Twenty-two genetic loci in COPD overlap with population-based lung function and pulmonary fibrosis loci

Brian D. Hobbs[1,2], Kim de Jong[3,4], Maxime Lamontagne[5], Yohan Bossé[5,6], Nick Shrine[7], María Soler Artigas[7], Victoria E. Jackson[7], Louise V. Wain[7], Ian P. Hall[8], Annah B. Wyss[9], Stephanie J. London[9], Kari E. North[10], Nora Franceschini[10], David P. Strachan[11], Terri H. Beaty[12], John E. Hokanson[13], James D. Crapo[14], Peter J. Castaldi[1,15], Robert P. Chase[1], Traci M. Bartz[16,17], Susan R. Heckbert[16,18,19], Bruce M. Psaty[16,19,20,], Sina A. Gharib[21], Pieter Zanen[22], Jan W. Lammers[23], Matthijs Oudkerk[24], H. J. Groen[25], Nick Locantore[26], Ruth Tal-Singer[26], Stephen I. Rennard[27,28], Wim Timens[29], Peter D. Paré[30], Jeanne C. Latourelle[31], Josée Dupuis[32,33], George T. O'Connor[33,34], Jemma B. Wilk[33], Woo Jin Kim[35], Mi Kyeong Lee[35], Yeon-Mok Oh[36], Judith M. Vonk[3,4], Harry J. de Koning[37], Shuguang Leng[38], Steven A. Belinsky[38], Yohannes Tesfaigzi[38], Ani Manichaikul[39,40], Xin-Qun Wang[40], Stephen S. Rich[39,40], R Graham Barr[41], David Sparrow[42], Augusto L. Litonjua[1,2], Per Bakke[43], Amund Gulsvik[43], Lies Lahousse[44,45], Guy G. Brusselle[44,45,46], Bruno H. Stricker[44,47,48,49], André G. Uitterlinden[44,48,49], Elizabeth J. Ampleford[50], Eugene R. Bleecker[50], Prescott G. Woodruff[51], Deborah A. Meyers[50], Dandi Qiao[1], David A. Lomas[52], Jae-Joon Yim[53], Deog Kyeom Kim[54], Iwona Hawrylkiewicz[55], Pawel Sliwinski[55], Megan Hardin[1,2], Tasha E. Fingerlin[56,57], David A. Schwartz[56,58,59], Dirkje S. Postma[4,25], William MacNee[60], Martin D. Tobin[7,61], Edwin K. Silverman[1,2], H. Marike Boezen[3,4], *Michael H. Cho[1,2], COPDGene Investigators, ECLIPSE Investigators, LifeLines Investigators, SPIROMICS Research Group, International COPD Genetics Network Investigators, UK BiLEVE Investigators, International COPD Genetics Consortium

1 Channing Division of Network Medicine, Brigham and Women's Hospital, Boston, MA, USA
2 Division of Pulmonary and Critical Care Medicine, Brigham and Women's Hospital, Boston, MA, USA
3 University of Groningen, University Medical Center Groningen, Department of Epidemiology, Groningen, the Netherlands
4 University of Groningen, University Medical Center Groningen, Groningen Research Institute for Asthma and COPD (GRIAC), Groningen, the Netherlands
5 Institut universitaire de cardiologie et de pneumologie de Québec, Québec, Canada
6 Department of Molecular Medicine, Laval University, Québec, Canada
7 Genetic Epidemiology Group, Department of Health Sciences, University of Leicester, Leicester, UK
8 Division of Respiratory Medicine, Queen's Medical Centre, University of Nottingham, Nottingham, UK
9 Epidemiology Branch, National Institute of Environmental Health Sciences, National Institutes of Health, Department of Health and Human Services, Research Triangle Park, NC, USA
10 Department of Epidemiology, University of North Carolina, Chapel Hill, NC, USA
11 St George's, University of London, Cranmer Terrace, London SW17 0RE, UK
12 Johns Hopkins University Bloomberg School of Public Health, Baltimore, MD, USA
13 Department of Epidemiology, University of Colorado Anschutz Medical Campus, Aurora, CO, USA
14 National Jewish Health, Denver, CO, USA
15 Division of General Internal Medicine, Brigham and Women's Hospital, Boston, MA, USA
16 Cardiovascular Health Research Unit, University of Washington, Seattle, WA, USA
17 Departments of Medicine and Biostatistics, University of Washington, Seattle, WA, USA
18 Department of Epidemiology, Unik of Washington, Seattle, WA, USA
19 Group Health Research Institute, Group Health Cooperative, Seattle, WA, USA
20 Departments of Epidemiology, Medicine and Health Services, University of Washington, Seattle, WA, USA
21 Computational Medicine Core, Center for Lung Biology, UW Medicine Sleep Center, Department of Medicine, University of Washington, Seattle, WA, USA

22 University Medical Center Utrecht, Department of Pulmonary Diseases, Utrecht, the Netherlands
23 Department of Pulmonology, University Medical Center Utrecht, University of Utrecht, Utrecht, the Netherlands
24 Department of Radiology, University Medical Center Groningen, University of Groningen, Groningen, the Netherlands
25 University of Groningen, University Medical Center Groningen, Department of Pulmonology, Groningen, the Netherlands
26 GSK R&D, King of Prussia, PA, USA
27 Pulmonary, Critical Care, Sleep and Allergy Division, Department of Internal Medicine, University of Nebraska Medical Center, Omaha, NE, USA
28 Clincal Discovery Unit, AstraZeneca, Cambridge, UK
29 Department of Pathology and Medical Biology, University of Groningen, University Medical Center Groningen, GRIAC Research Institute, Groningen, the Netherlands
30 University of British Columbia Center for Heart Lung Innovation and Institute for Heart and Lung Health, St Paul's Hospital, Vancouver, British Columbia, Canada
31 Department of Neurology, Boston University School of Medicine, Boston, MA, USA
32 Department of Biostatistics, Boston University School of Public Health, Boston, MA, USA
33 The National Heart, Lung, and Blood Institute's Framingham Heart Study, Framingham, MA, USA
34 Pulmonary Center, Department of Medicine, Boston University School of Medicine, Boston, MA, USA
35 Department of Internal Medicine and Environmental Health Center, School of Medicine, Kangwon National University, Chuncheon, South Korea
36 Department of Pulmonary and Critical Care Medicine, and Clinical Research Center for Chronic Obstructive Airway Diseases, Asan Medical Center, University of Ulsan College of Medicine, Seoul, South Korea
37 Department of Public Health, Erasmus Medical Center Rotterdam, Rotterdam, the Netherlands
38 Lovelace Respiratory Research Institute, Albuquerque, NM, USA
39 Center for Public Health Genomics, University of Virginia, Charlottesville, VA, USA
40 Department of Public Health Sciences, University of Virginia, Charlottesville, VA, USA
41 Department of Medicine, College of Physicians and Surgeons and Department of Epidemiology, Mailman School of Public Health, Columbia University, New York, NY, USA
42 VA Boston Healthcare System and Department of Medicine, Boston University School of Medicine, Boston, MA, USA
43 Department of Clinical Science, University of Bergen, Bergen, Norway
44 Department of Epidemiology, Erasmus Medical Center, Rotterdam, the Netherlands
45 Department of Respiratory Medicine, Ghent University Hospital, Ghent, Belgium
46 Department of Respiratory Medicine, Erasmus Medical Center, Rotterdam, the Netherlands
47 Netherlands Health Care Inspectorate, The Hague, the Netherlands
48 Department of Internal Medicine, Erasmus Medical Center, Rotterdam, the Netherlands
49 Netherlands Genomics Initiative (NGI)-sponsored Netherlands Consortium for Healthy Aging (NCHA), Leiden, the Netherlands
50 Center for Genomics and Personalized Medicine Research, Wake Forest University School of Medicine, Winston Salem, NC, USA
51 Cardiovascular Research Institute and the Department of Medicine, Division of Pulmonary, Critical Care, Sleep, and Allergy, University of California at San Francisco, San Francisco, CA, USA
52 University College London, London, UK
53 Division of Pulmonary and Critical Care Medicine, Department of Internal Medicine, Seoul National University College of Medicine, Seoul, South Korea
54 Seoul National University College of Medicine, SMG-SNU Boramae Medical Center, Seoul, South Korea

