Factors associated with lung function decline in a Korean population-based cohort

Chi Young Kim

Department of Medicine

The Graduate School, Yonsei University

Factors associated with lung function decline in a Korean population-based cohort

Chi Young Kim

Department of Medicine

The Graduate School, Yonsei University

Factors associated with lung function decline in a Korean population-based cohort

Directed by Professor Young Sam Kim

The Doctoral Dissertation
submitted to the Department of Medicine,
the Yonsei University Graduate School of Medicine
in partial fulfillment of the requirements for the degree of
Doctor of Philosophy

Chi Young Kim

December 2021

This certifies that the Doctoral Dissertation
of Chi Young Kim is approved.

Thesis Supervisor: Young Sam Kim

Thesis Committee Member #1: Chung Mo Nam

Thesis Committee Member #2: Sungho Won

Thesis Committee Member #3: Je Hyeong Kim

Thesis Committee Member #4: Ji Ye Jung

The Graduate School
Yonsei University

December 2021

ACKNOWLEDGEMENTS

I would like to express my special appreciation and thanks to my advisor Professor Young Sam Kim for his generous guidance during my doctoral studies. I would like to express my deepest gratitude and appreciation to Professor Sungho Won for guiding me and providing endless support for both education and research. In addition, is an honor for me to show my most sincere gratitude to Professor Chung Mo Nam, Professor Je Hyeong Kim, and Professor Ji Ye Jung, who continuously devoted attention to my subject and advised me on how to improve my work. In addition, I would like to thank Boram Park for all the helpful advice and statistical support.

I am extremely grateful to my parents for their love, prayers, and care, and for the sacrifices they made to educate and prepare me for the future. I am very much thankful to my wife and my son for their love, understanding, prayers and continuing support to complete this research work.

I would also like to thanks my friends and research colleagues, Dr. Eun Kyoung Goag, Professor Ji Eun Park, Professor Youngmok Park, Professor Eun hye Lee, Professor Jung Mo Lee, and Professor Sung Soo Ahn for their constant encouragement.

Although many of their names are not mentioned here, I am deeply grateful to all those who have shown a lot of help and affection in my doctoral program. Finally, my thanks go out to all who who supported me in completing this research, whether directly or indirectly.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

ABSTRACT

Factors associated with lung function decline in a Korean population-based cohort

Chi Young Kim

Department of Medicine
Yonsei University Graduate School of Medicine

(Directed by Professor Young Sam Kim)

Background : Pulmonary function measures are traits that predict morbidity and mortality and define chronic obstructive pulmonary disease (COPD). Several studies have shown that some genetic variants associated with COPD have been identified in genome-wide association studies (GWASs), especially in patients with moderate to severe COPD; however, genetic susceptibility for lung function decline in the general population has not been widely studied. This study aims to investigate the factors associated lung function decline with or without airflow obstruction, using data from a community-based cohort.

Methods : We evaluated characteristics of airflow obstruction and risk factors for the lung function decline of general population. We also conducted a genome-wide interaction study to identify the association between genetic variants and pulmonary function, and also examined how these variants relate to lung impairment in accordance with smoking status with amounts. Using community-based cohorts derived from the Korean Genome Epidemiology Study, we analyzed the association between genetic variants (single-nucleotide polymorphisms and haplotypes) and lung function (FEV_1, FEV_1/FVC ratio) using a linear mixed model for association and interaction to time effect.

Results : A total of 8845 subjects were recruited from two community-based

cohorts. When compared subjects with and without airflow obstruction clinical variable that such as age, sex, BMI, and smoking status were associated obstructive lung disease. We found annual mean FEV_1 declines of 41.7 mL for men and 33.4 mL for women, and the annual rate of decline in FEV_1 was fastest for current smokers. We also found a previously identified locus near *FAM13,* the most significant SNPs from the results of two LR tests for FEV_1/FVC. These selected SNPs were located in the upstream region of FAM13 on chromosome 4 and had similar minor allele frequencies (MAFs). Furthermore, we found that certain SNPs tended to have lower FEV_1/FVC values, and lung function decreased much faster with time interactions. The SNP most associated with lung function decline was the rs75679995 SNP on chromosome 7, and those SNPs located within the TAD of the *DNAH11* region and the eQTL of rs9991425 revealed a higher expression of *MFAP3L* and *AADAT* genes.

Conclusion: This is the first gene–time interaction study of lung function decline as a risk factor for COPD in the Korean population. In addition to replicating previously known signals for *FAM13A*, we identified two novel genomic regions (*DNAH11, AADAT*) involved in this gene-environmental interaction.

Key words: chronic obstructive pulmonary disease, genome-wide association study, single nucleotide polymorphism, airflow obstruction, *FAM13* gene, *DNAH11* gene, *AADAT* gene

Factors associated lung function decline in a Korean population-based cohort

Chi Young Kim

Department of Medicine
Yonsei University Graduate School of Medicine

(Directed by Professor Young Sam Kim)

I. INTRODUCTION

Lung function is an important trait of the respiratory system. Current research has provided evidence that lung health impairments, usually evidenced by low measures of the forced expiratory volume in one second (FEV_1), forced vital capacity (FVC), and the ratio of FEV_1 to forced vital capacity (FEV_1/FVC), are risk factors for morbidity and mortality in the general population.[1,2] Lung function decline can results from physiological lung aging, and numerous genetic and environmental factors can accelerate lung function decline.[3]

Lung function measurements of FEV_1 and the ratio of FEV_1/FVC are used as criteria for chronic obstructive pulmonary disease (COPD) diagnosis and the evaluation of pulmonary disease severity.[4,5] Notably, COPD, a disease characterized by airflow limitation, is one of the most common respiratory diseases and one of the leading causes of mortality worldwide.[6-8]

Multiple risk factors for COPD have been identified in the literature, and smoking has been recognized as the most significant risk factor for a rapid decline in lung function and consequent development of COPD.[9] However there are substantial differences in sensitivity to smoking among individuals, and these differences are partly attributable to genes and/or their interactions with smoking. Accordingly, genetic loci associated with lung function have

been shown to influence susceptibility to respiratory diseases including COPD. However, most identified variants through genome-wide association studies (GWASs) are common variants (minor allele frequency [MAF] > 5%) of the population.

Recently, multiple loci associated with pulmonary function and/or risk of COPD were identified using genome-wide association analyses.[10-19] A meta-analysis among four populations of the Cohorts for Heart and Aging Research in Genomic Epidemiology consortium reported that FEV_1/FVC is associated with chromosome 4. In particular, the chromosome 4q31 region near HHIP and the 4q22.1 region in FAM13A were related to FEV_1/FVC.[11,14-17,19] Moreover, a hereditary severe deficiency in alpha-1 antitrypsin, encoded by *SERPINA1* on chromosome 14, is a well-known genetic risk factor for the development of COPD, but severe alpha-1 antitrypsin deficiency accounts for only about 1% of patients with COPD.[20] Thus, improving our understanding of disease pathogenesis and progression would require studies on genetic susceptibility loci and their interactions with smoking.

These findings indicate that the biological mechanism explaining how the interaction of single-nucleotide polymorphisms (SNPs) and environmental factors (for example, smoking status) increases the prevalence of reduced pulmonary function and/or COPD is still unclear. In this study, we performed an evaluation of prevalence and incidence of COPD and identified clinical risk factors. Based on the clinical evidence for covariates, GWAS was conducted on the pulmonary function of two Korean population-based cohorts as part of the Korean Genome Epidemiology Study (KoGES). We investigated which genetic factors might contribute to pulmonary function as measured by FEV_1 and FEV_1/FVC and examined how they related to lung impairment and gene–time interactions

II. MATERIALS AND METHODS

1. Study Population

This study analyzed GWAS data from two independent cohorts from Ansan and Ansung those are included in the KoGES. Both cohorts enrolled Korean men and women aged 39–70 years between 2001–2002 to conduct a prospective investigation. Briefly, the KoGES collects epidemiological data (general information, disease history, questionnaire items such as diet and lifestyle, physical measurements, and the results of clinical examinations), genetic information (SNPs, etc.) and human resources (DNA, serum, plasma, etc.). The Ansung-Ansan cohort is a community-based, representative cohort in rural and urban areas of Korea, and the subjects included in the cohort were followed-up with every two years for the investigation of prospective region-based epidemiologic studies. The dataset collected in the Anseong-Ansan cohort was constructed through questionnaires and screening surveys, which is compatible with domestic epidemiological requirements. Detailed information on participant recruitment is available elsewhere.[21-23] A total of 5020 participants from Ansan and 5018 participants from Ansung took part in the baseline study from 2001 to 2003. Cohort members underwent a comprehensive health examination and a questionnaire-based interview, and biospecimens for assays were collected by health professionals at each study site. The questionnaire covered demographic characteristics, lifestyle choices and medical history.

2.Spirometric Lung Function Measurement

Pulmonary function tests were performed by a skilled technician using a portable spirometer (Vmax-2130, Sensor Medics,Yorba Linda, CA, USA) according to standardized protocols of the American Thoracic Society.[24] All participants performed a prebronchodilator spirometry test until completing at least three repeated measurements, and an acceptable measure was determined

to occur when the differences between the largest and the next largest FVC and FEV_1 values were within 0.15L. Calibration and quality control of spirometric examinations were also performed regularly based on American Thoracic Society guidelines.[24]

3. The Korea Association Resource Project

The Korea Association Resource Project (KARE), a multidisciplinary research consortium, began in 2007 to conduct a large-scale GWAS of Ansan and Ansung cohorts in the KoGES.[25] From the 10030 participants in both cohorts, we obtained genetic information from 10004 individuals on microarrays, but we excluded the data of 1476 individuals because they exhibited either a missing genotype call rate of <96%, heterozygosity <70%, gender consistencies (n=41), average pair-wise identity-by-state values higher than 0.8 as estimated values from first-degree relatives of Korean sib-pair samples, or medical history consisting of any kind of cancer. In addition, individuals who did not complete a pulmonary function test or anthropometric measurement and did not report smoking status were omitted.

4. Quality Control

For the discovery genome wide interaction study (GWIS) with KARE data, as well as for the validation analyses using the GENIE cohort, the QC of SNPs and subjects was conducted using PLINK23 and oneTOOL24. We excluded SNPs with P values on the Hardy–Weinberg equilibrium (HWE) analysis < 10^{-5}, minor allele frequencies (MAFs) < 0.05, and genotype call rates < 95%. Furthermore, we excluded subjects with missing genotype call rates > 5% or sex-based inconsistencies. After QC, 311,556 SNPs and 8554 participants were included for whole-genome imputation.

5. Genotype Imputation

For GWIS with KARE data, whole-genome imputation was performed using SHAPEIT2 and IMPUTE2 for pre-phasing data and genotype imputations. The 1000 Genomes Phase 3 was used as the reference panel. To maintain imputation quality, the estimated imputation accuracy for imputed SNPs was evaluated using the INFO metric, and any imputed SNPs with INFO < 0.5 were eliminated. The standard QC procedure was also applied for imputed SNPs, and 3,333,374 SNPs from 8554 participants were used for the GWIS discovery study (Figure 1). For the validation analyses, genotypes comprising the most significant SNPs were not originally genotyped, and target imputation was conducted. Target imputation for regions containing significantly associated SNPs was performed using IMPUTE2 with a buffer size of 5 million bp for each target SNP.

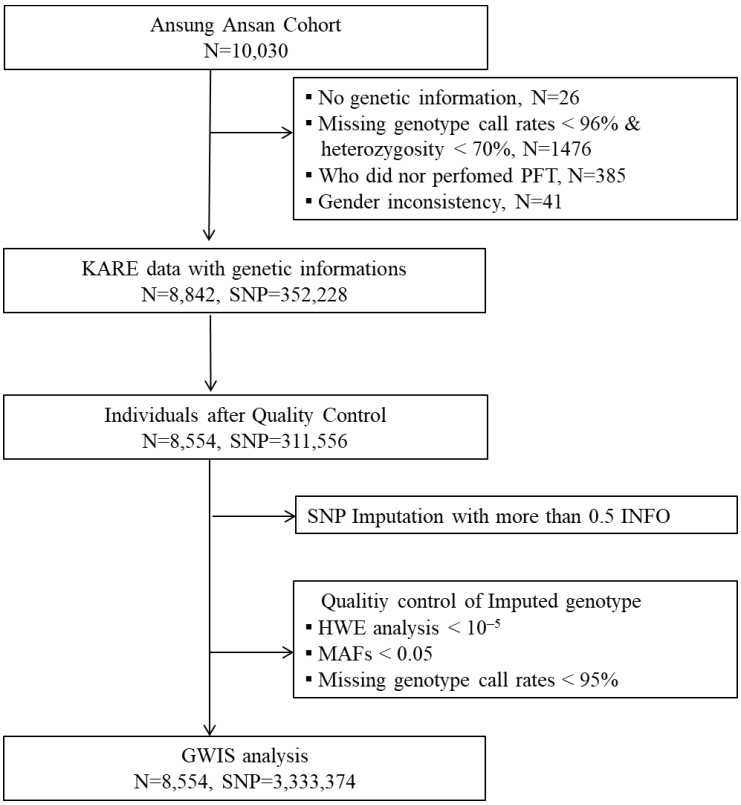


Figure 1. Flow diagram for KARE cohort. Fig. 1 explains how the individuals and SNPs were included and excluded.

