



A genome wide association study for lung function in the Korean population using an exome array

Kyu-Sun Lee¹, Kun Hee Kim², Yeon-Mok Oh³, Buhm Han⁴, and Woo Jin Kim¹

¹Department of Internal Medicine, Kangwon National University School of Medicine, Chuncheon;

²Department of Convergence Medicine and ³Department of Pulmonary and Critical Care Medicine and Clinical Research Center for Chronic Obstructive Airway Disease, Asan Medical Center, University of Ulsan College of Medicine, Seoul; ⁴Department of Medical Science, Seoul National University College of Medicine, Seoul, Korea

Received: June 20, 2019

Revised: November 6, 2019

Accepted: January 9, 2020

Correspondence to
Woo Jin Kim, M.D.

Department of Internal Medicine, Kangwon National University School of Medicine, 156 Baengnyeong-ro, Chuncheon 24289, Korea

Tel: +82-33-250-7815

Fax: +82-33-255-6567

E-mail: pulmo2@kangwon.ac.kr

https://orcid.org/0000-0003-

2927-370X

Background/Aims: Lung function is an objective indicator of diagnosis and prognosis of respiratory diseases. Many common genetic variants have been associated with lung function in multiple ethnic populations. We looked for coding variants associated with forced expiratory volume in 1 second (FEV₁) and FEV₁/forced vital capacity (FVC) in the Korean general population.

Methods: We carried out exome array analysis and lung function measurements of the FEV₁ and FEV₁/FVC in 7,524 individuals of the Korean population. We evaluated single variants with minor allele frequency greater than 0.5%. We performed look-ups for candidate coding variants associations in the UK Biobank, SpiroMeta, and CHARGE consortia.

Results: We identified coding variants in the *SMIM29* (*C6orf1*) ($p = 1.2 \times 10^{-5}$) and *HMGA1* locus on chromosome 6p21, the *GIT2* ($p = 6.5 \times 10^{-5}$) locus on chromosome 12q24, and the *ARHGEF40* ($p = 9.9 \times 10^{-5}$) locus on chromosome 14q11 as having a significant association with lung function (FEV₁). We also confirmed a previously reported association with lung function and chronic obstructive pulmonary disease in the *FAM13A* ($p = 4.54 \times 10^{-6}$) locus on chromosome 4q22, in *TNXB* ($p = 1.30 \times 10^{-6}$) and in *AGER* ($p = 1.09 \times 10^{-8}$) locus on chromosome 6p21.

Conclusions: Our exome array analysis identified that several protein coding variants were associated with lung function in the Korean population. Common coding variants in *SMIM29* (*C6orf1*), *HMGA1*, *GIT2*, *FAM13A*, *TNXB*, *AGER* and low-frequency variant in *ARHGEF40* potentially affect lung function, which warrant further study.

Keywords: Exome; Genetic variation; Respiratory function tests

INTRODUCTION

Lung function is an important trait of the respiratory system. Lung function measurements of the forced expiratory volume in one second (FEV₁) and the ratio of FEV₁ to forced vital capacity (FEV₁/FVC) are used as criteria for chronic obstructive pulmonary disease (COPD) diagnosis and severity evaluation for pulmonary disease [1,2]. Although environmental factors such as smoking, air pollution and particulate matter influence lung

function, the heritability of lung function has been reported to be around 40% [3,4]. Genome wide association studies (GWASs) for lung function have been reported in data on large populations [5,6]. As expected, genetic loci associated with lung function were shown to play roles in susceptibility to respiratory disease including COPD [7]. However, most identified variants through GWASs are common variants (minor allele frequency [MAF] > 5%) of the population. As in many other complex traits, despite the extensive discovery of associated loci from

GWAS, there are some limitations in understanding diseases risk or trait variability only through association of common variants [8]. This problem, so called missing heritability, might be explained by low-frequency and rare variants, and structural variation [8,9].

The exome array contains mostly variant that alter nonsynonymous, splice or stop codons that are likely to affect protein structure and function. The majority of variants are low-frequency ($1\% < \text{MAF} \leq 5\%$) and rare ($\text{MAF} < 1\%$) [8,9], which could explain additional disease risk and trait variability. Genotyping using an exome array can be a cost-effective and efficient strategy compared to whole exome sequencing [8]. GWAS results using exome arrays have been reported in COPD [9,10] and as meta-analysis for lung functions in persons with European ancestry [11]. However, these studies included only a small fraction of the Asian population samples. There was a study for exome chip quality control for variant analysis and several more loci were identified using exome array in Korean samples [12]. To gain further insight into genetic influence on lung function and to discover variants in coding regions associated with lung function in the Korean population, we carried out a GWAS using exome-based genotyping array.

METHODS

Study populations

We investigated an exome array for coding variants associated with lung function measurement in 7,524 individuals from the Korean Genome and Epidemiology Study (KoGES), which consists of six prospective cohort studies [13]. Among them, the Korea Association Resource cohort was a population-based cohort from the Ansong rural area and Ansan city in South Korea (KoGES Ansan and Ansong study) that was initiated in 2001. More than 260 traits were examined by means of epidemiological surveys, physical examinations and laboratory tests including a pulmonary function test [14]. Spirometry was carried out in accordance with American Thoracic Society/European Respiratory Society guidelines [15]. The baseline examinations have been previously described [14]. Written informed consents were provided by all participants in this study. The study was conducted with bioresources from National Biobank of Korea, the

Centers for Disease Control and Prevention, Republic of Korea (KBN 2017-003) and approved by the Institutional Review Board of Asan Medical Center (2015-1341).

Genotyping and quality of control

In this study, genomic DNAs isolated from peripheral blood were genotyped on the Infinium Human Exome BeadChip v1 (Illumina, San Diego, CA, USA). Genotyping process and quality control of the genotype dataset were previously reported [12]. After quality control, a total of 48,187 single nucleotide polymorphisms (SNPs) were used in the exome array analysis.

