scientific reports

Check for updates

OPEN

Genome wide interaction study of genetic variants associated with lung function decline

Chi Young Kim[1], Boram Park[2], Ji Ye Jung[1], Je Hyeong Kim[3], Chung Mo Nam[4], Jaehoon An[2,6], Sungho Won[2,5,6]✉ & Young Sam Kim[1]✉

Some genetic variants are associated with lung function decline and chronic obstructive pulmonary disease (COPD), but functional studies are necessary to confirm causality. We investigated the genetic susceptibility-associated lung function decline with or without COPD, using data from a community-based cohort (N = 8554). A genome-wide interaction study was conducted to identify the association between genetic variants and pulmonary function, and the way variants relate to lung impairment in accordance with smoking status and amount was examined. We further used a linear mixed model to examine the association and interaction to time effect. 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 the fastest for current smokers. We also found a previously identified locus near *FAM13A*, the most significant SNPs from the results of two likelihood ratio tests for FEV_1/FVC ($P = 1.56 \times 10^{-10}$). These selected SNPs were located in the upstream region of *FAM13A* 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 ($P = 2.28 \times 10^{-7}$ and 2.01×10^{-6}, respectively). This is the first study to investigate gene–time interactions in 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 genomic regions (DNAH11, AADAT) that are potentially involved in gene–environment interactions, warranting further investigation to confirm their roles.

Keywords Chronic obstructive pulmonary disease, Genome-wide association study, Single nucleotide polymorphism, Airflow obstruction, *FAM13A* gene, *DNAH11* gene, *AADAT* gene

Lung function is critical for the proper functioning of the respiratory system. Impaired lung function, usually represented by low measures of the forced expiratory volume (FEV_1) in one second, forced vital capacity (FVC), and FEV_1/FVC, is a risk factor for morbidity and mortality in the general population[1,2]. FEV_1/FVC is used as a diagnostic criterion for chronic obstructive pulmonary disease (COPD), and the FEV_1 is used to evaluate COPD severity[3,4].

Lung function can gradually decrease because of physiological aging of the lungs, and numerous genetic and environmental factors can accelerate this process[5]. Multiple risk factors for COPD have been identified; among them, smoking has been recognised as the most significant risk factor leading to rapid decline in lung function and consequent development of COPD[6]. However, sensitivity to smoking differs substantially among individuals, and these differences are partially attributed to genes and/or their interactions with smoking. Genetic loci associated with lung function have been shown to influence susceptibility to respiratory diseases, including COPD. However, most of the variants identified using genome-wide association studies (GWAS) are common variants (minor allele frequency [MAF] > 5%) in the population[7].

[1]Department of Internal Medicine, Yonsei University College of Medicine, Seoul, Republic of Korea. [2]Department of Public Health Sciences, School of Public Health, Seoul National University, Kwanak-Ro Kwanak-Gu, Seoul 151-742, South Korea. [3]Division of Pulmonology, Department of Internal Medicine, Korea University Ansan Hospital, Korea University College of Medicine, Ansan, Republic of Korea. [4]Department of Preventive Medicine and Public Health, Yonsei University College of Medicine, Seoul, Korea. [5]Institute of Health and Environment, Seoul National University, Seoul, South Korea. [6]RexSoft Corps, Seoul, South Korea. ✉email: won1@snu.ac.kr; ysamkim@yuhs.ac

Recently, multiple loci associated with pulmonary function and risk of COPD and lung function decline have been identified using GWAS[8–12]. A previous study revealed that the height-related gene *DLEU7* is linked to a decrease in FEV_1 among non-asthmatic European adults[13]. Tang et al. identified novel loci associated with the rate of FEV[1] change over time, specifically *IL16/STARD5/TMC3* on chromosome 15 and *ME3* on chromosome 11, using data from 14 population-based cohort studies[14]. Other studies utilizing data from the Framingham Heart Study have investigated lung function decline. Gottlieb et al.[15] observed a modest contribution of heritable factors to the variance in FEV_1, FVC, and the FEV_1/FVC ratio decline. Wilk et al.[16] identified *GSTO2* and *IL6R* as credible candidate genes associated with pulmonary function. They also identified several novel gene regions associated with pulmonary function; however, replication in other study samples and functional studies are needed to confirm causality.

Therefore, studies on genetic susceptibility loci and their temporal interactions with environmental factors are required to improve general understanding regarding the pathogenesis and progression of COPD. In this study, we investigated the genetic factors that might contribute to pulmonary function and examined how they are related to lung impairment and gene-time interactions in a community-based cohort.

Results

Baseline patient characteristics

In total, 8554 participants (57.1% men, 52.9% women) were included in the analysis and Table 1 shows the baseline characteristics of all participants. The mean participant age was 52.1 years. Smoking history was obtained using a questionnaire, and smoking status and pack years were used as covariates for association analyses. Smoking status was categorised into never smokers, former smokers, and current smokers, and the percentages were 60%, 15.4%, and 24.6%, respectively. Among the 8554 baseline subjects, COPD was observed in 738 subjects (8.6%) (Supplementary Table S2). During the 12-year follow-up period, 10.8% (n = 844) of patients developed COPD.

