

Multiethnic meta-analysis identifies new loci for pulmonary function

Annah B. Wyss¹, Tamar Sofer², Mi Kyeong Lee¹, Natalie Terzikhan^{3,4}, Jennifer N. Nguyen⁵, Lies Lahousse^{3,4}, Jeanne C. Latourelle⁶, Albert Vernon Smith⁷⁻⁹, Traci M. Bartz^{10,11}, Mary F. Feitosa¹², Wei Gao¹³, Tarunveer S. Ahluwalia^{14,15}, Wenbo Tang¹⁶, Christopher Oldmeadow¹⁷, Qing Duan¹⁸, Kim de Jong¹⁹, Mary K. Wojczynski¹², Xin-Qun Wang²⁰, Raymond Noordam²¹, Fernando Pires Hartwig^{22,23}, Victoria E. Jackson²⁴, Tianyuan Wang²⁵, Ma'en Obeidat²⁶, Brian D. Hobbs^{27,28}, Tianxiao Huan²⁹, Gleb Kichaev³⁰, Jianping Jin³¹, Mariaelisa Graff³², Tamara B. Harris³³, Ravi Kalhan³⁴, Susan R. Heckbert³⁵, Lavinia Paternoster²³, Kristin M. Burkart³⁶, Yongmei Liu³⁷, Elizabeth G. Holliday¹⁷, James G. Wilson³⁸, Judith M. Vonk¹⁹, Jason Sanders³⁹, R. Graham Barr^{40,41}, Renée de Mutsert⁴², Ana Maria Baptista Menezes²², Hieab H. H. Adams^{4,43}, Maarten van den Berge⁴⁴, Roby Joehanes⁴⁵, Lenore J. Launer³³, Alanna C. Morrison⁴⁶, Colleen M. Sitlani¹⁰, Juan C. Celedón⁴⁷, Stephen B. Kritchevsky⁴⁸, Rodney J. Scott^{17,49}, Kaare Christensen⁵⁰, Jerome I. Rotter⁵¹, Tobias N. Bonten^{52,53}, Fernando César Wehrmeister²², Yohan Bosse⁵⁴, Nora Franceschini³², Jennifer A. Brody¹⁰, Robert C. Kaplan⁵⁵, Kurt Lohman³⁷, Mark McEvoy¹⁷, Michael A. Province¹², Frits R. Rosendaal⁴², Kent D. Taylor⁵¹, David C. Nickle⁵⁶, International COPD Genetics Consortium Investigators, Vilundur Gudnason^{7,8}, Kari E. North³², Myriam Fornage⁵⁷, Bruce M. Psaty^{10,35,58,59}, Richard H. Myers⁶, George O'Connor^{60,61}, Torben Hansen¹⁴, Cathy C. Laurie², Pat Cassano^{16,62}, Joonhoon Sung⁶³, Woo Jin Kim⁶⁴, John R. Attia¹⁷, Leslie Lange⁶⁵, H. Marika Boezen¹⁹, Bharat Thyagarajan⁶⁶, Stephen S. Rich⁵, Dennis O. Mook-Kanamori^{42,52}, Bernardo Lessa Horta²², André G Uitterlinden⁶⁷, Don D. Sin^{26,68}, Hae Kyung Im⁶⁹, Michael H. Cho^{27,28}, Guy G. Brusselle^{3,4,70}, Sina A. Gharib⁷¹, Josée Dupuis^{13,60}, Ani Manichaikul⁵, and Stephanie J. London¹

1. Epidemiology Branch, National Institute of Environmental Health Sciences, National Institutes of Health, Department of Health and Human Services, Research Triangle Park, North Carolina, USA.
2. Department of Biostatistics, University of Washington, Seattle, Washington, USA.
3. Department of Respiratory Medicine, Ghent University Hospital, Ghent, Belgium.
4. Department of Epidemiology, Erasmus University Medical Center, Rotterdam, the Netherlands.
5. Center for Public Health Genomics, University of Virginia, Charlottesville, Virginia, USA.
6. Department of Neurology, Boston University School of Medicine, Boston, Massachusetts, USA.
7. Icelandic Heart Association, Kopavogur, Iceland.
8. Faculty of Medicine, University of Iceland, Reykjavik, Iceland.
9. Department of Biostatistics, University of Michigan, Ann Arbor, Michigan, USA.
10. Cardiovascular Health Research Unit, Department of Medicine, University of Washington, Seattle, Washington, USA.
11. Cardiovascular Health Research Unit, Department of Biostatistics, University of Washington, Seattle, Washington, USA.
12. Division of Statistical Genomics, Department of Genetics, Washington University School of Medicine, St Louis, Missouri, USA.
13. Department of Biostatistics, Boston University School of Public Health, Boston, Massachusetts, USA.
14. The Novo Nordisk Foundation Center for Basic Metabolic Research, Metabolic Genetics Section, Faculty of Health and Medical Sciences, University of Copenhagen, Copenhagen, Denmark.
15. Steno Diabetes Center Copenhagen, Gentofte, Denmark.
16. Division of Nutritional Sciences, Cornell University, Ithaca, New York, USA.
17. Hunter Medical Research Institute and Faculty of Health, University of Newcastle, NSW, Australia.
18. Department of Genetics, University of North Carolina, Chapel Hill, North Carolina, USA.

19. University of Groningen, University Medical Center Groningen, Department of Epidemiology, Groningen, the Netherlands.
20. Division of Biostatistics, Department of Public Health Sciences, University of Virginia, Charlottesville, Virginia, USA.
21. Department of Internal Medicine, Section Gerontology and Geriatrics, Leiden University Medical Center, Leiden, the Netherlands.
22. Postgraduate Program in Epidemiology, Federal University of Pelotas, Pelotas, Brazil.
23. Medical Research Council Integrative Epidemiology Unit, School of Social and Community Medicine, University of Bristol, Bristol, UK.
24. Department of Health Sciences, University of Leicester, Leicester, UK.
25. Integrative Bioinformatics Support Group, National Institute of Environmental Health Sciences, National Institutes of Health, Department of Health and Human Services, Research Triangle Park, North Carolina, USA.
26. The University of British Columbia Center for Heart Lung Innovation, St Paul's Hospital, Vancouver, British Columbia, Canada.
27. Channing Division of Network Medicine, Brigham and Women's Hospital, Boston, Massachusetts, USA.
28. Division of Pulmonary and Critical Care Medicine, Brigham and Women's Hospital, Boston, Massachusetts, USA.
29. National Heart, Lung, and Blood Institute, National Institutes of Health, Department of Health and Human Services, Bethesda, Maryland, USA.
30. University of California Los Angeles, Los Angeles, California, USA.
31. Westat, Durham, North Carolina, USA.
32. Department of Epidemiology, University of North Carolina, Chapel Hill, North Carolina, USA.
33. Laboratory of Epidemiology and Population Sciences, National Institute on Aging, National Institutes of Health, Department of Health and Human Services, Bethesda, Maryland, USA.
34. Division of Pulmonary and Critical Care Medicine, Northwestern University Feinberg School of Medicine, Chicago, Illinois, USA.
35. Cardiovascular Health Research Unit, Department of Epidemiology, University of Washington, Seattle, Washington, USA.
36. Division of Pulmonary, Allergy and Critical Care Medicine, Department of Medicine, College of Physicians and Surgeons, Columbia University, New York, New York, USA.
37. Wake Forest School of Medicine, Winston-Salem, North Carolina, USA.
38. Department of Physiology and Biophysics, University of Mississippi Medical Center, Jackson, Mississippi, USA.
39. Department of Medicine, Massachusetts General Hospital, Boston, Massachusetts, USA.
40. Department of Medicine, College of Physicians and Surgeons, Columbia University, New York, New York, USA.
41. Department of Epidemiology, Mailman School of Public Health, Columbia University, New York, New York, USA.
42. Department of Clinical Epidemiology, Leiden University Medical Center, Leiden, the Netherlands.
43. Department of Radiology, Erasmus University Medical Center, Rotterdam, the Netherlands.
44. University of Groningen, University Medical Center Groningen, Department of Pulmonary Diseases, Groningen, the Netherlands.
45. Hebrew SeniorLife, Harvard University, Boston, Massachusetts, USA.
46. Human Genetics Center, Department of Epidemiology, Human Genetics, and Environmental Sciences, School of Public Health, The University of Texas Health Science Center at Houston, Houston, Texas, USA.

47. Division of Pulmonary Medicine, Allergy, and Immunology, Department of Pediatrics, Children's Hospital of Pittsburgh of UPMC, University of Pittsburgh, Pittsburgh, Pennsylvania, USA.
48. Sticht Center on Aging, Wake Forest School of Medicine, Winston-Salem, North Carolina, USA.
49. Division of Molecular Medicine, Pathology North, NSW Health Pathology, Newcastle, NSW, Australia.
50. Department of Epidemiology, Biostatistics and Biodemography, University of Southern Denmark, Odense, Denmark.
51. Institute for Translational Genomics and Population Sciences, Los Angeles Biomedical Research Institute and Department of Pediatrics, Harbor-UCLA Medical Center, Torrance, California, USA.
52. Department of Public Health and Primary Care, Leiden University Medical Center, Leiden, the Netherlands.
53. Department of Pulmonology, Leiden University Medical Center, Leiden, the Netherlands.
54. Institut universitaire de cardiologie et de pneumologie de Québec, Department of Molecular Medicine, Laval University, Québec, Canada.
55. Department of Epidemiology and Population Health, Albert Einstein College of Medicine, Bronx, New York, USA.
56. Merck Research Laboratories, GpGx , Merck & Co., Inc., Kenilworth, New Jersey, USA.
57. Brown Foundation Institute of Molecular Medicine and Human Genetics Center, University of Texas Health Science Center at Houston, Houston, Texas, USA.
58. Cardiovascular Health Research Unit, Department of Health Services, University of Washington, Seattle, Washington, USA.
59. Kaiser Permanente Washington Health Research Institute, Seattle, Washington, USA.
60. National Heart, Lung, and Blood Institute's Framingham Heart Study, Framingham, Massachusetts, USA.
61. Pulmonary Center, Department of Medicine, Boston University School of Medicine, Boston, Massachusetts, USA.
62. Department of Healthcare Policy and Research, Weill Cornell Medical College, New York, New York, USA.
63. Department of Epidemiology, School of Public Health, Seoul National University, Seoul, South Korea.
64. Department of Internal Medicine and Environmental Health Center, Kangwon National University, Chuncheon, South Korea.
65. University of Colorado Denver, Denver, Colorado, USA.
66. Department of Laboratory Medicine and Pathology, University of Minnesota, Minneapolis, Minnesota, USA.
67. Department of Internal Medicine, Erasmus University Medical Center, Rotterdam, the Netherlands.
68. Respiratory Division, Department of Medicine, University of British Columbia, Vancouver, British Columbia, Canada.
69. Section of Genetic Medicine, The University of Chicago, Chicago, Illinois, USA.
70. Department of Respiratory Medicine, Erasmus University Medical Center, Rotterdam, The Netherlands.
71. Computational Medicine Core, Center for Lung Biology, UW Medicine Sleep Center, Department of Medicine, University of Washington, Seattle, Washington, USA.

These authors contributed equally to this work: ABW, TS, MKL, NT, JNN, LL

These authors jointly directed this work: GGB, SAB, JD, AM, SJL

Corresponding Author: Stephanie London, MD, DrPH, National Institute of Environmental Health Sciences, PO Box 12233, MD A3-05, Research Triangle Park, NC, 27709, email: london2@niehs.nih.gov, fax: 301-480-3290

Abstract

Nearly 100 loci have been identified for pulmonary function, almost exclusively in studies of European ancestry populations. We extend previous research by meta-analyzing genome-wide association studies of 1000 Genomes imputed variants in relation to pulmonary function in a multiethnic population of 90,715 individuals of European (N=60,552), African (N=8,429), Asian (N=9,959), and Hispanic/Latino (N=11,775) ethnicities. We identified over 50 novel loci at genome-wide significance in ancestry-specific and/or multiethnic meta-analyses. Recent fine mapping methods incorporating functional annotation, gene expression, and/or differences in linkage disequilibrium between ethnicities identified potential causal variants and genes at known and newly identified loci. Sixteen of the novel genes encode proteins with predicted or established drug targets, including *KCNK2* and *CDK12*.