Single variant analysis for association with lung function

Single variant association tests for FEV_1 and FEV_1/FVC were carried out using the linear mixed model. We used the likelihood ratio test implemented in the Genome wide Efficient Mixed Model Association (GEMMA) software package [16]. The fixed effects of each variant was tested after adjusting for age, sex, ever-smoking, pack-years, and height. A $p < 10^{-4}$ was the criterion for single variant association analysis. Variants analysis and annotation of genes was done with the GRCh37/hg19 database.

Gene-based testing for association with lung function

We carried out gene-based analysis using Sequence Kernel Association tests (SKAT) [17] to assess the joint effect of multiple low-frequency and rare genetic variants within genes on lung function traits. SKAT analyses identified the top 10 candidate genes associated ($p < 2.5 \times 10^{-5}$) with FEV_1 and FEV_1/FVC .

Replication study

We carried out look-up replication of the selected top nine variants for FEV_1 and FEV_1/FVC in 410,289 subjects in the UK Biobank study (<http://biobankengine.stanford.edu>), SpiroMeta and Cohorts for Heart and Aging Research in Genomic Epidemiology (CHARGE) (www.chargeconsortium.com) consortia [6,11,18-20]. A $p < 10^{-4}$ was the criterion for look-ups.

Characterization of findings

We assessed whether the identified loci contained variants associated with gene expression in various tissues by querying the expression quantitative trait loci (eQTL)

Table 1. Demographic characteristics of study populations

Characteristic	Value
Total sample	7,524
Sex	
Male	3,582 (47.6)
Female	3,942 (52.4)
Age, yr	52.1 ± 8.8
Current smoker	1,853 (24.6)
Former smoker	1,198 (15.9)
Never smoker	4,473 (59.5)
Pack-years	9.0 ± 15.9
Height, cm	160.1 ± 8.6
FEV ₁ , L	2.6 ± 1.9
FEV ₁ /FVC	78.0 ± 15.0

Values are presented as number (%) or mean ± SD. FEV₁, forced expiratory volume in 1 second; FVC, forced vital capacity.

database of the Genotype-tissue expression (GTEx project) (<https://gtexportal.org/home/>) [21]. Potentially deleterious coding variants were identified by Sorting Intolerant From Tolerant (SIFT) and PolyPhen-2 [22]. We searched for evidence of protein expression in the respiratory system by querying the Human Protein Atlas (www.proteinatlas.org) [23].

RESULTS

Cohort characteristics

Our analysis included 7,524 individuals (3,942 females; 52.4%) from KoGES with non-missing covariate and lung function phenotypes. The characteristics of the 7,524 individuals who were assessed for an exome array are shown in Table 1. The mean age was 52.1 years. A 40.5% (n = 3,051/7,524) had ever smoked with mean pack-years of 9.0.

Single variant analysis for association with lung function

We analyzed the association between SNPs from the KoGES exome array data and lung function measures, FEV₁ and FEV₁/FVC. For primary discovery analysis, we used genotyping data for the 7,524 subjects in the KoGES. First, we checked for sample quality. We detected 19

pairs of samples with genetic relatedness greater than 0.25, and removed one sample from each related pair. There was no sample failed for the genotype missingness test (missing rate < 5%). Next, we checked for marker quality. We removed 1 SNP which failed to pass 95% genotyping rate threshold, 74 SNPs which failed at Hardy-Weinberg Equilibrium (HWE) test in controls ($p < 0.0001$), and 3 SNPs which were detected as principal component analysis outliers. After quality controls, 77,397 SNPs remained. Plink version 1.07 was used for quality controls procedures and Genome-wide Complex Trait Analysis (GCTA) version 1.26 was used for calculating genetic relationship matrix. In the GEMMA analysis, only 31,571 SNPs were used which passed the internal filters of GEMMA in the default setting.

Finally, we isolate the 31,571 SNPs for FEV₁ and 16,616 SNPs for FEV₁/FVC. Among them, 10,513 of 31,571 (33.2%) SNPs for FEV₁ were rare and low-frequency variants and the rest 21,058 (66.7%) were common variants. Also, 5,294 of 16,616 (31.8%) SNPs for FEV₁/FVC were rare and low-frequency variants and the rest 11,322 (68.2%) SNPs were common variants. The top SNPs for FEV₁ and FEV₁/FVC ratio identified in the KoGES general population are listed in Table 2. Only one genotyped SNP met the exome-wide significance criteria ($p < 5 \times 10^{-8}$) in our exome array analyses. The strongest signal ($p = 1.2 \times 10^{-5}$) for FEV₁ was a variant rs1150781 in small integral membrane protein 29 (*SMIM29* [*C6orf1*]) on chromosome 6 (Figs. 1 and 2). Two SNPs (rs7742369, rs2780226) also on chromosome 6 (6p21) were located in or near *SMIM29* (*C6orf1*) and high-mobility group AT-hook 1 (*HMGAI1*). One of the top nine SNPs, variant rs114591848 was a low-frequency variant in rho guanine nucleotide exchange factor 40 (*ARHGEF40*) on chromosome 14 (MAF = 1.4%) and the others were common variants (Table 2).

