such as tobacco smoke components), AHR overexpression upregulated the expression of CYP1B1 in the early stage of lung adenocarcinoma, and elevated AHR expression in lung adenocarcinoma cells could enhance intracellular oxidative stress and promote cell growth [38]. Taken together, *AHR* is one of the strongest candidate genes for susceptibility to lung cancer.

The role of polymorphisms in the *AHR* gene in lung cancer risk has been examined in epidemiological studies in Japanese, French and Korean populations, in which three common SNPs (rs7796976 G/A in 5′-UTR, rs2074113 G/T in intron 7 and rs2066853 G/A in exon 10) were investigated but none of the genotypes was found to be associated with lung cancer risk [39–41]. In our study, we used a polymorphism selection strategy combing both hypotheses-driven (nsSNPs) and hypotheses-free (tagSNPs) approaches to identify the most representative SNPs in *AHR*. As, however, we chose SNPs from both the HapMap phase I and dbSNP databases, in which the specific LD among most SNPs of the *AHR* gene was not available at that time, we could only try to minimize the strong LD of any pair of candidate SNPs based on their intermarker distance and heterozygosity, which may be potentially redundant. In the CHB (45 Han Chinese in Bejing, northern China) genotyped data of *AHR*, released in phase II of the International HapMap Project (HapMap Data Rel 22/phase II Apr 07), 17 SNPs (rs7796976 and rs2074113 not included) of *AHR* have a minor allele frequency $\geq 5\%$. As shown in supplementary Fig. 1, 15 of the 17 SNPs are in LD, which comprise one LD block. Five SNPs in this block, two SNPs (rs713150 and rs6951212) in preblock region, and one reported SNP rs2074113 that was not included in phase II were selected in our study. The seven SNPs (rs2074113 excluded) were able to capture 15 of these 17 informative alleles with a minimal LD variable r^2 threshold of 0.7, and 14 (93%) of captured alleles have pairwise LD with $r^2 > 0.8$. Therefore, the genetic variants investigated in our study captured most of common variants in the gene in the Chinese population. However, one reported SNP rs7796976 was not included in this study. As there is no information regarding allele frequencies of this SNP in Chinese population up to now, we do not know whether or not

it is in LD with any of our investigated SNPs. To the best of our knowledge, except for SNPs rs2074113 and rs2066853, none of the other selected SNPs in our study has been previously investigated in cancer susceptibility.

Consistent with the earlier studies, no overall effects for rs2074113 and rs2066853 on lung cancer risk were observed in this study. However, two other SNPs (rs2158041 and rs7811989 in nearly complete LD), both residing in the intronic regions, were found to be significantly associated with increased risk of lung cancer. Accumulating evidence has indicated the importance of intronic polymorphisms as markers for cancer susceptibility. For example, intronic SNPs in *p*53 [42] and ataxia telangiectasia mutated [43] were associated with increased risk of lung cancer and breast cancer, respectively. These intronic SNPs could potentially produce alternative splicing of their gene products or they are in tight linkage with other causal loci or genes. However, the exact location and biological functions of the real causal SNPs in *AHR* is of great interest and warrant further investigation.

Interestingly, for some SNPs, we found that the risk effect seemed to be significant in the heterozygotes but not in the homozygotes. It is not uncommon that heterozygous genotypes seem to confer the greatest risk as reported in molecular epidemiological association studies with genotyping data in the literature. Although current knowledge has not provided a convincing explanation for such a finding, there are several possibilities that may explain these seemingly peculiar observations. Technically, such a bias toward the heterozygotes could result from a systematic genotyping error, but this is unlikely to occur in the Illumina assay that has a stringent quality control. It has been postulated that the heterozygous genotype may be in LD with other susceptibility loci, or some of homozygotes in combination with other adverse genotypes may have been embryogenically lethal or heterozygotes may have impaired functions because of the potential imbalance of the resultant protein structure [44]. These hypotheses, however, need to be further tested in in-depth molecular mechanistic studies in the future. Alternatively, because of the relatively small numbers of the variant homozygotes observed in both patients and controls, the effect of the variant homozygotes might more likely be affected by any selection bias inherent in such hospital-based retrospective study than the heterozygotes which are often present in a much larger number of observations, or simply there is not enough statistical power to detect any real effect among the variant homozygotes [45]. These can only be overcome by performing additional mechanistic studies and much larger association studies in the future.

