Genome-Wide Joint Meta-Analysis of SNP and SNP-by-Smoking Interaction Identifies Novel Loci for Pulmonary Function

The Harvard community has made this article openly available. Please share how this access benefits you. Your story matters

Citation

Published Version
doi:10.1371/journal.pgen.1003098

Citable link
http://nrs.harvard.edu/urn-3:HUL.InstRepos:10612881

Terms of Use
This article was downloaded from Harvard University’s DASH repository, and is made available under the terms and conditions applicable to Other Posted Material, as set forth at http://nrs.harvard.edu/urn-3:HUL.InstRepos:dash.current.terms-of-use#LAA
Genome-Wide Joint Meta-Analysis of SNP and SNP-by-Smoking Interaction Identifies Novel Loci for Pulmonary Function


1 Behavioral Health Epidemiology Program, Research Triangle Institute International, Research Triangle Park, North Carolina, United States of America, 2 Program in Molecular and Genetic Epidemiology, King’s College London, London, United Kingdom, 3 Science of Genomics, National Institute on Aging, National Institutes of Health, Bethesda, Maryland, United States of America, 4 Vermont Lung Center, University of Vermont, Burlington, Vermont, United States of America, 5 Department of Biostatistics and Bioinformatics, Duke University, Durham, North Carolina, United States of America, 6 Department of Geriatrics, University of Washington, Seattle, Washington, United States of America, 7 Program in Molecular and Genetic Epidemiology, University of Iowa, Iowa City, Iowa, United States of America, 8 Department of Clinical Genetics, University of Copenhagen, Copenhagen, Denmark, 9 Department of Clinical Genetics, King’s College London, London, United Kingdom, 10 Department of Epidemiology and Biostatistics, Imperial College London, London, United Kingdom, 11 Department of Biostatistics, Brown University, Providence, Rhode Island, United States of America, 12 Department of Medical and Molecular Genetics, University of Groningen, University Medical Center Groningen, Groningen, The Netherlands, 13 Division of Pulmonary, Critical Care and Sleep Medicine, University of Washington, Seattle, Washington, United States of America, 14 Department of Epidemiology, Erasmus Medical Center, Rotterdam, The Netherlands, 15 Department of Internal Medicine B, University Hospital Greifswald, Greifswald, Germany, 16 Department of Medical Sciences, Respiratory Medicine, Uppsala University, Uppsala, Sweden, 17 Program in Molecular and Genetic Epidemiology, University of North Carolina at Chapel Hill, Chapel Hill, North Carolina, United States of America, 18 Department of Biostatistics, University of Washington, Seattle, Washington, United States of America, 19 Department of Genetics, Washington University, St. Louis, Missouri, United States of America, 20 Division of Pulmonary and Critical Care Medicine, Department of Medicine, Veterans Administration Boston Healthcare System, Boston, Massachusetts, United States of America, 21 Department of Epidemiology and Biostatistics, School of Public Health, Imperial College London, London, United Kingdom, 22 Department of Biostatistical Sciences, Division of Public Health Sciences, Wake Forest School of Medicine, Winston-Salem, North Carolina, United States of America, 23 Department of Biostatistics, Harvard School of Public Health, Boston, Massachusetts, United States of America, 24 Department of Epidemiology, Harvard School of Public Health, Boston, Massachusetts, United States of America, 25 Department of Epidemiology and Biostatistics, Harvard School of Public Health, Boston, Massachusetts, United States of America, 26 Department of Biostatistics, Harvard School of Public Health, Boston, Massachusetts, United States of America, 27 Division of Pulmonary and Critical Care Medicine, Department of Medicine, University of Washington, Seattle, Washington, United States of America, 28 Department of Biostatistics, Harvard School of Public Health, Boston, Massachusetts, United States of America, 29 Department of Medicine, University of Washington, Seattle, Washington, United States of America, 30 Department of Medicine, Boston University School of Medicine, Boston, Massachusetts, United States of America, 31 Department of Respiratory Medicine, Ghent University Hospital, Ghent, Belgium, 32 Section of Pulmonary and Critical Care Medicine, Department of Medicine, University of California, Los Angeles, Los Angeles, California, United States of America, 33 Department of Pulmonary, Critical Care and Sleep Medicine, University of Washington, Seattle, Washington, United States of America, 34 Program in Medical and Population Genetics, Broad Institute, Cambridge, Massachusetts, United States of America, 35 Institute of Epidemiology, Helmholtz Zentrum München, Munich, Germany, 49 Medical School, Cedars-Sinai Medical Center, Los Angeles, California, United States of America, 60 Department of Medical Sciences, Respiratory Medicine, Uppsala University, Uppsala, Sweden, 36 Program in Medical and Population Genetics, Broad Institute, Cambridge, Massachusetts, United States of America, 37 Department of Pulmonary and Critical Care Medicine, Department of Medicine, University of California, Los Angeles, Los Angeles, California, United States of America, 38 Department of Molecular Biology, Massachusetts General Hospital, Boston, Massachusetts, United States of America, 39 Department of Medicine, University of California, Los Angeles, Los Angeles, California, United States of America, 40 Department of Genetics, Harvard Medical School, Boston, Massachusetts, United States of America, 41 Department of Pulmonology, University of Groningen, University Medical Center Groningen, Groningen, The Netherlands, 42 GRIAC Research Institute, University Medical Center Groningen, Groningen, The Netherlands, 43 Department of Thoracic Surgery and Division of Pulmonology, Vanderbilt University Medical Center, Nashville, Tennessee, United States of America, 44 Department of Radiology, Brigham and Women’s Hospital, Boston, Massachusetts, United States of America, 45 Division of Pulmonary and Critical Care Medicine, Department of Medicine, Cedars-Sinai Medical Center, Los Angeles, California, United States of America, 46 The Pulmonary Center, Department of Medicine, Boston University School of Medicine, Boston, Massachusetts, United States of America, 47 Institute of Epidemiology I, Helmholtz Zentrum München, Munich, Germany, 48 Department of Genetics, Cleveland Clinic, Cleveland, Ohio, United States of America.
Abstract

Genome-wide association studies have identified numerous genetic loci for spirometric measures of pulmonary function, forced expiratory volume in one second (FEV₁), and its ratio to forced vital capacity (FEV₁/FVC). Given that cigarette smoking adversely affects pulmonary function, we conducted genome-wide joint meta-analyses (JMA) of single nucleotide polymorphism (SNP) and SNP-by-smoking (ever-smoking or pack-years) associations on FEV₁ and FEV₁/FVC across 19 studies (total N = 50,047). We identified three novel loci not previously associated with pulmonary function. SNPs in or near DNER (smallest P_{JMA} = 5.00 × 10⁻⁵), HLA-DQB1 and HLA-DQA2 (smallest P_{JMA} = 4.35 × 10⁻⁵), and KCNJ2 and SOX9 (smallest P_{JMA} = 1.28 × 10⁻⁵) were associated with FEV₁/FVC or FEV₁ in meta-analysis models including SNP main effects, smoking main effects, and SNP-by-smoking (ever-smoking or pack-years) interaction. The HLA region has been widely implicated for autoimmunity and lung phenotypes, while the other novel loci, which have not been widely implicated. We evaluated DNER, KCNJ2, and SOX9 and found them to be expressed in human lung tissue. DNER and SOX9 further showed evidence of differential expression in human airway epithelium in smokers compared to non-smokers. Our findings demonstrated that joint testing of SNP and SNP-by-environment interaction identified novel loci associated with complex traits that are missed when considering only the genetic main effects.