Note : After quality controls and imputations, 8,845 participants and 3,333,374 SNPs were ultimately used for analyses.

6. Discovery analysis using KARE

Among these participants, 4,026 were men and 4,528 were women. We conducted GWAS on FEV_1 and FEV_1/FVC by linking the KARE data. The values of FEV_1 and FEV_1/FVC were observed up to seven times every two years (**Table 1.**), and at least one or more PFT observed individuals were included.

Genotype data were obtained using the Affymetrix Genome-Wide Human SNP array 5.0[25], and quality control analyses were performed participants underwent spirometry analysis, and their smoking history was recorded. After quality control (QC) process and imputation, 8554 indivisuals and 3,333,374 SNPs were included for the analyses.

Table 1. Number of follow-up spirometry measurements of KARE data

Follow up	1	2	3	4	5	6	7
N	1207	903	859	902	1154	1168	2361

Abbreviations: KARE, korea association resource Project

7. Validation analyses using GENIE cohort

The GENIE cohort comprised 7999 participants who visited Seoul National University Hospital Gangnam Center in 2014 and who agreed to provide blood samples and to participate in genetic studies.[26] Participants underwent genotype analysis using the Affymetrix Axiom KORV1.1–96 Array20, and genotype quality control (QC) was performed. The 4413 participants who had an

FEV_1/FVC ratio ≥ 70, were 40 years old or older, had available spirometric data, and had an available smoking history were included in the association analysis. Based on questionnaires, 2520 individuals were non-smokers, 1380 were former smokers, and 513 were current smokers. From this dataset, the 2520 non-smokers and 513 current smokers were included in the GWIS validation study.

8. Genome-wide interaction studies (GWIS) with KARE data.

FEV_1 and the FEV_1/FVC ratio were applied as spirometric measures, which were used to identify the genetic variants interacting with time. GWISs were conducted using KARE data. We conducted analyses using a linear mixed model with age, sex, height(Ht), body mass index (BMI), smoking pack year, and smoking status as covariates. To adjust for population substructure strictly, principal component (PC) analyses were applied to the genetic relationship matrix, and the first 10 PC scores were included as covariates. We let *yij* be the FEV_1 and FEV_1/FVC values for participant *i* at time point *j*, and they were assumed to follow a multivariate normal (*MVN*) distribution. The elapsed time from the baseline measurement and PC scores for participant *I* and component *k* were denoted as time*ij* and pc*i k*, respectively. Then, the linear mixed model for FEV_1, FEV_1/FVC was as follows.

$$FEV_{1_{ij}} = \beta_0 + \beta_1 age_i + \beta_2 sex_i + \beta_3 BMI_i + \beta_4 height + \beta_5 Smoking - status + \beta_6 Pack - years + \beta_7 TIME + \beta_8 SNP + \beta_9 TIME \cdot SNP + \Sigma_{K=1}^{10} \tau_k PC_i^k + b_{1i} Time + b_{0i} + \varepsilon_{ij}$$

$$\begin{bmatrix} b_0 \\ b_1 \end{bmatrix} \sim MVN(0, \Sigma),\ \left(\varepsilon_{i1}, \ldots, \varepsilon_{in_i}\right)^t \sim MVN(0, \Sigma')$$

$$FEV_1/FVC_{ij} = \beta_0 + \beta_1 age_i + \beta_2 sex_i + \beta_3 BMI_i + \beta_4 height + \beta_5 Smoking - status + \beta_6 Pack - years + \beta_7 TIME + \beta_8 SNP + \beta_9 TIME \cdot SNP + \Sigma_{K=1}^{10} \tau_k PC_i^k + b_{1i} Time + b_{0i} + \varepsilon_{ij}$$

$$\begin{bmatrix} b_0 \\ b_1 \end{bmatrix} \sim MVN(0, \Sigma),\ \left(\varepsilon_{i1}, \ldots, \varepsilon_{in_i}\right)^t \sim MVN(0, \Sigma')$$

It should be noted that pack years are 0 for never smokers and were included as covariates only for ever smokers. We compared several structures for Σ and Σ', and selected an unstructured covariance structure. The proposed models were applied to detect gene-time interactions for average FEV_1 or FEV_1/FVC levels. To identify SNPs interacting with time on spirometric measures, we considered H0: $\beta 8 = \beta 9 = 0$. This could be tested by summing a likelihood ratio test with 2 degrees of freedom (DF) for ever smokers and a likelihood ratio test. The most significant SNPs were selected for further analyses of gene-time interaction effects.

III. RESULTS

1. Baseline Patient Characteristics

Table 2 describes the baseline characteristics of all participants who were included in the analysis. In total, 8,554 participants (57.1% men, 52.9% women) were included in our analysis. The participants' mean age was 52.1 years old, and the percentages of never smokers, former smokers, and current smokers were 60%, 15.4%, and 24.6%, respectively. Pre-bronchodilator FEV_1, FVC, and FEV_1/FVC values were significantly different between males and females (Table 2.). Smoking history was obtained through a questionnaire, and smoking status and pack years were used for association analyses as covariates. Smoking status had three categories: never smokers, former smokers, and current smokers. Never smokers were defined as individuals who had never smoked, and former smokers were participants who had smoked previously but had stopped smoking prior to the survey. Current smokers were individuals who stated that they currently smoked during the investigation, or who had a record of smoking and did not belong to the other two categories. According to our categorization, there were 5,134 never smokers, 1,314 former smokers, and 2,016 current smokers in our cohort.

2. Rate of decline in FEV_1 in males and females.

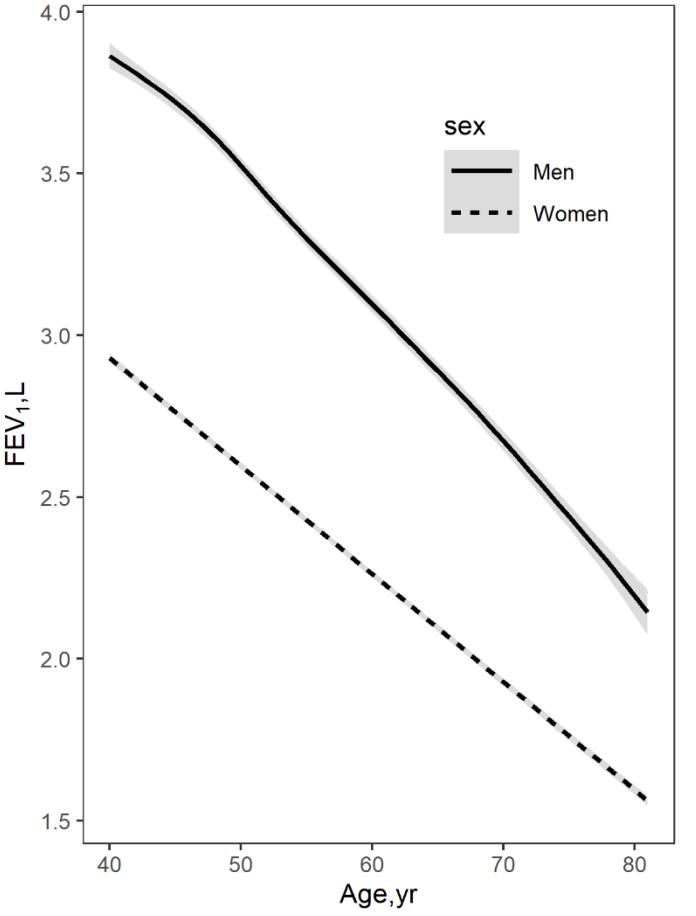
Figure 2 shows the decline in FEV_1 with age in males and females. Men exhibited higher absolute FEV_1 values than did women at all time points. The annual rate of decline was significantly more rapid for male patients (41.73 mL; 95% confidence interval [CI], 41.15–42.31) than for their female counterparts (33.36 mL; 95% CI, 33.12–33.59; $P < 0.001$).

Table 2. Baseline characteristics of male and female subjects included in the study

	Males (n=4,026)	Females (n=4,528)	Total (n=8,554)	p-value
Age, years, mean	51.7 ± 8.7	52.5 ± 9.0	52.1 ± 8.8	< 0.001
Body mass index, kg/m^2, mean	24.4±2.8	24.9±3.2	24.6±3.1	< 0.001
Height, cm	167.0 ± 5.8	153.8 ± 5.5	160.0 ± 8.7	< 0.001
Smoking status, N (%)				< 0.001
Never smoker	805 (20.0)	4329 (95.6)	5134 (60.0)	
Former smoker	1259 (31.6)	55 (1.2)	1314 (15.4)	
Current smoker	1962 (48.7)	144 (3.2)	2106 (24.6)	
Smoking, pack-year	24.2 ± 17.4	6.8 ± 9.9	22.9 ± 17.6	< 0.001
FEV_1, % predicted, mean	106.4 ± 16.2	116.6 ± 17.6	111.8 ± 17.7	< 0.001
FVC, % predicted, mean	102.0 ± 13.9	107.1 ± 15.0	104.7 ± 14.7	< 0.001
FEV_1/FVC, %	78.0 ± 13.9	81.7 ± 6.6	79.9 ± 7.7	< 0.001

Note : Data presented as mean ± standard deviation or n (%).

Abbreviations: FEV_1, forced expiratory volume in 1 second; FVC, forced vital capacity.



	FEV1 decline (ml/yr)*	P-value
Male	41.7 (41.2~42.3)	<0.001
Female	33.4 (33.1~33.6)	

Figure 2. Comparison of the annual mean decline in FEV_1 for male (n = 4,026) and female patients (n = 4,528).

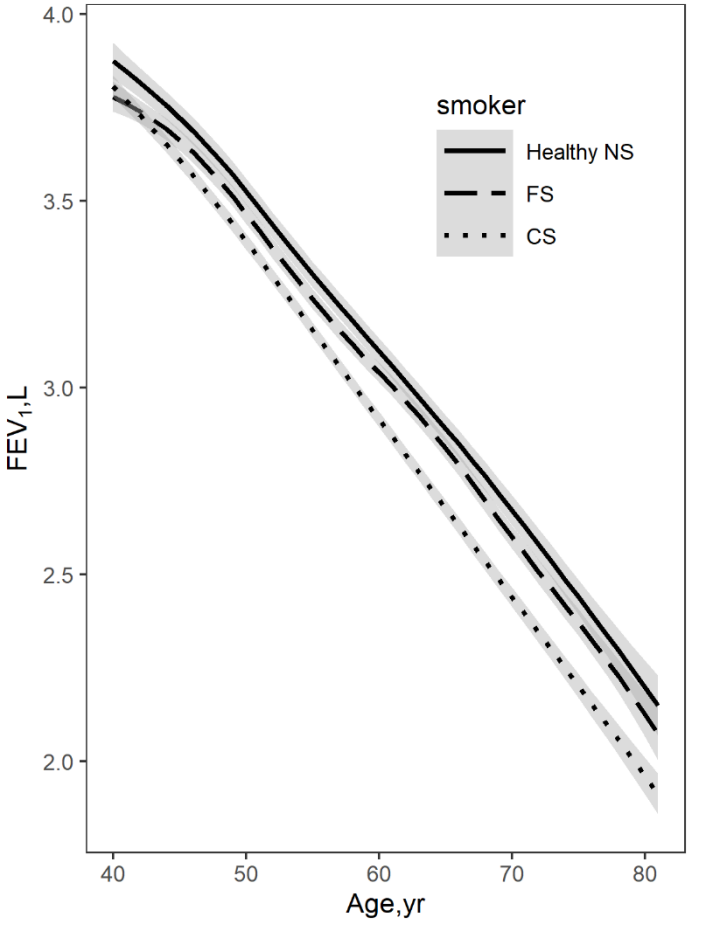
Note : *Data were presented as mean and 95% confidence intervals

Abbreviations: FEV_1, forced expiratory volume in 1 second.

3. Effect of smoking on the decline in lung function.

We compared the decline in FEV_1 among male healthy never smokers, former

smokers with and without obstructive lung disease (OLD), and current smokers with and without OLD.