55 2nd Department of Respiratory Medicine, Institute of Tuberculosis and Lung Diseases, Warsaw, Poland
56 Center for Genes, Environment and Health, National Jewish Health, Denver, CO, USA
57 Department of Biostatistics and Informatics, University of Colorado Denver, Aurora, CO, USA
58 Department of Medicine, School of Medicine, University of Colorado Denver, Aurora, CO, USA
59 Department of Immunology, School of Medicine, University of Colorado Denver, Aurora, CO, USA
60 University of Edinburgh, Edinburgh, UK
61 National Institute for Health Research (NIHR) Leicester Respiratory Biomedical Research Unit, Glenfield Hospital, Leicester, UK

*Corresponding author:
Michael H. Cho (remhc@channing.harvard.edu)
tel: 617-525-0897
fax: 888-487-1078

Chronic obstructive pulmonary disease (COPD) is a leading cause of mortality worldwide[1]. We performed a genetic association in 15,256 cases and 47,936 controls, with replication of select top results ($P < 5x10^{-6}$) in 9,498 cases and 9,748 controls. In the combined meta-analysis, we identified 22 loci at genome-wide significance, of which 15 have been associated with lung function in general population samples, and 4 (*EEFSEC*, *DSP*, *MTCL1*, and *SFTPD*) are novel. We noted 2 loci shared with pulmonary fibrosis (*FAM13A* and *DSP*) but with opposite risk alleles for COPD. None of our loci overlapped with genome-wide associations for asthma; however, one locus has been implicated in the joint susceptibility to asthma and obesity. We also identified genetic correlation between COPD and asthma. Our findings highlight novel loci, demonstrate the importance of specific lung function loci to COPD, and identify potential regions of genetic overlap between COPD and other respiratory diseases.

COPD is characterized by persistent and progressive airflow limitation diagnosed by lung function testing[1]. While cigarette smoking is the major risk factor, susceptibility is also influenced by genetics[2-4]. We established the International COPD Genetics Consortium (ICGC) to coordinate efforts to find susceptibility loci[5]. We defined cases based on pre-bronchodilator evidence of moderate-to-severe airflow limitation by modified GOLD criteria[6]; controls had normal spirometry, and all analyses were adjusted for age and cigarette smoking (pack-years and smoking status). We performed a two-stage genome-wide association study (Figure 1). In Stage 1, we combined 26 cohorts (Supplemental Table S1) containing 63,192 individuals (15,256 COPD cases and 47,936 controls). We selected 79 loci with $P < 5x10^{-6}$ and in analysis Stage 2, we tested them in the UK BiLEVE dataset (9,498 COPD cases and 9,748 controls) from the UK Biobank and performed an overall meta-analysis (Supplemental Table S3).

We identified 13 genome-wide significant ($P < 5x10^{-8}$) associations in Stage 1. Following the Stage 2 analysis, an additional 9 loci achieved genome-wide significance in the overall meta-

analysis (Table 1 and Figure 2). Of the 22 genome-wide significant loci described in our study, 9 have been previously described as genome- (or exome-) wide significant in studies of COPD[4,7-10]: *HHIP*, *CHRNA5*/15q25, *HTR4*, *FAM13A*, *RIN3*, *TGFB2*, *GSTCD/NPNT*, *CYP2A6*/19q13, and 16p11.2/*IL27*. An additional 8 loci: *ADGRG6*/*GPR126*, *THSD4*, *ADAM19*, *TET2*, *CFDP1*, *AGER*, *ARMC2*, and *RARB* have been previously described and replicated (Supplemental Table S6) in general population GWASs of two measures of lung function (FEV_1 and FEV_1/FVC) that are used in conjunction to diagnose COPD[11-17]. One locus near *PID1* was previously associated with FEV_1/FVC, but had not replicated in those studies[13,17]. Four loci are newly being described as genome-wide significant in association with either COPD or lung function: *EEFSEC*, *DSP, MTCL1,* and *SFTPD* (Figure 3).

To explore the potential function and causal genes for our novel loci, in addition to using publicly available datasets and prioritization tools (Supplemental Table S7), we also examined a larger set of lung expression quantitative trait loci (eQTL) in 1038 subjects, including subjects with COPD[18] (Supplemental Table S8). As eQTL are pervasive, we also attempted to determine whether our association signal co-localized[19] with an eQTL signal in lung tissue (Supplemental Table S9). We found strong evidence of co-localization (posterior probability > 0.8) for *DSP*, a major protein of desmosomes required for epidermal integrity [20], and *MTCL1*, important in epithelial-cell-specific microtubule stabilization[21,22], and expressed in respiratory epithelial cells[23]. Variants in strong LD with our top *MTCL1* variant rs647097 appear to have enhancer histone marks in fetal lung fibroblasts[24,25]. In contrast, we found no evidence of a strong eQTL signal or co-localization at our other two novel loci. At 3q21, *EEFSEC* is a potential candidate, as it is a paralog of *TUFM,* a top blood and lung eQTL gene for the 16p11.2/*IL27* COPD susceptibility locus[10], recently part of a novel COPD-related pathway involving *NLRX1*[26-28]. At 10q22, pulmonary surfactant-associated protein D (*SFTPD*) is the most likely candidate, as it is highly expressed in pneumocytes[23], and *sftpd* (-/-) mice develop pulmonary emphysema[29]. SFTPD has been explored as a COPD biomarker[30], and while

rs721917 is not an eQTL, polymorphisms in *SFTPD*, including rs721917, may lead to decreased surfactant protein D levels[31]; though the association of SFTPD polymorphisms with COPD susceptibility have been inconsistent. Our analysis also led to some additional insights into other previously described loci. We found evidence of COPD association and eQTL statistical co-localization in lung tissue (posterior probability > 0.8) for *THSD4*, *HHIP*, *AGER*, *CHRNA3*, and *RARB* (Supplemental Table S9). Additional data on eQTLs (Supplemental Table S8), cohort-specific associations at each locus (Supplemental Figures S1a-v), fine mapping (Supplemental Results and Table S20), and other supportive analysis for previously described and novel loci can be found in the Supplemental Materials.

We noted that our top variant at *DSP* (rs2076295) is also associated ($P = 1.1x10^{-19}$) with pulmonary fibrosis[32]. Recently, a re-sequencing study[33] at this locus identified a second fibrosis-associated variant (rs2744371). We performed additional analysis to investigate genetic overlap in this region using gwas-pw[34] (see Supplement). We found a posterior probability of > 0.99 for overlap at the *FAM13A* locus (top fibrosis SNP, rs2609255; $P_{fibrosis} = 2.20x10^{-11}$, $P_{COPD} = 1.9x10^{-7}$) and a posterior probability of 0.84 for overlap near *MAPT/KANSL1* (top $P_{fibrosis} = 8.87\times10^{-14}$, $P_{COPD} = 4.5x10^{-3}$); while the latter locus did not reach genome-wide significance in our study, we note its independent discovery in a genome-wide association in extremes of lung function[16]. Notably, for all four of these variants, the fibrosis risk allele is protective for development of COPD. Emphysema, a key component of COPD, and pulmonary fibrosis are both smoking-related lung diseases that have both shared and distinct pathophysiology[35-37], though genetic loci with opposing effects have not been previously described. Additional investigation of these loci as a well as a more comprehensive assessment of genetic overlap of COPD and pulmonary fibrosis may lead to insight into both disorders.

