Official Journal of the Asian Pacific Society of Respirology
Respirology

ORIGINAL ARTICLE

Genome-wide association studies identify locus on 6p21 influencing lung function in the Korean population

WOO JIN KIM,[1]* MI KYEONG LEE,[3]* CHOL SHIN,[4] NAM HAN CHO,[5] SANG DO LEE,[6] YEON-MOK OH[6] AND JOOHON SUNG[2,3]

[1]*Department of Internal Medicine and Environmental Health Center, Kangwon National University Hospital, School of Medicine, Kangwon National University, Chuncheon,* [2]*Complex Disease Genetic Epidemiology Branch, Department of Epidemiology, Graduate School of Public Health,* [3]*Institute of Health and Environment, Seoul National University,* [6]*Department of Pulmonary and Critical Care Medicine, and Clinical Research Center for Chronic Obstructive Airway Diseases, Asan Medical Center, University of Ulsan College of Medicine, Seoul,* [4]*Division of Pulmonary and Critical Care Medicine, Department of Internal Medicine, Korea University Ansan Hospital, Ansan, and* [5]*Department of Preventive Medicine, Ajou University School of Medicine, Suwon, Korea*

ABSTRACT

Background and objective: **Loss of lung function is an important chronic obstructive pulmonary disease phenotype and decreased forced expiratory volume in 1 s (FEV_1) is an independent risk factor of morbidity and mortality. Genome-wide association studies (GWAS) identifying genetic variants underlying lung function have been performed mostly in Caucasian populations. In this study, we aimed to identify genetic variants influencing lung function in a Korean population.**
Methods: **GWAS on lung function (FEV_1 and FEV_1/forced vital capacity (FVC) ratio) were performed in two cohort studies. A population-based cohort, the Korean Association Resource phase 3 (KARE3) (6223 subjects), served as a discovery set. The replication analysis was performed in a family-based cohort, the Healthy Twin Study (HTS; 2730 subjects). Dense single-nucleotide polymorphism array data from each study were imputed and used for genetic analysis.**
Results: **At the discovery phase, variants in 6p21 and 17q24 showed the strongest association with FEV_1/FVC ratio and FEV_1. Several variants in *FAM13A* on 4q22 locus exhibited positive association with FEV_1/FVC ratio. In the replication set, *PPT2* in the 6p21 region showed significant association with lung function in the HTS, although the 4q22 locus and the 17q24 locus were not replicated.**
Conclusions: **We identified that *PPT2* on chromosome 6p21 is associated with loss of lung function in the Korean population.**

Key words: genetics, genome-wide association study, lung function, single-nucleotide polymorphism.

Abbreviations: COPD, chronic obstructive pulmonary disease; FEV_1, forced expiratory volume in 1 s; FVC, forced vital capacity; GWAS, genome-wide association study; HTS, Healthy Twin Study; KARE3, Korean Association Resource phase 3; RAGE, receptor for advanced glycation end-products; SNP, single-nucleotide polymorphism.

SUMMARY AT A GLANCE

The aim of this study was to identify genetic determinants of lung function in a Korea population. Locus on chromosome 6p21 was identified to regulate lung function in two cohort studies. Additionally, chromosome 17q24 near *SOX9* showed a strong association in one study, although it was not replicated.

INTRODUCTION

Lung function measured by spirometry demonstrates a broad range of variation both in general population and in patients with respiratory illness. Forced expiratory volume in 1 s (FEV_1) and FEV_1/forced vital capacity (FVC) ratio have been the most frequently studied phenotypes among lung function, since these phenotypes are essential in both diagnosis and prognosis of lung and airway diseases. FEV_1/FVC ratio is an indicator of obstructive lung disease and also an essential criterion for the diagnosis of chronic obstructive

Correspondence: Joohon Sung, Complex Disease Genetic Epidemiology Branch, Department of Epidemiology, School of Public Health, Seoul National University, Seoul 151-742, Korea. Email: jsung@snu.ac.kr

*These authors contributed equally to this work.

Conflict of interest: Y.M.O. has been an investigator in university-sponsored studies (Asan Institute for Life Science, University of Ulsan College of Medicine) and an industry-sponsored study (MSD Korea and AstraZeneca Korea) and has participated as a speaker at scientific meetings organized and financed by various pharmaceutical companies (Handok, GlaxoSmithKline, AstraZeneca Korea, MSD Korea and Boehringer Ingelheim) and a magazine company (Korea Doctors' Weekly).

S.D.L. serves as a consultant to GlaxoSmithKline and has participated as a speaker at scientific meetings organized and financed by various pharmaceutical companies (GlaxoSmithKline, AstraZeneca Korea and Boehringer Ingelheim).

Received 23 April 2013; invited to revise 5 July and 12 September 2013; revised 30 July and 9 October 2013; accepted 15 October 2013 (Associate Editor: Robert Young).

pulmonary disease (COPD).[1] FEV_1 is used to determine the COPD severity and is an independent predictor of morbidity and mortality risk of all causes.[2] Spirometry measures of lung function have been demonstrated to have a genetic component. Genetic epidemiological studies of lung function have suggested that lung function phenotypes have significant degree of familial aggregations.[3] Recent reports on the heritability of FEV_1 and FEV_1/FVC ratio ranged between 30% and 60%,[4,5] which suggests a polygenic involvement in controlling lung function.

To date, four genome-wide association studies (GWAS) evaluating genetic relationships with lung function have been published.[6–9] So far, 28 loci have been identified including single-nucleotide polymorphism (SNP) on chromosome 4q22 in *FAM13A* and 6p21 near *AGER-PPT2*, which were genome-wide significant for FEV_1 or FEV_1/FVC in previous studies. Additional genes associated with lung function were identified after including interaction with smoking.[10] Recently, lung function genes were identified in a paediatric cohort.[11] The inventory of associated genes will provide an insight into the genetic architecture constituting lung function. Many common polygenic disorders may represent the extreme manifestation of continuous quantitative traits of which liability is normally distributed in a population. To support this hypothesis, it is shown that several 'lung function genes' identified through GWAS in the general population were shown to be associated with the risk of COPD as well. Those genes may provide a clue to find therapeutic targets or biomarkers of COPD across its very early to full-blown stages. However, as most GWAS of lung function were performed in Caucasian populations, studies in other ethnic groups are needed to fully capture the genetic architecture underlying lung function.

