Genetic Variants on Chromosome 15q25 Associated with Lung Cancer Risk in Chinese Populations

Chen Wu,[1] Zhibin Hu,[2] Dianke Yu,[1] Liming Huang,[1] Guangfu Jin,[2] Jie Liang,[2] Huan Guo,[3] Wen Tan,[1] Mingfeng Zhang,[2] Ji Qian,[4] Daru Lu,[4] Tangchun Wu,[3] Dongxin Lin,[1] and Hongbing Shen[2]

[1]Department of Etiology and Carcinogenesis and State Key Laboratory of Molecular Oncology, Cancer Institute and Hospital, Chinese Academy of Medical Sciences and Peking Union Medical College, Beijing, China; [2]Department of Epidemiology and Biostatistics, Cancer Center, Nanjing Medical University, Nanjing, China; [3]Ministry of Education Key Laboratory for Environment and Health, School of Public Health, Tongji Medical College, Huazhong University of Science and Technology, Wuhan, China; and [4]State Key Laboratory of Genetic Engineering, School of Life Sciences, Fudan University, Shanghai, China

Abstract

Recent three genome-wide association studies have mapped a lung cancer susceptibility locus to chromosome 15q25 in Caucasians. However, the reported risk single nucleotide polymorphisms (SNPs) are extremely rare in Asians, arguing against any of these being causative variants. This study sought to identify other variants on 15q25 associated with lung cancer susceptibility in Chinese. Two-stage case-control studies were conducted in subjects derived from both Northern and Southern China. The first-stage, consisting of 576 cases and 576 controls, was to discover novel risk variants using a haplotype-tagging SNP approach, and these variants were then replicated in the second-stage, consisting of 2,989 cases and 2,880 controls. Associations were estimated by logistic regression models, and function of the variants was examined by biochemical assays. We found that the three risk SNPs reported in Caucasians were not associated with lung cancer risk in Chinese. However, we identified four novel SNPs (rs2036534C>T, rs667282C>T, rs12910984G>A, and rs6495309T>C) that were associated with significantly increased lung cancer risk and smoking behavior, which were all confirmed in the replication analyses [odds ratios (95% confidence intervals) in the dominant model: 1.39 (1.23–1.57; $P = 2.3 \times 10^{-7}$), 1.52 (1.35–1.71; $P = 2.0 \times 10^{-12}$), 1.44 (1.28–1.63; $P = 2.7 \times 10^{-9}$), and 1.43 (1.27–1.61; $P = 2.6 \times 10^{-9}$), respectively]. We characterized the rs6495309T>C change in the *CHRNA3* promoter as a functional variant because it affected the Oct-1 binding ability, resulting in increased *CHRNA3* expression. These results support 15q25 as a susceptibility region for lung cancer in Chinese but underscore the difference in genetic markers among different ethnic populations. [Cancer Res 2009;69(12):5065–72]

Introduction

Lung cancer is the leading cause of cancer-related deaths all over the world. In many countries, including China, the morbidity and mortality of lung cancer have been increasing rapidly in recent years, which is believed to be mainly due to continuous increase in tobacco consumption (1, 2). Although over 80% of lung cancer are tobacco related (3), accumulating evidence supports that genetic factors are also risk factors for lung cancer because only a small fraction of smokers (usually <20%) develop this disease. Efforts have been made to discover the potential biomarkers (4–6), but lung cancer still lacks specific biomarkers for risk assessment, early detection, and prognosis.

Recently, three genome-wide association studies (GWAS) in Caucasian populations have mapped a lung cancer susceptibility locus to chromosome 15q25 containing nicotinic acetylcholine receptor genes *CHRNA3, CHRNA5*, and *CHRNB4* (7–9). Nicotinic acetylcholine receptors expressed in the key regions of the brain play an important role in controlling smoking behavior (10, 11). These receptors are also expressed in epithelial cells of the lung, where they execute signal transduction upon binding to nicotine and/or its carcinogenic derivatives [e.g., 4-(methylnitrosamino)-1-(3-pyridyl)-1-butanone, NNK], resulting in cell proliferation and neoplastic transformation (12–15). Therefore, it seems plausible that genetic variations, such as single nucleotide polymorphisms (SNP) in nicotinic acetylcholine receptors that affect the gene expression or protein function, would be associated with smoking behavior and the risk of smoking-related lung cancer.

The exciting results in the GWAS encouraged us to investigate the association between the SNPs reported in Caucasians (7–9) and lung cancer risk in Chinese populations. By examining the HapMap data, however, we found that the risk SNPs, rs8034191, rs1051730, and rs16969968 identified in previous GWAS are extremely rare in Asians, suggesting that the role these SNPs play in risk of lung cancer in Asian populations, if any, may not be as important as they do in Caucasians. We thus hypothesized that if chromosome 15q25 also plays a role in susceptibility to lung cancer in Chinese populations, there must be other risk variants. To test this hypothesis, we conducted a two-stage case-control studies of 3,565 cases with lung cancer and 3,456 controls derived from both Northern and Southern China. We discovered four novel SNPs associated with lung cancer risk in Chinese populations, which are different from those reported in the GWAS in Caucasian populations. In addition, we characterized a SNP in the promoter region of *CHRNA3* as a potential causative SNP because it affected the transcriptional factor Oct-1 binding ability, resulting in alteration of *CHRNA3* expression.

Note: Supplementary data for this article are available at Cancer Research Online (http://cancerres.aacrjournals.org/).

C. Wu and Z. Hu contributed equally to this work.

Requests for reprints: Hongbing Shen, Department of Epidemiology and Biostatistics, Cancer Center, Nanjing Medical University, Nanjing 210029, China. E-mail: hbshen@njmu.edu.cn or Dongxin Lin, Departments of Etiology and Carcinogenesis, Cancer Institute and Hospital, Chinese Academy of Medical Sciences, Beijing 100021, China. E-mail: lindx72@cicams.ac.cn.

doi:10.1158/0008-5472.CAN-09-0081

Table 1. Associations between candidate SNPs in 15q25 and risk of lung cancer in a Chinese population

