

Accepted Manuscript

Genome-wide association study on the FEV₁/FVC-ratio in never-smokers identifies *HHIP* and *FAM13A*

Diana A. van der Plaat, MSc, Kim de Jong, PhD, Lies Lahousse, PhD, Alen Faiz, PhD, Judith M. Vonk, PhD, Cleo C. van Diemen, PhD, Ivana Nedeljkovic, MD, Najaf Amin, PhD, Prof Guy G. Brusselle, PhD, Prof Albert Hofman, MD, Corry-Anke Brandsma, PhD, Yohan Bossé, PhD, Prof Don D. Sin, PhD, David C. Nickle, PhD, Prof Cornelia M. van Duijn, PhD, Prof Dirkje S. Postma, MD, Prof H Marika Boezen, PhD

PII: S0091-6749(16)30895-8

DOI: [10.1016/j.jaci.2016.06.062](https://doi.org/10.1016/j.jaci.2016.06.062)

Reference: YMAI 12321

To appear in: *Journal of Allergy and Clinical Immunology*

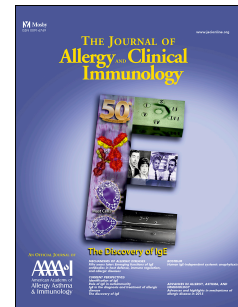
Received Date: 17 February 2016

Revised Date: 6 June 2016

Accepted Date: 10 June 2016

Please cite this article as: van der Plaat DA, de Jong K, Lahousse L, Faiz A, Vonk JM, van Diemen CC, Nedeljkovic I, Amin N, Brusselle GG, Hofman A, Brandsma C-A, Bossé Y, Sin DD, Nickle DC, van Duijn CM, Postma DS, Boezen HM, Genome-wide association study on the FEV₁/FVC-ratio in never-smokers identifies *HHIP* and *FAM13A*, *Journal of Allergy and Clinical Immunology* (2016), doi: 10.1016/j.jaci.2016.06.062.

This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.



Genome-wide association study on the FEV₁/FVC-ratio in never-smokers identifies *HHIP* and *FAM13A*

Authors: Diana A van der Plaat MSc^{1,2}, Kim de Jong PhD^{1,2}, Lies Lahousse PhD^{3,4}, Alen Faiz PhD^{5,2}, Judith M Vonk PhD^{1,2}, Cleo C van Diemen PhD⁶, Ivana Nedeljkovic MD³, Najaf Amin PhD³, Prof Guy G Brusselle PhD^{3,7,4}, Prof Albert Hofman MD³, Corry-Anke Brandsma PhD^{5,2}, Yohan Bossé PhD⁸, Prof Don D Sin PhD^{9,10}, David C Nickle PhD¹¹, Prof Cornelia M van Duijn PhD³, Prof Dirkje S Postma MD^{12,2}, Prof H Marika Boezen PhD^{1,2}

Author affiliation:

¹University of Groningen, University Medical Center Groningen, Department of Epidemiology, Groningen, the Netherlands. ²University of Groningen, University Medical Center Groningen, Groningen Research Institute for Asthma and COPD (GRIAC), Groningen, the Netherlands. ³Erasmus Medical Center, Department of Epidemiology, Rotterdam, the Netherlands. ⁴Ghent University Hospital, Department of Respiratory Medicine, Ghent, Belgium. ⁵University of Groningen, University Medical Center Groningen, Department of Pathology and Medical Biology, Groningen, the Netherlands. ⁶University of Groningen, University Medical Center Groningen, Department of Genetics, Groningen, the Netherlands. ⁷Erasmus Medical Center, Department of Respiratory Medicine, Rotterdam, the Netherlands. ⁸Laval University, Institut universitaire de cardiologie et de pneumologie de Québec, Department of Molecular, Medicine, Québec, Canada. ⁹University of British Columbia, Department of Medicine, Division of Respiratory Medicine, Vancouver, Canada. ¹⁰The University of British Columbia Center for Heart Lung Innovation, St Paul's Hospital, Vancouver, Canada. ¹¹Merck & Co., MRL, Seattle, WA, United States of America. ¹²University of Groningen, University Medical Center Groningen, Department of Pulmonary Diseases, Groningen, the Netherlands.

Correspondence to: Prof. H. Marika Boezen, PhD, University of Groningen, University Medical Center Groningen, Department of Epidemiology, Hanzeplein 1, 9700 RB Groningen, the Netherlands (e-mail: h.m.boezen@umcg.nl).

Contributors

DAvdP participated in the study design, analysis and interpretation of the data, and drafting of the manuscript, tables and figures. HMB, DSP, CCvD and CMVD obtained funding. KdJ, HMB, DSP, JMV, CCvD, CMVD, IN and NA determined the study design, participated in the analysis and interpretation of data, and critically supervised writing of the manuscript. LL, GGB and AH participated in collecting and analyzing the Rotterdam Study data. AF, CAB, YB, DDS and DCN participated in setting up the lung tissue database and analyses. All authors approved the final version of the manuscript.

Role of the funding source

This study is sponsored by grant number 4.1.13.007 of Lung Foundation (Longfonds), the Netherlands. DAVdP and KdJ are supported by grant number 4.1.13.007 of Longfonds. LL is a Postdoctoral Fellow of the Research Foundation-Flanders (FWO).

The LifeLines cohort study was supported by the Dutch Ministry of Health, Welfare and Sport; the Ministry of Economic Affairs, Agriculture and Innovation; the Province of Groningen; the European Union (regional development fund); the Northern Netherlands Provinces (SNN); the Netherlands Organization for Scientific Research (NWO); University Medical Center Groningen (UMCG); University of Groningen, de Nierstichting (the Dutch Kidney Foundation); and the Diabetes Fonds (the Diabetic Foundation).

The Vlagtwedde-Vlaardingens cohort study was supported by the Ministry of Health and Environmental Hygiene of the Netherlands and the Netherlands Asthma Fund (grant 187) and the Netherlands Asthma Fund grant no. 3.2.02.51, the Stichting Astma Bestrijding, BBMRI-NL (Complementation project), and the European Respiratory Society COPD research award 2011 to H.M. Boezen.

The Rotterdam Study was supported by the Erasmus MC and Erasmus University Rotterdam; the Netherlands Organisation for Scientific Research (NWO); the Netherlands Organisation for Health Research and Development (ZonMw); the Research Institute for Diseases in the Elderly (RIDE); the Netherlands Genomics Initiative (NGI); the Ministry of Education, Culture and Science; the Ministry of Health, Welfare and Sports; the European Commission (DG XII); and the Municipality of Rotterdam.

The Lung eQTL study was funded by Merck Research Laboratories.

The sponsors of this study played no role in the design of the study, data collection, analysis, interpretation or in the writing and submission of the manuscript.

66 **Disclosure of potential conflict of interest**

67 DDS reports personal fees from Amgen, grants and personal fees from AstraZeneca, personal
68 fees from Boehringer Ingelheim, grants from Novartis, outside of the submitted work. DCN is
69 an employee of Merck and Co. DSP reports grants to the university and consultancy fees to
70 the university from AstraZeneca, Boehringer Ingelheim, Ghiesi, GlaxoSmithKline, Takeda,
71 and TEVA outside of the submitted work. All other authors declare no competing interests.

Abstract

Background: Although a striking proportion (25-45%) of subjects with chronic obstructive pulmonary disease (COPD) are never-smokers, most genetic susceptibility studies have not focused on this group exclusively.

Objective: The aim of this study was to identify common genetic variants associated with the level of forced expiratory volume in one second (FEV₁) and its ratio to forced vital capacity (FEV₁/FVC) in never-smokers.

Methods: Genome-wide association studies (GWAS) were performed in 5,070 never-smokers of the identification cohort LifeLines, and results (p-value <10⁻⁵) were verified by a meta-analysis of the Vlagtwedde-Vlaardingen study and the Rotterdam Study I-III (total n=1,966). Furthermore, we aimed to assess the effects of the replicated variants in more detail by performing genetic risk score (GRS), expression quantitative trait loci (eQTL), and variant*ever-smoking interaction analyses.

Results: We identified associations between FEV₁/FVC and five common genetic variants in the identification cohort and two of these associations were replicated. The two variants annotated to the genes *HHIP* and *FAM13A* showed to have an additive effect on FEV₁/FVC level in the GRS analysis, were associated with gene expression of *HHIP* and *FAM13A* in lung tissue, respectively, and were genome-wide significant in a meta-analysis including both identification and four verification cohorts (p-value <2.19x10⁻⁷). Finally, we did not identify significant interactions between the variants and ever-smoking. Results of the FEV₁ identification analysis were not replicated.

Conclusion: The genes *HHIP* and *FAM13A* confer a risk for airway obstruction in general that is not driven exclusively by cigarette smoking, the main risk factor for COPD.

Clinical implications: The genes *HHIP* and *FAM13A* confer a risk for airway obstruction in general that is not exclusively driven by the main COPD risk factor, i.e. cigarette smoking.

Capsule Summary: The information provided in this article may be important for understanding worldwide COPD prevalence and may contribute to the understanding of mechanisms involved in the development of airway obstruction and COPD in general.

Key words: *Genome-Wide Association Study; never-smokers; genetics; pulmonary function; COPD*

Abbreviations used

COPD:	Chronic Obstructive Pulmonary Disease
eQTL:	Expression quantitative trait locus
<i>FAM13A</i> :	Family with sequence similarity 13 member A
FEV ₁ /FVC:	Ratio of forced expiratory volume in one second to forced vital capacity
GOLD:	Global initiative for chronic Obstructive Lung Disease
GRS:	Genetic Risk Score
GWAS:	Genome-wide association study
<i>HHIP</i> :	Hedgehog interacting protein
LD:	Linkage disequilibrium
MAF:	Minor allele frequency
SNP:	Single-nucleotide polymorphism

Introduction

Chronic obstructive pulmonary disease (COPD) is the third leading cause of death in the world and is characterized by impaired pulmonary function and progressive airway obstruction.^{1,2} It is estimated that 10% of people over the age of 40 years develop COPD.³ Impaired pulmonary function and airway obstruction are commonly diagnosed based on the spirometric measurements of forced expiratory volume in one second (FEV₁) and its ratio to forced vital capacity (FEV₁/FVC).⁴

Common genetic variants underlying pulmonary function have been identified by several genome-wide association (GWA) studies. Multiple single nucleotide polymorphisms (SNPs) were found to be associated with FEV₁ and/or FEV₁/FVC in population-based studies, including both ever- and never-smokers, amongst others SNPs annotated to the genes *ADAM13*, *AGAR*, *CHRNA3/5*, *FAM13A*, *GPR126*, *HHIP*, *HTR4*, *PID1*, *PTCH1*, and *THSD4*.⁵⁻⁷ Several of these genes, i.e. *CHRNA3/5*, *FAM13A* and *HHIP*, were also identified in case-control studies investigating the risk for COPD, including only ever-smokers with at least 2.5 pack-years of smoking.^{8,9} Cigarette smoking is considered the most important risk factor for developing airway obstruction in the industrialized world.² This explains the predominantly smoking-related focus of previous studies on COPD susceptibility genes, yet this ignores potential genetics underlying COPD development in never-smokers. Our previous candidate-gene based research suggested that some genes are related to pulmonary function impairment or COPD in both ever- and never-smokers,¹⁰⁻¹² however other genes seem to affect pulmonary function exclusively in never-smokers.¹³ Thus far, hardly any research explicitly focused on assessing the genetic background of airway obstruction in never-smokers, although a striking proportion of 25 to 45% of COPD patients worldwide are never-smokers.¹⁴ Few GWA studies on continuous FEV₁ and FEV₁/FVC levels, have presented stratified results by smoking status, yet only for their top hits selected based on analyses including both ever- and never-smokers.^{6,7,15} Two studies thus far have performed GWA analyses separately in ever and never-smokers, yet investigated pulmonary function as dichotomous outcome (airway obstruction or extremes of FEV₁, respectively) and not as continuous outcome.^{5,16}

The current study aimed to identify common genetic variants associated with pulmonary function (FEV₁ and FEV₁/FVC as continuous outcomes) in never-smokers and aimed to assess the effects of these variants in more detail by performing genetic risk score

156 (GRS) and expression quantitative trait loci (eQTL). Additionally, we also investigated the
157 genetic effects in ever-smokers and SNP*ever-smoking interactions.

