Genome-Wide Association Study of Lung Cancer in Korean Non-Smoking Women

Jin Hee Kim,[1,2] Kyunghee Park,[3]
Seon-Hee Yim,[4,5] Jin Eun Choi,[6]
Jae Sook Sung,[7,8] Ju-Yeon Park,[9,10]
Yi Young Choi,[6] Hyo-Sung Jeon,[6]
Jae Yong Park,[6,11] Hyoung Kyu Yoon,[12]
Yeul Hong Kim,[7,8] Byung Su Yoo,[9,10]
Young Tae Kim,[9,10] Hae-Jin Hu,[5,13]
Yeun-Jun Chung,[5,13] Ho Kim,[3]
Sook Whan Sung,[10,14] and Yun-Chul Hong[1,2]

[1]Department of Preventive Medicine, Seoul National University College of Medicine, Seoul; [2]Institute of Environmental Medicine, Seoul National University Medical Research Center, Seoul; [3]Department of Epidemiology and Biostatistics, Seoul National University School of Public Health, Seoul; [4]Department of Medical Humanities and Social Sciences, The Catholic University of Korea College of Medicine, Seoul; [5]Integrated Research Center for Genome Polymorphism, The Catholic University of Korea College of Medicine, Seoul; [6]Department of Biochemistry, School of Medicine, Kyungpook National University, Daegu; [7]Genomic Research Center for Lung and Breast/Ovarian Cancers, Korea University Anam Hospital, Seoul; [8]Division of Oncology/Hematology, Department of Internal Medicine, Korea University Anam Hospital, Seoul; [9]Department of Thoracic and Cardiovascular Surgery, Seoul National University Hospital, Seoul; [10]Department of Thoracic and Cardiovascular Surgery, Seoul National University College of Medicine, Seoul; [11]Department of Internal Medicine, School of Medicine, Kyungpook National University, Daegu; [12]Division of Pulmonary and Critical Care, Department of Internal Medicine, The Catholic University of Korea College of Medicine, Seoul; [13]Department of Microbiology, The Catholic University of Korea College of Medicine, Seoul; [14]Department of Thoracic and Cardiovascular Surgery, Seoul National University Bundang Hospital, Seongnam, Korea

Address for Correspondence:
Yun-Chul Hong, MD
Department of Preventive Medicine, Seoul National University College of Medicine, 103 Daehak-ro, Jongno-gu, Seoul 110-799, Korea
Tel: +82.2-740-8394, Fax: +82.2-747-4830
E-mail: ychong1@snu.ac.kr

Lung cancer in never-smokers ranks as the seventh most common cause of cancer death worldwide, and the incidence of lung cancer in non-smoking Korean women appears to be steadily increasing. To identify the effect of genetic polymorphisms on lung cancer risk in non-smoking Korean women, we conducted a genome-wide association study of Korean female non-smokers with lung cancer. We analyzed 440,794 genotype data of 285 cases and 1,455 controls, and nineteen SNPs were associated with lung cancer development ($P < 0.001$). For external validation, nineteen SNPs were replicated in another sample set composed of 293 cases and 495 controls, and only rs10187911 on 2p16.3 was significantly associated with lung cancer development (dominant model, OR of TG or GG, 1.58, $P = 0.025$). We confirmed this SNP again in another replication set composed of 546 cases and 744 controls (recessive model, OR of GG, 1.32, $P = 0.027$). OR and P value in combined set were 1.37 and < 0.001 in additive model, 1.51 and < 0.001 in dominant model, and 1.54 and < 0.001 in recessive model. The effect of this SNP was found to be consistent only in adenocarcinoma patients (1.36 and < 0.001 in additive model, 1.49 and < 0.001 in dominant model, and 1.54 and < 0.001 in recessive model). Furthermore, after imputation with HapMap data, we found regional significance near rs10187911, and five SNPs showed P value less than that of rs10187911 (rs12478012, rs4377361, rs13005521, rs12475464, and rs7564130). Therefore, we concluded that a region on chromosome 2 is significantly associated with lung cancer risk in Korean non-smoking women.

Key Words: Lung Neoplasms; Genome-Wide Association Study; Non-Smoking Women

INTRODUCTION

Lung cancer is the leading cause of cancer death worldwide with over 1 million deaths each year. Although cigarette smoking is the major cause of this malignancy, global statistics estimate that 15% of lung cancers in men and 53% in women are not directly attributable to smoking (1). The incidence of lung cancer in Korean women appears to be steadily increasing, as in Asian women generally, and the majority of these women are non-smokers. Moreover, lung cancer in Korean women appears to be predominantly adenocarcinoma. Gender, clinicopathological, and molecular differences in lung cancers arising in never-smokers and smokers indicate that lung cancer in non-smoking women might be a different disease (2, 3). Recent evidence for genetic influence on smoking behavior and nicotine dependence has prompted a search for susceptibility genes related with lung cancer development (4). However, no regions were replicated except for 5p15.33 as susceptibility loci for never-smoking lung cancer patients although three human genome regions at chromosome 5p15, 15q25, and 6p21 were found to be associated with susceptibility to lung cancer through several genome-wide association (GWA) studies (4-14). Furthermore, there was no information for Korean non-smoking women. For these reasons, we conducted a genome-wide association study (GWAS) of female non-smoking Koreans with lung cancer to identify low–penetrance alleles influencing lung cancer risk in Korean non-smoking women.