In this study, we performed GWAS of lung function in two cohorts of the general population in Korea: the Korean Association Resource phase 3 (KARE3), which is constituted of unrelated individuals, and the Healthy Twin Study (HTS), a family-based cohort.

METHODS

Study populations

We conducted a GWAS using data from the KARE3 (6223 subjects) and replicated the results in the HTS (2720 subjects). The Korean Association Resource project was initiated in 2007 to undertake GWAS analysis among 10 038 participants in the population-based cohort of Ansung rural area and Ansan city of South Korea. The cohort was initiated in 2001 and was designed to allow longitudinal prospective study. Participants were examined every 2 years, and the third follow-up study was completed in 2008. The KARE3 data were obtained from the third phenotype data. More than 260 traits have been examined through epidemiological surveys, physical examinations and laboratory tests. GWAS of the quantitative traits of the first phenotype in this cohort was published in 2009.[12]

The HTS is a family-based cohort of adult same-gender twin pairs and their family members based on the National Twin-Family Registry of Korea, all of whom completed a health examination and in-depth surveys at two general hospitals since 2005. The study design and protocols of the HTS were previously described in detail.[13,14] To ascertain the zygosity, either 16 microsatellites or standardized questionnaire was used. Family relationships taken by questionnaire and personal interviews were compared with information from the genetic marker. The studies were approved by the appropriate institutional review board at the respective institutions.

Genotyping

Genomic DNAs isolated from peripheral blood drawn from the KARE3 participants were genotyped on the Affymetrix Genome-Wide Human array 5.0 (Affymetrix, Inc., Santa Clara, CA, USA). The genotyping and quality control details of KARE3 have been previously described.[12]

Genotyping of twin-family participants was performed using an Affymetrix Genome Wide Human SNP Array 6.0 (Affymetrix, Inc., Santa Clara, CA, USA) platform. Genotyped markers with a Hardy-Weinberg Equilibrium ($P < 0.001$), low call rate (<95%) and low minor allele frequency (< 0.01) were excluded from the analysis. In addition to the Wellcome Trust Case–Control Consortium guideline for quality control, Mendelian and non-Mendelian errors mimicking double-recombinations were detected by PEDSTATS (University of Michigan, Ann Arbor, MI, USA)[15] and Merlin (University of Michigan, Ann Arbor, MI, USA).[16] This error-checking step further detected and deleted 63 777 erroneous markers. Finally, 516 452 SNP were used in the genetic analysis.

The untyped or missing markers in the KARE3 and the twin-family study were imputed using Beagle (University of Washington, Seattle, WA, USA).[17] Both HapMap3 phase2 (JPT+CHB; http://hapmap.ncbi.nlm.nih.gov/) and Korean HapMap (http://www.khapmap.org/) panel data were combined to serve as the reference. Imputation quality score ($0 \leq r^2 \leq 1$) was checked, and the r^2 cut-off for post-imputation SNP filtering was 0.5. Higher values of r^2 represent increased accuracy in genotype imputation.

Analytic strategies and statistical analysis

To test the association with lung function, we conducted association analyses in three steps (Fig. 1): (i) GWAS in the KARE3 served as the discovery phase; first, screening was performed using SNP markers which were actually typed; (ii) findings were further proved adding imputed markers in highly associated regions. SNP markers exceeding the significance level of 5×10^{-6} were selected for the validation phase. (iii) The HTS was used for the validation study, and nominal *P*-value of 0.05 was used to confirm the findings.

For the genome-wide association analysis, linear regression of age, gender, height and smoking pack-years was undertaken on FEV_1 and FEV_1/FVC. The residuals were used as the phenotype for association testing. SNP markers were tested for association by linear regression under an additive model in the

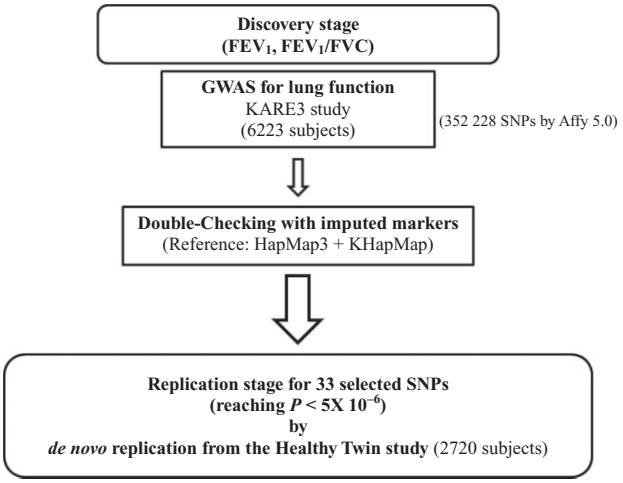

Figure 1 A schematic overview of the genetic analyses in this study.

Table 1 Baseline characteristics of subjects used for genome-wide association study of pulmonary function

Cohort	KARE3	Healthy twin study
Study design	Population-based	Family-based
Sample size	6223	2720
Age (years)	55.6 ± 8.7 (43–74)	44.2 ± 13.1 (17–81)
Female (%)	3225 (51.8%)	1660 (61.0%)
Smoking, pack-years	10.3 ± 17.5	5.3 ± 12.3
Smoking status		
Non-smoker	3862 (62%)	1774 (65%)
Ex-smoker	1200 (19%)	343 (13%)
Current smoker	1162 (19%)	603 (22%)
FEV_1 (L)	2.77 ± 0.68	2.92 ± 0.71
Height (cm)	160.1 ± 8.7	161.7 ± 8.5

Data represent the mean values ± standard deviation unless otherwise noted.

KARE3 population using PLINK (v1.07; Harvard University, Cambridge, MA, USA).[18] The family-based association analysis in the HTS was performed for a maximum-likelihood-based association test with variance-covariance matrix to consider phenotypic correlation between family members using Merlin.[16,19] Quantile-quantile plots, Manhattan plots and SNP annotation were performed using the WGA Viewer (Duke University, Durham, NC, USA).[20] Regional association results of significant SNP were plotted using LocusZoom (University of Michigan, Ann Arbor, MI, USA).[21] To estimate the proportion of the additive genetic variance explained by SNP, the proportion as $2p(1-p)\beta^2$ where p denotes minor allele frequency and β denotes the effect size.