158

Methods

Subject selection

All never-smokers with available genotype data from the Dutch population-based cohort study LifeLines were included as the identification cohort.¹⁷ To verify our findings, we analyzed never-smokers of four independent verification cohorts, namely the Vlagtwedde-Vlaardingen study¹⁸ and the three cohorts of the Rotterdam Study (RS I-III).^{19,20} Subjects were defined as 'never-smokers' based on self-reported never-smoking history and 0 pack-years. All subjects provided written informed consent and the studies were approved by the medical ethics committee of either the University Medical Center Groningen (LifeLines and Vlagtwedde-Vlaardingen cohorts) or the Erasmus University (RS I-III). Additional details on the cohorts are provided in the supplementary material.

Genotyping

In the LifeLines cohort study and the Vlagtwedde-Vlaardingen study, IlluminaCytoSNP-12 arrays were used to genotype blood samples.^{17,18} Illumina Human610-Quad BeadChips were used to genotype blood samples of the Rotterdam Study I-III.¹⁹ Single nucleotide polymorphism (SNPs) with a genotype call-rate $\geq 95\%$, minor allele frequency $\geq 1\%$ and Hardy-Weinberg p -value $\geq 10^{-4}$ were included. Non-Caucasian samples and first-degree relatives were excluded based on self-reporting, outlier (IBS) and principal components analysis.

Statistical analysis

After quality control, a total of 227,981 genotyped SNPs were included in the identification analysis. GWA analyses were performed on two spirometric measures of pulmonary function, i.e. pre-bronchodilator FEV₁ (mL) and the FEV₁/FVC ratio (%), measured in all cohorts according to American Thoracic Society (ATS) or European Respiratory Society (ERS) criteria.²¹ PLINK²² and GRIMP²³ software packages were used for the GWA analyses. Linear regression models (additive genetic model) were adjusted for sex, age and height. The major allele was used as reference allele and the reported effect estimates are for the minor allele (A1). Genotyped SNPs associated with pulmonary function, with a p -value $< 10^{-5}$ in the identification analysis were selected for verification. Analyses were performed in the four verification cohorts separately and then meta-analyzed using effect

estimates weighted by the inverse of the standard errors, using METAL.²⁴ Genotyped SNPs with a meta-analyzed fixed effects p-value <0.05 and with the same direction of effect in all cohorts were considered to be significantly replicated.

Genetic risk score analysis

To assess the cumulative effect of the replicated SNPs on pulmonary function, a genetic risk score (GRS) was constructed for each subject by calculating the total number of risk alleles of the replicated SNPs, for both the identification and replication cohorts separately. Using SPSS,²⁵ the linear regression models were adjusted for sex, age and height. Subsequently, we meta-analyzed the effect estimates of the four verification cohorts and all five cohorts.

Gene expression analysis (eQTL analysis)

We assessed if the selected SNPs were cis-acting expression quantitative trait loci (eQTL) by using a lung tissue database established by the lung eQTL consortium.²⁶ In total, 1,087 subjects were included, and probesets located within 2Mb of the replicated SNPs (4Mb window) were selected. The linear regression model was adjusted for disease status, age, gender, smoking status and cohort specific principal components. SNPs with a p-value below the Bonferroni corrected threshold ($p\text{-value}=0.05/\text{number of probes in window}$) were considered significantly associated with gene expression level (mRNA). As a sensitivity analysis, we performed the same analysis in only never-smoker of the lung tissue database (n=155) for the significant SNP probeset combinations. See supplementary material and GEO accession numbers GSE23546 and GPL10379 for additional information.

Ever-smokers

Ever-smokers of the identification cohort LifeLines were investigated to assess if the replicated SNPs were associated with pulmonary function in smokers as well. Ever-smoking was based on self-reporting smoking history and having smoked >5 pack-years. Linear regression models were adjusted for sex, age, height, pack-years, and current-smoking. In addition, we tested whether the associations of the replicated SNPs with pulmonary function were significantly different between ever- and never-smokers by adding the interaction term SNP*ever-smoking to the model (adjusted for sex, age, height, pack-years, and ever-smoking).

Results

Population characteristics

An overview of the study design is shown in Figure E1 in the supplementary material. Complete data on all covariates were available for 7,036 subjects of the five cohorts (identification cohort $n=5,070$ and verification cohorts $n=1,966$). The characteristics of never-smokers included in all cohorts are shown in Table 1, and those for ever-smokers of the identification cohort LifeLines in Table E1 of the supplementary material. Of interest, based on pre-bronchodilator measurements the identification cohort included 548 never-smokers (11%) with airway obstruction ($FEV_1/FVC < 70\%$) and 106 never-smokers (2%) with Global initiative for chronic Obstructive Lung Disease (GOLD) COPD stage two or higher ($FEV_1/FVC < 70\%$ and $FEV_1\%p < 80\%$). The genomic inflation factor lambda was low ($\lambda=1.02$) for both analyses on FEV_1 and FEV_1/FVC in the identification sample, indicating minimal population stratification. The quantile-quantile plots (Q-Q plots) can be found in Figure E2.

FEV_1

Only one SNP (rs17057154) was associated with FEV_1 level with a p-value $< 10^{-5}$ in never-smokers of the identification cohort (Table E2 and Figure E3 A). This SNP was associated with FEV_1 level with a p-value < 0.05 in the Vlagtwedde-Vlaardingen cohort, however, the effect estimate was in the opposite direction. Hence, this SNP was not further investigated in the Rotterdam Study or in additional analyses.

FEV_1/FVC

We identified five SNPs associated with FEV_1/FVC level in never-smokers in the identification cohort (p-value $< 10^{-5}$, Table 2 and Figure E3 B). The associations of these five SNPs with FEV_1 and FVC levels separately can be found in table E3. In the verification meta-analysis, two of these SNPs were associated with FEV_1/FVC level (fixed effect p-value < 0.05) and had the same direction of effect in all cohorts (Figure 1 A/B and Table E4). In addition, when meta-analyzing all five cohorts, both identification and verification cohorts, the same two SNPs reached genome-wide significance according to the Bonferroni-corrected threshold (p-value $< 2.19 \times 10^{-7}$; Table 2). The replicated SNPs are located in or near the genes Hedgehog Interacting Protein (*HHIP*, rs1512282) and Family With Sequence Similarity 13,

Member A (*FAM13A*; rs6849143). Regional association plots are shown in Figure E4. SNP rs1512282 is located 135Kb upstream from *HHIP* (4q31.21) and the minor allele (G, MAF 41%) was associated with higher FEV₁/FVC levels in never-smokers (Figure E5 A). SNP rs6849143 is located within an intron of *FAM13A* (4q22.1), with the minor allele (T, MAF 43%) being associated with lower FEV₁/FVC levels (Figure E5 B).

Genetic risk score analysis

We performed a genetic risk score analysis using the two replicated SNPs to assess their cumulative effect on FEV₁/FVC level in never-smokers of the identification and verification cohorts. In the identification cohort, the GRS indicated a strong significant association and an additive effect for each additional allele ($B = -0.58\%$ FEV₁/FVC, p -value = 2.87×10^{-11} , Figure E6). FEV₁/FVC levels in subjects carrying all four risk alleles were significantly lower (2.4%) compared to subjects not carrying any risk allele, an adjusted mean FEV₁/FVC level of 78.3% and 80.7% respectively. The GRS was also significant and had a similar effect estimate in the meta-analysis of the verification cohorts ($B = -0.53\%$) and all cohorts ($B = -0.57\%$, Figure 1C).

Gene expression analysis (eQTL analysis)

We found significant associations of the two replicated SNPs with gene expression levels (mRNA level) in lung tissue (Table 3 and Figure 2). The minor allele (G) of SNP rs1512282 was significantly associated with decreased levels of expression of three probesets located in *HHIP*. The minor allele (T) of SNP rs6849143 was significantly associated with increased expression of a probeset in the gene *TIGD2* (Tigger Transposable Element Derived 2), and with increased expression of three probesets located in *FAM13A*. When limiting the eQTL analysis to only never smokers ($n = 155$) we find comparable effect estimates for the most significant probesets in *HHIP* and *FAM13A* (Table 3).

Ever-smokers

To assess if the replicated SNPs were associated with FEV₁/FVC in ever-smokers as well, we investigated associations in ever-smokers (>5 pack-years) of the identification cohort (Table 4). The minor allele (G) of the SNP near *HHIP* (rs1512282) was associated with a higher FEV₁/FVC level in ever-smokers, with an effect estimate almost identical to the association in never-smokers. In line with these observations, the interaction analysis did not

show a significant interaction between the SNP and ever-smoking. The effect estimate for the *FAM13A* SNP (rs6849143) was lower in ever-smokers than in never-smokers, and not significant (p-value >0.05) in ever-smokers. The interaction analysis showed a borderline significant interaction between the SNP and ever-smoking (p-value=0.07).

Discussion

The current study assessed the genetic background of pulmonary function as continuous outcome (FEV_1 and FEV_1/FVC levels) in exclusively never-smokers. We identified two common genetic variants annotated to the genes *HHIP* and *FAM13A* to be associated with FEV_1/FVC levels in never-smokers. The risk alleles of the SNPs account for lower FEV_1/FVC levels and therefore confer a risk for airway obstruction. Our results imply that the genes *HHIP* and *FAM13A* are general susceptibility genes for the development of airway obstruction and confer an effect regardless of exposure to the most important risk factor of COPD, i.e. cigarette smoking.