Effect of smoking on lung function decline

The generalized additive mixed model demonstrates a significant difference in the decline of FEV1 with age between males and females. Males exhibit a faster annual rate of decline (41.73 mL; 95% CI 41.15–42.31) compared to females (33.36 mL; 95% CI 33.12–33.59; P < 0.001), independently of age, smoking pack years, and smoking status (Fig. 1a). The annual mean decline in FEV_1 among male healthy never smokers, former smokers, and current smokers was also compared (Fig. 1b). 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).

Genome-wide interaction study (GWIS) of FEV_1/FVC with Korea Association Research Project (KARE) data

The FEV1/FVC ratios were applied as spirometric measures, which were used to identify the genetic variants interacting with time. GWIS were conducted conducted for associations of 3,333,374 SNPs by applying the two DF test to KARE data based on the likelihood ratio test to assess whether both the main and interaction effects are equal to 0 simultaneously. Figure 2 shows the genome-wide significance levels were determined by using the quantile–quantile (QQ) and Manhattan (MH) plots of the results from the two likelihood ratio tests for SNP and SNP*Time effect.

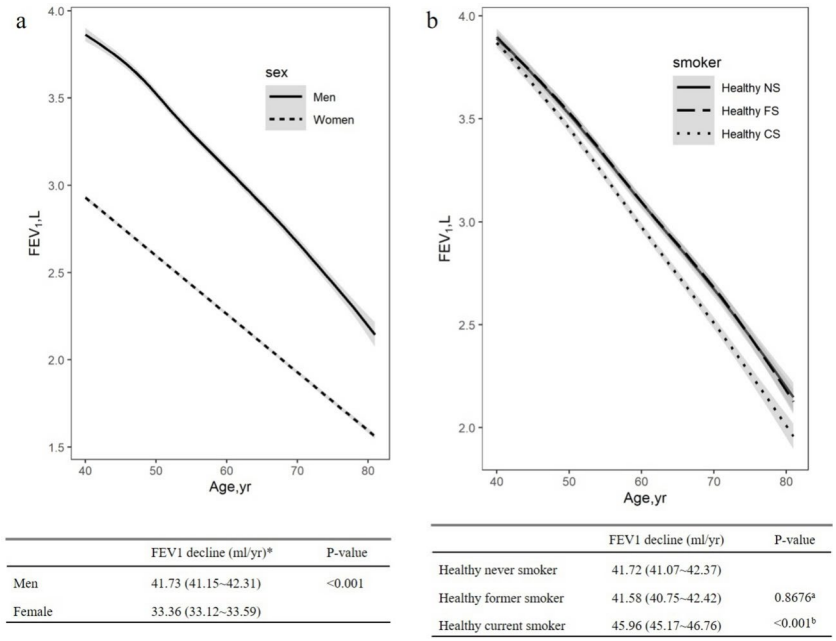
The MH plot showed that 89 SNPs associated with FEV_1/FVC at the genome-wide significance level were located on chromosome 4 in *FAM13A*. These selected SNPs were located in the upstream region of *FAM13A* on chromosome 4 and had similar MAFs. The SNP most associated with FEV_1/FVC was the rs2446304 SNP ($P = 1.56 \times 10^{-10}$). Figure 3 shows a regional plot (r^2) around rs2446304 SNPs created using LocusZoom. *FAM13A* within the topologically associated domain (TAD) is located between 89.6 to 90.2 Mb.

Effects of time and gene interactions on FEV_1/FVC

Figure 4a shows the QQ plot for the interaction effect tests, in which variance inflation factor (VIF) was 1.03, while Fig. 4b shows the MH plot for the same. To assess the significance of the identified interactions, statistical

	Males (n = 4026)	Females (n = 4528)	Total (n = 8554)	*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

Table 1. Baseline characteristics of male and female subjects included in the study. Data are presented as mean ± standard deviation or n (%). *FEV_1* forced expiratory volume in 1 s, *FVC* forced vital capacity.



	FEV1 decline (ml/yr)*	P-value
Men	41.73 (41.15~42.31)	<0.001
Female	33.36 (33.12~33.59)	

	FEV1 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[a]
Healthy current smoker	45.96 (45.17~46.76)	<0.001[b]

Fig. 1. Comparison of the annual mean decline of FEV_1. (**a**) FEV_1 for male (n = 4026) and female patients (n = 4528). (**b**) Comparison of the annual mean decline in FEV_1 among male healthy never smokers, former smokers, and current smokers. [a]*P*-value compared to healthy never smoker, [b]$P < 0.001$ compared to former smoker. *Data are presented as mean and 95% confidence intervals. *FEV_1* forced expiratory volume in 1 s.