Because our analysis relied on a spirometric definition of COPD alone, we did not specifically exclude other causes of airway obstruction such as asthma, which can overlap with

COPD in adults[38]. To define COPD, we used pre-bronchodilator spirometry, which was available across all cohorts, and we included at least moderately affected cases (FEV_1 < 80% predicted). We examined the top set of genome-wide significant results in a subset of our largest cohorts with both pre- and post-bronchodilator data and densely imputed genotypes; overall, the effect sizes (mean difference = 0.001) and P values (mean $\log_{10}$ P value difference = 0.18) were similar (Supplemental Table S10 and Supplemental Figures S4 and S5). In addition, a recent GWAS of FEV_1, FVC, and FEV_1/FVC did not find substantial differences including and excluding subjects with asthma[16]. In the 79 variants tested in Stage 2, we found no significant difference in the OR for COPD association when including and excluding individuals with asthma (Supplemental Figure S6). We examined COPD associations of genome-wide significant asthma (and asthma-associated traits) loci from the NHGRI-EBI GWAS Catalog[39] (Supplemental Table S11). We also compared our COPD association results to the GABRIEL asthma study[40] (Supplemental Tables S12). None of the genome-wide significant loci from asthma and COPD overlapped. Further, no asthma or COPD loci showed Bonferroni-adjusted (for number of look-ups) significant association with the other disease, though several loci showed nominal (P<0.05) significance (Supplemental Tables S11 and S12). The 16p11.2 (*CCDC101*) locus has been described in the joint susceptibility to asthma and obesity[41]. COPD susceptibility is strongly related to cigarette smoking. Two of our loci (15q25 and 19q13) have been previously associated with smoking behavior[42,43], though we found no additional evidence of overlap in genome-wide significant variants described in the NHGRI-EBI GWAS Catalog[39] and Tobacco and Genetics Consortium GWAS[43] (Supplemental Tables S13-S15). In contrast to minimal overlap in genome-wide significant results with asthma and smoking, we discovered a significant overall genetic correlation of COPD with asthma ($r_{genetic}$ = 0.38, P = 6.2×10^{-5}) using LD score regression in our white subjects[44,45]. We also assessed genetic correlation with population-based lung function, pulmonary fibrosis, smoking behavior, and two common COPD comorbidities, coronary artery disease and osteoporosis. We identified significant correlation of COPD with lung

function and two aspects of smoking behavior, but not with common comorbidities or with pulmonary fibrosis (Figure 4). The lack of significant correlation of COPD with pulmonary fibrosis may indicate our overlapping loci for COPD and pulmonary fibrosis are not representative of a broader disease correlation; alternatively, it could reflect limited sample size or a mix of positive and negative genetic correlations across the genome for the diseases. In potential support of this latter hypothesis, and in contrast to the loci we describe in this study, are recent descriptions of rare variants in telomere genes predisposing to both emphysema, a key feature of COPD, and pulmonary fibrosis[37,46]. Our analysis of partitioned heritability identified COPD genetic association enrichment in fetal lung tissue (coefficient $P = 3.5 \times 10^{-7}$); other analyses also support lung tissue or lung cell types (Supplemental Materials).

Our study is, to our knowledge, the largest genome-wide association study of COPD cases to date and includes over 60,000 subjects (including 15,256 COPD cases) in our Stage 1 analysis. We chose to combine subjects of different ethnicities, hypothesizing that the benefit of shared risk factors across ethnicities would outweigh power loss due to heterogeneity. While methods have been developed that can more rigorously assess the degree of overlap and provide additional power in this setting[47], none of our non-white cohorts were sufficiently sized or powered for these analyses. COPD is also a highly heterogeneous disease; whether a more precise phenotypic definition would result in greater power is not clear. We used a staged study design and examined overall meta-analysis P-values to determine genome-wide significance. Thus, 9 loci (*TET2*, *CFDP1*, *TGFB2*, *AGER*, *ARMC2*, *PID1*, *MTCL1*, *SFTPD*, and *CYP2A6*) from our Stage 1 analysis, which only reached genome-wide significance in either the Stage 2 UK BiLEVE analysis or the overall meta-analysis, should be further replicated. However, six of these 9 association signals are significant if we consider a Bonferroni correction ($P < 6.3 \times 10^{-4}$) for the 79 variants tested in Stage 2. Further, 8 of these 9 variants become more significant from the Stage 1 analysis to the overall meta-analysis,

with the exception of *RARB*, which has a previously reported association with both lung function[13] and airflow obstruction[15] (Table 1).

The majority of our significant loci overlap with lung function loci, strengthening the foundation for investigating the relationship of lung function variability in the general population to risk of developing COPD. These loci are unlikely to reflect susceptibility for asthma or for cigarette smoking; however, our association as a whole shows evidence of shared heritability with asthma (supporting investigation into shared genetic etiologies for these diseases) and cigarette smoking behavior (despite adjustment for smoking in our statistical model). We identified enrichment for fetal lung cells, supporting a role for early life events contributing to future risk of COPD. Finally, we identify loci that overlap with pulmonary fibrosis, but with opposite risk alleles. Our study highlights the important contribution of genetic association studies to understanding COPD, not only by identifying novel loci, but also illustrating relationships with other pulmonary traits and diseases.

Online Methods

Study Cohorts

We invited investigators from 22 studies with genome-wide association data and COPD case-control or general population samples with spirometry to participate in a genome-wide association meta-analysis. Additionally, we included four cohorts with Illumina HumanExome v1.2 and custom genotyping based primarily on prior top results from a previously published COPD GWAS[4], using results with $P < 1x10^{-4}$ using plink '--clump' on the COPDGene non-Hispanic whites to perform linkage disequilibrium pruning ($r^2 < 0.8$), preferentially retaining both an imputed and genotyped top SNP at each locus. An additional group of variants was a candidate panel, based on results from a previous candidate gene analysis[48], as well as variants identified in association with lung function (supplementing the existing content on the array, which included variants from previous genome-wide association studies), including the lead SNP and a 200kb region around that SNP pruned for variants with $P < 0.01$ and $r^2 < 0.8$, and additional top-ranked SNPs for COPDGene-specific analyses for lung function, bronchodilator responsiveness, exacerbations, and SNPs from candidate genes.

The baseline characteristics of these 26 cohorts can be seen in Supplemental Table S1. Each cohort obtained approval from appropriate ethical/regulatory bodies; informed consent was obtained for all individuals. (Further cohort-specific methods can be found in the Data Supplement.) As most of these cohorts did not have post-bronchodilator spirometry, we used a modified definition of GOLD criteria based on pre-bronchodilator spirometry: forced expiratory volume in 1 second (FEV_1) < 80% and FEV_1 to forced vital capacity (FVC) ratio of < 0.7 for cases, and FEV_1 > 80% and $FEV_1/FVC > 0.7$ for controls. Logistic regression was performed in each cohort, adjusting for age, sex, pack-years of smoking, ever-smoking status, current-smoking status, and ancestry-based principal components, as appropriate for each study. Summary statistics were

assessed using EasyQC[49] version 10.1. More detailed cohort information, including cohort-specific methods, can be found in the Data Supplement.