The genes *HHIP* and *FAM13A* are well-known genes associated with pulmonary function and COPD and have been previously identified in several GWA studies.⁵⁻⁹ These studies were, however, performed in ex- or current smokers only^{8,9} or in a general population adjusted for smoking.^{6,7} In the current study, SNPs annotated to *HHIP* and *FAM13A* were associated with continuous FEV_1/FVC levels in a sample of exclusively never-smokers and these associations were successfully replicated in a meta-analysis of four independent cohorts. Of interest, when meta-analyzing all five included cohorts (identification and replication) both *HHIP* and *FAM13A* SNPs reached genome-wide significance (Table 2).

As described in the introduction, two previous studies have performed GWA analyses separately in never and ever-smokers, yet investigated pulmonary function as dichotomous outcome. The study by Wilk *et al.* did not identify associations between *HHIP* or *FAM13A* and airway obstruction in never-smokers, but did so among all participants.⁵ Furthermore, no differences in associations were identified between heavy- and never-smokers for the *HHIP* locus when assessing low FEV_1 versus high FEV_1 levels in the study by Wain *et al.*¹⁶ Results of the *FAM13A* locus were not presented, since it was not identified as top hit.

In the current study, we found similar effect estimates of FEV_1/FVC level and the *HHIP* SNP in ever-smokers and never-smokers. Effect estimates of FEV_1/FVC level in ever-smokers compared to never-smokers for the *FAM13A* SNP were lower, yet without a significant interaction between ever-smoking and this SNP. In line with this finding, we also did not observe a significant SNP*ever-smoking interaction associated with gene expression of *HHIP* and *FAM13A* in our eQTL analysis (Table E5). Likewise, in the study by Soler Artigas *et al.* no significant differences in associations between FEV_1/FVC and *HHIP* or *FAM13A* SNPs were identified between ever- and never-smokers, yet this may have been expected since these SNPs were selected based on a meta-analysis of association in ever- and

never-smokers.¹⁵ In conclusion, results of the current study show that the genes *HHIP* and *FAM13A* play a role in airway obstruction in never-smokers, thus conferring an effect regardless of exposure to cigarette smoking.

The genetic risk score analysis showed that the combined effect estimate for the two independent SNPs annotated to *FAM13A* and *HHIP* accounted for a 2.4% lower FEV₁/FVC level, a credible clinically relevant effect. These findings are supported by the results of Soler Artigas *et al.*²⁷, who also showed a higher risk for COPD with more risk alleles for five known COPD genes, including *HHIP*. Interestingly, a substantial proportion of our study population, i.e. 6%, is carrying all four risk alleles. The worldwide prevalence of the risk alleles of the *HHIP* and *FAM13A* SNPs are 68% and 47%, respectively. It is thus likely that similar or even higher percentages of people worldwide are carrying all four risk alleles. Moreover, *HHIP* and *FAM13A* have been previously associated with airway obstruction and COPD in Chinese Han, Korean and Hispanic populations as well.²⁸⁻³⁰ Consequently, *HHIP* and *FAM13A* will therefore have an important role in worldwide COPD prevalence.

HHIP and *FAM13A* may have an independent effect on FEV₁/FVC levels since they are involved in different pathways. *HHIP* is known to play an important role during lung embryogenesis, involving fetal lung branching development. Hence, it is reasonable to assume that the effect of *HHIP* on pulmonary function can occur independently of smoking history. The identified *HHIP* SNP in this study is located in an uncharacterized non-coding RNA (ncRNA, <http://www.ncbi.nlm.nih.gov/SNP/>). Speculatively, this ncRNA might be the functional element regulating the *HHIP* gene through *cis*-interaction. In the current study, the minor allele (G) of the *HHIP* SNP was associated with lower gene expression of *HHIP* in lung tissue and a higher FEV₁/FVC level. Zhou *et al.*³¹ showed that *HHIP* expression, both at mRNA and protein level, is lower in COPD lung tissue compared to healthy subjects. They identified a block of genetic variants (± 79 kb) associated with *HHIP* expression about 51 kb away from *HHIP*. The genetic variant (rs1512282) identified in the current study is, however, located just outside this region, but is in moderate linkage disequilibrium (LD) with SNPs located in this block. This area near *HHIP* may function as an enhancer by interacting with the *HHIP* promoter via chromatin looping. Decreased *HHIP* expression, as seen in our lung eQTL analysis, induces an over activation of the hedgehog pathway.³¹ Silencing of *HHIP* leads to differential expression of about 300 genes enriched for cell growth, lung extracellular matrix and genes associated with COPD.³² Together these findings point towards a plausible role for *HHIP* as an airway obstruction susceptibility gene regardless of smoking.

It was suggested that *FAM13A* might play a role in Rho GTPase signaling based on the presence of a Rho-GAP (Ras homologous GTPase-activation protein) domain in the protein sequence.³³ It was shown that SNPs in LD with variants in the Rho-GAP domain of *FAM13A* confer a reduced risk for COPD and lung cancer.³⁴ Jiang *et al.* showed that *FAM13A* null mice were less susceptible to develop cigarette smoke induced emphysema and that loss of *FAM13A* resulted in increased cell proliferation and activation of β -catenin signaling.³⁵ They suggested that *FAM13A* may regulate β -catenin degradation, which may contribute to increased emphysema development. They also observed higher *FAM13A* and lower β -catenin protein levels in lung tissue of COPD patients compared to healthy ex-smokers. Our eQTL analysis showed that the minor allele of the *FAM13A* SNP rs6849143 is associated with higher gene expression of *TIGD2* and *FAM13A* in lung tissue. Our data thus suggests, in line with literature, that genetic variants in *FAM13A*, via higher expression of *FAM13A*, increase individual susceptibility for airway obstruction.

Of interest, we did not identify SNPs in the nicotine acetylcholine receptor (*CHRNA* 3/5) locus to be associated with FEV₁/FVC level in never-smokers. Two SNPs previously associated with COPD in the *CHRNA* locus, were in our study associated with FEV₁/FVC in ever-smokers but not in never-smokers, irrespectively of the comparable size of both groups (Table E6). These finding are in line with results from the only longitudinal study so far on the *CHRNA* locus, which showed that SNPs in the *CHRNA* locus did not have an effect on annual FEV₁ decline, but were however associated with smoking habits.^{36,37} Our current findings again suggest that SNPs in the *CHRNA* locus do not have a direct effect on COPD development, but act via nicotine dependence. In addition, we tested other genes previously identified in GWA studies on FEV₁/FVC in our cohort of never-smokers. The results of this analysis can be found in supplementary Table E7.

Our study has several potential limitations. First, the results of the identification GWA study on FEV₁ level were not replicated and therefore no additional analyses were performed. Second, since this is an exploratory study, we used a less stringent p-value of $<10^{-5}$ for our identification sample. However, we replicated two of our findings in a meta-analyses including four independent cohorts and reached genome-wide significance according to the Bonferroni-corrected threshold (p-value $< 2.19 \times 10^{-7}$) in a meta-analysis including all five cohorts for these same two SNPs, the p-value of the *HHIP* SNP being even lower than 5×10^{-8} (Table 2). This indicates the robustness of our findings. Third, we did not adjust our analysis for asthma or other risk factors for impaired pulmonary function, such as environmental

tobacco smoke (ETS) and occupational exposures.^{2,14} However, when we adjusted the linear models for asthma, occupational exposures (vapors, gases, dusts and fumes, or pesticides) or for ETS exposure (hours per day or *in utero*), we found similar effect estimates for the two replicated SNPs on FEV₁/FVC in never-smokers, suggesting there are no confounding effects in our study (Table E8). These findings also indicate that these variants may underlie impaired pulmonary function independently of these other exposures as well. This may be important information for countries where airway obstruction due to smoking is less prevalent.

In conclusion, in the current genome-wide association study on continuous FEV₁/FVC levels in exclusively never-smokers we identified *HHIP* and *FAM13A* as potential airway obstruction susceptibility genes. By strengthening our GWA findings by performing additional analyses on a genetic risk score and gene expression in lung tissue (eQTL analysis), we were able to assess the effects of these common genetic variants in more detail. These two genes confer a risk for airway obstruction in general, and have an effect regardless of the most important risk factor for COPD, i.e. cigarette smoking. Therefore, these two genes might have an impact on the prevalence of COPD worldwide, especially in countries where COPD due to smoking is less prevalent. This information may ultimately contribute to the understanding of cellular and molecular pathways involved in pulmonary function impairment and the development of COPD.

Acknowledgments

We thank Rob Bieringa, Joost Keers, René Oostergo, Rosalie Visser, and Jan Schouten for their work related to data collection and validation in the LifeLines cohort study and the Vlagtwedde-Vlaardingen study. The authors would like to thank the staff at the Respiratory Health Network Tissue Bank of the FRQS for their valuable assistance with the lung eQTL dataset at Laval University. We are grateful to the study participants and all staff involved in the LifeLines cohort study, Vlagtwedde-Vlaardingen Study, Rotterdam Study and lung eQTL database. We would like to thank Anis Abuseiris, Karol Estrada, Dr. Tobias A. Knoch, and Rob de Graaf as well as their institutions Biophysical Genomics, Erasmus MC Rotterdam, The Netherlands, and especially the national German MediGRID and Services@MediGRID part of the German D-Grid, both funded by the German Bundesministerium fuer Forschung und Technology under grants #01 AK 803 A-H and # 01 IG 07015 G for access to their grid resources.

References

1. The top 10 causes of death, fact sheet N°310 [homepage on the Internet]. Updated May 2014.
2. Viegi G, Pistelli F, Sherrill DL, Maio S, Baldacci S, Carrozzi L. Definition, epidemiology and natural history of COPD. *Eur Respir J*. 2007 Nov;30(5):993-1013.
3. Halbert RJ, Natoli JL, Gano A, Badamgarav E, Buist AS, Mannino DM. Global burden of COPD: Systematic review and meta-analysis. *Eur Respir J*. 2006 Sep;28(3):523-32.
4. Standards for the diagnosis and management of patients with COPD [internet]. version 1.2. new york: American thoracic society [homepage on the Internet]. New York: American Thoracic Society. 2004 [updated 2005 September 8]. Available from: <http://www.thoracic.org/go/copd>.
5. Wilk JB, Shrine NR, Loehr LR, Zhao JH, Manichaikul A, Lopez LM, *et al*. Genome-wide association studies identify CHRNA5/3 and HTR4 in the development of airflow obstruction. *Am J Respir Crit Care Med*. 2012 Oct 1;186(7):622-32.
6. Hancock DB, Eijgelsheim M, Wilk JB, Gharib SA, Loehr LR, Marcianti KD, *et al*. Meta-analyses of genome-wide association studies identify multiple loci associated with pulmonary function. *Nat Genet*. 2010 Jan;42(1):45-52.
7. 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 Jan;42(1):36-44.
8. Pillai SG, Ge D, Zhu G, Kong X, Shianna KV, Need AC, *et al*. A genome-wide association study in chronic obstructive pulmonary disease (COPD): Identification of two major susceptibility loci. *PLoS Genet*. 2009 Mar;5(3):e1000421.
9. 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 Mar;42(3):200-2.
10. van Diemen CC, Postma DS, Vonk JM, Bruinenberg M, Nolte IM, Boezen HM. Decorin and TGF-beta1 polymorphisms and development of COPD in a general population. *Respir Res*. 2006 Jun 16;7:89.
11. Siedlinski M, Boezen HM, Boer JM, Smit HA, Postma DS. ABCC1 polymorphisms contribute to level and decline of lung function in two population-based cohorts. *Pharmacogenet Genomics*. 2009 Sep;19(9):675-84.
12. de Jong K, Boezen HM, Hacken NH, Postma DS, Vonk JM, LifeLines cohort study. GST-omega genes interact with environmental tobacco smoke on adult level of lung function. *Respir Res*. 2013 Aug 9;14:83,9921-14-83.
13. Siedlinski M, van Diemen CC, Postma DS, Vonk JM, Boezen HM. Superoxide dismutases, lung function and bronchial responsiveness in a general population. *Eur Respir J*. 2009 May;33(5):986-92.
14. Salvi SS, Barnes PJ. Chronic obstructive pulmonary disease in non-smokers. *The Lancet*. 2009;374(9691):733-43.
15. 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 Sep 25;43(11):1082-90.

