Genome-wide association and large-scale follow up identifies 16 new loci influencing lung function

Citation for published version:

Digital Object Identifier (DOI):
10.1038/ng.941

Link:
Link to publication record in Edinburgh Research Explorer

Document Version:
Peer reviewed version

Published In:
Nature Genetics

Publisher Rights Statement:

General rights
Copyright for the publications made accessible via the Edinburgh Research Explorer is retained by the author(s) and / or other copyright owners and it is a condition of accessing these publications that users recognise and abide by the legal requirements associated with these rights.

Take down policy
The University of Edinburgh has made every reasonable effort to ensure that Edinburgh Research Explorer content complies with UK legislation. If you believe that the public display of this file breaches copyright please contact openaccess@ed.ac.uk providing details, and we will remove access to the work immediately and investigate your claim.

Download date: 14. Apr. 2019
Genome-wide association and large scale follow-up identifies 16 new loci influencing lung function

Maria Soler Artigas1,2,128, Daan W Loth3,4,128, Louise V Wain1,2,128, Sina A Gharib5,6,128, Ma’en Obeidat7,128, Wenbo Tang8,128, Guanguz Zha10,12, Jing Hua Zhao11, Albert Vernon

128A full list of members is provided in the Supplementary Note.
129These authors contributed equally to this work. Correspondence should be addressed to M.D.T. (mt47@leicester.ac.uk) or S.J.L. (london2@niehs.nih.gov).

AUTHOR CONTRIBUTIONS
Author contributions are listed in alphabetical order. See Supplementary Note for definitions of study acronyms.


COMPETING FINANCIAL INTERESTS:
Inês Barroso and spouse own stock in Incyte Ltd and GlaxoSmithKline. Fredrik Nyberg is employed by AstraZeneca R&D, 431 83 Mölndal, Sweden. Professor Postma has received unrestricted research grants from and has been consultant to AstraZeneca, Boehringer Ingelheim, Chiesi, GlaxoSmithKline, Nycomed, TEVA. Clyde Francks is a full-time employee of the company GlaxoSmithKline (GSK) and GSK also funded several aspects of the study as detailed in the ACKNOWLEDGMENTS section for Ox-GSK.

1Department of Health Sciences, University of Leicester, Leicester, UK 2Department of Genetics, University of Leicester, Leicester, UK 3Department of Epidemiology, Erasmus MC, Rotterdam, The Netherlands 4Inspectorate of Healthcare, The Hague, The Netherlands 5Center for Lung Biology, University of Washington, Seattle, Washington, USA 6Department of Medicine, University of Washington, Seattle, Washington, USA 7Division of Therapeutics and Molecular Medicine, Nottingham Respiratory Biomedical Research Unit, University Hospital of Nottingham, Nottingham, London, UK 8Division of Nutritional Sciences, Cornell University Ithaca, New York, USA 9Department of Twin Research and Genetic Epidemiology, King’s College London, London, UK 10Discipline of Genetics, Faculty of Medicine, Memorial University of Newfoundland, St. John’s, Newfoundland, Canada 11Medical Research Council (MRC) Epidemiology Unit, Institute of Metabolic Science, Addenbrooke’s Hospital, Cambridge, UK 12Icelandic Heart Association, Kopavogur, Iceland 13University of Iceland, Reykjavik, Iceland 14MRC Human Genetics Unit,
School of Women’s and Infants’ Health, The University of Western Australia, Perth, Western Australia, Australia 57Department of Medicine, Columbia University Medical Center, New York, New York, USA 58Department of Surgery and Pathology, University of Western Australia, Nedlands, Western Australia, Australia 59PathWest Laboratory Medicine of Western Australia, Queen Elizabeth II Medical Centre, Nedlands, Western Australia, Australia 60Faculty of Medicine, University of Split, Croatia, Split, Croatia 61Occupational and Environmental Medicine, Department of Public Health and Community Medicine, Institute of Medicine, Sahlgrenska Academy, University of Gothenburg, Gothenburg, Sweden 62Division of Epidemiology and Public Health, School of Community Health Sciences and Nottingham Respiratory Biomedical Research Unit, University of Nottingham, Nottingham, UK 63Department of Respiratory Medicine, Ghent University Hospital, Ghent, Belgium 64MRC Lifecourse Epidemiology Unit, University of Southampton, Southampton, UK 65Chronic Disease Epidemiology, Swiss Tropical and Public Health (TPH) Institute, Basel, Switzerland 66University of Basel, Switzerland 67Non-Communicable Diseases Epidemiology Unit, Department of Epidemiology and Population Health, London School of Hygiene and Tropical Medicine, London, UK 68South Asia Centre for Chronic Disease, New Delhi, India 69Max Planck Institute for Psycholinguistics, Nijmegen, The Netherlands 70MRC Unit for Lifelong Health and Ageing, London, UK 71School of Social and Community Medicine, University of Bristol, Bristol, UK 72Hankinson Consulting, Inc, Athens, Georgia, USA 73Medical Genetics Section, The University of Edinburgh, Edinburgh, UK 74Department of Epidemiology and Public Health, University College London, London, UK 75Netherlands Genomics Initiative (NGI)-sponsored Netherlands Consortium for Healthy Aging (NCHA), The Netherlands 76Telethon Institute for Child Health Research and Centre for Child Health Research, The University of Western Australia, Perth, Western Australia, Australia 77School of Pathology and Laboratory Medicine, The University of Western Australia, Nedlands, Western Australia, Australia 78School of Population Health, University of Western Australia, Nedlands, Western Australia, Australia 79Molecular Medicine Centre, Institute of Genetics and Molecular Medicine, University of Edinburgh, Edinburgh, UK 80Division of Genetic Epidemiology, Department of Medical Genetics, Molecular and Clinical Pharmacology, Innsbruck Medical University, Innsbruck, Austria 81Department of Statistics, University of Oxford, Oxford, UK 82Medical Research Institute, The University of Dundee, Dundee, UK 83Department of Pulmonology, University Medical Center Groningen, University of Groningen, Groningen, The Netherlands 84Department of Internal Medicine, Erasmus MC, Rotterdam, The Netherlands 85Division of Pulmonary Medicine, University Hospitals of Geneva, Geneva, Switzerland 86Department of Medicine, University of New Mexico, Albuquerque, New Mexico, USA 87Geriatric Medicine Unit, The University of Edinburgh, Edinburgh, UK 88Department of Epidemiology, University Medical Center Groningen, University of Groningen, Groningen, The Netherlands 89Department of Primary Care and Population Health, University College London, London, UK 90Division of Population Health Sciences and Education, St George’s University of London, London, UK 91Department of Genetics, University Medical Center Groningen, University of Groningen, Groningen, The Netherlands 92Division of Preventive Medicine, University of Alabama at Birmingham, Birmingham, USA 93College of Pharmacy, University of Tennessee Health Science Center, Memphis, Tennessee, USA 94Centre for Clinical Research, Haukeland University Hospital, Bergen, Norway 95Obesity Research Unit, Department of Medicine, Division of Internal Medicine, Helsinki University Central Hospital, Helsinki, Finland 96Hjelt Institute, Department of Public Health, University of Helsinki, Helsinki, Finland 97Institute for Molecular Medicine, University of Helsinki, Helsinki, Finland 98Pédiatrie, Centre Hospitalier Universitaire (CHU), Grenoble, France 99Inserm U823, Centre de Recherche Albert Bonniot, Grenoble, France 100Université Joseph Fourier, Grenoble, France 101Gen-Info Ltd, Zagreb, Croatia 102Department of Children, Young People and Families, National Institute for Health and Welfare, Oulu, Finland 103Institute of Clinical Medicine, University of Oulu, Oulu, Finland 104Gerontology Research Centre, Department of Health Sciences, University of Jyväskylä, Jyväskylä, Finland 105Public
Abstract