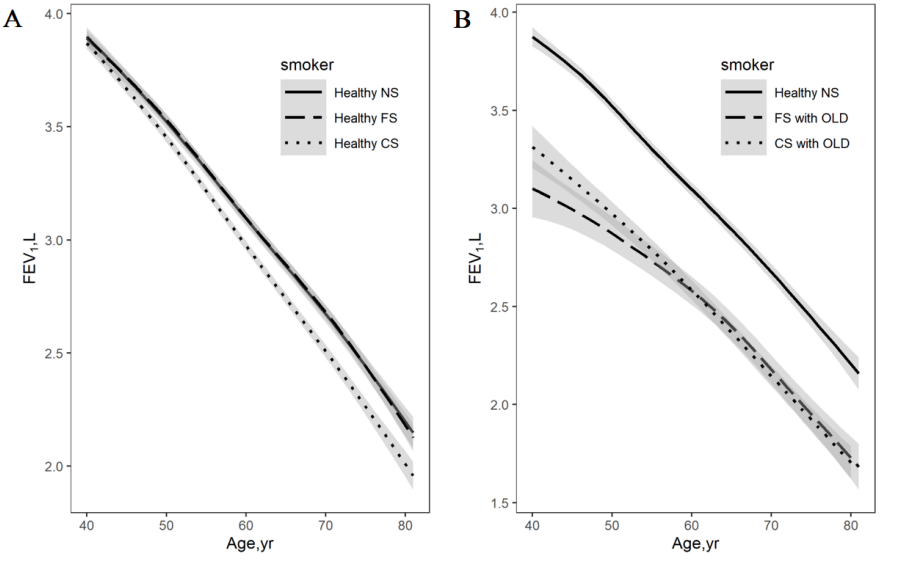
	FEV_1 decline (ml/yr)[a]	P-value[b]
Healthy never smoker	41.8 (41.1~42.5)	
Former smoker	41.8 (41.0~42.7)	0.948
Current smoker[c]	46.3 (45.5~47.1)	<0.001

Figure 3. Comparison of the annual mean decline in FEV_1 among male healthy never smokers, former smokers, and current smokers.

Note : [a]Data were presented as mean and 95% confidence intervals
[b]*P*-value compared to healthy never smoker, [c]*P*<0.001 compared to former smoker. **Abbreviations:** FEV_1, forced expiratory volume in 1 second.

Figure 3 presents comparisons of the annual mean decline in FEV_1 among healthy never smokers (without OLD), former smokers (with and without OLD), and current smokers (with and without OLD). The annual rate of decline in FEV_1 was fastest for current smokers (46.3 mL; 95% CI, 45.2–46.8; $P < 0.001$ vs. healthy never smokers and former smokers), while it was more rapid for former smokers (41.8 mL; 95% CI, 41.0–42.7; $P = 0.9482$) than for healthy never smokers (41.8 mL; 95% CI, 41.1–42.5; Fig. 3). Current smokers (46.3 mL) showed the fastest annual rate of decline in FEV_1, followed by former smokers (41.8 mL) and healthy never smokers (41.7 mL), in that order.

Figure 4 shows comparisons among male healthy never smokers, healthy former smokers, healthy current smokers, former smokers with OLD, and current smokers with OLD. Figure 4A compares the decline in FEV_1 among male healthy never smokers, healthy former smokers, and healthy current smokers. The annual rate of decline in FEV_1 was fastest for healthy current smokers (46.0 mL; 95% CI, 45.2–46.8; $P < 0.001$) vs. healthy never smokers, while it was statistically indistinguishable between healthy former smokers (41.8 mL; 95% CI, 40.8–42.4 mL) and healthy never smokers ($P = 0.948$). Figure 4B compares the decline in FEV_1 among male healthy never smokers, former smokers with OLD, and current smokers with OLD. The annual rate of decline in FEV_1 is slowest for former smokers with OLD (36.6 mL).



	FEV_1 decline (ml/yr)	P-value
Healthy never smoker	41.72 (41.07~42.37)	
Healthy former smoker	41.58 (40.75~42.42)	0.8676
Healthy current smoker	45.96 (45.17~46.76)	<0.001

	FEV_1 decline (ml/yr)	P-value
Healthy never smoker	41.53 (40.82~42.26)	
Former smoker with OLD	36.58 (34.75~38.41)	0.0067
Current smoker with OLD	41.76 (40.27~43.25)	0.8802

Figure 4. Comparison of the annual mean decline in FEV_1 according to smoking status and the presence or absence of OLD in men.

Note : Data were presented as mean and 95% confidence intervals. *P*-value compared to healthy never smokers

(**A**) Comparison of the annual mean decline in FEV_1 among healthy never smokers, healthy former smokers, and healthy current smokers. The annual rate of decline in FEV_1 is fastest for healthy current smokers (46.0 mL). (**B**) Comparison of the annual mean decline in FEV_1 among healthy never smokers, former smokers with OLD, and current smokers with OLD.

Abbreviations: FEV_1, forced expiratory volume in 1 second; OLD, obstructive lung disease.

4. The Prevalance of COPD and Baseline Characteristics of the Subjects

Among the 8,554 baseline subjects, COPD was observed in 738 (8.6 %) (**Table 3**). The mean age of the COPD group was 58.5± 8.7, which was older than that of the non-COPD group (51.6±8.7, $p<0.001$). The proportion of males was higher in the COPD group compared with the non-COPD group. (74.1% vs. 44.5%, $P < 0.001$). The mean % predicted values of FEV_1 and FEV_1/FVC were significantly lower in the COPD group than in the non-COPD group ($p<0.001$). The proportion of former and current smokers was higher in the COPD group ($p<0.001$), and smoking amount of the COPD group (30.7±18.8 pack-year) was larger than that of the non-COPD group (22.2±16.8 pack-year) ($p<0.001$). Average body mass indexes (BMIs) were 23.7±3.1 kg/m^2 and 24.7±3.0 kg/m^2 in the COPD and non-COPD groups, respectively, with BMIs being significantly lower in the COPD group. In addition, lung function was significantly lower in the COPD group.

5. Risk factors for incidence of COPD

During the follow-up period, 10.8% (n=844) of patients developed COPD. **Table 4** shows the results of a univariate logistic regression analysis of the risk factors for COPD. Univariate analysis revealed that age, sex, BMI, smoking, status, and socioeconomic status including education and income were independent prognostic factors.

Table 3. Baseline characteristics of airflow obstruction subjects included in the study

	Normal (n=7,816)	COPD (n=738)	Total (n=8,554)	p-value
Age, years, mean	51.6 ± 8.7	58.5 ± 8.1	52.1 ± 8.8	< 0.001
Sex, male, N (%)	3479 (44.5)	547 (74.1)	4026 (100)	< 0.001
Body mass index, kg/m^2, mean	24.7 ± 3.0	23.7 ± 3.1	24.6 ± 3.1	< 0.001
Height, cm	159.8 ± 8.7	162.4 ± 7.8	160.0 ± 8.7	< 0.001
Smoking status, N (%)				< 0.001
Never smoker	4801 (62.0)	220 (30.1)	5021 (59.3)	
Former smoker	1144 (14.8)	165 (22.6)	1309 (15.5)	
Current smoker	1795 (23.2)	346 (47.3)	2141 (25.3)	
Smoking, pack-year	22.2 ± 16.8	30.7 ± 18.8	23.4 ± 17.4	< 0.001
FEV_1, % predicted, mean	113.7 ± 16.4	91.6 ± 17.5	111.8 ± 17.7	< 0.001
FVC, % predicted, mean	104.6 ± 14.4	106.0 ± 17.3	104.7 ± 14.7	0.027
FEV_1/FVC, %	81.5 ± 5.5	62.8 ± 6.6	79.9 ± 7.7	< 0.001

Abbreviations: COPD, chronic obstructive pulmonary disease; FEV_1, forced expiratory volume in 1 second; FVC, forced vital capacity.

Table 4. Univariate logistic regression analysis of the risk factors for COPD

Variables	HR (95% CI)	P-value
Age, years	1.05 (1.05–1.06)	< 0.001
Male sex	3.61 (3.09–4.23)	< 0.001
Body mass index, kg/m^2	0.91 (0.89–0.94)	< 0.001
Height, cm	1.05 (1.04–1.06)	< 0.001
Smoking status	1.92 (1.77–2.08)	< 0.001
Smoking, Pack-Years	1.02 (1.01–1.02)	0.028
Education, highest quartile	0.85 (0.73–0.99)	0.036
Income, highest quartile	0.68 (0.58–0.78)	< 0.001
Area, Rural	0.48 (0.37–0.61)	< 0.001
HTN	0.87 (0.71–1.07)	0.201
DM	1.02 (0.76–1.34)	0.899
CAOD	1.07 (0.44–2.21)	0.860
Dyslipidemia	1.09 (0.69–1.65)	0.683

Abbreviations: HTN, hypertension; DM, diabetes mellitus; CHF; congestive heart failure, CAOD; coronary artery occlusive disease.

Figure 5. shows the results of a multivariate logistic regression analysis of the risk factors for COPD. After adjusting for confounding variables of age, sex, BMI, and smoking status were confirmed risk factor for develop COPD.

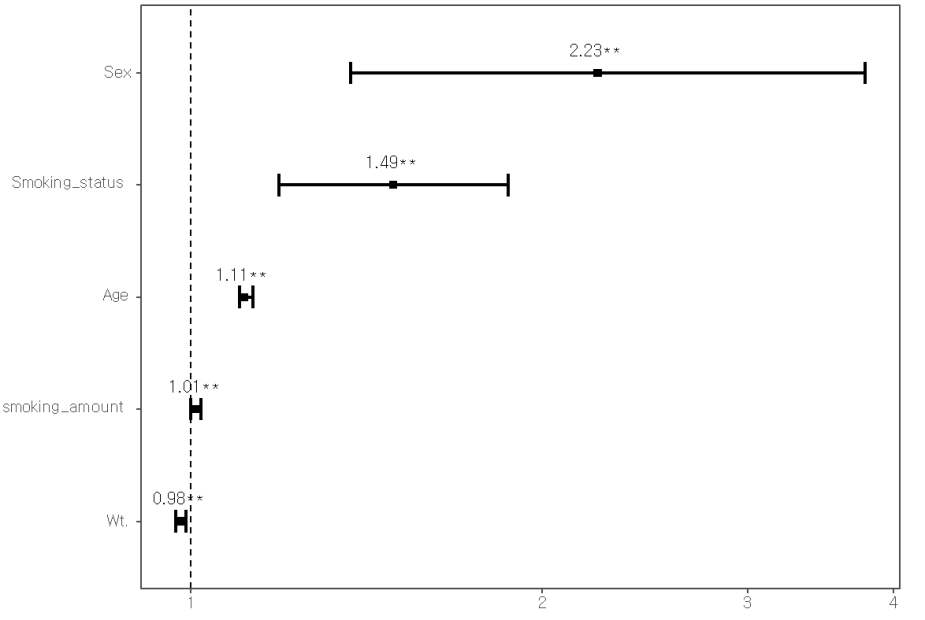


Figure 5. Odds ratio plot of multivariate logistic regression analysis of the risk factors for COPD.

Note : ***P*<0.001 in multivariate logistic regression analysis

6. GWISs of FEV_1 with KARE data

GWISs were conducted associations of 3.333,515 SNPs were tested by applying the 2 DF test to KARE data. For the GWISs of FEV_1**. Figure 6A** presents the quantile-quantile (QQ) plot for 2 LR tests, and variance-inflation factors (VIF) was estimated at 1.04 for FEV_1. There was no suggestion of an inflated type I error with a genomic inflation factor. **Figure 6B** is a Manhattan (MH) plot of results from the two LR tests. Logarithms of the 2 DF P-values of 3.333,515 SNPs were plotted against its physical chromosomal position. The red and blue horizontal lines represent 1.71×10^{-7}(Bonferroni adjusted 0.05 significance level) and 10^{-6} respectively. A Manhattan plot of results from the 2 LR test shows that four SNPs associated with FEV_1 to the GWAS significance level are located on

chromosome 3 and chromosome 17.

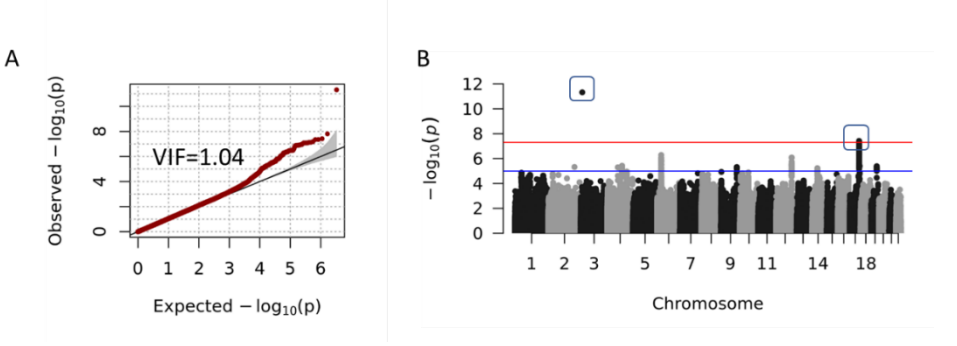


Figure 6. QQ plot and MH Plot (SNP, SNP*Time 2df LRT).