16. Wain LV, Shrine N, Miller S, Jackson VE, Ntalla I, Artigas MS, *et al.* Novel insights into the genetics of smoking behaviour, lung function, and chronic obstructive pulmonary disease (UK BiLEVE): A genetic association study in UK biobank. *Lancet Respir Med.* 2015 Oct;3(10):769-81.
17. Stolk RP, Rosmalen JG, Postma DS, de Boer RA, Navis G, Slaets JP, *et al.* Universal risk factors for multifactorial diseases: LifeLines: A three-generation population-based study. *Eur J Epidemiol.* 2008;23(1):67-74.
18. van der Lende R, Rijksuniversiteit te Groningen, Gezondheidsorganisatie T.N.O. Epidemiology of chronic non-specific lung disease (chronic bronchitis). A critical analysis of three field surveys of CNSLD carried out in the netherlands [dissertation]. Assen: Van Gorcum; 1969.
19. Hofman A, Darwish Murad S, van Duijn CM, Franco OH, Goedegebure A, Ikram MA, *et al.* The rotterdam study: 2014 objectives and design update. *Eur J Epidemiol.* 2013 Nov;28(11):889-926.
20. Hofman A, Grobbee DE, de Jong PT, van den Ouweland FA. Determinants of disease and disability in the elderly: The rotterdam elderly study. *Eur J Epidemiol.* 1991 Jul;7(4):403-22.
21. Miller MR, Hankinson J, Brusasco V, Burgos F, Casaburi R, Coates A, *et al.* Standardisation of spirometry. *Eur Respir J.* 2005 Aug;26(2):319-38.
22. Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira MA, Bender D, *et al.* PLINK: A tool set for whole-genome association and population-based linkage analyses. *Am J Hum Genet.* 2007 Sep;81(3):559-575.
23. Estrada K, Abuseiris A, Grosveld FG, Uitterlinden AG, Knoch TA, Rivadeneira F. GRIMP: A web- and grid-based tool for high-speed analysis of large-scale genome-wide association using imputed data. *Bioinformatics.* 2009 Oct 15;25(20):2750-2.
24. Willer CJ, Li Y, Abecasis GR. METAL: Fast and efficient meta-analysis of genomewide association scans. *Bioinformatics.* 2010 Sep 1;26(17):2190-1.
25. IBM corp. released 2013. IBM SPSS statistics for windows, version 22.0. armonk, NY: IBM corp.
26. Hao K, Bosse Y, Nickle DC, Pare PD, Postma DS, Laviolette M, *et al.* Lung eQTLs to help reveal the molecular underpinnings of asthma. *PLoS Genet.* 2012;8(11):e1003029.
27. 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 Oct 1;184(7):786-95.
28. Wang B, Zhou H, Yang J, Xiao J, Liang B, Li D, *et al.* Association of HHIP polymorphisms with COPD and COPD-related phenotypes in a chinese han population. *Gene.* 2013 Nov 15;531(1):101-5.
29. Kim S, Kim H, Cho N, Lee SK, Han BG, Sull JW, *et al.* Identification of FAM13A gene associated with the ratio of FEV1 to FVC in korean population by genome-wide association studies including gene-environment interactions. *J Hum Genet.* 2015 Mar;60(3):139-45.
30. Chen W, Brehm JM, Manichaikul A, Cho MH, Boutaoui N, Yan Q, *et al.* A genome-wide association study of chronic obstructive pulmonary disease in hispanics. *Ann Am Thorac Soc.* 2015 Mar;12(3):340-8.

31. Zhou X, Baron RM, Hardin M, Cho MH, Zielinski J, Hawrylkiewicz I, *et al.* Identification of a chronic obstructive pulmonary disease genetic determinant that regulates HHIP. *Hum Mol Genet.* 2012 Mar 15;21(6):1325-35.
32. Zhou X, Qiu W, Sathirapongsasuti JF, Cho MH, Mancini JD, Lao T, *et al.* Gene expression analysis uncovers novel hedgehog interacting protein (HHIP) effects in human bronchial epithelial cells. *Genomics.* 2013 May;101(5):263-72.
33. Corvol H, Hodges CA, Drumm ML, Guillot L. Moving beyond genetics: Is FAM13A a major biological contributor in lung physiology and chronic lung diseases? *J Med Genet.* 2014 Oct;51(10):646-9.
34. Young RP, Hopkins RJ, Hay BA, Whittington CF, Epton MJ, Gamble GD. FAM13A locus in COPD is independently associated with lung cancer - evidence of a molecular genetic link between COPD and lung cancer. *Appl Clin Genet.* 2010 Dec 22;4:1-10.
35. Jiang Z, Lao T, Qiu W, Polverino F, Gupta K, Guo F, *et al.* A chronic obstructive pulmonary disease susceptibility gene, FAM13A, regulates protein stability of beta-catenin. *Am J Respir Crit Care Med.* 2016 Feb 10.
36. Budulac SE, Vonk JM, Postma DS, Siedlinski M, Timens W, Boezen MH. Nicotinic acetylcholine receptor variants are related to smoking habits, but not directly to COPD. *PLoS One.* 2012;7(3):e33386.
37. Saccone SF, Hinrichs AL, Saccone NL, Chase GA, Konvicka K, Madden PA, *et al.* Cholinergic nicotinic receptor genes implicated in a nicotine dependence association study targeting 348 candidate genes with 3713 SNPs. *Hum Mol Genet.* 2007 Jan 1;16(1):36-49.
38. Quanjer PH, Tammeling GJ, Cotes JE, Pedersen OF, Peslin R, Yernault JC. Lung volumes and forced ventilatory flows. *Eur Respir J.* 1993 Mar;6 Suppl 16:5-40.

Tables

Table 1. Characteristics of never-smoking subjects included in the identification and the verification cohorts

	<i>LifeLines (Identification)</i>	<i>Vlagtwedde- Vlaardingen</i>	<i>Rotterdam Study I</i>	<i>Rotterdam Study II</i>	<i>Rotterdam Study III</i>
N with no missing data	5,070	432	408	379	747
Males , N (%)	1,942 (38)	103 (23)	81 (20)	124 (33)	308 (41)
Age (yrs), median (min-max)	46 (18-89)	54 (36-79)	78 (72-94)	71 (65-98)	62 (51-93)
Height (cm), mean (SD)	174 (9)	166 (9)	163 (9)	166 (9)	171 (9)
Pulmonary function , mean (SD)					
FEV ₁ (L)	3.45 (0.84)	2.65 (0.78)	2.09 (0.60)	2.50 (0.67)	3.00 (0.77)
FEV ₁ %predicted (%) [*]	104 (14)	98 (15)	109 (22)	108 (18)	106 (17)
FEV ₁ /FVC (%) [†]	78 (7)	76 (7)	77 (7)	79 (6)	78 (6)
FEV ₁ /FVC <70%, N (%) [†]	548 (11)	77 (18)	51 (13)	27 (7)	75 (10)
GOLD COPD stage 2 <, N (%) [‡]	106 (2)	26 (6)	21 (5)	9 (2)	16 (2)

FEV₁=forced expiratory volume in one second, FVC=forced vital capacity, GOLD=Global initiative for chronic Obstructive Lung Disease

** FEV₁%predicted is based on the reference equation by Quanjer, 1993³⁸*

† FEV₁%IVC (inspiratory vital capacity) for the Vlagtwedde-Vlaardingen cohort

‡ FEV₁/FVC <70% and FEV₁%p <80% based on pre-bronchodilator measurements

548 **Table 2. SNPs associated with FEV₁/FVC (%) in the identification analysis and verification by meta-analysis of four**
 549 **independent cohorts**

SNP	Chr	A1	Annotation Nearest gene	Identification LifeLines (n=5,070)				Meta-analysis verification 4 cohorts (n=1,966)*				Meta-analysis all 5 cohorts (n=7,036)				Direction of effect in independent cohorts†
				MAF	B	SE	P	B	SE	P	I ²	B	SE	P	I ²	
rs16861512	2	C	LINC00276 (Intronic)	2%	-2.18	0.44	6.16 x 10 ⁻⁷	0.93	0.74	0.204	0.0	-1.37	0.37	2.66 x 10 ⁻⁴	70.4	-++++
rs1512282	4	G	HHIP (135Kb 5')	41%	0.60	0.12	1.23 x 10 ⁻⁶	0.55	0.21	0.008	0.0	0.58	0.11	3.15 x 10⁻⁸	0.0	+++++
rs3768948	2	C	AGAP1 (Intronic)	7%	1.07	0.23	4.32 x 10 ⁻⁶	0.26	0.37	0.487	0.0	0.84	0.20	1.94 x 10 ⁻⁵	34.4	+++--+
rs9396752	6	A	OFCC1 (Intronic)	26%	-0.64	0.14	4.34 x 10 ⁻⁶	-0.07	0.23	0.767	2.6	-0.49	0.12	3.99 x 10 ⁻⁷	46.3	-+--+
rs6849143	4	T	FAM13A (Intronic)	43%	-0.54	0.12	8.65 x 10 ⁻⁶	-0.58	0.20	0.005	0.0	-0.55	0.10	1.36 x 10⁻⁷	0.0	-----

550
 551 SNPs were selected in the identification cohort based on a p-value <10⁻⁵. Associations between SNPs and FEV₁/FVC (%) were
 552 analyzed using linear regression adjusted for sex, age and height. The results of the separate verification studies can be found in
 553 Table E4. SNPs with a p-value <Bonferroni corrected p-value of 2.19x10⁻⁷ were considered genome-wide significant. A1=minor
 554 allele (effect allele), MAF=minor allele frequency, B=effect estimate, SE=standard error and P=p-value, I² = heterogeneity
 555 measure.*Verification cohorts: Vlagtwedde-Vlaardingen and Rotterdam Study I to III. † Order: LifeLines, Vlagtwedde-Vlaardingen,
 556 and Rotterdam Study I to III. + represents a positive effect estimate and - represents a negative effect estimate.