#	Identity	Chromosome position	Base change	MAF	Genotype (576 cases and 576 controls)				*P*
					Common*	Heterozygous*	Rare*	OR† (95% CI)	
1	rs7181486	76528673	C>T	0.147	77.7/72.6	21.1/25.3	1.2/2.1	0.77 (0.60–0.99)	0.036
2	rs2656065	76537604	C>T	0.257	55.9/54.7	39.4/39.4	4.7/5.9	0.94 (0.77–1.14)	0.507
3	rs9788721	76589924	T>C	0.289	42.7/48.5	48.2/45.1	9.1/6.4	1.22 (1.02–1.47)	0.029
4	rs7164594	76590112	T>C	0.446	23.3/30.5	52.4/49.8	24.3/19.7	1.27 (1.07–1.50)	0.005
5	rs10519203	76601101	A>G	0.112	81.3/78.3	17.8/21.0	0.9/0.7	0.86 (0.65–1.13)	0.259
6	rs931794	76613235	A>G	0.308	44.2/47.9	41.5/42.6	14.3/9.5	1.21 (1.01–1.45)	0.032
7	rs2036534	76614003	T>C	0.461	36.2/30.3	48.7/47.0	15.1/22.7	0.76 (0.64–0.90)	0.001
8	rs3813571	76619847	C>A	0.177	67.0/67.2	30.4/30.3	2.6/2.5	1.01 (0.81–1.26)	0.929
9	rs12901682	76620278	C>A	0.131	72.2/75.0	26.7/23.7	1.1/1.3	1.12 (0.87–1.43)	0.367
10	rs8024878	76630466	C>G	0.175	72.8/68.8	24.7/27.6	2.5/3.6	0.83 (0.66–1.05)	0.102
11	rs880395	76631411	G>A	0.158	64.9/70.9	32.0/26.8	3.1/2.3	1.27 (1.02–1.59)	0.030
12	rs684513	76645455	C>G	0.280	51.0/51.8	45.0/40.3	4.0/7.9	0.92 (0.76–1.12)	0.406
13	rs667282	76650527	T>C	0.489	32.6/26.4	50.0/44.8	17.4/28.8	0.70 (0.59–0.83)	2.4×10^{-5}
14	rs17486278	76654537	A>C	0.269	50.8/53.6	39.8/39.1	9.4/7.3	1.13 (0.93–1.36)	0.198
15	rs680244	76658343	G>A	0.252	49.7/56.4	44.3/37.2	6.0/6.4	1.17 (0.97–1.42)	0.097
16	rs692780	76663560	C>G	0.210	57.6/61.2	37.7/35.6	4.7/3.2	1.16 (0.95–1.42)	0.145
17	rs555018	76666297	T>C	0.188	63.0/65.6	32.9/31.2	4.1/3.2	1.12 (0.91–1.39)	0.273
18	rs12910984	76678682	A>G	0.476	34.9/28.5	48.9/47.8	16.2/23.7	0.76 (0.64–0.90)	0.001
19	rs3743078	76681814	G>C	0.254	56.4/54.7	37.6/39.7	6.0/5.6	0.97 (0.80–1.18)	0.751
20	rs3743077	76681951	G>A	0.210	60.8/62.0	34.0/34.1	5.2/3.9	1.08 (0.88–1.32)	0.459
21	rs1317286	76683184	A>G	0.094	80.1/81.9	19.2/16.9	0.7/1.2	1.07 (0.81–1.42)	0.616

*% of cases/% of controls.
†Allelic OR calculated by logistic regression.

Materials and Methods

Study subjects. In this study, we performed three independent case-control analyses. The pilot analysis, which was to replicate the association between the SNPs reported in Caucasian populations (7–9) and lung cancer risk in Chinese populations, consisted of 1,152 cases and 1,152 controls equally derived from both Northern and Southern China. The first-stage (discovery-phase) fine-mapping analysis, which was to discover novel risk variants in Chinese populations, consisted of 576 cases and 576 controls from Northern China. The second-stage (replication-phase) analysis, which was to confirm the results obtained from the first-stage analysis, consisted of 2,989 cases and 2,880 controls derived from both Northern and Southern cohorts. All subjects in this study were unrelated ethnic Han Chinese who were used in our previously published association studies (16–18). In the Northern Chinese population, lung cancer cases (n = 2,221) recruited at the Cancer Hospital, Chinese Academy of Medical Sciences (Beijing, China), were from Beijing city and surrounding provinces between 1997 and 2008, with a response rate of 94%. Controls (n = 2,112) were cancer-free individuals selected from a community nutritional survey of 6,450 individuals in the same region during the same period as cases were collected. In the Southern Chinese population, lung cancer cases (n = 1,344) were recruited at the Cancer Hospital of Jiangsu Province, the First Affiliated Hospital of Nanjing Medical University, and the Nanjing Thoracic Hospital (Nanjing, China) between 2002 and 2008 with a response rate of 91%. In the same time period as case patients were recruited, 1,344 cancer-free control subjects were randomly selected from a pool of 30,000 individuals participated in a community-based screening program for noninfectious diseases conducted in Jiangsu Province. All the control subjects were selected on the basis of physical examinations and frequency-matched for age and sex to each set of the lung cancer patients. We collected smoking information through interviews; those who had smoked an average of less than 1 cigarette per day and <1 y in their lifetime were defined as nonsmokers; otherwise, they were considered as smokers; those smokers who quit for >1 y before the recruitment were considered as former smokers. The distributions of selected characteristics among cases and controls used for the pilot discovery-phase and replication-phase association studies are shown in Supplementary Table S1 and S2. This study was approved by the institutional Review Boards of Chinese Academy of Medical Sciences Cancer Institute and Nanjing Medical University.

SNP identification and selection. The 100-kb linkage disequilibrium (LD) block on 15q25 (chr15:76589924-76683184) identified in Caucasians splits into 4 blocks in Asians (Supplementary Fig. S1*A*). Haplotype-tagging SNPs (htSNP) were chosen from genotyped SNPs in a Han Chinese population (CHB) in the HapMap Project [Rel 21a/phaseII Jan07, on National Center for Biotechnology Information B35 assembly, dbSNP b125, minor allele frequency (MAF) of ≥0.05, Hardy-Weinberg equilibrium P value of ≥0.05, and call rate of ≥90%] with Haploview 4.0 software on a block-by-block basis (r^2 threshold, 0.8). In addition, because Hung and colleagues (7) reported in their study that 25 SNPs other than rs8034191 and rs1051730 within and flanking the 100-kb region achieved evidence of association exceeding the GWAS significance level of 5×10^{-7}, we added additional 4 preblock or postblock SNPs based on the pair-wise LD analysis (r^2 threshold = 0.8) using HapMap CHB data. As a result, a total of 21 SNPs were analyzed in the discovery stage (Supplementary Fig. S1*B*). To infer potentially functional SNPs tagged by rs2036534, rs667282, and rs12910984 SNPs, which were found to be associated with lung cancer risk in our discovery stage, we searched all common (MAF, ≥0.05) SNPs located in the coding region, 5′flanking region, 5′untranslated region (5′UTR), and 3′UTR using HapMap CHB data around ~200-kb region on 15q25. We found that only rs6495309 SNP in the *CHRNA3* promoter region (−1923T>C) was in strong LD ($r^2 > 0.8$) with rs12910984 SNP, and this SNP was also chosen for genotyping all cases and controls (Supplementary Table S3; Supplementary Fig. S1*C*).