Editor: Greg Gibson, Georgia Institute of Technology, United States of America

Received July 3, 2012; Accepted October 1, 2012; Published December 20, 2012

This is an open-access article, free of all copyright, and may be freely reproduced, distributed, transmitted, modified, built upon, or otherwise used by anyone for any lawful purpose. The work is made available under the Creative Commons CC0 public domain dedication.

Funding: This work was supported, in part, by the Intramural Research Program of the National Institutes of Health (NIH), the National Institute of Environmental Health Sciences (NIEHS, Z01ES043012). The CHARGE Pulmonary Working Group acknowledges funding from the National Heart, Lung, and Blood Institute (NHLBI) (HL105756) and organizational support from the CHARGE Consortium. From SpirioMeta, MD Tobin was supported by UK MRC Senior Clinical Fellowship G0902313. IP Hall and the laboratory work supporting exposure profiling were supported by MRC (G1000861). P Kraft and H Aschard were supported by R21DK084529. The Age, Gene/Environment Susceptibility (AGES)-Reykjavik Study is funded by NIH contract number N01-AU-12100, Hajteraven (the Icelandic Heart Association), and the Althingi (the Icelandic Parliament). The Atherosclerosis Risk in Communities (ARIC) Study is carried out as a collaborative study supported by NHLBI contracts (HHSN268201100006C, HHSN268201100007C, HHSN268201100008C, HHSN268201100009C, HHSN268201100100C, HHSN268201100101C, and HHSN26820110012). R19HL087651, R19HL95367, and R19HL086694; National Human Genome Research Institute (NHGRI) contract U01HG004402; and NIH contract HHSN2682006225226C. Infrastructure was partly supported by grant number UL1RR025051, a component of the NIH and NIH Roadmap for Medical Research. The British 1958 Cohort (BSBC) DNA collection was funded by the Wellcome Trust grant 0000914. Genotyping for the Wellcome Trust Case Control Consortium of BSBC was funded by the Wellcome Trust grant 076113/B/04/Z. The Type 1 Diabetes Genetics Consortium of BSBC was supported by the National Institute of Diabetes and Digestive and Kidney Diseases (NIDDK), National Institute of Allergy and Infectious Diseases, NIH, National Institute of Child Health and Human Development, and Juvenile Diabetes Research Foundation International and supported by U01 DK04318. Genome-wide data was deposited by the Diabetes and Inflammation Laboratory, Cambridge Institute for Medical Research (CIMR), University of Cambridge, which is funded by Juvenile Diabetes Research Foundation International, the Wellcome Trust, and the National Institute for Health Research Cambridge Biomedical Research Centre and is in receipt of a Wellcome Trust Strategic Award (079895). The Coronary Artery Risk Development in Young Adults (CARDIA) study was funded by contracts N01-HC-95095, N01-HC-98047, N01-HC-98048, N01-HC-98049, N01-HC-98050, N01-HC-05134, N01-HC-05187, N01-HC-45204, and N01-HC-45204 from NHLBI to the CARDIA Investigators. Genotyping of the CARDIA participants was supported by grants U01-HG-004729, U01-HG-004446, and U01-HG-004424 from the NHGRI. Statistical analyses were supported by grants U01-HG-004729 and R01-HL-08409 to M Fornage Cardiovascular Health Study (CHS) was supported by NHLBI contracts N01-HC-85239, N01-HC-85079 through N01-HC-85086, N01-HC-35129, N01-HC-15103, N01-HC-55222, N01-HC-75150, N01-HC-45133, and HHSN26820120036C and by NHLBI grants HL080295, HL075366, HL087652, and HL105756 with additional contribution from
Introduction

Spirometric measures of pulmonary function, particularly forced expiratory volume in one second (FEV1) and its ratio to forced vital capacity (FEV1/FVC), are important clinical tools for diagnosing pulmonary disease, classifying its severity, and evaluating its progression over time. These measures also predict other morbidities and mortality in the general population [1–3]. Genetic factors likely play a prominent role in determining the maximal level of pulmonary function [18] makes it a good candidate for such an approach, since interactive effects alone [15–17]. The well-documented and relatively uncommon deficiency of HHIP genetic factors may have heterogeneous effects on pulmonary function [8–10]. Inhaled pollutants, especially cigarette smoking, can have important adverse effects on pulmonary function. Candidate gene studies have not consistently identified interactions with cigarette smoking in relation to pulmonary function. Despite the importance of smoking and other environmental factors in the etiology of many complex human diseases and traits, few GWAS have incorporated gene-by-environment interactions [11–14]. Meta-analyses are generally necessary to provide sufficient sample size to detect moderate effects, and methods for joint testing of single nucleotide polymorphism (SNP) main effects and SNP-by-environment interactions in the meta-analysis setting have only recently been developed [15,16]. This strategy has the potential to identify novel loci that would not emerge from analyses based on the SNP main or interactive effects alone [15–17]. The well-documented and consistent deleterious effect of cigarette smoking on pulmonary function [18] makes it a good candidate for such an approach, since genetic factors may have heterogeneous effects on pulmonary function depending on smoking exposure. We conducted genome-wide association studies (GWAS) of FEV1/FVC 6, 7, GWAS meta-analyses for FEV1/FVC and FEV1 from the Cohorts for Heart and Aging Research in Genomic Epidemiology (CHARGE) and SpiroMeta Consortia have together identified 26 additional novel loci in or near the following genes:

ADAM19, AGER-PPT2, ARMC2, C10orf11, C3DC8, CDC123, CDFP1, FAM13A, GPR126, HDAC4, HTRA1, INTS12-GSTCD-NPNT, KCNE2, LRPI, MECOM (EVT1), MFAP2, MMP13, NCR3, PDI, PTCH1, RARB, SPATA9, TGFBR2, THSD4, TNS1, and ZSCAN43 [8–10].

Inhaled pollutants, especially cigarette smoking, can have important adverse effects on pulmonary function. Candidate gene studies have not consistently identified interactions with cigarette smoking in relation to pulmonary function. Despite the importance of smoking and other environmental factors in the etiology of many complex human diseases and traits, few GWAS have incorporated gene-by-environment interactions [11–14]. Meta-analyses are generally necessary to provide sufficient sample size to detect moderate effects, and methods for joint testing of single nucleotide polymorphism (SNP) main effects and SNP-by-environment interactions in the meta-analysis setting have only recently been developed [15,16]. This strategy has the potential to identify novel loci that would not emerge from analyses based on the SNP main or interactive effects alone [15–17]. The well-documented and consistent deleterious effect of cigarette smoking on pulmonary function [18] makes it a good candidate for such an approach, since genetic factors may have heterogeneous effects on pulmonary function depending on smoking exposure. We conducted genome-wide association studies (GWAS) of FEV1/FVC 6, 7, GWAS meta-analyses for FEV1/FVC and FEV1 from the Cohorts for Heart and Aging Research in Genomic Epidemiology (CHARGE) and SpiroMeta Consortia have together identified 26 additional novel loci in or near the following genes:
Author Summary

Measures of pulmonary function provide important clinical tools for evaluating lung disease and its progression. Genome-wide association studies have identified numerous genetic risk factors for pulmonary function but have not considered interaction with cigarette smoking, which has consistently been shown to adversely impact pulmonary function. In over 50,000 study participants of European descent, we applied a recently developed joint meta-analysis method to simultaneously test associations of gene and gene-by-smoking interactions in relation to two major clinical measures of pulmonary function. Using this joint method to incorporate genetic main effects plus gene-by-smoking interaction, we identified three novel gene regions not previously related to pulmonary function: (1) DNER, (2) HLA-DQB1 and HLA-DQA2, and (3) KCNJ2 and SOX9. Expression analyses in human lung tissue from ours or prior studies indicate that these regions contain genes that are plausibly involved in pulmonary function. This work highlights the utility of employing novel methods for incorporating environmental interaction in genome-wide association studies to identify novel genetic regions.