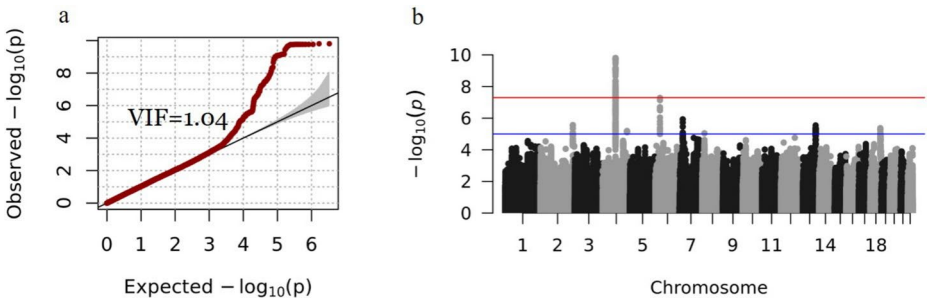


Fig. 2. QQ plot and MH plot from GWIS in KARE data. (**a**) Comparison between observed and expected *P*-values of quantiles under uniform distribution (null hypothesis). The VIF was 1.04. (**b**) Plot of the logarithms of the *P*-values of 3,333,374 SNPs against the physical chromosomal position. The red line represents genome-wide significance level (5×10^{-8}), which was met by several SNPs located at chromosome 4. *QQ* Quantile–quantile, *MH* Manhattan, *VIF* variance inflation factor, *GWIS* genome-wide interaction study, *KARE* Korea Association Resource Project.

significance thresholds that are routinely used for GWAS were applied. A genome-wide significance threshold of 5×10^{-8} was specified, while a more liberal genome-wide suggestive threshold of 1×10^{-5} was also applied to include interactions with P-values that were slightly above the genome-wide significance threshold but may represent genuine interactions.

The SNPs most associated with lung function decline in SNP-by-time interaction were found to be rs75679995 within *DNAH11*, and rs9991425 located near *AADAT* and *MFAP3L*, with $P = 2.28 \times 10^{-7}$ and 2.01×10^{-6}, respectively. Figure 5 shows the linkage disequilibrium and the regional plot for the most significant SNP region. The SNPs were located within the TAD of the *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* (Fig. 6). Therefore, our results indicated that *DNAH11* may be functionally related to decrease in lung function.

Several identified loci, including *MFAP3L* and *AADAT,* were plausible candidates for lung function decline characterised by changes in the FEV_1/FVC ratio (Fig. 7). We evaluated whether the identified SNPs were associated with gene expression using the GTEx Portal platform (https://gtexportal.org/). Two SNPs were identified as cis-expression quantitative trait loci (eQTLs) in lung tissue: one for MFAP3L in muscle tissue and

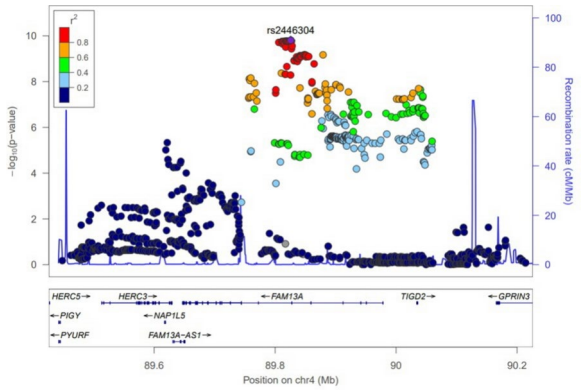


Fig. 3. Regional plot for chromosome 4 SNPs considering SNP × time interaction. Regional association plots at the most significant loci on chromosome 4 show associations with lung function, as measured by FEV1/FVC and SNPs for SNP Analysis Incorporating SNP × Time Interaction.

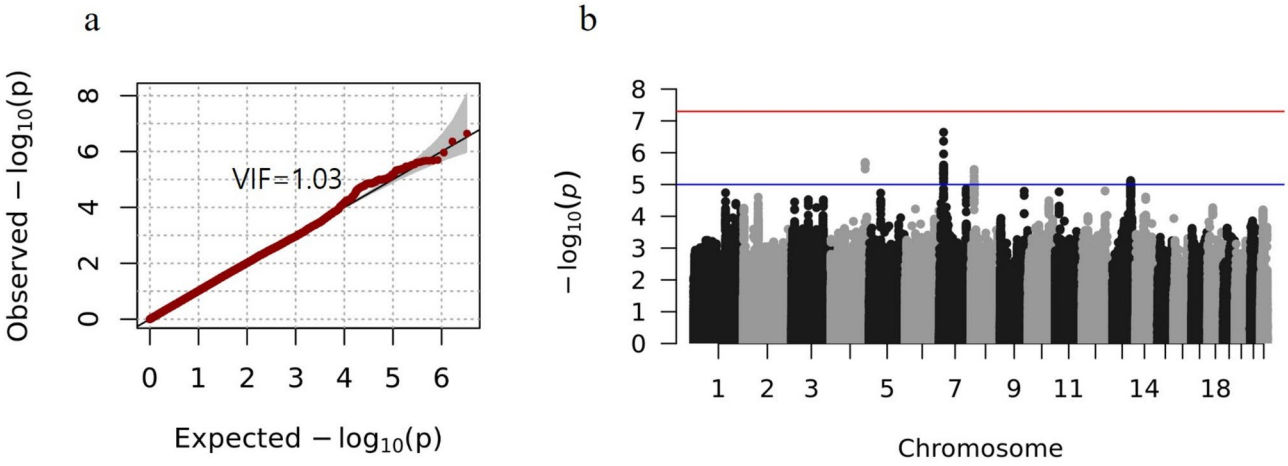


Fig. 4. QQ plot and MH plot (interaction effects) (**a**) QQ plot. (**b**) MH plot of the results of interaction effect on FEV_1/FVC. The blue line indicating the genome-wide suggestive level (5×10^{-5}). *QQ* Quantile–quantile, *MH* Manhattan, *VIF* variance inflation factor.

another for AADAT in cultured fibroblasts. The exon-level expressions of *MFAP3L* and *AADAT* showed 6.286 and 1 transcript per million kilobases in lung tissues, respectively (Supplementary Fig. S3).