Genome-wide association quality control

Summary statistics, including effect allele and other allele oriented to the + strand, effect allele frequency, chromosome and position (hg19), and imputation quality were uploaded to a secure site at the Brigham and Women's Hospital / Channing Division of Network Medicine. Quality control assessments included assessing allele frequencies versus 1000 Genomes reference, standard error versus sample size, and quantile-quantile plots. Variants with an imputation quality metric of < 0.3 (provided a higher threshold for imputation quality was not already implemented), a minor allele count (MAC) of < 20 using the effective sample size or the number of cases and adjusted for imputation quality where applicable, were set to missing. Variants were included for meta-analysis if they were present in at least 13 studies (those with European ancestry and at least 7 million markers passing all quality control filters).

Staged GWAS meta-analysis

In Stage 1 of the analysis, we used Metal[50,51] version 2011-03-25 to perform a fixed-effects meta-analysis of genome-wide data from 22 studies and four additional COPD cohorts genotyped on an Illumina HumanExome v1.2 platform with custom content; this content included a set of COPD candidate genes and regions identified from prior COPD GWAS efforts[4]. We adjusted for inflation using genomic control correction in each study. We included study populations with subjects of non-European ancestry in the overall analysis, and additionally examined results limited to study populations of European ancestry. To identify variants to test for association in Stage 2 in the UK BiLEVE study, we selected top results ($P < 5x10^{-6}$) from the Stage 1 meta-analysis. We selected one lead variant from the chromosome 15q25, *FAM13*, and *HHIP* regions, as all of these have been described in multiple COPD GWASs[4,7,8,15]. For the remainder of the regions, we performed linkage disequilibrium pruning using the plink2 --clump procedure with an r^2 of 0.5, additionally

examining these SNPs for the number of cohorts with passing quality control at each variant and including SNPs in strong LD (i.e., part of the same clump) with a lower degree of missingness. To identify independent results, we used GCTA-COJO[52,53] on the Stage 1 meta-analysis for variants with $P < 5x10^{-6}$ using the default distance of 10Mb. We used the COPDGene non-Hispanic whites (as the largest representative population) as the reference population for these analyses. An overall meta-analysis across the Stage 1 and Stage 2 (UK BiLEVE) cohorts was performed and variants with $P < 5x10^{-8}$ were considered genome-wide significant (Figure 1).

Lung eQTL analysis

Lung expression quantitative trait loci (eQTL) were calculated from 1,111 human subjects who underwent lung surgery at three academic sites, Laval University, University of British Columbia (UBC), and University of Groningen, henceforth referred to as Laval, UBC, and Groningen, respectively. This lung eQTL dataset has been described previously[18,54]. Briefly, 66.7% to 91.2% of the individuals in this study were current or former smokers and 24.2% to 35.3% had moderate to severe COPD (GOLD spirometry grade 2 to 4). Whole-genome gene expression profiling in the lung was performed on a custom Affymetrix array (GPL10379). Microarray pre-processing and quality controls were described previously[18,55,56]. Probe sequences were mapped to the human genome (hg19) using Bowtie[57] and probes not mapping to a coding region or having a common SNP (MAF ≥ 5%) in their sequence were removed. Expression data were adjusted for age, sex, and smoking status using residuals obtained with the robust fitting of linear models function (rlm) in the R statistical package MASS. Residual values deviating from the median by more than three standard deviations were filtered as outliers. Genotyping was carried on the Illumina Human 1M-Duo BeadChip array.

Twenty-one out of the 22 SNPs (in main manuscript Table 1) were genotyped or imputed in the three cohorts, i.e. Laval, UBC, and Groningen. One of the SNPs, rs7186831, was not well-imputed; a proxy, rs11865296 in modest linkage disequilibrium ($r^2 = 0.54$, 1000 genomes phase 3,

EUR) was used instead. These variants were tested for association with adjusted expression traits (43,465 probe sets) in the lung. SNPs within 1 Mb up and downstream of the transcription probe set were considered as local-eQTL. Distant-acting eQTLs were further than 1 Mb away or on a different chromosome. Association tests were carried with PLINK1.9[58,59] in each cohort and then meta-analyzed using Fisher's method. All local eQTL with nominal P value < 0.05 in the meta-analysis were considered. To provide an additional overall estimate of eQTL significance, we considered a Bonferroni correction threshold ([0.05/(22 SNPs x 43,465 probe sets) = P value < 5.2 x 10^{-8}]). Statistical analyses were performed in R3.2.3[60].

Co-localization Analysis

Co-localization of statistical signals between COPD genetic association and eQTL were examined using the coloc R package[19]. We used phenotypic summary statistics from whites with genome-wide association data and all eQTL results and examined 500kb flanks around the top 22 genome-wide significant associations found in the overall meta-analysis (Table 1).

Sensitivity Analysis

To estimate the effect of using pre- instead of post-bronchodilator lung function on our results, we examined the top set of genome-wide significant results in our largest cohorts with both pre- and post-bronchodilator data and densely imputed genotypes (COPDGene NHW and AA, ECLIPSE, NETT-NAS, and Norway / GenKOLS). Since subjects from these cohorts (except for COPDGene) were included based on post-bronchodilator values, including all subjects with COPD based on post-bronchodilator spirometry would lead to larger sample sizes and make comparison of P-values more difficult. Thus, we chose a random sample of post-bronchodilator cases and controls that matched the number of pre-bronchodilator cases and controls. We performed logistic regression using these equal sized set of pre- and post-bronchodilator cohorts, and meta-analyzed the results.

Asthma overlap analysis

We assessed the overlap between our results and known asthma susceptibility loci. We downloaded information on genome-wide significant ($P < 5x10^{-8}$) associations with asthma and asthma-related traits including asthma and hay fever, asthma (childhood onset), asthma (corticosteroid response), bronchodilator response in asthma, pulmonary function decline, and severe asthma in the NHGRI-EBI GWAS Catalog[39]. Additionally, we examined top associated variants (which were not genome-wide significant) in the susceptibility to the asthma-COPD overlap syndrome[61]. In all, we assessed the association statistics of 49 unique asthma-associated trait loci across 26 genomic regions in our Stage 1 meta-analysis results. We also examined the asthma association statistics of our top COPD loci from overall meta-analysis using publically available asthma GWAS data from the GABRIEL Consortium[40]. For COPD loci not present in the GABRIEL Consortium asthma GWAS data, we attempted to examine proxy SNPs in LD ($r^2 > 0.5$, 1000 genomes phase 1 CEU) with our top COPD loci.

To examine the genetic correlation[45] of COPD and asthma over the entire genome, we performed LD score regression[44] using summary statistics from publically available asthma GWAS data from the GABRIEL Consortium[40]. For all comparisons using LD score regression, we filtered to HapMap3 variants, limited to white subjects with genome-wide data, and filtered on missingness using default parameters in munge_sumstats.py. For the GABRIEL data, we required a variant to be present in at least 35 of the studies.

Smoking behavior overlap analysis

We downloaded information on genome-wide significant ($P < 5x10^{-8}$) associations with the traits "nicotine dependence" and "smoking behaviour" in the NHGRI-EBI GWAS Catalog[39]. We assessed these top smoking-associated SNPs in our Stage 1 meta-analysis results. We also assessed overlap of smoking and COPD in the publically available summary statistics from the 2010 Tobacco and Genetics Consortium GWAS[43]. We evaluated our top COPD loci associations from overall meta-analysis with both cigarettes per day and ever-smoking traits. For COPD risk SNPs not directly

analyzed in the Tobacco and Genetics Consortium GWAS, we attempted to examine proxy SNPs in LD ($r^2 > 0.5$, 1000 genomes phase1 CEU) with our top COPD loci.