Note : Figure 6A is a QQ plot of the study results showing observed versus expected log10(P) values for forced expiratory volume in 1 s (FEV_1). Figure 6B is a Manhattan plot of results from the 2 LR test from the proposed 2 DF tests on FEV_1.

Based on likelihood 2 LR tests, 4 SNPs was identified as significant. The top 10 most significant p-values obtained after analysis are presented in Table 5. Regional association plots at the most significant loci on chromosomes 3 and 17 associated with lung function and FEV_1 are shown in **Figure 7.** The SNP most associated with lung function decline as measured by with FEV_1 by SNP and SNP-by-time LRT was rs2272402, which is found within the *SLC6A1*, *LSC6A11*, *HRH1*, *ATG7* genes on chromosome 3, and rs2190426, which is found within the CASC17 gene on chromosome 17. Genomic region on chromosome 3 near rs2272102 was located at gene dense region. However, the significant SNPs were determined by considering the false discovery rate (FDR), which controls for the expected proportion of false rejection of the null hypothesis. We used gene expression profiling data from a previous RNA-seq experiment using lung tissues from 56 participants with COPD.[27] The FDR rate of candidate gene was above 0.05 considering there was an undetermined significant SNP–time interaction compared other COPD gene microarray study in GEO database.

Table 5. Association results for the top 10 significant SNPs (SNP, SNP*Time 2LRT)

Chr	dbSNP.RS.ID	BP	Minor	MAF	HWE	MISS	Info	SNP Beta (SE)	SNP P-value	Interaction Beta (SE)	Interaction P-value	LR
3	rs2272402	11075461	A/G	0.075	0.064	0	1	0.024(0.012)	0.046586	-0.016(0.002)	5.50E-13	4.76E-12
17	rs2190456	69211366	T/C	0.382	0.569	0.016	0.993	-0.035(0.007)	9.93E-08	-0.001(0.001)	0.574398	3.81E-08
17	rs983085	69212061	G/A	0.382	0.6	0.014	0.994	-0.035(0.007)	1.24E-07	-0.001(0.001)	0.544067	4.42E-08
17	rs9905278	69211683	T/C	0.382	0.616	0.014	0.993	-0.035(0.007)	1.23E-07	-0.001(0.001)	0.548927	4.45E-08
17	rs72873967	69212318	A/G	0.381	0.616	0.012	0.994	-0.035(0.007)	1.71E-07	-0.001(0.001)	0.550932	6.52E-08
17	rs11650277	69212795	A/T	0.378	0.714	0.014	0.993	-0.035(0.007)	1.78E-07	-0.001(0.001)	0.57192	7.37E-08
17	rs4313845	69205148	C/T	0.378	0.664	0.01	0.996	-0.035(0.007)	1.39E-07	-0.001(0.001)	0.660005	7.44E-08
17	rs1107305	69200643	T/G	0.378	0.648	0.01	0.996	-0.035(0.007)	1.65E-07	-0.001(0.001)	0.626724	8.07E-08
17	rs8066924	69209694	C/T	0.378	0.681	0.011	0.995	-0.035(0.007)	1.71E-07	-0.001(0.001)	0.619475	8.23E-08
17	rs8066934	69209923	T/A	0.378	0.698	0.011	0.995	-0.035(0.007)	1.74E-07	-0.001(0.001)	0.615505	8.29E-08

Abbreviations: BP, physical position; MAF, minor allele frequencies; HWE, Hardy-Weinberg equilibrium; SE, standard error.

The top 10 most significant p-values obtained after analysis.

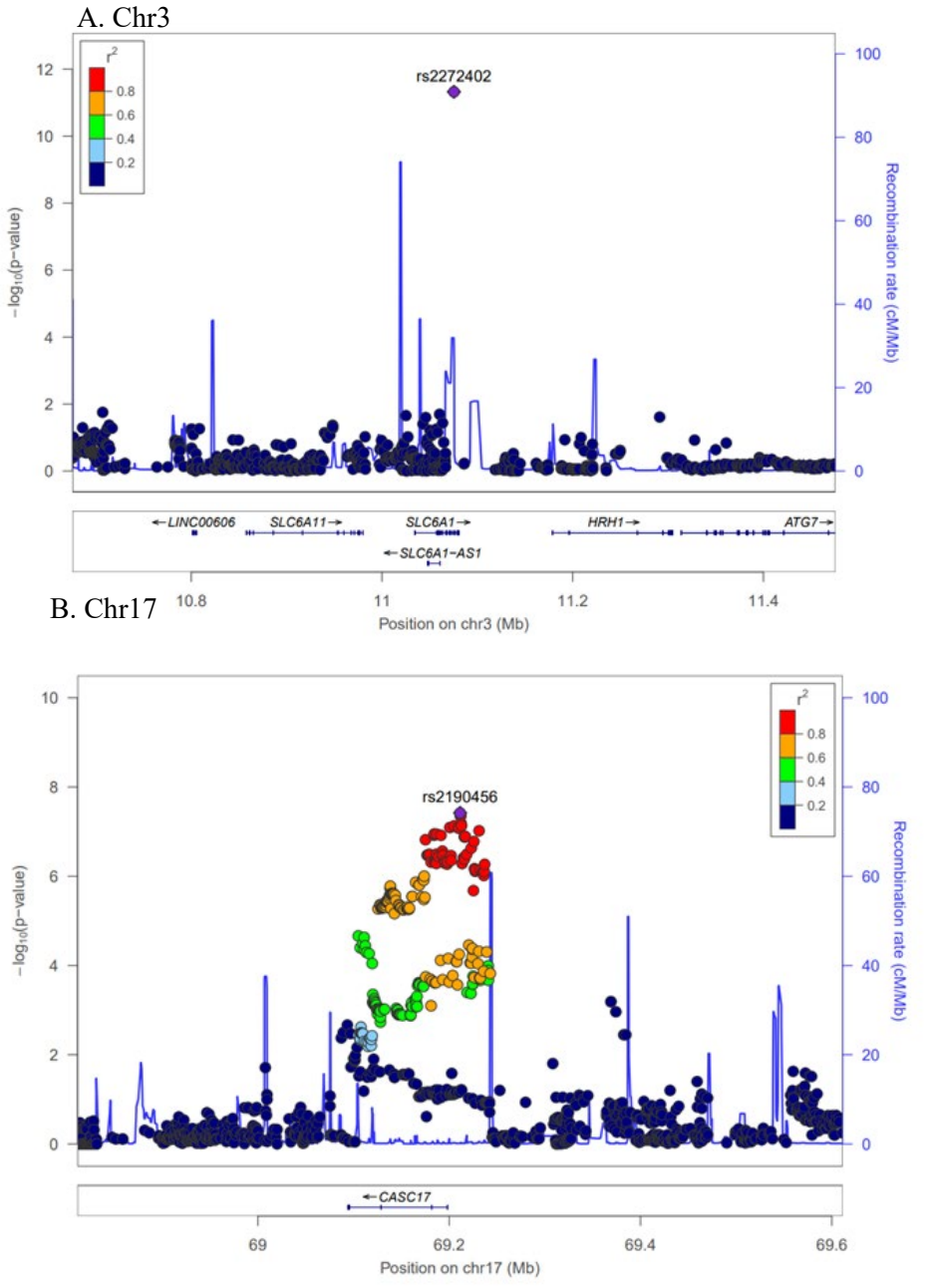


Figure 7. Regional plot (CHR4) - SNP, SNP x Time 2LRT.

Note : Regional association plots at the most significant loci associated with lung

function. The plots of genetic loci on chromosomes 3 (A) and 17 (B) are shown as they were created using LocusZoom.

7. Effects of time and gene interactions on FEV_1

Figure 8A shows the QQ-plot for interaction effect tests, and it shows that VIF was 1.01. Figure 8B is a Manhattan plot of results from interaction effect. To assess the significance of the identified interactions, we applied statistical significance thresholds that are routinely used for genome wide association studies. We specified a genome-wide significance threshold of $5x10^{-8}$, while we also applied a more liberal genome-wide suggestive threshold of $1x10^{-05}$ to include interactions with P-values that were slightly above genome-wide significance but that may represent genuine interactions. We also investigated time interactions and meeting a P-value significance threshold of 0.05 (nominal significance) were included in the interaction analysis. From GWISs with KARE, gene–time interaction effects were conducted, and the most significant SNPs are summarized in Table 6.

A B.

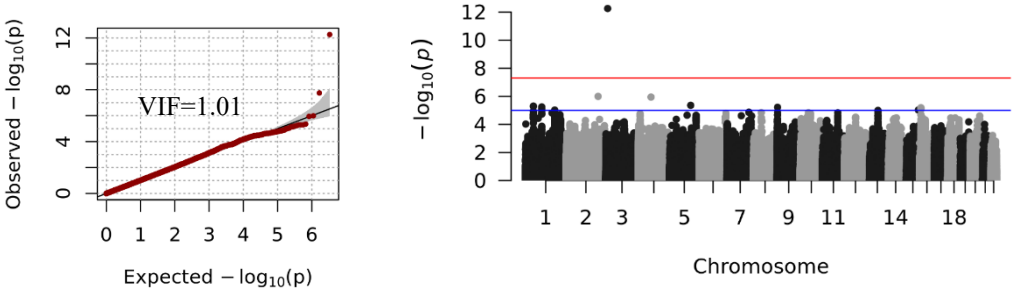


Figure 8. QQ & MH Plot (Interaction Effects).

Note : Figure 8A is a QQ plot and Figure 8B is a Manhattan plot of the results of interaction effect on FEV_1.

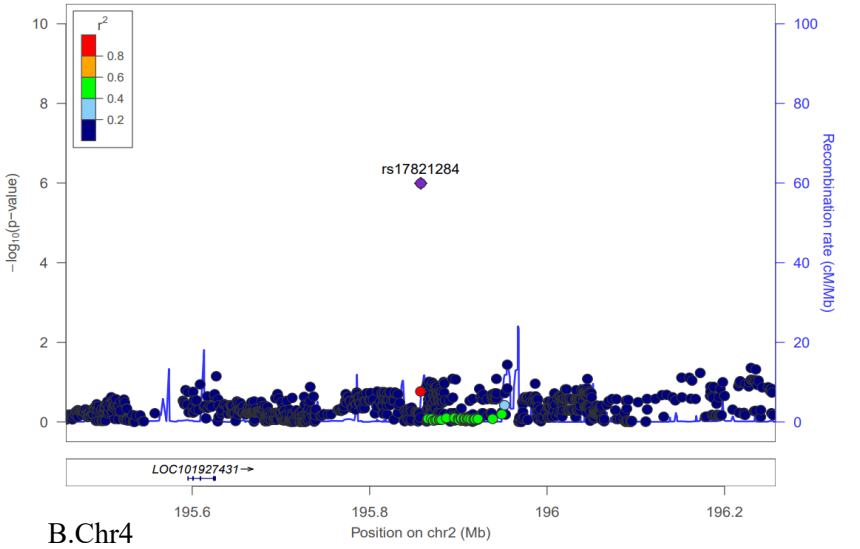
Table 6. Top Significant SNPs (Interaction Effect)

Chr	dbSNP.RS.ID	BP	Minor	MAF	HWE	MISS	Info	SNP Beta (SE)	SNP P-value	Interaction Beta (SE)	Interaction P-value	LR
3	rs2272402	11075461	A/G	0.073	0.02	0.038	1	0.024(0.012)	0.046586	-0.016(0.002)	5.50E-13	4.76E-12
2	rs17821284	195857628	C/G	0.065	0.15	0.042	1	0.011(0.014)	0.410568	-0.012(0.002)	1.02E-06	4.77E-06
4	rs4241605	78161506	T/C	0.053	0.017	0.032	1	0.035(0.015)	0.020302	-0.013(0.003)	1.14E-06	5.07E-06
5	rs465040	129173776	A/G	0.051	0.001	0.03	1	0.027(0.019)	0.148793	-0.013(0.003)	4.40E-06	2.57E-05
1	rs71647734	49746671	T/C	0.346	0.382	0.011	0.998	0.002(0.007)	0.78424	-0.006(0.001)	5.10E-06	1.41E-05
1	rs34201294	49754798	G/A	0.346	0.382	0.011	0.998	0.002(0.007)	0.786524	-0.006(0.001)	5.12E-06	1.41E-05
1	rs34002388	49758840	A/G	0.346	0.382	0.011	0.999	0.002(0.007)	0.786524	-0.006(0.001)	5.12E-06	1.41E-05
1	rs2809818	101005855	T/C	0.131	0.107	0.034	0.972	0.01(0.01)	0.316284	-0.008(0.002)	5.71E-06	3.03E-05
9	rs2210369	2453208	G/A	0.438	0.217	0.042	0.972	-0.019(0.006)	0.004217	0.005(0.001)	6.09E-06	1.18E-05
16	rs59245292	8449193	G/A	0.414	0.894	0.031	0.969	-0.011(0.007)	0.086176	0.005(0.001)	6.50E-06	3.72E-05
16	rs73495481	8448026	T/C	0.413	0.894	0.025	0.987	-0.012(0.007)	0.069853	0.005(0.001)	7.23E-06	3.97E-05

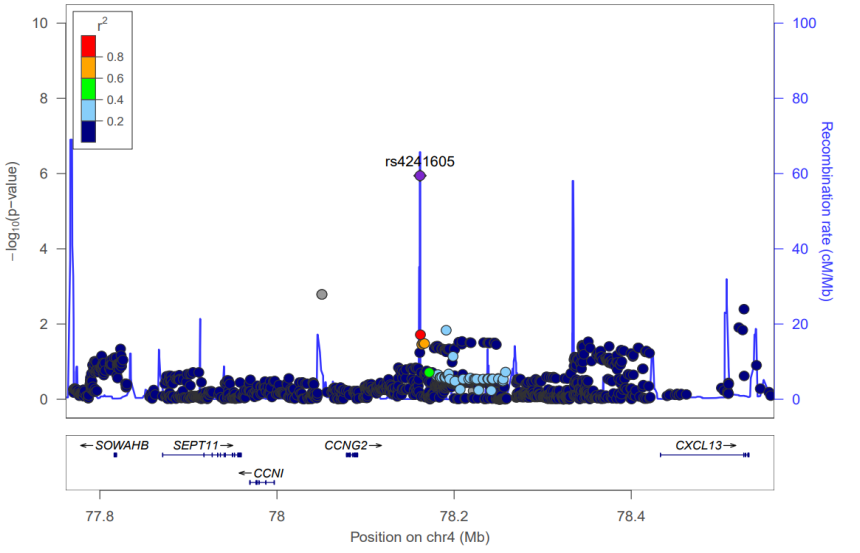
Abbreviations: BP; physical position; MAF, minor allele frequencies; HWE Hardy-Weinberg, equilibrium; SE, standard error.