Results

Table S1 presents characteristics of the 50,047 participants from 19 studies contributing to our analyses. As expected, mean FEV1 and FVC values were lower in studies with the oldest participants. Standardized residuals of FEV1 and FEV1/FVC (see Methods) were used as the phenotypes for the JMA, in order to maximize comparability with our recent GWAS meta-analysis from the CHARGE and SpiroMeta Consortia [10]. Our original GWAS meta-analyses, conducted separately in CHARGE and SpiroMeta, showed that we were able to identify replicable genetic loci whether using actual pulmonary function measures [8] or their standardized residuals [9]. The standardized residual approach was similarly used in GWAS of other complex quantitative traits, such as height and body mass index from the Genetic Investigation of ANthropometric Traits (GIANT) Consortium [19,20].

In each of the 19 studies, four regression models with differing SNP-by-smoking interaction terms were run: (1) SNP-by-ever-smoking for standardized FEV1/FVC residuals, (2) SNP-by-pack-years for standardized FEV1/FVC residuals, (3) SNP-by-ever-smoking for standardized FEV1 residuals, and (4) SNP-by-pack-years for standardized FEV1 residuals. Study-specific genomic inflation factors (\(k_{\text{gc}}\)) were calculated for the 1 degree-of-freedom (d.f.) SNP-by-smoking interaction term, to ensure that there was no substantial inflation due to the main effect of smoking being misspecified [21]. All study-specific results had 1 d.f. \(k_{\text{gc}} \approx 1.09\) (Table S2), which is of comparable magnitude to other studies with large sample sizes [10,19,22,23].

The study-specific regression coefficients from each of the four models were then combined in JMA, and the resulting \(k_{\text{gc}}\) values from the 2 d.f. JMA, calculated across all SNPs, ranged from 1.056 to 1.064. The quantile-quantile plots (Figure S1) show substantial deviation from expectation for SNPs having low \(P\) values from the JMA \(P_{\text{JMA}}\). The JMA results corresponding to the top SNP from each previously implicated locus [8–10] are presented in Table S3. To identify novel loci among the genome-wide significant loci implicated by our JMA models, the genomic regions surrounding the most significant SNP from each of the 27 previously implicated loci [8–10] (500 kb upstream to 500 kb downstream of each SNP) were removed from consideration (Table S3). Following the removal of all previously implicated loci [8–10], the quantile-quantile plots show that some deviation remained between observed and expected \(P\) values for high-signal SNPs suggesting the presence of novel signals.

In the JMA of SNP and SNP-by-smoking in relation to FEV1/FVC, we observed two novel loci containing several significant SNP associations at the standard genome-wide Bonferroni-corrected threshold of \(P_{\text{JMA}} < 5 \times 10^{-8}\) when considering interaction with ever-smoking (Figure 1A) or pack-years (Figure 1B). The SNP associations from both loci also exceeded the more conservative genome-wide significance threshold of \(P_{\text{JMA}} < 1.25 \times 10^{-8}\), based on additional Bonferroni correction for the four JMA models.

The most statistically significant result was for rs7594321, an intronic SNP located in DNER (delta/notch-like EGF-related receptor) on chromosome 2, which gave \(P_{\text{JMA}} = 2.64 \times 10^{-9}\) (corresponding \(P_{\text{INT}} = 0.27\) in the ever-smoking model and \(P_{\text{JMA}} = 5.00 \times 10^{-11}\) (corresponding \(P_{\text{INT}} = 0.0069\) in the pack-years model (Table 1). For the ever/never-smoking interaction model, the observed level of significance for the JMA is plausible in the presence of a nominally significant SNP main effect and a nonsignificant interactive effect, as detailed in Text S1. The rs7594321 T allele had a positive \(\beta\) coefficient for the genetic main association and a negative \(\beta\) coefficient for the interaction (Table 1, Table S4 for study-specific results). The regression coefficients correspond to a per allele change of 0.049 (95% CI: 0.030, 0.068) in never-smokers and 0.035 (95% CI: 0.016, 0.053) in ever-smokers. A conserved binding site for the Zic1 transcription factor is located 115 kb base pairs away from rs7594321. Further, rs7594321 is located upstream of the previously implicated PIDD1 gene (Figure 2A), but it is 713 kb away from the previously implicated SNP rs1435867, which is located downstream of PIDD1. There is no linkage disequilibrium (LD) between rs7594321 and rs1435867 \((r^2 = 0, \Delta r = 0)\).

Our next most statistically significant SNP (rs7764819) is intergenic between two human leukocyte antigen (HLA) genes, HLA-DQB1 and HLA-DQA2, on chromosome 6 (Figure 2B). The HLA-DQ region is highly variable, and the association signal in this region is largely driven by two SNPs that are in high LD with one another (rs7764819 and rs7765379, \(r^2 = 1\)) but only low to moderate LD with all other genotyped and imputed SNPs. A GWAS meta-analysis of asthma implicating the HLA-DQ region similarly found highly significant associations with only a few SNPs [24]. Our top SNP rs7764819 gave \(P_{\text{JMA}} = 4.39 \times 10^{-8}\) in the ever-smoking model and \(P_{\text{JMA}} = 4.35 \times 10^{-7}\) in the pack-years model for FEV1/FVC (Table 1). The corresponding INT values were >0.05 (see Text S1). The rs7764819 T allele had negative \(\beta\) coefficients for both the main association and interaction (Table 1, Table S5 for study-specific results), which correspond to a SNP effect of −0.060 (95% CI: −0.09, −0.031) in never-smokers and −0.070 (95% CI: −0.10, −0.042) in ever-smokers. Although rs7764819 is located 529 kb away from a previously implicated AGER SNP rs2070600, there is some LD between the two SNPs \((r^2 = 0.29, \Delta r = 0.81)\). Conserved binding sites for two transcription factors, HTF and Lmo2, are located within 100 kb of rs7764819.