Pulmonary function measures reflect respiratory health and predict mortality, and are used in the diagnosis of chronic obstructive pulmonary disease (COPD). We tested genome-wide association with the forced expiratory volume in 1 second (FEV$_1$) and the ratio of FEV$_1$ to forced vital capacity (FVC) in 48,201 individuals of European ancestry, with follow-up of top associations in up to an additional 46,411 individuals. We identified new regions showing association (combined $P < 5 \times 10^{-8}$) with pulmonary function, in or near MFAP2, TGFβ2, HDAC4, RARB, MECOM (EVI1), SPATA9, ARMC2, NCR3, ZKSCAN3, CDC123, C10orf11, LRP1, CCDC38, MMP15, CFDP1, and KCNE2. Identification of these 16 new loci may provide insight into the molecular mechanisms regulating pulmonary function and into molecular targets for future therapy to alleviate reduced lung function.

Introduction

Pulmonary function, reliably measurable by spirometry, is a heritable trait reflecting the physiological state of the airways and lungs$^1$. Pulmonary function measures are important predictors of population morbidity and mortality$^2$-$^4$, and are used in the diagnosis of chronic obstructive pulmonary disease (COPD), which ranks among the leading causes of death in developed and developing countries$^5$-$^6$. A reduced ratio of forced expiratory volume in 1 second (FEV$_1$) to forced vital capacity (FVC) is used to define airway obstruction, and a reduced FEV$_1$ is used to grade the severity of airway obstruction$^7$.

Recently, two large genome-wide association studies (GWAS), each comprising discovery sets of more than 20,000 individuals of European ancestry identified novel loci for lung
function\textsuperscript{8,9}. Recognizing the need for larger datasets to increase the power to detect loci of individually modest effect size, we conducted a meta-analysis of 23 lung function GWAS comprising a total of 48,201 individuals of European ancestry (Stage 1) and followed-up potentially novel loci in 17 further studies comprising up to 46,411 individuals (Stage 2). We identified 16 additional novel loci for lung function, and provided evidence corroborating association of loci previously associated with lung function\textsuperscript{8-11}. Our findings implicate a number of different mechanisms underlying regulation of lung function and highlight loci shared with complex traits and diseases, including height, lung cancer, and myocardial infarction.

**Results**

**Genome-wide analysis (stage 1)**

Meta-analyses for cross-sectional lung function measures were undertaken for approximately 2.5 million genotyped or imputed SNPs across 23 studies with a combined sample size of 48,201 adult individuals of European ancestry. Characteristics of the cohort participants and the genotyping are shown in Supplementary Tables 1A and 1B. FEV\textsubscript{1} and FEV\textsubscript{1}/FVC were adjusted for ancestry principal components, age, age\textsuperscript{2}, sex, and height as covariates. Association testing of the inverse-normal transformed residuals for FEV\textsubscript{1} and FEV\textsubscript{1}/FVC assumed an additive genetic model and was stratified by ever-smoking (versus never-smoking) status. Meta-analyses of the smoking strata within study, and the study-specific results, were undertaken using inverse variance weighting (the inverse of the standard error squared was used as the weight). We applied genomic control twice at study level (to each smoking stratum separately and to the study level pooled estimates) and also at meta-analysis level to avoid inflation of test statistics due to cryptic population structure or relatedness (see Supplementary Table 1A for study level estimates). Our application of genomic control at the three stages is likely to be overly conservative because it has recently been shown that in large meta-analyses, test statistics are expected to be elevated under polygenic inheritance even when there is no population structure\textsuperscript{12}. Test statistic inflation ($\lambda_{GC}$) prior to applying genomic control at meta-analysis level was 1.12 for FEV\textsubscript{1} and 1.09 for FEV\textsubscript{1}/FVC. Genomic inflation estimates increase with sample size, as has been shown for other traits\textsuperscript{13-15}; standardised estimates to a sample of 1000 individuals ($\lambda_{GC\_1000}$) were 1.002 for FEV\textsubscript{1} and 1.002 for FEV\textsubscript{1}/FVC. Plots of meta-analysis $P$-values for FEV\textsubscript{1} and FEV\textsubscript{1}/FVC against a uniform distribution of $P$ values expected under the null hypothesis showed deviations which were attenuated, but persisted, after removal of SNPs in loci reported previously, consistent with additional loci being associated with lung function (Supplementary Figure 1A).

**Follow-up analysis (stage 2)**

Twenty-nine new loci showing evidence of association with lung function ($P<3\times10^{-6}$) in Stage 1 were followed up in Stage 2 by utilizing in silico data from seven studies, and by undertaking additional genotyping in 10 studies for the 10 highest-ranked SNPs (Figure 1). Full details of the SNP selection are given in the Online Methods. Inverse variance weighting meta-analysis was performed across Stages 1 and 2, and two-sided $p$-values were obtained for the pooled estimates. Sixteen new loci reached genome-wide significance ($P<5\times10^{-8}$) and showed consistent direction of effects in both stages, comprising 12 new loci for FEV\textsubscript{1}/FVC, 3 new loci for FEV\textsubscript{1}, and one new locus reaching genome-wide significance for both traits (Figure 2, Table 1). To assess the heterogeneity across studies included in Stage 1 and Stage 2, Chi-square tests were undertaken for all 16 SNPs; and none of them was statistically significant after applying a Bonferroni correction for 16 tests. The sentinel SNPs at these loci were in or near the genes MFAP2 (1p36.13), TGFBR2/LYPLAL1 (1q41), HDAC4/FLJ43879 (2q37.3), RARB (3p24.2), MECOM (EVII) (3q26.2), SPATA9/
RHOBTB3 (5q15), ARMC2 (6q21), NCR3/AIF1 (6p21.33), ZKSCAN3 (6p22.1), CDC123 (10p13), C10orf11 (10q22.3), LRP1 (12q13.3), CCDC38 (12q22), MMP15 (16q13), CFDP1 (16q23.1) and KCNE2/C21orf82 (21q22.11) (Supplementary Figures 1B and 1C). The strongest signals in AGER (rs2070600)\(^8,9\) and two of the novel signals (rs6903823 in ZKSCAN3 and rs2857595, upstream of NCR3) lie within a 3.8Mb interval at 6p21.32-22.1 which is characterised by long-range linkage disequilibrium. Nevertheless, the leading SNPs in these regions which are within the major histocompatibility complex (MHC) were statistically independent (Supplementary Note).

**Gene expression**

We investigated mRNA expression of the nearest gene for each of the 16 novel loci in human lung tissue and in a range of human primary cells including lung, brain, airway smooth muscle cells and bronchial epithelial cells. Transcripts were detected for all selected genes in lung tissue except CCDC38 and transcripts for most genes were also detected in airway smooth muscle cells and in bronchial epithelial cells (Table 2). As we were unable to detect expression of CCDC38 in any tissue, we also examined expression of SNPRF, which is the adjacent gene (Table 2), and found expression in all four cell types. TGFB2, MFAP2, EVI1 and MMP15 were expressed in one or more lung cell types but not in peripheral blood mononuclear cells providing evidence that these genes may exhibit tissue-specific expression.