Introduction

Pulmonary function traits (PFTs), including forced expiratory volume in the first second (FEV₁) and forced vital capacity (FVC), and their ratio FEV₁/FVC, are important clinical measures for assessing the health of the lungs, diagnosing chronic obstructive pulmonary disease (COPD), and monitoring the progression and severity of various other lung conditions. Further, even when within the normal range, these parameters are related to mortality, independently of standard risk factors¹⁻³.

In addition to lifestyle and environmental factors, such as smoking and air pollution, there is a demonstrated genetic component to pulmonary function⁴⁻⁶. Previous genome-wide association studies (GWAS) have collectively identified nearly 100 loci associated with PFTs. These analyses have been primarily conducted using HapMap imputed data among European ancestry populations⁷⁻¹². Recently, the UK BiLEVE Study (N=48,943) and SpiroMeta Consortium (N=38,199) have also examined associations between 1,000 Genomes imputed variants and PFTs, but only among Europeans¹³⁻¹⁵.

The present Cohorts for Hear and Aging Research in Genomic Epidemiology (CHARGE) meta-analysis builds upon previous studies by examining PFTs in relation to the more comprehensive 1000 Genomes panel in a larger study population (90,715 individuals from 22 studies, Table 1) comprised of multiple ancestral populations: European (60,552 individuals from 18 studies), African (8,429 individuals from 7 studies), Asian (9,959 individuals from 2 studies), and Hispanic/Latino (11,775 individuals from 6 ethnic background groups in 1 study). Along with look-up of our top findings in existing analyses of lung function traits and COPD, we additionally investigated correlation with gene expression in datasets from blood and lung tissue and assessed the potential biological and clinical relevance of our findings using recently developed methods integrating linkage disequilibrium (LD), functional annotation, gene expression and the multiethnic nature of our data. Finally, we identified known or potential drug targets for newly identified lung function associated loci.

Results

Ancestry-Specific Meta-Analyses

Each study used linear regression to model the additive effect of variants on PFTs, adjusting for age, sex, height, cigarette smoking, center (if multicenter study), and ancestral principal components, including a random familial effect to account for family relatedness when appropriate. Ancestry-specific fixed-effects inverse-variance weighted meta-analyses of study-specific results, with genomic control correction, were conducted in METAL (<http://www.sph.umich.edu/csg/abecasis/metal/>). Meta-analyses included approximately 11.1 million variants for European ancestry, 18.1 million for African ancestry, 4.2 million variants for Asian ancestry, and 13.8 million for Hispanic/Latino ethnicity. See Methods for full methods description.

European ancestry meta-analysis identified 17 novel loci (defined as more than 500kb in either direction from a known locus)^{16,17} which were significantly (defined as $p < 5.0 \times 10^{-8}$)^{14,18} associated with pulmonary function: 2 loci for FEV₁ only, 6 loci for FVC only, 7 loci for FEV₁/FVC only, and 2 loci for both FEV₁ and FVC (Table 2, Figure 1, Supplemental Figures 2-3). The African ancestry meta-analysis identified 8 novel loci significantly associated with pulmonary function: 2 loci for FEV₁, 1 locus for FVC, and 5 loci for FEV₁/FVC (Table 2, Supplemental Figure 1-3). Five of these loci were also significant at a stricter $p < 2.5 \times 10^{-8}$ threshold as has been suggested for populations of African ancestry¹⁸. Six of the African ancestry loci were identified based on low frequency variants (allele frequencies 0.01 to 0.02). In the Hispanic/Latino ethnicity meta-analysis, we identified one novel locus for FVC (Table 2, Supplemental Figure 1-3). Another locus was significantly associated with FEV₁, but this region was recently reported by HCHS/SOL¹⁹. For FEV₁/FVC among Hispanics/Latinos, all significant variants were in loci identified in previous studies of European ancestry populations. In the Asian ancestry meta-analysis, all variants significantly associated with PFTs were also in loci previously identified among European ancestry

populations (Supplemental Figure 1). Within each ancestry, variants discovered for one PFT were also looked-up for associations with the other two PFTs (Supplemental Table 1).

Multiethnic Meta-analysis

In multiethnic fixed effects meta-analyses of 10.9 million variants, we identified 47 novel loci significantly associated with pulmonary function. Thirteen of these loci were also identified in the ancestry-specific meta-analyses, while 34 were uniquely identified in the multiethnic meta-analysis: 11 loci for FEV₁ only, 14 loci for FVC only, 7 loci for FEV₁/FVC only, 1 locus for FEV₁ and FEV₁/FVC, and 1 locus for all three phenotypes (Table 2, Figure 1, Supplemental Figure 2-3). Although many of the 34 loci uniquely identified in the multiethnic meta-analysis were just shy of significance in the European ancestry meta-analysis, and therefore benefited from the additional sample size of the multiethnic meta-analysis, some multiethnic loci contained variants near genome-wide significance in at least one other ancestry-specific meta-analysis with some nominal significance ($p < 0.05$) in the remaining ancestry-specific meta-analyses (Supplemental Table 2). For example, rs7899503 in *JMJD1C* associated with FEV₁ had an I^2 value of 0 with a heterogeneity p-value of 0.40 and meta-analysis p-values of 1.35×10^{-5} for European ancestry, 4.56×10^{-7} for Asian ancestry, 0.002 for Hispanic/Latino ethnicity and 0.03 for African ancestry, with a multiethnic p-value of 8.70×10^{-14} .

In addition to the fixed-effects multiethnic meta-analysis, we conducted a random-effects meta-analysis using the Han and Eskin method²⁰ in METASOFT (<http://genetics.cs.ucla.edu/meta/>) as a sensitivity analysis. In instances where significant heterogeneity is present, the Han-Eskin method mitigates power loss²⁰. In the Han-Eskin random-effects model, 37 of the 47 loci identified in the fixed-effects model at $p < 5 \times 10^{-8}$ had a p-value below the same threshold (Supplemental Table 3). Among the 10 loci that did not, 8 loci still gave a $p < 5 \times 10^{-7}$ in the Han-Eskin random-effects model (*PIK3C2B*, *SUZ12P1*, *NCOR2/SCARB1*, *CTAGE1/RBBP8*, *C20orf112*, *COMTD1/ZNF503-AS1*, *EDAR*, and *RBMS3*) while only two did not (*CRADD* and *CCDC41*) (Supplemental Table 3). In addition, there were 6 loci for

FEV₁/FVC that were genome-wide significant in the Han-Eskin random-effects model that had not quite achieved genome-wide significance in the fixed-effects model: *GSTO1/GSTO2* (chr10, rs10883990), *FRMD4A* (chr10, rs1418884), *ETFA/SCAPER* (chr15, rs12440815), *APP* (chr21, rs2830155), *A4GNT* (chr3, rs9864090), *UBASH3B* (chr11, rs4935813) (Supplemental Table 3).

X-Chromosome Meta-Analysis

Within 13 of the 22 studies we conducted linear regression to model the effects of X-chromosome variants on PFTs by sex. Fixed-effects inverse-variance weighted meta-analyses were conducted separately in males and females using METAL and the resulting sex-specific results were combined using a weighted sums approach. No X-chromosome variants were associated with PFTs at genome-wide significance in any ancestral population (data not shown).

Look-up Replication of Novel Loci

Look-up replication was conducted in the UK BiLEVE study (N=48,943)¹⁴. Since this study only included individuals of European ancestry, we sought replication only for the 52 novel loci (excluding the major histocompatibility complex, MHC) identified in either the European ancestry or multiethnic discovery meta-analyses. Data for the lead variant was available in the UK BiLEVE study for 51 loci, including 49 loci with a consistent direction of effect between our results and those from UK BiLEVE (Table 2). Based on a two-sided $p < 9.6 \times 10^{-4}$ (0.05/52), 15 loci replicated for the same trait based on the lead variant from our analysis: *DCBLD2/MIR548G*, *SUZ12P1*, *CRHR1*, *WNT3*, *ZNF337*, *ALX1/RASSF9*, *MED1/CDK12*, *EYA2*, *RBMS3*, *LINC00340*, *FLJ35282/ELAVL2*, *DDHD1/MIR5580*, *TSHZ3*, *KLHL22/MED15*, *FAM168A* (Table 2). It was recently demonstrated that using one-sided replication p-values in GWAS, guided by the direction of association in the discovery study, increases replication power while being protective against type 1 error compared to the two-sided p-values²¹; under this criterion, an additional 4 loci replicated for the same trait based on the lead variant: *RAB5B*, *JMJD1C*, *AGMO*, and *C20orf112* (Table 2). Finally, one locus replicated for the same trait based on a different significant variant in the

locus (*DCC*, rs8085778, replication $p=2.80 \times 10^{-4}$, R^2 with lead variant=0.9) and one locus replicated for a different pulmonary function trait based on the lead variant (*PDXDC2P*, rs3973397, discovered for FVC replicated for FEV₁, replication $p=3.99 \times 10^{-4}$). In summary, we found evidence of replication in UK BiLEVE for 21 novel loci.

Overlap of Newly Identified Lung Function Loci with COPD

Pulmonary function measures are the basis for the diagnosis of COPD, an important clinical outcome; therefore, we also looked-up the 52 novel loci identified in the European ancestry or multiethnic meta-analyses in the International COPD Genetics Consortium (ICGC). This consortium recently published a meta-analysis of 1000 Genomes imputed variants and COPD primarily among individuals of European ancestry (N= 15,256 cases and 47,936 controls), including some of the same individuals included in the present lung function analysis²². Ten lead variants representing 8 novel loci were associated with COPD at $p < 9.6 \times 10^{-4}$: *RBMS3*, *OTUD4/SMAD1*, *TMEM38B/ZNF462*, *NCOR2/SCARB1*, *SUZ12P1*, *WNT3*, *SOGA2*, *C20orf112* (Supplemental Table 4). Directions of effects were consistent between our results and the COPD findings for these variants (i.e. variants associated with increased pulmonary function values were associated with decreased odds of COPD and vice-versa). Our top variant in *SOGA2* (also known as *MTCL1*) is in LD ($R^2=0.8$) with the top variant for COPD as reported by the ICGC Consortium²².

Expression Quantitative Trait Loci (eQTL) and Methylation Quantitative Trait Loci (mQTL) Signals of Pulmonary Function Variants

To query whether novel loci contained variants associated with gene expression (eQTLs), we looked-up variants from all 60 novel loci identified in ancestry-specific or multiethnic meta-analyses in the following datasets: 1) lung samples from 278 individuals in GTEx²³; 2) lung samples from 1,111 individuals participating in studies from the Lung eQTL Consortium including Laval University, the University of Groningen and the University of British Columbia²⁴⁻²⁶; 3) whole blood samples from 5,257

individuals participating in the Framingham Heart Study²⁷; 4) peripheral blood samples from 5,311 individuals participating in studies from EGCUT, InCHIANTI, Rotterdam Study, Fehrmann, HVH, SHIP-TREND and DILGOM²⁸; and 5) peripheral blood samples from 2,116 individuals participating in 4 Dutch studies collectively known as BIOS^{29,30}. The Lung eQTL Consortium study used a 10% FDR cut-off, while all other studies used a 5% FDR cut-off (see Supplemental Methods for further study descriptions and methods).

A significant cis-eQTL in at least one dataset was found for 34 lead variants representing 27 novel loci (Supplemental Table 5). Of these, 13 loci had significant cis-eQTLs only in datasets with blood samples and 3 loci only in datasets with lung samples (*COMTD1/ZNF503-AS1*, *FAM168A*, *SOGA2*). Eleven loci had significant cis-eQTLs in both blood and lung samples based on lead variants, with 1 locus having a significant cis-eQTL across all five datasets (*SMAD3*) and another 4 loci having a significant cis-eQTL in four datasets (*RAB5B*, *CRHR1*, *WNT3*, *ZNF337*). Significant trans-eQTLs in at least one dataset were found for 7 lead variants representing 4 novel loci (*TMEM38B/ZNF462*, *RAB5B*, *CRHR1*, and *WNT3*, Supplemental Table 5).