To examine the genetic correlation[45] of COPD and smoking behaviours (cigarettes per day and ever-smoking status) over the entire genome, we performed LD score regression[44] using summary statistics from our current COPD study as noted above and publically available summary statistics from the 2010 Tobacco and Genetics Consortium GWAS[43].

Table 1. Overall study results showing 22 loci with genome-wide significant P values in overall meta-analysis following UK BiLEVE Stage 2 analysis.

rsID	Closest Gene	Locus	Risk Allele	Alt Allele	Risk Allele Mean Freq (Range)	Stage 1 OR (95% CI)	Stage 1 P Value	UK BiLEVE (Stage 2) OR (95% CI)	UK BiLEVE (Stage 2) P Value	**Overall Meta-Analysis P Value**
rs13141641	*HHIP*	4q31.21	T	C	0.594 (0.524-0.886)	1.231 (1.183-1.281)	1.16E-24	1.213 (1.16-1.267)	8.15E-18	**9.10E-41**
rs17486278	*CHRNA5*	15q25.1	C	A	0.351 (0.244-0.442)	1.224 (1.177-1.273)	2.61E-24	1.126 (1.077-1.178)	2.35E-07	**1.77E-28**
rs7733088	*HTR4*	5q32	G	A	0.602 (0.47-0.685)	1.178 (1.129-1.229)	4.40E-14	1.177 (1.127-1.229)	1.78E-13	**5.33E-26**
rs9399401	*ADGRG6*	6q24.1	T	C	0.724 (0.615-0.748)	1.141 (1.095-1.189)	3.59E-10	1.174 (1.119-1.232)	6.18E-11	**1.81E-19**
rs1441358	*THSD4*	15q23	G	T	0.332 (0.194-0.546)	1.132 (1.09-1.177)	2.06E-10	1.121 (1.071-1.172)	6.87E-07	**8.22E-16**
rs6837671	*FAM13A*	4q22.1	G	A	0.405 (0.364-0.582)	1.157 (1.115-1.2)	1.02E-14	1.066 (1.021-1.114)	3.75E-03	**7.48E-15**
rs11727735	*GSTCD*	4q24	A	G	0.936 (0.926-0.988)	1.266 (1.167-1.374)	1.55E-08	1.246 (1.144-1.357)	4.93E-07	**3.84E-14**
rs754388	*RIN3*	14q32.12	C	G	0.821 (0.804-0.865)	1.196 (1.136-1.258)	7.07E-12	1.108 (1.05-1.169)	1.85E-04	**4.96E-14**
rs113897301	*ADAM19*	5q33.3	AT	A	0.175 (0.052-0.187)	1.195 (1.13-1.263)	4.52E-10	1.127 (1.066-1.192)	2.79E-05	**1.58E-13**
rs2047409	*TET2*	4q24	A	G	0.618 (0.222-0.649)	1.101 (1.059-1.146)	1.58E-06	1.136 (1.087-1.188)	1.95E-08	**2.46E-13***
rs2955083	*EEFSEC*	3q21.3	A	T	0.881 (0.854-0.893)	1.195 (1.123-1.272)	2.00E-08	1.167 (1.093-1.246)	4.01E-06	**4.16E-13**
rs7186831	*CFDP1*	16q23.1	A	G	0.435 (0.226-0.472)	1.125 (1.07-1.183)	3.54E-06	1.116 (1.069-1.165)	6.63E-07	**1.12E-11***
rs10429950	*TGFB2*	1q41	T	C	0.731 (0.216-0.773)	1.117 (1.071-1.164)	1.83E-07	1.096 (1.044-1.15)	1.94E-04	**1.66E-10***
rs2070600	*AGER*	6p21.32	C	T	0.955 (0.85-0.987)	1.276 (1.151-1.414)	3.54E-06	1.208 (1.105-1.319)	2.96E-05	**5.94E-10***
rs17707300	*CCDC101*	16p11.2	C	T	0.373 (0.109-0.433)	1.122 (1.079-1.167)	6.24E-09	1.063 (1.018-1.111)	6.10E-03	**6.75E-10**
rs2806356	*ARMC2*	6q21	C	T	0.184 (0.052-0.242)	1.125 (1.071-1.181)	2.84E-06	1.116 (1.057-1.178)	6.88E-05	**8.34E-10***
rs16825267	*PID1*	2q36.3	C	G	0.929 (0.874-0.942)	1.242 (1.149-1.342)	5.22E-08	1.131 (1.045-1.224)	2.27E-03	**1.68E-09***
rs2076295	*DSP*	6p24.3	T	G	0.554 (0.442-0.581)	1.107 (1.067-1.149)	4.95E-08	1.061 (1.016-1.107)	7.45E-03	**3.97E-09**
rs647097	*MTCL1*	18p11.22	C	T	0.269 (0.259-0.399)	1.106 (1.06-1.154)	3.03E-06	1.089 (1.038-1.142)	4.66E-04	**6.14E-09***
rs1529672	*RARB*	3p24.2	C	A	0.831 (0.675-0.862)	1.162 (1.106-1.221)	2.37E-09	1.048 (0.991-1.109)	9.95E-02	**2.47E-08**
rs721917	*SFTPD*	10q22.3	G	A	0.422 (0.392-0.631)	1.092 (1.053-1.132)	2.11E-06	1.068 (1.023-1.115)	2.60E-03	**2.49E-08***
rs12459249	*CYP2A6*	19q13.2	C	T	0.662 (0.617-0.702)	1.126 (1.071-1.183)	2.89E-06	1.077 (1.029-1.127)	1.35E-03	**3.42E-08***

OR = odds ratio, CI = confidence interval. *Genome-wide significant in overall meta-analysis only

Figure 1. Study design showing cohorts used in each stage of the analysis.

Stage 1:

26 studies represented. European ancestry studies are shaded.

Study	COPD Cases	Controls
ARIC	1060	6164
B58	205	3665
CHS EA	736	1586
COPACETIC	397	1906
COPDGene NHW	3068	2110
ECLIPSE	1741	149
EOCOPD*	394	495
ICGN*	1852	557
EQTL	252	224
FHS	701	5110
LifeLines	466	9863
Lovelace	259	641
MESA Caucasian	167	754
NETT-NAS	376	435
Norway/GenKOLS	846	695
RS1	112	815
RS2	94	811
RS3	106	1596
SPIROMICS	571	175
TCGS-Poland *	307	311
CHS AA	138	292
COPDGene AA	910	1556
KARE	199	6741
MESA AA	94	532
MESA Hispanic	52	548
TCGS-Korea *	153	205
TOTAL	**15256**	**47936**

Stage 2:

Top results from Stage 1 analysis tested in UK BiLEVE study.