We assessed whether SNPs in these new regions, or their proxies \((r^2>0.6)\), were associated with gene expression using a database of expression-associated SNPs in lymphoblastoid cell lines\(^{16}\). Four loci showed regional \((cis)\) effects on expression \((P<1\times10^{-7})\), Supplementary Note). A proxy for our sentinel SNP in CFDP1, rs2865531, coincided with the peak of the expression signal for CFDP1 and the strongest proxy for rs6903823 in ZKSCAN3 coincided with the peak of expression for ZSCAN12.

**Plausible pathways for lung function involving new loci**

The putative function of the genes within, or closest to, the association peaks identify a range of plausible mechanisms for impacting lung function. The most statistically significant new signal for FEV\(_1\)/FVC \((P=7.5\times10^{-16})\) was in the gene encoding MFAP2, an antigen of elastin-associated microfibrils\(^{17}\), although correlated SNPs in the region potentially implicate other genes that could plausibly influence lung function, such as CROCC, which encodes rootletin, a component of cilia\(^{18}\). Our second strongest new signal, also for FEV\(_1\)/FVC, was in the gene encoding the retinoic acid receptor beta \((RARB)\). Rarb-null knockout mice exhibit premature alveolar septation\(^{19}\). The third most statistically significant new signal for FEV\(_1\)/FVC, and the most statistically significant new signal for FEV\(_1\), was in CDC123. This was the only novel region to show genome-wide association with both traits. CDC123 encodes a homologue of a yeast cell division cycle protein which plays a critical role in modulating Eukaryotic initiation factor 2 in times of cell stress\(^{20}\). The fourth signal for FEV\(_1\)/FVC is downstream of HDAC4 which encodes a histone deacetylase; reductions in the expression of other histone deacetylases (specifically HDAC2, HDAC5 and HDAC8) have been noted in COPD\(^{21}\). The regions we observed in the MHC are much more difficult to localize with multiple genes being tagged by the top SNP, including non-synonymous SNPs in ZKSCAN3, PGBD1, ZSCAN12, ZNF323, TCF19, LTA, C6orf15 and GPANK1 (also known as BAT4) (Supplementary Table 2). At 6p21.33, the strongest association with lung function was observed for rs2857595, which is in linkage disequilibrium \((LD, r^2=0.47)\) with a non-synonymous SNP in LTA (encoding lymphotoxin alpha) and with a SNP in the upstream promoter region of TNFA (encoding tumour necrosis factor alpha, \(r^2=0.86)\), both of which are plausible candidates\(^{22,23}\). Our top SNP in MMP15 is in strong LD \((r^2=1)\) with a non-synonymous SNP (rs3743563, which has an association with FEV\(_1\)/FVC at
\( P = 1.8 \times 10^{-7} \) within the same gene. Plausible mechanisms implicated by the other novel signals of association with lung function reported here include TGF-beta signalling; \( TGFB2 \) expression is upregulated in bronchial epithelial cells in asthma\(^{24}\). The putative function of key genes (as defined by LD with the leading SNP) in each of the 16 loci, and relevant findings from animal models, are summarised in Table 2 and detailed in Supplementary Table 2.

**Associations with lung function in children**

Alleles representing 11 of the 16 novel loci showed directionally consistent effects on lung function in 6,281 children (7 to 9 years of age) (Supplementary Table 3A) suggesting that genetic determination of lung function in adults may in part act via effects on lung development, or alternatively, that some genetic determinants of lung growth and lung function decline are shared.

**Association of lung function loci with other traits**

Although we stratified for ever-smoking versus never-smoking, we did not adjust for the amount smoked. In order to investigate the possibility that the associations at any of our 16 novel regions were driven by an effect of the SNP on smoking behaviour, we evaluated *in silico* data for associations with smoking amount from the Ox-GSK consortium\(^{25}\) for the leading SNPs in these 16 regions. None of these 16 SNPs showed statistically significant association with the number of cigarettes smoked per day (Supplementary Table 3B).

In addition, in our Stage 1 and Stage 2 datasets combined, we assessed whether the estimated effect sizes of the variants on lung function phenotypes differed substantially between ever-smokers and never-smokers (Supplementary Table 4) across the 16 loci. For the most strongly associated trait at each locus, we tested the SNP interaction with ever-smoking (versus never-smoking). None of the 16 novel loci showed a significant interaction (Bonferroni corrected threshold for 16 independent SNPs \( P = 0.003125 \)). These analyses suggest that the genetic effects we have identified underlie lung function variability irrespective of smoking exposure.

Our lung function associations were adjusted for height, but there are some overlaps between loci associated with height and those associated with lung function. Therefore, we evaluated *in silico* data for height associations of our novel regions in the GIANT consortium\(^{14}\) dataset. The G allele of rs2284746 (MFAP2, intron), which was associated with decreased FEV\(_1\)/FVC was associated with increased height (Supplementary Table 3C).

Given reported associations between lung cancer and either COPD or lung function decline, we also assessed *in silico* data for sentinel or proxy SNPs in these 16 regions for associations with lung cancer in the International Lung Cancer Consortium (ILCCO) GWAS meta-analysis\(^{26}\). Alleles associated with reduced lung function were associated with risk of lung cancer at the strongest available proxy SNP for rs2857595 (upstream of \( NCR3 \)) at 6p21.33 (rs3099844, \( r^2 = 0.67 \)), and the strongest proxy SNP for rs6903823 (SNP in intron of \( ZKSCAN3 \) and \( ZNF323 \)) at 6p22.1 (rs209181, \( r^2 = 0.69 \)) (lung cancer associations, \( P = 2.2 \times 10^{-7} \) and \( P = 3.4 \times 10^{-5} \), respectively, Supplementary Table 3D). No significant associations with lung cancer were seen at the other new loci (proxy SNPs were available for 15 of the 16 loci, Bonferroni corrected \( P < 0.0033 \)).

In addition to the effects on height, smoking and lung cancer described above, we examined the literature for evidence for associations with other traits for each of the 16 new loci (detailed in Supplementary Table 2). Genome-wide significant associations (\( P < 5 \times 10^{-8} \)) have been reported in \( KCNE2 \) with myocardial infarction\(^{27}\), and at 6p21.33 near \( NCR3/ AIF1 \), with neonatal lupus\(^{28}\) and with systemic lupus erythematosus\(^{29}\). Other significant
complex disease associations have also been noted in the regions of CDC123 (type 2 diabetes\textsuperscript{30}), CFDP1 (type 1 diabetes\textsuperscript{31}) and MECOM (blood pressure\textsuperscript{32,33}), but with weaker LD ($r^2<0.3$) between the reported SNP and the sentinel SNP for lung function in the region (Supplementary Table 2).

**Proportion of variance explained by loci discovered to date**

Associations in 10 loci previously reported for lung function\textsuperscript{8,9} reached genome-wide significance ($P<5\times10^{-8}$) in our Stage 1 data, namely loci in or near TNS1, FAM13A, GSTCD/NPNT, HHIP, HTR4, ADAM19, AGER, GPR126, PTCH1, and TSHD4 (Supplementary Table 5A). Thus, a total of 26 regions showed genome-wide significant association with lung function in our study. In aggregate, variants at these 26 regions explain approximately 3.2% of the additive polygenic variance for FEV\textsubscript{1}/FVC and 1.5% for FEV\textsubscript{1} (see Supplementary Note). Following the approach described by Park et al.\textsuperscript{34} we estimated that there is a total of 102 (95% CI 57-155) independent variants with similar effect sizes to the 26 variants we report. In combination these 102 variants, comprising 26 discovered variants and 76 putative undiscovered variants, collectively explain around 7.5% of the additive polygenic variance for FEV\textsubscript{1}/FVC and 3.4% for FEV\textsubscript{1} (see Supplementary Table 6, Online Methods and Supplementary Note).

