A Second Genetic Variant on Chromosome 15q24-25.1 Associates with Lung Cancer

Pengyuan Liu[1*], Ping Yang[2*], Xifeng Wu[3*], Haris G. Vikis[1], Yan Lu[1], Yian Wang[1], Ann G. Schwartz[4], Susan M. Pinney[5], Mariza de Andrade[2], Adi Gazdar[6], Colette Gaba[7], Diptasri Mandal[8], Juwon Lee[5], Elena Kupert[5], Daniela Seminara[9], John Minna[6], Joan E. Bailey-Wilson[10], Margaret Spitz[3], Christopher I. Amos[3], Marshall W. Anderson[5], and Ming You[1]

Supplementary Data

Supplementary Figure legend

Figure S1 Two possible mechanisms underlying lung carcinogenesis in smokers.

Figure S1

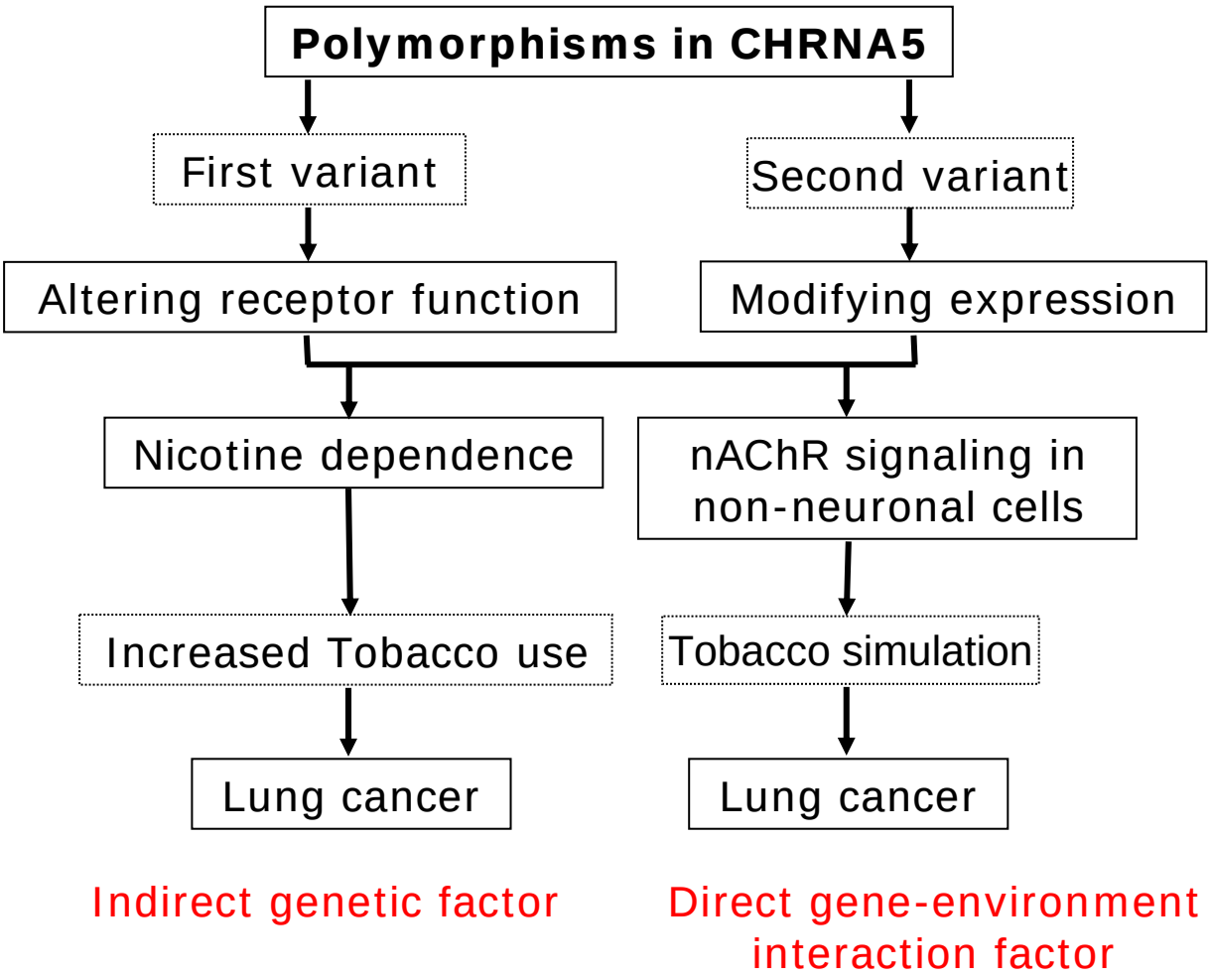
Polymorphisms in CHRNA5
First variant
Second variant
Altering receptor function
Modifying expression
Nicotine dependence
nAChR signaling in non-neuronal cells
Increased Tobacco use
Tobacco simulation
Lung cancer
Lung cancer
Indirect genetic factor
Direct gene-environment interaction factor