Study	COPD Cases	Controls
UK BiLEVE Never Smokers	3737	4871
UK BiLEVE Heavy Smokers	5761	4877
TOTAL	**9498**	**9748**

ARIC = Atherosclerosis Risk in Communities, B58 = British 1958 Birth Cohort, CHS = Cardiovascular Health Study, COPACETIC = COPD Pathology: Addressing Critical gaps, Early Treatment & Diagnosis and Innovative Concepts, ECLIPSE = Evaluation of COPD Longitudinally to Identify Predictive Surrogate End-points, eQTL = Lung Expression Quantitative Trait Loci Study, FHS = Framingham Heart Study, KARE = Korean Association Resource project, MESA = Multi-Ethnic Study of Atherosclerosis, NETT-NAS = National Emphysema Treatment Trial / Normative Aging Study, RS = Rotterdam Study, SPIROMICS = Subpopulations and intermediate outcome measures in COPD study , EOCOPD = Boston Early-Onset COPD Study, ICGN = International COPD Genetics Network, TCGS = Transcontinental COPD Genetics Study, UK BiLEVE = UK Biobank Lung Exome Variant Evaluation; NHW = Non-Hispanic white, AA = African American, EA = European American.* Studies without genome-wide array genotyping (custom genotyping)

Figure 2. P values for Stage 1 analysis (small open diamonds) with overlay of overall meta-analysis P values for SNPs analyzed in UK BiLEVE Stage 2 analysis (filled circles). Gene names in gray are previously described COPD or lung function (FEV_1 or FEV_1/FVC) loci; black are novel loci discovered in this study. The Stage 1 cohorts with available genotyping data (Supplemental Figures S1a-v) and the UK BiLEVE cohort determined the sample size for each top variant.

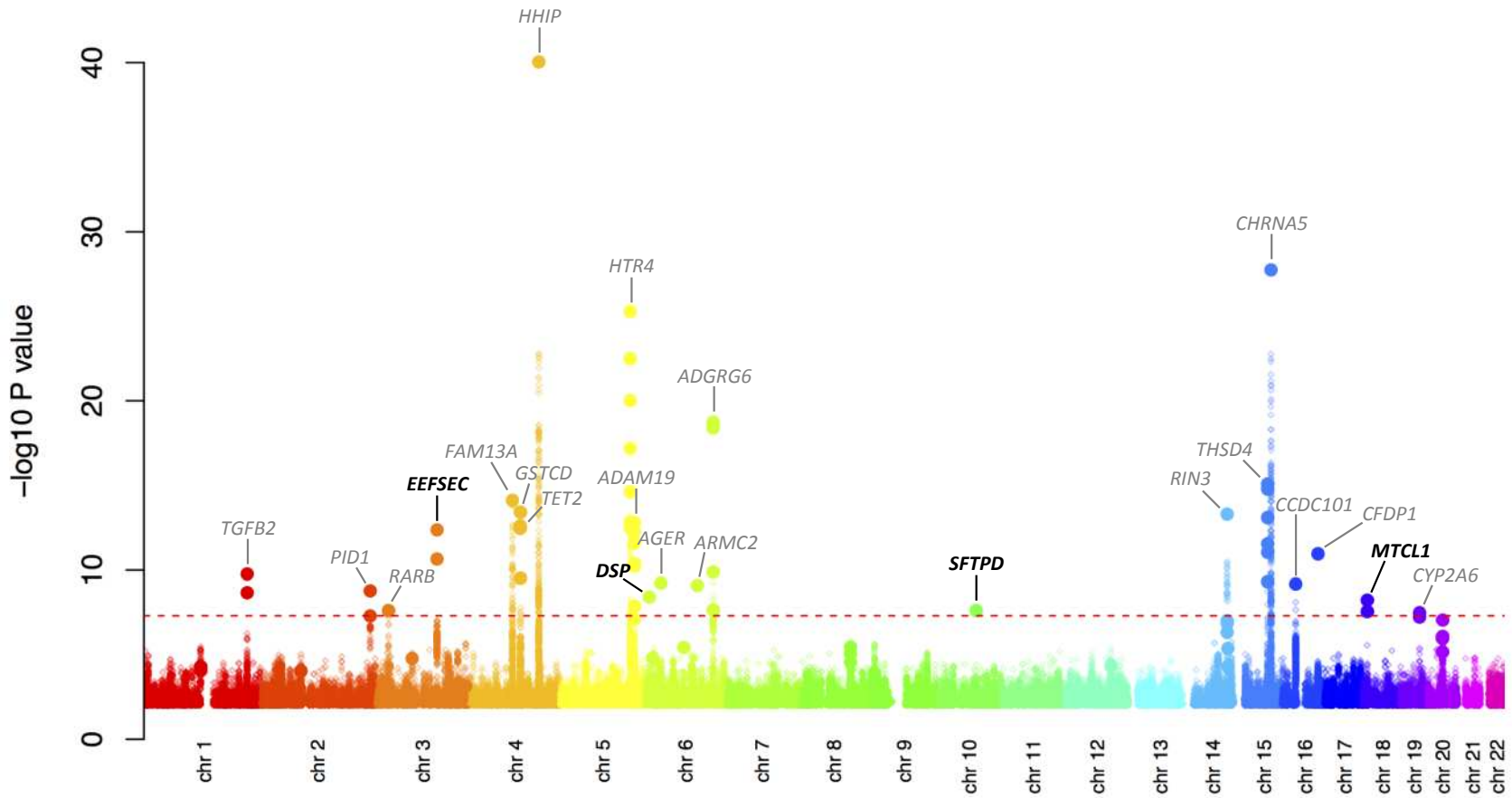


Figure 3a-d. Regional association for novel loci. LocusZoom plots showing regional association of variants at the four novel COPD loci. The point size is proportional to the sample size, where Stage 1 cohorts with available genotyping data (Supplemental Figures S1a-v) and the UK BiLEVE cohort determined the sample size for each top variant.

Fig 3a. rs2955083 (3q21.3, *EEFSEC*)

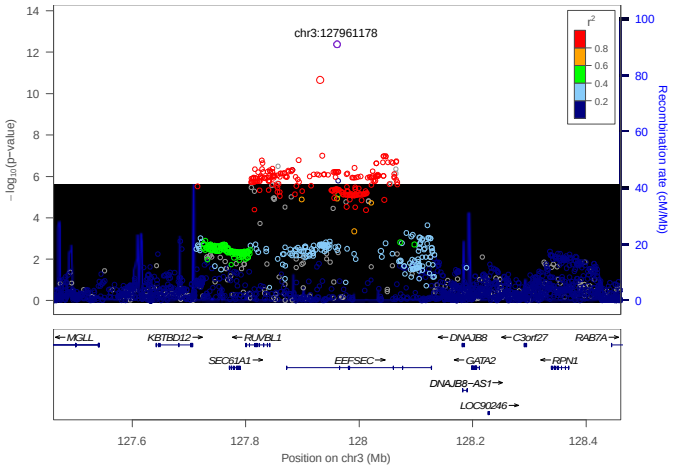


Fig 3b. rs2076295 (6p24.3, *DSP*)

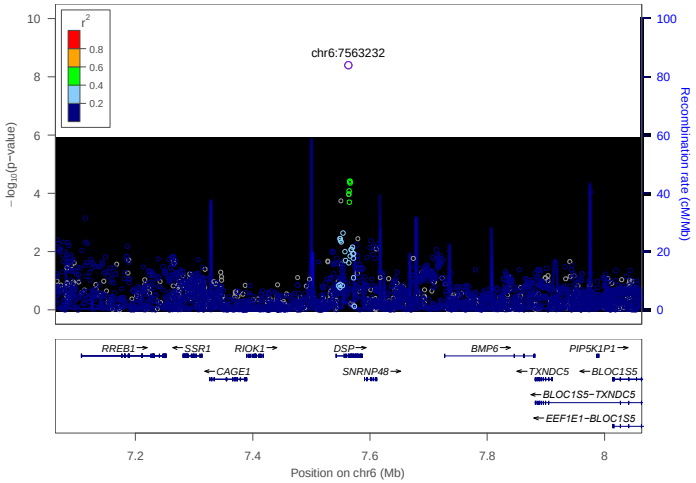


Fig 3c. rs647097 (18p11.22, *MTCL1*)