Genotype determination. We determined genotypes using a GenomeLab SNPstream 12-plex genotyping platform (Beckman Coulter). The overall call rate was 98.3%, with all loci having call rates of >95%. For quality

control, we developed two PCR-restriction fragment length polymorphism (PCR-RFLP) assays to determine two randomly selected SNPs, rs880395 and rs3813571, in 192 random samples and compared with the results by the SNPstream. The reproducibility was 99.5%. In the replication-phase, a PCR-RFLP method was developed to determine genotypes of rs6495309 SNP. To validate this PCR-RFLP method, we analyzed 540 random samples (330 cases and 210 controls) by both TaqMan (Applied Biosystems) and PCR-RFLP assays, and the concurrence rate of these 2 methods was 99.2%, indicating that the PCR-RFLP method was reliable. Genotypes were also confirmed by direct sequencing of PCR products. Genotyping of the four SNPs in the replication stage was performed independently in the two laboratories (Chinese Academy of Medical Sciences and Nanjing Medical University). A 5% random sample was reciprocally tested by different persons in the 2 laboratories, and the reproducibility was 100%. Laboratory technicians who performed genotyping experiments were blinded to case/control status. All sequences for PCR primers, probes, and enzymes are available upon request.

Construction of luciferase reporter plasmids. Three DNA fragments corresponding to the *CHRNA3* 5′-flanking region were generated by PCR (primers are available upon request) and subcoloned into the pGL3-Basic vector (Promega), and the resultant plasmids were designated as pC-2512, p-1642, or p-775, respectively. The pC-2512 construct containing C at nucleotide -1923 position relative to the transcriptional start site (rs6495309C allele) was then site-specifically mutated to create the construct pT-2512, which contains –1923T (rs6495309T allele). All constructs used in this study were restriction mapped and sequenced to confirm their authenticity.

Luciferase assays. We seeded 5×10^5 human lung cancer cells H1299, A549, and H446 per well in 12-well plates and transfected them with an empty pGL3-basic (a promoterless control) or pGL3-basic construct with a different *CHRNA3* promoter region or different rs6495309C/T alleles. pRL-SV40 plasmid (Promega) was cotransfected as a normalizing control. All transfections were carried out in triplicate. Cigarette smoke condensate was added into the medium at a final concentration of 5 ng/mL when the cells were cultured for 24 h. After incubation for another 6 h, cells were collected and analyzed for the luciferase activity with a Dual-Luciferase Reporter Assay System (Promega).

Analysis of *CHRNA3* RNA. Total RNA was isolated from surgically removed normal lung tissues adjacent to tumors of 55 lung cancer patients and then converted to cDNA using oligo $(dT)_{15}$ primer and Superscript II (Invitrogen). *CHRNA3* RNA was measured by real-time quantitative reverse transcription-PCR (RT-PCR) in triplicate using the StepOne Real-Time PCR System (Applied Biosystems) based on the SYBR-Green method. The measurement of individual *CHRNA3* RNA expression was calculated relative to that of *β-actin* expression using a modification of the method described by Lehmann and Kreipe (19). The primers used for detecting *CHRNA3* RNA are available upon request.

Electrophoretic mobility-shift assays. Synthetic double-stranded and 3′ biotin–labeled oligonucleotides and H1299 cell nuclear extracts were incubated at room temperature for 20 min using the LightShift Chemiluminescent EMSA kit (Pierce). The reaction mixture was separated on 8% PAGE, and the products were detected by Stabilized Streptavidin-Horseradish Peroxidase Conjugate (Pierce). Unlabeled oligonucleotides at 100-fold molar excess were added to the reaction for competition. We confirmed the identity of the DNA-binding protein in assays using antibodies specific to Oct-1 or nonspecific rabbit IgG (Santa Cruz Biotechnology).

Chromatin immunoprecipitation assays. H1299 cells carrying the *CHRNA3* rs6495309CC genotype were cross linked in 1% formaldehyde for 10 min. DNA from the fixed-chromatin cells was then subjected to immunoprecipitation using a chromatin immunoprecipitation (ChIP) assay kit (Upstate Biotechnology) and antibodies against Oct-1 or nonspecific rabbit IgG, according to the manufacturer's protocol (Santa Cruz). Purified DNA was analyzed by PCR with the primers used for rs6495309 SNP genotyping.

Statistical analysis. We used χ^2 or Student's t tests to examine the differences in the distributions of demographic characteristics, selected variables, and genotypes between cases and controls. Light and heavy smokers

Table 2. Genotype frequencies of the four SNPs in 15q25 in patients and controls and their associations with lung cancer risk

Genotype	Beijing replication (1645/1536)*			Nanjing replication (1344/1344)*			Pooled (2989/2880)*		
	%†	OR‡ (95% CI)	*P*	%†	OR‡ (95% CI)	*P*	%†	OR‡ (95% CI)	*P*
rs2036534									
CC	16.4/19.9	1.00 (Reference)		15.1/20.2	1.00 (Reference)		15.8/20.1	1.00 (Reference)	
CT	50.6/49.3	1.25 (1.03–1.51)	0.025	48.6/47.6	1.37 (1.11–1.70)	0.004	49.7/48.5	1.30 (1.13–1.50)	3.3×10^{-4}
TT	33.0/30.8	1.30 (1.05–1.59)	0.014	36.3/32.2	1.52 (1.21–1.90)	3.1×10^{-4}	34.5/31.4	1.40 (1.20–1.62)	1.9×10^{-5}
P_{trend}		0.024			6.7×10^{-4}			6.4×10^{-5}	
rs667282									
CC	16.9/23.8	1.00 (Reference)		18.0/22.8	1.00 (Reference)		17.4/23.4	1.00 (Reference)	
CT	52.8/48.3	1.54 (1.28–1.86)	4.1×10^{-6}	50.8/47.8	1.34 (1.10–1.64)	0.004	51.9/48.0	1.45 (1.26–1.66)	9.6×10^{-8}
TT	30.3/27.9	1.53 (1.25–1.88)	3.8×10^{-5}	31.2/29.4	1.34 (1.07–1.67)	0.009	30.7/28.6	1.44 (1.24–1.68)	1.4×10^{-6}
P_{trend}		1.7×10^{-4}			0.018			1.1×10^{-5}	
rs12910984									
GG	15.6/22.0	1.00 (Reference)		16.7/20.6	1.00 (Reference)		16.1/21.3	1.00 (Reference)	
GA	53.0/48.4	1.55 (1.28–1.87)	7.2×10^{-6}	49.3/48.4	1.26 (1.02–1.55)	0.032	51.4/48.4	1.41 (1.22–1.62)	1.9×10^{-6}
AA	31.4/29.6	1.48 (1.21–1.83)	1.9×10^{-4}	34.0/31.0	1.36 (1.09–1.70)	0.007	32.5/30.3	1.43 (1.23–1.66)	3.9×10^{-6}
P_{trend}		1.2×10^{-3}			0.011			3.5×10^{-5}	
rs6495309									
TT	15.4/21.8	1.00 (Reference)		17.6/21.4	1.00 (Reference)		16.4/21.6	1.00 (Reference)	
TC	53.0/48.9	1.54 (1.27–1.87)	8.9×10^{-6}	52.6/50.2	1.28 (1.05–1.57)	0.017	52.8/49.5	1.41 (1.23–1.62)	1.1×10^{-6}
CC	31.6/29.3	1.52 (1.24–1.87)	7.4×10^{-5}	29.8/28.4	1.28 (1.03–1.60)	0.029	30.8/28.9	1.41 (1.21–1.64)	1.1×10^{-5}
P_{trend}		4.8×10^{-4}			0.048			8.5×10^{-5}	

*Number of cases/number of controls.
†% of cases/% of controls.
‡Data were calculated by logistic regression analysis with adjustment for age, sex and center.