Figure 9 shows the regional plot for most significant SNPs on chromosome 2, chromosome 4, chromosome 5, chromosome 1, chromosome 9, and chromosome 16. Unfortunately, our study's GWIS data of our study was not met FDR P value with previous miccroarray gene expression study of COPD.

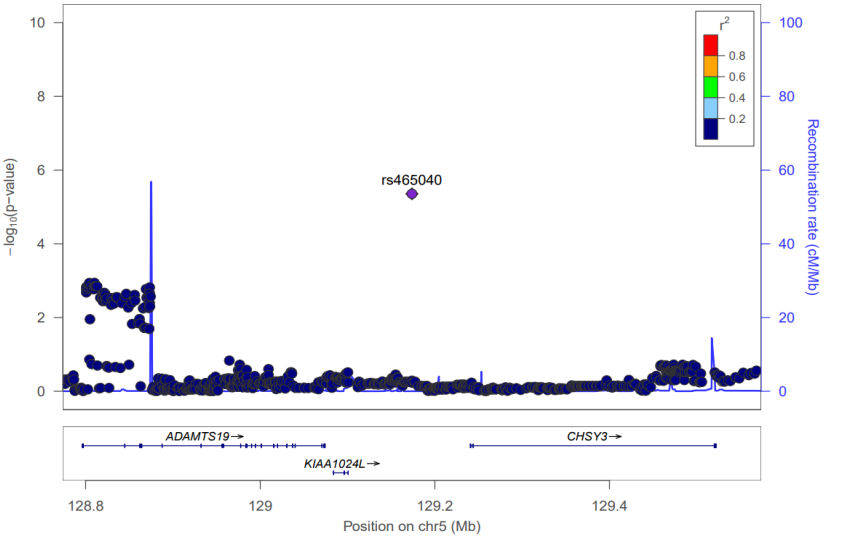
A.Chr2



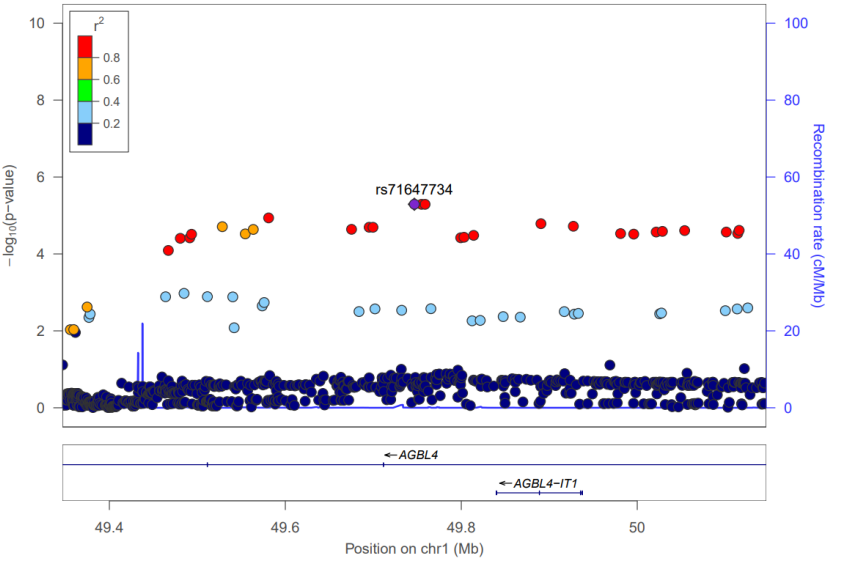
B.Chr4



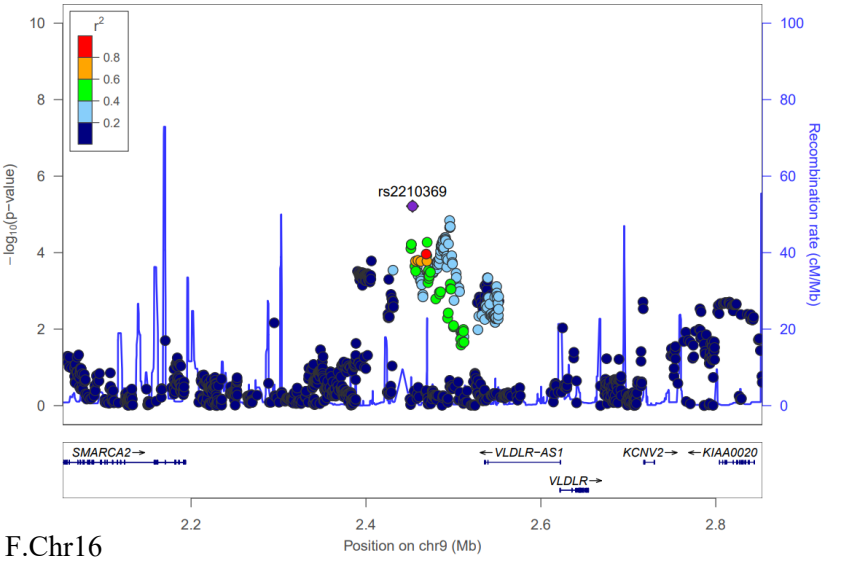
C.Chr5



D.Chr1



E. Chr9



F.Chr16

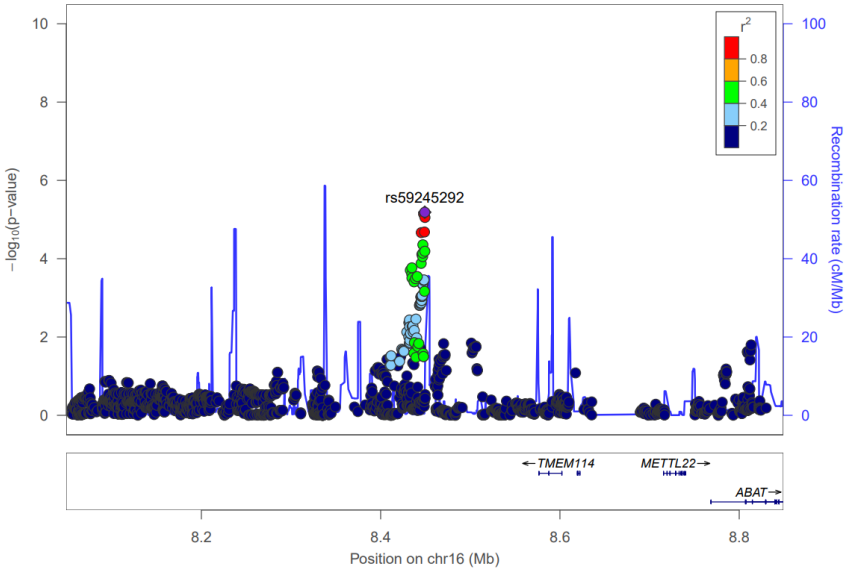


Figure 9. Regional Plot (Interaction Effect).

Note : Regional association plots at the most significant loci associated with lung function as measured by FEV_1 by SNP-by-time interaction for chromosomes 2 (A), 4 (B), 5 (C), 1 (D), and 9 (E).

8. GWISs of FEV_1/FVC with KARE data

GWISs were conducted associations of 3.333,515 SNPs were tested by applying the 2 DF test to KARE data**. Figure 10** shows the quantile-quantile (QQ) plot and the Manhattan (MH) plot of results from the 2 LR test. A Manhattan plot of results from the 2 LR test shows that 89 SNPs associated with FEV_1/FVC at the genome-wide significance level were located on chromosome 4 in the FAM13A gene.

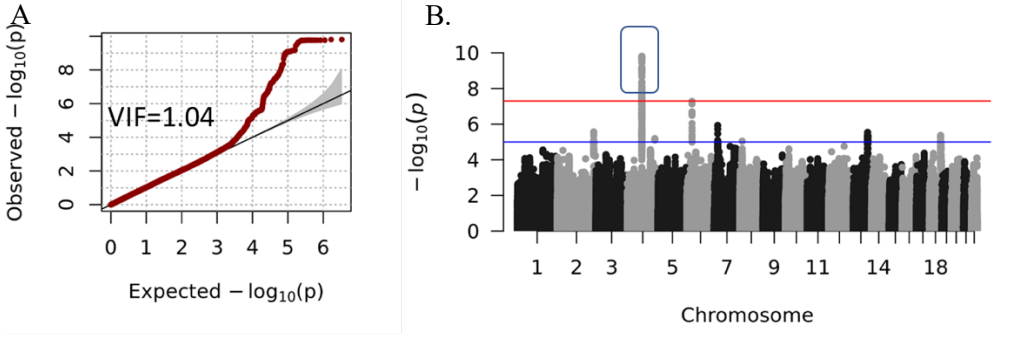


Figure 10. QQ Plot & MH Plot (SNP, SNP*Time 2df LRT).

Note : Figure 10A is a QQ plot of the study results showing observed versus expected log10(P) values for FEV_1/FVC. Figure 10B is a Manhattan plot of results from the 2 LR test from the proposed 2 DF tests on FEV_1/FVC.

Table 7 shows the 10 most significant SNPs from the results of two LR tests for FEV_1/FVC. These selected SNPs were located in the upstream region of *FAM13* on chromosome 4 and had similar minor allele frequencies (MAFs). In our study, the SNP most associated with FEV_1/FVC was the rs2446304 SNP, yielding $p = 1.56 \times 10^{-10}$. Figure 11 is a regional plot r^2 around rs2446304 SNPs created with LocusZoom. *FAM13* within the topologically associated domain (TAD) is located between 89.6 to 90.2 Mb.

Table 7. Top 10 Significant SNPs (SNP, SNP*Time 2LRT)

CHR	SNP	BP	Minor	MAF	HWE	MISS	Info	SNP BETA (SE)	SNP P-value	Interaction Beta (SE)	Interaction P-value	LR
4	rs2446304	89825872	C/A	0.477	0.83	0.003	0.999	0.665(0.104)	1.40E-10	-0.012(0.024)	0.614241	1.56E-10
4	rs2446303	89826058	T/C	0.477	0.83	0.003	0.999	0.665(0.104)	1.40E-10	-0.012(0.024)	0.614241	1.56E-10
4	rs2704581	89815404	A/G	0.477	0.814	0.003	0.999	0.664(0.104)	1.51E-10	-0.012(0.024)	0.611959	1.71E-10
4	rs2464528	89817664	C/A	0.477	0.814	0.003	0.999	0.664(0.104)	1.51E-10	-0.012(0.024)	0.611959	1.71E-10
4	rs2704582	89818521	T/C	0.477	0.814	0.003	0.999	0.664(0.104)	1.51E-10	-0.012(0.024)	0.611959	1.71E-10
4	rs2704583	89818643	A/G	0.477	0.814	0.003	0.999	0.664(0.104)	1.51E-10	-0.012(0.024)	0.611959	1.71E-10
4	rs1246629	89819423	T/C	0.477	0.814	0.003	0.999	0.664(0.104)	1.51E-10	-0.012(0.024)	0.611959	1.71E-10
4	rs1246631	89820287	G/A	0.477	0.814	0.003	0.999	0.664(0.104)	1.51E-10	-0.012(0.024)	0.611959	1.71E-10
4	rs1246632	89822166	C/T	0.477	0.814	0.003	0.999	0.664(0.104)	1.51E-10	-0.012(0.024)	0.611959	1.71E-10
4	rs1246634	89823798	G/A	0.477	0.814	0.003	0.999	0.664(0.104)	1.51E-10	-0.012(0.024)	0.611959	1.71E-10

Abbreviations: BP, physical position; MAF, minor allele frequencies; HWE, Hardy-Weinberg equilibrium; SE, standard error.