RESULTS

Characteristics of study population

Genotype data for the 6223 subjects in the KARE3 were available for the primary discovery analysis after quality control. Replication of identified 33 SNP was assessed in 2720 subjects in the HTS. Healthy Twin subjects consisted of 485 monozygotic and 108 dizygotic twin pairs and their family members. There was no subject in the HTS who was included in the KARE3 population cohort. The main demographical characteristics of the study cohorts are summarized in Table 1. The mean age was 55.6 years in the KARE3 population cohort and 44.2 years in the HTS. There were 598 subjects whose pre-bronchodilator FEV_1/FVC was less than 0.7 in the KARE3 population and 165 subjects in the HTS. The average amount of smoking was 10.3 pack-years and 5.3 pack-years, respectively.

GWAS of lung function in the discovery phase

In the discovery phase, a total of 352 228 genotyped SNP were included in the analysis after quality control. Several SNP showed probable associations with FEV_1/FVC ratio and FEV_1; results shown by the quantile-quantile plots suggest the presence of multiple loci with modest effects (Fig. 2). The Manhattan plots using Haploview are shown in Figure 3. Top 10 SNP for FEV_1 and FEV_1/FVC ratio identified in the KARE3 population are listed in Table 2. No genotyped SNP met genome-wide significance criteria ($P < 5 \times 10^{-8}$) for association with FEV_1/FVC or FEV_1. Top 10 SNP according to the smoking status are listed in Table S1 in the online supporting information.

After imputing the regions showing the top significance with Beagle, 14 SNP with FEV_1/FVC and 19 SNP with FEV_1 were shown to be significant at the level of *P*-value below 5×10^{-6} (Table 3). Chromosome 6p21 region exhibited the strongest association with FEV_1/FVC ratio, and one SNP (rs10947233 in *PPT2*) showed genome-wide significance ($P = 3.4 \times 10^{-9}$) (Fig. 4b). SNP in *FAM13A* on chromosome 4q22 were also associated with FEV_1/FVC ratio (Fig. 4a). SNP on the chromosome 17q24 region upstream of *SOX9* was most significantly associated with FEV_1. Two SNP (rs4793538 and rs11655567) in this region showed near genome-wide significance with FEV_1 ($P = 5.6 \times 10^{-8}$) (Fig. 4c). The lowest *P*-values of genes that were associated with lung function in previous GWAS are listed in Table S2 in the online supporting information.

Replication study in the Healthy Twin Study

Thirty-three potential SNP of significance below 5×10^{-6} were analysed for replication. We evaluated these potential SNP for lung function in the HTS cohort using Merlin software. Only SNP in 6p21 region including rs10947233 in *PPT2* showed nominally significant associations with FEV_1/FVC ($P = 0.0028$–0.045) (Table 4). The effect of minor allele was positive which is consistent with beta in the KARE3 results. When we performed a sensitivity analysis excluding those with asthmatic history or under age 25, the results were similar (Supplementary Table S3).

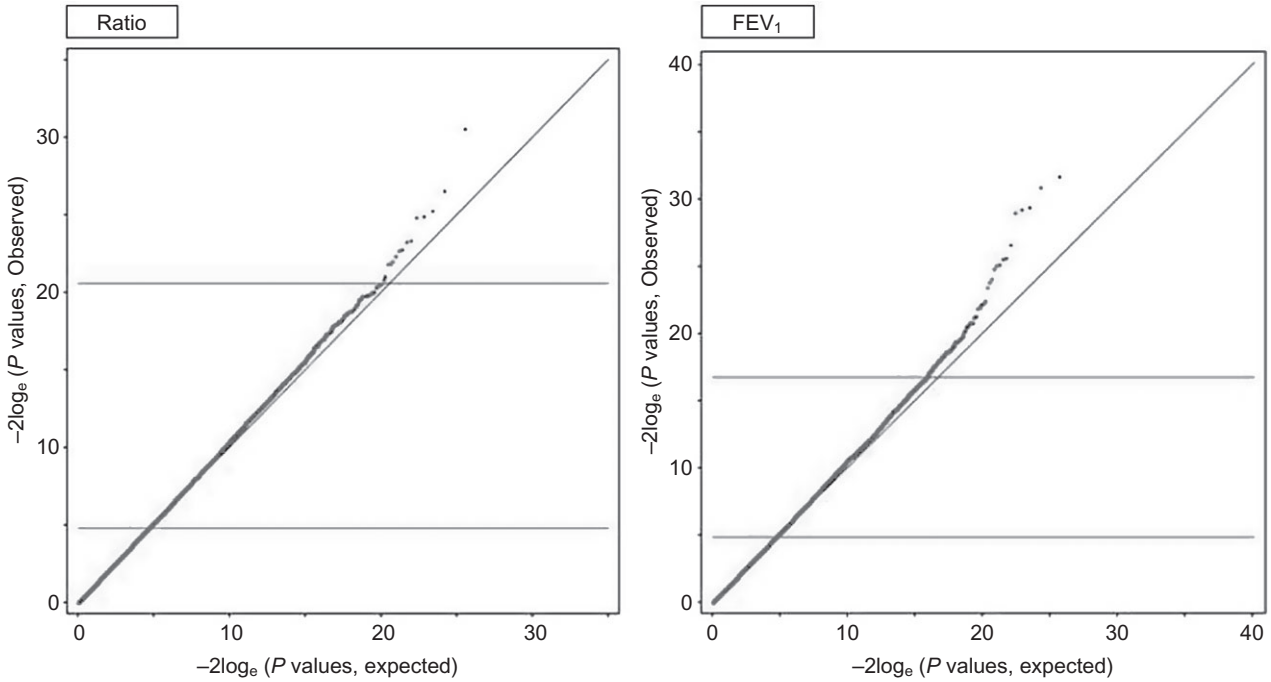

Figure 2 Quantile-quantile (QQ) plot of the study results. QQ plot of observed versus expected log10(*P*) values in forced expiratory volume in 1 s (FEV_1)/forced vital capacity (FVC) ratio and FEV_1.