In addition, mQTL data were available from FHS and BIOS. Significant cis-mQTLs and trans-mQTLs in at least one dataset were found for 52 lead variants (43 novel loci) and 1 lead variant (1 novel locus), respectively (Supplemental Table 5).

Heritability and Genetic Correlation with Smoking and Height (LD Score Regression)

We used LD score regression³¹ to estimate the variance explained by genetic variants investigated in our European ancestry meta-analysis, also known as SNP heritability. Across the genome, the SNP heritability (narrow-sense) was estimated to be 20.7% (SE 1.5%) for FEV₁, 19.9% (SE 1.4%) for FVC and 17.5% (SE 1.4%) for FEV₁/FVC.

We also partitioned heritability by functional categories to investigate whether particular subsets of common variants were enriched³². We found significant enrichment in 6 functional categories

for all three PFTs: conserved regions in mammals, DNase I hypersensitive sites (DHS), super enhancers, the histone methylation mark H3K4me1 and histone acetylation marks H3K9Ac and H3K27Ac (Supplemental Figure 4). Another 7 categories showed enrichment for at least one PFT (Supplemental Figure 5). We observed the largest enrichment of heritability (14.5-15.3 fold) for conserved regions in mammals, which has ranked highest in previous partitioned heritability analyses for other GWAS traits (Supplemental Figure 5)³².

Since both height and smoking are important determinants of pulmonary function, and have been associated with many common variants in previous GWAS, we also used LD score regression to investigate genetic overlap³³ between our FEV₁, FVC and FEV₁/FVC results and publicly available GWAS results of ever smoking³⁴ and height³⁵. No significant genetic correlation was found between PFTs and smoking or height (Supplemental Table 6), indicating our findings are independent of those traits.

Functional Annotation of Pulmonary Function Variants

For functional annotation, we considered all novel variants with $p < 5 \times 10^{-8}$ from the 60 loci discovered in our ancestry-specific and multiethnic meta-analyses, plus significant variants from the MHC region, two loci previously discovered in the CHARGE exome chip study (*LY86/RREB1* and *SEC24C*)³⁶ and *DDX1*. Using Ensembl VEP³⁷, we found 6 missense variants in 4 loci outside of the MHC region and 22 missense variants in the MHC region (Supplementary Table 7). Of the 28 total missense variants, two (chr15:67528374 in *AAGAB* and chr6:30899524 in the MHC region) appear to be possibly damaging based on SIFT³⁸ and PolyPhen-2³⁹ scores (Supplementary Table 7). Using CADD⁴⁰, we found an additional 28 deleterious variants from 15 loci based on having a scaled C-score greater than 15 (Supplementary Table 8). In the MHC region, we found another 11 deleterious variants based on CADD. Based on RegulomeDB⁴¹, which annotates regulatory elements especially for non-coding regions, there were 52 variants from 18 loci with predicted regulatory functions (Supplementary Table 8). In the MHC region, there were an additional 72 variants with predicted regulatory functions.

Pathway Enrichment Analysis (DEPICT and IPA)

Gene set enrichment analyses conducted using DEPICT⁴² on genes annotated to variants with $p < 1 \times 10^{-5}$ based on the European ancestry meta-analysis results revealed 218 significantly enriched pathways (FDR < 0.05) (Supplementary Table 9). The enriched pathways were dominated by fundamental developmental processes, including many involved in morphogenesis of the heart, vasculature, and lung. Tissue and cell type analysis noted significant enrichment (FDR < 0.05) of smooth muscle, an important component of the lung (Supplemental Table 10 and Supplemental Figure 6). We found 8, 1, and 82 significantly prioritized genes (FDR < 0.05) for FEV₁, FVC, and FEV₁/FVC, respectively (Supplemental Table 11). Given the generally smaller p-values for variants associated with FEV₁/FVC, enriched pathways and tissue/cell types as well as prioritized genes were predominantly discovered from DEPICT analyses of FEV₁/FVC.

Due to extended LD across the MHC locus on chromosome 6 (positions 25,000,000-35,000,000), DEPICT excludes this region⁴². Standard Ingenuity Pathway Analysis (IPA) run without excluding the MHC highlighted 16 enriched networks, including three involved in inflammatory diseases or immunity; 33 of the 84 genes in these three networks are in the MHC region (Supplementary Table 12).

Identification of Potential Causal Variants by Incorporating Trans-ethnic Data and Functional Annotation (PAINITOR)

Using a multiethnic fine-mapping analysis incorporating strength of association, variation in genetic background across major ethnic groups, and functional annotations in PAINITOR⁴³, we examined 40 loci that contained at least 5 genome-wide significant variants in the European ancestry and multiethnic meta-analyses or at least 1 significant variant in the African or Hispanic/Latino ancestry meta-analyses. We identified 15 variants representing 13 loci as having high posterior probabilities of causality (>0.8): 3 for FEV₁, 3 for FVC, and 9 for FEV₁/FVC (Supplemental Table 13). Of the 15 putative casual variants, 11 showed high posterior probabilities of causality (>0.8) before considering annotations

and 4 were identified by adding functional annotations. Nine were the top SNPs at that locus from the fixed-effects meta-analysis (loci: *WNT3*, *PMFBP1/ZFH3*, *EN1/MARCO*, *C2orf48/HPCAL1*, *CPT1C*, *CADPS*, *LOC283867/CDH5*, *HDC*, and *CDC7/TGFBR3*), while 6 were not (loci: *CDK2/RAB5B*, *BMS1P4*, *PMFBP1/ZFH3*, *FLJ35282/ELAVL2*, *HDC*, and *COL8A1*).

Identification of Independent Signals in Known and Novel Loci for Pulmonary Function (FINEMAP)

We used FINEMAP⁴⁴ to identify variants with a high posterior probability of causality (>0.6) independent of 118 lead variants in pulmonary function loci identified in the current or previous studies¹⁴. We identified 37 independent variants for 23 previously identified loci and one independent variant at each of two novel loci (*LINC00340* and *SLC25A51P1/BAI3*; Supplementary Table 14).

Gene-based Analysis of GWAS Summary Statistics (S-PrediXcan)

Among the novel loci identified in the current GWAS of PFTs, we identified 7 variants corresponding to 9 genes demonstrating genome-wide significant evidence of association with lung or whole blood tissue-specific expression (Supplemental Table 15) based on the gene-based S-PrediXcan approach⁴⁵. Bayesian colocalization analysis⁴⁶ indicated the following associations demonstrated at least 50% probability of shared SNPs underlying both gene expression and PFTs: *ARHGEF17* and *FAM168A* in analysis of multiethnic GWAS for FEV₁/FVC based on GTEx whole blood models, and *WNT3* in analysis of multiethnic GWAS for FVC based on GTEx lung models (Supplemental Table 16).

Druggable Targets

To investigate whether the genes identified in our study encode proteins with predicted drug targets, we queried the ChEMBL database (<https://www.ebi.ac.uk/chembl/>). In addition, we incorporated an Ingenuity Pathway Analysis (IPA) to identify potential upstream targets. Genes associated with pulmonary function, but not included in the drug target analysis performed by Wain et al¹⁴, were evaluated for a total of 139 genes outside of the MHC: 110 genes representing the 60 novel loci identified in our fixed-effects ancestry-specific and multiethnic meta-analysis, 13 genes representing

the 6 novel loci identified in our random effects meta-analysis²⁰, 3 genes representing an additional 3 loci near significance in the African ancestry meta-analysis (*BAZ2B*, *NONE/PCDH10*, and *ADAMTS17*), 9 genes representing 2 loci identified in a recent CHARGE analysis of exome variants³⁶ which were also significant in our 1000 Genomes analysis (*LY86/RREB1* and *SEC24C*), and 4 genes representing one locus identified at genome-wide significance in a separate publication from one of our participating studies (*HCHS/SOL*)¹⁹ but also significant in our analysis (*ADORA2B/ZSWIM7/TTC19/NCOR1*). In the ChEMBL database, 17 of these genes encode proteins with predicted or known drug targets: *NR5A2*, *KCNK2*, *EDAR*, *KCNJ3*, *NR4A2*, *BAZ2B*, *A4GNT*, *GSTO1*, *GSTO2*, *NCOR2*, *SMAD3*, *NCOR1*, *CDK12*, *WNT3*, *PYGB*, *NANP*, *EYA2* (Supplemental Table 17). Of these, two genes (*KCNK2* and *CDK12*) have approved drug targets. Using IPA, four additional genes were predicted as drug targets (*ADORA2B*, *APP*, *CRHR1*, and *MAP3K1*; Supplemental Table 18) and 37 genes had drugs or chemicals as upstream regulators (Supplemental Table 19).

Discussion

By conducting a GWAS meta-analysis in a large multiethnic population we increased the number of known loci associated with pulmonary function by over 50%. In total, we identified 60 novel genetic regions (outside of the MHC region): 17 from European ancestry, 8 from African ancestry, 1 from Hispanic/Latino ethnicity, and 34 from multiethnic meta-analyses.

Just under half of the novel loci identified in our European ancestry and multiethnic meta-analyses replicated in look-up in a smaller independent sample of Europeans from the UK BiLEVE study¹⁴: *DCBLD2/MIR548G*, *SUZ12P1*, *CRHR1*, *WNT3*, *ZNF337*, *ALX1/RASSF9*, *MED1/CDK12*, *EYA2*, *RBMS3*, *LINC00340*, *FLJ35282/ELAVL2*, *DDHD1/MIR5580*, *TSHZ3*, *KLHL22/MED15*, *FAM168A*, *RAB5B*, *JMJD1C*, *AGMO*, *C20orf112*, *DCC*, and *PDXDC2P*. Among those loci which did not directly replicate for PFTs in the UK BiLEVE, the lead variants in an additional 4 European or multiethnic loci were significantly associated in the ICGC Consortium with COPD, which was defined using PFT measures²²:

OTUD4/SMAD1, *TMEM38B/ZNF462*, *NCOR2/SCARB1*, and *SOGA2*. Since we were not able to seek external replication for loci identified in the African ancestry meta-analyses, we used PAINTOR to query whether variants in the African-specific loci had a high probability of causality based on association statistics, LD, and functional annotations across ancestral groups⁴³. Interestingly, PAINTOR identified the lead variants for 6 out of 8 novel loci identified in African ancestry as having a high probability for causality: *C2orf48/HPCAL1*, *EN1/MARCO*, *CADPS*, *HDC*, *LOC283867/CDH5*, and *CPT1C*. PAINTOR or other new genomic methods such as FINEMAP and S-PrediXcan also produced evidence for causality for 4 European ancestry and multiethnic loci which had not replicated in UK BiLEVE or ICGC: *DCAF8*, *AFAP1*, *SLC25A51P1/BAI3* and *SMAD3*. Therefore, we found evidence for look-up replication of 25 loci in the UK BiLEVE study or ICGC COPD consortium and support for validation of an additional 10 loci using PAINTOR, FINEMAP, or S-PrediXcan.

Our analysis also sheds light on additional potential causal genes at a complex locus (chromosome 17 near positions 43600000 to 44300000, hg19) previously discovered from GWAS of FEV₁ which identified *KANSL1* in European populations as the top finding for this region^{14,15}. With the exception of a single INDEL in *KANSL1* in our European ancestry meta-analysis (17:44173680:T_TC, $p=1.03 \times 10^{-10}$), we found *CRHR1* as the strongest gene associated with FEV₁ in this region. Although some variants in *CRHR1* identified in our study are within 500kb of *KANSL1* (e.g., rs16940672, 17:43908152, $p=2.07 \times 10^{-10}$), a number of significant variants in this gene are more than 500kb away from previously identified hits [our definition of novel] (e.g., rs143246821, 17:43685698, $p=9.06 \times 10^{-10}$). In our multiethnic meta-analysis, several variants in *CRHR1* were associated with FEV₁ at smaller p-values than variants in *KANSL1*. Definitive assessment of the causal variants at this locus, as well as other multigenic GWAS loci, will likely require additional data from ongoing large scale sequencing studies to enable detailed fine mapping.