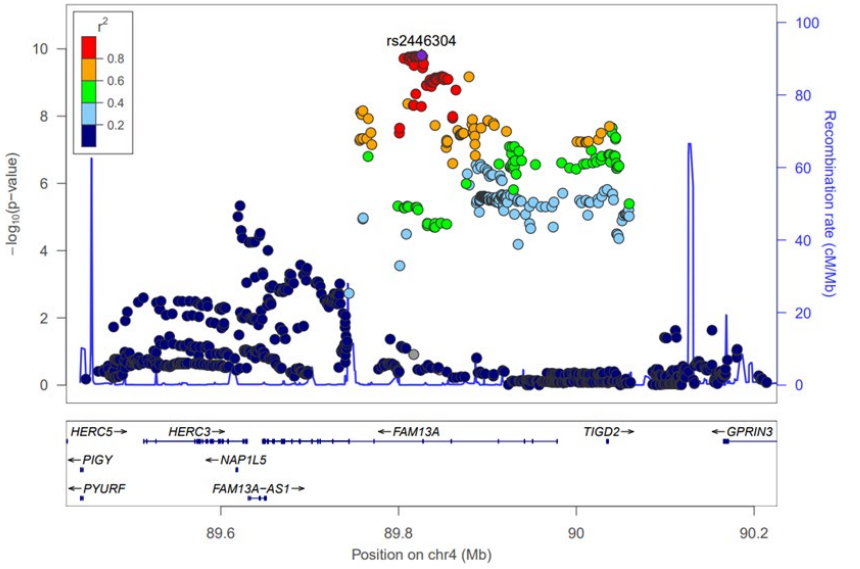


Figure 11. Regional Plot (CHR4) - SNP, SNP x Time 2LRT.

Note : Regional association plots at the most significant loci associated with lung function as measured by FEV_1/FVC by SNP-by-time interaction for chromosomes 4.

9. Effects of time and gene interactions on FEV_1/FVC

Figure 12A shows the QQ-plot for the interaction effect tests, and it shows that VIF was 1.03. **Figure 12B** is a Manhattan plot of results from interaction effect. To assess the significance of the identified interactions, we applied statistical significance thresholds that are routinely used for genome wide association studies. We specified a genome-wide significance threshold of $5x10^{-8}$, while we also applied a more liberal genome-wide suggestive threshold of $1x10^{-05}$ to include interactions with P-values that were slightly above genome-wide significance but that may represent genuine interactions. We also investigated time interactions and meeting a P-value significance threshold of 0.05 (nominal significance) were included in the interaction analysis. From GWISs with KARE, data–time interaction effects were conducted, and the most significant SNPs are

summarized in **Table 8.**

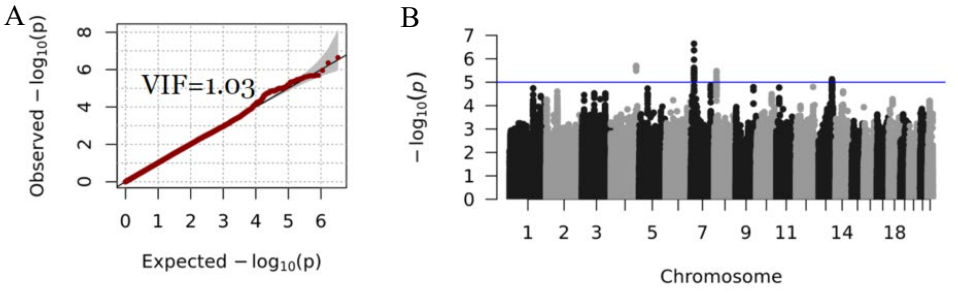


Figure 12. QQ Plot and MH Plot (Interaction Effects).

Note : Figure 12A is a QQ plot and Figure 12B is a Manhattan plot of the results of interaction effect on FEV_1/FVC.

The SNPs most associated with lung function decline by SNP-by-time interaction were found to be rs75679995, within the *DNAH11* gene, and rs9991425, located near the *AADAT* and *MFAP3L* gene, yielding p =2.28 x 10^{-7} and p = 2.01x 10^{-6}, respectively. **Figure 13** shows the regional plot for most significant SNPs region. Those SNPs are located within the TAD of *DNAH11* region (21.6~22.2 Mb of chromosome 7). DNA sequences within a TAD physically interact with each other more frequently than with sequences outside the TAD, and thus our most significant SNPs may affect the expression of *DNAH 11* (Figure 13A). Therefore, our results indicate that *DNAH11* may be functionally related to lung function decline.

Additionally, several identified loci were plausible candidates for lung function decline with FEV_1/FVC ratio, including *MFAP3L* and *AADAT* (Figure 13B). We assessed whether replicated SNPs were associated with gene expression by analyzing whether they were cis acting expression quantitative trait loci (eQTL) in lung tissue. Two SNPs were cis-eQTLs for MFAP3L in muscle and AADAT in cultured fibroblast. The exon-level expressions of *MFAP3L* and *AADAT* show 6.286 and 1 transcripts per kilobase million in lung tissues, respectively

Table 8. Top 10 significant SNPs (Interaction Effects)

CHR	dbSNP.RS.ID	BP	minor	MAF	HWE	MISS	info	SNP BETA (SE)	SNP P-value	Interaction Beta (SE)	Interaction P-value	LR
7	rs75679995	21880268	G/A	0.224	0.576	0.021	0.98	0.151(0.125)	0.228528	-0.149(0.029)	2.28E-07	1.18E-06
7	rs74676405	21878748	A/G	0.208	0.896	0.021	0.97	0.126(0.129)	0.328204	-0.149(0.03)	4.37E-07	1.89E-06
7	rs10274154	21881548	G/A	0.215	0.949	0.005	0.992	0.083(0.126)	0.511169	-0.14(0.029)	1.10E-06	3.44E-06
4	rs9991425	171175255	G/A	0.083	0.318	0.04	0.945	0.524(0.191)	0.006037	-0.206(0.043)	2.01E-06	6.92E-06
4	rs10026067	171175090	A/T	0.083	0.318	0.04	0.945	0.526(0.191)	0.005779	-0.206(0.043)	2.10E-06	7.06E-06
4	rs55764142	171176154	T/G	0.083	0.392	0.04	0.945	0.522(0.19)	0.006134	-0.205(0.043)	2.13E-06	7.31E-06
4	rs55999620	171176227	T/A	0.083	0.392	0.04	0.945	0.522(0.19)	0.006134	-0.205(0.043)	2.13E-06	7.31E-06
4	rs4692810	171177863	T/C	0.083	0.392	0.04	0.944	0.521(0.191)	0.006284	-0.205(0.043)	2.21E-06	7.63E-06
4	rs10014884	171175473	C/A	0.083	0.392	0.039	0.945	0.525(0.19)	0.005804	-0.204(0.043)	2.36E-06	7.82E-06
7	rs73279813	21889459	A/T	0.212	0.897	0.017	0.986	0.101(0.128)	0.430423	-0.138(0.029)	2.44E-06	9.12E-06

Abbreviations: BP, physical position; MAF, minor allele frequencies; HWE, Hardy-Weinberg equilibrium; SE, standard error

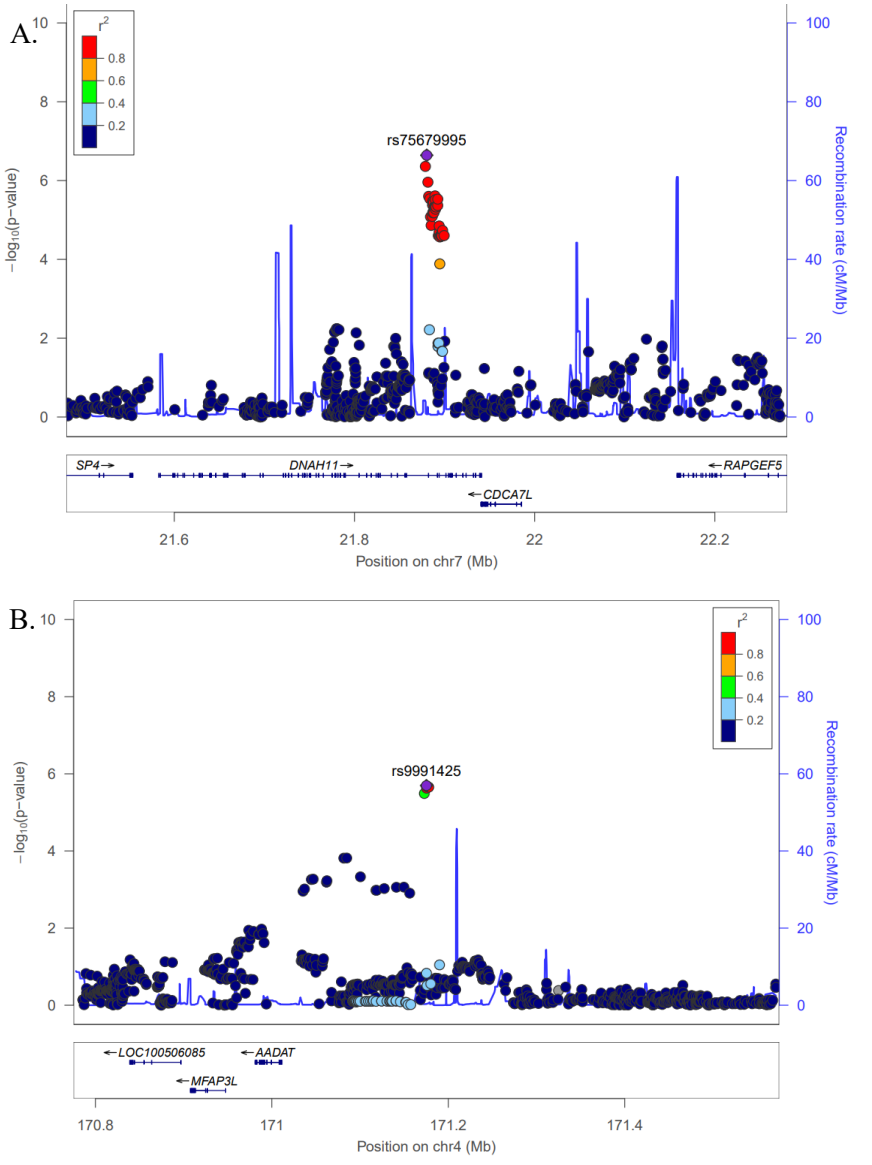


Figure 13. Regional Plot (A.CHR7, B.CHR4, Interaction Effects).

Note : Regional association plots at the most significant loci associated with lung function as measured by FEV_1/FVC by SNP-by-time interaction for chromosomes 7 (A), 4 (B).

Declines of lung function along age Changes in FEV_1/ FVC along age according to time and rs75679995 (*DNAH 11*) were plotted by generalized additive models**. Figure 14** suggests that rs75679995 has a significant effect on FEV_1/FVC ratios to cohort participatant for time to gene interaction rapid of lung function decline.

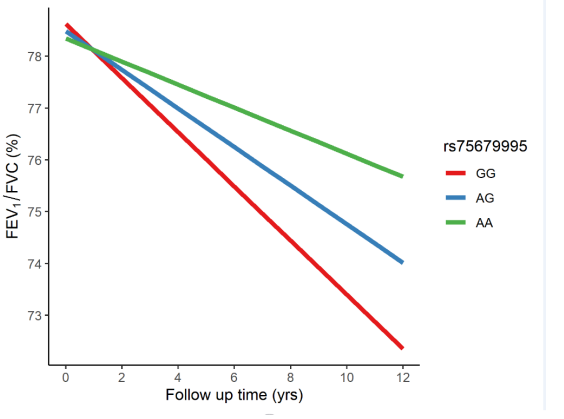


Figure 14. Estimated FEV_1/FVC According to Time Interction and rs 75679995.

10. Replication of most significant SNP in GENIE cohort from KARE

Table 9. Results for the most significant SNPs as identified by from discovery GWIS. For FEV_1/FVC, GWIS was performed on a GENIE cohort, and Table 9 summarizes the genome-wide significant results. SNP and INT are the coefficients for the main SNP and for the interaction effects between the SNP and time, respectively. Overall effects indicate P values (PLR) for testing the null hypotheses H0 : βSNP = βSNP−time = 0 by F test.