Changes in FEV_1/FVC with time and rs75679995 (*DNAH 11*) status were plotted using generalised additive models. The estimated FEV1/FVC ratio was generated using gene-by-time analysis and KARE data (Fig. 7). The result suggested that rs75679995 significantly affected the FEV_1/FVC ratios of the cohort participants for time-to-gene interactions, resulting in rapid decline in lung function; the minor allele G negatively impacted the increment in time interaction.

Replication of the most significant SNPs in the GENIE cohort of KARE

Results for the most significant SNPs were identified in GWIS. For FEV_1/FVC, GWIS was performed on a GENIE cohort, and the significant results are summarised in Table S3. SNP and INT are the coefficients for the main SNP and interaction effects between SNP and time, respectively. Overall effects are indicated by the *P*-values for testing the null hypotheses, H0: βSNP = βSNP – time = 0, obtained using the F test. Although we identified susceptibility loci associated with decreased lung function in KARE data, these findings were unfortunately not replicated in the GENIE study.

Discussion

In the present study, the annual decline in lung function was estimated for a population-based cohort in Korea. Current smokers showed significantly higher rate of decline in FEV_1 than healthy never smokers. We also performed GWIS for genetic variants associated with COPD susceptibility over time. The interaction was evaluated using a joint test near the previously described *FAM13A, MFAP3L*, and *AADAT* loci on chromosome 4 and at the *DNAH11* locus on chromosome 7. Importantly, we found evidence for temporal interaction between *DNAH11*, *MFAP3L*, and *AADAT* and FEV_1/FVC; minor alleles of the selected SNPs near *DNAH11*, *MFAP3L*,

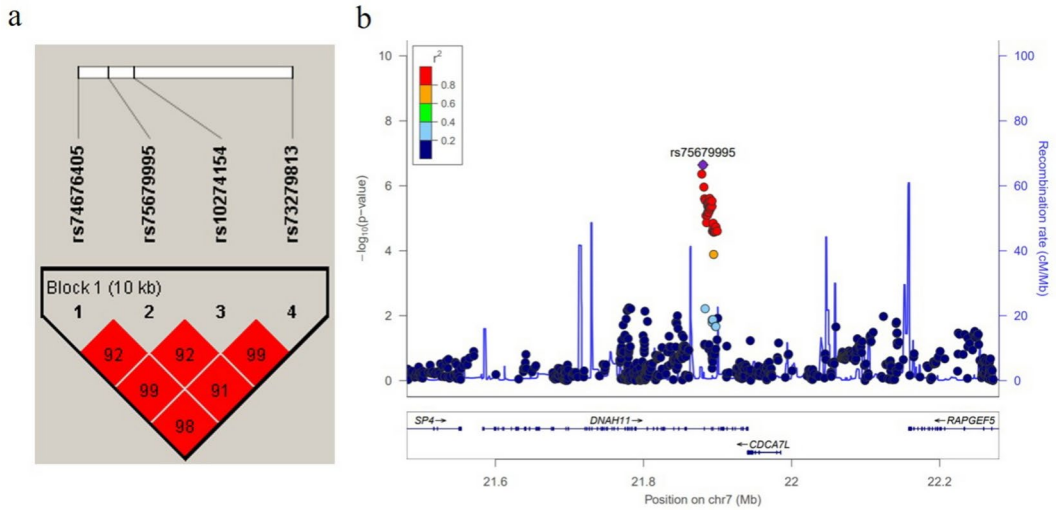


Fig. 5. LD among the four most significant SNPs and regional plots (chromosome 7, interaction effects). LD plot and regional association plots at the most significant loci showing association with lung function as measured by FEV_1/FVC and SNP-by-time interaction for chromosome 7. *LD* linkage disequilibrium, *SNP* single nucleotide polymorphism, FEV_1 forced expiratory volume in 1 s, *FVC* functional vital capacity.

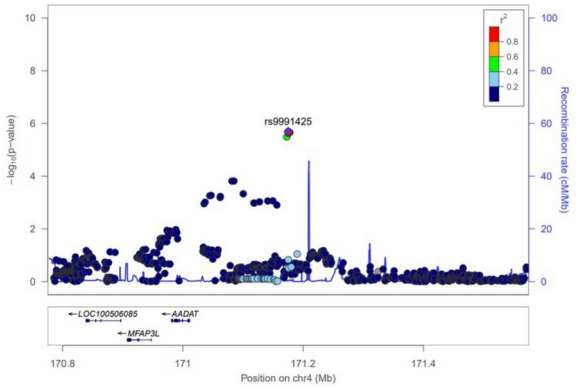


Fig. 6. Regional plot (chromosome 4, interaction effects).