The strongest signal ($p = 1.0 \times 10^{-8}$) for FEV₁/FVC was a rs2070600 in advanced glycosylation end-product specific receptor (*AGER*) on chromosome 6, which previously reported as a locus associated with lung function and COPD [18,19]. The second strongest signal ($p = 1.3 \times 10^{-6}$) was rs2239688 in tenascin XB (*TNXB*) on chromosome 6. *TNXB* encoded extracellular matrix glycoproteins, which are associated with organizing and maintaining the structure of tissues that support the body's muscles, joints, organs, and skin. This gene was previously reported to be associated with IPF [24] and COPD, lung

Table 2. Top nine coding variants associated with FEV₁, FEV₁/FVC and results of look-up analyses

Trait	KoGES				UK biobank			CHARGE consortium		SpiroMeta consortium	
	SNP ID	Chr:Pos	Gene	MAF	OR or beta (QT)	p value	OR or beta (QT)	OR or beta (QT)	p value	OR or beta (QT)	p value
FEV ₁	rs7742369	6:34,165,721	-	0.171	3.75E-02	7.11 × 10 ⁻⁵	0.028	4.00E-31	-	-	-
	rs2780226	6:34,199,092	-	0.165	3.94E-02	3.86 × 10 ⁻⁵	-0.035	1.15E-26	-	-	-
	rs1150781	6:34,214,322	C6orf1 (SMIM29)	0.181	4.04E-02	1.20 × 10 ⁻⁵	-0.033	5.65E-24	-	-	-
	rs1130,247,700	11:130,247,700	-	0.464	-2.82E-02	5.36 × 10 ⁻⁵	-	-	-	-	-
	rs12110,390,979	12:110,390,979	GIT2/TCHP	0.078	-5.31E-02	6.55 × 10 ⁻⁵	-	-	-	-	-
FEV ₁ /FVC	rs114591848	14:21,550,212	ARHGGEF40	0.014	1.17E-01	9.95 × 10 ⁻⁵	-	-	-	-	-
	rs7143633 ^a	14:21,547,088	ARHGGEF40	-	-	-	2.33	2.36E-03	-	-	-
	rs7671167	4:89,883,979	FAM13A	0.493	4.74E-01	4.54 × 10 ⁻⁶	-	0.397	3.60E-09	0.043	5.87E-06
	rs2045517	4:89,870,964	FAM13A	-	-	-	-0.047	2.00E-11	0.438	-0.047	2.00E-11
	rs2239688	6:32,054,212	TNXB	0.175	7.03E-01	1.30 × 10 ⁻⁶	-	-	-	-	-
rs433061 ^b	6:32,054,212	TNXB	-	-	-	-	0.127	2.00E-04	-	-	
rs1150754 ^c	6:32,050,758	TNXB	-	-	-	-	0.111	3.00E-03	-	-	
rs2070600	6:32,151,443	AGER	0.158	8.75E-01	1.09 × 10 ⁻⁸	0.126	9.07E-15	0.186	1.50E-08	2.56E-02	

Total 31,571 SNPs for FEV₁ and total 16,616 SNPs for FEV₁/FVC analyzed. Results presented for the trait and chromosome (Chr) and position (Pos) in build 37 are given for each SNP.

FEV₁, forced expiratory volume in 1 second; FVC, forced vital capacity; KoGES, The Korean Genome Epidemiology Study; CHARGE, Cohorts for Heart and Aging Research in Genomic Epidemiology; SNP, single nucleotide polymorphism; Chr, chromosome; Pos, position; MAF, minor allele frequency; OR, odds ratio; SMIM29, small integral membrane protein 29; GIT2, G protein-coupled receptor kinase interacting ArfGAP2; TCHP, trichoplein keratin filament binding; ARHGGEF40, rho guanine nucleotide exchange factor 40; FAM13A, family with sequence similarity 13 member A; TNXB, tenascin XB; AGER, advanced glycosylation end-product specific receptor.

^ars7143633 is only identified in UK Biobank.

^brs433061 and ^crs1150754 are only identified in CHARGE consortium data set.

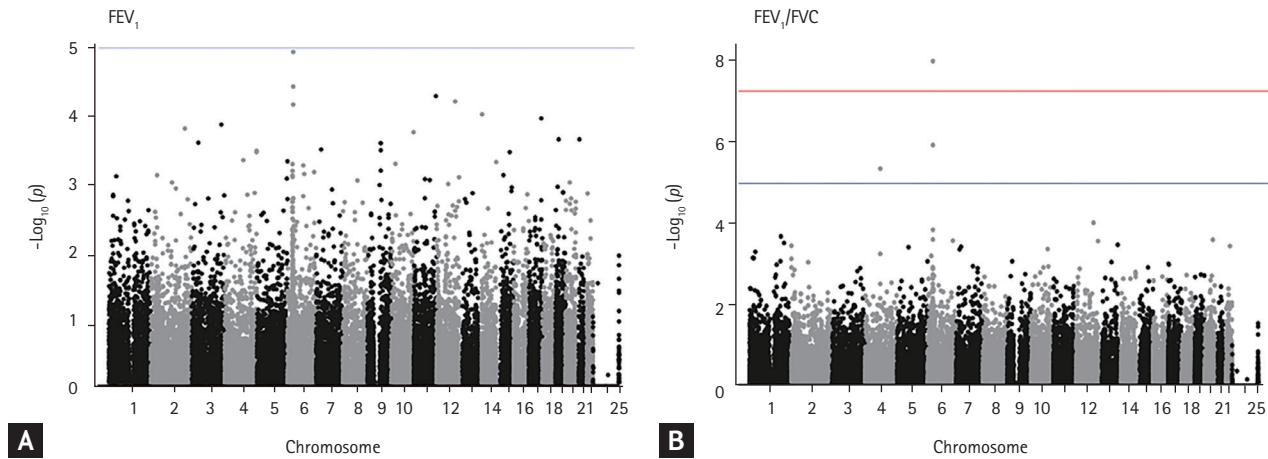


Figure 1. Manhattan plots of association results for forced expiratory volume in 1 second (FEV₁) and FEV₁/forced vital capacity (FVC). The manhattan plots for (A) FEV₁ and (B) FEV₁/FVC are ordered by chromosome position. Single nucleotide polymorphisms for which $-\log_{10} p > 5$ are indicated in blue line.

function [25]. To gain further insight into the associated variants, we assessed whether the candidate variants, or their proxies were associated with gene expression in various tissues by using GTEx and eQTL analyses (Supplementary Table 1). Among them, sentinel variants or their close proxies of rs114591848, rs7671167, and rs2070600 variants were eQTL in lung for *ARHGEF40*, family with sequence similarity 13 member A (*FAM13A*), and *AGER*. The protein and mRNA expression profiles of all implicated genes from the single variant association analyses are shown in Supplementary Table 2.