**Discussion**

In meta-analysis of 23 studies comprising 48,201 individuals of European ancestry and follow-up in 17 studies comprising up to 46,411 individuals, we report genome-wide significant associations with an additional 12 regions for FEV\textsubscript{1}/FVC, an additional three regions for FEV\textsubscript{1} and one additional region associated with both FEV\textsubscript{1} and FEV\textsubscript{1}/FVC. We also confirm genome-wide association with 10 regions previously associated with lung function, bringing to 26 the total number of loci associated with lung function in these data. Most of the new loci are in regions not previously suspected to have been involved in lung development, the control of pulmonary function or risk of developing COPD. Elucidating the mechanisms through which these regions influence lung function should lead to a more complete understanding of lung function regulation and the pathogenesis of COPD. Four of the new loci (MFAP2, ZKSCAN3, near NCR3 and near KCNE2) we show to be associated with lung function are also associated with other complex traits and diseases ($P<5\times10^{-8}$ for the other trait at a SNP with $r^2>0.3$ with the top lung function SNP in the region).

Understanding the intermediates underlying these pleiotropic effects could also reveal crucial insights into the pathophysiology of lung disease. One potential explanation is that these loci underlie control of the mechanisms regulating the development and resolution of inflammation and subsequent tissue remodelling in a range of tissues.

The effect sizes of the variants in the 26 loci associated with lung function explain a modest proportion of the additive genetic variance in FEV\textsubscript{1}/FVC and in FEV\textsubscript{1}, even after accounting for putative undetected variants with a similar distribution of effect sizes\textsuperscript{34}. Our findings are consistent with those from other common complex traits, where it is thought that many as yet unidentified common and rare sequence variants, and potentially structural variants could explain the remaining heritability\textsuperscript{35}. That our study more than doubled the number of loci known to be associated with lung function underlines the utility of large sample sizes to achieve the power to detect common variants associated with complex traits. Nevertheless, it is likely that additional variants with similar effect sizes remain undiscovered\textsuperscript{14}. In addition, our study was not designed to detect rare variants or structural variants associated with lung function. Identification of rare variants associated with lung function could be helpful in narrowing the scope of ongoing functional work to those genes most likely to be causally related to the association signals we detected.
Our study focused on cross-sectional measures of lung function. Adult lung function at a particular time point is influenced by the peak lung function achieved by 25-35 years of age, as well as the rate of decline of lung function after that peak. The 26 loci now confirmed to be associated with lung function could affect either pre- or post-natal lung development and growth or decline in lung function during adulthood, or both. We showed consistent directions of estimated effects on lung function between adults and children at 7-9 years of age for SNPs at 11 of the 16 new loci, and eight of 10 previously reported loci (Supplementary Table 3A). The results we show for lung function in children provide some indication that these loci affect lung function development, although studies in larger populations of children would provide greater clarity for SNPs in the new loci. Further investigations will be required in large populations with longitudinal data to delineate the influence of these variants on the rates of development of, and decline in, lung function and on the risk of developing COPD.

Of the sentinel SNPs at the 16 new loci associated with lung function, only rs2284746 (MFAP2) was associated with height in the GIANT consortium dataset. The G allele of rs2284746 was associated with both increased height and reduced lung function. A similar relationship between lung function and height was previously reported for the G allele of rs3817928 in GPR126, which is associated with decreased height, but with increased FEV1/FVC. A further three of the 180 loci found to be associated with height showed association (for 180 loci, Bonferroni corrected threshold $P=2.8\times10^{-4}$) with either FEV1 (CLIC4 and BMP6) or FEV1/FVC (PIP4K2B) (Supplementary Table 3E). In each case, the allele associated with an increase in height was associated with a decrease in lung function. This is not the case for the association of rs1032296 near HHIP, which has shown consistent directions of effects on lung function and height. However, the strongest SNP associated with height in the HHIP region lies within an intron of HHIP but shows no association with FEV1 or FEV1/FVC. Furthermore, while height is an important predictor of FEV1, this is not true for its ratio to FVC. These observations argue against the associations with lung function at these loci being simply due to incomplete adjustment for height.

We stratified by ever- and never-smoker status in our analyses and in our investigation of amount smoked in the Ox-GSK consortium none of the sentinel SNPs in the 16 new regions showed association with the number of cigarettes smoked per day. Additionally, none of these regions was associated with ever-smoking in the Ox-GSK consortium data (Supplementary Table 3B). Thus the SNP associations with lung function we observed are unlikely to have arisen simply as a consequence of inadequate adjustment for smoking.

We did not observe any interactions with ever-smoking for any of the sentinel SNPs in the 16 new regions that exceeded a Bonferroni-corrected significance level (for 16 SNPs). Thus, the effects on lung function of the novel variants we identified are apparent in both ever-smokers and in never-smokers, and the effects of smoking and of these genetic variants may be independent and additive.

In other common complex diseases, follow up studies that incorporate common genetic risk variants into models to predict disease have not been shown to add substantially to existing risk models, particularly when such models already include family history. The same may also prove to be true for the 26 genetic variants described in this paper, as the effect size of any individual variant is small, but further work is required in this area. The major utility of our findings will be in the knowledge they provide about previously unknown pathways underlying lung function. Elucidating the mechanisms that these genes are involved in will lead to improved understanding of the regulation of lung function and potentially to new therapeutic targets for COPD.
Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

We thank the many colleagues who contributed to collection and phenotypic characterization of the clinical sampling, genotyping, and analysis of the data. We especially thank those who kindly agreed to participate in the studies.