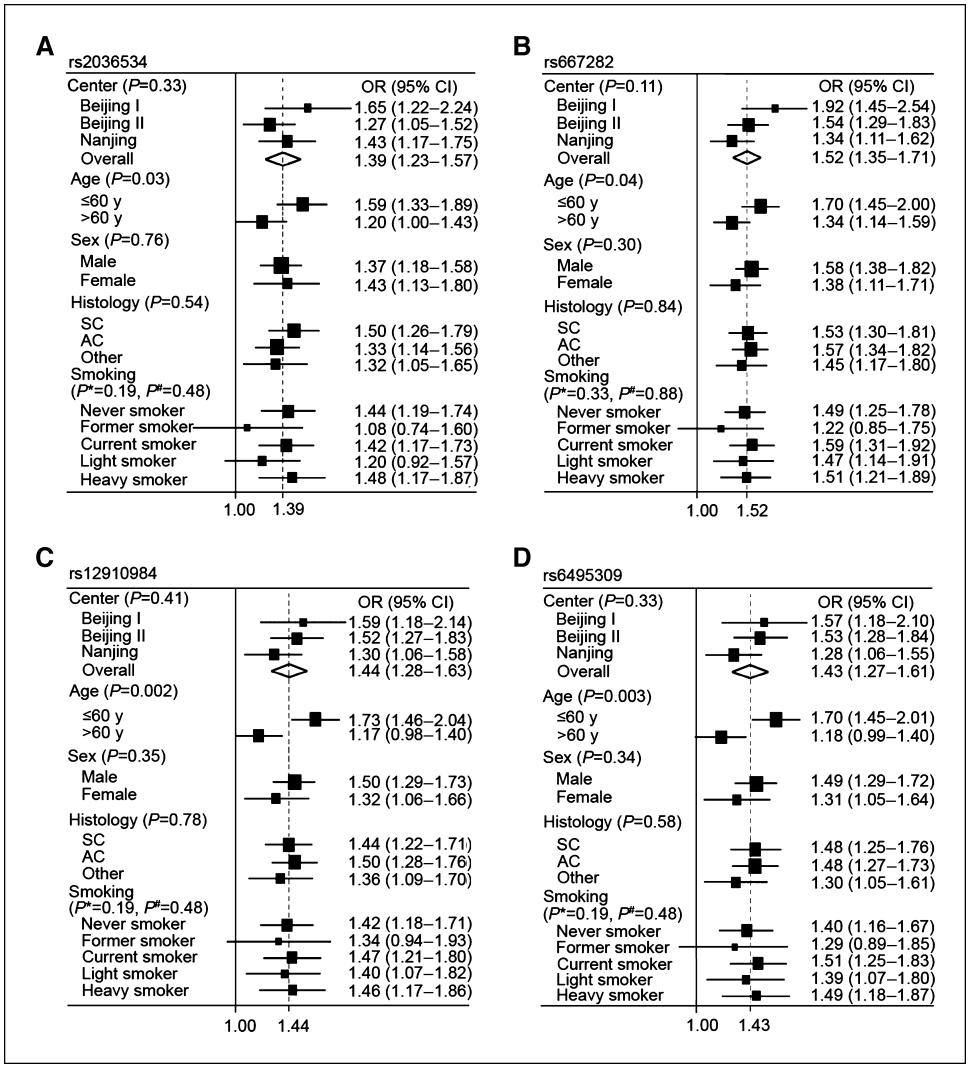

Figure 1. Forest plot representing lung cancer risk and the four variants in the 15q25 region assuming dominant effects, adjusted by age, sex, and center (*A*) rs2036534, (*B*) rs667282, (*C*) rs12910984, and (*D*) rs6495309. *SC,* squamous cell carcinoma; *AC,* adenocarcinoma; other includes small cell lung cancer, mixed cell carcinoma, and bronchioalveolar carcinoma. *P* values are from heterogeneity tests. *, never smokers versus former smokers versus current smokers; †, never smokers versus light smokers versus heavy smokers.

were categorized by using the 50th percentile pack-year values [pack-years = (cigarettes per day/20) × (years smoked)] of the controls as the cutoff points (i.e., ≤25 pack-years and >25 pack-years). LD blocks were determined by the method of Gabriel and colleagues (20) in Haploview 4.0 software with default settings. Within-block htSNP haplotype frequencies were computed by using the EM algorithm in SAS 9.1.3 PROC HAPLOTYPE. Logistic regression was used to calculate odds ratios (OR) and 95% confident intervals (CI) to estimate the association between a single locus and lung cancer risk, with adjustment for age, sex, center, and smoking status, where appropriate. A more-than-multiplicative gene-environment interaction was evaluated by logistic regression analysis including main effect variables and their product terms. When the test for multiplicative interaction was not rejected, further tests for an additive interaction were performed by a bootstrapping method for 1000 times. A more-than-additive interaction was indicated when $OR_{11}>OR_{10}+OR_{01}-1$, where OR_{11} = OR when both factors were present, OR_{01} = OR when only factor 1 was present, OR_{10} = OR when only factor 2 was present. Statistical heterogeneity between studies was assessed with the χ^2-based Q test (21). To assess the significance, we performed both Bonferroni correction and permutation procedure to correct the *P* value of single-locus association results for multiple testing. All analyses were performed using the SAS (version 9.1.3; SAS Institute) and Stata software (version 10.0; StataCorp LP). *P* value of <0.05 was used as the criterion of statistical significance, and all statistical tests were two sided.

Results

We first investigated the association between the three risk SNPs (rs8034191, rs1051730, and rs16969968) previously identified in Caucasians and risk of lung cancer in a pilot study consisting of 1,152 cases with lung cancer and 1,152 controls. We found that the variant genotypes were extremely rare (all MAF, <0.05) in our study population, which is consistent with the genotyping data for CHB in the HapMap database. Furthermore, these SNPs were not significantly associated with lung cancer risk in Chinese populations (Supplementary Table S4), and the statistical power was limited due to the rarity of the polymorphisms (assuming an allelic OR of 1.30, the power is 24.2%, 22.9%, or 22.2% for rs8034191, rs1051730, or rs16969968, respectively). In the discovery-phase analysis consisting of 576 lung cancer cases and 576 controls, we found that three common SNPs, rs2036534, rs667282, and rs12910984, were significantly associated with lung cancer risk (all *P* value of <0.002 after Bonferroni corrections; Table 1).