In both our European and multiethnic meta-analyses we also noted a significant association with *WNT3* on chromosome 17 near position 44800000 (hg19) which is more than 500kb from *KANSL1* or *CRHR1* [our definition of novel]. We found that the top variant in *WNT3* for FEV₁ among individuals of European ancestry (rs916888, 17:44863133, p=3.76x10⁻⁹) had a high probability for causality based on PAINTOR, an analysis which integrates functional annotations along with association statistics and LD for each ethnicity⁴³. We also found evidence that *WNT3* may be the causal gene at this locus using S-PrediXcan, a gene level association test that prioritizes potentially causal genes while filtering out LD-induced false positives^{45,46}. Notably, S-PrediXcan implicated *WNT3* as a likely mediating gene for FVC based on the top variant in our multiethnic meta-analyses (rs199525, 17:44847834, p=7.52x10⁻⁹), which is an eQTL SNP for *WNT3* in lung and other tissues. Further, the lead *WNT3* variants for both FEV₁ and FVC (rs916888 and rs199525) were significantly associated with COPD in a look-up of a large published meta-analysis dataset²². In addition, other genes in the *WNT* signaling pathway, a fundamental development pathway, have been implicated as influencing pulmonary function⁴⁷. This pathway was also one of the significant pathways identified in our analysis. In a previous pathway analysis of asthma, *SMAD3* has been shown to interact with the *WNT* signaling pathway⁴⁸. Finally, *WNT3* also emerged as having a potential druggable target, and incorporation of pathway analysis to identify upstream regulators found an additional four drugs in clinical use for which *WNT3* is a target molecule (chemotherapeutic agents doxorubicin and paclitaxel, the hormone beta-estradiol and LGK-974, a novel agent that targets a WNT-specific acyltransferase)⁴⁹. Again, further evaluation of this interesting and complex locus which contains many significant variants in LD will benefit from data being generated in ongoing large-scale sequencing studies.

Some genes identified in our study play key roles in inflammation, immunity and pulmonary biology. For example, *MARCO* (macrophage receptor with collagenous structure) has been shown in murine models to be required for lung defense against pneumonia as well as inhaled particles^{50,51}.

SMAD3 is part of the SMAD family of proteins which are signal transducers and transcriptional modulators that mediate multiple signaling pathways. *SMAD3* is activated by transforming growth factor beta (TGF- β) which plays a key role in airway remodeling. *SMAD3* has a predicted drug target and SNPs in *SMAD3* are significantly associated with asthma in GWAS^{52,53}.

Other genes identified in our study which are targeted by approved drugs include *CDK12* and *KCNK2*. *CDK12* drug targets include AT-7519, Roniciclib, AZD-5438, and PH.A-793887. Roniciclib has been used in clinical trials including lung cancer patients⁵⁴. *KCNK2* (potassium channel subfamily K member 2) is targeted by 5 inhalational anesthetic agents. These agents have anti-inflammatory effects both systemically⁵⁵ and in the lungs⁵⁶ and meta-analysis of clinical studies shows protection against pulmonary complications after cardiac surgery⁵⁷. A recent trial suggested that one of these inhalation agents, sevoflurane, offers promise for reducing epithelial injury and improving outcomes in patients with acute respiratory distress syndrome⁵⁸.

In addition to querying commonly used genome databases for functional annotation of variants, we sought to narrow down causal variants in implicated loci using recently developed methods that incorporate LD, functional data and/or the multiethnic analysis done in this paper. In particular, PAINTOR is a useful tool to identify potential causal variants in our novel loci as it leverages LD across ancestral groups along with association statistics and functional annotations⁴³. PAINTOR identified 15 putative causal variants from 13 loci, including 6 novel loci in African ancestry meta-analyses and 7 loci uniquely identified in the multiethnic meta-analyses such as *PMFBP1/ZFH3* and *COL8A1* (part of the *DCBLD2* loci). Notably, 8 of the 15 putative causal variants from PAINTOR were the top SNPs from the fixed-effects meta-analysis (e.g., rs916888 *WNT3*). Similarly, FINEMAP has been shown to be an accurate and efficient tool for investigating whether lead SNPs for a given loci are driven by independent variants in the same region, especially when annotation information is not available⁴⁴. Among previous and novel loci identified in individuals of European ancestry, we identified 37 independent variants for 23

previously identified loci and two lead variants for two novel loci (rs1928168 *LINC00340* and rs9351637 *SLC25A51P1/BAI3*) with a high probability of causality. Finally, we ran S-PrediXcan a gene level association test that prioritizes potentially causal genes⁴⁵. Seven of our novel loci contained putative causal genes based on S-PrediXcan for lung or whole blood tissues, including *NRBF2* (part of the *JMJD1C* locus) and *WNT3*. S-PrediXcan also highlighted the region around chromosome 11 position 73280000 (hg19), noting strong evidence for both *FAM168A* and *ARHGEF17* which was further supported by the co-localization analysis. Interestingly, DEPICT also prioritized *ARHGEF17*, a member of the guanine nucleotide exchange factor (GEF) family of genes which can mediate actin polymerization and contractile sensitization in airway smooth muscle^{59,60}.

Rather than conducting a standard gene-based pathway analysis, we performed a newer integrative method, DEPICT, that incorporates cell and tissue-specific functional data into a pathway analysis to prioritize genes within implicated loci⁴². In addition to identifying potential causal variants, this approach revealed a number of fundamental development processes, including pathways related to lung development, regulation of growth, and organ morphogenesis. The *WNT* signaling pathway was also highlighted along with processes relevant to the pathogenesis of COPD including extracellular matrix structure and collagen networks. Tissue/cell type enrichment results highlighted smooth muscle which is highly relevant for lung function. DEPICT excludes the MHC due to extended LD in this region, which likely explains the relative paucity of inflammation-related pathways identified compared to previous pathway analyses in GWAS of PFTs^{24,47}. Indeed, standard IPA analysis of our data including the MHC region, found that 33 of 84 genes (39%) in the 3 (out of 16) enriched networks involved in immune or inflammatory processes are in the MHC. The predominance of fundamental pathways related to lung growth, differentiation and structure is consistent with recent work⁶¹ that has rekindled interest in the observation made 40 years ago⁶² that individuals can cross the threshold for diagnosis of COPD either by rapid decline in adulthood or by starting from a lower baseline of maximal pulmonary function attained

during growth. Within this context, understanding the genetic (and environmental) factors that influence the variability in maximal lung function attained during the first three decades of life is essential to reducing the public health burden of COPD⁶³.

In summary, our study extends existing knowledge of the genetic landscape of PFTs by utilizing the more comprehensive 1000 Genomes imputed variants, increasing the sample size, including multiple ancestries and ethnicities, and employing newly developed computational applications to interrogate implicated loci. We discovered 60 novel loci associated with pulmonary function and replicated many in an independent sample. We found evidence that several variants in these loci were missense mutations and had possible deleterious or regulatory effects, and many had significant eQTLs. Further, using new genomic methods that incorporate LD, functional data and the multiethnic structure of our data, we shed light on potential causal genes and variants in implicated loci. Finally, several of the newly identified genes linked to lung function are druggable targets, highlighting the clinical relevance of our integrative genomics approach.

Methods

Studies

Member and affiliate studies from The Cohorts for Hearth and Aging Research in Genomic Epidemiology (CHARGE) consortium with pulmonary function and 1,000 Genomes imputed genetic data were invited to participate in the present meta-analysis. Participating studies included: AGES, ALHS, ARIC, CARDIA, CHS, FamHS, FHS, GOYA, HCHS/SOL, HCS, Health ABC, Healthy Twin, JHS, KARE3, LifeLines, LLFS, MESA, NEO, 1982 PELOTAS, RSI, RSII, RIII. Characteristics of these studies are provided in Supplemental Table 20 and descriptions of study designs are provided in the Supplemental Methods.

Pulmonary Function

Spirometry measures of pulmonary function (FEV₁, FVC, and the ratio FEV₁/FVC) were collected by trained staff in each study according to American Thoracic Society (ATS) or European Respiratory Society guidelines (see cohort descriptions in Supplemental Methods for more details).

Variants

Studies used various genotyping platforms, including Affymetrix Human Array 6.0, Illumina Human Omni Chip 2.5, and others, as described in cohort descriptions in the Supplemental Methods. Using MACH, MINIMAC, or IMPUTE2, studies then used genotyped data to impute ~38 million variants based on the 1000 Genomes Integrated phase 1 version 3 reference panel (released March 2012). Two studies (Hunter Community and CARDIA) imputed to the 1000 Genomes European phase 1 version 3 reference panel; sensitivity analyses excluding these two studies showed no material differences in results (see Supplemental Results).

Statistical Analysis

Within each study, linear regression was used to model the additive effect of variants on PFTs. FEV₁ and FVC were modeled as milliliters and FEV₁/FVC as a proportion. Studies were asked to adjust analyses for age, age², sex, height, height², smoking status (never, former, current), pack-years of smoking, center (if multicenter study), and ancestral principal components, including a random familial effect to account for family relatedness when appropriate⁶⁴. Models of FVC were additionally adjusted for weight. Analyses were conducted using ProbABEL, PLINK, FAST, or the R kinship package as described in the cohort descriptions of the Supplemental Methods.

Ancestry-specific and multiethnic fixed effects meta-analyses using inverse variance weighting of study-specific results with genomic control correction were conducted in Meta Analysis Helper (METAL, <http://www.sph.umich.edu/csg/abecasis/metal/>). Multiethnic random effects meta-analyses using the four ancestry-specific fixed effects meta-analysis results were conducted using the Han-Eskin

model²⁰ in METASOFT (<http://genetics.cs.ucla.edu/meta/>). Only variants with p-values for association <0.05 or p-values for heterogeneity <0.1 from the fixed-effects model were included in the random-effects model.

Variants with imputation quality scores (r^2) less than 0.3 and/or a minor allele count (MAC) less than 20 were excluded from each study prior to meta-analysis. Following meta-analysis, we also excluded variants with less than 1/3 the total sample size or less than the sample size of the largest study for a given meta-analysis to achieve the following minimal sample sizes: 20,184 for European ancestry; 2,810 for African ancestry; 7,862 for Asian ancestry; 4,435 for Hispanic/Latino ethnicity and 30,238 for Multiethnic.

Significance was defined as $p < 5 \times 10^{-8}$ ^{14,18}. Novel variants were defined as being more than +/- 500kb from the top variant of a loci identified in a previous GWAS of pulmonary function^{16,17}. We used the list of 97 known variants as published in the recent UK BiLEVE paper¹⁴ with the following modifications: added variants in *DDX1*, *DNER*, *CHRNA5* since listed in GWAS catalog; added variants in *LCT*, *FGF10*, *LY86/RREB1*, *SEC24C*, *RPAP1*, *CASC17*, and *UQCC1* since identified in exome chip paper³⁶; added variant in *TMEM163* identified in Loth et al paper¹⁰; used 17:44339473 instead of 17:44192590 to represent *KANSL1* since 17:44339473 was the original variant listed for *KANSL1* in Wain et al 2015¹⁵; and used 12:28283187 instead of 12:28689514 to represent *PTHLH* since 12:28283187 was the original variants listed for *PTHLH* in Soler Artigas et al 2015¹³.

Genomic inflation factors (lambda values) from quantile-quantile plots of observed and expected p-values for ancestry- and phenotype-specific meta-analyses are presented in Supplemental Table 21. Lambda values were slightly higher in European and multiethnic meta-analyses (range of lambda 1.12 to 1.16) than in other ancestry-specific meta-analyses (range of lambda 1.01 to 1.06) likely due to the much larger sample sizes of the European and multiethnic meta-analyses⁶⁵.