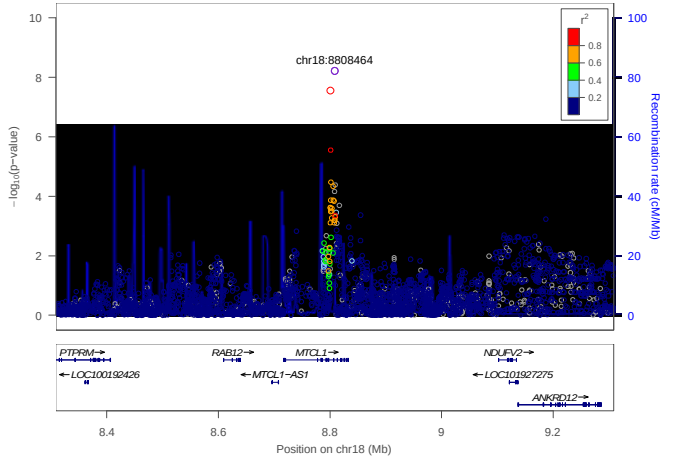


Fig 3d. rs721917 (10q22.3, *SFTPD*)

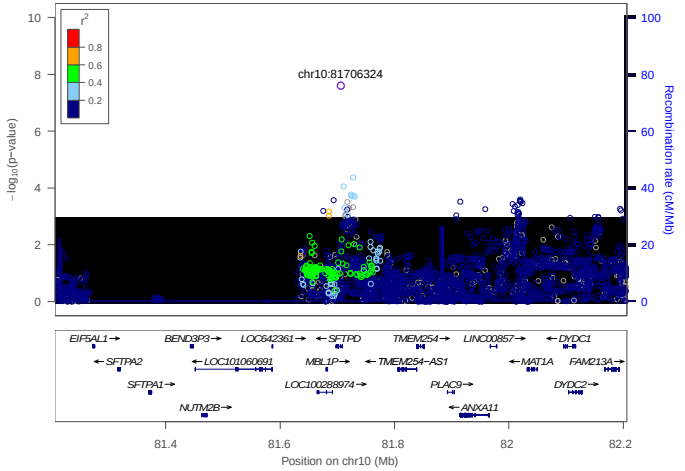
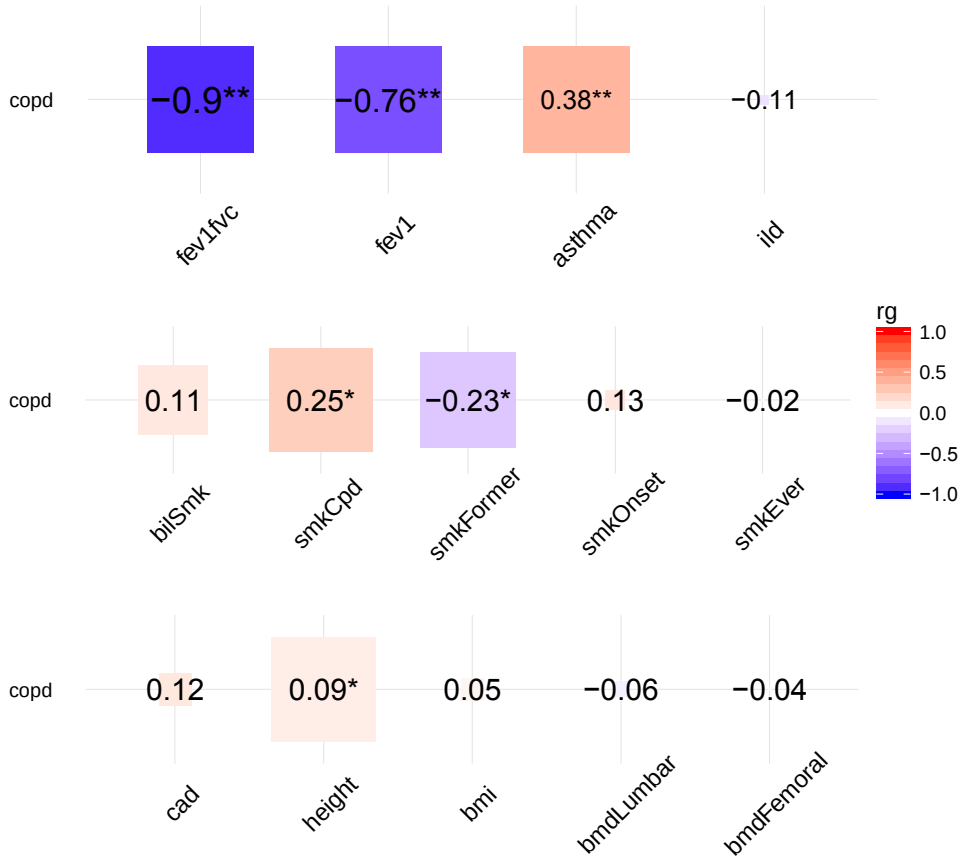


Figure 4 Genetic correlation (using LD score regression) between COPD and other traits. Shading and numbers represents strength of correlation. * indicates nominal (P < 0.05) significance, ** indicates significant after Bonferroni correction for number of pairwise comparisons. fev1fvc and fev1 = lung function (FEV_1FVC ratio and FEV_1 from CHARGE/SpiroMeta[13], asthma taken from the asthma GWAS by the GABRIEL Consortium[40], ild = pulmonary fibrosis from Fingerlin et al.[32,62], bilSmk = subset of smokers in the UK BiLEVE study[16], smkCpd = cigarettes per day smoking from the Tobacco and Genetics (TAG) Consortium[43], smkFormer = current versus former smokers from TAG, smkOnset = age of smoking initiation from TAG, smkEver = ever versus never smoking from TAG. cad = coronary artery disease from the CARDIoGRAM study[63], height[64] and bmi (body mass index)[65] from the GIANT consortium, bmdLumbar and bmdFemoral = lumbar and femoral bone mineral density, respectively, from the Genetic Factors for Osteoporosis (GeFOS) Consortium[66].



Author Contributions:

B.D.H. & M.H.C. contributed to the study concept and design, data analysis, statistical support, and manuscript writing. K.d.J., S.J.L., & D.P.S. contributed to the study concept and design & data analysis. N.S. & M.S.A. contributed to the data analysis and statistical support. T.H.B. & J.E.H. contributed to the study concept and design and statistical support. L.L. contributed to the data collection, data analysis, and statistical support. K.E.N. contributed to data collection and data analysis. J.D.C., B.M.P., R.T.S., G.T.O., Y.T., R.G.B., S.I.R., P.B., A.G., P.G.W., D.A.M., D.A.S, & E.K.S. contributed to the study concept and design and data collection. D.Q., T.A.F., M.L., Y.B., N.S., A.B.W., N.F., P.J.C., R.P.C., T.M.B., S.A.G., J.C.L., J.D., J.B.W., M.K.L., S.L., A.M., X.W., & E.J.A. contributed to the data analysis. L.V.W., I.P.H., P.D.P., D.S.P., W.M., M.D.T., & H.M.B. contributed to the study concept and design. S.R.H., P.A.D., W.J.K., Y.O., S.S.R., D.S., A.L.L., G.G.B., B.H.S., E.R.B., D.A.L., J.J.Y., D.K.K., I.H., P.S. & M.H. contributed to the data collection. All authors, including those whose initials are not listed above, contributed to the critical review and editing of the manuscript and approved the final version of the manuscript.

Funding:

Please refer to the Supplement for funding information.