The interplay between SNPs of the *AHR* gene and cigarette smoking with respect to lung cancer risk has not been investigated in the earlier studies. With the advantage of detailed measures of cigarette smoking, our results indicated that the effect of the p.Arg554Lys (rs2066853) seemed to be strongly modified by cigarette smoking and that the Lys/Lys genotype conferred a statistically significantly high risk of lung cancer in the heavy smokers who smoked more than 30 pack-years. As the potential interaction between rs2066853 and smoking was, however, gradually developed by a list of testing procedures, future replication studies in other ethnic populations should be performed to substantiate this observation. As indicated in the literature [46], a possible definition of gene–environment interaction is that it occurs when a genetic and environmental exposure collectively cause a disease outcome in some or all patients. As this association was restricted to the heavy smokers, a role for the p.Arg554Lys in such a subgroup inferred a causal association between PAHs and lung cancer. Several lines of evidence may support this finding. The arginine to lysine substitution at codon 554 exists in D/E-rich transactivation domain of AHR, which may permit finely tuned modulation in downstream gene expression including the CYP1A1 induction triggered by smoking [33]. CYP1A1 and its correlated catalytic activities are enhanced in the lungs of most cigarette smokers, presumably resulting from concentrated exposure of the lung to PAHs and other AHR ligands in smoke [33]. Furthermore, more than a dozen laboratories independently found correlations between the high CYP1A1 inducibility phenotype and lung cancer in cigarette smokers [8]. Therefore, the variant of rs2066853 SNP may be heavily dependant on the cumulative dose of the cigarette smoking for its carcinogenic potential, that is, there is no increased risk for the variant genotype in the absence of high dose of cigarette smoking. Smart and Daly [47] reported that cultured lymphocytes from individuals carrying at least one Lys allele had significantly elevated CYP1A1 activity and expression in a Caucasian population. Contrary to the observation of Smart and Daly, several studies, however, did not find any evidence for altered CYP1A1 inducibility associated with p.Arg554Lys [39,40,48,49]. This discrepancy may be due to inclusion of different ethnic groups and the limitation of the methods for assessing the effect of the *AHR* genotype on the induction phenotype in humans. Therefore, the extent of influence of p.Arg554Lys on the gene functions remains to be elucidated in further studies.

Recently, three genome-wide association studies simultaneously identified an association between SNPs of a nicotinic receptor gene locus at 15q24/15q25.1 and risk of lung cancer [50–52]. These studies, however, did not identify an association between the *AHR* locus and lung cancer risk due to several reasons. First, this may be due to ethnic difference, because only European populations were included in these studies. Second, because of the

stringent threshold of genome-wide significance, to some extent, some loci may be missed because the causal variants have smaller overall effects or they are in low LD with SNPs on the genotyping arrays [53]. Further genome-wide scans in populations with detailed data on smoking exposure and intensity of smoking with diverse ethnic groups [54] are needed to unravel possible association between the *AHR* locus and lung cancer risk or related gene–environment interaction.

Several limitations are inherent in this type of case–control study and must be noted. First of all, this study was a hospital-based case–control study with patients recruited from hospitals and controls selected from the community, and thus selection bias may lead to spurious findings. Potential confounding bias was, however, minimized by frequency-matching of the controls to the patients by age, sex, and area and further adjustment for potential confounding factors in the analyses. To reduce potential spurious findings because of multiple testing, we conducted permutation testing to guide interpretation of nominally statistically significant associations. Second, the sample size of this study was relatively small, which may produce false-positive or false-negative results, whereas the genotype frequency was low or stratified analyses were performed. Therefore, further prospective studies will be needed in a larger and ethnically diverse population to validate the findings. Finally, our study lacked information on other exposures such as dietary consumption, occupational and environmental exposure, which might act as potential confounding factors. Further studies are needed to address whether these factors also modify genetic susceptibility to lung cancer.

In conclusion, in this case–control study in a Chinese population, we provided evidence that some common genetic variants of *AHR* may modulate risk of lung cancer. In particular, our results support the notion that some genetic effect that may not be detectable in a single-locus analysis can be unraveled when gene–environment interaction is considered. However, large prospective studies with ethnically diverse populations and functional evaluation are warranted to further elucidate the impact of genetic variation in *AHR* on lung cancer susceptibility.

Supplementary data

Supplementary data available online at the *Journal Pharmacogenetics and Genomics* (*www.pharmacogenetics and genomics.com*).