LD Score Regression

The SNP heritability, i.e. the variance explained by genetic variants, was calculated from the European ancestry GWAS summary statistics (with genomic control off) using LD score regression (<https://github.com/bulik/ldsc>)³¹. Partitioned heritability was also calculated using the method described by Finucane and colleagues³². In total, 28 functional annotation classes were used for this analysis, including: coding regions, regions conserved in mammals, CCCTC-binding factor (CTCF), DNase genomic foot printing (DGF), DNase I hypersensitive sites (DHS), fetal DHS, enhancer regions; including super-enhancers and active enhancers from the FANTOM5 panel of samples, histone marks including two versions of acetylation of histone H3 at lysine 27 (H3K27ac and H3K27ac2), histone marks monomethylation (H3K4me1), trimethylation of histone H3 at lysine 4 (H3K4me) and acetylation of histone H3 at lysine 9 (H3K9ac5). In addition to promotor and intronic regions, transcription factor binding site (TFBS), transcription start site (TSS) and untranslated regions (UTR3 and UTR5). A p-value of 0.05/28 classes $< 1.79 \times 10^{-3}$ was considered statistically significant. Genetic correlation between our pulmonary function (FEV₁, FVC and FEV₁/FVC) results and publicly available GWAS of ever smoking³⁴ and height³⁵ was also investigated using LD score regression³³.

Functional Annotation

To find functional elements in novel genome-wide significant signals, we annotated SNPs using various databases. We used Ensembl Variant Effect Predictor (VEP)³⁷ (Accessed 17 Jan 2017) and obtained mapped genes, transcripts, consequence of variants on protein sequence, Sorting Intolerant from Tolerant (SIFT)³⁸ scores, and Polymorphism Phenotyping v2 (PolyPhen-2)³⁹ scores. We checked if there were deleterious variants using Combined Annotation Dependent Depletion (CADD) v1.3⁴⁰ which integrates multiple annotations, compares each variant with possible substitutions across the human genome, ranks variants, and generates raw and scaled C-scores. A variant having a scaled C-score of 10 or 20 indicates that it is predicted to be in the top 10% or 1% deleterious changes in human genome,

respectively. We used a cutoff of 15 to provide deleterious variants since it is the median for all possible splice site changes and non-synonymous variants (<http://cadd.gs.washington.edu/info>, Accessed 18 Jan 2017). To find potential regulatory variants, we used RegulomeDB⁴¹ (Accessed 17 Jan 2017) which integrates DNA features and regulatory information including DNAase hypersensitivity, transcription factor binding sites, promoter regions, chromatin states, eQTLs, and methylation signals based on multiple high-throughput datasets and assign a category to each variant. Variants having RegulomeDB categories 1 or 2, meaning 'likely to affect binding and linked to expression of a gene target' or 'likely to affect binding,' were considered as regulatory variants.

Pathway Analysis (DEPICT and IPA)

For gene prioritization and identification of enriched pathways and tissues/cell types, we used Data-driven Expression Prioritized Integration for Complex Traits (DEPICT)⁴² with association results for FEV₁, FVC, and FEV₁/FVC. We used association results from our European ancestry meta-analysis and the LD structure from 1000 Genomes European (CEU, GBR, and TSI) reference panel. The software excludes the major histocompatibility complex (MHC) region on chromosome 6 due to extended LD structure in the region. We ran a version of DEPICT for 1000 Genomes imputed meta-analysis results using its default parameters with an input file containing chromosomal location and p-values for variants having unadjusted p-values <1x10⁻⁵. For gene set enrichment analyses, DEPICT utilizes 14,461 reconstituted gene sets generated by genes' co-regulation patterns in 77,840 gene expression microarray data. For tissue/cell type enrichment analysis, mapped genes were tested if they are highly expressed in 209 medical subject headings (MeSH) annotations using 37,427 microarray data. Gene prioritization analysis using co-functionality of genes can provide candidate causal genes in associated loci even if the loci are poorly studied or a gene is not the closest gene to a genome-wide significant variant. We chose FDR<0.05 as a cutoff for statistical significance in these enrichment analyses and gene prioritization results. Because DEPICT excludes the MHC, we also ran a pathway analysis with Ingenuity Pathway

Analysis (IPA) (Ingenuity Systems, Redwood City, CA, USA, <http://www.ingenuity.com/>) on genes to which variants with $p < 1 \times 10^{-5}$ annotated.

PAINTOR

To identify causal variants in novel genome-wide significant loci, we used a trans-ethnic functional fine mapping method⁴³ implemented in PAINTOR (https://github.com/gkichaev/PAINTOR_FineMapping, Accessed on May 2, 2016). This method utilizes functional annotations along with association statistics (Z-scores) and linkage disequilibrium (LD) information for each locus for each ethnicity. We included our ancestry-specific meta-analysis results and used the African, American, European, and East Asian individuals from 1000 Genomes to calculate linkage disequilibrium (LD)⁶⁶. From the novel loci we identified in our ancestry-specific and multiethnic fixed effects meta-analyses, we selected 40 high priority loci which had at least five variants meeting significance: 10 loci for FEV₁, 15 loci for FVC, and 17 loci for FEV₁/FVC. This list included 6 loci which overlapped with the UK BiLEVE 1000 Genomes paper¹⁴ and 1 locus with the CHARGE exome paper³⁶, since we ran PAINTOR prior those publications. To reduce computational burden, we limited flanking regions to ± 100 kilobase (kb) from the top single nucleotide polymorphisms (SNPs) and included variants with absolute value of Z-score greater than 1.96.

We used 269 publicly available annotations relevant to 'lung', 'bronch', or 'pulmo' from the following: hypersensitivity sites (DHSs)⁶⁷, super enhancers⁶⁸, Fantom5 enhancer and transcription start site regions⁶⁹, immune cell enhancers⁷⁰, and methylation and acetylation marks ENCODE⁷¹. We ran PAINTOR for each phenotype separately to prioritize annotations based on likelihood-ratio statistics^{72,73}. We included minimally correlated top annotations (less than five for each phenotype) to identify causal variants.

For the 40 loci from the fixed-effects meta-analysis, we used PAINTOR to construct credible sets of causal variants using a Bayesian meta-analysis framework. To obtain 95% credible sets for each locus,

we ranked SNPs based on posterior probabilities of causality (high to low) and then took the SNPs filling in 95% of the summed posterior probability. We computed the median number of SNPs in the credible sets for ancestry-specific and multi-ethnic analyses of each trait.

FINEMAP

We used FINEMAP⁴⁴ to identify signals independent of lead variants for pulmonary function loci identified in the current or previous studies¹⁴. We used a reference population from the Rotterdam Study (N=6,291). SNPs with MAF of <1% were excluded, leaving 118 SNPs for analysis. 10 SNPs for FEV₁ and FVC and 20 SNPs for FEV₁/FVC were further excluded because the LD matrix of the reference file from the Rotterdam Study did not represent the correlation matrix of the total study population. We allowed up to 10 causal SNPs per loci in FINEMAP analyses. To reduce the chance of false positive findings, we also conducted sensitivity analyses allowing up to 15 causal SNPs for loci with more than 4 SNPs with posterior probabilities of >0.8.

S-PrediXcan

S-PrediXcan is a novel summary statistics based approach for gene-based analysis⁴⁵ that was derived as an extension of the PrediXcan method for integration of GWAS and reference transcriptome data⁷⁴. We used the S-PrediXcan approach to prioritize potentially causal genes, coupled with a Bayesian colocalization procedure⁴⁶ used to filter out LD-induced false positives. S-PrediXcan was used to analyze both European ancestry and multi-ethnic GWAS summary data for pulmonary function traits from the current study.

S-PrediXcan analysis was performed using the following publicly available tissue-specific expression models (<http://predictdb.org>) from the Genotype-Tissue Expression (GTEx) project v6p²³: (1) GTEx Lung (278 samples) and (2) GTEx Whole blood (338 samples). Approximately 85% of participants in GTEx are white, 12% African American, and 3% of other races/ethnicities. Gene-based S-PrediXcan results were filtered on the following: (1) Proportion of SNPs used = (n SNPs available in GWAS summary

data)/ (n SNPs in prediction model) > 0.6, and (2) prediction performance R-squared > 0.01. Following application of S-PrediXcan to each of the GWAS summary data sets, we computed Bonferroni-corrected p-values derived as the nominal p-value for each gene-based test divided by the number of genes passing specified filters in each analysis to test whether genetically regulated gene expression was associated with the trait of interest. The genome-wide S-PrediXcan results were then merged with novel loci from the current GWAS study by identifying all matches in which the novel locus SNP was within 500kb of the start of the gene.

We further incorporated a Bayesian colocalization approach⁴⁶ to interpret the extent to which S-PrediXcan results may have been influenced by linkage disequilibrium within the region of interest. The Bayesian colocalization procedure was run using the following priors: $p_1 = 1e-4$; prior probability SNP associated to trait 1, $p_2 = 1e-4$; prior probability SNP associated to trait 2, $p_{12} = 1e-5$; prior probability SNP associated to both traits. The procedure generated posterior probabilities that correspond to one of the following hypotheses: a region is (H0) has no association with neither trait, (H1) associated with PFT phenotype but not gene expression, (H2) associated with gene expression but not PFT phenotype, (H3) associated with both traits, due to two independent SNPs, (H4) associated with both traits, due to one shared SNP.

Druggable Targets

We searched annotated gene lists against the ChEMBL database (v22.1, updated on November 15, 2016) to identify genes as targets of approved drugs or drugs in development. In addition, we used the Ingenuity Pathway Analysis (IPA, www.ingenuity.com, content of 2017-06-22) to identify drug targets and upstream regulators of the gene lists. We reported the upstream regulators in the following categories, biologic drug, chemical - endogenous mammalian, chemical - kinase inhibitor, chemical – other, chemical drug, chemical reagent, and chemical toxicant.

Data Availability

Upon acceptance at a journal, complete meta-analysis results will be deposited in the database of Genotypes and Phenotypes (dbGaP) under the CHARGE acquisition number phs000930. GWAS data for US studies are already available in dbGaP. For non-US studies, please send requests to the study PI or Stephanie London (london2@niehs.nih.gov) who will forward them to the relevant party.