MATERIALS AND METHODS

Study populations

A total of 286 non-smoking women with newly diagnosed lung cancer at Kyungpook National University Hospital (30 cases), Korea University Anam Hospital (39 cases), Seoul National University Hospital (140 cases), and Seoul National University Bundang Hospital (77 cases) in Korea between 2001 and 2008 were recruited for GWAS. There were no age, histologic, or stage restrictions, and all lung cancer cases were histologically confirmed. The control subjects for GWAS were 1,462 non-cancer participants admitted to the Korean Association REsource (KARE) of Korea Centers for Disease Control and Prevention (KCDC). They were matched to the cases by sex and smoking status (case: control ≒ 1:5). For the replication studies, we recruited new population, 293 cases and 495 controls for replication 1 (70 cases and 100 controls in Kyungpook National University Hospital; 64 cases in Korea University Anam Hospital; 115 cases in Seoul National University Hospital; and 44 cases and 395 controls in Seoul National University Bundang Hospital) and 546 cases and 744 controls for replication 2 (84 cases and 43 controls in Kyungpook National University Hospital; 44 cases in Korea University Anam Hospital; 90 cases in Seoul National University

Hospital; 57 cases and 232 controls in Seoul National University Bundang Hospital; and 271 cases and 469 controls in Korea National Cancer Center). All participants were never-smoking women and information on age, sex, smoking history, and disease status for them was collected by trained interviewers working in each center.

GWA genotyping

DNA was extracted from whole blood sample using a QIAamp® DNA Blood Midi Kit (Qiagen, Valencia, CA, USA). SNP genotyping was performed using Affymetrix Genome-Wide Human SNP array 5.0 comprising 440,794 genome-wide SNPs (Affymetrix, Santa Clara, CA, USA) according to the manufacturer's protocol. Total genomic DNA (500 ng) were digested with *Nsp* I and *Sty* I restriction enzymes and ligated to specific adaptors which incorporate a universal PCR priming sequence. PCR amplification was performed with universal primers in a reaction optimized to amplify fragments between 200-1,100 base pairs. A fragmentation step then reduced the PCR product to segments of approximately 25-50 bp, which were then end-labeled using biotinylated nucleotides. The labeled product was hybridized to a chip, washed and detected. The images were analyzed using GCOS software (Affymetrix, Santa Clara, CA, USA). For quality control (QC), two algorithms were used: the QC call rates for each array exceeded 86% using the Dynamic Model algorithm, and genotype calls for each site were more than 98% using the birdseed v2 genotyping call algorithm.

GWA data mining

We genotyped 440,794 SNPs in 286 cases using Affymetrix GeneChip 5.0 and obtained chip data of 1,462 controls from KARE of KCDC. First, we checked relationships among subjects using graphical representation of relationship errors (GRR) software and excluded 8 subjects (7 controls from KARE and 1 case from Seoul National University Bundang Hospital) with mean of identical-by-state (IBS) allele sharing over a number of polymorphic loci for each pair of individuals more than 1.5 and finally analyzed 285 cases and 1,455 controls using PLINK v1.07 program. Cutoff P values of Hardy-Weinberg Equilibrium (HWE), call rate, and minor allele frequency were 0.001, 0.95, and 0.01, respectively, and 331,088 SNPs were used in the analysis. Because of our small sample size, we used a 10-fold cross-validation method (90% randomly repeated selection of total population for a validation, 10 trials) to protect against false positives in logistic regression analysis. In all analyses, we adjusted for age in the model. Finally we selected 19 SNPs under P value 0.0001 for replication in another sample sets after considering each genotype scatter plot. All nineteen SNPs were consistent in all ten times cross-validation under P value 0.001 and confirmed to be called accurately. A quantile-quantile plot using P values obtained from additive model in logistic regression anal-

yses revealed no difference of the observed *P* values and those expected by chance (Fig. 1). A small inflation factor (λ) of 1.019 (or 1.000, based on the bottom 90% of quantile-quantile plot) indicated that the association we found is unlikely to have resulted from population stratification. Population structure was evaluated by PLINK v1.07, and the quantile-quantile plot was generated using R 2.13.0. We used the MACH 1.0.16 software to impute untyped SNPs using the linkage disequilibrium information from the HapMap phase II database (CHB+JPT was used as reference set, released 17 July 2006).

Replication genotyping

From the GWA analysis, we selected 19 SNPs (rs10187911, rs-13005521, rs1484322, rs755927, rs2191876, rs2378688, rs303451, rs41393946, rs7928853, rs7305016, rs8041151, rs17258206, rs17-258247, rs9939608, rs7241996, rs10485620, rs2884026, rs925338, and rs8068946) associated with lung cancer development under cutoff *P* value 0.0001 and genotyped those SNPs in samples obtained from new populations (replication set 1 and replication set 2).
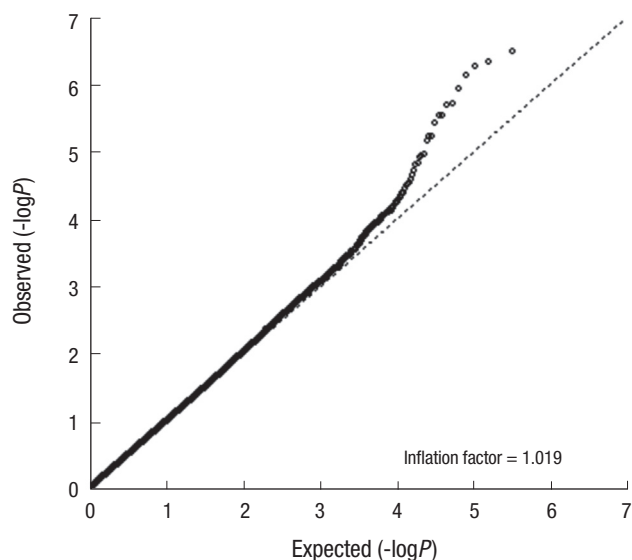

Fig. 1. Quantile-quantile plot of *P* value.