We also carried out a look-up in the publicly available UK Biobank results as adjusted for sex and ancestral principle components, SpiroMeta, and CHARGE consortia data in order to confirm the novelty of our results and presence or absence of difference of genetic variation in lung function among different ethnicities. SNPs associated with lung function within ± 1.5 Mb regions from selected variants were presented. This look-up showed evidence of replication for variant rs1150781 in or near *SMIM29* (*C6orf1*) and *HMGA1* on chromosome 6, proxies of variants rs114591848 in *ARHGEF 40* on chromosome 14 for FEV₁, and variant rs7671167 in *FAM 13A* and variant rs2070600 in *AGER* on chromosome 6 for FEV₁/FVC (Table 2).

Gene-based analysis for gene association with lung function

For gene-based analysis, we carried out the SKAT meth-

Table 3. Association results for all genes identified in SKAT analyses (KoGES)

Chr	Gene	p value	Trait
1	<i>DFFA</i>	8.01E-08	FEV ₁
10	<i>CHST15</i>	2.25E-05	FEV ₁
1	<i>DFFA</i>	5.81E-18	FEV ₁ /FVC
8	<i>DEFB135</i>	1.58E-10	FEV ₁ /FVC
11	<i>MOGAT2</i>	7.96E-10	FEV ₁ /FVC
11	<i>BSX</i>	1.09E-07	FEV ₁ /FVC
11	<i>SPTY2D1</i>	1.38E-07	FEV ₁ /FVC
14	<i>DEFB118</i>	2.86E-07	FEV ₁ /FVC
20	<i>PTGDR</i>	2.15E-06	FEV ₁ /FVC
20	<i>MYBL2</i>	3.25E-06	FEV ₁ /FVC

Results are given as chromosome, trait and p values ($p < 2.5 \times 10^{-3}$).

SKAT, Sequence Kernel Association Test; KoGES, Korean Genome and Epidemiology Study; Chr, chromosome; *DFFA*, DNA fragmentation factor subunit alpha; FEV₁, forced expiratory volume in 1 second; *CHST15*, carbohydrate sulfotransferase 15; FVC, forced vital capacity; *DEFB135*, defensin beta 135; *MOGAT2*, monoacylglycerol O-acyltransferase 2; *BSX*, brain specific homeobox; *SPTY2D1*, SPT2 chromatin protein domain containing 1; *DEFB118*, defensin beta 118; *PTGDR*, prostaglandin D2 receptor; *MYBL2*, MYB proto-oncogene like 2.

od to assess the joint effects of variants within genes on lung function traits. The top 10 most significant genes and p values of lung function are shown in Table 3. Our

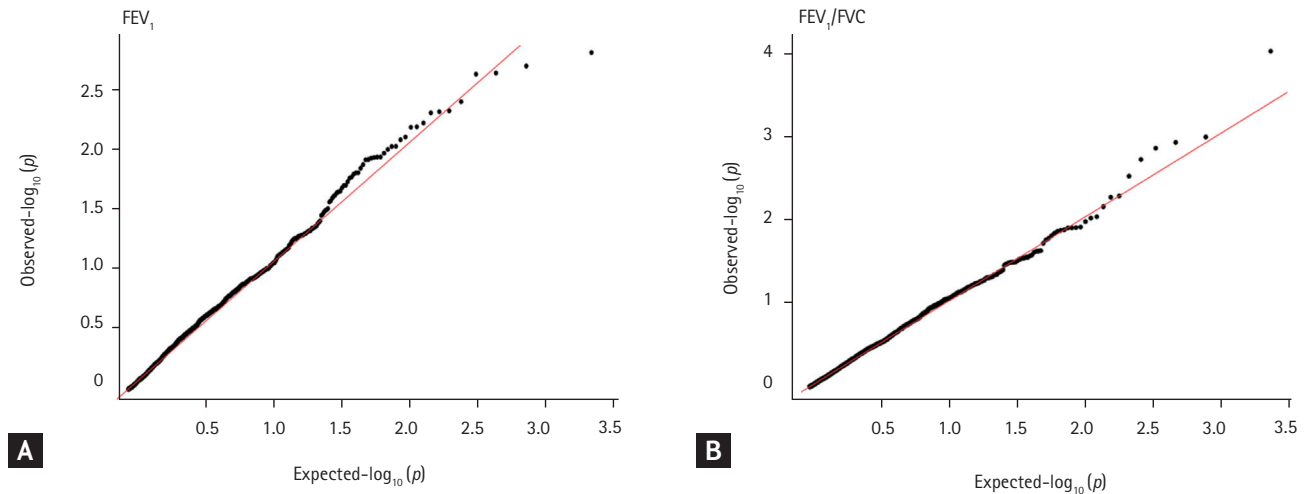


Figure 2. Quantile-Quantile (QQ) plots show $-\log_{10}(p)$ of observed genome wide association results against expected association results for (A) forced expiratory volume in 1 second (FEV_1) and (B) FEV_1 /forced vital capacity (FVC). Genomic control inflation factors (λ_{GC}) before genomic control was 0.90 for FEV_1 and 0.84 for FEV_1/FVC .

top association was in the gene DNA fragmentation factor subunit alpha (*DFFA*) ($p = 8 \times 10^{-8}$ for FEV_1 , $p = 5.8 \times 10^{-18}$ for FEV_1/FVC). However, we confirmed that the SKAT analysis and the candidate SNPs in or near target genes did not match.