References

1. Hole, D.J. *et al.* Impaired lung function and mortality risk in men and women: findings from the Renfrew and Paisley prospective population study. *BMJ* **313**, 711-5; discussion 715-6 (1996).
2. Young, R.P., Hopkins, R. & Eaton, T.E. Forced expiratory volume in one second: not just a lung function test but a marker of premature death from all causes. *Eur Respir J* **30**, 616-22 (2007).
3. Schunemann, H.J., Dorn, J., Grant, B.J., Winkelstein, W., Jr. & Trevisan, M. Pulmonary function is a long-term predictor of mortality in the general population: 29-year follow-up of the Buffalo Health Study. *Chest* **118**, 656-64 (2000).
4. Wilk, J.B. *et al.* Evidence for major genes influencing pulmonary function in the NHLBI family heart study. *Genet Epidemiol* **19**, 81-94 (2000).
5. Givelber, R.J. *et al.* Segregation analysis of pulmonary function among families in the Framingham Study. *Am J Respir Crit Care Med* **157**, 1445-51 (1998).
6. Chen, Y., Horne, S.L., Rennie, D.C. & Dosman, J.A. Segregation analysis of two lung function indices in a random sample of young families: the Humboldt Family Study. *Genet Epidemiol* **13**, 35-47 (1996).
7. Repapi, E. *et al.* Genome-wide association study identifies five loci associated with lung function. *Nat Genet* **42**, 36-44 (2010).
8. Hancock, D.B. *et al.* Meta-analyses of genome-wide association studies identify multiple loci associated with pulmonary function. *Nat Genet* **42**, 45-52 (2010).
9. Hancock, D.B. *et al.* Genome-wide joint meta-analysis of SNP and SNP-by-smoking interaction identifies novel loci for pulmonary function. *PLoS Genet* **8**, e1003098 (2012).
10. Loth, D.W. *et al.* Genome-wide association analysis identifies six new loci associated with forced vital capacity. *Nat Genet* **46**, 669-77 (2014).
11. Soler Artigas, M. *et al.* Genome-wide association and large-scale follow up identifies 16 new loci influencing lung function. *Nat Genet* **43**, 1082-90 (2011).
12. Tang, W. *et al.* Large-scale genome-wide association studies and meta-analyses of longitudinal change in adult lung function. *PLoS One* **9**, e100776 (2014).
13. Soler Artigas, M. *et al.* Sixteen new lung function signals identified through 1000 Genomes Project reference panel imputation. *Nat Commun* **6**, 8658 (2015).
14. Wain, L.V. *et al.* Genome-wide association analyses for lung function and chronic obstructive pulmonary disease identify new loci and potential druggable targets. *Nat Genet* **49**, 416-425 (2017).
15. Wain, L.V. *et al.* Novel insights into the genetics of smoking behaviour, lung function, and chronic obstructive pulmonary disease (UK BiLEVE): a genetic association study in UK Biobank. *Lancet Respir Med* **3**, 769-81 (2015).
16. Liu, C.T. *et al.* Trans-ethnic Meta-analysis and Functional Annotation Illuminates the Genetic Architecture of Fasting Glucose and Insulin. *Am J Hum Genet* **99**, 56-75 (2016).
17. Mahajan, A. *et al.* Genome-wide trans-ancestry meta-analysis provides insight into the genetic architecture of type 2 diabetes susceptibility. *Nat Genet* **46**, 234-44 (2014).
18. Pe'er, I., Yelensky, R., Altshuler, D. & Daly, M.J. Estimation of the multiple testing burden for genomewide association studies of nearly all common variants. *Genet Epidemiol* **32**, 381-5 (2008).
19. Burkart, K.M. *et al.* Genome-Wide Association Study (GWAS) of Lung Function Among Hispanic/Latino Individuals of Diverse Backgrounds. The Hispanic Community Health Study/Study of Latinos (HCHS/SOL). *American Journal of Respiratory and Critical Care Medicine* **191**, A1069.

20. Han, B. & Eskin, E. Random-effects model aimed at discovering associations in meta-analysis of genome-wide association studies. *Am J Hum Genet* **88**, 586-98 (2011).
21. Sofer, T. *et al.* A powerful statistical framework for generalization testing in GWAS, with application to the HCHS/SOL. *Genet Epidemiol* **41**, 251-258 (2017).
22. Hobbs, B.D. *et al.* Genetic loci associated with chronic obstructive pulmonary disease overlap with loci for lung function and pulmonary fibrosis. *Nat Genet* **49**, 426-432 (2017).
23. The Genotype-Tissue Expression (GTEx) project. *Nat Genet* **45**, 580-5 (2013).
24. Obeidat, M. *et al.* Molecular mechanisms underlying variations in lung function: a systems genetics analysis. *Lancet Respir Med* **3**, 782-95 (2015).
25. Obeidat, M. *et al.* GSTCD and INTS12 regulation and expression in the human lung. *PLoS One* **8**, e74630 (2013).
26. Hao, K. *et al.* Lung eQTLs to help reveal the molecular underpinnings of asthma. *PLoS Genet* **8**, e1003029 (2012).
27. Joehanes, R. *et al.* Integrated genome-wide analysis of expression quantitative trait loci aids interpretation of genomic association studies. *Genome Biol* **18**, 16 (2017).
28. Westra, H.J. *et al.* Systematic identification of trans eQTLs as putative drivers of known disease associations. *Nat Genet* **45**, 1238-43 (2013).
29. Bonder, M.J. *et al.* Disease variants alter transcription factor levels and methylation of their binding sites. *Nat Genet* **49**, 131-138 (2017).
30. Zhernakova, D.V. *et al.* Identification of context-dependent expression quantitative trait loci in whole blood. *Nat Genet* **49**, 139-145 (2017).
31. Bulik-Sullivan, B.K. *et al.* LD Score regression distinguishes confounding from polygenicity in genome-wide association studies. *Nat Genet* **47**, 291-5 (2015).
32. Finucane, H.K. *et al.* Partitioning heritability by functional annotation using genome-wide association summary statistics. *Nat Genet* **47**, 1228-35 (2015).
33. Bulik-Sullivan, B. *et al.* An atlas of genetic correlations across human diseases and traits. *Nat Genet* **47**, 1236-41 (2015).
34. Genome-wide meta-analyses identify multiple loci associated with smoking behavior. *Nat Genet* **42**, 441-7 (2010).
35. Wood, A.R. *et al.* Defining the role of common variation in the genomic and biological architecture of adult human height. *Nat Genet* **46**, 1173-86 (2014).
36. Jackson, V.E. *et al.* Meta-analysis of exome array data identifies six novel genetic loci for lung function. *bioRxiv* (2017).
37. McLaren, W. *et al.* The Ensembl Variant Effect Predictor. *Genome Biol* **17**, 122 (2016).
38. Kumar, P., Henikoff, S. & Ng, P.C. Predicting the effects of coding non-synonymous variants on protein function using the SIFT algorithm. *Nature Protocols* **4**, 1073-1082 (2009).
39. Adzhubei, I.A. *et al.* A method and server for predicting damaging missense mutations. *Nature Methods* **7**, 248-249 (2010).
40. Kircher, M. *et al.* A general framework for estimating the relative pathogenicity of human genetic variants. *Nat Genet* **46**, 310-5 (2014).
41. Boyle, A.P. *et al.* Annotation of functional variation in personal genomes using RegulomeDB. *Genome Research* **22**, 1790-1797 (2012).
42. Pers, T.H. *et al.* Biological interpretation of genome-wide association studies using predicted gene functions. *Nat Commun* **6**, 5890 (2015).
43. Kichaev, G. & Pasaniuc, B. Leveraging Functional-Annotation Data in Trans-ethnic Fine-Mapping Studies. *Am J Hum Genet* **97**, 260-71 (2015).
44. Benner, C. *et al.* FINEMAP: efficient variable selection using summary data from genome-wide association studies. *Bioinformatics* **32**, 1493-501 (2016).

45. Barbeira, A. *et al.* Integrating tissue specific mechanisms into GWAS summary results. *bioRxiv* (2017).
46. Giambartolomei, C. *et al.* Bayesian test for colocalisation between pairs of genetic association studies using summary statistics. *PLoS Genet* **10**, e1004383 (2014).
47. Gharib, S.A. *et al.* Integrative pathway genomics of lung function and airflow obstruction. *Hum Mol Genet* **24**, 6836-48 (2015).
48. Barreto-Luis, A. *et al.* A pathway-based association study reveals variants from Wnt signalling genes contributing to asthma susceptibility. *Clin Exp Allergy* **47**, 618-626 (2017).
49. Liu, J. *et al.* Targeting Wnt-driven cancer through the inhibition of Porcupine by LGK974. *Proc Natl Acad Sci U S A* **110**, 20224-9 (2013).
50. Arredouani, M. *et al.* The scavenger receptor MARCO is required for lung defense against pneumococcal pneumonia and inhaled particles. *J Exp Med* **200**, 267-72 (2004).
51. Dorrington, M.G. *et al.* MARCO is required for TLR2- and Nod2-mediated responses to *Streptococcus pneumoniae* and clearance of pneumococcal colonization in the murine nasopharynx. *J Immunol* **190**, 250-8 (2013).
52. Moffatt, M.F. *et al.* A large-scale, consortium-based genomewide association study of asthma. *N Engl J Med* **363**, 1211-1221 (2010).
53. Torgerson, D.G. *et al.* Meta-analysis of genome-wide association studies of asthma in ethnically diverse North American populations. *Nat Genet* **43**, 887-92 (2011).
54. Bahleda, R. *et al.* Phase I dose-escalation studies of roniciclib, a pan-cyclin-dependent kinase inhibitor, in advanced malignancies. *Br J Cancer* **116**, 1505-1512 (2017).
55. Yu, Q.B., Li, H.M., Li, L.L., Wang, S.Y. & Wu, Y.B. Sevoflurane downregulates interleukin-6 and interleukin-8 levels in patients after cardiopulmonary bypass surgery: a meta-analysis. *Genet Mol Res* **14**, 19016-27 (2015).
56. Cho, E.J., Yoon, J.H., Hong, S.J., Lee, S.H. & Sim, S.B. The effects of sevoflurane on systemic and pulmonary inflammatory responses after cardiopulmonary bypass. *J Cardiothorac Vasc Anesth* **23**, 639-45 (2009).
57. Uhlig, C. *et al.* Effects of Volatile Anesthetics on Mortality and Postoperative Pulmonary and Other Complications in Patients Undergoing Surgery: A Systematic Review and Meta-analysis. *Anesthesiology* **124**, 1230-45 (2016).
58. Jabaudon, M. *et al.* Sevoflurane for Sedation in Acute Respiratory Distress Syndrome. A Randomized Controlled Pilot Study. *Am J Respir Crit Care Med* **195**, 792-800 (2017).
59. Hirshman, C.A. & Emala, C.W. Actin reorganization in airway smooth muscle cells involves Gq and Gi-2 activation of Rho. *Am J Physiol* **277**, L653-61 (1999).
60. Billington, C.K. & Penn, R.B. Signaling and regulation of G protein-coupled receptors in airway smooth muscle. *Respir Res* **4**, 2 (2003).
61. Lange, P. *et al.* Lung-Function Trajectories Leading to Chronic Obstructive Pulmonary Disease. *N Engl J Med* **373**, 111-22 (2015).
62. Fletcher, C. & Peto, R. The natural history of chronic airflow obstruction. *Br Med J* **1**, 1645-8 (1977).
63. Martinez, F.D. Early-Life Origins of Chronic Obstructive Pulmonary Disease. *N Engl J Med* **375**, 871-8 (2016).
64. Hankinson, J.L., Odencrantz, J.R. & Fedan, K.B. Spirometric reference values from a sample of the general U.S. population. *Am J Respir Crit Care Med* **159**, 179-87 (1999).
65. Yang, J. *et al.* Genomic inflation factors under polygenic inheritance. *Eur J Hum Genet* **19**, 807-12 (2011).
66. Genomes Project, C. *et al.* An integrated map of genetic variation from 1,092 human genomes. *Nature* **491**, 56-65 (2012).

67. Maurano, M.T. *et al.* Systematic localization of common disease-associated variation in regulatory DNA. *Science* **337**, 1190-5 (2012).
68. Hnisz, D. *et al.* Super-enhancers in the control of cell identity and disease. *Cell* **155**, 934-47 (2013).
69. Forrest, A.R.R. *et al.* A promoter-level mammalian expression atlas. *Nature* **507**, 462-+ (2014).
70. Farh, K.K.H. *et al.* Genetic and epigenetic fine mapping of causal autoimmune disease variants. *Nature* **518**, 337-343 (2015).
71. Dunham, I. *et al.* An integrated encyclopedia of DNA elements in the human genome. *Nature* **489**, 57-74 (2012).
72. Kichaev, G. *et al.* Integrating Functional Data to Prioritize Causal Variants in Statistical Fine-Mapping Studies. *Plos Genetics* **10**(2014).
73. Pickrell, J.K. Joint analysis of functional genomic data and genome-wide association studies of 18 human traits. *Am J Hum Genet* **94**, 559-73 (2014).
74. Gamazon, E.R. *et al.* A gene-based association method for mapping traits using reference transcriptome data. *Nat Genet* **47**, 1091-8 (2015).