Besides the DNER and HLA-DQB1/HLA-DQA2 loci, SNPs from 12 other chromosomal regions having \(P_{\text{JMA}}\) values between...
The most significant SNP (rs11654749) from both smoking models is intergenic between KCNJ2 (a potassium inwardly-rectifying channel also known as Kir2.1) and SOX9 (sex determining region Y-box 9) (Figure 2C). Conserved binding sites for four transcription factors (HNF-1, CP2, Cdc5, and FOXF2) are located within 100 kb upstream or downstream of rs11654749. The rs11654749 SNP gave $\log_{10} P_{\text{JMA}} = 1.26 \times 10^{-6}$ in the ever-smoking model and $P_{\text{JMA}} = 6.03 \times 10^{-8}$ in the pack-years model (Table 1). The corresponding $P_{\text{INT}}$ values were >0.05 (see Text S1). The rs11654749 T allele had negative $\beta$ coefficients for both the main association and interaction (Table 1, Table S6 for study-specific results). These estimates correspond to a SNP effect of $-0.028$ (95% CI: $-0.047$, $-0.010$) in never-smokers and $-0.046$ (95% CI: $-0.063$, $-0.029$) in ever-smokers. To better understand the magnitude of these $\beta$ estimates, we compared our results with those observed in one of our previous GWAS meta-analyses of SNP main effects [9], where standardized residuals of the pulmonary function measures were similarly computed. For a SNP with MAF around 40%, an absolute $\beta$ value of 0.028 would be equivalent to 19 mL per copy of the risk allele (comparable to a year of FEV$_1$ decline in healthy never-smokers), and an absolute $\beta$ value of 0.046 would be equivalent to 31 mL per copy of the risk allele (comparable to a year and a half of FEV$_1$ decline in healthy never-smokers) [23].

Besides this KCNJ2/SOX9 locus, SNPs from five other chromosomal regions have $P_{\text{JMA}}$ values between $5 \times 10^{-8}$ and $1 \times 10^{-6}$ from either smoking model in relation to FEV$_1$, as shown in Table S6. In secondary meta-analyses of the interaction product terms, there were no SNP-by-smoking (ever-smoking or pack-years) interactions implicated at genome-wide statistical significance with FEV$_1$. SNPs from four chromosomal regions had $P_{\text{INT}}$ values between $5 \times 10^{-8}$ and $1 \times 10^{-6}$ in relation to FEV$_1$, as shown in Table S7.

None of the most significant SNPs from the three novel loci we identified by the JMA were associated with FEV$_1$/FVC or FEV$_1$/FVC$\delta$ at or near genome-wide significance in our previous GWAS meta-analysis of 48,201 participants from the CHARGE andSpiroMeta Consortia. In fact, the lowest $P$ value observed for these SNPs was $1.04 \times 10^{-5}$ (Table 2) [10].

To evaluate whether the three novel loci identified by the JMA were related to smoking, we evaluated their SNP associations with ever-smoking and cigarettes per day using GWAS meta-analysis results from the Oxford-GlaxoSmithKline (Ox-GSK) Consortium (N = 41,150) [26]. None of our implicated SNPs were associated with these smoking measures at the standard threshold ($P<5 \times 10^{-8}$ as indicated by the solid red line). Names of the novel gene (or closest genes) are provided.
with these smoking phenotypes at \( P < 0.05 \) (Table S9), adding confidence that our JMA-implicated SNP associations were not simply reflective of smoking main effects.

**Expression analyses**

Three genes (DNER, KCNJ2, and SOX9) harboring or flanking novel genome-wide significant SNPs were selected for follow-up mRNA expression profiling in human lung tissue and a series of primary cells. Transcripts of all three genes were found in lung tissue, airway smooth muscle, and bronchial epithelial cells; DNER and KCNJ2 transcripts were also found in peripheral blood cells (Table S10).

In a separate line of investigation, using the publicly available Gene Expression Omnibus repository [27,28], we found that the expression profiling of DNER and SOX9 showed differential expression in human airway epithelium of smokers compared to non-smokers (Figure S2A and S2B) [29]. Expression profiling of KCNJ2 did not show statistically significant differential expression by smoking status (Figure S2C) [29]. We also identified novel genome-wide significant SNPs in the HLA-DQ region, but we did not examine HLA-DQ expression given the known expression of class II MHC antigens on a range of airway cell types [30,31]. However, the lead SNP in this region (rs7764819) was associated with statistically significant effects on HLA-DQ1 expression (\( P = 1.2 \times 10^{-11} \)), according to an eQTL analysis database of lymphoblastoid cell lines [32].

**Discussion**

Few GWAS have accounted for potential interaction with environmental risk factors. To identify novel genetic risk factors that are missed when considering only genetic main effects [33], we used the newly available JMA method [15] to simultaneously summarize regression coefficients for the main SNP and SNP-by-smoking interactive effects in 50,047 participants from 19 studies, based on models that were fully saturated for the main effect of smoking. This study represents the most comprehensive analysis to date of gene-by-smoking interaction in relation to pulmonary function. We identified two novel loci (DNER and HLA-DQB1/HLA-DQA2) having highly significant evidence for association with FEV1/FVC. A third novel locus (KCNJ2/SOX9) was associated with FEV1. For the most significant SNPs at each of these three loci, there was no evidence for heterogeneity across the studies (smallest heterogeneity \( P = 0.59 \), indicating that the associations were not driven by one or a few studies and thus reflect accumulation of evidence across the studies. None of these three loci had previously been associated with pulmonary function. The comparison of results with our prior GWAS meta-analysis of SNP main effects [10], using a comparable sample size, suggested that the SNP associations for our top SNPs were weaker in our previous analyses that examined only genetic main effects. However, our analyses and those of Manning et al. [14] suggest that some of the benefit of using the joint test for some findings comes from the careful adjustment for the environmental main effect. Thus, future studies aimed at replicating these findings may wish to jointly test the SNP main and interactive effects [15,16,33] instead of implementing a standard test of only the SNP main effects. If there is no evidence for interaction at a given locus, the SNP associations for our top SNPs were weaker in our previous analyses that examined only genetic main effects. However, our analyses and those of Manning et al. [14] suggest that some of the benefit of using the joint test for some findings comes from the careful adjustment for the environmental main effect. Thus, future studies aimed at replicating these findings may wish to jointly test the SNP main and interactive effects [15,16,33] instead of implementing a standard test of only the SNP main effects. If there is no evidence for interaction at a given locus, the saturation of the main effect of the environmental factor may be important. The joint testing is applicable for both candidate gene [15] and genome-wide [14] approaches. Further, there was minimal overlap in the top SNPs associated with FEV1/FVC and FEV1, as similarly observed in our previous GWAS meta-analyses of SNP main effects [8–10]. Given that the biological underpinnings of these discrepant association findings remain unknown, future studies should evaluate these genetic loci in the context of the pulmonary function measure for which they were originally implicated.

Given that pulmonary function is a phenotype for which numerous genetic loci have been identified in GWAS and smoking is clearly associated with pulmonary function, it might seem surprising that none of the genome-wide significant SNPs

---

**Table 1. Genome-wide significant SNPs from the joint meta-analysis (JMA) of SNP and SNP-by-smoking (ever-smoking or pack-years) interaction in relation to pulmonary function.**

<table>
<thead>
<tr>
<th>SNP (coded allele)</th>
<th>Chr</th>
<th>Gene/closest gene(s)</th>
<th>Coded allele frequency*</th>
<th>JMA results</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Smoking metric</td>
</tr>
<tr>
<td><strong>SNPs implicated in relation to FEV1/FVC</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>rs7594321 (T)</td>
<td>2q36.3</td>
<td>DNER</td>
<td>0.35</td>
<td>Ever-smoking</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Pack-years</td>
</tr>
<tr>
<td>rs7764819 (T)</td>
<td>6p21.32</td>
<td>HLA-DQB1/HLA-DQA2</td>
<td>0.89</td>
<td>Ever-smoking</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Pack-years</td>
</tr>
<tr>
<td><strong>SNPs implicated in relation to FEV1</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>rs11654749 (T)</td>
<td>17q24.3</td>
<td>KCNJ2/SOX9</td>
<td>0.39</td>
<td>Ever-smoking</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Pack-years</td>
</tr>
</tbody>
</table>

After removing SNPs with known associations with FEV1/FVC or FEV1, three novel loci with genome-wide significant SNPs (standard threshold of \( P < 5 \times 10^{-8} \)) remained from the JMA testing in the current study. The most significant SNP from each locus is shown.