The associations of lung cancer risk with these three newly identified SNPs were further replicated in the replication phase by two independent case-control cohorts. The first replication cohort consisted of 1,645 cases and 1,536 controls derived from a Northern

Chinese population (Beijing); the second replication cohort included 1,344 cases and 1,344 controls derived from a Southern Chinese population (Nanjing, Jiangsu Province). Genotyping results showed that these three SNPs were all statistically significantly associated with lung cancer risk in both Northern and Southern Chinese populations (Table 2). Although *P* values for the log-additive model were statistically significant for all the three SNPs, the scales of the ORs were similar for heterozygotes and risk homozygotes. We therefore used dominant genetic model hereinafter. Final pooled analysis showed that the adjusted (for age, sex, and center) ORs and 95% CIs for the 3 SNPs in the dominant model were 1.39 (95% CI, 1.23-1.57; $P = 2.3 \times 10^{-7}$) for the rs2036534TT/TC, 1.52 (95% CI, 1.35-1.71; $P = 2.0 \times 10^{-12}$) for the rs667282TT/TC, and 1.44 (95% CI, 1.28-1.63; $P = 2.7 \times 10^{-9}$) for the rs12910984AA/AG genotypes compared with the rs2036534CC, rs667282CC, and rs12910984GG genotype, respectively (Fig. 1*A–C*). Further adjustment for age, sex, center, and smoking status did not notably change the ORs and significance levels (Supplementary Fig. S2).

From the HapMap CHB database, we found another SNP (rs6495309T>C) in the promoter region of *CHRNA3* that is in strong LD with rs12910984 ($r^2 = 0.85$). Because computational analysis suggested that this SNP is within a transcriptional factor Oct-1 binding site and thus may have functional consequence, we therefore genotyped it in all case-control sets. We found that subjects with at least one copy of the rs6495309C allele (CC or CT genotype) had significantly elevated risk of lung cancer compared with the rs6495309TT genotype (OR, 1.43; 95% CI, 1.27-1.61; $P = 2.6 \times 10^{-9}$; Fig. 1*D*).

Although the four newly identified lung cancer risk-associated SNPs are not within the same LD block (Supplementary Fig. S1*B*), they are in strong LD in our study populations (all pare-wise r^2>0.45 and all pare-wise *D*'>0.70 in this study population versus all pare-wise r^2>0.75 and all pare-wise *D*'>0.88 in Caucasians; Supplementary Table S5). We thus constructed the 4-site haplotypes (rs2036534/rs667282/rs12910984/rs6495309) and found 7 common haplotypes (MAF, >0.01) in both cases and controls (Supplementary Table S6). We observed that the CCGT haplotype frequency was significantly lower in cases than in controls, whereas the TTAC haplotype was significantly higher in cases than in controls (both $P < 0.0001$, 1,000 time permutation test). In a stepwise logistic regression analysis, the most common haplotype TTAC was retained in the final model along with age, sex, smoking, and a rare haplotype TTAT. Subjects carrying the TTAC haplotype had a 1.47-fold (95% CI, 1.28-1.68; $P = 3.1 \times 10^{-8}$) increased lung cancer risk compared with the other haplotypes. To further assess the independence of the four SNPs, we also conducted stepwise logistic regression analysis with all these SNPs included in the model, and found that only rs667282 was retained in the final model along with sex and smoking status (all $P < 0.001$).

Stratification analyses using pooled case-control sets showed that the risk associated with the four SNPs were all significantly different between the subgroups by age (test for heterogeneity, $P = 0.03$ for rs2036534, $P = 0.04$ for rs667282, $P = 0.002$ for rs12910984, and $P = 0.003$ for rs6495309), but not the subgroups by sex, histology, smoking, or research centers (Fig. 1). Although the associations between the 4 SNPs and lung cancer risk were found in both smokers and nonsmokers, they were all additively interacted with smoking status (never-smokers versus ever-smokers) to contribute to lung cancer risk ($P = 0.061$ for rs2036534, $P = 0.003$ for rs667282, $P = 0.009$ for rs12910984, and $P = 0.002$ for rs6495309). We next examined the association between these four SNPs and smoking behavior in all 7,021 subjects (Table 3). The frequencies of risk homozygotes for the 4 SNPs were all significantly higher in current smokers ($P_{trend} = 0.004$, 2.1×10^{-5}, 2.5×10^{-5}, or 5.6×10^{-6}, respectively). The analyses stratified by cigarettes per day and duration of smoking showed that subjects carrying the risk homozygotes were also more likely to become heavy smokers (consume ≥30 cigarettes per day) and smoked longer duration (≥40 years; Table 3). Similar trend was observed for cases and controls separately and for males and females separately (detailed data not shown).

Table 3. Distribution of genotypes of the four SNPs in 15q25 by smoking status in Chinese populations

	rs2036534			rs667282			rs12910984			rs6495309		
	CC	CT	TT	CC	CT	TT	GG	GA	AA	TT	TC	CC
Smoking status												
Never (n = 3296)	18.9	49.8	31.3	22.6	49.2	28.3	20.1	50.1	29.8	20.5	52.1	27.4
Former (n = 763)	18.3	49.1	32.6	21.9	51.3	26.8	21.2	50.5	28.2	20.6	52.0	27.4
Current (n = 2962)	17.1	47.8	35.1	18.6	49.6	31.9	16.8	49.0	34.2	17.7	49.1	33.2
P_{trend}*		0.004			2.1×10^{-5}			2.5×10^{-5}			5.6×10^{-6}	
Smoking level (cigarettes per day)†												
≤15 (n = 1079)	18.6	48.3	33.1	20.3	49.8	29.9	19.2	48.1	32.7	19.1	48.4	32.5
15–30 (n = 1654)	18.2	48.3	33.5	20.3	49.6	30.1	18.7	50.0	31.3	20.1	49.2	30.7
≥30 (n = 992)	14.4	47.7	37.9	16.4	50.6	33.0	14.4	49.6	36.0	14.4	51.8	33.8
P_{trend}*		4.2×10^{-3}			0.022			5.1×10^{-3}			0.027	
Duration of smoking (y)†												
≤20 (n = 982)	19.9	47.4	32.7	21.4	51.3	27.3	19.8	50.5	29.7	20.6	49.9	29.5
20–40 (n = 1613)	16.3	48.5	35.2	18.9	48.9	32.2	17.7	47.8	34.6	18.0	48.8	33.2
≥40 (n = 1130)	16.5	48.3	35.2	17.7	50.3	31.9	16.1	50.5	33.5	16.8	50.7	32.5
P_{trend}*		0.046			8.2×10^{-3}			0.014			0.021	

*Trend test, homozygote comparison.

†Categorized by using the 25th and 75th percentile daily cigarettes consumption and years smoked in all subjects as the cutoff points.