Author Contributions

Study-level Design and Data Collection: ABW, TS, MKL, LL, JCL, AVS, TMB, MFF, WG, TSA, WT, CO, QD, KdJ, MKW, XQW, RN, FPH, MG, TBH, RK, SRH, LP, KMB, YL, EGH, JGW, JMV, JS, RGB, RdM, AMBM, LIL, ACM, CMS, JCC, SBK, RJS, KC, JIR, TNB, FCW, NF, JAB, RCK, KL, MM, MAP, FRR, KDT, VG, KEN, MF, BMP, RHM, GO, TH, CCL, PC, JS, WJK, JRA, LL, HMB, BT, SSR, DOMK, BLH, AGU, GGB, SAG, JD, AM, SJL

Study-level Data Analysis: ABW, TS, MKL, LL, JCL, AVS, TMB, MFF, WG, TSA, WT, CO, QD, KdJ, MKW, XQW, RN, FPH

Meta-analysis: ABW, TS, JJ, JD, SJL

Pulmonary Function Look-up Replication in the UK BiLEVE Study: VEJ

COPD Look-up in the ICGC Consortium: BDH, MHC, International COPD Genetics Consortium Investigators

eQTL Look-up: ABW, LL, MO, TH, MvdB, RJ, YB, DCN, DDS

Functional Data Analysis (LD Score, DEPICT, PAINTOR, FINEMAP, S-PrediXcan, or Druggable Targets): MKL, NT, JNN, TW, GK, HHH, HKI, AM

Writing Group: ABW, TS, MKL, NT, JNN, LL, GGB, SAB, JD, AM, SJL

Manuscript Review: all authors

Acknowledgements

We thank Huling Li for expert technical assistance and Dr. Frank Day for computational support, both from the National Institute of Environmental Health Sciences (NIEHS), and Dr. Louise Wain, University of Leicester, for critical reviews of the manuscript. Supported in part by the Intramural Research Program of the National Institutes of Health, NIEHS. Infrastructure for the CHARGE Consortium is supported in part by the National Heart, Lung, and Blood Institute grant R01HL105756. Study specific funding and acknowledgments can be found in the Supplemental Material.

Disclosures

J.C.L. is currently an employee of GNS Healthcare. W.T. is currently an employee of Boehringer Ingelheim Pharmaceuticals. D.C.N. is an employee of Merck Research Laboratories. B.M.P. serves on the DSMB of a clinical trial funded by the manufacturer (Zoll LifeCor) and on the Steering Committee of the Yale Open Data Access Project funded by Johnson & Johnson. M.H.C. has received grant support from GlaxoSmithKline.

**Table 1. Sample size and location of studies included in the CHARGE consortium
1000 Genomes and pulmonary function meta-analysis**

Study ^a	Country	Sample Size by Ancestry			
		European	African	Hispanic/Latino	Asian
AGES	Iceland	1620			
ALHS	United States	2844			
ARIC	United States	8878	1837		
CARDIA	United States	1580	883		
CHS	United States	3135	566		
FamHS	United States	1679			
FHS	United States	7689			
GOYA	Denmark	1456			
HCHS/SOL	United States			11775	
HCS	Australia	1822			
Health ABC	United States	1472	943		
Healthy Twin	South Korea				2098
JHS	United States		2015		
KARE3	South Korea				7861
Lifelines	Netherlands	11851			
LLFS	United States and Denmark	3787			
MESA	United States	1339	863		
NEO	Netherlands	5460			
1982 Pelotas	Brazil	1357	1322		
RS I	Netherlands	1232			
RS II	Netherlands	1135			
RS III	Netherlands	2216			
	Total	60552	8429	11775	9959

^aAGES Age Gene Environment Susceptibility Study; ALHS Agricultural Lung Health Study; ARIC Atherosclerosis Risk in Communities Study; CARDIA Coronary Artery Risk Development in Young Adults; CHS Cardiovascular Health Study; FamHS Family Heart Study; FHS Framingham Heart Study; GOYA Genetics of Overweight Young Adults Study; HCHS/SOL Hispanic Community Health Study/Study of Latinos; HCS Hunter Community Study; JHS Jackson Heart Study; KARE3 Korean Association Resource Phase 3 Study; LLFS Long Life Family Study; MESA Multi-Ethnic Study of Atherosclerosis; NEO Netherlands Epidemiology of Obesity Study; RS Rotterdam Study

Table 2. Top Variants from Novel Loci Discovered in the CHARGE Consortium Meta-Analysis of Pulmonary Function

Nearest Gene(s) ^a	Phenotype ^b	Population ^c	Top Variant	Chr:Pos	Coded Allele ^d	Coded Allele Freq	CHARGE Discovery			UK BiLEVE Replication	
							Beta ^e	SE	P-value	Direction of Effect Concordant with Discovery ^f	P-value ^g
Top Variants from Novel Loci Discovered in the European and Multiethnic Meta-analyses and Tested for Replication in the UK BiLEVE Study											
<i>LOC728989</i>	FVC	Eur	rs12724426	1:146494027	a	0.21	-36.75	6.63	2.95E-08	Yes	8.10E-01
<i>DCAF8</i>	FEV ₁ /FVC	ME	rs11591179	1:160206067	t	0.45	-0.002	0.0003	3.48E-08	Yes	3.69E-01
<i>NR5A2</i>	FVC	ME	rs2821332	1:200085714	a	0.47	14.50	2.51	7.65E-09	Yes	1.20E-02
<i>PIK3C2B</i>	FEV ₁	ME	rs12092943	1:204434927	t	0.74	-14.57	2.67	4.83E-08	Yes	1.94E-01
<i>CENPF,KCNK2</i>	FVC	Eur	rs512597	1:215095003	t	0.81	-24.26	4.12	3.92E-09	Yes	9.06E-02
<i>C1orf140,DUSP10</i>	FVC	Eur	rs6657854	1:221630555	a	0.72	-19.89	3.49	1.18E-08	Yes	1.07E-02
		ME	rs12046746	1:221635207	c	0.71	-16.99	2.81	1.41E-09	Yes	1.11E-02
	FEV ₁	ME	1:221765779:C_CA	1:221765779	i	0.12	-36.25	6.57	3.38E-08	Yes	2.07E-01
<i>RYR2</i>	FVC	ME	1:237929787:T_TCA	1:237929787	i	0.11	-37.17	6.79	4.46E-08	Yes	2.29E-01
<i>PKDCC, EML4</i>	FEV ₁	ME	rs963406	2:42355947	a	0.12	-23.13	4.18	3.17E-08	Yes	3.38E-03
<i>EDAR</i>	FVC	ME	rs17034666	2:109571508	a	0.23	-27.93	4.96	1.81E-08	Yes	4.83E-01
<i>KCNJ3, NR4A2</i>	FEV ₁ /FVC	ME	rs72904209	2:157046432	t	0.88	0.003	0.0005	3.09E-08	Yes	2.36E-02
<i>RBMS3</i>	FEV ₁ /FVC	Eur	rs17666332	3:29469675	t	0.72	0.003	0.0005	4.76E-08	Yes	6.89E-05
		ME	rs28723417	3:29431565	a	0.74	0.002	0.0004	1.77E-08	Yes	3.44E-04
<i>DNAH12</i>	FEV ₁	ME	rs79294353	3:57494433	a	0.92	-29.56	5.05	4.82E-09	Yes	2.94E-01
		ME	rs6778584	3:98815640	t	0.70	12.98	2.37	4.51E-08	Yes	2.93E-04
<i>DCBLD2,MIR548G</i>	FVC	ME	rs1404098	3:98806782	a	0.71	15.93	2.73	5.45E-09	Yes	4.66E-05
	FEV ₁ /FVC	ME	rs80217917	3:99359368	t	0.88	-0.003	0.0005	2.58E-08	Yes	1.15E-01
<i>AFAP1</i>	FEV ₁ /FVC	Eur	rs28520091	4:7846240	t	0.48	0.003	0.0004	2.17E-09	Yes	2.95E-02
		ME	rs28520091	4:7846240	t	0.44	0.0021	0.0004	8.40E-09	Yes	2.95E-02
<i>OTUD4,SMAD1</i>	FEV ₁	ME	rs111898810	4:146174040	a	0.20	-20.24	3.61	2.14E-08	Yes	1.16E-02
<i>AP3B1</i>	FEV ₁	Eur	rs252746	5:77392117	a	0.78	20.05	3.45	6.19E-09	Yes	2.04E-01
		Eur	rs12513481	5:77450828	c	0.23	-25.01	3.74	2.15E-11	Yes	3.84E-02
		ME	rs72776440	5:77440196	c	0.21	-21.30	3.21	3.20E-11	Yes	3.24E-02
<i>LINC00340</i>	FEV ₁ /FVC	Eur	rs1928168	6:22017738	t	0.51	0.003	0.0004	6.74E-14	Yes	7.95E-05
		ME	rs9350408	6:22021373	t	0.51	-0.003	0.0003	7.45E-14	Yes	1.97E-05
<i>SLC25A51P1,BAI3</i>	FEV ₁ /FVC	Eur	rs9351637	6:67863782	t	0.61	0.002	0.0004	2.89E-08	No	3.63E-01
<i>CENPW,RSPO3</i>	FVC	ME	rs11759026	6:126792095	a	0.72	-20.20	3.44	4.35E-09	Yes	1.51E-01
<i>AGMO</i>	FVC	ME	rs55905169	7:15506529	c	0.31	-17.57	3.09	1.28E-08	Yes	1.70E-03
<i>CNTNAP2</i>	FEV ₁ /FVC	Eur	rs1404154	7:146651409	t	0.99	-0.03	0.006	2.80E-08	NA	NA
<i>DMRT2,SMARCA2</i>	FVC	Eur	rs771924	9:1555835	a	0.42	-18.4	3.18	7.16E-09	Yes	1.75E-01
		ME	rs9407640	9:1574877	c	0.42	-16.82	3.03	2.87E-08	Yes	6.87E-02
	FEV ₁	ME	rs9407640	9:1574877	c	0.41	-14.48	2.65	4.77E-08	Yes	8.91E-02
<i>FLJ35282,ELAVL2</i>	FEV ₁ /FVC	ME	rs10965947	9:23588583	t	0.39	0.002	0.0004	2.70E-09	Yes	1.07E-04