DISCUSSION

In this study, we identified three loci (chromosome 4, 6, and 14) associated with lung function. A look-up study revealed that our novel SNPs in 6p21 and 14q11 loci replicated the association with FEV_1 from the UK Biobank. Some of the candidate SNPs (rs7742369, rs2780226, rs1150781, rs2239688, and rs2070600) were located on 6p21. We previously reported that this locus on 6p21 influences lung function in the Korean population [14]. The variant rs1150781 (MAF = 18%, $p = 1.2 \times 10^{-5}$, Gly150Ala, PolyPhen prediction: benign) (Supplementary Table 3) is a missense variant in *SMIM29* (*C6orf1*), which encodes an integral membrane and is expressed in brain, skin, thyroid, spleen, and lungs. This protein consists of 102 amino acids with molecular weight is of 11.5 kDa and is detected in human fetal lung cell lysate and respiratory epithelial cells. The expression of this protein was reported to increase in some non-small cell lung cancer patients, especially for adenocarcinoma and squamous cell lung cancer [23]. However, gain and loss

of functional studies of this gene are lacking. For this reason, although our exome array analysis identified the missense variant rs1150781 and nonsynonymous substitution (Gly150Ala) of the *SMIM29* (*C6orf1*) protein, to determine whether variant rs1150781 affects protein function, further validation of the association and functional studies of *SMIM29* (*C6orf1*) will be required. Variant rs1150781 and their proxy (rs2780226, LD, $r^2 = 0.99$) were located in or near *SMIM29* (*C6orf1*) and *HMGA1*. *SMIM29* (*C6orf1*) was located downstream of *HMGA1*, and these two genes are related to genetic linkage. *HMGA1* encodes a protein related to epigenetic modification and functions as a dynamic regulator of chromatin structure and transcription, which is localized in the cell nucleus. The *HMGA1* protein is expressed in human lung macrophages and respiratory epithelial cells [23]. Recently, Zhang et al. [26] reported that the protein and mRNA of *HMGA1* were highly expressed in intact human airway epithelia and their basal cells. In a loss of function study with *HMGA1* siRNA, they demonstrated that *HMGA1* down regulation in human airway basal cells led to increase expression of airway remodeling related genes. The NHGRI-EBI catalog of published GWASs shows that variants in or near *HMGA1* are associated with body height, BMI and smoking behavior (Supplementary Table 4) [27]. Also, the *HMGA1* protein is a key regulator of the insulin pathway [28] and variants of the *HMGA1* gene are associated with type 2 diabetes mellitus [29].

We also identified a nonsynonymous variant rs114591848 in the *ARHGEF40* locus on chromosome 14. This variant is a low-frequency (MAF = 1.4%) missense variant and resulted in an amino acid change (Arg1062Gln, PolyPhen prediction: possibly damaging) (Supplementary Table 3). *ARHGEF40* encodes Rho guanine nucleotide exchange factor is directly responsible for the activation of Rho-family GTPase, and regulates numerous cellular responses such as proliferation, differentiation, and cytoskeletal organization [30].

Moreover, we detected nominal levels of significance with two intronic SNPs rs7671167 in the *FAM13A* on chromosome 4q22.1, and rs2239688 in *TNXB* on chromosome 6p21.3 and one exonic SNP rs2070600 in the *AGER* on chromosome 6p21.3. These variants were previously reported to be loci associated with lung function and pulmonary diseases [6,10,19,20]. The *FAM13A* isoform 1 protein has a Rho GTPase-activating protein (GAP) domain and participates in the Rho GTPase signaling pathway [31]. Also, *ARHGEF40* encodes the Rho guanine nucleotide exchange factor. These results suggest that the Rho GTPase signaling pathway might play a role in lung function and COPD.

By means of our exome array analysis, we have tried to identify the low-frequency and rare variants potentially associated with lung function in order to uncover the missing heritability of lung function. However, our discovery analyses did not identify many rare and low-frequency coding variants that are responsible for the lung function trait in the Korean population, probably because of our small sample size and limited statistical power. Further confirmation of these associations in a large sample is needed.

We additionally investigated the joint effects of low-frequency and rare variant within genes, on lung function traits, by using the SKAT gene-based test. In these analyses, we identified an exome-wide significant signal ($p = 8 \times 10^{-8}$ for FEV_1 , $p = 5.8 \times 10^{-8}$ for FEV_1/FVC) in *DFFA*, which is also known to be an inhibitor of caspase-activated DNase. *DFFA* protein product is the substrate for caspase-3 and triggers DNA fragmentation during apoptosis [32]. However, this gene was not replicated in the UK Biobank data.

Our study has some potential limitations. First, the sample size is relatively small, and lack of statistical power may be a limitation. Second, we did not provide

further evidence for the biological role of the *SMIM29* (*C6orf1*), *HMGAI* and *ARHGEF40* in lung function. Finally, our exome array identified only coding variants, but cannot provide the roles of noncoding variants in lung function. To date, many studies and meta-analyses including SpiroMeta, CHARGE consortia, and UK Biobank studies have reported nearly 100 loci and many variants associated with lung function and COPD [11,20,33,34]. However, these studies have been exclusively carried out among populations whereas Asian ancestry populations participate with relatively smaller sample size. Therefore, there is a need to perform GWAS information from many people with Asian ancestry in order to better understand the genetic architecture of lung function.

In conclusion, we have newly identified a common coding variant in or near *SMIM29* (*C6orf1*), *HMGAI*, and one missense low-frequency variant in *ARHGEF40*, that are associated with lung function. Although a large sample size may be required to strengthen our results, we present additional evidence to support the notion that the genetic contribution to lung function includes polygenic architecture with low-frequency and common genetic variants in the Korean population.

KEY MESSAGE

1. We identified novel single nucleotide polymorphisms associated with lung function in the Korean population. There are: Common coding variant rs1150781 in or near *SMIM29* (*C6orf1*), *HMGAI* located on 6p21 and low-frequency variant rs114591848 in *ARHGEF40* locus on 14q11, which were associated with FEV_1 .
2. Common coding variant rs2070600 in *AGER* located on 6p21.3 associated with forced expiratory volume in 1 second/forced vital capacity with exome-wide significant threshold as previously reported in loci associated with lung function and chronic obstructive pulmonary disease.