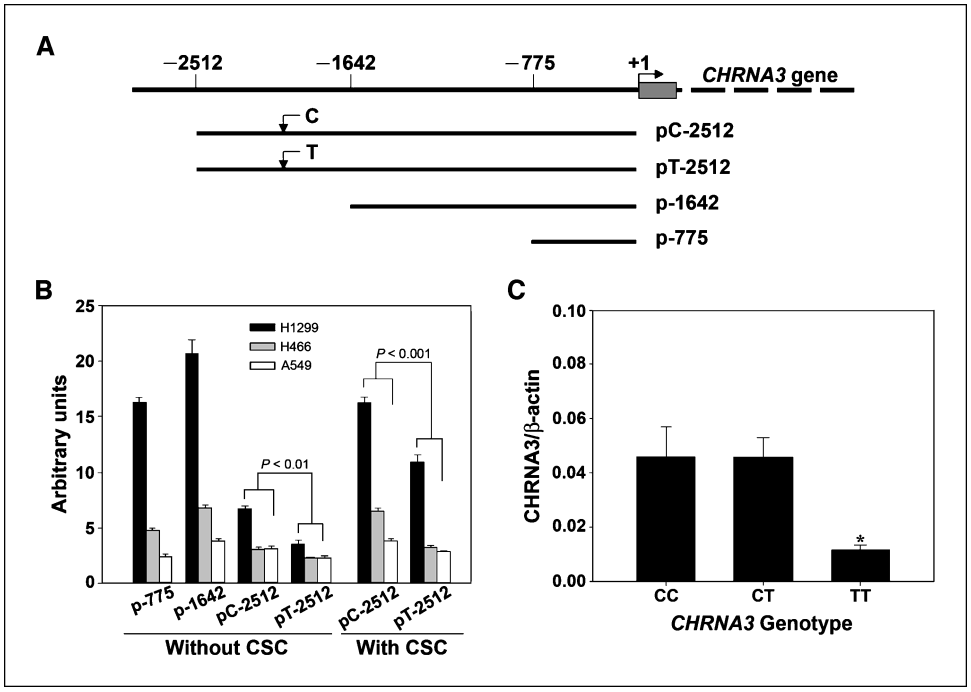

Figure 2. Promoter activity of *CHRNA3* with different deletion constructs or constructs containing different alleles of rs6495309T>C SNP. *A*, schematic representation of the *CHRNA3* 5′-flanking region and reporter gene constructs used in this study. *B*, luciferase expression of deletion constructs and constructs containing rs6495309C or rs6495309T in H1299 cells with or without cigarette smoke condensate (*CSC*) stimulation. All constructs were cotransfected with pRL-SV40 to standardize the transfection efficiency. Luciferase levels of pGL3-Basic and pRL-SV40 were determined in triplicate. Fold increase was measured by defining the activity of the empty pGL3-Basic vector as 1. *Columns*, means from three independent experiments, each in triplicate; *bars*, SD. *C*, levels of *CHRNA3* RNA expression in lung tissues as a function of rs6495309T>C genotype. *Columns*, means normalized to β-actin; *bars*, SE. Expression levels among the CC and CT genotypes were significantly higher than that among the TT genotype; *, $P < 0.05$.

Because the rs6495309 SNP is located within the *CHRNA3* promoter region and computational analysis showed that it might affect the transcriptional factor Oct-1 binding ability, we therefore examined whether the SNP has an impact on the *CHRNA3* promoter activity by a set of luciferase reporter gene constructs, which were used to transfect transiently human lung cancer cell line H1299, H446, and A549 cells (Fig. 2*A*). Intriguingly, we observed a significantly higher promoter activity for constructs p-775 and p-1642, which contain a shorter 5′-flanking region, compared with constructs pC-2512 and pT-2512, which contain a longer 5′-flanking region encompassing the rs6495309C/T site (Fig. 2*A* and *B*). The results indicated that there may exist a negative regulatory element between −1642 and −2512 from the transcriptional start site. As expected, the construct pC-2512 containing the C allele drove a significantly higher reporter gene expression compared with the T allele ($P < 0.01$; Fig. 2*B*). Further experiments showed that the reporter gene expression driven by either the rs6495309C or T allelic *CHRNA3* promoter could be comparably induced by cigarette smoke condensate (Fig. 2*B*). The *CHRNA3* promoter constructs exhibited similar activity patterns in the three cell lines, showing a common regulatory mechanism for *CHRNA3* expression in the lung epithelial cells.

We also evaluated the association between rs6495309T>C genotypes and *CHRNA3* RNA levels in surgically removed normal lung tissues adjacent to the tumors from 55 lung cancer patients using quantitative RT-PCR. As shown in Fig. 2*C*, the *CHRNA3* RNA level (mean ± SE) was significantly higher in rs6495309 CC (0.459 ± 0.011; $n = 17$) and CT genotype carriers (0.459 ± 0.007; $n = 28$) than that in TT genotype carriers (0.012 ± 0.001; $n = 10$; $P = 0.031$ and 0.010, respectively). These results further confirmed that the rs6495309C-containing *CHRNA3* may have higher promoter activity in a dominant manner.

Electrophoretic mobility-shift assays were then conducted to distinguish the differences in binding capacity between the rs6495309C and T alleles to transcriptional factor (Fig. 3*A*). The results showed that Oct-1–containing nuclear extracts bound more abundantly to the biotin-labeled oligonucleotide probe with the T allele than to that with the C allele. A 100-fold excess of unlabeled oligonucleotide or the Oct-1 consensus sequences competed with this binding activity. Furthermore, we verified that the protein bound to the T allele was indeed Oct-1 by using specific antibodies against Oct-1, which abolished the binding when it was added to the assays (Fig. 3*B*). We also examined whether the binding of Oct-1 to the *CHRNA3* promoter occurred *in vivo* in cells using ChIP assays in H1299 cells with the rs6495309CC genotype. Indeed, the DNA was precipitated specifically with the Oct-1 antibody but not nonspecific IgG (Fig. 3*C*).

Discussion

In the present study, we examined whether the most significant SNPs on chromosome 15q25 identified in Caucasians (7–9) are also associated with risk of lung cancer in Chinese populations. Unfortunately, on the basis of analysis of 1,152 lung cancer cases and 1,152 controls derived from both Northern and Southern Chinese populations, we found that these SNPs were not significantly different between cases and controls. However, by using the htSNP approach, we identified other four novel SNPs that were associated with risk of lung cancer in Chinese populations. To summarize, we found that the rs2036534C>T, rs667282C>T, rs12910984G>A, and rs6495309T>C variants in the *CHRNA3-CHRNB4-CHRNA5* gene cluster located in chromosome 15q25 were significantly associated with increased risk of lung cancer and smoking behavior. In addition, we showed that the rs6495309T>C change in the *CHRNA3* promoter region results in up-regulation of *CHRNA3* expression due to influencing its binding ability to the transcriptional factor Oct-1, suggesting that it may be a causal SNP for susceptibility to lung cancer. Together, our results support chromosome 15q25 as a susceptibility region for lung cancer and underscore the difference in genetic markers

among different ethnic populations and utility of conducting genetic studies in diverse populations.