557

558 **Table 3. Results of the gene expression analysis in lung tissue for the two replicated SNPs annotated to *HHIP* (rs1512282)**
 559 **and *FAM13A* (rs6849143)**

SNP	A1	Probe	Gene	Threshold*	Complete database (n=1,087)			Never-smokers (n=155)		
					Estimate	SE	P	Estimate	SE	P
rs1512282	G	100139086_TGI_at	<i>HHIP</i> (AK024689)	7.25×10^{-4}	-0.12	0.02	7.89×10^{-10}	-0.10	0.07	0.14
		100148028_TGI_at	<i>HHIP</i>		-0.11	0.03	1.23×10^{-5}	-0.02	0.10	0.84
		100311674_TGI_at	<i>HHIP</i>		-0.09	0.02	3.38×10^{-5}	-0.04	0.10	0.68
rs6849143	T	100148014_TGI_at	<i>TIGD2</i>	6.58×10^{-4}	0.07	0.01	6.98×10^{-11}	0.002	0.03	0.96
		100312038_TGI_at	<i>FAM13A</i>		0.09	0.02	1.99×10^{-9}	0.09	0.04	0.03
		100151492_TGI_at	<i>FAM13A</i> (AK023526)		0.10	0.02	6.88×10^{-6}	0.03	0.07	0.63
		100160946_TGI_at	<i>FAM13A</i>		0.05	0.01	4.52×10^{-4}	0.04	0.04	0.30

560
 561 *A1=minor allele (effect allele), SE=standard error and P=p-value. *Bonferroni corrected threshold was calculated as p-value=*
 562 *0.05/number probes within 4Mb window around the SNP.*

563 **Table 4. Associations of the replicated SNPs with FEV₁/FVC level stratified according to smoking status in the identification**
 564 **cohort LifeLines**

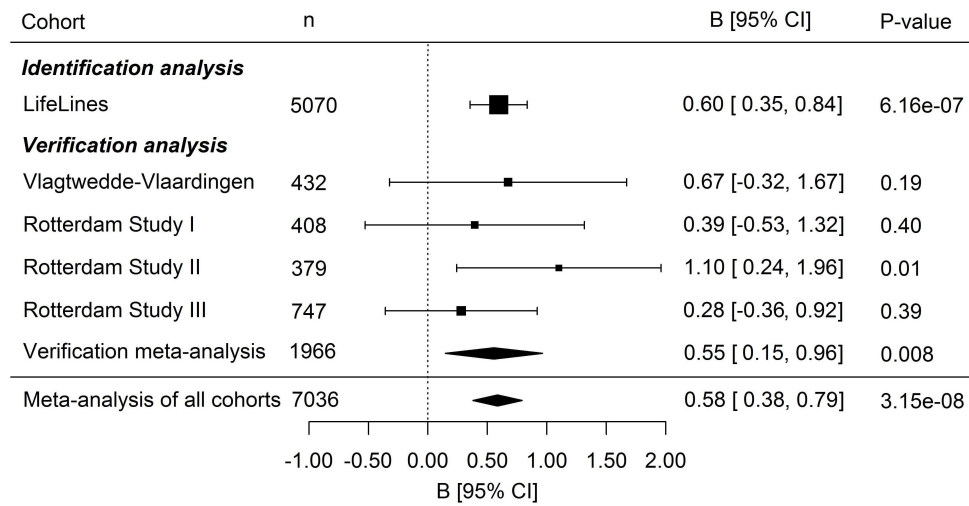
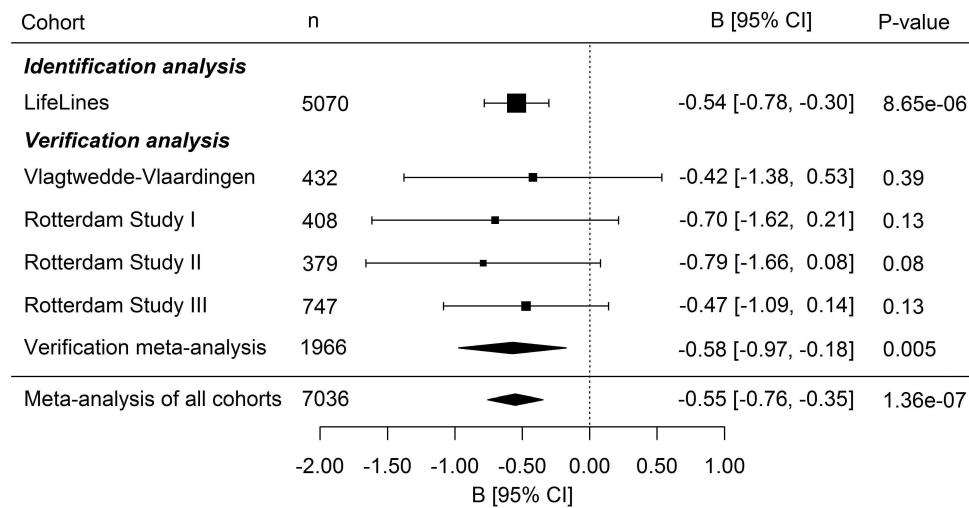
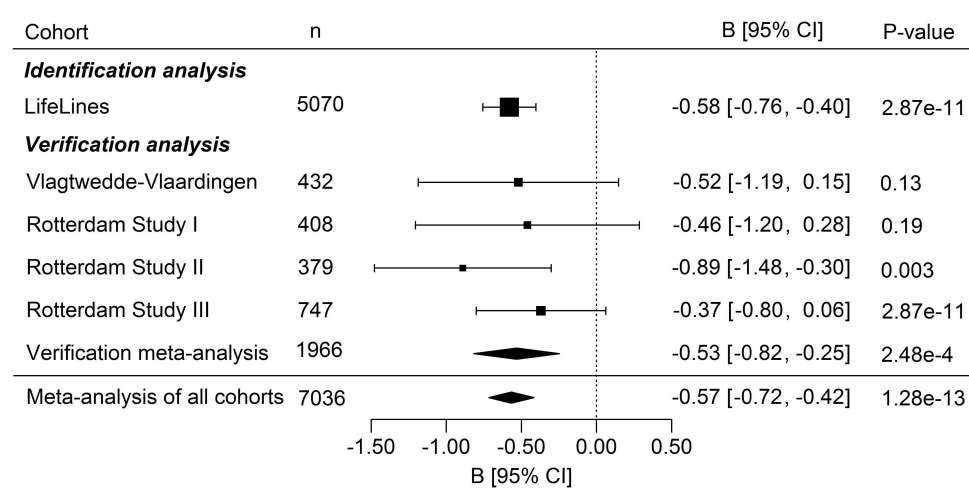
SNP	Chr	A1	Annotation Nearest gene	Never-smokers (n=5,070)			Ever-smokers (>5 py) (n=4,855)			SNP*ever-smoking interaction (n=9,925)		
				B	SE	P	B	SE	P	B	SE	P
rs1512282	4	G	HHIP (135Kb 5')	0.60	0.12	1.23 x 10 ⁻⁶	0.62	0.15	2.56 x 10 ⁻⁵	0.01	0.19	0.96
rs6849143	4	T	FAM13A (Intronic)	-0.54	0.12	8.65 x 10 ⁻⁶	-0.17	0.15	0.23	0.35	0.19	0.07

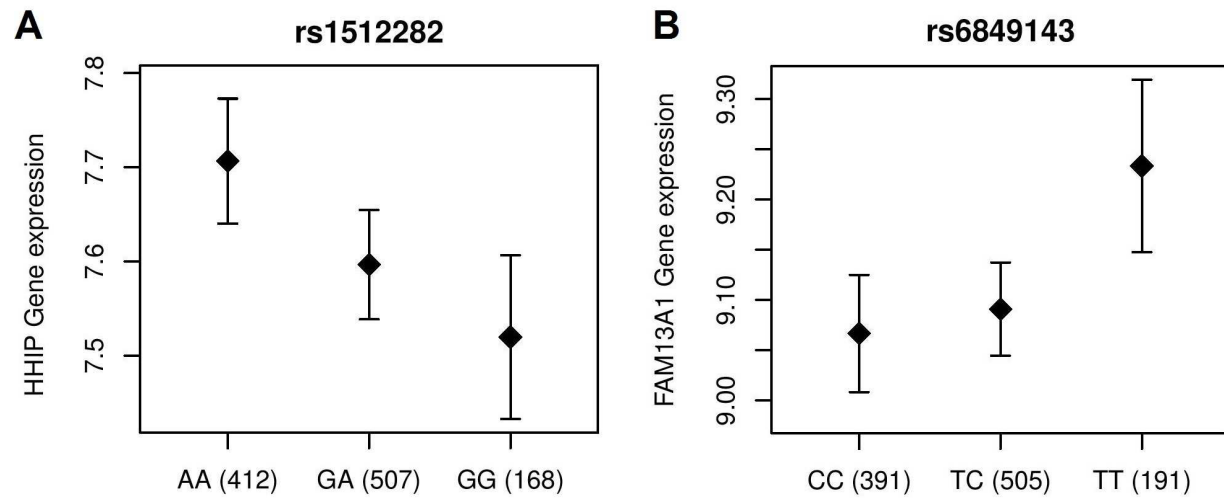
565
 566 *Linear regression models assessing associations between SNPs and FEV₁/FVC (%) were adjusted for sex, age and height in the analysis in*
 567 *never-smokers, and additionally adjusted for pack-years, and current-smoking in the analysis in ever-smokers. SNP*ever-smoking*
 568 *interactions were tested in a linear regression model adjusted for sex, age, height, pack-years and ever-smoking.*
 569

Figures legends

Figure 1. Forrest plots showing the associations of the two replicated SNPs and the genetic risk score with FEV₁/FVC levels (%) in never-smokers. A) rs1512282 located near *HHIP*; B) rs6849143 located within an intron of *FAM13A*; and C) genetic risk score (GRS), which was created by adding up the number of alleles (risk alleles) for both *HHIP* and *FAM13A* SNPs associated with a lower FEV₁/FVC level. A value of 0 means no risk alleles for both SNPs, and a value of 4 means a subjects is carrying all four risk alleles of both SNPs.. B and 95% CI represent the mean difference in % FEV₁/FVC level for each additional allele relative to the wild-type (adjusted for sex, age and height).

Figure 2. Raw data plots of gene expression levels in lung tissue according to genotype for *HHIP* (100139086_TGI_at, A) and *FAM13A* (100312038_TGI_at, B). Results were obtained from a meta-analysis using data of three cohorts included in the lung tissue database (see Supplementary material). The meta-analyzed unadjusted mean and 95% CI are plotted. The genotypes are shown on the x-axis and on the y-axis the log2 microarray intensity. A forest plot showing the adjusted data of all cohorts separately and the meta-analysis is shown in Figure E7.