Conflict of interest

No potential conflict of interest relevant to this article was reported.

Acknowledgments

This research was supported by the National Research Foundation of Korea (2017R1A2B4003790).

REFERENCES

1. Fromer L, Cooper CB. A review of the GOLD guidelines for the diagnosis and treatment of patients with COPD. *Int J Clin Pract* 2008;62:1219-1236.
2. Park YB, Rhee CK, Yoon HK, et al. Revised (2018) COPD clinical practice guideline of the Korean academy of tuberculosis and respiratory disease: a summary. *Tuberc Respir Dis (Seoul)* 2018;81:261-273.
3. Wilk JB, Djousse L, Arnett DK, et al. Evidence for major genes influencing pulmonary function in the NHLBI family heart study. *Genet Epidemiol* 2000;19:81-94.
4. Palmer LJ, Knuiman MW, Divitini ML, et al. Familial aggregation and heritability of adult lung function: results from the Busselton Health Study. *Eur Respir J* 2001;17:696-702.
5. Soler Artigas M, Wain LV, Miller S, et al. Sixteen new lung function signals identified through 1000 Genomes Project reference panel imputation. *Nat Commun* 2015;6:8658.
6. Soler Artigas M, Loth DW, Wain LV, et al. Genome-wide association and large-scale follow up identifies 16 new loci influencing lung function. *Nat Genet* 2011;43:1082-1090.
7. Hobbs BD, de Jong K, Lamontagne M, et al. Genetic loci associated with chronic obstructive pulmonary disease overlap with loci for lung function and pulmonary fibrosis. *Nat Genet* 2017;49:426-432.
8. Lee S, Abecasis GR, Boehnke M, Lin X. Rare-variant association analysis: study designs and statistical tests. *Am J Hum Genet* 2014;95:5-23.
9. Jackson VE, Ntalla I, Sayers I, et al. Exome-wide analysis of rare coding variation identifies novel associations with COPD and airflow limitation in MOCS3, IFIT3 and SERPINA12. *Thorax* 2016;71:501-509.
10. Hobbs BD, Parker MM, Chen H, et al. Exome array analysis identifies a common variant in IL27 associated with chronic obstructive pulmonary disease. *Am J Respir Crit Care Med* 2016;194:48-57.
11. Jackson VE, Latourelle JC, Wain LV, et al. Meta-analysis of exome array data identifies six novel genetic loci for lung function. *Wellcome Open Res* 2018;3:4.
12. Park TJ, Heo L, Moon S, et al. Practical calling approach for exome array-based genome-wide association studies in Korean population. *Int J Genomics* 2015;2015:421715.
13. Kim Y, Han BG; KoGES group. Cohort profile: the Korean genome and epidemiology study (KoGES) consortium. *Int J Epidemiol* 2017;46:1350.
14. Kim WJ, Lee MK, Shin C, et al. Genome-wide association studies identify locus on 6p21 influencing lung function in the Korean population. *Respirology* 2014;19:360-368.
15. Miller MR, Hankinson J, Brusasco V, et al. Standardisation of spirometry. *Eur Respir J* 2005;26:319-338.
16. Zhou X, Stephens M. Genome-wide efficient mixed-model analysis for association studies. *Nat Genet* 2012;44:821-824.
17. Wu MC, Lee S, Cai T, Li Y, Boehnke M, Lin X. Rare-variant association testing for sequencing data with the sequence kernel association test. *Am J Hum Genet* 2011;89:82-93.
18. Hancock DB, Eijgelsheim M, Wilk JB, et al. Meta-analyses of genome-wide association studies identify multiple loci associated with pulmonary function. *Nat Genet* 2010;42:45-52.
19. Repapi E, Sayers I, Wain LV, et al. Genome-wide association study identifies five loci associated with lung function. *Nat Genet* 2010;42:36-44.
20. Wain LV, Shrine N, Artigas MS, et al. Genome-wide association analyses for lung function and chronic obstructive pulmonary disease identify new loci and potential druggable targets. *Nat Genet* 2017;49:416-425.
21. GTEx Consortium. The genotype-tissue expression (GTEx) project. *Nat Genet* 2013;45:580-585.
22. Flanagan SE, Patch AM, Ellard S. Using SIFT and PolyPhen to predict loss-of-function and gain-of-function mutations. *Genet Test Mol Biomarkers* 2010;14:533-537.
23. Uhlen M, Oksvold P, Fagerberg L, et al. Towards a knowledge-based human protein atlas. *Nat Biotechnol* 2010;28:1248-1250.
24. Garner IM, Evans IC, Barnes JL, et al. Hypomethylation of the TNXB gene contributes to increased expression and deposition of tenascin-X in idiopathic pulmonary fibrosis. *Am J Respir Crit Care Med* 2014;189:A3378.
25. Artigas MS, Wain LV, Shrine N, et al. Targeted sequencing of lung function loci in chronic obstructive pulmonary disease cases and controls. *PLoS One* 2017;12:e0170222.
26. Zhang H, Yang J, Walters MS, et al. Mandatory role of HMGA1 in human airway epithelial normal differentiation and post-injury regeneration. *Oncotarget*