Replication 1. The 19 SNPs were evaluated in replication study. Of the 19 SNPs, 14 variants (rs10187911, rs13005521, rs2884026, rs1484322, rs755927, rs2191876, rs2378688, rs303451, rs41393946, rs7928853, rs7305016, rs8041151, rs17258206, rs17258247, rs-925338, rs9939608, rs8068946, rs7241996, and rs10485620) were selected by MassARRAY AssayDesign software package (v3.1) and genotyped using SEQUENOM's MassArray iPLEX technology (SEQUENOM I') with negative controls following the manufacturer's instructions. Genotype calls were made using the default post-processing calling parameters in SEQUENOM's Typer 4.0 software. The genotypes of five SNPs (rs13005521, rs-2191876, rs2884026, rs8068946, and rs925338) were performed using a polymerase chain reaction (PCR) and restriction fragment length polymorphism (RFLP) (Table 1) with negative controls. For quality control, the genotyping analysis was performed blind with regards to the subjects. In addition, approximately 5% of the samples were randomly selected to be genotyped again by SEQUENOM's MassArray iPLEX technology and PCR-RFLP, and the results showed 100% concordance.

Replication 2. Because only rs10187911 of replicated 19 SNPs showed statistical significance in replication set 1, only rs10-187911 was genotyped using the TaqMan fluorogenic 5' nuclease assay (AB I') in replication set 2. The final volume of polymerase chain reaction (PCR) was 5 μL, containing 10 ng of genomic DNA and 2.5 μL TaqMan Universal PCR Master Mix, with 0.13 μL of 40X Assay Mix (Assay ID, C_30098973_10). Thermal cycle conditions were as follows: 50℃ for 2 min to activate the uracil N-glycosylase and to prevent carry-over contamination, 95℃ for 10 min to activate the DNA polymerase, followed by 45 cycles of 95℃ for 15 sec and 60℃ for 1 min. All PCR were performed using 384-well plates by a Dual 384-Well GeneAmp PCR System 9700 (AB I') and the endpoint fluorescent readings were performed on an ABI PRISM 7900 HT Sequence Detection System (AB I'). Duplicate samples and negative controls were included to ensure accuracy of genotyping.

Statistical analyses for replication study

Unconditional logistic regression was used to estimate odds ratios (OR) and 95% confidence interval (CI) after adjustment for

Table 1. Primers and PCR conditions for amplification and RFLP

rs N	Direction	Sequence	Annealing temperature (℃)	Restriction enzyme
rs13005521	Forward Reverse	5´-GGATTTTCATATAGAAGAGAAGGAG-3´ 5´-CTGTGGGAATGTAGGTTAGTTC-3´	58	*Nla*IV
rs2191876	Forward Reverse	5´-CGCAAGAATCTTAAAACCTCCTCTG-3´ 5´-GAGTGTCATCAGAAAGATACATCAGTC-3´	60	*HPyCH4*III
rs2884026	Forward Reverse	5´-TGCAGATATTAATCTGGTTATC-3´ 5´-AATTGTTCATAACTCGATAGTC-3´	52.8	*Hpy188*III
rs8068946	Forward Reverse	5´-TGCAGATATTAATCTGGTTATC-3´ 5´-AATTGTTCATAACTCGATAGTC-3´	56	*Hpy188*III
rs925338	Forward Reverse	5´-CTTTCATCCTCTTCTCTAGTGTG-3´ 5´-CTTCCAGGACAGCTTCTGC-3´	60	*Msp*I

PCR, polymerase chain reaction; RFLP, restriction fragment length polymorphism.

age in lung cancer patients and control subjects. To find out whether each SNP site are on the HWE, the distributions of observed genotype frequency and expected genotype frequency calculated from observed allele frequency were compared using χ^2 test. SAS, version 9.1 (SAS institute Inc., Cary, NC, USA), was used for statistical analysis.

Ethics statement

The study protocol was approved by the institutional review board at Seoul National University Hospital (IRB No. C-0903-037-274), and written informed consent was provided by all study participants.

RESULTS

We analyzed 285 cases and 1,455 controls using Affymetrix Gene-Chip 5.0. To protect against false positive, we used 0.001, 0.95, and 0.01 as cutoff P values of HWE, call rate, and minor allele frequency, respectively, as well as 10-fold cross-validation method, and then considered each genotype scatter plot. Finally we selected 19 SNPs under P value 0.0001 for replication in another sample sets. Table 2 shows summary of 19 SNPs significantly associated with lung cancer development in GWAS. For validation, the nineteen SNPs were replicated in another sample set, replication 1, which was composed of 293 cases and 495 controls, and only one SNP (rs10187911 on p16.3 of chromosome 2) was significantly associated with lung cancer development in dominant model (OR of TG or GG, 1.58, P = 0.025) (Ta-

ble 3). We reconfirmed this SNP in another sample set, replication 2, composed of 546 cases and 744 controls (total 1,290) (recessive model, OR of GG, 1.32, P = 0.027) (Table 4). OR and P value in combined set (total set) were 1.37 and < 0.001 for additive model, 1.51 and < 0.001 for dominant model, and 1.54 and < 0.001 for recessive model (Table 4). We also assessed the effect of the SNP by center (Fig. 2). All five centers showed consistent results both in total population and in adenocarcinoma cases (additive model, OR$_{total}$, 1.37 vs OR$_{adenocarcinoma}$, 1.36).