Major funding for this work is from the following sources (alphabetical): Academy of Finland (project grants 104781, 120315, 129269, 1114194, Center of Excellence in Complex Disease Genetics (213506 and 129680) and SALVE); Althingi (Icelandic Parliament); Arthritis Research Campaign; Asthma UK; AstreaZeneca; AXA Research Fund; Biotechnology and Biological Sciences Research Council (BBSRC) (BB/F019394/1, G20234); British Heart Foundation (PG/07/012, PG/06/154/22043, FS/05/125); British Lung Foundation; Canadian Institutes of Health Research (Grant ID MOP-82893); Cancer Research United Kingdom; Chief Scientist Office, Scottish Government Health Directorate (CZD/166); Croatian Institute for Public Health; UK Department of Health; Dutch Kidney Foundation; Erasmus Medical Center and Erasmus University, Rotterdam; Estonian Genome Center, University of Tartu, Estonia (SF0180142208); EU funding (GABRIEL GRANT Number: 081996, ECRHS II Coordination Number: QLK4-CT-1999-01237); European Commission (DG XII, EURO-BLCS, FP-5 QLGI1-CT-2000-01643, FP-6 LSHB-CT-2006-018996 GABRIEL), FP-6 LSHG-CT-2006-018947 (EUROSPAN), FP-6 GenomEUtwin project QLG2-CT-2000-01254, FP-7/2007-2013: HEALTH-F2-2008-201865, GEFOS, HEALTH-F4-2008-35627, TREAT-OA, HEALTH-F4-2007-201413 (ENGAGE); Finnish Foundation for Cardiovascular Research; Flight Attendant Medical Research Institute (FAMRI); German Asthma and COPD Network (COSYCONET: BMBF grant 01GO883); German Bundesministerium fuer Forschung und Technologie (01 AK 803 A-H, 01 IG 07015 G); German Federal Ministry of Education and Research (BMBF) (03ZIK012, 01ZZ9603, 01ZZ0103, and 01ZZ0403); German National Genome Research Network (NGFN-2 and NGFN-plus); German Ministry of Cultural Affairs; GlaxoSmithKline; Gyllenberg Foundations; Healthway, Western Australia; Helmholtz Zentrum Munich, German Research Center for Environmental Health, Neuherberg, Germany; Healthcare and Bioscience iNet (funded by the East Midlands Development Agency, part-financed by the European Regional Development Fund, delivered by Medlink East Midlands); Higher Education Funding Council for England (HEFCE); Hjartarven (Icelandic Heart Association); Innsbruck Medical University; Institute for Anthropological Research in Zagreb; International Osteoporosis Foundation; Intramural Research Program of the NIH, National Institute on Aging and National Institute of Environmental Health Sciences; Jalmari and Rauha Ahokas Foundation; Juvenile Diabetes Research Foundation International (JDRF); Lifelong Health and Wellbeing Initiative (G0700704/84698); Medical Research Council UK (G1000861, G0501942, G0902313, G0000934, G0800582, G050539, G0600705, PrevMetSyn/ SALVE, G9901462); Medical Research Fund of the Tampere University Hospital; Ministry of Science, Education and Sport of the Republic of Croatia (108-1080315-0302); Medical Research Council Human Genetics Unit; Medisearch-The Leicester Medical Research Foundation; Munich Center for Health Sciences (MC Health) as part of LMIinnovativ; National Health and Medical Research Council of Australia (Grant ID 403981 and ID 003209); National Human Genome Research Institute (NHGRI) (U01-HG-004729, U01-HG-004402); National Institute for Health Research (NIHR) Comprehensive Biomedical Research Centres (Guy’s & St Thomas’ NHS Foundation Trust in partnership with King’s College London and Cambridge University Hospitals NHS Foundation Trust in partnership with the University of Cambridge); Netherlands Genomics Initiative (Netherlands Organisation for Scientific Research (NWO) (050-060-810); Netherlands Organization for the Health Research and Development (ZonMw); Netherlands Organization of Scientific Research NOW (175.01.02007006, 175.01.02005.01, 911-03-012); Northern Netherlands Collaboration of Provinces (SNN); Norwegian University of Science and Technology; Novo Nordisk; Ontario Institute of Cancer Research and Canadian Cancer Society Research Institute (CCSR1 020214); Republic of Croatia Ministry of Science, Education and Sports research grants (108-1080315-0302); Research Institute for Diseases in the Elderly (RIDE) (014-93-015: RIDE2); Research Into Ageing (251); Siemens Healthcare, Erlangen, Germany and the Federal State of Mecklenburg-West Pomerania; Social Ministry of the Federal State of Mecklenburg-West Pomerania; Structure Enhancing Fund (FES) of the Dutch government; Swedish Heart and Lung Foundation Grant 20050561; Swedish Research Council for Worklife and Social research (FAS), grants 2001-0263, 2003-0139; Swiss National Science Foundation (grant no 4026-28099.3347/CO-108796, 3247/BO-104283, 3247/BO-104288, 3247/BO-104284, 32-68061.01, 32-59302.99, 32-52720.97, 32-4253.94); The Asthma, Allergy and Inflammation Research Trust; The Great Wine Estates of the Margaret River region of Western Australia (SF0180142208); The Netherlands’ Ministry of Education, Culture and Science and Ministry for Health, Welfare and Sports; The Royal Society; The University of Split and Zagreb Medical Schools; Tromsø University; U01 DK062418; UBS Wealth Foundation Grant BA29s8Q7-DDZ; UK Department of Health Policy Research Programme; University Hospital Oulu, Biocenter, University of Oulu, Finland (75617); University Medical Center Groningen; University of Bristol; University of Leicester HEFCE CIF award; University of Nottingham; US National Institute of Health (NIH) (1P50 CA70907, RO1 CA121197, U19 CA148127, CA55769, CA17219, R01HL071051, R01HL071205, R01HL071250, R01HL071252, R01HL071258, R01HL071259, UL1RR025005, contracts HHSN268200625226C, HHSN268200782096C, R01-HL084099); US NIH National Cancer Institute (RO1CA111703); US NIH National...
APPENDIX

Online Methods

Study design

The study consisted of two stages. Stage 1 was a meta-analysis conducted on directly genotyped and imputed SNPs from individuals of European ancestry in 23 studies, with a total of 48,201 individuals. Supplementary Table 1A gives details of these studies. Thirty-four SNPs selected according to the results in Stage 1 were followed up in Stage 2. The ten leading SNPs were followed up in up to 46,411 individuals of European origin and the remaining 24 SNPs were followed up in a subset of up to 21,674 individuals (Figure 1).

Stage 1 samples

A total of 23 studies, 17 from the SpiroMeta consortium and six studies from the CHARGE consortium, formed Stage 1: AGES, ARIC, B58C T1DGC, B58C WTCCC, BHS1, CHS, ECRHS, EPIC (obese cases and population-based studies), the EUROSPAN studies (CROATIA-Korcula, ORCADES and CROATIA-Vis), FHS, FTC (incorporating the FinnTwin16 and Finnish Twin Study on Aging), Health 2000, Health ABC, KORA F4, KORA S3, NFBC1966, RS-I, RS-II, SHIP and TwinsUK-I (see Supplementary Table 1 for definitions of abbreviations). Measurements of spirometry for each study are described in the Supplementary Note. The genotyping platforms and quality-control criteria implemented by each study are described in Supplementary Table 1B.

Imputation

Imputation of non-genotyped SNPs was undertaken with MACH47, IMPUTE48 or BIMBAM49 with pre-imputation filters and parameters as shown in Supplementary Table 1B. SNPs were excluded if the imputation information, assessed using r^2.hat (MACH), .info (IMPUTE) or OEvar (BIMBAM), was <0.3. In total, 2,706,349 SNPs were analyzed.

Transformation of data and genotype-phenotype association analysis

Linear regression of age, age^2, sex, height and ancestry principal components was undertaken on FEV₁ (milliliters) and FEV₁/FVC (percent). The residuals were transformed to ranks and then transformed to normally distributed z-scores. These transformed residuals were then used as the phenotype for association testing under an additive genetic model, separately for ever-smokers and for never-smokers. The software used is specified in Supplementary Table 1B. Appropriate tests for association in related individuals were applied where necessary, as described in the Supplementary Note.
Meta-analysis of Stage 1 data

All Stage 1 study effect estimates, both for ever-smokers and never-smokers were corrected using genomic control and were oriented to the forward strand of the NCBI build 36 reference sequence of the human genome, consistently using the alphabetically higher allele as the coded allele. Study-specific lambda estimates are shown in Supplementary Table 1. For each study, effect estimates and standard errors for ever-smokers and never-smokers were meta-analysed using inverse variance weighting. Genomic control was applied again to the pooled effect-size estimates for each study. Finally effect-size estimates and standard errors were combined across studies using inverse variance weighting meta-analysis, and genomic control was applied to the pooled effect-size estimates. To describe the effect of imperfect imputation on power, for each SNP we report the effective sample size ($N_{\text{effective}}$); the sum of the study-specific products of the sample size and the imputation quality metric. Meta-analysis statistics and figures were produced using R version 2.9.2.