When this article was in preparation, a study conducted in Japan was published, reporting that the rs8034191, rs1051730, and rs16969968 SNPs, the risk variants in European Caucasians described in previous GWAS (7–9), are also associated with lung cancer risk in a Japanese population (22). These findings are inconsistent with our results in Chinese populations in the present study and might also reflect the difference among different populations. However, it should be noted that the genotype frequencies for these SNPs reported in the Japanese study do not conform to those expected under Hardy-Weinberg equilibrium ($P < 0.000001$ for rs8034191 and $P = 0.064$ for rs1051730 calculated using the data presented in the article), albeit it was claimed that all genotypes in the study were consistent with Hardy-Weinberg equilibrium (22), suggesting that the results in that study might have some selection bias. On the other hand, the Japanese study genotyped only seven SNPs in this chromosome region, which is not as comprehensive as ours. Furthermore, because of the extreme rarity of the rs8034191, rs1051730, and rs16969968 variants in Asian populations (22 and this study), these variants play little roles, if any, in risk of lung cancer and nicotine dependence in Asians populations including Chinese and Japanese populations. In contrast, our data showed that the rs6495309C allele associated with higher *CHRNA3* expression was prevalent in both smokers and lung cancer patients. This SNP has also been shown to be associated with statistically significantly increased lung cancer risk in Caucasians (8), suggesting that it may be a causal variant across diverse ethnic populations.

In the present study, we showed using a set of biochemical assays that the rs6495309T>C change considerably influenced the *CHRNA3* promoter activity, resulting in significant increase in the *CHRNA3* RNA expression with rs6495309C allele compared with the rs6495309T allele. This effect may result from reduced ability of the rs6495309C allele to bind Oct-1, a transcriptional factor that has been shown to be able to repress gene transcription (23–25). Consistent with this finding, our results showed that the reporter gene constructs containing the rs6495309C allele had a notably lower promoter activity compared with the constructs without rs6495309C allele. Although the *CHRNA3* RNA levels related to the genotypes were measured in lung tissues, it may represent those in the neural tissues because the SNP is a germline mutation. We speculated that the higher CHRNA3 receptor production due to the rs6495309C allele may render individuals more sensitive and/or tolerant to nicotine and thus become nicotine-dependent, leading to higher levels of exposure to smoking, a well-established etiologic factor for lung cancer. It is conceivable that individuals consumed more cigarettes per day and smoked longer duration would be expected to be at higher risk for developing lung cancer. However, we observed that nonsmokers also had increased risk associated with these SNPs on chromosome 15q25 region. This result may be explained by environmental smoking. Exposure to environmental smoke is almost unavoidable, especially in China, where there are >300 million smokers, and 530 million people were estimated to be exposed to environmental smoke (26).

In addition to exposure to tobacco carcinogens because of smoking addiction due to genetic polymorphisms in chromosome 15q25, other mechanisms may also exist to explain the association between higher production of nicotinic acetylcholine receptors resulted from genetic polymorphism and risk of lung cancer. Accumulating evidence suggests that nicotine may contribute directly to lung carcinogenesis by stimulating nicotinic acetylcholine receptors in nonneuronal cells (12–15). It has been shown that nicotine and its carcinogenic derivatives such as NNK acting through specific nicotinic acetylcholine receptors expressed in human airway epithelial cells cause cells' loss of contact inhibition and resistance to apoptosis (12–15). Numerous studies have also shown that both nicotine and NNK are able to promote lung cancer cell proliferation and suppress apoptosis induced by diverse stimuli via nicotinic acetylcholine receptors mediated activation of multiple signaling pathways (27–30).

In our study populations, the association with lung cancer risk of htSNP rs12910984 located in intron 5 of *CHRNA3* may be explained by functional SNP rs6495309 located in the promoter of the gene because these two SNPs are in strong LD. However, in stepwise logistic regression analysis, only rs667282 was included in the final model and this SNP had higher risk scale compared with rs6495309. Therefore, we cannot rule out the possibility of having other causal SNPs in highly LD with the three SNPs (especially

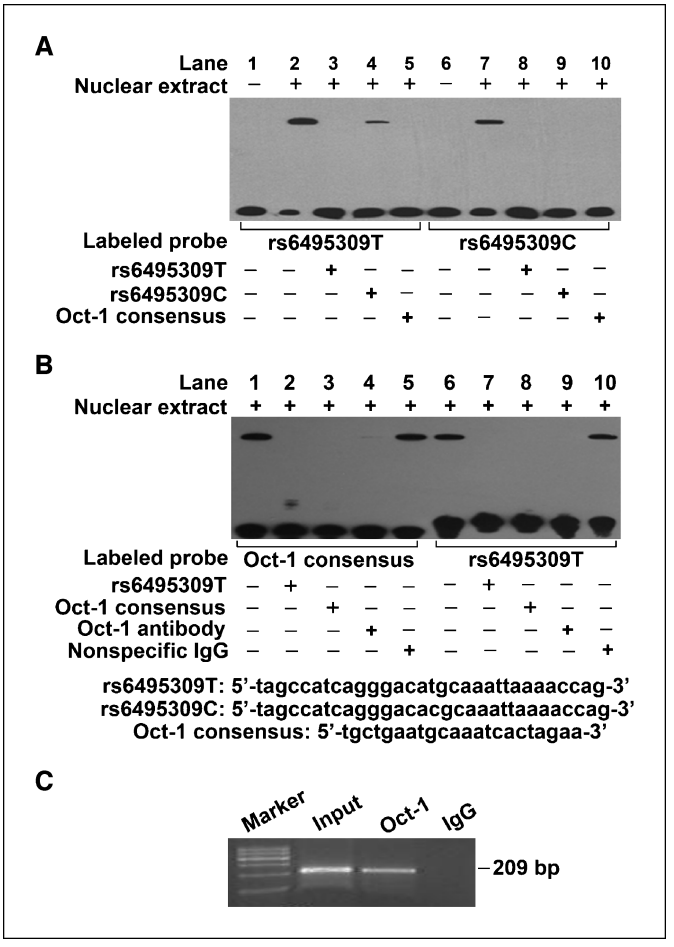

Figure 3. Abolishment of an Oct-1 binding site in the *CHRNA3* promoter by the rs6495309T>C change. *A,* electrophoretic mobility-shift assay with biotin-labeled rs6495309T or rs6495309C probes and H1299 nuclear extract. *Lanes 1* and *6,* probe only; *lanes 2* and *7,* probe and nuclear extracts; *lanes 3* to *5* and *8* to *10,* probe and nuclear extracts plus 100× unlabeled rs6495309T (*lanes 3* and *8*), rs6495309C (*lanes 4* and *9*), or Oct-1 consensus probes (*lanes 5* and *10*). *B,* electrophoretic mobility-shift assay with antibody. *Lanes 4* and *9,* with antibody to Oct-1; *lanes 5* and *10,* with rabbit IgG. *C,* ChIP assays using H1299 cells carrying the rs6495309CC genotype. The presence of Oct1-binding *CHRNA3* promoter was verified by PCR.

rs667282 in *CHRNA5*) within 15q25 region. Further investigations including resequencing the 15q25 region in Asian populations and functional assays to identify the causal SNPs tagged by these htSNPs are warranted.