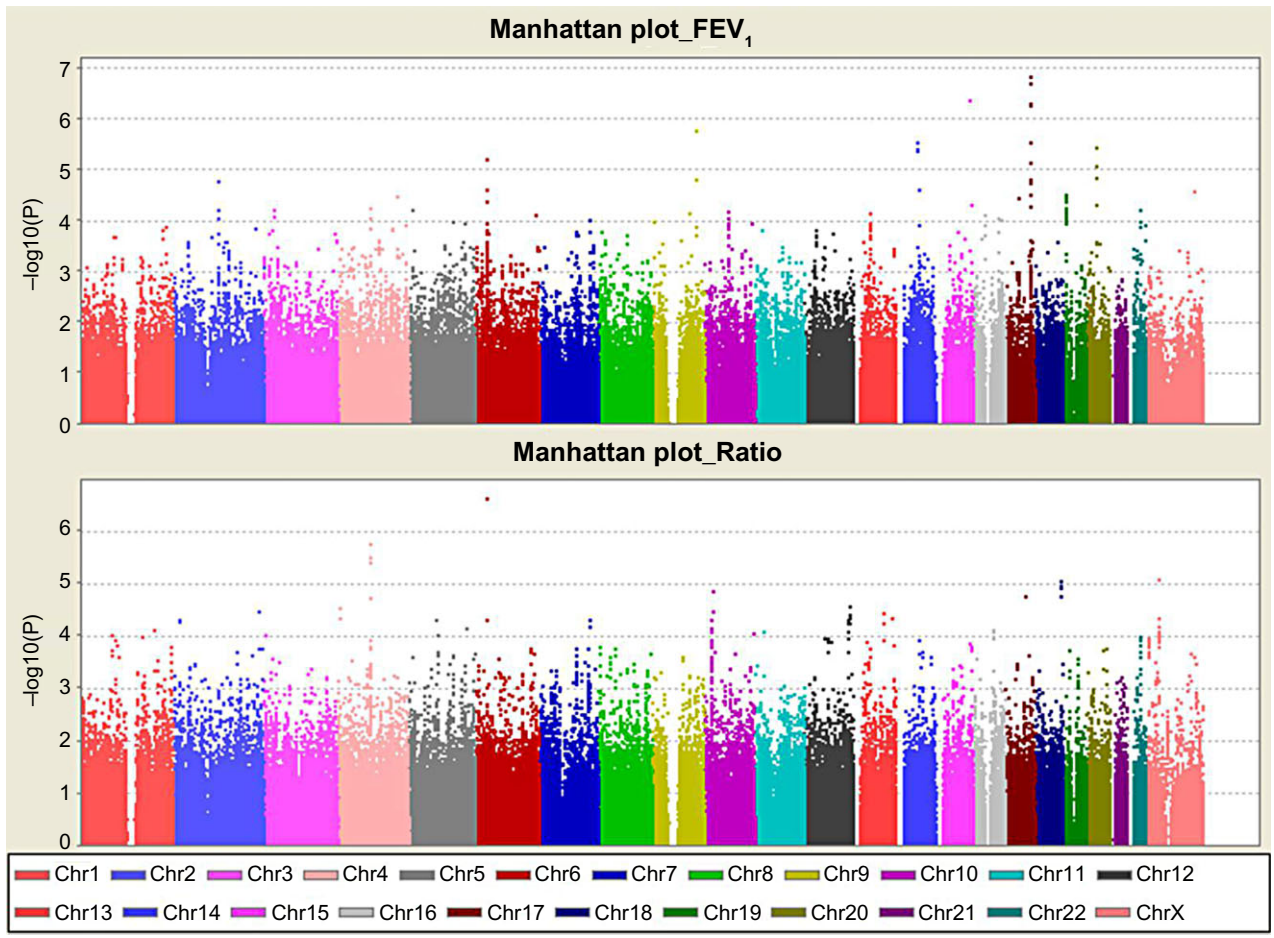

Figure 3 Manhattan plots of the genome-wide association signals with lung function from the Korean Association Resource phase 3 (KARE3) population. The –log10(*P*) values are plotted against the chromosomal base-pair positions.

Table 2 Association results for the top 10 genotyped SNP identified in KARE3 for their association to lung function measurements of FEV_1/FVC ratio and FEV_1

SNP	CHR	Position	Gene	Allele	Frequency	Beta	SE	*P*-value
FEV_1/FVC								
rs2021783	6	32152829	*TNXB*	T	0.16	0.776	0.15	2.28E-07
rs2609261	4	90054508	*FAM13A*	A	0.47	0.529	0.11	1.72E-06
rs2609260	4	90055842	*FAM13A*	A	0.56	0.518	0.111	3.23E-06
rs2609264	4	90047103	*FAM13A*	T	0.52	0.51	0.11	3.88E-06
rs1458551	4	90031265	*FAM13A*	T	0.47	0.51	0.111	4.06E-06
rs2628125	18	70722593	*ZNF407*	A	0.49	0.496	0.112	9.02E-06
rs8085262	18	70725167	*ZNF407*	A	0.33	0.516	0.118	1.15E-05
rs2628123	18	70728913	*ZNF407*	A	0.50	–0.479	0.109	1.18E-05
rs6415963	10	21280349	*NEBL*	A	0.45	0.479	0.11	1.43E-05
rs4891199	18	70729556	*ZNF407*	T	0.33	0.506	0.118	1.71E-05
FEV_1								
rs17765644	17	66691087	*SOX9*	T	0.62	–0.037	0.007	1.41E-07
rs17178251	17	66688474	*SOX9*	C	0.39	–0.037	0.007	2.06E-07
rs8031759	15	91599814	*RGMA*	A	0.08	–0.063	0.013	4.39E-07
rs11870732	17	66706836	*SOX9*	T	0.61	–0.036	0.007	4.84E-07
rs4793541	17	66739190	*SOX9*	A	0.61	–0.035	0.007	5.30E-07
rs3748172	9	116875752	*TNC*	T	0.11	–0.052	0.01	1.75E-06
rs9674957	17	66622693	*SOX9*	T	0.59	0.032	0.007	2.87E-06
rs1535574	14	57210327	*SLC35F4*	T	0.59	–0.033	0.006	2.94E-06
rs6132862	20	25669248	*ZNF337*	A	0.68	–0.034	0.007	3.60E-06
rs11851949	14	57209749	*SLC35F4*	C	0.41	–0.032	0.007	3.80E-06

CHR, Chromosome.