A**rs1512282 (HHIP)****B****rs6849143 (FAM13A)****C****Genetic risk score**



Supplementary material

Genome-wide association study on non-smoking related airway obstruction identifies

HHIP and *FAM13A*

Authors: Diana A van der Plaat MSc^{1,2}, Kim de Jong PhD^{1,2}, Lies Lahousse PhD^{3,4}, Alen Faiz PhD^{5,2}, Judith M Vonk PhD^{1,2}, Cleo C van Diemen PhD⁶, Ivana Nedeljkovic MD³, Najaf Amin PhD³, Prof Guy G Brusselle PhD^{3,7,4}, Prof Albert Hofman MD³, Corry-Anke Brandsma PhD^{5,2}, Yohan Bossé PhD⁸, Prof Don D Sin PhD^{9,10}, David C Nickle PhD¹¹, Prof Cornelia M van Duijn PhD³, Prof Dirkje S Postma MD^{12,2}, Prof H Marika Boezen PhD^{1,2}

Author affiliation: ¹University of Groningen, University Medical Center Groningen, Department of Epidemiology, Groningen, the Netherlands. ²University of Groningen, University Medical Center Groningen, Groningen Research Institute for Asthma and COPD (GRIAC), Groningen, the Netherlands. ³Erasmus Medical Center, Department of Epidemiology, Rotterdam, the Netherlands. ⁴Ghent University Hospital, Department of Respiratory Medicine, Ghent, Belgium. ⁵University of Groningen, University Medical Center Groningen, Department of Pathology and Medical Biology, Groningen, the Netherlands. ⁶University of Groningen, University Medical Center Groningen, Department of Genetics, Groningen, the Netherlands. ⁷Erasmus Medical Center, Department of Respiratory Medicine, Rotterdam, the Netherlands. ⁸Laval University, Institut universitaire de cardiologie et de pneumologie de Québec, Department of Molecular Medicine, Québec, Canada. ⁹University of British Columbia, Department of Medicine, Division of Respiratory Medicine, Vancouver, Canada. ¹⁰The University of British Columbia Center for Heart Lung Innovation, St Paul's Hospital, Vancouver, Canada. ¹¹Merck & Co. MRL, Seattle, WA, United States of America. ¹²University of Groningen, University Medical Center Groningen, Department of Pulmonary Diseases, Groningen, the Netherlands.

Supplementary Methods

Cohorts and pulmonary measurements

The LifeLines cohort study is a multi-disciplinary prospective population-based cohort study of the Northern provinces of the Netherlands. It was specifically designed to study the role of genes, exposures and their interaction in the onset of common complex diseases like COPD. All subjects provided written informed consent. Spirometry was measured using a Welch Allyn Version 1.6.0.489, PC-based SpiroPerfect with Ca Workstation software.¹

The Vlagtwedde-Vlaardingen study was the first population-based prospective cohort study in the Netherlands, aiming to gain more insight in chronic airway diseases. Subjects were included from Vlagtwedde (a rural area) and Vlaardingen (an urban area) and were followed for 25 years. Written informed consent was provided by all participants and the study protocol was approved by the local university medical hospital ethics committee, University of Groningen, University Medical Center Groningen, The Netherlands. For the current study we used questionnaire data, spirometry, and DNA isolated from neutrophil depots in peripheral blood samples drawn during the last survey in 1989/1990. FEV₁ as a percentage of inspiratory vital capacity (IVC) was used instead of FEV₁/FVC and data were collected by performing a slow inspiratory maneuver, using a water-sealed spirometer (Lode instruments, Groningen, the Netherlands).²

Both the LifeLines cohort study and the Vlagtwedde-Vlaardingen studies were approved by the Medical Ethics Committee of the University Medical Center Groningen, Groningen, the Netherlands.

The Rotterdam elderly study is a prospective population-based study investigating chronic diseases in elderly living in the Ommoord district in the city of Rotterdam in the Netherlands. Subject aged 55 years and older were recruited starting in 1990 (RS I). In 2000, new subjects of 55 years and older were recruited for RS II, and in 2006, a third expansion (RS III) followed by recruiting subjects of 45 years and older. All subjects provided written informed consent and the study was approved by the medical ethics committee of Erasmus University. Spirometric data were collected up to 2009 using a SpiroPro portable spirometer (Erich Jaeger, Hoechberg, Germany), and since 2009 using the Jaeger Masterscreen PFT (Care Fusion, the Netherlands).^{3,4}

In all cohorts, pre-bronchodilator spirometric measurements (FEV₁, FEV₁/FVC) were performed according to the ATS/ERS criteria.⁵

Gene expression analysis (eQTL analysis)

We assessed if the selected SNPs were cis-acting expression quantitative trait loci (eQTL) in a lung tissue database established by the lung eQTL consortium, including the University of Groningen (GRN), Laval University (Laval) and University of British Columbia (UBC).⁶ Lung tissue samples were obtained from patients during their lung resection surgery at the three sites. Written informed consent was provided by all patients. The ethics committees of the Institut universitaire de cardiologie et de pneumologie de Québec (Laval) and the UBC-Providence Health Care Research Institute Ethics Board (UBC) approved the study, and the study protocol was consistent with the Research Code of the University Medical Center Groningen and Dutch national ethical and professional guidelines.

Genetic profiles were obtained using Illumina Human1M-Duo BeadChip arrays and gene expression profiles were obtained using a custom Affymetrix array (GEO accession number GPL10379 and GSE23546). In total 1087 subjects with no missing data were included in the analysis, and probes located within 2 Mb (in both directions) of the replicated SNPs were selected (4 MB window). First, residuals from linear regression models on log₂ transformed gene expression levels (of each probe separately) adjusted for disease status (alpha-1 antitrypsin deficiency, idiopathic pulmonary fibrosis, pulmonary hypertension, cystic fibrosis, other disease), age, gender and smoking status (never/ever/unknown) were used to calculate cohort specific (GRN, Laval and UBC) principal components (PCs). 14 PCs for GRN and Laval, and 16 for UBC were included in the main analysis, which explained at least one percent of the total variance. Second, associations between the SNPs and log₂ transformed gene expression levels (mRNA level) were tested in each cohort separately using linear regression (additive genetic model) adjusted for disease status, age, gender, smoking status and a cohort specific number of PCs. Third, effect estimates for the three cohorts were meta-analyzed using fixed effects models with effect estimates weighted by the reciprocal of the estimated variance. SNPs with a p-value below the Bonferroni corrected threshold ($p=0.05/\text{number of probes in 4Mb window}$) in the meta-analysis were considered significantly associated with gene expression level.

Supplementary references

1. Stolk RP, Rosmalen JG, Postma DS, de Boer RA, Navis G, Slaets JP, *et al.* Universal risk factors for multifactorial diseases: LifeLines: A three-generation population-based study. *Eur J Epidemiol.* 2008;23(1):67-74.
2. van der Lende R, Rijksuniversiteit te Groningen, Gezondheidsorganisatie T.N.O. Epidemiology of chronic non-specific lung disease (chronic bronchitis). A critical analysis of three field surveys of CNSLD carried out in the netherlands [dissertation]. Assen: Van Gorcum; 1969.
3. Hofman A, Darwish Murad S, van Duijn CM, Franco OH, Goedegebure A, Ikram MA, *et al.* The rotterdam study: 2014 objectives and design update. *Eur J Epidemiol.* 2013 Nov;28(11):889-926.
4. Hofman A, Grobbee DE, de Jong PT, van den Ouweland FA. Determinants of disease and disability in the elderly: The rotterdam elderly study. *Eur J Epidemiol.* 1991 Jul;7(4):403-22.
5. Miller MR, Hankinson J, Brusasco V, Burgos F, Casaburi R, Coates A, *et al.* Standardisation of spirometry. *Eur Respir J.* 2005 Aug;26(2):319-38.
6. Hao K, Bosse Y, Nickle DC, Pare PD, Postma DS, Laviolette M, *et al.* Lung eQTLs to help reveal the molecular underpinnings of asthma. *PLoS Genet.* 2012;8(11):e1003029.
7. Quanjer PH, Tammeling GJ, Cotes JE, Pedersen OF, Peslin R, Yernault JC. Lung volumes and forced ventilatory flows. *Eur Respir J.* 1993 Mar;6 Suppl 16:5-40.
8. Pillai SG, Ge D, Zhu G, Kong X, Shianna KV, Need AC, *et al.* A genome-wide association study in chronic obstructive pulmonary disease (COPD): Identification of two major susceptibility loci. *PLoS Genet.* 2009 Mar;5(3):e1000421.
9. Cho MH, McDonald ML, Zhou X, Mattheisen M, Castaldi PJ, Hersh CP, *et al.* Risk loci for chronic obstructive pulmonary disease: A genome-wide association study and meta-analysis. *Lancet Respir Med.* 2014 Mar;2(3):214-25.
10. Pruim RJ, Welch RP, Sanna S, Teslovich TM, Chines PS, Gliedt TP, *et al.* LocusZoom: Regional visualization of genome-wide association scan results. *Bioinformatics.* 2010 Sep 15;26(18):2336-7.

Supplementary Tables

Table E1. Descriptive characteristics of the never- and ever-smoking subjects of the LifeLines cohort study

	<i>Never-smokers (0 py)</i>	<i>Ever-smokers (>5 py)</i>
N with non-missing data	5,070	4,855
Males , N (%)	1,942 (38)	2,312 (48)
Age (yrs), median (min-max)	46 (18-89)	49 (22-85)
Height (cm), mean (sd)	174 (9)	175 (9)
Pulmonary function , mean (sd)		
FEV ₁ (L)	3.45 (0.84)	3.34 (0.82)
FEV ₁ %predicted (%)†	104 (14)	100 (15)
FEV ₁ /FVC (%)	78 (7)	75 (8)
FEV ₁ /FVC <70%, N (%)	548 (11)	1,107 (23)
GOLD COPD stage 2<, N (%)‡	106 (2)	310 (6)
Smoking Status , N (%)		
Never-smoker	5,070 (100)	0 (0)
Ever-smoker (>5 py)	0 (0)	4,855 (100)
Ex-smoker	0 (0)	2,684 (55)
Current-smoker	0 (0)	2,171 (45)
Pack-years, median (min-max)	0 (0-0)	14.3 (5-100)

FEV₁=forced expiratory volume in one second, FVC=forced vital capacity,
 GOLD=Global initiative for chronic Obstructive Lung Disease, py=pack-
 years

† FEV₁%predicted calculation based on reference equation by Quanjer,
 1993⁷

‡ FEV₁/FVC <70% and FEV₁%p <80%

Table E2. The SNP associated with FEV₁ (mL) (P-value <10⁻⁵) in the identification analysis and verification in the independent Vlagtwedde-Vlaardingen cohort

SNP	Chr	A1	Annotation	Identification LifeLines (n=5,070)				Vlagtwedde-Vlaardingen (n=432)			
				MAF	B	SE	P	MAF	B	SE	P
rs17057154	4	C	RP11-97E7.2 (Exonic)	16%	58	12	9.37 x 10 ⁻⁷	14%	-76	37	0.04

B=effect estimate, SE=standard error and P=p-value.