Selection of SNPs for Stage 2

All regions selected for follow-up in Stage 2 contained a lead SNP with novel evidence of association (all with $P<3\times 10^{-6}$) with FEV$_1$ and/or FEV$_1$/FVC, an $N_{\text{effective}} \geq 70\%$ of the total Stage 1 sample size and the association signals from surrounding SNPs were consistent with their correlation (linkage disequilibrium) with the leading SNP. Twenty-nine independent regions with a leading SNP meeting these criteria were assessed in Stage 2. Regions were defined as independent if the leading SNP from one region was >500kb from the leading SNP of any other region. Long range linkage disequilibrium was also investigated between leading SNPs of regions in or near the MHC on chromosome 6 (Supplementary Note). For two regions, the leading SNP had an $N_{\text{effective}} \geq 70\%$ but $<80\%$ of the Stage 1 sample size and therefore a proxy SNP ($r^2=1$ and 0.97) was also taken forward. For three regions, there were different leading SNPs for FEV$_1$ and FEV$_1$/FVC and so both leading SNPs were assessed. A total of 34 SNPs were analysed in Stage 2 and are listed in Supplementary Table 5B. Previously reported regions were not followed up. We present association test statistics in Stage 1 only for relevant SNPs from previously reported regions in Supplementary Table 5A.

Stage 2 samples

The 34 SNPs were followed up in up to 11,275 individuals from seven studies with in silico data; CARDIA, CROATIA-Split, LifeLines, LBC1936, MESA-Lung, RS-III and TwinsUK-II (Supplementary Table 1). SNP rs2647044 was not available from TwinsUK-II.

The 34 SNPs were ranked by $P$ value (for association with either FEV$_1$ or FEV$_1$/FVC) and the top ten leading SNPs selected for follow up by genotyping in up to 35,136 individuals from ADONIX, BHS2, BRHS, BWHHS, Gedling, GS:SFHS, HCS, Nottingham Smokers, NSHD and SAPALDIA (Supplementary Table 1). If a SNP within the top ten had an $N_{\text{effective}} <80\%$, only the proxy SNP was included in the top ten for follow up. For regions which showed association with both FEV$_1$ and FEV$_1$/FVC, only the leading SNP with the lowest $P$ value for either trait was included if it was within the top ten SNPs. The study design is illustrated in Figure 1.

Meta-analysis of Stage 2 data

All Stage 2 studies provided effect estimates for ever-smokers and never-smokers, apart from Nottingham Smokers since they only had smokers. Studies with family data (BHS2 and GS:SFHS) analysed ever- and never-smokers together to account for the family correlation, adding the smoking status as a covariate in the model, and therefore provided smoking adjusted effect estimates. All Stage 2 study effect estimates were oriented to the
forward strand of the NCBI build 36 reference sequence of the human genome, consistently using the alphabetically higher allele as the coded allele. For each study with separate results for ever- and for never-smokers, effect estimates and standard errors for ever- and never-smokers were meta-analysed using inverse variance weighting. Genomic control was applied to the pooled effect sizes of those studies with in silico data that undertook the analysis genome-wide. Effect estimates and standard errors were combined across Stage 2 studies using inverse variance weighting meta-analysis.

Combined analysis of Stage 1 and Stage 2 samples

Meta-analysis of Stage 1 and Stage 2 results was undertaken using inverse variance weighting. We described associations as genome-wide significant if $P < 5 \times 10^{-8}$.

PCR expression profiling

The mRNA expression profiles of $\text{TGFB2, MFAP2, HDAC4, EVII, RARB, SPATA9, ARMC2, NCR3, CDC123, LRP1, CCDC38, SNRPF, MMP15, CFDP1, ZKSCAN3, KCNE2}$ and $\text{C10orf11}$ were determined in human lung tissue and primary cell samples using RT-PCR, including RNA from lung (Ambion/ABI), brain, airway smooth muscle cells and human bronchial epithelial cells (Clonetics42). Primer sequences are listed in Supplementary Table 2. Full details are provided in the Supplementary Note.

Lung function associations in our data of SNPs previously associated with lung function

In order to permit comparison of findings with recent studies of relevance to the field, we present association test statistics (in Stage 1 only) for relevant SNPs from previously reported regions (Supplementary Table 5A). We included regions: (i) reported as showing genome-wide significant association ($P < 5 \times 10^{-8}$) with lung function; (ii) reported as showing genome-wide significant association with COPD, providing that there was additional evidence of association with lung function and; (iii) $\text{DAAM2}$, which reached borderline significance in the SpiroMeta consortium$^9$. Within each of these regions, if multiple SNPs had been reported, we included all relevant SNPs and also the SNP that showed the strongest association in our data.

Association to other traits of lung function associated SNPs

Regions associated ($P < 5 \times 10^{-8}$) with lung function or COPD (and also associated with lung function) were looked up for other traits. Where multiple SNPs were reported for different traits or by different investigators, we aimed to include all relevant SNPs, except those with $r^2 > 0.9$ with another SNP in the region. We also included the SNPs that showed the strongest association in our data for each region. The following related traits were assessed: (i) lung function in children (Supplementary Table 3A); (ii) smoking amount and ever-smoking versus never-smoking in the Ox-GSK consortium$^{25}$ dataset (Supplementary Table 3B); (iii) height in the GIANT consortium$^{14}$ dataset (Supplementary Table 3C and 3E) and (iv) lung cancer in the International Lung Cancer Consortium (ILCCO) GWAS meta-analysis$^{26}$ (Supplementary Table 3D).

Estimation of the number of undiscovered variants and calculation of the proportion of variance explained

We used the approach proposed by Park et al.$^{34}$ to estimate the number of independent variants associated with lung function measures that have similar effect sizes to the variants already reported, and to calculate the proportion of the variance explained by them. We excluded discovery data when estimating effect sizes to avoid winner’s curse bias, and obtained the number of undiscovered variants using the discovery power to detect the unbiased effect sizes (Supplementary Table 6 and Supplementary Note).
Additional analyses

The top SNPs from our novel loci, and proxies, were searched for correlation with known common copy number variants and expression SNPs. Analyses to identify common pathways underlying the association signals for lung function were undertaken using MAGENTA v2\textsuperscript{53} and GRAIL\textsuperscript{54}. Full methods and results are given in the Supplementary Note.

References


Nat Genet. Author manuscript; available in PMC 2012 May 1.