In conclusion, we identified four novel SNPs in the *CHRNA3-CHRNB4-CHRNA5* gene cluster on chromosome 15q25 that are significantly associated with risk of lung cancer and smoking tendency in Chinese populations. Among them, the rs6495309T>C SNP located in the *CHRNA3* promoter region alters the ability to bind transcriptional factor Oct-1, resulting in increased *CHRNA3* expression, which might be the underlying mechanism in conferring susceptibility to lung cancer.

Disclosure of Potential Conflicts of Interest

No potential conflicts of interest were disclosed.

Acknowledgments

Received 1/9/09; revised 3/23/09; accepted 4/9/09; published OnlineFirst 6/2/09.

Grant support: National Natural Science Foundation grant 30730080 (H. Shen) and State Key Basic Research Program grants 2004CB518701 (D. Lin) and 2002CB512902 (H. Shen).

The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked *advertisement* in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.

We thank Drs. Q. Wei (The University of Texas M. D. Anderson Cancer Center) and P. Brennan (The IARC) for critical review and editing of the manuscript.

References

1. Corrao MA, Guindon GE, Sharma N. Building the evidence base for global tobacco control. Bull World Health Organ 2000;78:884–90.
2. Zhang H, Cai B. The impact of tobacco on lung health in China. Respirology 2003;8:17–21.
3. Parkin DM, Pisani P, Lopez AD, Masuyer E. At least one in seven cases of cancer is caused by smoking: global estimates for 1985. Int J Cancer 1994;59:494–504.
4. Wu X, Zhao H, Amos CI. et al. p53 Genotype and haplotypes associated with lung cancer susceptibility and ethnicity. J Natl Cancer Inst 2002;94:681–90.
5. Zhan X, Miao X, Liang G, et al. Polymorphisms in DNA base excision repair genes ADPRT and XRCC1 and risk of lung cancer. Cancer Res 2005;65:722–6.
6. Wang H, Tan W, Hao B, et al. Substantial reduction in risk of lung adenocarcinoma associated with genetic polymorphism in CYP2A13, the most active cytochrome P450 for the metabolic activation of tobacco-specific carcinogen NNK. Cancer Res 2003;63:8057–61.
7. Hung RJ, McKay JD, Gaborieau U, et al. A susceptibility locus for lung cancer maps to nicotinic acetylcholine receptor subunit genes on 15q25. Nature 2008;452:633–7.
8. Thorgeirsson TE, Geller F, Sulem P, et al. A variant associated with nicotine dependence, lung cancer and peripheral arterial disease. Nature 2008;452:638–42.
9. Amos CI, Wu X, Broderick P, et al. Genome-wide association scan of tag SNPs identifies a susceptibility locus for lung cancer at 15q25.1. Nat Genet 2008;40:616–22.
10. Saccone SF, Hinrichs AL, Saccone NL, et al. Cholinergic nicotinic receptor genes implicated in a nicotine dependence association study targeting 348 candidate genes with 3713 SNPs. Hum Mol Genet 2007;16:36–49.
11. Bierut LJ, Stitzel JA, Wang JC, et al. Variants in nicotinic receptors and risk for nicotine dependence. Am J Psychiatry 2008;165:1163–71.
12. West KA, Brognard J, Clark AS, et al. Rapid Akt activation by nicotine and a tobacco carcinogen modulates the phenotype of normal human airway epithelial cells. J Clin Invest 2003;111:81–90.
13. Minna JD. Nicotine exposure and bronchial epithelial cell nicotinic acetylcholine receptor expression in the pathogenesis of lung cancer. J Clin Invest 2003;111:31–3.
14. Tsurutani J, Castillo SS, Brognard J, et al. Tobacco components stimulate Akt-dependent proliferation and NFκB-dependent survival in lung cancer cells. Carcinogenesis 2005;26:1182–95.
15. Schuller HM. Is cancer triggered by altered signalling of nicotinic acetylcholine receptors? Nat Rev Cancer 2009;9:195–205.
16. Sun T, Zhou Y, Yang M, et al. Functional genetic variations in cytotoxic T-lymphocyte antigen 4 and susceptibility to multiple types of cancer. Cancer Res 2008;68:7025–34.
17. Sun T, Gao Y, Tan W, et al. A six-nucleotide insertion-deletion polymorphism in the CASP8 promoter is associated with susceptibility to multiple cancers. Nat Genet 2007;39:605–13.
18. Zhang X, Miao X, Guo Y, et al. Genetic polymorphisms in cell cycle regulatory genes MDM2 and TP53 are associated with susceptibility to lung cancer. Hum Mutat 2006;27:110–7.
19. Lehmann U, Kreipe H. Real-time PCR analysis of DNA and RNA extracted from formalin-fixed and paraffin-embedded biopsies. Methods 2001;25:409–18.
20. Gabriel SB, Schaffner SF, Nguyen H, et al. The structure of haplotype blocks in the human genome. Science 2002;296:2225–9.
21. Lau J, Ioannidis JP, Schmid CH. Quantitative synthesis in systematic reviews. Ann Intern Med 1997;127:820–6.
22. Shiraishi K, Kohno T, Kunitoh H, et al. Contribution of nicotine acetylcholine receptor polymorphisms to lung cancer risk in a smoking-independent manner in the Japanese. Carcinogenesis 2009;30:65–70.
23. Hitomi T, Matsuzaki Y, Yasuda S, et al. Oct-1 is involved in transcriptional repression of $p15^{ink4b}$ gene. FEBS Lett 2007;581:1087–109.
24. Mesplede T, Island ML, Christeff N, Petek F, Doly J, Navarro S. The POU transcription factor Oct-1 represses virus-induced interfron A gene expression. Mol Cell Biol 2005;25:8717–31.
25. Wu GD, Lai EJ, Huang N, Wen X. Oct-1 and CCAAT/enhancer-binding protein (C/EBP) bind to overlapping elements within the interleukin-8 promoter. The role of Oct-1 as a transcriptional repressor. J Biol Chem 1997;272:2396–403.
26. Yang G, Kong L, Zhao W, et al. Emergence of chronic non-communicable diseases in China. Lancet 2008;372:1697–705.
27. Maneckjee R, Minna JD. Opioids induce while nicotine suppresses apoptosis in human lung cancer cells. Cell Growth Differ 1994;5:1033–40.
28. Heusch WL, Maneckjee R. Signaling pathways involved in nicotine regulation of apoptosis of human lung cancer cells. Carcinogenesis 1998;19:551–6.
29. Ho YS, Chen CH, Wang YJ, et al. Tobacco-specific carcinogen 4-(methylnitrosamino)-1-(3-pyridyl)-1-butanone (NNK) induces cell proliferation in normal human bronchial epithelial cells through NFκB activation and cyclin D1 up-regulation. Toxicol Appl Pharmacol 2005;205:133–48.
30. Dasgupta P, Kinkade R, Joshi B, Decook C, Haura E, Chellappan S. Nicotine inhibits apoptosis induced by chemotherapeutic drugs by up-regulating XIAP and survivin. Proc Natl Acad Sci U S A 2006;103:6332–7.