Because we confirmed the effect of rs10187911 on lung cancer development in all sets of GWAS, replication 1, replication 2, and total population, we tried to show regional plot near rs10-187911 with genotypes estimated using HapMap data to increase the spectrum of variants tested for associations in this region (Fig. 3). After imputation, we found regional significance near rs10187911, and all rs10187911 and six SNPs existing in the same intron of *NRXN1* gene with rs10187911 showed P value < 0.001. Among these six SNPs existing in the same intron of *NRXN1* gene with rs10187911, five SNPs (rs12478012, rs4377361, rs13005521, rs12475464, and rs7564130) showed P value less than that of rs10187911 and one SNP (rs10048731) showed P value more than that of rs10187911.

DISCUSSION

We found a new SNP (rs10187911) at 2p16.3 as a susceptibility marker to lung cancer in never-smoking Korean females. Since most cases in our population were adenocarcinoma patients

Table 2. Summary of nineteen SNPs significantly associated with lung cancer development in GWAS

rs N	Chromosome cytoband	Gene	Base pair position in chromosome	Ref. allele	Additive model			Dominant model			Recessive model		
					OR	95% CI (Lower-Upper)	P value	OR	95% CI (Lower-Upper)	P value	OR	95% CI (Lower-Upper)	P value
rs10187911	2p16.3	*NRXN1*	50647947	G	1.47	1.22-1.78	< 0.001	1.71	1.25-2.34	< 0.001	1.66	1.21-2.26	0.002
rs13005521	2p16.3	*NRXN1*	50595856	C	1.52	1.25-1.84	< 0.001	1.75	1.25-2.46	0.001	1.73	1.28-2.33	< 0.001
rs2884026	2q32.3	*TMEFF2*	192930360	C	1.38	1.13-1.69	0.002	1.84	1.36-2.48	< 0.001	1.12	0.76-1.65	0.576
rs1484322	4p14	*UBE2K*, C4orf34	39671142	C	0.67	0.55-0.82	< 0.001	0.75	0.57-1.00	0.048	0.34	0.21-0.54	< 0.001
rs755927	4p16.2	*CYTL1, MSX1*	4943456	A	0.53	0.39-0.72	< 0.001	0.51	0.36-0.70	< 0.001	0.41	0.12-1.40	0.154
rs2191876	7p14.3	*PDE1C*	31882928	A	2.09	1.45-3.00	< 0.001	2.12	1.46-3.09	< 0.001	3.60	0.35-37.63	0.284
rs2378688	9q21.33	*AGTPBP1, NTRK2*	88118037	C	0.66	0.54-0.81	< 0.001	0.66	0.50-0.87	0.003	0.44	0.28-0.68	< 0.001
rs303451	10p11.23	*MAP3K8, LYZL2*	30777280	A	2.67	1.68-4.22	< 0.001	2.67	1.68-4.22	< 0.001	-	-	-
rs41393946	11p14.1	*KCNA4, METT5D1*	29017564	C	1.57	1.27-1.95	< 0.001	1.83	1.39-2.40	< 0.001	1.53	0.92-2.56	0.105
rs7928853	11p14.1	*KCNA4, METT5D1*	29029556	C	1.53	1.24-1.89	< 0.001	1.78	1.35-2.34	< 0.001	1.46	0.87-2.44	0.149
rs7305016	12q13.13	*SMAGP*	51642543	C	0.65	0.53-0.81	< 0.001	0.64	0.49-0.84	0.001	0.44	0.27-0.72	0.001
rs8041151	15q26.2	*MCTP2, RGMA*	94626455	A	1.51	1.24-1.83	< 0.001	1.79	1.31-2.46	< 0.001	1.67	1.21-2.30	0.002
rs17258206	16q21	LOC644649, *GOT2*	59381595	A	0.71	0.58-0.87	< 0.001	0.56	0.42-0.74	< 0.001	0.86	0.58-1.28	0.463
rs17258247	16q21	LOC644649, *GOT2*	59394794	A	0.70	0.57-0.86	< 0.001	0.56	0.43-0.74	< 0.001	0.83	0.56-1.24	0.368
rs925338	16q21	LOC644649, *GOT2*	59349532	A	0.71	0.58-0.87	0.001	0.56	0.43-0.74	< 0.001	0.88	0.59-1.30	0.513
rs9939608	16q21	LOC644649, *GOT2*	59392020	C	0.70	0.57-0.85	< 0.001	0.56	0.42-0.74	< 0.001	0.81	0.54-1.22	0.319
rs8068946	17q24.3	*SLC39A11*	70712513	A	0.73	0.59-0.89	0.002	0.57	0.43-0.75	< 0.001	0.93	0.63-1.38	0.721
rs7241996	18q12.2	LOC647946, *MIR4318*	35490215	A	0.53	0.40-0.70	< 0.001	0.52	0.38-0.71	< 0.001	0.20	0.05-0.87	0.032
rs10485620	20q13.2	*ZFP64, SALL4*	50565168	C	2.00	1.53-2.61	< 0.001	2.07	1.54-2.79	< 0.001	3.76	1.56-9.08	0.003

OR, odds ratio; CI, confidence interval; GWAS, genome-wide association study.