Table 9. Validation Analyses on the Most Significant SNPs identified by Discovery GWIS

Marker	Minor	MAF	HWE	MISS	SNP.BETA	SNP.p	INT.BETA	Int.p	LR
rs75679995	G/A	0.21	0.337	0.089	0.058(0.127)	0.647635	0.023(0.018)	0.210252	0.35024
rs74676405	A/G	0.194	0.433	0.085	0.117(0.13)	0.371256	0.019(0.019)	0.306255	0.313309
rs10274154	G/A	0.208	0.556	0.067	0.113(0.126)	0.372234	0.011(0.018)	0.556916	0.492053
rs9991425	G/A	0.074	0.534	0.074	0.243(0.195)	0.213421	-0.01(0.028)	0.731486	0.459079
rs10026067	A/T	0.074	0.592	0.075	0.268(0.195)	0.170278	-0.013(0.028)	0.646054	0.384121
rs55764142	T/G	0.074	0.534	0.074	0.244(0.195)	0.211523	-0.01(0.028)	0.728262	0.455996
rs55999620	T/A	0.074	0.534	0.074	0.244(0.195)	0.211523	-0.01(0.028)	0.728262	0.455996
rs4692810	T/C	0.074	0.535	0.074	0.249(0.195)	0.201726	-0.01(0.028)	0.72666	0.440702
rs10014884	C/A	0.074	0.534	0.074	0.243(0.195)	0.213421	-0.01(0.028)	0.731486	0.459079
rs73279813	A/T	0.209	0.718	0.038	0.13(0.124)	0.293917	0.01(0.018)	0.586104	0.425723

Note : βSNP and βINT are the coefficients for the main SNP and interaction effects between SNP and time, respectively.
Abbreviations: BP, physical position; MAF, minor allele frequencies; HWE, Hardy-Weinberg equilibrium; SE, standard error.

IV. Discussion

In the present study, the annual decline in lung function was estimated for a population-based cohort in Korea. Current smokers showed a significantly increased rate of decline in FEV_1 than healthy never smoker. The rate of decline in FEV_1 was similar between former smokers and never smokers among men without OLD and between male healthy current smokers and male current smokers with OLD. We also conducted GWISs for genetic variants associated with decreases in lung function (FEV_1, FEV_1/FVC) over time. The interaction was evaluated using a joint test near the previously described *FAM13, MFAP3L*, and *AADAT* loci on chromosome 4 and at the *DNAH11* locus on chromosome 7. Importantly, we found evidence for the time interaction between the *DNAH11, MFAP3L*, and *AADAT* genes and FEV_1/FVC; minor alleles of the selected SNPs near *DNAH11, MFAP3L*, and *AADAT* tended to have lower FEV_1/FVC values. In addition, the coefficient of gene-time interaction showed the same direction as the main SNP effect, and its amount was comparable. Finally, it was also demonstrated that the statistical model for the analysis of gene-time interactions should be carefully selected. This was because the effects of time on FEV_1/FVC were very strong, and the means and variances could differ according to genetic variants.

Studies of the decline in lung function in population-based cohorts have been previously reported[28-30]. Compared to previous study by *Leem et al.*[30]which using same community-based cohort 4 year follow up, the mean decline of FEV_1 was increased in healthy never smoker. We found annual mean FEV_1 declines of 41.7 mL for men and 33.4 mL for women, while *Leem at el.* reported respective decreases of 31.6 mL and 27.0 mL26. It was confirmed that the decline in lung function was increased as get older during the 14 years long term follow-up period. The results described in the present study are consistent with several earlier investigation with lung function decline of old age[31].

Previous study reveal consistent finding on the rate lung function decline is increased with higher age[31]. The average annual decline in FEV_1 in current smokers in our study was consistent with that in the Swiss Study on Air Pollution and Lung Diseases in Adults, where the annual mean decline in FEV1 was 43.8 mL for male current smokers and 34.7 mL for their female counterparts (43.7 mL for men and 32.0 mL for women in our study)[28]. Omori et al. reported similar declines in FEV_1 in the Japanese male population stratified by age and the smoking status[32].

FAM13A was reported to be associated with FEV_1/FVC ratio in the CHARGE consortium[11] and was strongly associated with COPD susceptibility in the previous GWAS.[12,33-35] In this study, *FAM13A* was strongly associated with FEV_1/FVC ratio. The associated SNPs are in 1.40×10^{-10}. Although this locus did not reach genome-wide significance nor was replicated in the GINIE cohort, this finding suggests that this previously reported region could be associated with lung function in the Korean population. *FAM13* have 2 splice variants of Variant-1 encodes a 683-amino acid protein with two coiled-coil domains and three nuclear localization signals. Variant-2 encodes a deduced 1023 amino acid protein that has an N-terminal extension containing a RhoGAP domain that is absent in Variant-1. The 5-kb Variant-1 transcript is ubiquitously expressed, while its expression is highest in the lung, skeletal muscle, thymus and brain. On the other hand, the 6-kb Variant-2 transcript is less abundant and is predominantly detected in the lung, thymus, kidney, pancreas, and liver. *FAM13A* has a putative role in signal transduction, and those SNPs that lie in an intronic region downstream of a Rho GTPase-activating protein(RhoGAP) domain may especially be involved in COPD susceptibility[36]. Although little is known about *FAM13A* function, gene expression analyses in cell lines from several tissues demonstrated a consistent increase in response to hypoxia. Differences in respiratory epithelial cell expression of *FAM13A* have been also noted during differentiation into pulmonary type II cells in vitro and in mild versus severe

cystic fibrosis patients. [37,38]

The second most significant SNP in our cohort was the rs75679995 SNP found at the intron of *DNAH11,* which is one of the MAP kinases. The biological importance of dyneins (microtubule motor protein complexes) in *DNAH11* is well understood. They are required for the initiation of MAPK3/6 and p38, which in turn regulates a large number of biological process such as cell survival, differentiation, migration[39,40], and immune and inflammatory responses[41,42], and thus has implications for the progression of different types of cancers. Meanwhile, primary ciliary dyskinesia (PCD) is a rare autosomal recessive genetic disorder characterized by dysfunction of motile cilia.

DNAH11 mutations are reported in 6% of all PCDs, and 22% of those with normal ultrastructure.[43-45] Generally, ciliary dysmotility causes poor mucociliary clearance and leads to impairment of pulmonary function and severe respiratory infections. The mechanism of lung function decline that was observed in the cohort could be attributed to ciliary dysfunction.

The eQTL result of rs9991425 revealed a higher expression of *MFAP3L* and *AADAT* genes. We further analyzed *MFAP3L* and *AADAT* expression using GTEx V8 data, and this revealed that *MFAP3L* was upregulated in muscle while *AADAT* was expressed in cultured fibroblast. The numbers of transcripts per million kilobases for *MFAP3L* and *AADAT* in lung tissue were 1 and 6.286, respectively. Even though this result implies that it is unclear which gene, *MFAP3L* or *AADAT*, contributes to the significant effect of rs9991425 on FEV_1/FVC ratio, *AADAT* might be a more promising candidate gene due to its higher level of expression in lung tissue compared to that of *MFAP3L.* This gene has been described as an important player in thyroid hormone regulation[46] and lung cancer[47], and as a major negative regulator of amino acid metabolism involving mitochondria in cultured fibroblast, suggesting that *AADAT* could play a role in the pathogenesis of COPD or in lung function decline.

There is increasing evidence that mitochondria serve cellular functions beyond

oxygen sensing and energy production. Mitochondria can sense upstream processes such as inflammation, infection, tobacco smoke, and environmental insults important in these diseases and in turn can respond to such stimuli through altered mitochondrial protein expression and structure (and resultant dysfunction). On the other hand, mitochondrial dysfunction has downstream influences on cytosolic and mitochondrial calcium regulation, airway contractility, gene and protein housekeeping, responses to oxidative stress, proliferation, apoptosis, fibrosis, and certainly metabolism, which are all key aspects of airway disease pathophysiology.[48] To the best of our knowledge, our study is the first to demonstrate a potential novel COPD susceptibility locus in the *AADAT* gene on chromosome 4.

It has been reported that more than 70% of patients with COPD have GOLD stage 1 (mild) or 2 (moderate) with no apparent respiratory symptoms such as dyspnea on exertion.[49] Previous studies have shown that tiotropium could ameliorate the annual decline in the FEV_1 and lower the frequency of acute exacerbation when compared with placebo-treated patients with GOLD stage 1 or 2 COPD.[50,51] Therefore, earlier detection of COPD or lung function decline could lead to early effective intervention, reducing disease progression and socioeconomic burden. The clinical implication of our study is that we provided meaningful insights into this genetic risk. We proposed a method that illustrates the complexity of gene-time interaction analyses, identified consistent gene-time interactions, and proposed a statistical model association. Such early detection has been an interest in recent years in the United States, where COPD monitoring is performed through the National Lung Health Education Program. However, given that there are individuals who are expected to suffer a rapid decline of lung function, it may be cost effective to perform selective PFT only in patients with genetic susceptibility to COPD, rather than in all patients. The results of our study are especially noteworthy given that there has been no study investigating the genetic risk factors for early COPD or lung function decline despite substantial

progress in evaluating the genetic susceptibility to COPD.

Although we were able to find susceptibility foci related to decreases in lung function, unfortunately, these data were not identically replicated in the GINIE study. For example, an interaction between time and SNPs was found for the KARE data, but for the GINIE data, the interaction between pack years and SNPs was not significant. However, this seems to be attributed to the fact that in our replication study the GENIE study the characteristics of the patients were quite different with that of KARE data. Of note, the KARE data were based on rural and urban community populations, while the GENIE data comprised participants who underwent regular health screening and received routine medical care. Medical care and routine health check-ups are often positively related to socioeconomic status, which could have resulted in selection bias and influenced the results of our study.[52]

Our study provides novel findings, but it has several limitations that should be noted. First, the definition of COPD is based on pre-bronchodilator data. Indeed, it is possible that asthma patients might have been included among those defined as having COPD. Second, we did not perform functional investigation of whether these genetic variants affect early COPD via an *in vitro* analysis. Further studies are needed to validate our findings.

V. CONCLUSION

In conclusion, our study suggests that several SNPs located on chromosome 4 or *FAM13* gene are possibly associated with early COPD. We also identified that the *DNAH11 and AADAT* genes might play a role in susceptibility to lung function decline or COPD. Further studies are needed to validate our findings.

Reference

1. Sorlie PD, Kannel WB, O'Connor G. Mortality associated with respiratory function and symptoms in advanced age. The Framingham Study. Am Rev Respir Dis 1989;140:379-84.
2. Knuiman MW, James AL, Divitini ML, Ryan G, Bartholomew HC, Musk AW. Lung function, respiratory symptoms, and mortality: results from the Busselton Health Study. Ann Epidemiol 1999;9:297-306.
3. Postma DS, Bush A, van den Berge M. Risk factors and early origins of chronic obstructive pulmonary disease. Lancet 2015;385:899-909.
4. Pauwels RA, Rabe KF. Burden and clinical features of chronic obstructive pulmonary disease (COPD). Lancet 2004;364:613-20.
5. Viegi G, Pistelli F, Sherrill DL, Maio S, Baldacci S, Carrozzi L. Definition, epidemiology and natural history of COPD. Eur Respir J 2007;30:993-1013.
6. Singh D, Agusti A, Anzueto A, Barnes PJ, Bourbeau J, Celli BR, et al. Global Strategy for the Diagnosis, Management, and Prevention of Chronic Obstructive Lung Disease: the GOLD science committee report 2019. Eur Respir J 2019;53.
7. Vogelmeier CF, Criner GJ, Martinez FJ, Anzueto A, Barnes PJ, Bourbeau J, et al. Global Strategy for the Diagnosis, Management, and Prevention of Chronic Obstructive Lung Disease 2017 Report. GOLD Executive Summary. Am J Respir Crit Care Med 2017;195:557-82.
8. Schünemann HJ, Dorn J, Grant BJ, Winkelstein W, Jr., Trevisan M. Pulmonary function is a long-term predictor of mortality in the general population: 29-year follow-up of the Buffalo Health Study. Chest 2000;118:656-64.
9. Hobbs BD, de Jong K, Lamontagne M, Bossé Y, Shrine N, Artigas MS, et al. Genetic loci associated with chronic obstructive pulmonary disease overlap with loci for lung function and pulmonary fibrosis. Nat Genet

2017;49:426-32.