Table S1 Characteristics of study populations*

Characteristics	Age, y	Sex, %		Smoking status, %	
	Mean (± SD)	Female	Male	Nonsmoker	Smoker
GELCC					
Cases (n = 194)	61.3 (10.4)	57.7%	42.3%	14.4%	86.6%
Controls (n = 219)	75.4 (8.6)	41.1%	58.9%	6.4%	93.6%
MCC					
Cases (n = 890)	66.1 (8.8)	41.2%	58.8%	0.0%	100.0%
Controls (n = 865)	65.3 (9.2)	45.0%	55.0%	0.0%	100.0%
TXC					
Cases (n = 1466)	61.7 (11.1)	46.3%	53.7%	16.2%	83.8%
Controls (n = 1389)	61.3 (9.2)	47.4%	52.6%	13.5%	86.5%
TXA					
Cases (n = 268)	61.6 (10.6)	44.4%	55.6%	10.1%	89.9%
Controls (n = 293)	57.1 (10.1)	58.0%	42.0%	24.9%	75.1%

* Cohorts: GELCC=Genetic Epidemiology of Lung Cancer Consortium; MCC=Mayo Clinic Caucasians; TXC=Texas Caucasians; TXA=Texas African Americans. Cases in the GELCC are lung cancer patient with family history of lung cancer, and cases in MCC, TXC and TXA are lung cancer patient without family history.

Table S2 Results from four populations were combined using Mantel-Hazenszel tests.

Samples/SNP*	Alleles †	Frequency		P-values†	Marginal OR	95%CI
		Cases	Controls			
All subjects in single-marker analysis (2818 cases and 2794 controls)						
rs1051730	G/**A**	0.4094	0.2938	1.48×10^{-7}	1.24	1.14-1.34
rs481134	T/**C**	0.5632	0.5714	0.512	0.97	0.90-1.05
Subjects with genotypes GG at SNP rs1051730 (1144 cases and 1288 controls) ‡						
rs481134	T/**C**	0.3694	0.4227	7.93×10^{-4}	0.82	0.72-0.92
Subjects with genotypes CC at SNP rs481134 (969 cases and 926 controls) §						
rs1051730	G/**A**	0.6232	0.5093	2.98×10^{-11}	1.59	1.51-1.68

* Subjects are combined four populations: GELCC=Genetic Epidemiology of Lung Cancer Consortium; MCC=Mayo Clinic Caucasians; TXC=Texas Caucasians; and TXA=Texas African American.

† Bold letter is reference allele used for estimating OR in Mantel-Hazenszel tests.

‡ In subgroup analysis of subjects with genotypes GG at SNP rs1051730, only SNP rs481134 was tested.

§In subgroup analysis of subjects with genotypes GG at SNP rs481134, only SNP rs1051730 was tested.

Table S3 Population attributable risks for SNPs rs481134 and rs1051730 on chromosome 15q24-25.1

Genotypes	**Freq in controls (p_i)**	**ORs***	**p_i (OR_i-1) (%)**
rs1051730			
GG	0.4406	1.00 (ref.) §	/
AA	0.1036	2.21	12.52
GA	0.4559	1.49	22.17
PAR (%)	/	/	25.76
rs481134			
CC	0.3067	1.00 (ref.)	/
TT	0.1824	1.36	6.63
TC	0.5109	1.17	8.57
PAR (%)	/	/	13.19
rs1051730-rs481134†			
G_C/G_C‡	0.0553	1.00 (ref.)	/
G_C/A_C	0.1482	1.42	6.20
G_C/G_T	0.2043	1.16	3.33
A_C/A_C	0.1032	2.01	10.45
G_T/A_C	0.3060	1.65	19.88
G_T/G_T	0.1829	1.35	6.44
PAR (%)	/	/	31.64

* Frequencies of genotypes and marginal ORs in the TXC population were used for estimating PAR (%) for single locus.

† Frequencies of genotypes and haplotype-specific ORs in the TXC population were used for jointly estimating PAR (%) for two loci (rs1051730 and rs481134).

‡ h1/h2 indicates a genotype composing two haplotypes h1 and h2.

§ The lowest-risk genotype was used as the reference to estimate PAR (%). See the Method section for details.

Table S4 Results from the logistic regression analysis modeling both SNPs rs1051730 and rs481134

SNPs	P-values			
	GELCC	MCC	TXC	TXA
Model-1*				
rs1051730	2.48×10^{-4}	4.70×10^{-3}	7.40×10^{-4}	5.13×10^{-4}
rs481134	0.029	0.055	0.321	0.171
Model-2†				
rs1051730	8.10×10^{-4}	5.82×10^{-3}	4.03×10^{-5}	7.69×10^{-5}
rs481134	0.036	0.049	0.050	0.044
rs1051730:rs481134	0.750	0.566	0.017	0.039
Model comparison‡				
AIC-Model-1§	338.0	2400.5	3833.7	709.8
AIC-Model-2	340.0	2402.2	3830.0	707.5
P-values††	0.752	0.567	0.016	0.039

* SNPs rs1051730 and rs481134 were incorporated into the logistic regression model simultaneously.

†Both main effect of SNPs rs1051730 and rs481134 and their interaction were incorporated into the logistic regression model simultaneously.

‡The goodness-of-fit of these two models was evaluated based on AIC and Chi-square test.

§ Akaike information criterion (AIC) for each model was estimated. Smaller values of AIC indicate the preferred model.

††P values from Chi-square testd by comparing model-1 with model-2. Model-2 generally fits better for the data than Model-1 based on Chi-square tests.