Table 3. Summary of nineteen SNPs in another sample set, replication 1

rs N	Chromosome cytoband	Gene	Base pair position in chromosome	Ref. allele	Additive model OR	Additive model P value	Dominant model OR	Dominant model P value	Recessive model OR	Recessive model P value
rs10187911	2p16.3	NRXN1	50647947	G	1.23	0.067	1.58	0.025	1.16	0.392
rs13005521	2p16.3	NRXN1	50595856	C	1.08	0.477	1.16	0.416	1.07	0.724
rs2884026	2q32.3	TMEFF2	192930360	C	1.05	0.667	1.05	0.818	1.08	0.660
rs1484322	4p14	UBE2K, C4orf34	39671142	C	0.93	0.544	0.88	0.549	0.93	0.681
rs755927	4p16.2	CYTL1, MSX1	4943456	A	0.99	0.967	1.24	0.644	0.96	0.821
rs2191876	7p14.3	PDE1C	31882928	A	1.06	0.798	1.13	0.591	-	-
rs2378688	9q21.33	AGTPBP1, NTRK2	88118037	C	0.93	0.529	0.96	0.817	0.82	0.381
rs303451	10p11.23	MAP3K8, LYZL2	30777280	A	0.72	0.332	-	-	0.72	0.332
rs41393946	11p14.1	KCNA4, METT5D1	29017564	C	0.98	0.875	1.29	0.487	0.93	0.619
rs7928853	11p14.1	KCNA4, METT5D1	29029556	C	1.00	0.971	1.40	0.338	0.92	0.621
rs7305016	12q13.13	SMAGP	51642543	C	1.18	0.129	1.26	0.144	1.24	0.322
rs8041151	15q26.2	MCTP2, RGMA	94626455	A	0.92	0.452	0.87	0.417	0.93	0.710
rs17258206	16q21	LOC644649, GOT2	59381595	A	0.89	0.338	0.90	0.662	0.85	0.306
rs17258247	16q21	LOC644649, GOT2	59394794	A	0.87	0.245	0.84	0.457	0.84	0.278
rs925338	16q21	LOC644649, GOT2	59349532	A	0.92	0.450	0.90	0.640	0.89	0.471
rs9939608	16q21	LOC644649, GOT2	59392020	C	1.07	0.562	1.14	0.437	1.01	0.983
rs8068946	17q24.3	SLC39A11	70712513	A	0.94	0.601	0.92	0.584	0.94	0.780
rs7241996	18q12.2	LOC647946, MIR4318	35490215	A	1.12	0.415	1.73	0.189	1.07	0.689
rs10485620	20q13.2	ZFP64, SALL4	50565168	C	0.93	0.669	0.89	0.538	2.01	0.419

OR, odds ratio.

**Table 4.** Summary of SNP (rs10187911) replicated in other validation sets, replication 1 and replication 2

Set	Geno-type	No. (%) of cases	No. (%) of adenocarcinoma	No. (%) of controls	Model	All cases OR	All cases 95% CI (Lower-Upper)	All cases P value	Adenocarcinoma only OR	Adenocarcinoma only 95% CI (Lower-Upper)	Adenocarcinoma only P value
GWAS (285 cases and 1,455 controls)	GG	80 (28.2)	67 (27.1)	283 (19.5)	Additive	1.47	1.22-1.78	< 0.001	1.40	1.15-1.71	0.001
	GT	138 (48.6)	119 (48.2)	699 (48.0)	Dominant	1.71	1.25-2.34	< 0.001	1.59	1.14-2.20	0.006
	TT	66 (23.2)	61 (24.7)	473 (32.5)	Recessive	1.66	1.21-2.26	0.002	1.56	1.12-2.18	0.008
Replication 1 (293 cases and 495 controls)	GG	93 (33.3)	87 (33.0)	150 (30.7)	Additive	1.23	0.99-1.53	0.067	1.24	0.99-1.55	0.060
	GT	140 (50.2)	135 (51.1)	223 (45.7)	Dominant	1.58	1.06-2.35	0.025	1.65	1.10-2.48	0.016
	TT	46 (16.5)	42 (15.9)	115 (23.6)	Recessive	1.16	0.83-1.61	0.392	1.15	0.82-1.61	0.428
Replication 2 (546 cases and 744 controls)	GG	180 (33.6)	172 (34.3)	207 (28.2)	Additive	1.15	0.98-1.34	0.082	1.16	0.99-1.36	0.064
	GT	240 (44.9)	222 (44.2)	360 (49.0)	Dominant	1.08	0.83-1.42	0.573	1.08	0.82-1.42	0.578
	TT	115 (21.5)	108 (21.5)	167 (22.8)	Recessive	1.32	1.03-1.67	0.027	1.35	1.06-1.72	0.017
Total (1,124 cases and 2,694 controls)	GG	353 (32.1)	326 (32.2)	640 (23.9)	Additive	1.37	1.23-1.51	< 0.001	1.36	1.23-1.52	< 0.001
	GT	518 (47.2)	476 (47.0)	1,282 (47.9)	Dominant	1.51	1.26-1.79	< 0.001	1.49	1.25-1.79	< 0.001
	TT	227 (20.7)	211 (20.8)	755 (28.2)	Recessive	1.54	1.31-1.81	< 0.001	1.54	1.31-1.82	< 0.001

OR, odds ratio; CI, confidence interval; GWAS, genome-wide association study.

(92.4%), this SNP was significantly associated only with adenocarcinoma.