10. Jackson VE, Latourelle JC, Wain LV, Smith AV, Grove ML, Bartz TM, et al. Meta-analysis of exome array data identifies six novel genetic loci for lung function. Wellcome Open Res 2018;3:4.
11. Hancock DB, Eijgelsheim M, Wilk JB, Gharib SA, Loehr LR, Marciante KD, et al. Meta-analyses of genome-wide association studies identify multiple loci associated with pulmonary function. Nat Genet 2010;42:45-52.
12. Cho MH, Boutaoui N, Klanderman BJ, Sylvia JS, Ziniti JP, Hersh CP, et al. Variants in FAM13A are associated with chronic obstructive pulmonary disease. Nat Genet 2010;42:200-2.
13. Pillai SG, Kong X, Edwards LD, Cho MH, Anderson WH, Coxson HO, et al. Loci identified by genome-wide association studies influence different disease-related phenotypes in chronic obstructive pulmonary disease. Am J Respir Crit Care Med 2010;182:1498-505.
14. Wilk JB, Chen TH, Gottlieb DJ, Walter RE, Nagle MW, Brandler BJ, et al. A genome-wide association study of pulmonary function measures in the Framingham Heart Study. PLoS Genet 2009;5:e1000429.
15. Repapi E, Sayers I, Wain LV, Burton PR, Johnson T, Obeidat M, et al. Genome-wide association study identifies five loci associated with lung function. Nat Genet 2010;42:36-44.
16. Cho MH, Castaldi PJ, Wan ES, Siedlinski M, Hersh CP, Demeo DL, et al. A genome-wide association study of COPD identifies a susceptibility locus on chromosome 19q13. Hum Mol Genet 2012;21:947-57.
17. Soler Artigas M, Wain LV, Repapi E, Obeidat M, Sayers I, Burton PR, et al. Effect of five genetic variants associated with lung function on the risk of chronic obstructive lung disease, and their joint effects on lung function. Am J Respir Crit Care Med 2011;184:786-95.
18. Soler Artigas M, Loth DW, Wain LV, Gharib SA, Obeidat M, Tang W,

et al. Genome-wide association and large-scale follow up identifies 16 new loci influencing lung function. Nat Genet 2011;43:1082-90.

19. Kim WJ, Lee MK, Shin C, Cho NH, Lee SD, Oh YM, et al. Genome-wide association studies identify locus on 6p21 influencing lung function in the Korean population. Respirology 2014;19:360-8.
20. Lieberman J, Winter B, Sastre A. Alpha 1-antitrypsin Pi-types in 965 COPD patients. Chest 1986;89:370-3.
21. Kim CY, Kim BK, Kim YJ, Lee SH, Kim YS, Kim JH. Longitudinal Evaluation of the Relationship Between Low Socioeconomic Status and Incidence of Chronic Obstructive Pulmonary Disease: Korean Genome and Epidemiology Study (KoGES). Int J Chron Obstruct Pulmon Dis 2020;15:3447-54.
22. Shin C, Kim J, Kim J, Lee S, Shim J, In K, et al. Association of habitual snoring with glucose and insulin metabolism in nonobese Korean adult men. Am J Respir Crit Care Med 2005;171:287-91.
23. Baik I, Kim J, Abbott RD, Joo S, Jung K, Lee S, et al. Association of snoring with chronic bronchitis. Arch Intern Med 2008;168:167-73.
24. Graham BL, Steenbruggen I, Miller MR, Barjaktarevic IZ, Cooper BG, Hall GL, et al. Standardization of Spirometry 2019 Update. An Official American Thoracic Society and European Respiratory Society Technical Statement. Am J Respir Crit Care Med 2019;200:e70-e88.
25. Cho YS, Go MJ, Kim YJ, Heo JY, Oh JH, Ban HJ, 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.
26. Lee C, Choe EK, Choi JM, Hwang Y, Lee Y, Park B, et al. Health and Prevention Enhancement (H-PEACE): a retrospective, population-based cohort study conducted at the Seoul National University Hospital Gangnam Center, Korea. BMJ Open 2018;8:e019327.
27. Bhattacharya S, Srisuma S, Demeo DL, Shapiro SD, Bueno R, Silverman

EK, et al. Molecular biomarkers for quantitative and discrete COPD phenotypes. Am J Respir Cell Mol Biol 2009;40:359-67.

28. Downs SH, Brändli O, Zellweger JP, Schindler C, Künzli N, Gerbase MW, et al. Accelerated decline in lung function in smoking women with airway obstruction: SAPALDIA 2 cohort study. Respir Res 2005;6:45.
29. Chinn S, Jarvis D, Melotti R, Luczynska C, Ackermann-Liebrich U, Antó JM, et al. Smoking cessation, lung function, and weight gain: a follow-up study. Lancet 2005;365:1629-35; discussion 00-1.
30. Leem AY, Park B, Kim YS, Chang J, Won S, Jung JY. Longitudinal decline in lung function: a community-based cohort study in Korea. Sci Rep 2019;9:13614.
31. Luoto J, Pihlsgård M, Wollmer P, Elmståhl S. Relative and absolute lung function change in a general population aged 60-102 years. Eur Respir J 2019;53.
32. Omori H, Nonami Y, Morimoto Y. Effect of smoking on FEV decline in a cross-sectional and longitudinal study of a large cohort of Japanese males. Respirology 2005;10:464-9.
33. Siedlinski M, Tingley D, Lipman PJ, Cho MH, Litonjua AA, Sparrow D, et al. Dissecting direct and indirect genetic effects on chronic obstructive pulmonary disease (COPD) susceptibility. Hum Genet 2013;132:431-41.
34. Ziółkowska-Suchanek I, Mosor M, Gabryel P, Grabicki M, Żurawek M, Fichna M, et al. Susceptibility loci in lung cancer and COPD: association of IREB2 and FAM13A with pulmonary diseases. Sci Rep 2015;5:13502.
35. Xie J, Wu H, Xu Y, Wu X, Liu X, Shang J, et al. Gene susceptibility identification in a longitudinal study confirms new loci in the development of chronic obstructive pulmonary disease and influences lung function decline. Respir Res 2015;16:49.
36. Cohen M, Reichenstein M, Everts-van der Wind A, Heon-Lee J, Shani M, Lewin HA, et al. Cloning and characterization of FAM13A1--a gene

near a milk protein QTL on BTA6: evidence for population-wide linkage disequilibrium in Israeli Holsteins. Genomics 2004;84:374-83.

37. Chi JT, Wang Z, Nuyten DS, Rodriguez EH, Schaner ME, Salim A, et al. Gene expression programs in response to hypoxia: cell type specificity and prognostic significance in human cancers. PLoS Med 2006;3:e47.
38. Wang B, Liang B, Yang J, Xiao J, Ma C, Xu S, et al. Association of FAM13A polymorphisms with COPD and COPD-related phenotypes in Han Chinese. Clin Biochem 2013;46:1683-8.
39. Cuadrado A, Nebreda AR. Mechanisms and functions of p38 MAPK signalling. Biochem J 2010;429:403-17.
40. Cuenda A, Rousseau S. p38 MAP-kinases pathway regulation, function and role in human diseases. Biochim Biophys Acta 2007;1773:1358-75.
41. Wang CY, Chang K, Petralia RS, Wang YX, Seabold GK, Wenthold RJ. A novel family of adhesion-like molecules that interacts with the NMDA receptor. J Neurosci 2006;26:2174-83.
42. Lee JC, Laydon JT, McDonnell PC, Gallagher TF, Kumar S, Green D, et al. A protein kinase involved in the regulation of inflammatory cytokine biosynthesis. Nature 1994;372:739-46.
43. Knowles MR, Daniels LA, Davis SD, Zariwala MA, Leigh MW. Primary ciliary dyskinesia. Recent advances in diagnostics, genetics, and characterization of clinical disease. Am J Respir Crit Care Med 2013;188:913-22.
44. Leigh MW, Pittman JE, Carson JL, Ferkol TW, Dell SD, Davis SD, et al. Clinical and genetic aspects of primary ciliary dyskinesia/Kartagener syndrome. Genet Med 2009;11:473-87.
45. Pifferi M, Michelucci A, Conidi ME, Cangiotti AM, Simi P, Macchia P, et al. New DNAH11 mutations in primary ciliary dyskinesia with normal axonemal ultrastructure. Eur Respir J 2010;35:1413-6.
46. Teumer A, Chaker L, Groeneweg S, Li Y, Di Munno C, Barbieri C, et al.

Genome-wide analyses identify a role for SLC17A4 and AADAT in thyroid hormone regulation. Nat Commun 2018;9:4455.

47. Hsu CC, Yang AY, Chen JY, Tsai HH, Lin SH, Tai PC, et al. Lysine Deprivation Induces AKT-AADAT Signaling and Overcomes EGFR-TKIs Resistance in EGFR-Mutant Non-Small Cell Lung Cancer Cells. Cancers (Basel) 2021;13.

48. Prakash YS, Pabelick CM, Sieck GC. Mitochondrial Dysfunction in Airway Disease. Chest 2017;152:618-26.

49. Zhong N, Wang C, Yao W, Chen P, Kang J, Huang S, et al. Prevalence of chronic obstructive pulmonary disease in China: a large, population-based survey. Am J Respir Crit Care Med 2007;176:753-60.

50. Zhou Y, Zhong NS, Li X, Chen S, Zheng J, Zhao D, et al. Tiotropium in Early-Stage Chronic Obstructive Pulmonary Disease. N Engl J Med 2017;377:923-35.

51. van der Molen T, Kirenga BJ. COPD: early diagnosis and treatment to slow disease progression. Int J Clin Pract 2015;69:513-4.

52. Mackenbach JP, Stronks K, Kunst AE. The contribution of medical care to inequalities in health: differences between socio-economic groups in decline of mortality from conditions amenable to medical intervention. Soc Sci Med 1989;29:369-76.

ABSTRACT (IN KOREAN)

한국인 지역사회코호트에서 폐기능 저하에 영향을 미치는 요인에 대한 분석

< 지도교수 김 영 삼 >

연세대학교 대학원 의학과

김 치 영

배경 및 목적: 폐기능은 만성폐쇄성폐질환과 같은 호흡기병으로의 이환율과 사망률을 예측할 수 있는 가장 중요한 호흡기계의 특성이다. 이전의 몇 가지 연구들에서 전장 유전체 분석을 통해 중등도 이상의 만성폐쇄성폐질환과 관련된 유전 변이들이 알려져 있다. 그러나 일반 인구 집단에서의 폐기능 저하와 관련된 유전적 감수성에 대해서는 널리 알려지지 않았다.

목적: 본 연구에서는 지역 사회 기반 코호트를 통해 일반인구에서의 폐기능 저하와 관련된 요인을 확인하고자 하였다.

방법: 우리는 폐기능 저하와 관련된 요인들을 분석하기 위해 지역사회 기반 코호트를 이용하여 임상적 표현형 및 특성과 함께 FEV_1 및 FEV_1/FVC등의 폐기능을 이용하여 관련된 유전자 변이와 (유전자 다형성과 일배체형) 함께 선형혼합모형을 이용하여 전장 유전체 연관성 분석을 통하여 확인하였다.

결과: 총 8845명의 참가자들이 분석에 포함되었다. 매년 평균 폐기능저하는 남성에서는 41.7 mL 그리고 여성에서는 33.4 mL 로

확인되었으며, 흡연자에서 폐기능저하의 속도가 가장 빠른 것이 확인되었다. 기류 제한의 여부에 따라 환자를 분석하여 보았을 때 나이, 성별, 체질량 지수, 그리고 흡연력 등이 만성 폐쇄성 폐질환과 연관이 있었다. 우리는 전장 유전체 연관성 분석을 통해 이 임상적인 결과들과 함께 분석을 진행하였다. 이전에 밝혀져 있던 *FAM13A* 유전자와 연관된 유전 변이 지역을 확인하였으며, 이 부위는 염색체 4번에 존재하는 유전자로 비슷한 대립유전자형 빈도를 보였다. 게다가 우리는 FEV_1/FVC값의 저하가 특정 유전 돌연변이와 연관되어 시간이 지날수록 빨리 감소한다는 것을 확인할 수 있었다. 폐기능 저하와 가장 연관된 단일 염기 다형성 부위는 염색체 7번에 있는 rs75679995와 그 단일 염기 다형성부위가 위치한 TAD 인 *DNAH11* 부위로 확인되었다. rs9991425 이*MFAP3L*와 *AADAT* 유전자 발현에 관여하는 것이 함께 확인 되었으며, 폐기능 저하 특히 FEV_1/FVC값과 연관성이 있음이 확인되었다.

결론: 일반인구에서의 폐기능저하의 속도는 흡연과 관련이 있으며, 반복 측정한 폐기능과 선형혼합모형을 이용하여 전장 유전체 연관분석을 시행하였을 때 염색체 4번에 있는 특정 단일 염기 다형성부위들과 관련된유전자 *FAM13A*와 염색체 7번에 특정 단일 염기 다형성부위들에 위치한 *DNAH11*, *AADAT* 유전자가 폐기능저하와 연관성이 있는 것이 확인 할 수 있었다.

핵심 되는 말: 만성폐쇄성폐질환, 폐기능 저하, 전장유전체연관성분석, 단일 염기 다형성부위, 기류조절장애, *FAM13A*, *DNAH11*, *AADAT* 유전자