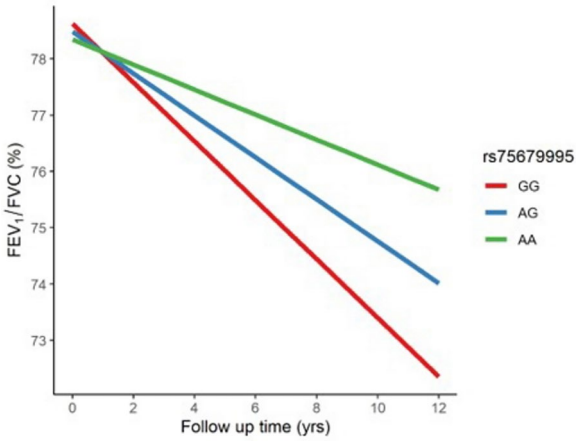


Fig. 7. Estimated FEV_1/FVC according to time interaction and rs75679995 The estimated FEV1/FVC ratio was generated by gene-by-time analysis using KARE data. Minor allele G has negative impact on increment of time interaction.

and *AADAT* tended to have low FEV_1/FVC values. However, the statistical model for the analysis of gene-time interactions should be selected carefully. This was because the effects of time on FEV_1/FVC were very strong, and the means and variances may differ according to genetic variants.

FAM13A was reported to be associated with the FEV_1/FVC ratio in the CHARGE consortium[9] and was strongly associated with COPD susceptibility in the previous GWAS[8,17,18]. In this study, *FAM13A* was strongly associated with the FEV_1/FVC ratio. Although this locus neither reached genome-wide significance nor was replicated in the GENIE cohort, this previously reported region could be associated with lung function in the Korean population.

The rs75679995 SNP was the second-most significant SNP in our cohort, which was found in the intron of *DNAH11*, a dynein-coding gene. 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 signal transduction, which in turn regulate many biological processes, such as cell survival, differentiation, migration[19,20], and immune and inflammatory responses[21,22].

Primary ciliary dyskinesia (PCD) is a rare autosomal recessive genetic disorder characterised by dysfunction of motile cilia. *DNAH11* mutations are reported in 6% of all PCDs, and 22% of those with normal ultrastructure[23–25]. Generally, ciliary dysmotility causes poor mucociliary clearance and leads to the 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 results of eQTL analysis of rs9991425 revealed high expression of both *MFAP3L* and *AADAT*. We further analysed *MFAP3L* and *AADAT* expression using the GTEx portal, which revealed that *MFAP3L* was upregulated in muscle while *AADAT* was expressed in cultured fibroblasts. The numbers of transcripts per million kilobases for *MFAP3L* and *AADAT* in lung tissue were 1 and 6.286, respectively. Thus, it was not clear as to which gene contributed to the significant effect of rs9991425 on FEV_1/FVC ratio; however, *AADAT* might be a more promising candidate, as its expression in lung tissue is higher than that of *MFAP3L*. This gene plays an important role in thyroid hormone regulation[26] and lung cancer[27] and is a major negative regulator of amino acid metabolism involving the mitochondria in cultured fibroblasts, suggesting that *AADAT* may be involved in the pathogenesis of COPD or in lung function decline.

Increasing evidence indicates that the mitochondria are involved in cellular functions beyond oxygen sensing and energy production. Mitochondria can sense upstream processes such as inflammation, infection, presence of tobacco smoke, and environmental insults[28–30], which play important roles in these diseases, and can subsequently respond to these stimuli by altering mitochondrial protein expression and structure (and resultant dysfunction). In contrast, mitochondrial dysfunction affects cytosolic and mitochondrial calcium regulation, airway contractility, expression of housekeeping genes, responses to oxidative stress, proliferation, apoptosis, fibrosis, and metabolism, which are all key aspects of airway disease pathophysiology[31]. To the best of our knowledge, this is the first study to demonstrate a potentially novel COPD susceptibility locus in *AADAT* on chromosome 4.

A report has shown that > 70% patients with COPD have GOLD stage 1 (mild) or 2 (moderate) disease with no apparent respiratory symptoms such as dyspnoea on exertion[32]. Previous studies have shown that tiotropium can ameliorate the annual decline in FEV_1 and reduce the frequency of acute exacerbation compared to that observed in placebo-treated patients with GOLD stage 1 or 2 COPD[33,34]. Therefore, earlier detection of COPD or lung function decline may lead to early effective intervention, reducing disease progression and the associated socioeconomic burden. Our observations are clinically important as they provide meaningful insights regarding the genetic risk of lung function decline in COPD. We proposed a method that revealed the complexity of gene–time interaction analyses, identified consistent gene–time interactions, and proposed a statistical model for association. Interest in early detection of the disease has been increasing in the USA in recent years, where COPD monitoring is performed by the National Lung Health Education Program. However, considering that individuals may suddenly undergo rapid decline in lung function, performing selective PFT only in patients with genetic susceptibility to COPD, rather than in all patients, may be cost-effective. The results of our study are especially noteworthy as studies investigating the genetic risk factors for early COPD or lung function decline are lacking despite substantial progress in evaluating the genetic susceptibility to COPD.