Acknowledgements

The authors thank Shuhua Xu, Zhengwen Jiang, Rui Li, and Wenqing Fu for their helpful comments and discussion. This study was supported by the National Key Basic Research Program of China (2002CB512902), High-Tech Program of China (2006AA020706), National Outstanding Youth Science Foundation of China (30425001 and 30625019), Shanghai Science and Technology Committee (06DZ19501, 06XD14015 and 05DZ22201), and Shanghai Leading Academic Discipline Project (B111).

References

1 Shields PG. Molecular epidemiology of smoking and lung cancer. *Oncogene* 2002; **21**:6870–6876.
2 Tsou JA, Hagen JA, Carpenter CL, Laird-Offringa IA. DNA methylation analysis: a powerful new tool for lung cancer diagnosis. *Oncogene* 2002; **21**:5450–5461.
3 Nebert DW, McKinnon RA, Puga A. Human drug-metabolizing enzyme polymorphisms: effects on risk of toxicity and cancer. *DNA Cell Biol* 1996; **15**:273–280.
4 Taningher M, Malacarne D, Izzotti A, Ugolini D, Parodi S. Drug metabolism polymorphisms as modulators of cancer susceptibility. *Mutat Res* 1999; **436**:227–261.
5 Bartsch H, Nair U, Risch A, Rojas M, Wikman H, Alexandrov K. Genetic polymorphism of CYP genes, alone or in combination, as a risk modifier of tobacco-related cancers. *Cancer Epidemiol Biomarkers Prev* 2000; **9**:3–28.
6 Hecht SS. Tobacco smoke carcinogens and lung cancer. *J Natl Cancer Inst* 1999; **91**:1194–1210.
7 Rowland JC, Gustaffson JA. Aryl hydrocarbon receptor-mediate signal transduction. *Crit Rev Toxicol* 1997; **27**:109–134.
8 Nebert DW, Dalton TP, Okey AB, Gonzalez FJ. Role of aryl hydrocarbon receptor-mediated induction of the CYP1 enzymes in environmental toxicity and cancer. *J Biol Chem* 2004; **279**:23847–23850.
9 Uno S, Dalton TP, Dragin N, Curran CP, Derkenne S, Miller ML, *et al.* Oral benzo[a]pyrene in Cyp1 knockout mouse lines: CYP1A1 important in detoxication, CYP1B1 metabolism required for immune damage independent of total-body burden and clearance rate. *Mol Pharmacol* 2006; **69**:1103–1114.
10 Hayes JD, Pulford DJ. The glutathione S-transferase supergene family: regulation of GST and the contribution of the isoenzymes to cancer chemoprotection and drug resistance. *Crit Rev Biochem Mol Biol* 1995; **30**:445–600.
11 Schrenk D. Impact of dioxin-type induction of drug-metabolizing enzymes on the metabolism of endo- and xenobiotics. *Biochem Pharmacol* 1998; **55**:1155–1162.
12 Phillips DH. Fifty years of benzo(a)pyrene. *Nature* 1983; **303**:468–472.
13 Stansbury KH, Flesher JW, Gupta RC. Mechanism of aralkyl-DNA adduct formation from benzo[a]pyrene in vivo. *Chem Res Toxicol* 1994; **7**:254–259.
14 Shimizu Y, Nakatsuru Y, Ichinose M, Takahashi Y, Kume H, Mimura J, *et al.* Benzo[a]pyrene carcinogenicity is lost in mice lacking the aryl hydrocarbon receptor. *Proc Natl Acad Sci U S A* 2000; **97**:779–782.
15 Revel A, Raanani H, Younglai E, Xu J, Rogers I, Han R, *et al.* Resveratrol, a natural aryl hydrocarbon receptor antagonist, protects lung from DNA damage and apoptosis caused by benzo[a]pyrene. *J Appl Toxicol* 2003; **23**:255–261.
16 Okey AB, Vella LM, Harper PA. Detection and characterization of a low affinity form of cytosolic Ah receptor in livers of mice nonresponsive to induction of cytochrome P1-450 by 3-methylcholanthrene. *Mol Pharmacol* 1989; **35**:823–830.
17 Nebert DW. The Ah locus: genetic differences in toxicity, cancer, mutation, and birth defects. *Crit Rev Toxicol* 1989; **20**:153–174.
18 Poland A, Palen D, Glover E. Analysis of the four alleles of the murine aryl hydrocarbon receptor. *Mol Pharmacol* 1994; **46**:915–921.
19 Shields PG. Molecular epidemiology of lung cancer. *Ann Oncol* 1999; **10 (Suppl 5)**:S7–S11.
20 Harris CC, Autrup H, Connor RD, Barrett LA, McDowell EM, Trump BF. Interindividual variation in binding of benzo[a]pyrene to DNA in cultured human bronchi. *Science* 1976; **194**:1067–1069.
21 Willey JC, Coy EL, Frampton MW, Torres A, Apostolakos MJ, Hoehn G, *et al.* Quantitative RT-PCR measurement of cytochromes P450 1A1, 1B1, and 2B7, microsomal epoxide hydrolase, and NADPH oxidoreductase expression in lung cells of smokers and nonsmokers. *Am J Respir Cell Mol Biol* 1997; **17**:114–124.