This SNP (rs1884396) is located on intron of NRXN1 gene, one of the neurexins that function as cell adhesion molecules during synaptogenesis and take part in intracellular signaling (15). NRXN1 is extraordinarily large, occupying 1,112,039 bp of DNA – nearly 0.04% of the entire human genome. Neurexin transcripts undergo alternative splicing at several sites (15). Each site of alternative splicing is defined as a position in the mRNA where sequence variations have been observed. Therefore, the SNP (rs1884396) on intron of NRXN1 gene may be responsible for regulation of splicing events, leading to protein diversity. In fact, NRXN1 is related with hemangiomas as well as autism spectrum disorders, mental retardation, language development delays, hypotonia, and the VACTERL association (16-19). Moreover, Nussbaum et al. suggested involvement of NRXN1 in nicotine dependence (20), associated with the 15q25 region in a previous lung cancer GWA study. Mutation of epidermal growth factor receptor (EGFR) is known to be important in chemotherapy of Asian lung cancer patients. Interestingly, NRXN1 encodes alpha type of isoform using upstream promoter and this isoform has an epidermal growth factor (EGF)-like sequence (16) which can bind to EGFR. Furthermore, lung injury by several chemicals may in part be inferred via curated interactions between NRXN1 and chemicals such as tetrachlorodibenzodioxin (TCDD), ozone (O_3), and chlorine (Cl) (Comparative Toxicogenomics Database, http://ctd.mdibl.org/detail.go?type = gene&acc = 9378; http://ctd.mdibl.org/de-
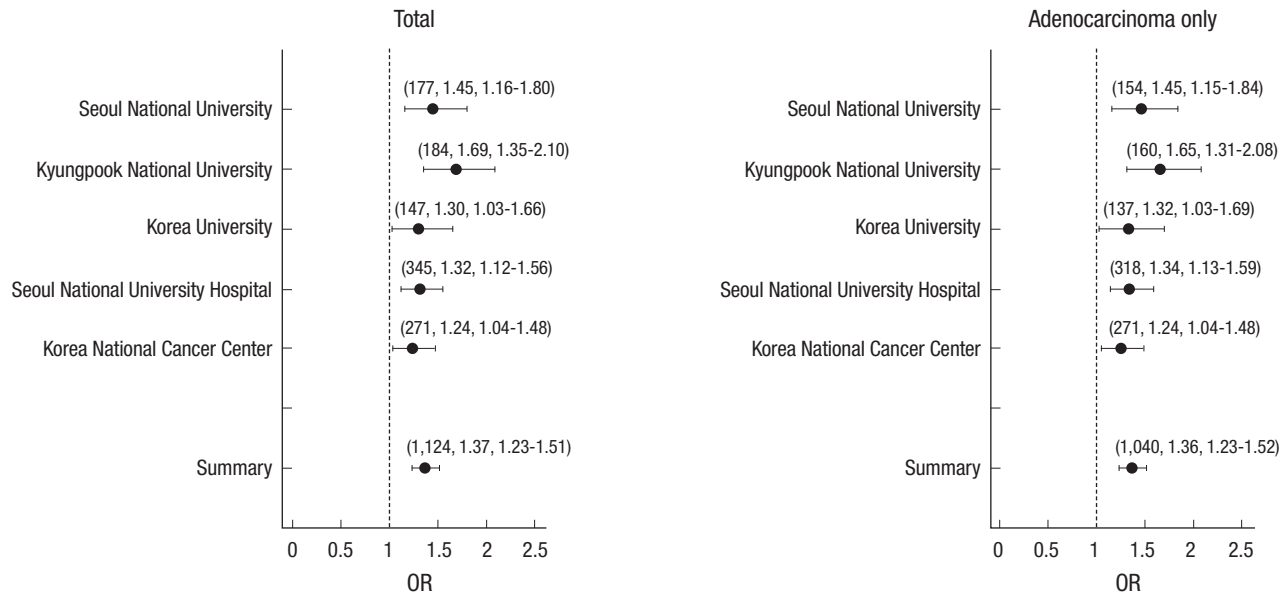
Fig. 2. Summary of odds ratio (OR) and confidence interval (CI) by center. The number of cases, OR, and 95% CI were depicted in order in brackets.
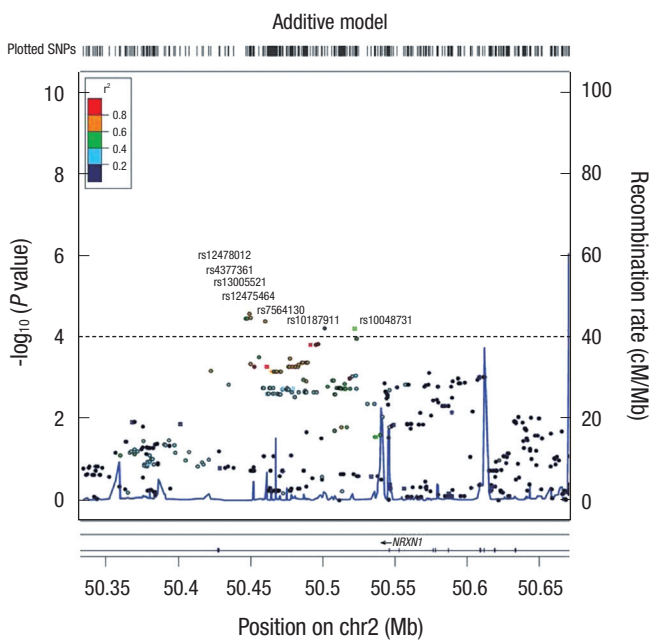
Fig. 3. Regional plot surrounding rs10187911.

tail.go?type = gene&acc = 9378&view = ixn).

In this study, we imputed other SNPs not explored in our GWAS on lung cancer development using HapMap data and found significant associations with other SNPs surrounding rs10187911 (existing in the same intron) in *NRXN1*, indicating that the rs10187911 we found is a true indicator related with lung cancer development. However, we need to confirm which one among these significant SNPs is a real marker of lung cancer development in the future.