<i>TMEM38B,ZNF462</i>	FEV ₁ /FVC	ME	rs2451951	9:109496630	t	0.47	0.002	0.0003	2.36E-08	Yes	3.72E-02
<i>JMJD1C</i>	FEV ₁	ME	rs7899503	10:65087468	c	0.25	21.16	2.84	8.70E-14	Yes	1.64E-03
	FEV ₁ /FVC	ME	rs75159994	10:64916064	t	0.77	-0.003	0.0004	6.09E-09	Yes	2.63E-02
<i>COMTD1,ZNF503-AS1</i>	FVC	ME	10:77002679:TC_T	10:77002679	d	0.22	22.36	4.10	4.89E-08	Yes	5.88E-01
<i>HTRA1</i>	FEV ₁ /FVC	ME	rs2293871	10:124273671	t	0.23	0.002	0.0004	1.51E-08	Yes	4.00E-02
<i>FAM168A</i>	FEV ₁ /FVC	ME	11:73280955:GA_G	11:73280955	d	0.20	0.004	0.0006	2.74E-08	Yes	2.32E-04
<i>KIRREL3-AS3,ETS1</i>	FVC	ME	rs73025192	11:127995904	t	0.12	-24.18	4.28	1.63E-08	Yes	2.53E-01
<i>RAB5B</i>	FEV ₁	ME	rs772920	12:56390364	c	0.72	13.86	2.49	2.48E-08	Yes	1.26E-03
	FEV ₁	Eur	rs10779158	12:85724096	a	0.34	15.89	2.90	4.36E-08	Yes	1.60E-03
<i>ALX1,RASSF9</i>	FVC	Eur	rs10779158	12:85724096	a	0.34	18.72	3.31	1.52E-08	Yes	1.28E-03
		ME	rs7971039	12:85724305	a	0.26	16.36	2.88	1.44E-08	Yes	4.37E-04
<i>CRADD</i>	FVC	ME	rs11107184	12:94184082	t	0.34	14.89	2.71	3.87E-08	Yes	1.84E-01
<i>CCDC41</i>	FVC	ME	rs10859698	12:94852628	a	0.21	21.19	3.84	3.49E-08	Yes	6.70E-01
<i>NCOR2,SCARB1</i>	FEV ₁	ME	rs11057793	12:125230287	t	0.75	17.66	3.24	4.78E-08	Yes	9.93E-02
<i>DDHD1,MIR5580</i>	FEV ₁ /FVC	ME	rs4444235	14:54410919	t	0.54	0.002	0.0004	4.03E-08	Yes	4.32E-05
<i>SQRDL,SEMA6D</i>	FVC	ME	rs4775429	15:46722435	t	0.17	40.23	7.21	2.45E-08	No	3.04E-01
<i>SMAD3</i>	FVC	ME	rs8025774	15:67483276	t	0.29	-20.87	2.92	9.34E-13	Yes	9.10E-03
<i>PDXDC2P</i>	FVC	ME	rs3973397	16:70040398	a	0.48	-22.38	4.05	3.31E-08	Yes	5.90E-03
<i>PMFBP1,ZFXH3</i>	FVC	ME	rs55771535	16:72252097	a	0.13	-29.88	4.83	6.38E-10	Yes	7.95E-01
<i>SUZ12P1</i>	FEV ₁	ME	rs62070631	17:29087285	a	0.15	20.26	3.64	2.57E-08	Yes	3.19E-04
<i>MED1,CDK12</i>	FVC	ME	rs8067511	17:37611352	t	0.80	18.30	3.20	1.08E-08	Yes	2.12E-05
		Eur	rs143246821	17:43685698	a	0.79	30.58	4.99	9.06E-10	Yes	1.31E-12
<i>LOC644172,CRHR1</i>		ME	rs186806998	17:43682323	t	0.82	29.50	4.70	3.47E-10	Yes	3.72E-12
	FVC	ME	rs150741403	17:43682405	c	0.85	35.83	5.97	1.94E-09	Yes	4.15E-10
<i>WNT3</i>		Eur	rs916888	17:44863133	t	0.75	20.53	3.48	3.76E-09	Yes	5.19E-07
	FEV ₁	ME	rs199525	17:44847834	t	0.80	18.85	3.08	9.59E-10	Yes	4.69E-11
	FVC	ME	rs199525	17:44847834	t	0.80	20.32	3.52	7.52E-09	Yes	3.35E-11
<i>SOGA2</i>	FEV ₁	ME	rs513953	18:8801351	a	0.29	-14.50	2.58	1.96E-08	Yes	2.23E-03
<i>CTAGE1,RBBP8</i>	FEV ₁	ME	rs7243351	18:20148531	t	0.45	12.31	2.25	4.69E-08	Yes	9.62E-02
<i>CABLES1</i>	FVC	ME	rs7238093	18:20728158	a	0.22	18.15	3.13	6.78E-09	Yes	6.76E-02
<i>DCC</i>		Eur	rs8089865	18:50957922	a	0.59	20.57	3.23	1.95E-10	Yes	3.10E-03
	FVC	ME	rs8089865	18:50957922	a	0.53	15.81	2.57	7.38E-10	Yes	3.10E-03
<i>TSHZ3</i>		Eur	rs1353531	19:31846907	t	0.14	-0.003	0.0006	4.53E-08	Yes	2.69E-05
	FEV ₁ /FVC	ME	rs9636166	19:31829613	a	0.86	0.003	0.0005	3.25E-09	Yes	1.38E-06
<i>ZNF337</i>	FEV ₁	ME	rs6138639	20:25669052	c	0.79	17.91	2.85	3.17E-10	Yes	1.54E-04
<i>C20orf112</i>	FEV ₁	ME	rs1737889	20:31042176	t	0.22	-16.82	3.07	4.17E-08	Yes	1.69E-03
<i>EYA2</i>	FVC	Eur	rs2236519	20:45529571	a	0.38	-18.06	3.28	3.51E-08	Yes	1.15E-04
<i>KLHL22,MED15</i>		Eur	rs4820216	22:20854161	t	0.15	-0.004	0.0006	1.53E-09	Yes	2.22E-04
	FEV ₁ /FVC	ME	rs4820216	22:20854161	t	0.13	-0.003	0.0005	2.61E-10	Yes	2.22E-04

Top Variants from Novel Loci Discovered in the African and Hispanic/Latino Meta-analyses

<i>RYR2</i>	FEV ₁	Afr	rs3766889	1:237941781	t	0.82	52.21	9.52	4.12E-08	NA	NA
<i>C2orf48, HPCAL1</i>	FEV ₁ /FVC	Afr	rs139215025	2:10418806	a	0.01	-0.07	0.01	9.03E-11	NA	NA
<i>EN1, MARCO</i>	FVC	Afr	rs114962105	2:119660943	a	0.98	178.48	32.44	3.77E-08	NA	NA
<i>CADPS</i>	FEV ₁ /FVC	Afr	rs1111793843	3:62386350	t	0.01	-0.05	0.008	1.97E-08	NA	NA
<i>ANKRD55, MAP3K1</i>	FEV ₁	Afr	rs11748173	5:55922145	t	0.21	67.07	10.72	3.91E-10	NA	NA
<i>HDC</i>	FEV ₁ /FVC	Afr	rs180930492	15:50555681	t	0.01	-0.07	0.01	2.59E-09	NA	NA
<i>LOC283867, CDH5</i>	FEV ₁ /FVC	Afr	rs144296676	16:66060569	t	0.99	-0.03	0.006	5.35E-09	NA	NA
<i>CPT1C</i>	FEV ₁ /FVC	Afr	rs147472287	19:50213396	t	0.01	-0.05	0.009	3.25E-08	NA	NA
<i>DKFZp686O1327, PABPC1P2</i>	FVC	His	rs6746679	2:147046592	a	0.56	-37.36	6.67	2.17E-08	NA	NA

^aNearest gene: indicates gene either harboring the variant or nearest to it. *CRHR1* locus also includes *ARHGAP27, LOC644172, MGC57346, CRHR1-IT1, LRRC37A4P*. *SMAD3* locus also includes *AAGAB, IQCH*. *MED1/CDK12* locus also includes *FBXL20*. *HTRA1* locus also includes *DMBT1*. *JMJD1C* locus also includes *EGR2, NRBF2, JMJD1C-AS1, REEP3*. *RAB5B* locus also includes *SOUX*. *ZNF337* locus also includes *ABHD12, PYGB, GINS1, NINL, NANP, FAM182B, LOC100134868*. *KLHL22/MED15* locus also includes *ZNF74, SCARF2*. *C1orf140/DUSP10* locus also includes *HLX*.

^bPhenotypes: FEV₁ forced expiratory volume in 1 second (in ml), FVC forced vital capacity (in ml), Ratio FEV₁/FVC (as a proportion)

^cAncestral/ethnic populations: Afr African, His Hispanic/Latino, Eur European, ME Multiethnic

^dAlleles for INDELS: I Insertion, D Deletion

^eAdditive effect of variant on pulmonary function, adjusting for age, age², sex, height, height², smoking status, pack-years of smoking, center (if multi-center study), ancestral principal components, and weight (for FVC only)

^fUK BiLEVE estimates are on a different scale (inverse normal transformation) therefore we report only whether the direction of effect is concordant between discovery and replication studies.

^gBased on a two-sided $p < 9.6 \times 10^{-4}$, 15 loci replicated for the same trait based on the lead variant from our analysis: *DCBLD2/MIR548G, SUZ12P1, CRHR1, WNT3, ZNF337, ALX1/RASSF9, MED1/CDK12, EYA2, RBMS3, LINC00340, FLJ35282/ELAVL2, DDHD1/MIR5580, TSHZ3, KLHL22/MED15, FAM168A*. Based on a one-sided p-value, an additional 4 loci replicated for the same trait based on the lead variant: *RAB5B, JMJD1C, AGMO*, and *C20orf112*. One other loci replicated for the same trait based on a different significant variant in the locus (*DCC*, rs8085778, replication $p = 2.80 \times 10^{-4}$, R^2 with lead variant=0.9) and another loci replicated for a different pulmonary function trait based on the lead variant (*PDXDC2P* discovered for FVC replicated for FEV₁, replication $p = 3.99 \times 10^{-4}$).

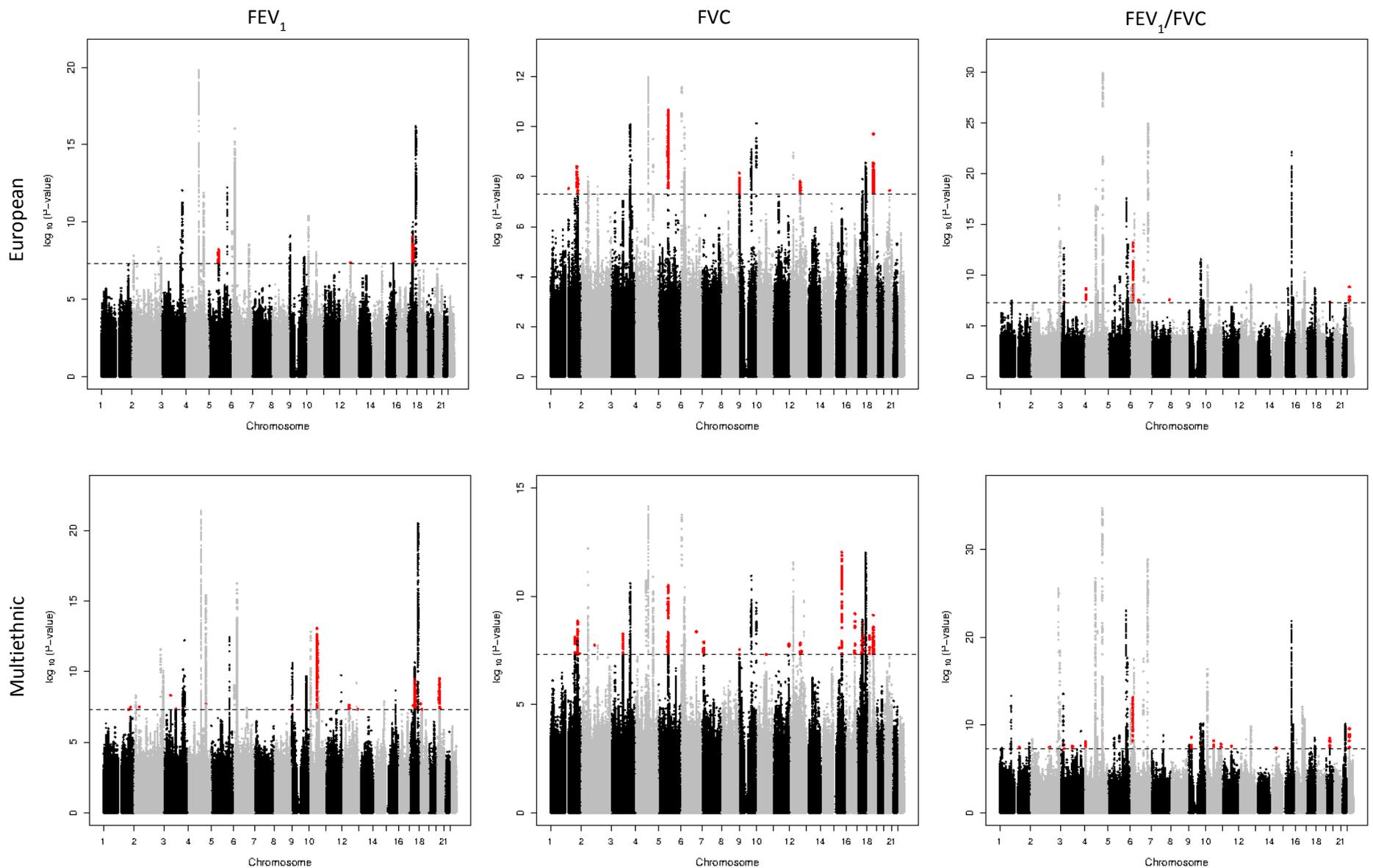


Figure 1. Manhattan plots for genome-wide association results for pulmonary function (FEV₁, FVC, and FEV₁/FVC) from European and multiethnic meta-analyses in CHARGE. Novel loci indicated by red. Significance level (5×10^{-8}) indicated by dashed line.