FEV1, forced expiratory volume in the first second; FVC, forced vital capacity; JMA, joint meta-analysis; SE, standard error; SNP, single nucleotide polymorphism.

*Weighted average coded allele frequency across the 19 studies. The coded allele refers to the effect allele.

\( \beta_{SNP} \), per allele change in the FEV1/FVC standardized residual due to the SNP main association.

\( \beta_{INT} \), per allele change in the FEV1/FVC standardized residual due to the interaction between SNP and smoking.

doi:10.1371/journal.pgen.1003098.t001
Findings demonstrate the value of applying the newly developed broad public health campaigns to curb smoking. Instead, our established harmful effects of cigarette smoking nor the need for smoking status or pack-years [10]. Some of our novel JMA findings compared with our previous GWAS findings may reflect, in part, the saturated modeling of the smoking main effect rather than the interaction per se.

The current analysis of 50,047 participants included only 1,846 more participants than our previous GWAS meta-analysis of SNP main effects [10]. To evaluate the likelihood that this 3.8% increase in sample size above that in our previous meta-analysis of pulmonary function was sufficient to explain our identification of these three novel loci at genome-wide statistical significance in the current JMA, we calculated the statistical power to detect genetic main associations (QUANTO [34]) with minor allele frequency (MAF) and β estimates comparable to the three genome-wide significant SNPs presented in Table 1. The current study (total N = 50,047 participants) had only 0.7% to 4.2% more statistical power than our previous GWAS meta-analysis (total N = 48,201 participants) [10], suggesting that the JMA-implicated SNPs are not merely reflective of increased power to detect genetic main effects. Instead, our novel JMA findings demonstrate an advantage of the method used to jointly test the SNP and SNP-by-smoking interactive effects, including the benefit of the saturated modeling of the smoking main effect.

SNPs located in the DNER gene were significantly associated with FEV1/FVC, even at the more conservative P-value threshold of 1.25×10^{-6}. The JMA results for DNER SNPs were driven by both smoking-adjusted main effects and interaction with quantitative smoking history. The DNER protein product is a ligand of the Notch signaling pathway that has been implicated in neuronal differentiation and maturation [35,36], adipogenesis [37], and hair-cell development [38]. The Notch pathway is a critical controller of cellular differentiation in multiple organs including the lung [39,40]. Interestingly, the expression levels of many members of the Notch signaling cascade are significantly altered in the lung [39,40].

Figure 2. Regional association plots of novel loci implicated for pulmonary function. Three novel loci contained SNPs associated with FEV1/FVC or FEV1 at the standard genome-wide significance threshold (P<5×10^{-8}) in joint meta-analyses of SNP and SNP-by-smoking interaction. SNPs are shown within 500 kb of the most significant SNPs on chromosomes (A) 2q36.3 associated with FEV1/FVC, (B) 6p21.32 associated with FEV1/FVC, and (C) 17q24.3 associated with FEV1. Pairwise r^2 values were based on the HapMap CEU population, and progressively darker shades of red indicate higher r^2 values. Estimated recombination rates from HapMap are shown as background lines.

doi:10.1371/journal.pgen.1003098.g002

implicated by the JMA demonstrated a substantial interaction per se. The lack of strong interactive effects does not negate the well-established harmful effects of cigarette smoking nor the need for broad public health campaigns to curb smoking. Instead, our findings demonstrate the value of applying the newly developed joint methods to uncover novel genetic risk factors that might shed light on the mechanisms leading to reduced pulmonary function.

Our pattern of SNP main and interactive results resemble the patterns seen in another recent application of the same JMA method to incorporate the interaction with body mass index (BMI) into GWAS of type 2 diabetes traits (fasting insulin and blood glucose) [14]. In that study with a sample size of 96,453, nearly double that of ours, the top JMA finding had a corresponding interaction P-value of 1.6×10^{-4} [14]. In our study, the smallest interaction P-value for our top JMA finding was 6.9×10^{-3}. In both our GWAS of smoking and pulmonary function and the recent GWAS of BMI and diabetes traits [14], the SNPs newly implicated by the JMA had marginally significant associations with the trait under study in models with no interaction term, but they became genome-wide significant when accounting for the environmental factor (cigarette smoking or BMI) and the SNP-by-environment interaction. Our JMA included careful modeling of the environmental factor to saturate the environmental main effects along with the interaction testing. In the GWAS of diabetes traits [14], the careful modeling of the environmental factor appeared to account for some of the novel findings from the JMA, consistent with the modest evidence for interaction [14]. Although our previous GWAS meta-analysis was conducted in ever/never-smoking strata, the regression models were not adjusted for smoking status or pack-years [10]. Some of our novel JMA findings with FEV1/FVC or FEV1 at the standard genome-wide significance threshold.
The lead SNP is associated with expression of wide significance threshold. The eQTL analyses indicated that the associated with FEV1 at the standard meta-analysis, because the genetic main associations were not most likely that this locus was not identified in our previous GWAS. We found little evidence for interaction between HLA variants and cigarette smoking has been previously implicated. We found little evidence for interaction with smoking at this locus, suggesting that the JMA results were primarily driven by smoking-adjusted genetic main effects. It is most likely that this locus was not identified in our previous GWAS meta-analysis, because the genetic main associations were not evaluated with careful adjustment for smoking status and pack-years. Adjustment for smoking in the current analysis may have removed residual variance in the outcome that is not attributable to genetic variation [14], thus making the identification of the newly associated SNPs possible.

Intergenic SNPs between KCNJ2 and SOX9 were significantly associated with FEV1 at the standard P value threshold, but not the more conservative threshold. Similar to the HLA region, it appears that the JMA results for the KCNJ2/SOX9 region were primarily driven by smoking-adjusted genetic main effects. This region is enriched for long-range regulatory elements for SOX9, although the possibility of this region containing regulatory elements cannot be discounted [47]. KCNJ2 is a member of the inwardly-rectifying potassium channel family, which regulates membrane potential and cell excitability and is expressed in many tissues including myocardium, neurons, and vasculature. This potassium channel also affects human bronchial smooth muscle tone and airflow limitation [48]. Dominant negative mutations in KCNJ2 cause the Andersen syndrome, characterized by ventricular arrhythmias, periodic paralysis, and a number of skeletal and cardiac abnormalities [49]. SOX9 is a transcription factor that is essential for cartilage formation, [50] but it is also abundantly expressed in other tissues including the respiratory epithelium during development [51]. Sox9<sup>−/−</sup> and Sox9<sup>+/−</sup> mice have multiple skeletal anomalies and severe tracheal cartilage malformations and die prematurely from respiratory insufficiency [50,52]. Mutations in SOX9 cause campomelic dysplasia characterized by skeletal defects and autosomal sex reversal [53]. These individuals develop respiratory distress due to chest wall abnormalities, narrowed airways resulting from tracheobronchial defects and hypoplastic lungs [54]. We confirmed that KCNJ2 and SOX9 transcripts were present in human lung tissue and peripheral cells. Using publicly available microarray data [29], we established that SOX9 is expressed in human airway epithelial cells and its expression is significantly down-regulated in smokers relative to non-smoking adults (Figure S2B). Taken together, these results suggest that SOX9 may be involved in cigarette smoke-induced airflow obstruction, but further investigation is required to elucidate putative mechanisms.