Proportion of heritability explained by each SNP was calculated as additive genetic variance explained by the SNP divided by heritability and total lung function variation with same samples. As a result, proportion of heritability explained by rs10947233 was 0.86%, assuming heritability of FEV_1/FVC is 0.5 in HTS. Other SNP including those in *FAM13A* and near *SOX9* were not replicated in HTS.

DISCUSSION

In this study, we conducted a population-based study for the discovery of loci associated with lung function. The regions identified were 6p21 for FEV_1/FVC and 17q24 for FEV_1. A replication study in a family-based cohort revealed that SNP in the 6p21 region replicated the association with FEV_1/FVC, whereas the 17q24 region did not show any significant association in the HTS.

The loci identified for FEV_1/FVC in the Korean population included 6p21 and 4q22, which are located near regions previously reported by GWAS in Caucasian populations for their associations with lung function. The 6p21 region included *TNXB*, *PPT2*, *AGER* and *NOTCH4* genes. The most significant association was shown in the intronic SNP (rs10947233) of *PPT2*. This SNP was tightly linked to missense SNP in *AGER* (rs2070600) and to several intronic SNP in *NOTCH4* using HaploReg (http://www.broadinstitute.org/mammals/haploreg/haploreg.php). Among them, rs2070600 in *AGER* is the only potentially functional SNP. This SNP was significantly associated with lung function in the SpiroMeta[7] and CHARGE studies.[8] SNP in *AGER* were also associated with COPD in a case–control study.[22] Both of the lung function GWAS showed positive beta of the minor allele, and the minor allele was protective in the COPD case–control analysis. The current study showed positive beta of the minor allele which is consistent with the results of European data. AGER protein, a receptor for advanced glycation end-products (RAGE), is a multi-ligand receptor of the immunoglobulin super-family and interacts with distinct molecules implicated in homeostasis, development, inflammation, diabetes and neurodegeneration.[23] RAGE signals depend on the cell type and the context. RAGE expression increases following cigarette smoke exposure and is partially responsible for inducing the pro-inflammatory signalling pathway.[24] A transgenic mouse model developed to conditionally upregulate RAGE showed apoptosis and pronounced inflammation in the peripheral lung,[25] while RAGE knockout mice showed decreased alveolar macrophage-mediated inflammation by tobacco smoke exposure. In cystic fibrosis, the promoter variant of *AGER* was associated with poor lung function and increased the RAGE expression.[26] The CC genotype of rs2070600 was associated with higher circulating levels of soluble forms of RAGE,[27] thus suggesting that this SNP may have a functional effect.

In the GWAS of asthma in the Japanese population, the most significant association was shown in rs404860 located in the 6p21 *NOTCH4* near *AGER*, but that was in weak linkage disequilibrium with rs2070600.[28] Another study in asthma subjects showed

Table 3 SNP highly associated ($P < 5 \times 10^{-6}$) with lung function in KARE3 cohort following imputation

CHR	SNP (r^2)	Position	Gene	Beta	SE	*P*-value
FEV_1/FVC ratio						
4	rs2609255 (0.99)	90030218	FAM13A	0.511	0.111	3.85E-06
4	rs2464528 (1.00)	90036687	FAM13A	0.51	0.11	3.84E-06
4	rs2609265 (1.00)	90045989	FAM13A	0.51	0.11	3.84E-06
4	rs2609264 (1.00)	90047103	FAM13A	0.511	0.11	3.72E-06
4	rs2609262 (1.00)	90054461	FAM13A	0.529	0.11	1.68E-06
4	rs2609261 (0.98)	90054508	FAM13A	0.531	0.11	1.59E-06
4	rs2609260 (1.00)	90055842	FAM13A	0.518	0.111	3.14E-06
4	rs1246642 (0.93)	90083469	FAM13A	0.511	0.111	3.94E-06
6	rs9368704 (0.99)	32149351	TNXB	0.777	0.1499	2.23E-07
6	rs2021783 (1.00)	32152829	TNXB	0.776	0.1499	2.28E-07
6	rs9348878 (0.61)	32217272	PRRT1	1	0.1909	1.68E-07
6	rs10947233 (0.97)	32232402	PPT2	1.149	0.1942	3.42E-09[†]
18	rs1863417 (0.88)	70747030	ZNF407	0.539	0.1113	1.29E-06
18	rs9960454 (1.00)	70798925	ZNF407	0.513	0.111	3.89E-06
FEV_1						
9	rs3748172 (1.00)	116875752	TNC	–0.052	0.001	1.82E-06
14	rs2146621 (0.99)	57195940	SLC35F4	–0.032	0.007	3.39E-06
14	rs11851949 (1.00)	57209749	SLC35F4	–0.032	0.007	3.80E-06
14	rs1535574 (1.00)	57210327	SLC35F4	–0.033	0.006	2.94E-06
15	rs8031759 (1.00)	91599814	RGMA	–0.063	0.013	4.39E-07
17	rs9674957 (1.00)	66622693	SOX9	0.032	0.007	2.95E-06
17	rs17178251 (1.00)	66688474	SOX9	–0.037	0.007	2.07E-07
17	rs17765644 (0.99)	66691087	SOX9	–0.037	0.007	2.54E-07
17	rs740408 (1.00)	66692691	SOX9	–0.036	0.004	2.38E-07
17	rs4793535 (1.00)	66698476	SOX9	–0.036	0.007	2.93E-07
17	rs1008348 (1.00)	66702911	SOX9	–0.036	0.007	3.10E-07
17	rs6501452 (1.00)	66704882	SOX9	–0.036	0.007	3.38E-07
17	rs11870732 (1.00)	66706836	SOX9	–0.036	0.007	4.67E-07
17	rs6501455 (1.00)	66713406	SOX9	–0.036	0.007	3.80E-07
17	rs983085 (0.99)	66723656	SOX9	–0.036	0.007	3.41E-07
17	rs4793538 (1.00)	66727523	SOX9	–0.038	0.007	5.64E-08
17	rs11655567 (0.99)	66728282	SOX9	–0.038	0.007	5.64E-08
20	rs6138650 (0.99)	25663507	ZNF337	–0.034	0.007	3.61E-06
20	rs6132862 (1.00)	25669248	ZNF337	–0.034	0.007	3.61E-06

[†]Genome-wide significant ($< 5 \times 10^{-8}$).
CHR, chromosome; r^2, imputation quality score.

that SNP in *NOTCH4* was associated with lung function in asthmatic subjects, though rs2070600 was not associated.[29] One potential explanation for these results may be that the lung function determinant is the functional SNP of *AGER* and the asthma association is driven by *NOTCH4*. However, further research will be required in order to determine the role of this region in asthma pathogenesis.