Although we were able to identify susceptibility loci related to decrease in lung function, unfortunately, these data were not identically replicated in the GENIE study. For example, although an interaction between time and SNPs was found for the KARE data, the interaction between pack years and SNPs was not significant for the GENIE data. However, this may be attributed to the fact that in our replication of the GENIE study, the characteristics of the patients differed from that obtained from the KARE data. In particular, 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[35].

Our study provides novel findings but has several notable limitations. First, the definition of COPD is based on pre-bronchodilator data. Second, we did not perform functional in vitro analysis to determine whether these genetic variants affect early COPD. Additionally, while some of the gene–environment interactions identified did not reach traditional levels of statistical significance, they remain suggestive based on clinical observations and trends across multiple analyses.

In conclusion, our study suggests that several SNPs located on chromosome 4 and in FAM13A may be associated with early COPD. Additionally, we propose that DNAH11 and AADAT could be involved in susceptibility to lung function decline or COPD. Further large-scale studies are needed to validate these findings.

Materials and methods

Study populations and genetic information

Data from two independent community-based cohorts of Ansan (urban) and Ansung (rural), included in the Korean Genome and Epidemiology Study (KoGES), were analysed. Korean men and women aged 39–70 years between 2001 and 2002 were enrolled in both cohorts for a prospective 12-year investigation, with assessment every two years (see Supplementary Text 1 for detailed information on cohorts). Genetic information was obtained from KARE, a multidisciplinary research consortium that began in 2007, to conduct a large-scale GWAS (see Supplementary Text 2 for detailed consortium information) of the Ansan and Ansung cohorts in the KoGES[36]. In addition, individuals who did not complete a pulmonary function test (PFT), genetic information of whom were lacking, or did not report smoking status were excluded.

The Korea Centre for Disease Control and Prevention obtained written informed consent from all participants regarding the collection of their data, and the Institutional Review Board of Korea Ansan Hospital (IRB No. 2019AS0102) and Seoul National University (IRB No. E1605/E002-003) approved the study. All methods were performed in accordance with the approved protocol and the relevant guidelines and regulations.

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 standardised protocols of the American Thoracic Society[37]. All participants underwent a prebronchodilator spirometry test until completing at least three repeated measurements, and an acceptable measurement was considered to occur when the differences between the largest and the next largest FVC and FEV_1 values were within 0.15 L. Calibration and quality control of spirometric examinations were also performed regularly based on the guidelines of the American Thoracic Society.

Discovery analysis using KARE

Decrease in lung function was assessed using KARE data on FEV_1 and GWIS data on FEV_1/FVC. The values of FEV_1 and FEV_1/FVC were observed up to seven times every two years, and at least one or more individuals who had undergone PFT were included (Supplementary Table S1). Genotype data were obtained using the Affymetrix genome-wide human SNP array 5.0[36], and quality control (QC) analyses were performed for participants who underwent spirometry analysis, and their smoking history was recorded. After QC process and imputation (see Supplementary Text 3 for detailed procedures about QC and genotypic imputation), 8,554 individuals and 3,333,374 SNPs were included for the analyses (Fig. 8).

Validation analyses using the genomics evidence neoplasia information exchange (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 participate in genetic studies[38]. The genotype of the participants was analysed using the Affymetrix Axiom KORV1.1–96 Array20, and genotype QC was performed. The 4413 participants who had FEV_1/FVC ratio ≥ 70 were ≥ 40 years old, and the spirometric data and smoking history of whom were available 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.

GWIS with KARE data

The FEV_1/FVC ratios were applied as spirometric measures, which were used to identify the genetic variants interacting with time. GWIS were conducted using KARE data. Analyses using a linear mixed model were performed with age, sex, height, body mass index, smoking pack year, and smoking status as covariates. To adjust for population substructure strictly, principal component analyses (PCA) were applied to the genetic relationship matrix, and the first 10 PCA scores were included as covariates. *yij* was the FEV_1/FVC value 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/FVC was as follows.

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

$$\begin{bmatrix} b_0 \\ b_1 \end{bmatrix} \sim MVN(0, \Sigma), (\varepsilon_{i1}, \ldots, \varepsilon_{in_i})^t \sim MVN(0, \Sigma\prime)$$

It should be noted that never smokers were included in the model, with the pack-year variable set to 0 for these individuals. We compared several structures for Σ and Σ′, and selected an unstructured covariance structure for Σ and AR1 structures for Σ′. The proposed models were applied to detect gene-time interactions for average FEV_1/FVC levels. To identify SNPs interacting with time on spirometric measures, we considered H0: β8 = β9 = 0. The most significant SNPs were selected for further analyses of gene-time interaction effects.