Most of the previously implicated SNPs had genome-wide significant (or nearly significant) associations with pulmonary function in the JMA, but some were associated with pulmonary function at P values that did not approach the genome-wide statistical significance threshold in the JMA analysis. This pattern has two possible explanations. First, the identification of these SNPs at genome-wide statistical significance in our most recent analysis [10] required a sample size of nearly 95,000 individuals, which was obtained by combining discovery and replication cohorts, including additional genotyping on thousands of participants from studies without GWAS data. In the current analysis, the sample size is greatly reduced because of the need for detailed quantitative smoking data and because we were unable to perform additional genotyping in studies without GWAS data. Second, Manning et al.[15] showed that a meta-analysis of main SNP effects has slightly greater power than the JMA under the scenario of no interaction, so it is not surprising that a few of the prior SNP findings had varying levels of significance between our prior GWAS meta-analyses [8–10] and the current JMA study. While our sample size of over 50,000 study participants is large, and the study of Manning et al. [14] examining SNP-by-BMI interaction in relation to fasting insulin is nearly twice as large, identification of interactions is challenging from a statistical power perspective. Given the multiple testing issues in genome interaction testing, even larger sample sizes will likely be needed to identify gene-by-environment interactions with rare variants or with the modest effect sizes that we generally expect. Nonetheless, our findings exemplify the greater power achieved by using the joint methods, such as those reported by Manning et al. [15] and Kraft et al. [16,33], to incorporate interaction with a clearly associated environmental risk factor. The novel genetic loci identified here for pulmonary function would have remained unknown using standard GWAS approaches.

### Methods

**Ethics statement**

Nineteen independent studies contributed to our analyses. All study protocols were approved by the respective local Institutional

**Table 2.** Look-up evaluation of SNP main associations with FEV<sub>I</sub>/FVC and FEV<sub>I</sub> using data generated by our previous genome-wide association study meta-analysis (N = 48,201), for the most significant SNP from each of the three novel loci implicated at genome-wide significance in the joint meta-analysis.

<table>
<thead>
<tr>
<th>SNP (coded allele)</th>
<th>Gene/closest gene(s)</th>
<th>FEV&lt;sub&gt;I&lt;/sub&gt;/FVC</th>
<th>FEV&lt;sub&gt;I&lt;/sub&gt;</th>
</tr>
</thead>
<tbody>
<tr>
<td>rs7594321 (T)</td>
<td>DNER</td>
<td>0.032</td>
<td>1.04×10&lt;sup&gt;−5&lt;/sup&gt;</td>
</tr>
<tr>
<td>rs7764819 (T)</td>
<td>HLA-DQB1/HLA-DQA2</td>
<td>−0.044</td>
<td>8.79×10&lt;sup&gt;−5&lt;/sup&gt;</td>
</tr>
<tr>
<td>rs11654749 (T)</td>
<td>KCNJ2/SOX9</td>
<td>−0.023</td>
<td>0.0015</td>
</tr>
</tbody>
</table>

**FEV<sub>I</sub>, forced expiratory volume in the first second; FVC, forced vital capacity; SE, standard error; SNP, single nucleotide polymorphism.**

<sup>1</sup>β<sub>SE</sub> per allele change in the FEV<sub>I</sub>/FVC standardized residual due to the SNP main association.

doi:10.1371/journal.pgen.1003098.t002
Cohort studies

Of the 19 studies contributing to our analyses, 18 studies came from the CHARGE [8,55] or SpiroMeta [9] Consortium: Age, Gene, Environment, Susceptibility (AGES) – Reykjavik Study [56]; Atherosclerosis Risk in Communities (ARIC) Study [57]; British 1958 Birth Cohort (B58C) [58]; Coronary Artery Risk Development in Young Adults (CARDIA) [59,60]; Cardiovascular Health Study (CHS) [61]; European Community Respiratory Health Survey (ECRHS) [62]; European Prospective Investigation into Cancer and Nutrition (EPIC, obese cases and population-based subsets) [63]; Framingham Heart Study (FHS) [64,65]; Health, Aging, and Body Composition (Health ABC) Study [66]; Northern Finland Birth Cohort of 1966 (NFBC1966) [67,68]; Multi-Ethnic Study of Atherosclerosis (MESA) [69,70]; Rotterdam Study (RS-I, RS-II, and RS-III) [71]; Swiss Study on Air Pollution and Lung Diseases in Adults (SAPALDIA) [72]; Study of Health in Pomerania (SHIP) [73]; and TwinsUK [74]. We reached out to other population-based studies with GWAS genotyping and data available on cigarette smoking and pulmonary function, resulting in the inclusion of LifeLines [75]. Given the greater power needed to detect novel genetic loci with subtle gene-environment interaction regardless of the statistical method used [16], we chose to maximize statistical power to discover novel genetic loci by combining all available participants and to use the regression coefficients across the many different component studies as evidence for consistency. This approach was similarly taken by another large-scale GWAS consortium for discovering SNP main effects [24].

Pulmonary function measurements and smoking information

All studies were included in our previous GWAS meta-analysis of pulmonary function or the follow-up replication analyses, wherein their pulmonary function testing protocols were described [10]. For studies with spirometry at a single visit (B58C, LifeLines, MESA, NFBC1966, SHIP, RS-I, RS-II, and RS-III), we analyzed FEV1/FVC and FEV1 measured at that visit. For studies with spirometry at more than one visit, we analyzed measurements from the baseline visit (AGES, ARIC, CARDIA, CHS, ECRHS, EPIC, EPIC obese cases, EPIC population-based, Health ABC, and SAPALDIA) or the most recent examination with spirometry data (FHS and TwinsUK).

Smoking history (current-, past-, and never-smoking) was ascertained by questionnaire at the time of pulmonary function testing. Pack-years of smoking were calculated for current and past smokers by multiplying smoking amount (packs/day) and duration. Pack-years of smoking were calculated for current and past smokers, pack-years of smoking (continuous variable), and a SNP-by-smoking interaction product term. Two of the 19 studies (meta-analysis $\beta = -0.0030$ and corresponding $P<1 \times 10^{-6}$ for pack-years of smoking).

The FEV1/FVC and FEV1 standardized residuals were used as the phenotypes for genome-wide association testing with linear regression models, which included the following predictor variables: imputed SNP genotype dosages, smoking history (dichotomous variable, 0 = never-smokers and 1 = ever-smokers), smoking status (dichotomous variable, 0 = never- and past-smokers and 1 = current-smokers), pack-years of smoking (continuous variable), and a SNP-by-smoking interaction product term. Two of the 19 studies (FHS and TwinsUK) had much relatedness among participants, and we took appropriate account of relatedness in the association testing (Table S12). Four regression models with interaction terms for ever-smoking or pack-years were specified in relation to standardized residuals for FEV1/FVC or FEV1. As it has long been advised in studying interactions, the regression models were designed to fully saturate the main smoking effect on pulmonary function, so that the interaction terms do not capture residual main effects [77]. In each of the 19 studies, the genome-wide analyses were implemented with robust variance estimation using the software packages indicated in Table S12.