22 Kellermann G, Cantrell E, Shaw CR. Variations in extent of aryl hydrocarbon hydroxylase induction in cultured human lymphocytes. *Cancer Res* 1973; **33**:1654–1656.

23 Nebert DW, Jensen NM. The Ah locus: genetic regulation of the metabolism of carcinogens, drugs, and other environmental chemicals by cytochrome P-450-mediated monooxygenases. *CRC Crit Rev Biochem* 1979; **6**:401–437.

24 Kouri RE, McKinney CE, Slomiany DJ, Snodgrass DR, Wray NP, McLemore TL. Positive correlation between high aryl hydrocarbon hydroxylase activity and primary lung cancer as analyzed in cryopreserved lymphocytes. *Cancer Res* 1982; **42**:5030–5037.

25 Hu Z, Wang H, Shao M, Jin G, Sun W, Wang Y, *et al*. Genetic variants in MGMT and risk of lung cancer in Southeastern Chinese: a haplotype-based analysis. *Hum Mutat* 2007; **28**:431–440.

26 Chen D, Jin G, Wang Y, Wang H, Liu H, Liu Y, *et al*. Genetic variants in peroxisome proliferator-activated receptor-gamma gene are associated with risk of lung cancer in a Chinese population. *Carcinogenesis* 2008; **29**:342–350.

27 Lewontin RC. On measures of gametic disequilibrium. *Genetics* 1988; **120**:849–852.

28 Gabriel SB, Schaffner SF, Nguyen H, Moore JM, Roy J, Blumenstiel B, *et al*. The structure of haplotype blocks in the human genome. *Science* 2002; **296**:2225–2229.

29 Barrett JC, Fry B, Maller J, Daly MJ. Haploview: analysis and visualization of LD and haplotype maps. *Bioinformatics* 2005; **21**:263–265.

30 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**:425–434.

31 Hahn ME. The aryl hydrocarbon receptor: a comparative perspective. *Comp Biochem Physiol C Pharmacol Toxicol Endocrinol* 1998; **121**:23–53.

32 Karchner SI, Franks DG, Powell WH, Hahn ME. Regulatory interactions among three members of the vertebrate aryl hydrocarbon receptor family: AHR repressor, AHR1, and AHR2. *J Biol Chem* 2002; **277**:6949–6959.

33 Harper PA, Wong JY, Lam MS, Okey AB. Polymorphisms in the human AH receptor. *Chem Biol Interact* 2002; **141**:161–187.

34 Martey CA, Baglole CJ, Gasiewicz TA, Sime PJ, Phipps RP. The aryl hydrocarbon receptor is a regulator of cigarette smoke induction of the cyclooxygenase and prostaglandin pathways in human lung fibroblasts. *Am J Physiol Lung Cell Mol Physiol* 2005; **289**:L391–L399.

35 Marlowe JL, Puga A. Aryl hydrocarbon receptor, cell cycle regulation, toxicity, and tumorigenesis. *J Cell Biochem* 2005; **96**:1174–1184.

36 Senft AP, Dalton TP, Nebert DW, Genter MB, Puga A, Hutchinson RJ, *et al*. Mitochondrial reactive oxygen production is dependent on the aromatic hydrocarbon receptor. *Free Radic Biol Med* 2002; **33**:1268–1278.