In this study, we did not confirm the previously reported SNPs

from GWA studies in Asian population and found inconsistent results among several reports (21-25). A study of lung cancer in never-smoking Asian women (21) demonstrated that common genetic variants (rs2736100) in the TERT-CLPTM1L locus on chromosome 5p15.33 are associated with risk for lung adeno-carcinoma. Li et al. (22) demonstrated novel genetic variants at 13q31.3 as susceptibility marker to never-smoker lung cancer in mostly American subjects. However, this study did not repli-cate significant association of 5p15.33 with lung cancers in nev-er-smokers. Ahn et al. (23) suggested the 18p11.22 region as novel lung cancer susceptibility locus in never-smokers in Ko-rean population. This study observed a *P* value of 0.008 for rs-2736100 with no genome-wide significance level. Unfortunate-ly, we could not evaluate the significance of rs2736100 because the locus was not included in 497,345 markers of our GWA study. We tried to estimate the effect of SNPs (rs2352028 at 13q31.3 and rs11080466 and rs11663246 at 18p11.22) reported from the Li et al.'s and Ahn et al.'s studies. However, rs2352028 and rs11-663246 did not reach significance in our data (*P* = 0.838 for rs-2352028 and *P* = 0.421 for rs11663246), and rs11080466 was not included in our markers. Another study of lung cancer in Han Chinese (24) suggested three new lung cancer susceptibility SNPs, rs753955 at 13q12.12 and rs17728461 and rs36600 at 22q-12.2. However, rs753955 and rs17728461 did not reach signifi-cance in our data (*P* = 0.382 for rs753955 and *P* = 0.537 for rs-17728461), and rs36600 was not included in our markers. A re-cent study of lung cancer in never-smoking women in Asia (25) indicated that rs7086803 of the *VTI1A* gene on chromosome 10q25.2 is associated with lung carcinogenesis. The inconsis-tent observations in several GWA studies might be attributed to different environmental exposures, ethnic differences (21, 22,

24), mixed histology (23, 24), sex (22-24), and smoking history (23, 24), or different SNP markers (21, 22, 25).

The incidence of lung cancer in non-smoking Asian women appears to be increasing and is predominantly adenocarcinoma. Based on clinico-pathological and molecular characteristics of lung cancer arising in never-smoking females, adenocarcinoma has been considered a different disease from other lung cancer types. In the present study, all of our subjects were Korean females and most of the subjects had adenocarcinoma. Therefore, rs10187911 located on intron of NRXN1 gene at 2p16.3 found in this study may be an important marker for adenocarcinoma development in never-smoking females.

There were some limitations to this study. First, the sample size in the present study was small. However, to protect against false positives resulting from small sample size, we used a 10-fold cross-validation method and confirmed that rs10187911 was replicated in all 10 trials with randomly repeated selection of total population. Furthermore, we replicated the same analyses in other sample sets, replication 1 and 2, and obtained consistent results for rs10187911. Second, we did not adjust for center and histology regardless of differences in histology by center. However, when we stratified the total population in the present study by center only for adenocarcinoma, we did not find any differences by center in the association between rs10187911 and lung cancer risk. Moreover, we found a small inflation factor indicating that the association we found is unlikely to result from difference by center.

Although the association of NRXN1 and lung cancer has not been explored in epidemiologic study, previously reported data (15-20) support a role for this gene in the tumorigenesis of lung cancer. Nonetheless, the exact import of NRXN1 in lung cancer development remains largely unknown, and further biological and epidemiologic studies are needed.

In conclusion, we have identified a novel genetic locus at 2p16.3 region associated with susceptibility of adenocarcinoma in Korean never-smoking females. Further replication studies in larger populations are necessary to clarify our hypothesis.

DISCLOSURE

The authors have no conflicts of interest to disclose.

REFERENCES

1. Parkin DM, Bray F, Ferlay J, Pisani P. *Global cancer statistics, 2002.* CA Cancer J Clin 2005; 55: 74-108.

2. Sun S, Schiller JH, Gazdar AF. *Lung cancer in never smokers: a different disease. Nat Rev Cancer 2007; 7: 778-90.*

3. Torok S, Hegedus B, Laszlo V, Hoda MA, Ghanim B, Berger W, Klepetko W, Dome B, Ostoros G. *Lung cancer in never smokers. Future Oncol 2011; 7: 1195-211.*

4. Spitz MR, Amos CI, Dong Q, Lin J, Wu X. *The CHRNA5-A3 region on chromosome 15q24-25.1 is a risk factor both for nicotine dependence and for lung cancer. J Natl Cancer Inst 2008; 100: 1552-6.*

5. Amos CI, Wu X, Broderick P, Gorlov IP, Gu J, Eisen T, Dong Q, Zhang Q, Gu X, Vijayakrishnan J, 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.*

6. Hung RJ, McKay JD, Gaborieau V, Boffetta P, Hashibe M, Zaridze D, Mukeria A, Szeszenia-Dabrowska N, Lissowska J, Rudnai P, et al. *A susceptibility locus for lung cancer maps to nicotinic acetylcholine receptor subunit genes on 15q25. Nature 2008; 452: 633-7.*

7. McKay JD, Hung RJ, Gaborieau V, Boffetta P, Chabrier A, Byrnes G, Zaridze D, Mukeria A, Szeszenia-Dabrowska N, Lissowska J, et al. *Lung cancer susceptibility locus at 5p15.33. Nat Genet 2008; 40: 1404-6.*