Our analyses were aimed at finding novel loci associated with pulmonary function when considering an interaction with cigarette smoking, so we chose to implement JMA of SNP main and interactive SNP-by-smoking effects (two d.f. test of the null hypothesis $\beta_{SNP} = 0$ and $\beta_{INT} = 0$) [15]. Manning et al. previously compared the joint methods, such as JMA, with other methods that incorporate gene-environment interaction (such as screening for main effects [78] or conducting a 1 d.f. meta-analysis of the interaction product term), and they found that the joint methods offer optimal statistical power over a range of scenarios for SNP main and interactive effects [15,33]. Therefore, our analyses centered on the JMA method, which simultaneously estimates regression coefficients for the SNP and SNP-by-smoking interaction terms, while accounting for their covariance, to generate a joint test of significance [15]. It also accounts for the unequal variances from studies of different sample sizes. Secondarily, we implemented meta-analyses of just the $\beta$ coefficient from the interaction term for comparison with the JMA results. Of note, the two-step gene-environment interaction study designs by Murcray et al. [79,80] and Gauderman et al. [81] are applicable to case-control or case-parent trio studies, respectively, and were thus not considered for our population-based studies of continuous traits.

The JMA was conducted with fixed effects on approximately 2.5 million SNPs using METAL software (version 2010-02-08) [82] and patch source code provided by Manning et al. [82]. Genomic control correction was applied by computing $\lambda_{gc}$ as the ratio of the observed and expected (2 d.f.) median chi-square statistics and control metrics (call rate, discordance with prior genotyping, and genotypic and phenotypic sex mismatch among others), missing pulmonary function data, or missing covariate data (Table S13).

Statistical analysis

Our analyses included 50,047 participants from 19 studies who passed their study-specific quality control and had complete data on pulmonary function and smoking. Each study transformed the pulmonary function measures to residuals using linear regression of FEV1/FVC (%) and FEV1 (mL) on age, age2, sex, and standing height as predictors. Principal component eigenvectors and recruitment site were also included as covariates to adjust for population stratification (if applicable). The residuals were converted to z scores (henceforth referred to as standardized residuals). We confirmed that smoking was inversely associated with the FEV1/FVC and FEV1 standardized residuals in all 19 studies (meta-analysis $\beta = -0.0030$ and corresponding $P<1 \times 10^{-6}$ for pack-years of smoking).

Of the 19 studies contributing to our analyses, 18 studies came from the CHARGE [8,55] or SpiroMeta [9] Consortium: Age, Gene, Environment, Susceptibility (AGES) – Reykjavik Study [56]; Atherosclerosis Risk in Communities (ARIC) Study [57]; British 1958 Birth Cohort (B58C) [58]; Coronary Artery Risk Development in Young Adults (CARDIA) [59,60]; Cardiovascular Health Study (CHS) [61]; European Community Respiratory Health Survey (ECRHS) [62]; European Prospective Investigation into Cancer and Nutrition (EPIC, obese cases and population-based subsets) [63]; Framingham Heart Study (FHS) [64,65]; Health, Aging, and Body Composition (Health ABC) Study [66]; Northern Finland Birth Cohort of 1966 (NFBC1966) [67,68]; Multi-Ethnic Study of Atherosclerosis (MESA) [69,70]; Rotterdam Study (RS-I, RS-II, and RS-III) [71]; Swiss Study on Air Pollution and Lung Diseases in Adults (SAPALDIA) [72]; Study of Health in Pomerania (SHIP) [73]; and TwinsUK [74]. We reached out to other population-based studies with GWAS genotyping and data available on cigarette smoking and pulmonary function, resulting in the inclusion of LifeLines [75]. Given the greater power needed to detect novel genetic loci with subtle gene-environment interaction regardless of the statistical method used [16], we chose to maximize statistical power to discover novel genetic loci by combining all available participants and to use the regression coefficients across the many different component studies as evidence for consistency. This approach was similarly taken by another large-scale GWAS consortium for discovering SNP main effects [24].

Pulmonary function measurements and smoking information

All studies were included in our previous GWAS meta-analysis of pulmonary function or the follow-up replication analyses, wherein their pulmonary function testing protocols were described [10]. For studies with spirometry at a single visit (B58C, LifeLines, MESA, NFBC1966, SHIP, RS-I, RS-II, and RS-III), we analyzed FEV1/FVC and FEV1 measured at that visit. For studies with spirometry at more than one visit, we analyzed measurements from the baseline visit (AGES, ARIC, CARDIA, CHS, ECRHS, EPIC, EPIC obese cases, EPIC population-based, Health ABC, and SAPALDIA) or the most recent examination with spirometry data (FHS and TwinsUK).

Smoking history (current-, past-, and never-smoking) was ascertained by questionnaire at the time of pulmonary function testing. Pack-years of smoking were calculated for current and past smokers by multiplying smoking amount (packs/day) and duration (years smoked). Table S11 presents the specific questions used to ascertain smoking history and pack-years in each of the 19 studies.

Genotyping, quality control, and imputation

Study participants were genotyped on various genotyping platforms, and standard quality control filters for call rate, Hardy-Weinberg equilibrium p-value, MAF, and other measures were applied to the genotyped SNPs (Table S12). To generate a common set of SNPs for meta-analysis, imputation was conducted with reference haplotype panels from HapMap phase II subjects of European ancestry (CEU) (Table S12) [76]. Imputed genotype dosage values (estimated reference allele count with a fractional value ranging from 0 to 2.0) were generated for approximately 2.5 million autosomal SNPs. Among participants with genome-wide SNP genotyping data, exclusions were made due to standard quality control metrics (call rate, discordance with prior genotyping, and genotypic and phenotypic sex mismatch among others), missing pulmonary function data, or missing covariate data (Table S13).
dividing the observed chi-square statistics by $\lambda_{m}$. SNPs having $P_{\text{JMA}}<3\times10^{-6}$ (the standard Bonferroni-adjusted $P$ value) were considered statistically significant [93]. Further correction for the four different (albeit related) JMA models yielded a conservative $P_{\text{JMA}}$ threshold of $1.25\times10^{-7}$. In addition to reporting the $P_{\text{JMA}}$ for the most significant SNP from each novel locus, we used the $\beta$ and standard error (SE) estimates from the JMA results to calculate the $P$ values corresponding to the SNP main association ($P_{\text{SNP}}$) and the SNP-by-ever-smoking interaction ($P_{\text{INT}}$) [13].

Bioinformatics analysis

Gene annotation was performed using the gene prediction tracks “UCSC Genes” and “RefSeq Genes” in the UCSC browser (http://genome.ucsc.edu). The “sno/mRNA” track from the UCSC browser was used to search for any microRNA within 100 kb upstream or downstream of each SNP, and the “TFBS Conserved” track was used to search for conserved transcription factor binding sites (TFBSs) at or near the most significant SNPs. The SNAP program [84] was used to infer LD patterns, based on the HapMap phase II CEU population.