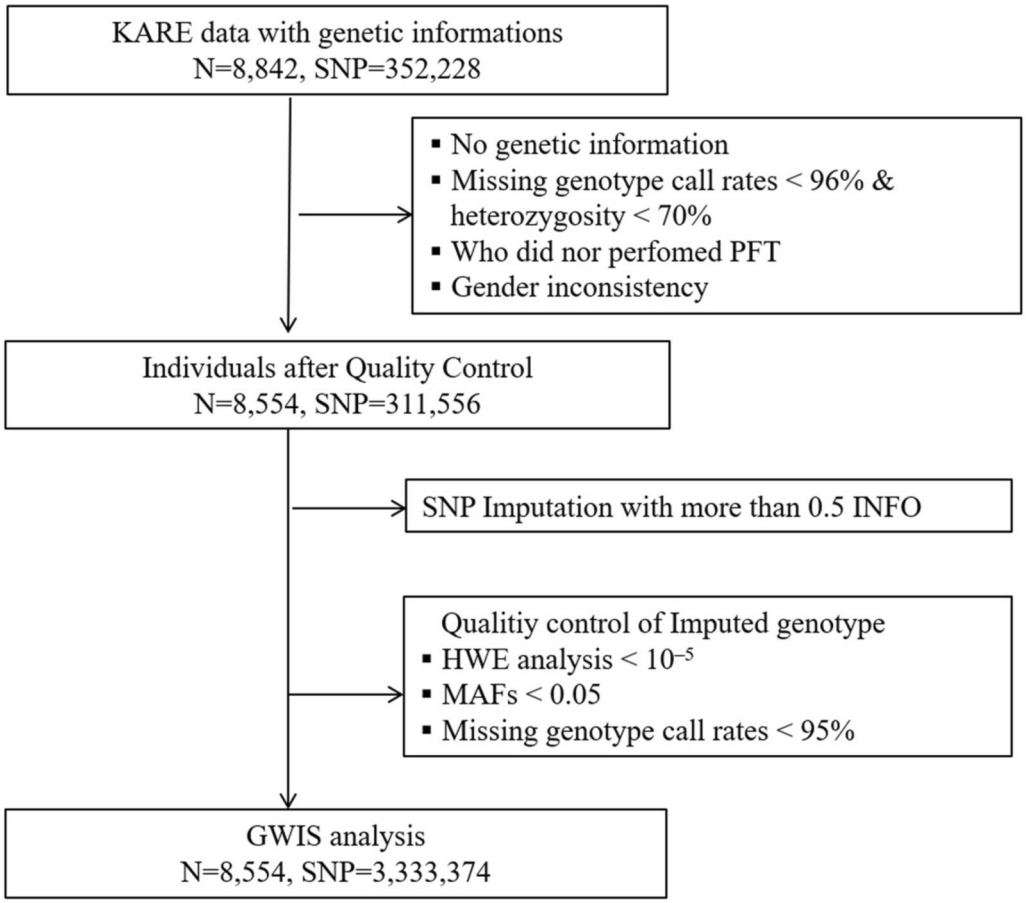


Fig. 8. Flow diagram for the KARE cohort. Flow chart showing how the individuals and SNPs were included and excluded in the study.

Data availability

Genotype and clinical data for KARE can be downloaded from the Korea Disease Control and Prevention Agency (KDCA) website (https://www.kdca.go.kr/), subject to an approval process by the Korean NIH. For more information on the approval process, please contact biobank@korea.kr.

Received: 6 April 2024; Accepted: 5 March 2025
Published online: 21 March 2025

References

1. Sorlie, P. D., Kannel, W. B. & O'Connor, G. Mortality associated with respiratory function and symptoms in advanced age: The Framingham Study. *Am. Rev. Respir. Dis.* **140**, 379–384 (1989).
2. Knuiman, M. W. et al. Lung function, respiratory symptoms, and mortality: Results from the Busselton Health Study. *Ann. Epidemiol.* **9**, 297–306 (1999).
3. Pauwels, R. A. & Rabe, K. F. Burden and clinical features of chronic obstructive pulmonary disease (COPD). *Lancet* **364**, 613–620 (2004).
4. Viegi, G. et al. Definition, epidemiology and natural history of COPD. *Eur. Respir. J.* **30**, 993–1013 (2007).
5. Postma, D. S., Bush, A. & van den Berge, M. Risk factors and early origins of chronic obstructive pulmonary disease. *Lancet* **385**, 899–909 (2015).
6. Venkatesan, P. GOLD COPD report: 2023 update. *Lancet Respir. Med.* **11**, 18 (2023).
7. Shrine, N. et al. New genetic signals for lung function highlight pathways and chronic obstructive pulmonary disease associations across multiple ancestries. *Nat. Genet.* **51**, 481–493 (2019).
8. Cho, M. H. et al. Variants in FAM13A are associated with chronic obstructive pulmonary disease. *Nat. Genet.* **42**, 200–202 (2010).
9. Hancock, D. B. et al. Meta-analyses of genome-wide association studies identify multiple loci associated with pulmonary function. *Nat. Genet.* **42**, 45–52 (2010).
10. Pillai, S. G. 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.* **182**, 1498–1505 (2010).
11. Repapi, E. et al. Genome-wide association study identifies five loci associated with lung function. *Nat. Genet.* **42**, 36–44 (2010).
12. Soler Artigas, M. 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.* **184**, 786–795 (2011).
13. Imboden, M. et al. Genome-wide association study of lung function decline in adults with and without asthma. *J. Allergy Clin. Immunol.* **129**, 1218–1228 (2012).