Figure 1.
Study design. A total of 34 SNPs showing novel evidence of association ($P<3\times10^{-6}$) with FEV$_1$ and/or FEV$_1$/FVC in a meta-analysis of the Stage 1 studies were followed up in Stage 2. Studies with a combined total of 24,737 individuals undertook genotyping and association testing of the top 10 SNPs. Seven studies (*) with a combined total of 11,275 individuals had genome-wide association data and provided results for up to 34 SNPs. GS:SFHS undertook genotyping on a 32-SNP multiplex genotyping platform and so included the 32 top ranking SNPs (including proxies and both SNPs from regions which showed association with both FEV$_1$ and FEV$_1$/FVC). This assay failed for one SNP (rs3769124) which was subsequently replaced with the 33rd SNP (rs4762767). SNP rs2284746 was excluded due to poor clustering. Although rs3743563 was chosen as proxy for rs12447804 which had N effective <80% in the Stage 1 meta-analysis, BHS2 were unable to genotype rs3743563 and so undertook genotyping for rs12447804 instead. See Table 1 for definitions of abbreviations.
Figure 2.
Manhattan plots of association results for (a) FEV₁/FVC and (b) FEV₁. Manhattan plots ordered by chromosome position. SNPs for which $-\log_{10}P > 5$ are indicated in red. Novel regions which reached genome-wide significance after Stage 1 + Stage 2 are labelled.
Loci associated with lung function. Shown are FEV$_1$ and FEV$_1$/FVC results for the leading SNPs, ordered by chromosome and position, for each independent locus associated ($P < 5 \times 10^{-8}$) with FEV$_1$ or FEV$_1$/FVC in a joint analysis of up to 94,612 individuals of European ancestry from the SpiroMeta-CHARGE GWAS (Stage 1) and follow-up (Stage 2). Two-sided $p$-values are given for Stage 1, Stage 2 and the joint meta-analysis of all stages. $p$-values reaching genome-wide significance ($P < 5 \times 10^{-8}$) in the joint meta-analysis of all stages are indicated in bold. SNPs reaching independent replication in Stage 2 ($P=0.05/34 = 1.47 \times 10^{-3}$) are indicated with their Stage 2 $p$-value in bold. The sample sizes ($N$) shown are the effective sample sizes. Effective sample size within each study is the product of sample size and imputation quality metric. Joint meta-analysis includes data from Stage 1 and Stage 2. Beta values reflect effect-size estimates on an inverse-normal transformed scale after adjustments for age, age$^2$, sex, height and ancestry principal components. The estimated proportion of the variance explained by each SNP can be found in Supplementary Table 6.

<table>
<thead>
<tr>
<th>Chr.</th>
<th>SNP_ID/NCBI36 position (nt), function</th>
<th>Coded allele freq.</th>
<th>MEASURE</th>
<th>Stage 1</th>
<th>Stage 2</th>
<th>Joint meta-analysis of all stages</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td>Beta (s.e.m.)</td>
<td>N</td>
<td>Beta (s.e.m.)</td>
<td>N</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>P</td>
<td></td>
<td>P</td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>rs2284746(17179262), MFAP2(intron)</td>
<td>G</td>
<td>−0.042 (0.007)</td>
<td>2.47E-09</td>
<td>−0.038 (0.007)</td>
<td>2.64E-07</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>0.516</td>
<td>4944</td>
<td>0.006 (0.007)</td>
<td>3.70E-01</td>
</tr>
<tr>
<td>1</td>
<td>rs93925(216926691), TGFB2(downstream)</td>
<td>T</td>
<td>0.04 (0.007)</td>
<td>2.54E-07</td>
<td>0.023 (0.01)</td>
<td>1.76E-02</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>0.308</td>
<td>42402</td>
<td>0.003 (0.007)</td>
<td>7.29E-01</td>
</tr>
<tr>
<td>2</td>
<td>rs1247731(423954385), HDAC4(downstream)</td>
<td>T</td>
<td>0.025 (0.008)</td>
<td>4.48E-09</td>
<td>0.031 (0.008)</td>
<td>8.41E-05</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>0.202</td>
<td>45585</td>
<td>0.025 (0.007)</td>
<td>1.82E-04</td>
</tr>
<tr>
<td>3</td>
<td>rs1529672(224595386), RARB(ntroon)</td>
<td>C</td>
<td>−0.06 (0.009)</td>
<td>7.75E-10</td>
<td>−0.038 (0.009)</td>
<td>1.16E-05</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>0.829</td>
<td>40624</td>
<td>−0.011 (0.007)</td>
<td>9.33E-02</td>
</tr>
<tr>
<td>3</td>
<td>rs144555(117028193), MECOM(ntroon)</td>
<td>T</td>
<td>−0.019 (0.008)</td>
<td>2.61E-02</td>
<td>−0.017 (0.012)</td>
<td>1.55E-01</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>0.205</td>
<td>40667</td>
<td>−0.025 (0.009)</td>
<td>6.44E-03</td>
</tr>
<tr>
<td>5</td>
<td>rs153916(95062456), SPATA9(upstream)</td>
<td>T</td>
<td>−0.033 (0.007)</td>
<td>2.06E-06</td>
<td>−0.025 (0.009)</td>
<td>6.67E-03</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>0.532</td>
<td>47350</td>
<td>0.004 (0.007)</td>
<td>6.22E-01</td>
</tr>
<tr>
<td>6</td>
<td>rs803832(23403275), ZKSCAN3(ntroon)</td>
<td>G</td>
<td>−0.027 (0.008)</td>
<td>2.28E-03</td>
<td>−0.013 (0.013)</td>
<td>2.34E-01</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>0.209</td>
<td>47057</td>
<td>−0.029 (0.008)</td>
<td>4.73E-04</td>
</tr>
<tr>
<td>6</td>
<td>rs285759(13676448), NCR3(utroon)</td>
<td>G</td>
<td>0.049 (0.009)</td>
<td>7.68E-08</td>
<td>0.026 (0.008)</td>
<td>5.36E-04</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>0.809</td>
<td>45540</td>
<td>0.007 (0.007)</td>
<td>9.41E-03</td>
</tr>
<tr>
<td>6</td>
<td>rs2798641(10374713), ARMCl(ntroon)</td>
<td>T</td>
<td>−0.047 (0.009)</td>
<td>2.81E-07</td>
<td>−0.03 (0.012)</td>
<td>1.57E-02</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>0.183</td>
<td>46369</td>
<td>−0.009 (0.01)</td>
<td>3.33E-01</td>
</tr>
<tr>
<td>10</td>
<td>rs7065966(22137998), CDC123(ntroon)</td>
<td>T</td>
<td>0.045 (0.007)</td>
<td>1.28E-10</td>
<td>0.023 (0.006)</td>
<td>3.86E-04</td>
</tr>
<tr>
<td>Chr.</td>
<td>SNP_ID (NCBI36 position), function</td>
<td>Coded allele freq.</td>
<td>Measure</td>
<td>Stage 1</td>
<td>Stage 2</td>
<td>Joint meta-analysis of all stages</td>
</tr>
<tr>
<td>------</td>
<td>----------------------------------</td>
<td>------------------</td>
<td>---------</td>
<td>---------</td>
<td>---------</td>
<td>----------------------------------</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Beta (s.e.m.)</td>
<td>P</td>
<td>Beta (s.e.m.)</td>
<td>P</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Coded allele freq.</td>
<td></td>
<td>Coded allele freq.</td>
<td></td>
</tr>
<tr>
<td>10</td>
<td>rs11001819(77985320), C10orf11(intron)</td>
<td>G</td>
<td>FEV1</td>
<td>0.04 (0.007)</td>
<td>1.19E-08</td>
<td>0.022 (0.005)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>FEV1/PVC</td>
<td>0.019 (0.007)</td>
<td>6.50E-03</td>
<td>-0.006 (0.006)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>FEV1</td>
<td>0.04 (0.007)</td>
<td>1.42E-08</td>
<td>-0.022 (0.005)</td>
</tr>
<tr>
<td>12</td>
<td>rs11172113(55813550), LRPI(intron)</td>
<td>T</td>
<td>FEV1/PVC</td>
<td>-0.035 (0.007)</td>
<td>1.36E-06</td>
<td>-0.026 (0.007)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>FEV1</td>
<td>-0.021 (0.007)</td>
<td>3.55E-03</td>
<td>-0.003 (0.007)</td>
</tr>
<tr>
<td>12</td>
<td>rs3036429(49595559), CCDC38(intron)</td>
<td>T</td>
<td>FEV1/PVC</td>
<td>0.049 (0.008)</td>
<td>1.24E-08</td>
<td>0.028 (0.008)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>FEV1</td>
<td>0.01 (0.008)</td>
<td>2.67E-01</td>
<td>0.004 (0.006)</td>
</tr>
<tr>
<td>16</td>
<td>rs12447804(96632783), MMP15(intron)</td>
<td>T</td>
<td>FEV1/PVC</td>
<td>-0.053 (0.009)</td>
<td>7.12E-08</td>
<td>-0.021 (0.007)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>FEV1</td>
<td>-0.017 (0.009)</td>
<td>8.02E-02</td>
<td>0.004 (0.007)</td>
</tr>
<tr>
<td>16</td>
<td>rs2665531(7394787), CFDP1(intron)</td>
<td>T</td>
<td>FEV1/PVC</td>
<td>0.039 (0.007)</td>
<td>2.30E-08</td>
<td>0.024 (0.006)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>FEV1</td>
<td>0.024 (0.007)</td>
<td>6.30E-04</td>
<td>0.011 (0.005)</td>
</tr>
<tr>
<td>21</td>
<td>rs978142(34574109), KCNE2(upstream)</td>
<td>T</td>
<td>FEV1/PVC</td>
<td>-0.048 (0.009)</td>
<td>8.23E-07</td>
<td>-0.031 (0.013)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>FEV1</td>
<td>-0.012 (0.009)</td>
<td>2.47E-01</td>
<td>-0.015 (0.013)</td>
</tr>
</tbody>
</table>
Expression profiling of candidate genes in the lung and periphery. (+) indicates the gene is expressed in the cell type used and (−) indicates the gene expression at mRNA level is not detected following 40 cycles of PCR. PCR profiling of gene transcripts in the human lung demonstrates expression of all candidates except \( \text{CCDC38} \), for which two sets of primers were designed and tested under different optimization conditions. None of these assays detected expression of \( \text{CCDC38} \) in the cell types analysed. The \( \text{SNRPF} \) gene neighbouring \( \text{CCDC38} \) and harbouring SNPs with strong LD with \( \text{CCDC38} \)'s sentinel SNP was assayed instead. All PCR products were sequence verified. Glyceraldehyde-3-phosphate dehydrogenase gene (\( \text{GAPDH} \)) was used as a positive control for the cDNA and was expressed in all tissues. HASM: Human airway smooth muscle. HBEC: Human bronchial epithelial cells. PBMC: peripheral blood mononuclear cells.