Expression analyses

We used separate types of expression analyses to confirm the biologic plausibility of our findings. First, we carried out mRNA expression profiling to show whether or not the implicated genes are expressed in human tissues relevant to pulmonary function. The mRNA expression profiles of implicated genes were determined using reverse transcription polymerase chain reaction (RT-PCR). RNA was sourced from lung (Ambion/ABI), human bronchial epithelial cells (Clonetics) [85], and peripheral blood mononuclear cells (3H Biomedica). RNA from human airway smooth muscle cells, cultured as previously described from tissue obtained at thoracotomy [86], was extracted using a commercially available kit (Qiagen). Ethical approval for the use of primary cells was obtained from the local ethics committees. cDNA was generated using 1 $\mu$g of RNA template using random hexamers and a SuperScript kit (Invitrogen) as directed by the manufacturer. PCR assays were designed to cross intron-exon boundaries, where possible and where splice variation was known, in order to detect all variants. The GAPDH gene was used as a positive control for the cDNA quality, and water was used as a negative control. Primer sequences and primers for the genes of interest are used as a positive control for the cDNA quality, and water was used as a negative control.

We used another publicly available data repository to investigate whether any of the implicated genes showed evidence for differential expression depending on smoking history. The gene expression profiles of human airway epithelium from healthy smokers ($N=10$) and nonsmokers ($N=12$) were obtained from the Gene Expression Omnibus site (http://www.ncbi.nlm.nih.gov/geo/) [27,28], based on robust multichip average processing of probe intensities from Affymetrix HG-U133 Plus 2.0 microarrays [29], with the y-axes reflecting the probe intensities of each gene transcript. The x-axes reflect the gene expression profiles of human airway epithelium from healthy smokers versus nonsmokers. Expression profiles of 10 smokers (indicated in blue) and 12 nonsmokers (indicated in red) were obtained for (A) DNER, (B) SOX9, and (C) KCNJ2, using microarray data from the Gene Expression Omnibus site (http://www.ncbi.nlm.nih.gov/geo/) (GSE4498). The two degree-of-freedom genomic inflation factors ($\lambda_{m}$) were shown, as calculated across all SNPs before the exclusion of previously implicated SNPs. The JMA results of all SNPs were plotted (in blue), along with the SNP values remaining after exclusion of the 27 previously implicated loci (in black).

Supporting Information

Figure S1 Quantile–quantile plots for the genome-wide joint meta-analysis (JMA) of SNP and SNP-by-smoking interaction in relation to pulmonary function. The plots compare the observed vs. expected $P$ values for JMA testing of SNPs by (A) ever-smoking in relation to FEV1/FVC, (B) pack-years of smoking in relation to FEV1/FVC, (C) ever-smoking in relation to FEV1, and (D) pack-years of smoking in relation to FEV1. The corresponding two degree-of-freedom genomic inflation factors ($\lambda_{m}$) are shown, as calculated across all SNPs before the exclusion of previously implicated SNPs. The JMA results of all SNPs were plotted (in blue), along with the SNP values remaining after exclusion of the 27 previously implicated loci (in black).

Table S1 Characteristics of study participants (total $N=50,047$) at the time of pulmonary function testing.

Table S2 Genomic inflation factors ($\lambda_{m}$) for study-specific results (corresponding to the 1 degree of freedom SNP-by-smoking product term) in each of the four regression models.

Table S3 Regions surrounding the most significant SNP from each of 27 previously implicated loci (500 kb upstream to 500 kb downstream of each SNP). These loci were excluded when identifying novel loci from the joint meta-analysis (JMA) of SNP and SNP-by-smoking interaction. The smallest $P$ value from the JMA ($P_{\text{JMA}}$) is shown, along with the corresponding JMA model from which the result was obtained.

Table S4 Study-specific results for the genome-wide significant SNP rs7594321 (coded allele: T), located in the DNER gene. $\beta$ estimates and $P$ values are shown for the SNP main association ($P_{\text{SNP}}$ and $P_{\text{SNP}}$) and interactive association ($P_{\text{INT}}$ and $P_{\text{INT}}$) by smoking (ever-smoking and pack-years) in relation to FEV1/FVC. The $P$ values corresponding to the joint test of SNP main and interactive associations are also shown.

Table S5 Study-specific results for the genome-wide significant SNP rs7764819 (coded allele: T), located between the HLA-DQB1 and HLA-DQA1 genes. $\beta$ estimates and $P$ values are shown for the SNP main association ($P_{\text{SNP}}$ and $P_{\text{SNP}}$) and interactive association ($P_{\text{INT}}$ and $P_{\text{INT}}$) by smoking (ever-smoking and pack-years) in relation to FEV1/FVC. The $P$ values corresponding to the joint test of SNP main and interactive associations are also shown.
**References**


---

**Table S6** SNPs from each of 16 chromosomal regions with P values between $5 \times 10^{-8}$ and $1 \times 10^{-6}$ for the joint meta-analysis of SNP and SNP-by-smoking (ever-smoking or pack-years) in relation to pulmonary function (FEV$_1$/FVC or FEV$_1$). A hyphen (‘’’) indicates $P > 1 \times 10^{-6}$. For each regression model, the SNP having the smallest $P_{MA}$ from each locus is shown.

**Table S7** SNPs with $P < 1 \times 10^{-6}$ from the 1 degree-of-freedom meta-analysis of regression coefficients corresponding to the SNP-by-smoking (ever-smoking or pack-years) interaction term in relation to FEV$_1$/FVC. No SNPs exceeded the standard genome-wide significance threshold ($P < 5 \times 10^{-8}$). A hyphen (‘’’) indicates $P > 1 \times 10^{-6}$. For each regression model, the SNP having the smallest $P_{INT}$ from each locus is shown.

**Table S8** Study-specific results for the genome-wide significant SNP rs11654749 (coded allele: T), located between the KCNJ2 and SQX9 genes. β estimates and P values are shown for the SNP main association ($β_{SNP}$ and $P_{SNP}$) and interactive association ($β_{INT}$ and $P_{INT}$) by ever-smoking in relation to FEV$_1$. The P values corresponding to the joint test of SNP main and interactive associations are also shown.

**Table S9** Look-up evaluation of main SNP associations with cigarette smoking phenotypes using data generated by the Oxford-GlaxoSmithKline Consortium (N = 41,150), for the most significant SNP from each of the three novel loci implicated at genome-wide significance in the joint meta-analysis.

**Table S10** mRNA expression profiling of three candidate genes in the human lung and periphery. Primer sequences are provided in Table S5. A ‘‘+’’ sign indicates presence of the transcript, and ‘‘−’’ indicates its absence. All products were sequence verified.

**Table S11** Questionnaire data used to ascertain cigarette smoking history (ever-smoking), amount, and duration across the 19 studies. Smoking amount and duration were used together to calculate pack-years.

**Table S12** Details of single nucleotide polymorphism (SNP) genotyping, quality control (QC), imputation, and statistical analysis across the 19 studies.

**Table S13** Study participants of European descent and quality control (QC) across the 19 studies. Participants passing QC filters and having acceptable spirometry data and complete covariate data were included in the meta-analyses.

**Table S14** Primers for mRNA expression profiling.

**Text S1** Detailed explanation of joint meta-analysis significance levels, in relation to main and interactive significance.

---

**Acknowledgments**

The authors thank the other investigators, staff, and participants of all cohort studies for their valuable contributions. The authors from Lifecare are also grateful to the Medical Biobank Northern Netherlands and participating general practitioners and pharmacists. A full list of principal Cardiovascular Health Study (CHS) investigators and institutions can be found at http://www.chs-nhlbi.org/pi.htm. A full list of participating Multi-Ethnic Study of Atherosclerosis (MESA) investigators and institutions can be found at http://www.mesa-nhlbi.org.

**Author Contributions**


310.