37 Poland A, Palen D, Glover E. Tumour promotion by TCDD in skin of HRS/J hairless mice. *Nature* 1982; **300**:271–273.

38 Chang JT, Chang H, Chen PH, Lin SL, Lin P. Requirement of aryl hydrocarbon receptor overexpression for CYP1B1 up-regulation and cell growth in human lung adenocarcinomas. *Clin Cancer Res* 2007; **13**:38–45.

39 Kawajiri K, Watanabe J, Eguchi H, Nakachi K, Kiyohara C, Hayashi S. Polymorphisms of human Ah receptor gene are not involved in lung cancer. *Pharmacogenetics* 1995; **5**:151–158.

40 Cauchi S, Stucker I, Solas C, Laurent-Puig P, Cenee S, Hemon D, *et al*. Polymorphisms of human aryl hydrocarbon receptor (AhR) gene in a French population: relationship with CYP1A1 inducibility and lung cancer. *Carcinogenesis* 2001; **22**:1819–1824.

41 Kim JH, Kim H, Lee KY, Kang JW, Lee KH, Park SY, *et al*. Aryl hydrocarbon receptor gene polymorphisms affect lung cancer risk. *Lung Cancer* 2007; **56**:9–15.

42 Wu X, Zhao H, Amos CI, Shete S, Makan N, Hong W, *et al*. p53 Genotypes and haplotypes associated with lung cancer susceptibility and ethnicity. *J Natl Cancer Inst* 2002; **94**:681–690.

43 Lee KM, Choi JY, Park SK, Chung HW, Ahn B, Yoo KY, *et al*. Genetic polymorphisms of ataxia telangiectasia mutated and breast cancer risk. *Cancer Epidemiol Biomarkers Prev* 2005; **14**:821–825.

44 Hu Z, Shao M, Yuan J, Xu L, Wang F, Wang Y, *et al*. Polymorphisms in DNA damage binding protein 2 (DDB2) and susceptibility of primary lung cancer in the Chinese: a case–control study. *Carcinogenesis* 2006; **27**:1475–1480.

45 Ma H, Xu L, Yuan J, Shao M, Hu Z, Wang F, *et al*. Tagging single nucleotide polymorphisms in excision repair cross-complementing group 1 (ERCC1) and risk of primary lung cancer in a Chinese population. *Pharmacogenet Genomics* 2007; **17**:417–423.

46 Brennan P. Gene-environment interaction and etiology of cancer: what does it mean and how can we measure it? *Carcinogenesis* 2002; **23**:381–387.

47 Smart J, Daly AK. Variation in induced CYP1A1 levels: relationship to CYP1A1, Ah receptor and GSTM1 polymorphisms. *Pharmacogenetics* 2000; **10**:11–24.

48 Smith GB, Harper PA, Wong JM, Lam MS, Reid KR, Petsikas D, *et al*. Human lung microsomal cytochrome P4501A1 (CYP1A1) activities: impact of smoking status and CYP1A1, aryl hydrocarbon receptor, and glutathione S-transferase M1 genetic polymorphisms. *Cancer Epidemiol Biomarkers Prev* 2001; **10**:839–853.

49 Anttila S, Lei XD, Elovaara E, Karjalainen A, Sun W, Vainio H, *et al*. An uncommon phenotype of poor inducibility of CYP1A1 in human lung is not ascribable to polymorphisms in the AHR, ARNT, or CYP1A1 genes. *Pharmacogenetics* 2000; **10**:741–751.

50 Hung RJ, McKay JD, Gaborieau V, Boffetta P, Hashibe M, Zaridze D, *et al*. A susceptibility locus for lung cancer maps to nicotinic acetylcholine receptor subunit genes on 15q25. *Nature* 2008; **452**:633–637.

51 Thorgeirsson TE, Geller F, Sulem P, Rafnar T, Wiste A, Magnusson KP, *et al*. A variant associated with nicotine dependence, lung cancer and peripheral arterial disease. *Nature* 2008; **452**:638–642.

52 Amos CI, Wu X, Broderick P, Gorlov IP, Gu J, Eisen T, *et al*. Genome-wide association scan of tag SNPs identifies a susceptibility locus for lung cancer at 15q25.1. *Nat Genet* 2008; **40**:616–622.

53 Altshuler D, Daly M. Guilt beyond a reasonable doubt. *Nat Genet* 2007; **39**:813–815.

54 Chanock SJ, Hunter DJ. Genomics: when the smoke clears. *Nature* 2008; **452**:537–538.