FAM13A was reported to be associated with FEV_1/FVC ratio in the CHARGE consortium[8] and was strongly associated with COPD susceptibility in the previous GWAS.[30] One previous GWAS used only the initial phase of KARE3 to identify the genes influencing lung function. In the study, *FAM13A* showed the most significant association with lung function (unpubl. data). In the current study, *FAM13A* was modestly associated with FEV_1/FVC ratio. The associated SNP are in $1.6–3.9 \times 10^{-6}$. They are located between two SNP identified in CHARGE (rs2869967 and rs6830970), and are 50 kb upstream from the location of COPD-associated SNP (rs1903003 and rs7671167). Although this locus did not reach genome-wide significance nor was it replicated in the family cohort, this finding suggests that this previously reported region was associated with lung function in the Korean population.

There was a genome-wide genetic association of the chromosome 17q24 region upstream of *SOX9* with FEV_1 in the discovery set, although this association was not replicated in the family cohort. Recently, the same locus was associated with FEV_1 using joint meta-analysis method to test associations of gene and gene-by-smoking interaction.[10] *SOX9* was upregulated in adenocarcinoma of the lung, and the gene expression was associated with cell growth.[31] *SOX9* was known to be associated with lung development. *SOX5* and *SOX9* are members of the SOX trio, which is important for differentiation of chondrocytes. Of note, a previous study reported that the genetic variation of *SOX5* was associated with COPD susceptibility as a result of the

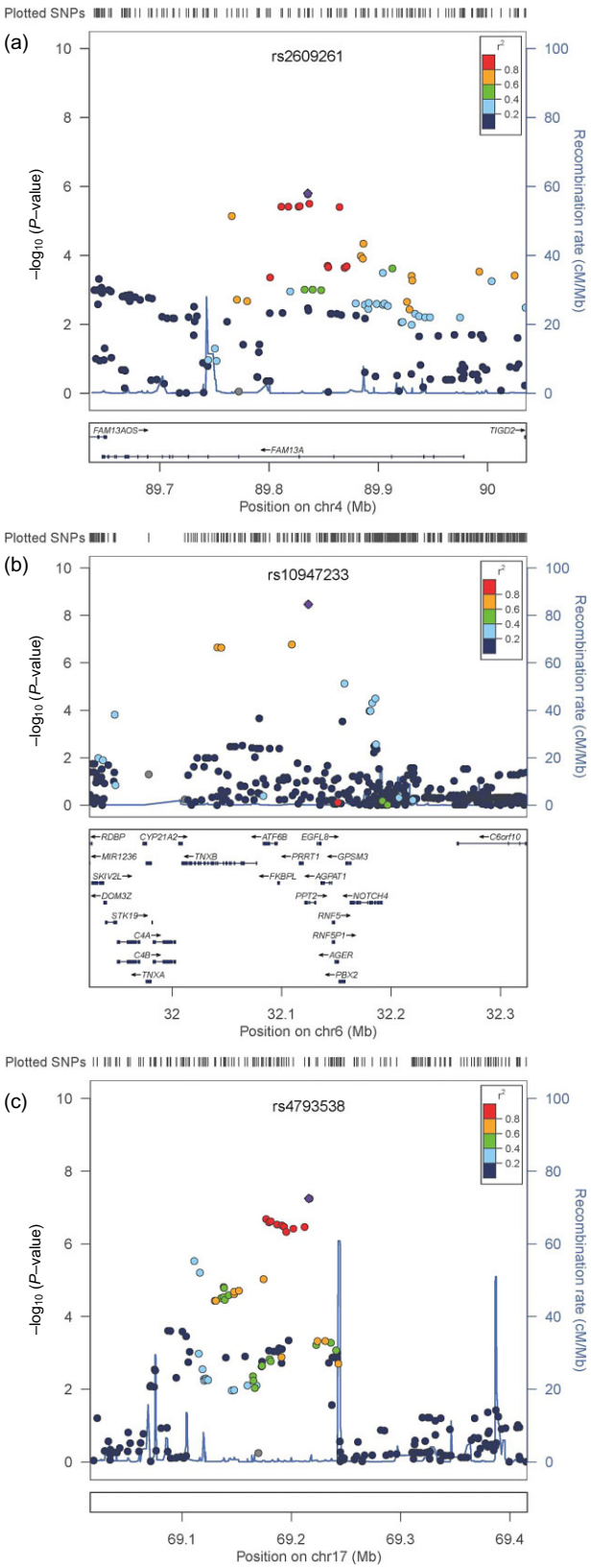

Figure 4 Regional association plots at the most significant loci associated with lung function. The plots of genetic loci on chromosome 4 (a), 6 (b) and 17 (c) are shown as they were created using LocusZoom.

fine mapping of the previous linked chromosomal region.[32] They also reported that *SOX5* gene expression decreased in severe COPD and that the *SOX5* knockout mouse showed abnormal lung morphology. Genes involved in lung development including *HHIP*, *SOX5* and *SOX9* may be important for determining lung function. This genetic association may be confined to a specific ethnic group. Further studies of replication in another population will be necessary.