- 2018;9:14324-14337.
27. MacArthur J, Bowler E, Cerezo M, et al. The new NHGRI-EBI catalog of published genome-wide association studies (GWAS catalog). *Nucleic Acids Res* 2017;45:D896-D901.
 28. Chiefari E, Nevolo MT, Arcidiacono B, et al. HMGA1 is a novel downstream nuclear target of the insulin receptor signaling pathway. *Sci Rep* 2012;2:251.
 29. Chiefari E, Tanyolac S, Paonessa F, et al. Functional variants of the HMGA1 gene and type 2 diabetes mellitus. *JAMA* 2011;305:903-912.
 30. Fujiwara S, Matsui TS, Ohashi K, Mizuno K, Deguchi S. Keratin-binding ability of the N-terminal Solo domain of Solo is critical for its function in cellular mechanotransduction. *Genes Cells* 2019;24:390-402.
 31. Corvol H, Hodges CA, Drumm ML, Guillot L. Moving beyond genetics: is FAM13A a major biological contributor in lung physiology and chronic lung diseases? *J Med Genet* 2014;51:646-649.
 32. Liu X, Zou H, Slaughter C, Wang X. DFF, a heterodimeric protein that functions downstream of caspase-3 to trigger DNA fragmentation during apoptosis. *Cell* 1997;89:175-184.
 33. Wain LV, Shrine N, Miller S, et al. Novel insights into the genetics of smoking behaviour, lung function, and chronic obstructive pulmonary disease (UK BiLEVE): a genetic association study in UK Biobank. *Lancet Respir Med* 2015;3:769-781.
 34. Wyss AB, Sofer T, Lee MK, et al. Multiethnic meta-analysis identifies ancestry-specific and cross-ancestry loci for pulmonary function. *Nat Commun* 2018;9:2976.

Supplementary Table 1. Single variant association with eQTL analysis in GTEx

Trait	SNP ID (KoGES)	Proxy	Chr:Pos	Variant ID	Genecode ID	Gene symbol	p value	NES (effect size)	Tissue
FEV ₁	rs7742369	rs7742369	6:34,165,721	6_34165721_A_G_b37	ENSG00000186577.7	SMIM29 (C6orf1)	0.000032	0.13	Skin
	rs7742369	rs7742369	6:34,165,721	6_34165721_A_C_b37	ENSG00000186577.7	SMIM29 (C6orf1)	0.000073	0.14	Thyroid
	rs2780226	rs2780226	6:34,199,092	6_34199092_C_T_b37	ENSG00000186577.7	SMIM29 (C6orf1)	0.000024	-0.17	Skin
	rs2780226	rs2780226	6:34,199,092	6_34199092_C_T_b37	ENSG00000186577.7	SMIM29 (C6orf1)	0.000042	-0.21	Esophagus-mucosa
	rs1150781	rs1150781	6:34,214,322	6_34214322_C_G_b37	ENSG00000186577.7	SMIM29 (C6orf1)	0.000012	-0.18	Skin
	rs1130,247,700	none	11:130,247,700	-	-	-	-	-	-
	rs1130,247,700	rs4937505	11:130,247,781	11_130247781_T_A_b37	-	-	-	-	-
	rs12110,390,979	None	12:110,390,979	-	-	-	-	-	-
	rs12110,390,979	rs11069018	12:110386611	12_110386611_G_A_b37	ENSG00000139437.16	GIT2	4.60E-07	-0.28	Artery - tibial
	rs12110,390,979	rs11069018	12:110389404	12_110389404_T_C_b37	ENSG00000139437.16	GIT2	0.000034	-0.32	Testis
FEV ₁ /FVC	rs12110,390,979	rs2292354	12:110368201	12_110368201_G_A_b37	ENSG00000139437.13	TCHP	0.000022	-0.31	Colon - sigmoid
	rs12110,390,979	rs1106939	12:110413252	12_110413252_G_C_b37	ENSG00000139437.13	TCHP	0.000055	-0.33	Colon - sigmoid
	rs114591848	rs114591848	14:21,550,212	-	-	-	-	-	-
	rs114591848	rs398117960	14:21545490	14_21545490_C_CT_b37	ENSG00000165801.5	ARHGEF40	7.40E-09	-0.20	Lung
	rs114591848	rs3827911	14:21548159	14_21548159_G_A_b37	ENSG00000165801.5	ARHGEF40	1.10E-07	-0.18	Lung
	rs114591848	rs3827913	14:21548448	14_21548448_G_A_b37	ENSG00000165801.5	ARHGEF40	3.20E-08	-0.19	Lung
	rs114591848	rs1958395	14:21551058	14_21551058_G_A_b37	ENSG00000165801.5	ARHGEF40	4.30E-08	-0.19	Lung
	rs17671167	rs17671167	4:89,883,979	4_89883979_C_T_b37	ENSG00000138640.10	FAM13A	0.000012	0.16	Lung
	rs2239688	rs2239688	6:32,054,212	6_32054212_C_G_b37	-	-	-	-	-
	rs2239688	rs79770049	6:31,997,493	6_31977493_C_T_b37	ENSG00000168477.13	TNXB	3.50E-13	0.63	Adrenal gland
FEV ₁ /FVC	rs2239688	rs2746414	6:31,996,966	6_31996966_G_A_b37	ENSG00000168477.13	TNXB	2.70E-07	0.62	Adrenal gland
	rs2239688	rs2852706	6:31,972,370	6_31972370_G_C_b37	ENSG00000168477.13	TNXB	0.000014	0.50	Adrenal gland
	rs2239688	rs76096058	6:31,997,549	6_31977548_G_A_b37	ENSG00000168477.13	TNXB	0.000027	0.54	Adrenal gland
	rs2070600	rs2070600	6:32,151,443	6_32151443_C_T_b37	ENSG00000204305.9	AGER	0.000061	-0.28	Lung

Means that no significant eQTL were found for SNP in eQTL tissues.

eQTL, expression quantitative trait loci; GTEx, genotype-tissue expression; SNP, single nucleotide polymorphism; KoGES, Korean Genome and Epidemiology Study; Chr, chromosome; Pos, position; NES, normalized effect size; SMIM29, small integral membrane protein 29; GIT2, G protein-coupled receptor kinase interacting, Arf-GAP 2; TCHP, trichoplein keratin filament binding protein; ARHGEF40, rho guanine nucleotide exchange factor 40; FAM13A, family with sequence similarity 13 member A; TNXB, tenascin XB; AGER, advanced glycosylation end-product specific receptor.