14. Tang, W. et al. Large-scale genome-wide association studies and meta-analyses of longitudinal change in adult lung function. *PLoS ONE* **9**, e100776 (2014).
15. Gottlieb, D. J. et al. Heritability of longitudinal change in lung function: The Framingham study. *Am. J. Respir. Crit. Care Med.* **164**, 1655–1659 (2001).
16. Wilk, J. B., Walter, R. E., Laramie, J. M., Gottlieb, D. J. & O'Connor, G. T. Framingham heart study genome-wide association: Results for pulmonary function measures. *BMC Med. Genet.* **8**, S8 (2007).
17. Siedlinski, M. et al. Dissecting direct and indirect genetic effects on chronic obstructive pulmonary disease (COPD) susceptibility. *Hum. Genet.* **132**, 431–441 (2013).
18. Ziółkowska-Suchanek, I. et al. Susceptibility loci in lung cancer and COPD: Association of IREB2 and FAM13A with pulmonary diseases. *Sci. Rep.* **5**, 13502 (2015).
19. Cuadrado, A. & Nebreda, A. R. Mechanisms and functions of p38 MAPK signalling. *Biochem. J.* **429**, 403–417 (2010).
20. Cuenda, A. & Rousseau, S. p38 MAP-kinases pathway regulation, function and role in human diseases. *Biochim. Biophys. Acta* **1773**, 1358–1375 (2007).
21. Wang, C. Y. et al. A novel family of adhesion-like molecules that interacts with the NMDA receptor. *J. Neurosci.* **26**, 2174–2183 (2006).
22. Lee, J. C. et al. A protein kinase involved in the regulation of inflammatory cytokine biosynthesis. *Nature* **372**, 739–746 (1994).
23. Knowles, M. R., Daniels, L. A., Davis, S. D., Zariwala, M. A. & Leigh, M. W. Primary ciliary dyskinesia: Recent advances in diagnostics, genetics, and characterization of clinical disease. *Am. J. Respir. Crit. Care Med.* **188**, 913–922 (2013).
24. Leigh, M. W. et al. Clinical and genetic aspects of primary ciliary dyskinesia/Kartagener syndrome. *Genet. Med.* **11**, 473–487 (2009).
25. Pifferi, M. et al. New DNAH11 mutations in primary ciliary dyskinesia with normal axonemal ultrastructure. *Eur. Respir. J.* **35**, 1413–1416 (2010).
26. Teumer, A. et al. Genome-wide analyses identify a role for SLC17A4 and AADAT in thyroid hormone regulation. *Nat. Commun.* **9**, 4455 (2018).
27. Hsu, C. C. et al. Lysine deprivation induces AKT-AADAT signaling and overcomes EGFR-TKIs resistance in EGFR-mutant non-small cell lung cancer cells. *Cancers* **13**, 272 (2021).
28. Hoffmann, R. F. et al. Prolonged cigarette smoke exposure alters mitochondrial structure and function in airway epithelial cells. *Respir. Res.* **14**, 97 (2013).
29. Tiku, V., Tan, M. W. & Dikic, I. Mitochondrial functions in infection and immunity. *Trends Cell Biol.* **30**, 263–275 (2020).
30. Duarte-Hospital, C. et al. Mitochondrial dysfunction as a hallmark of environmental injury. *Cells* **11**, 110 (2021).
31. Prakash, Y. S., Pabelick, C. M. & Sieck, G. C. Mitochondrial dysfunction in airway disease. *Chest* **152**, 618–626 (2017).
32. Zhong, N. et al. Prevalence of chronic obstructive pulmonary disease in China: A large, population-based survey. *Am. J. Respir. Crit. Care Med.* **176**, 753–760 (2007).
33. Zhou, Y. et al. Tiotropium in early-stage chronic obstructive pulmonary disease. *N. Engl. J. Med.* **377**, 923–935 (2017).
34. van der Molen, T. & Kirenga, B. J. COPD: Early diagnosis and treatment to slow disease progression. *Int. J. Clin. Pract.* **69**, 513–514 (2015).
35. Mackenbach, J. P., Stronks, K. & Kunst, A. E. 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.* **29**, 369–376 (1989).
36. Cho, Y. S. et al. A large-scale genome-wide association study of Asian populations uncovers genetic factors influencing eight quantitative traits. *Nat. Genet.* **41**, 527–534 (2009).
37. Miller, M. R. et al. Standardisation of spirometry. *Eur. Respir. J.* **26**, 319–338 (2005).
38. Lee, C. 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* **8**, e019327 (2018).

Acknowledgements

Data in this study were obtained from the Korean Genome and Epidemiology Study (4851-302), National Research Institute of Health, Centres for Disease Control and Prevention, Ministry for Health and Welfare, South Korea.

Author contributions

Concept and design: YSK, SW, and YSK Data management: CYK and BP Analysis and interpretation: All authors Drafting the manuscript for important intellectual content: CYK and BP Critical revision and final approval: All authors.

Declarations

Competing interests

The authors declare no competing interests.

Ethics approval

The Korea Centre for Disease Control and Prevention obtained written informed consent from all participants regarding the collection of their data, and the Institutional Review Board of Korea Ansan Hospital (IRB No. 2019AS0102) and Seoul National University (IRB No. E1605/E002-003) approved the study. All methods were performed in accordance with the approved protocol and the relevant guidelines and regulations.

Additional information

Supplementary Information The online version contains supplementary material available at https://doi.org/10.1038/s41598-025-93147-6.

Correspondence and requests for materials should be addressed to S.W. or Y.S.K.