Our study provides novel findings although it has several limitations which should be noted. In our study, we analysed genetic associations in a population-based cohort in the discovery set and replicated the study in a family-based cohort. These two analyses showed inconsistent results. The lack of replication may relate to the different statistical approach and the difference in demographical characteristics between two cohorts. The Healthy Twin family subjects were younger and smoked less. If the age effect is more environmental, such as longer duration of smoking, our study should have been largely adjusted by considering pack-year of the subjects. If different genes are expressed to affect lung functions depending on age, the power to detect replicated variants should have been reduced substantially. It is unlikely, however, that the replicated variants in this study stem from the age difference between two populations. Second, replication was based on nominal levels of association in the replication set. Formal adjustment for multiple statistical testing may have weakened the association. Further validation of the association in a larger study population will be required. Imputation method is known to generally enrich genetic marker information,[33] but some degree of errors might be introduced through the imputed markers. For our study, however, it is unlikely that replications based on imputed markers are false positive because the imputed markers are common variants, and imputation quality was very high (quality score 0.99~1.00). Additionally, our imputation method utilized family relationships to exclude markers which are Mendelian-inconsistent, which made the imputation results more accurate than conventional approach.

In conclusion, the chromosome 6p21 locus showed association with reduced FEV_1/FVC ratio and the chromosome 17q24 locus near *SOX9* with FEV_1 in the discovery set. *FAM13A* was also associated with FEV_1/FVC. In the replication set, SNP in *PPT2* showed the most significant association. We confirmed the association of 6p21 with lung function in the Korean population.

Acknowledgements

This work was funded by the Consortium for Large Scale Genome Wide Association Study III (2011E7300400), which was supported by the genotyping data (the Korean Genome Analysis Project, 4845-301) and the phenotype data (the Korean Genome Epidemiology Study, 4851-302) and a grant from the Centers for Disease Control and Prevention of Korea (budget 2012-E71011-00, 2011-E71011-00 and 2010-E71010-00) and also supported by the National Project for Personalized Genomic Medicine (A111218-11-GM02 and A111218-12-GM10).

Table 4 Genetic association results with lung function in Healthy Twin Study cohort using Merlin

CHR	SNP	Position	Nearest gene	Allele	Effect	MAF	*P*-value (HTS)
6	rs9368704	32149351	*TNXB*	A	0.004	0.1677	0.045
6	rs2021783	32152829	*TNXB*	T	0.004	0.1677	0.045
6	rs9348878	32217272	*PRRT1*	G	0.005	0.1046	0.02
6	rs10947233	32232402	*PPT2*	T	0.006	0.1492	0.0028

CHR, chromosome; MAF, minor allele frequency.

REFERENCES

1 Rabe KF, Hurd S, Anzueto A, Barnes PJ, Buist SA, Calverley P, Fukuchi Y, Jenkins C, Rodriguez-Roisin R, van Weel C *et al*. Global strategy for the diagnosis, management, and prevention of chronic obstructive pulmonary disease. *Am. J. Respir. Crit. Care Med.* 2007; **176**: 532–55.

2 Young RP, Hopkins R, Eaton TE. Forced expiratory volume in one second: not just a lung function test but a marker of premature death from all causes. *Eur. Respir. J.* 2007; **30**: 616–22.

3 Chen Y. Genetic epidemiology of pulmonary function. *Thorax* 1999; **54**: 818–24.

4 Ingebrigtsen T, Thomsen S, Sluis S, Miller M, Christensen K, Sigsgaard T, Backer V. Genetic influences on pulmonary function: a large sample twin study. *Lung* 2011; **189**: 323–30.

5 Hukkinen M, Kaprio J, Broms U, Viljanen A, Kotz D, Rantanen T, Korhonen T. Heritability of lung function: a twin study among never-smoking elderly women. *Twin Res. Hum. Genet.* 2011; **14**: 401–7.

6 Wilk JB, Chen TH, Gottlieb DJ, Walter RE, Nagle MW, Brandler BJ, Myers RH, Borecki IB, Silverman EK, Weiss ST *et al*. A genome-wide association study of pulmonary function measures in the Framingham Heart Study. *PLoS Genet.* 2009; **5**: e1000429.

7 Repapi E, Sayers I, Wain LV, Burton PR, Johnson T, Obeidat M, Zhao JH, Ramasamy A, Zhai G, Vitart V *et al*. Genome-wide association study identifies five loci associated with lung function. *Nat. Genet.* 2010; **42**: 36–44.

8 Hancock DB, Eijgelsheim M, Wilk JB, Gharib SA, Loehr LR, Marciante KD, Franceschini N, van Durme YM, Chen TH, Barr RG *et al*. Meta-analyses of genome-wide association studies identify multiple loci associated with pulmonary function. *Nat. Genet.* 2010; **42**: 45–52.

9 Artigas MS, Loth DW, Wain LV, Gharib SA, Obeidat ME, Tang W, Zhai G, Zhao JH, Smith AV, Huffman JE *et al*. Genome-wide association and large-scale follow up identifies 16 new loci influencing lung function. *Nat. Genet.* 2011; **43**: 1082–90.

10 Hancock DB, Artigas MS, Gharib SA, Henry A, Manichaikul A, Ramasamy A, Loth DW, Imboden M, Koch B, McArdle WL *et al*. Genome-wide joint meta-analysis of SNP and SNP-by-smoking interaction identifies novel loci for pulmonary function. *PLoS Genet.* 2012; **8**: e1003098.

11 Ong BA, Li J, McDonough JM, Wei Z, Kim C, Chiavacci R, Mentch F, Caboot JB, Spergel J, Allen JL *et al*. Gene network analysis in a pediatric cohort identifies novel lung function genes. *PLoS ONE* 2013; **8**: e72899.

12 Cho YS, Go MJ, Kim YJ, Heo JY, Oh JH, Ban H-J, Yoon D, Lee MH, Kim D-J, Park M *et al*. A large-scale genome-wide association study of Asian populations uncovers genetic factors influencing eight quantitative traits. *Nat. Genet.* 2009; **41**: 527–34.

13 Sung J, Cho SI, Lee K, Ha M, Choi EY, Choi JS, Kim H, Kim J, Hong KS, Kim Y *et al*. Healthy twin: a twin-family study of Korea—protocols and current status. *Twin Res. Hum. Genet.* 2006; **9**: 844–8.

14 Gombojav B, Song YM, Lee K, Yang S, Kho M, Hwang YC, Ko G, Sung J. The healthy twin study, Korea updates: resources for omics and genome epidemiology studies. *Twin Res. Hum. Genet.* 2013; **16**: 241–5.