8. Rafnar T, Sulem P, Stacey SN, Geller F, Gudmundsson J, Sigurdsson A, Jakobsdottir M, Helgadottir H, Thorlacius S, Aben KK, et al. *Sequence variants at the TERT-CLPTM1L locus associate with many cancer types. Nat Genet 2009; 41: 221-7.*

9. Truong T, Hung RJ, Amos CI, Wu X, Bickeböller H, Rosenberger A, Sauter W, Illig T, Wichmann HE, Risch A, et al. *Replication of lung cancer susceptibility loci at chromosomes 15q25, 5p15, and 6p21: a pooled analysis from the International Lung Cancer Consortium. J Natl Cancer Inst 2010; 102: 959-71.*

10. Landi MT, Chatterjee N, Yu K, Goldin LR, Goldstein AM, Rotunno M, Mirabello L, Jacobs K, Wheeler W, Yeager M, et al. *A genome-wide association study of lung cancer identifies a region of chromosome 5p15 associated with risk for adenocarcinoma. Am J Hum Genet 2009; 85: 679-91.*

11. Wang Y, Broderick P, Webb E, Wu X, Vijayakrishnan J, Matakidou A, Qureshi M, Dong Q, Gu X, Chen WV, et al. *Common 5p15.33 and 6p21.33 variants influence lung cancer risk. Nat Genet 2008; 40: 1407-9.*

12. Rafnar T, Sulem P, Besenbacher S, Gudbjartsson DF, Zanon C, Gudmundsson J, Stacey SN, Kostic JP, Thorgeirsson TE, Thorleifsson G, et al. *Genome-wide significant association between a sequence variant at 15q15.2 and lung cancer risk. Cancer Res 2011; 71: 1356-61.*

13. Kohno T, Kunitoh H, Shimada Y, Shiraishi K, Ishii Y, Goto K, Ohe Y, Nishiwaki Y, Kuchiba A, Yamamoto S, et al. *Individuals susceptible to lung adenocarcinoma defined by combined HLA-DQA1 and TERT genotypes. Carcinogenesis 2010; 31: 834-41.*

14. Pande M, Spitz MR, Wu X, Gorlov IP, Chen WV, Amos CI. *Novel genetic variants in the chromosome 5p15.33 region associate with lung cancer risk. Carcinogenesis 2011; 32: 1493-9.*

15. Rowen L, Young J, Birditt B, Kaur A, Madan A, Philipps DL, Qin S, Minx P, Wilson RK, Hood L, et al. *Analysis of the human neurexin genes: alternative splicing and the generation of protein diversity. Genomics 2002; 79: 587-97.*

16. Ching MS, Shen Y, Tan WH, Jeste SS, Morrow EM, Chen X, Mukaddes NM, Yoo SY, Hanson E, Hundley R, et al. *Deletions of NRXN1 (neurexin-1) predispose to a wide spectrum of developmental disorders. Am J Med Genet B Neuropsychiatr Genet 2010; 153B: 937-47.*

17. Glessner JT, Wang K, Cai G, Korvatska O, Kim CE, Wood S, Zhang H, Estes A, Brune CW, Bradfield JP, et al. *Autism genome-wide copy number variation reveals ubiquitin and neuronal genes. Nature 2009; 459: 569-73.*

18. Bucan M, Abrahams BS, Wang K, Glessner JT, Herman EI, Sonnenblick

LI, Alvarez Retuerto AI, Imielinski M, Hadley D, Bradfield JP, et al. *Genome-wide analyses of exonic copy number variants in a family-based study point to novel autism susceptibility genes. PLoS Genet 2009; 5: e1000536.*

19. Pedrosa E, Kaushik S, Lachman HM. *ChIP-chip analysis of neurexins and other candidate genes for addiction and neuropsychiatric disorders. J Neurogenet 2010; 24: 5-17.*

20. Nussbaum J, Xu Q, Payne TJ, Ma JZ, Huang W, Gelernter J, Li MD. *Significant association of the neurexin-1 gene (NRXN1) with nicotine dependence in European- and African-American smokers. Hum Mol Genet 2008; 17: 1569-77.*

21. Hsiung CA, Lan Q, Hong YC, Chen CJ, Hosgood HD, Chang IS, Chatterjee N, Brennan P, Wu C, Zheng W, et al. *The 5p15.33 locus is associated with risk of lung adenocarcinoma in never-smoking females in Asia. PLoS Genet 2010; 6: e1001051.*

22. Li Y, Sheu CC, Ye Y, de Andrade M, Wang L, Chang SC, Aubry MC, Aakre JA, Allen MS, Chen F, et al. *Genetic variants and risk of lung cancer in never smokers: a genome-wide association study. Lancet Oncol 2010; 11: 321-30.*

23. Ahn MJ, Won HH, Lee J, Lee ST, Sun JM, Park YH, Ahn JS, Kwon OJ, Kim H, Shim YM, et al. *The 18p11.22 locus is associated with never smoker non-small cell lung cancer susceptibility in Korean populations. Hum Genet 2012; 131: 365-72.*

24. Hu Z, Wu C, Shi Y, Guo H, Zhao X, Yin Z, Yang L, Dai J, Hu L, Tan W, et al. *A genome-wide association study identifies two new lung cancer susceptibility loci at 13q12.12 and 22q12.2 in Han Chinese. Nat Genet 2011; 43: 792-6.*

25. Lan Q, Hsiung CA, Matsuo K, Hong YC, Seow A, Wang Z, Hosgood HD 3rd, Chen K, Wang JC, Chatterjee N, et al. *Genome-wide association analysis identifies new lung cancer susceptibility loci in never-smoking women in Asia. Nat Genet 2012; 44: 1330-5.*