Table E3. Associations between the five identified SNPs and levels of FEV₁/FVC (%), FEV₁ (mL) and FVC (mL) separately, in never-smokers of LifeLines

SNP	Chr	A1	Annotation	FEV ₁ /FVC (%)			FEV ₁ (mL)			FVC (mL)		
				B	SE	P	B	SE	P	B	SE	P
rs16861512	2	C	LINC00276	-2.18	0.44	6.16x10 ⁻⁷	-90	32	4.40x10 ⁻³	14	38	0.72
rs1512282	4	G	HHIP	0.60	0.12	1.23x10 ⁻⁶	29	9	1.16x10 ⁻³	2	11	0.88
rs3768948	2	C	AGAP1	1.07	0.23	4.32x10 ⁻⁶	31	17	0.07	-23	20	0.25
rs9396752	6	A	OFCC1	-0.64	0.14	4.34x10 ⁻⁶	-31	10	2.19x10 ⁻³	-5	12	0.68
rs6849143	4	T	FAM13A	-0.54	0.12	8.65x10 ⁻⁶	-2	9	0.78	30	11	5.05x10 ⁻³

All analyses were adjusted for sex, age and height.

144 **Table E4. Associations between SNPs and FEV₁/FVC (%) in never-smokers in the separate verification cohorts**

SNP	Chr	A1	Annotation Nearest gene	Vlagtwedde-Vlaardingen (n=432)				Rotterdam Study I (n=408)				Rotterdam Study II (n=379)				Rotterdam Study III (n=747)			
				%	B	SE	P	%	B	SE	P	%	B	SE	P	%	B	SE	P
rs16861512	2	C	LINC00276 (Intronic)	2	0.50	1.78	0.78	2	0.51	1.50	0.74	2	1.35	1.39	0.33	2	1.13	1.33	0.39
rs1512282	4	G	HHIP (135Kb 5')	41	0.67	0.51	0.19	40	0.39	0.47	0.40	41	1.10	0.44	0.01	41	0.28	0.33	0.39
rs3768948	2	C	AGAP1 (Intronic)	9	1.29	0.83	0.12	8	0.17	0.79	0.83	7	-0.78	0.99	0.43	8	0.16	0.56	0.77
rs9396752	6	A	OFCC1 (Intronic)	23	0.02	0.54	0.97	25	-0.65	0.54	0.23	25	0.61	0.51	0.23	24	-0.19	0.36	0.60
rs6849143	4	T	FAM13A (Intronic)	44	-0.42	0.49	0.39	43	-0.70	0.47	0.13	44	-0.79	0.44	0.08	44	-0.47	0.31	0.13

145
146 % = minor allele frequency (MAF), B = effect estimate, SE = standard error and P = p-value.
147
148

Table E5. Results of the SNP*ever-smoking interaction using gene expression data obtained from lung tissue for the two replicated SNPs annotated to *HHIP* (rs1512282) and *FAM13A* (rs6849143)

SNP	A1	Probe	Annotation	Estimate	SE	P
rs1512282	G	100139086_TGI_at	AK024689 (<i>HHIP</i>)	-0.05	0.06	0.41
		100148028_TGI_at	<i>HHIP</i>	-0.11	0.09	0.20
		100311674_TGI_at	<i>HHIP</i>	-0.08	0.08	0.27
rs6849143	T	100148014_TGI_at	<i>TIGD2</i>	0.03	0.03	0.35
		100312038_TGI_at	<i>FAM13A</i>	-0.01	0.05	0.87
		100151492_TGI_at	AK023526 (<i>FAM13A</i>)	0.03	0.07	0.65
		100160946_TGI_at	<i>FAM13A</i>	0.00	0.04	0.96

A1=minor allele (effect allele), SE = standard error, P=p-value and n=1,087.

Table E6. Associations between two SNPs in the *CHRNA* locus and FEV₁/FVC level stratified according to smoking status in the identification cohort LifeLines

SNP	Chr	A1	Annotation nearest gene	Never-smokers (n=5,070)			Ever-smokers (>5 py) (n=4,855)			SNP*ever-smoking interaction (n=9,925)		
				B	SE	P	B	SE	P	B	SE	P
rs8034191 ⁸	15	C	<i>IREB2</i>	0.06	0.13	0.63	-0.50	0.16	1.40x10 ⁻³	-0.53	0.20	8.81x10 ⁻³
rs12914385 ⁹	15	T	<i>CHRNA3</i>	0.17	0.13	0.18	-0.31	0.15	3.90x10 ⁻²	-0.46	0.20	1.93x10 ⁻²

LD of the SNPs $r^2=0.68$, $D'=0.88$. Linear regression models assessing associations between SNPs and FEV₁/FVC (%) were adjusted for sex, age and height in the analysis in never-smokers, and additionally adjusted for pack-years, and current-smoking in the analysis in ever-smokers. SNP*ever-smoking interactions were tested in a linear regression model adjusted for sex, age, height, pack-years and ever-smoking.

Table E7. Associations with FEV₁/FVC in never-smokers of genes previously identified in GWA studies on FEV₁/FVC.

Gene	Chr	SNP	B	SE	P
MFAP2	1	rs9435732	0.49	0.15	9.87E-04
TGFB2	1	rs1418555	-0.38	0.14	6.32E-03
PID1	2	rs4321351	0.32	0.13	1.51E-02
DNER	2	rs6436859	0.31	0.15	3.78E-02
HDAC4	2	rs3791480	0.36	0.15	1.42E-02
RARB	3	rs1881708	0.22	0.13	9.65E-02
SPATA9	5	rs169883	0.28	0.12	2.17E-02
HTR4	5	rs11168048	0.26	0.12	3.30E-02
ADAM19	5	rs3734031	-0.31	0.20	1.28E-01
NCR3	6	rs2857595	0.02	0.14	8.84E-01
AGER	6	rs2856437	0.65	0.32	4.47E-02
HLA-DQB1, HLA-DQA2	6	rs4947342	-0.40	0.15	8.22E-03
ARMC2	6	rs11153132	-0.38	0.13	2.67E-03
GPR126	6	rs1319584	-0.18	0.12	1.48E-01
PTCH1	9	rs16912133	0.23	0.12	6.22E-02
CDC123	10	rs10906097	-0.16	0.12	1.87E-01
LRP1	12	rs10783815	-0.22	0.13	7.83E-02
CCDC38	12	rs11610816	-0.35	0.13	8.45E-03
THSD4	15	rs11632680	0.47	0.13	2.41E-04
MMP15	16	rs12929002	-0.23	0.14	1.18E-01
CFDP1	16	rs10514393	0.23	0.12	6.87E-02
KCNE2	21	rs4817646	0.35	0.19	6.02E-02

We used the "GWAS catalog" from United States' National Human Genome Research Institute (<http://www.ebi.ac.uk/gwas/>) to select genome-wide significant SNPs (p -value $< 5 \times 10^{-8}$) from previously published genome wide association studies (GWAS) on FEV₁/FVC. For each of the SNPs we used the annotated genes and looked in our own data for a SNP in this gene region with the lowest p -value in the association with FEV₁/FVC in our never-smokers cohort.

Table E8. Linear regression models showing associations of *HHIP* (rs1512282) and *FAM13A* (rs6849143) with FEV₁/FVC (%) in never-smokers of the identification cohort, in the basic model and with additional adjustments for asthma, occupational exposures (VGDFs or pesticides) or for ETS (hours per day or *in utero*).

<i>HHIP</i>				<i>FAM13A</i>			
Basic model	B	SE	P	Basic model	B	SE	P
Constant	110.48	2.79	7.91E-299	Constant	111.69	2.79	2.46E-304
Sex	-0.85	0.25	7.14E-4	Sex	-0.86	0.25	6.77E-4
Age (yrs)	-0.23	0.01	1.25E-171	Age (yrs)	-0.23	0.01	5.25E-174
Height (cm)	-0.12	0.01	8.49E-19	Height (cm)	-0.12	0.01	4.89E-19
rs1512282	0.60	0.12	1.23E-6	rs6849143	-0.54	0.12	8.65E-6
Model + Asthma	B	SE	P	Model + Asthma	B	SE	P
Constant	111.79	2.73	0.00E+00	Constant	112.94	2.73	0.00E+00
Sex	-0.93	0.25	1.80E-4	Sex	-0.93	0.25	1.66E-4
Age (yrs)	-0.23	0.01	2.01E-175	Age (yrs)	-0.24	0.01	61.57E-177
Height (cm)	-0.12	0.01	1.79E-20	Height (cm)	-0.12	0.01	9.92E-21
rs1512282	0.56	0.12	2.59E-6	rs6849143	-0.52	0.12	1.34E-5
Asthma	-4.42	0.30	7.01E-48	Asthma	-4.40	0.30	1.98E-47
Model + VGDF	B	SE	P	Model + VGDF	B	SE	P
Constant	111.22	2.81	3.12E-299	Constant	112.42	2.81	1.33E-304
Sex	-1.02	0.26	1.01E-4	Sex	-1.01	0.26	1.08E-4
Age (yrs)	-0.23	0.01	1.16E-171	Age (yrs)	-0.23	0.01	4.31E-174
Height (cm)	-0.12	0.01	2.62E-19	Height (cm)	-0.12	0.01	1.51E-19
rs1512282	0.60	0.12	9.60E-7	rs6849143	-0.55	0.12	8.34E-6
VGDF, Low	-0.12	0.19	5.30E-1	VGDF, Low	-0.14	0.19	4.70E-1
VGDF, high	-0.76	0.30	1.12E-2	VGDF, high	-0.73	0.30	1.45E-2
Model + Pesticides	B	SE	P	Model + Pesticides	B	SE	P
Constant	110.61	2.79	2.15E-299	Constant	111.82	2.79	6.77E-305
Sex	-0.91	0.25	3.19E-4	Sex	-0.92	0.25	3.12E-4
Age (yrs)	-0.23	0.01	2.64E-170	Age (yrs)	-0.23	0.01	9.61E-173
Height (cm)	-0.12	0.01	7.46E-19	Height (cm)	-0.12	0.01	4.26E-19
rs1512282	0.60	0.12	9.74E-7	rs6849143	-0.55	0.12	8.36E-6
Pesticides, Low	-0.80	0.45	7.78E-2	Pesticides, Low	-0.72	0.45	1.11E-1
Pesticides, High	-1.00	0.83	2.30E-1	Pesticides, High	-1.06	0.83	2.03E-1
Model + ETS	B	SE	P	Model + ETS	B	SE	P
Constant	111.18	2.93	4.40E-274	Constant	112.28	2.93	8.91E-279
Sex	-0.91	0.27	6.20E-4	Sex	-0.90	0.27	6.96E-4
Age (yrs)	-0.23	0.01	1.18E-145	Age (yrs)	-0.23	0.01	2.17E-147
Height (cm)	-0.12	0.01	3.56E-18	Height (cm)	-0.12	0.01	3.39E-18
rs1512282	0.61	0.13	1.89E-6	rs6849143	-0.61	0.13	2.46E-6
ETS (hours/day)	0.03	0.26	9.22E-1	ETS (hours/day)	0.02	0.26	9.37E-1
Model + <i>In utero</i> ETS	B	SE	P	Model + <i>In utero</i> ETS	B	SE	P
Constant	110.88	2.82	4.49E-294	Constant	112.15	2.82	6.57E-300
Sex	-0.85	0.25	7.93E-4	Sex	-0.86	0.25	7.01E-4
Age (yrs)	-0.23	0.01	6.08E-157	Age (yrs)	-0.23	0.01	3.11E-159
Height (cm)	-0.12	0.01	8.64E-19	Height (cm)	-0.12	0.01	4.49E-19
rs1512282	0.59	0.12	1.65E-6	rs6849143	-0.58	0.12	2.07E-6
<i>In utero</i> ETS, yes	-0.60	0.29	3.38E-2	<i>In utero</i> ETS, yes	-0.62	0.29	3.02E-2
<i>In utero</i> ETS, unknown	-0.20	0.25	4.31E-1	<i>In utero</i> ETS, unknown	-0.22	0.25	3.73E-1