15 Wigginton JE, Abecasis GR. PEDSTATS: descriptive statistics, graphics and quality assessment for gene mapping data. *Bioinformatics* 2005; **21**: 3445–7.

16 Abecasis GR, Cherny SS, Cookson WO, Cardon LR. Merlin-rapid analysis of dense genetic maps using sparse gene flow trees. *Nat. Genet.* 2002; **30**: 97–101.

17 Browning BL, Browning SR. A unified approach to genotype imputation and haplotype-phase inference for large data sets of trios and unrelated individuals. *Am. J. Hum. Genet.* 2009; **84**: 210–23.

18 Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira MAR, Bender D, Maller J, Sklar P, de Bakker PIW, Daly MJ *et al*. PLINK: a tool set for whole-genome association and population-based linkage analyses. *Am. J. Hum. Genet.* 2007; **81**: 559–75.

19 Chen W-M, Abecasis GR. Family-based association tests for genomewide association scans. *Am. J. Hum. Genet.* 2007; **81**: 913–26.

20 Ge D, Zhang K, Need AC, Martin O, Fellay J, Urban TJ, Telenti A, Goldstein DB. WGAViewer: software for genomic annotation of whole genome association studies. *Genome Res.* 2008; **18**: 640–3.

21 Pruim RJ, Welch RP, Sanna S, Teslovich TM, Chines PS, Gliedt TP, Boehnke M, Abecasis GR, Willer CJ. LocusZoom: regional visualization of genome-wide association scan results. *Bioinformatics* 2010; **26**: 2336–7.

22 Castaldi PJ, Cho MH, Litonjua AA, Bakke P, Gulsvik A, Lomas DA, Anderson W, Beaty TH, Hokanson JE, Crapo JD *et al*. The association of genome-wide significant spirometric loci with COPD susceptibility. *Am. J. Respir. Cell Mol. Biol.* 2011; **45**: 1147–53.

23 Sorci G, Riuzzi F, Giambanco I, Donato R. RAGE in tissue homeostasis, repair and regeneration. *Biochim. Biophys. Acta Mol. Cell. Res.* 2013; **1833**: 101–9.

24 Robinson AB, Stogsdill JA, Lewis JB, Wood TT, Reynolds PR. RAGE and tobacco smoke: insights into modeling chronic obstructive pulmonary disease. *Front. Physiol.* 2012; **3**: 1–11.

25 Fineschi S, De Cunto G, Facchinetti F, Civelli M, Imbimbo BP, Carnini C, Villetti G, Lunghi B, Stochino S, Gibbons DL *et al*. RAGE contributes to postnatal pulmonary development and adult lung maintenance program in mice. *Am. J. Respir. Cell Mol. Biol.* 2013; **48**: 164–71.

26 B Beucher J, Boëlle P-Y, Busson P-F, Muselet-Charlier C, Clement A, Corvol H, French CFMGSI. AGER -429T/C is associated with an increased lung disease severity in cystic fibrosis. *PLoS ONE* 2012; **7**: e41913.

27 Gaens KHJ, Ferreira I, van der Kallen CJH, van Greevenbroek MMJ, Blaak EE, Feskens EJM, Dekker JM, Nijpels G, Heine RJ, 't Hart LM *et al*. Association of polymorphism in the receptor for advanced glycation end products (RAGE) gene with circulating RAGE levels. *Clin. Endocrin. Metabol.* 2009; **94**: 5174–80.

28 Hirota T, Takahashi A, Kubo M, Tsunoda T, Tomita K, Doi S, Fujita K, Miyatake A, Enomoto T, Miyagawa T *et al*. Genome-wide association study identifies three new susceptibility loci for adult asthma in the Japanese population. *Nat. Genet.* 2011; **43**: 893–6.

29 Li X, Howard TD, Moore WC, Ampleford EJ, Li H, Busse WW, Calhoun WJ, Castro M, Chung KF, Erzurum SC *et al*. Importance of hedgehog interacting protein and other lung function genes in asthma. *J. Allergy Clin. Immunol.* 2011; **127**: 1457–65.

30 Cho MH, Boutaoui N, Klanderman BJ, Sylvia JS, Ziniti JP, Hersh CP, DeMeo DL, Hunninghake GM, Litonjua AA, Sparrow D *et al*. Variants in FAM13A are associated with chronic obstructive pulmonary disease. *Nat. Genet.* 2010; **42**: 200–2.
31 Jiang S, Fang W, Hou Y, Huang S, Yen B, Chang J, Li S, Liu H, Liu Y, Huang C. Upregulation of SOX9 in lung adenocarcinoma and its involvement in the regulation of cell growth and tumorigenicity. *Clin. Cancer Res.* 2010; **16**: 4363–73.
32 Hersh CP, Silverman EK, Gascon J, Bhattacharya S, Klanderman BJ, Litonjua AA, Lefebvre V, Sparrow D, Reilly JJ, Anderson WH *et al*. SOX5 is a candidate gene for chronic obstructive pulmonary disease susceptibility and is necessary for lung development. *Am. J. Respir. Crit. Care Med.* 2011; **183**: 1482–9.
33 Marchini J, Howie B. Genotype imputation for genome-wide association studies. *Nat. Rev. Genet.* 2010; **11**: 499–511.

Supporting information

Additional Supporting Information may be found in the online version of this article at the publisher's web-site:

Table S1-1 Association results for the top 10 genotyped single-nucleotide polymorphism (SNP) identified in non-smoker subjects of the Korean Association Resource phase 3 (KARE3) for their association to lung function measurements of forced expiratory volume in 1 s (FEV_1)/forced vital capacity (FVC) ratio and FEV_1

Table S1-2 Association results for the top 10 genotyped single-nucleotide polymorphism (SNP) identified in ever-smokers of the Korean Association Resource phase 3 (KARE3) for their association to lung function measurements of forced expiratory volume in 1 s (FEV_1)/forced vital capacity (FVC) ratio and FEV_1

Table S2 Lowest *P* values of genes which were associated with lung function in previous genome-wide association study (GWAS).

Table S3 Sensitivity analysis excluding 49 subjects with asthma history or 35 subjects with under age 25.