<table>
<thead>
<tr>
<th>Sentinel SNP (relationship to gene)</th>
<th>Chr.</th>
<th>Gene</th>
<th>Putative Function of encoded protein</th>
<th>Tissue</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Lung</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>HASM</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>HBEC</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>PBMC</td>
</tr>
<tr>
<td>rs993925 (intron)</td>
<td>1</td>
<td>( \text{TGFB2} )</td>
<td>Cytokine with roles in pro-fibrotic cytokine modulating epithelial repair mechanisms and extracellular matrix homeostasis including collagen deposition(^b)</td>
<td>+</td>
</tr>
<tr>
<td>rs2284746 (intron)</td>
<td>1</td>
<td>( \text{MFAP2} )</td>
<td>Major antigen of elastin-associated microfibrils(^{17}) and a candidate for involvement in the etiology of inherited connective tissue diseases.</td>
<td>+</td>
</tr>
<tr>
<td>rs12477314 (downstream)</td>
<td>2</td>
<td>( \text{HDAC4} )</td>
<td>Deacetylase of histone surrounding DNA thus influencing transcription factor access to the DNA possibly repressing gene transcription.</td>
<td>+</td>
</tr>
<tr>
<td>rs1344555 (intron)</td>
<td>3</td>
<td>( \text{EVII} )</td>
<td>Zinc finger transcription factor, encoded as part of MDS1/EVII complex locus (MECOM).</td>
<td>+</td>
</tr>
<tr>
<td>rs1529672 (intron)</td>
<td>3</td>
<td>( \text{RARB} )</td>
<td>Nuclear retinoic acid receptor responsive to retinoic acid, a vitamin A derivative and which also controls cell proliferation and differentiation.</td>
<td>+</td>
</tr>
<tr>
<td>rs153916 (intron)</td>
<td>5</td>
<td>( \text{SPATA9} )</td>
<td>Initially identified as a mediator of spermatogenesis, other family members may have a role in pancreatic development and β-cell proliferation(^{11})</td>
<td>+</td>
</tr>
<tr>
<td>rs2798641 (intron)</td>
<td>6</td>
<td>( \text{ARMC2} )</td>
<td>Function unknown although other family members have been identified as having roles in cell signalling, protein degradation and cytoskeleton functions(^{42})</td>
<td>+</td>
</tr>
<tr>
<td>rs2857595 (upstream)</td>
<td>6</td>
<td>( \text{NCR3} )</td>
<td>Required for efficient cytotoxicity responses by natural killer cells against normal cells and tumours(^{43})</td>
<td>+</td>
</tr>
<tr>
<td>Sentinel SNP (relationship to gene)</td>
<td>Chr.</td>
<td>Gene</td>
<td>Putative Function of encoded protein</td>
<td>Tissue</td>
</tr>
<tr>
<td>----------------------------------</td>
<td>------</td>
<td>------</td>
<td>-------------------------------------------------------------------------------------------------------</td>
<td>--------</td>
</tr>
<tr>
<td>rs6903823 (intron)</td>
<td>6</td>
<td>ZKSCAN3</td>
<td>Transcription factor involved in cell growth/cell cycle/signal transduction</td>
<td>+</td>
</tr>
<tr>
<td>rs7068966 (intron)</td>
<td>10</td>
<td>CDC123</td>
<td>Homologue in yeast shown to be a critical control protein modulating Eukaryotic initiation factor 2 in times of cell stress</td>
<td>+</td>
</tr>
<tr>
<td>rs11001819 (intron)</td>
<td>10</td>
<td>C10orf11</td>
<td>Function unknown</td>
<td>+</td>
</tr>
<tr>
<td>rs11172113 (intron)</td>
<td>12</td>
<td>LRP1</td>
<td>Potentially diverse roles including cell signalling and migration</td>
<td>+</td>
</tr>
<tr>
<td>rs1036429 (intron)</td>
<td>12</td>
<td>CCDC38</td>
<td>Function unknown although other family members involved in a diverse array of functions skeletal and motor function</td>
<td>-</td>
</tr>
<tr>
<td>rs1036429 (rs&lt;sup&gt;2&lt;/sup&gt;=0.96 with rs4762633 in SNRPF)</td>
<td>12</td>
<td>SNRPF</td>
<td>Small nuclear ribonucleoprotein F</td>
<td>+</td>
</tr>
<tr>
<td>rs12447804 (intron)</td>
<td>16</td>
<td>MMP15</td>
<td>Member of a large protease family with diverse functional roles via protease activity and specificity including tissue remodelling, wound healing, angiogenesis, and tumor invasion</td>
<td>+</td>
</tr>
<tr>
<td>rs2865531 (intron)</td>
<td>16</td>
<td>CFDP1</td>
<td>Craniofacial Development Protein 1</td>
<td>+</td>
</tr>
<tr>
<td>rs9978142 (upstream)</td>
<td>21</td>
<td>KCNE2</td>
<td>KCNQ1-KCNE2 K+ channels may modulate transepithelial anion secretion in Calu3 airway epithelial cells</td>
<td>-</td>
</tr>
<tr>
<td>Reference gene</td>
<td>12</td>
<td>GAPDH</td>
<td></td>
<td>+</td>
</tr>
</tbody>
</table>