<i>HHIP</i>				<i>FAM13A</i>			
Full model	B	SE	P	Full model	B	SE	P
Constant	113.35	2.91	5.18E-286	Constant	114.43	2.91	2.12E-290
Sex	-1.03	0.27	1.31E-04	Sex	-1.02	0.27	1.58E-04
Age (yrs)	-0.24	0.01	8.10E-152	Age (yrs)	-0.24	0.01	1.33E-153
Height (cm)	-0.13	0.01	1.95E-20	Height (cm)	-0.13	0.01	2.09E-20
rs1512282	0.64	0.13	3.26E-07	rs6849143	-0.57	0.13	6.78E-06
Asthma	-4.55	0.32	1.15E-44	Asthma	-4.52	0.32	4.47E-44
VGDF, Low	-0.11	0.20	5.87E-01	VGDF, Low	-0.12	0.20	5.36E-01
VGDF, high	-0.46	0.37	2.21E-01	VGDF, high	-0.45	0.37	2.26E-01
Pesticides, Low	-0.45	0.54	4.07E-01	Pesticides, Low	-0.35	0.54	5.12E-01
Pesticides, High	0.52	0.94	5.85E-01	Pesticides, High	0.42	0.94	6.57E-01
ETS (hours/day)	0.27	0.26	2.87E-01	ETS (hours/day)	0.27	0.26	2.86E-01
<i>In utero</i> ETS, yes	-0.62	0.29	3.35E-02	<i>In utero</i> ETS, yes	-0.64	0.29	2.90E-02
<i>In utero</i> ETS, unknown	-0.16	0.26	5.26E-01	<i>In utero</i> ETS, unknown	-0.19	0.26	4.54E-01

Asthma=doctor diagnosed or having both two or more symptoms and using asthma medication. VGDF=Vapors, Gases, Dusts, or Fumes and ETS=Environmental Tobacco Smoke.

Supplementary figures legends

Figure E1. Flow chart of the study design. We performed GWA studies on FEV₁ and FEV₁/FVC in never-smokers of the LifeLines cohort study. SNPs with a p-value <10⁻⁵ were selected for verification. SNPs with a fixed effects p-value <0.05 in the verification meta-analysis of four cohorts (effect estimates weighted by the inverse of the SE) were further investigated. Additional analyses included genetic risk score, expression quantitative trait loci (eQTL) and pathway analyses. Finally, we investigated the associations in ever-smokers and performed a SNP*ever-smoking interaction analysis in the identification cohort.

Figure E2. Q-Q plot showing the distribution of observed p-values relative to the expected p-values from the analysis on FEV₁ (A) and FEV₁/FVC (B). Both plots show an estimated inflation factor (λ) of 1.02, indicating no significant overall effect of population stratification.

Figure E3. Manhattan plot showing associations between SNPs and levels of FEV₁ (A) or FEV₁/FVC (B) in never-smokers of the identification cohort. The plot shows the chromosomal position of the SNPs on the horizontal axis and the corresponding (negative log₁₀) p-values on the vertical axis. The line represents a p-value of 10⁻⁵ (selection p-value). SNPs selected for verification are indicated with a diamond.

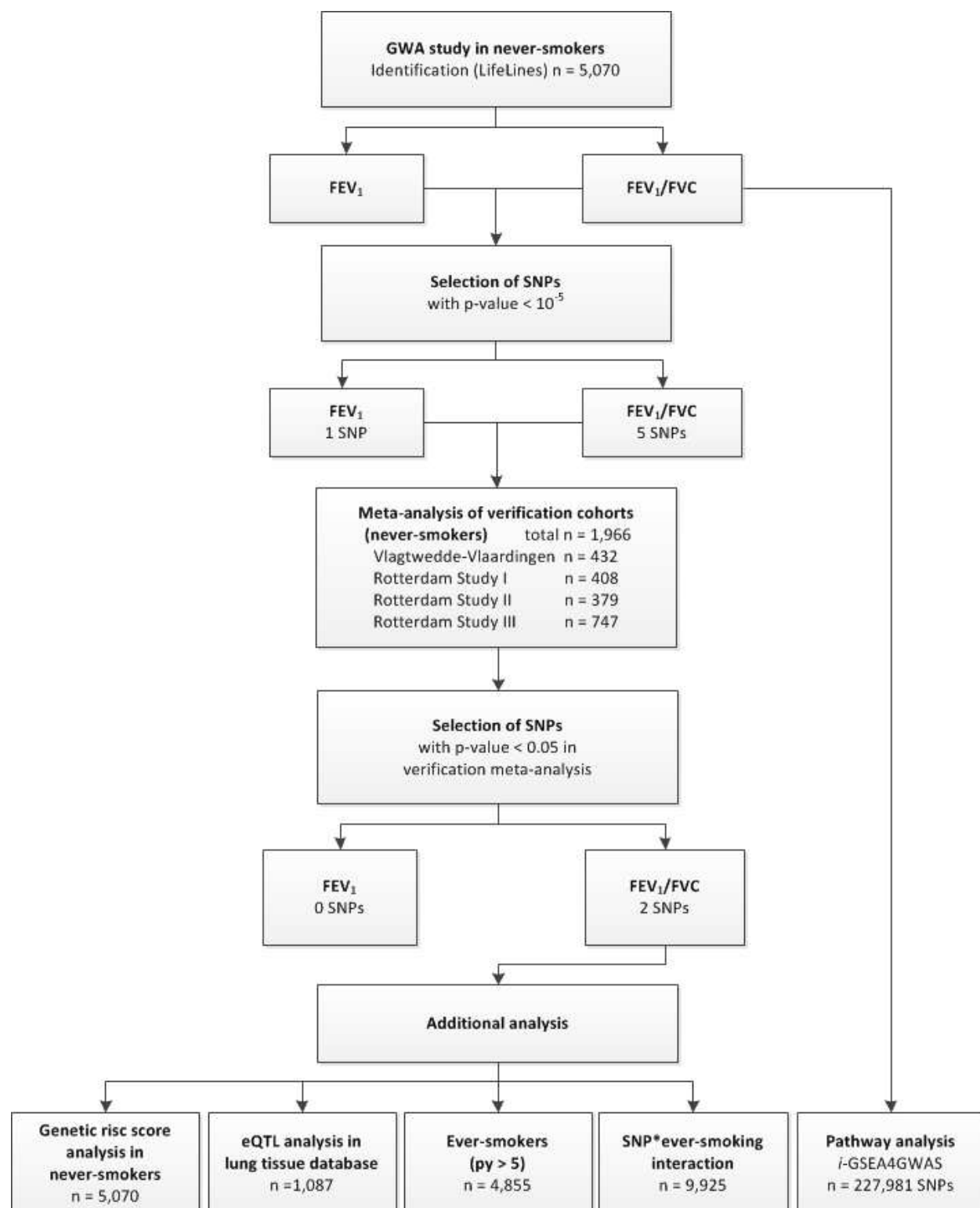
Figure E4. Regional association plots showing the p-values of the associations between SNPs and FEV₁/FVC level in never-smokers of the identification cohort, for A) *HHIP* (rs1512282) and B) *FAM13A* (rs6849143). The plots were created using LocusZoom.¹⁰ The x-axis represents the megabase (Mb) position on chromosome 4 and the y-axis, the negative log₁₀ of the p-values. The linkage disequilibrium (r^2) of other SNPs with the selected SNPs (purple diamond) is illustrated by different colors according to the inset legend.

Figure E5. Associations between SNPs genotyping groups and FEV₁/FVC level in the identification cohort. The model was adjusted for sex, age and height. A) rs1512282 (*HHIP*); B) rs6849143 (*FAM13A*). For both SNPs we can confirm an additive genetic effect.

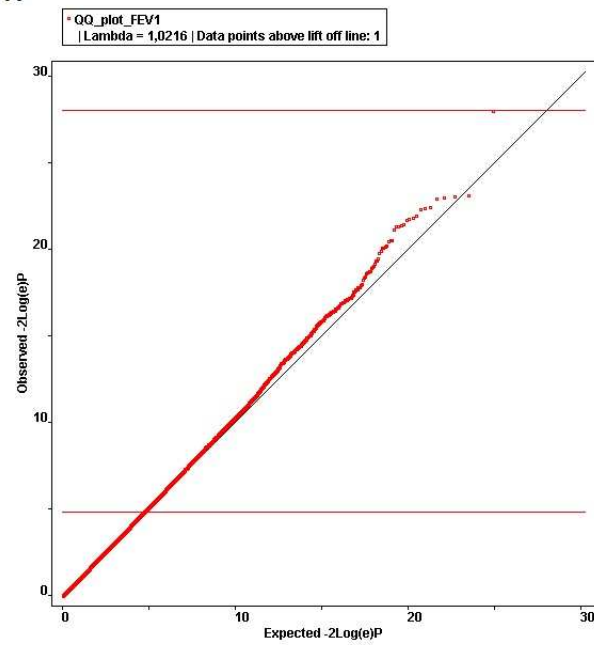
Figure E6. Genetic risk score (GRS) analysis of the risk alleles in *FAM13A* and *HHIP* on FEV₁/FVC level in the identification cohort. Genetic risk scores were created by adding

up the number of alleles (risk alleles) for both *HHIP* and *FAM13A* SNPs associated with a lower FEV₁/FVC level. A value of 0 means no risk alleles for both SNPs, and a value of 4 means a subjects is carrying all four risk alleles of both SNPs. The mean FEV₁/FVC levels and 95% CI are given for each GRS based on a median age and a mean height.

Figure E7. Forrest plots showing the associations of the *HHIP* SNP (rs1512282, A) and *FAM13A* SNP (rs6849143, B) with gene expression levels in lung tissue (lung eQTL analysis). For each significantly associated probe, the adjusted mean residuals with 95% CI are plotted of the three cohorts separately and of the meta-analysis.



A



B