Supplementary Table 2. Protein and mRNA expression profiles of implicated genes from single association analyses

Gene	Protein expression				RNA profile			
	Human protein atlas				Human protein atlas		GTEx	
	Nasopharynx	Bronchus	Lung macrophages	Lung pneumocytes	TPM	Category	RPKM	Category
<i>SMIM29(C6orf1)</i>	High	Not detected	Medium	Medium	21.2	Medium	14.8	Medium
<i>HMGA1</i>	Medium	Low	Medium	Not detected	25.9	Medium	27.1	Medium
<i>GIT2</i>	High	Low	High	Low	25	Medium	8.5	Low
<i>TCHP</i>	Medium	Medium	Low	Low	11	Medium	5.8	Low
<i>ARHGGEF40</i>	NA	NA	NA	NA	10.2	Medium	15.2	Medium
<i>FAM13A</i>	NA	Medium	Medium	Medium	17.8	Medium	3.4	Low
<i>TNXB</i>	NA	NA	NA	NA	16.4	Medium	20.3	Medium
<i>AGER</i>	Not detected	Not detected	High	Medium	889.2	Medium	424.4	High

Implicated genes were those located at or close to the position of the candidate top 9 single nucleotide polymorphisms. Protein expression are qualitative antibody based protein profiles in the human lung from the Human Protein Atlas. RNA expression is quantitative data estimating the transcript abundance of each protein-coding gene by RNA-seq from the Human Protein Atlas and GTEx.

GTEx, genotype-tissue expression; TPM, transcript per million; RPKM, reads per kilobase gene model and million mapped reads; *SMIM29*, small integral membrane protein 29; *HMGA1*, high mobility group AT-hook 1; *GIT2*, G protein-coupled receptor kinase interacting ArfGAP 2; *TCHP*, trichoplein keratin filament binding protein; *ARHGGEF40*, rho guanine nucleotide exchange factor 40; NA, not available; *FAM13A*, family with sequence similarity 13 member A; *TNXB*, tenascin XB; *AGER*, advanced glycosylation end-product specific receptor.

Supplementary Table 3. SIFT/PolyPhen predictions for rs1150781 and rs114591848

SNP ID	Chr:Pos	Allele	Consequence	Gene	Transcript ID	Codons (aminoacid)	SIFT consequence (score)	PolyPhen consequence (score)
rs1150781	6:34,214,322	G	Missense variant	<i>C6orf1 (SMIM29)</i>	ENSG00000186577	GGG(Gly150)→ GCG(Ala150)	Deleterious (0)	Benign (0.015)
rs114591848	14:21,550,212	G	Missense variant	<i>ARHGEF40</i>	ENSG00000165801	CGG(Arg1062)→ CAG(Gln1062)	Tolerated9 (0.08)	possibly damaging (0.679)

All SNPs were annotated and imported from dbSNP (release 151) in European Molecular Biology Laboratory-European Bioinformatics Institute. SIFT, sorting intolerant from tolerant; PolyPhen, polymorphism phenotyping; SNP, single nucleotide polymorphism; Chr, chromosome; Pos, position; SMIM29, small integral membrane protein 29; ARHGEF40, advanced glycosylation end-product specific receptor.

Supplementary Table 4. Association between variants in or near its target genes and disease and traits in NHGRI-EBI GWAS

Trait	SNP	Chr:Pos	Phenotype, disease, and trait	Reported gene(s)	Associated allele	p value	β Coefficient
FEV ₁	rs7742369	6:34,165,721	Birth weight	HMGAI	G	1.00E-08	0.0268 unit increase
			Height	HMGAI	G	1.00E-13	0.11 cm increase
Height	rs2780226	6:34,199,092	Height	HMGAI	T	8.00E-28	0.076 unit decrease
			Waist circumference adjusted for BMI	CYCSSL1, HMGAI	C	8.00E-10	0.0504 cm increase
Height	rs1150781	6:34,214,322	Height	HMGAI, C6orf1	C	2.00E-08	0.0048 unit increase
	11:130,247,700	11:130,247,700	-	-	-	-	-
	12:110,390,979	12:110,390,979	-	-	-	-	-
FEV ₁ /FVC	rs114591848	14:21,550,212	-	-	-	-	-
	rs7671167	4:89,883,979	COPD	FAM13A	-	1.00E-11	Odd ratio = 1.32
Post bronchodilator FEV ₁			Post bronchodilator FEV ₁	FAM13A	T	9.00E-08	0.046 unit decrease
			Post bronchodilator FEV ₁ /FVC ratio	FAM13A	T	2.00E-10	0.011 unit decrease
rs2239688	6:32,054,212	6:32,054,212	-	-	-	-	-
	rs2070600	6:32,151,443	COPD	AGER	C	6.00E-10	Odd ratio = 1.24
Emphysema imaging measurement			Emphysema imaging measurement	AGER	T	5.00E-09	0.14 unit decrease
			FEV ₁ /FVC ratio	AGER	T	3.00E-25	0.14 unit increase
FEV ₁ /FVC ratio, forced expiratory volume, PF measurement			FEV ₁ /FVC ratio, forced expiratory volume, PF measurement	AGER, PPT2	T	3.00E-14	Odd ratio = 1.0

NHGRI-EBI, National Human Genome Research Institute-European Bioinformatics Institute; GWAS, genome-wide association study; SNP, single nucleotide polymorphism; Chr, chromosome; Pos, position; FEV₁, forced expiratory volume in 1 second; HMGAI, high mobility group AT-hook 1; BMI, body mass index; CYCSSL1, cytochrome C, somatic-like 1; FVC, forced vital capacity; COPD, chronic obstructive pulmonary disease; FAM13A, family with sequence similarity 13 member A; AGER, advanced glycosylation end-product specific receptor; PF, pulmonary function; PPT2, palmitoyl-protein thioesterase 2.