Acknowledgments:

Please refer to the Supplement for full acknowledgements.

Conflict of Interest Statements:
I.P.H. has received grant support from Pfizer.
B.P. serves on the DSMB of a clinical trial funded by the manufacturer and on the Steering Committee of the Yale Open Data Access Project funded by Johnson & Johnson.
N.L. and R.T-S. are shareholders and employees of GSK.
S.I.R. is a current employee and shareholder at AstraZeneca. He has served as a consultant, participated in advisory boards, received honorarium for speaking or grant support from: American Board of Internal Medicine, Advantage Healthcare, Almirall, American Thoracic Society, AstraZeneca, Baxter, Boehringer Ingelheim, Chiesi, ClearView Healthcare, Cleveland Clinic, Complete Medical Group, CSL, Dailchi Sankyo, Decision Resources, Forest, Gerson Lehman, Grifols, GroupH, Guidepoint Global, Haymarket, Huron Consulting, Inthought, Johnson and Johnson, Methodist Health System – Dallas, NCI Consulting, Novartis, Pearl, Penn Technology, Pfizer, PlanningShop, PSL FirstWord, Qwessential, Takeda, Theron and WebMD.
J.C.L. is currently an employee of GNS Healthcare in Cambridge, MA.
J.B.W. was employed by Pfizer during the time this research was performed.
P.B. has received consulting and lecture fees from AstraZeneca, Boehringer Ingelheim, Chiesi, Novartis, and Teva.
L.L. has performed consultancy for Boehringer Ingelheim GmbH, received an AstraZeneca Scientific Award and travel support from Novartis, European Respiratory Society, and the Belgian Respiratory Society.
P.G.W. has consulted for Amgen, Sanofi, Novartis, Genentech/Roche, Boehringer-Ingelheim, Neostem and has had research grants from Pfizer and Genentech.
D.L. received grant support, honoraria and consultancy fees from GSK for work on the ICGN and ECLIPSE studies, and was a member and then Chaired the GSK Respiratory Therapy Area Board (2009-2015).

D.S.P. - The University of Groningen has received money for Professor Postma regarding a grant for research from Astra Zeneca, Chiesi, Genentec, GSK and Roche. Fees for consultancies were given to the University of Groningen by Astra Zeneca, Boehringer Ingelheim, Chiesi, GSK, Takeda and TEVA.E.K.S. has received honoraria and consulting fees from Merck, grant support and consulting fees from GSK, and honoraria and travel support from Novartis.
M.H.C. has received grant support from GSK.

References

1. Vestbo, J. *et al.* Global strategy for the diagnosis, management, and prevention of chronic obstructive pulmonary disease: GOLD executive summary. *Am J Respir Crit Care Med* **187**, 347-65 (2013).
2. Laurell, C.-B. & Eriksson, S. The electrophoretic alpha-1-globulin pattern of serum in alpha-1-antitrypsin deficiency. *Scandinavian Journal of Clinical and Laboratory Investigation* **15**, 132-140 (1963).
3. Silverman, E.K. *et al.* Genome-wide linkage analysis of severe, early-onset chronic obstructive pulmonary disease: airflow obstruction and chronic bronchitis phenotypes. *Hum Mol Genet* **11**, 623-32 (2002).
4. Cho, M.H. *et al.* Risk loci for chronic obstructive pulmonary disease: a genome-wide association study and meta-analysis. *Lancet Respir Med* **2**, 214-25 (2014).
5. Silverman, E.K. *et al.* Opportunities and challenges in the genetics of COPD 2010: an International COPD Genetics Conference report. *COPD* **8**, 121-35 (2011).
6. Mannino, D.M. & Buist, A.S. Global burden of COPD: risk factors, prevalence, and future trends. *Lancet* **370**, 765-73 (2007).
7. Pillai, S.G. *et al.* A genome-wide association study in chronic obstructive pulmonary disease (COPD): identification of two major susceptibility loci. *PLoS Genet* **5**, e1000421 (2009).
8. Cho, M.H. *et al.* Variants in FAM13A are associated with chronic obstructive pulmonary disease. *Nat Genet* **42**, 200-2 (2010).
9. Cho, M.H. *et al.* A genome-wide association study of COPD identifies a susceptibility locus on chromosome 19q13. *Hum Mol Genet* **21**, 947-57 (2012).
10. Hobbs, B.D. *et al.* Exome Array Analysis Identifies a Common Variant in IL27 Associated with Chronic Obstructive Pulmonary Disease. *Am J Respir Crit Care Med* **194**, 48-57 (2016).
11. Wilk, J.B. *et al.* A genome-wide association study of pulmonary function measures in the Framingham Heart Study. *PLoS Genet* **5**, e1000429 (2009).
12. 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).
13. 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).
14. 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).
15. Wilk, J.B. *et al.* Genome-wide association studies identify CHRNA5/3 and HTR4 in the development of airflow obstruction. *Am J Respir Crit Care Med* **186**, 622-32 (2012).
16. 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).
17. Soler Artigas, M. *et al.* Sixteen new lung function signals identified through 1000 Genomes Project reference panel imputation. *Nat Commun* **6**, 8658 (2015).

18. Hao, K. *et al.* Lung eQTLs to help reveal the molecular underpinnings of asthma. *PLoS Genet* **8**, e1003029 (2012).
19. Giambartolomei, C. *et al.* Bayesian test for colocalisation between pairs of genetic association studies using summary statistics. *PLoS Genet* **10**, e1004383 (2014).
20. Vasioukhin, V., Bowers, E., Bauer, C., Degenstein, L. & Fuchs, E. Desmoplakin is essential in epidermal sheet formation. *Nat Cell Biol* **3**, 1076-85 (2001).
21. Sato, Y. *et al.* The novel PAR-1-binding protein MTCL1 has crucial roles in organizing microtubules in polarizing epithelial cells. *J Cell Sci* **126**, 4671-83 (2013).
22. Sato, Y. *et al.* MTCL1 crosslinks and stabilizes non-centrosomal microtubules on the Golgi membrane. *Nat Commun* **5**, 5266 (2014).
23. Uhlen, M. *et al.* Proteomics. Tissue-based map of the human proteome. *Science* **347**, 1260419 (2015).
24. Ward, L.D. & Kellis, M. HaploReg: a resource for exploring chromatin states, conservation, and regulatory motif alterations within sets of genetically linked variants. *Nucleic Acids Res* **40**, D930-4 (2012).
25. Ward, L.D. & Kellis, M. HaploReg v4: systematic mining of putative causal variants, cell types, regulators and target genes for human complex traits and disease. *Nucleic Acids Res* (2015).
26. Lei, Y. *et al.* The mitochondrial proteins NLRX1 and TUFM form a complex that regulates type I interferon and autophagy. *Immunity* **36**, 933-46 (2012).
27. Lei, Y., Wen, H. & Ting, J.P. The NLR protein, NLRX1, and its partner, TUFM, reduce type I interferon, and enhance autophagy. *Autophagy* **9**, 432-3 (2013).
28. Kang, M.J. *et al.* Suppression of NLRX1 in chronic obstructive pulmonary disease. *J Clin Invest* **125**, 2458-62 (2015).
29. Wert, S.E. *et al.* Increased metalloproteinase activity, oxidant production, and emphysema in surfactant protein D gene-inactivated mice. *Proc Natl Acad Sci U S A* **97**, 5972-7 (2000).
30. Lomas, D.A. *et al.* Serum surfactant protein D is steroid sensitive and associated with exacerbations of COPD. *Eur Respir J* **34**, 95-102 (2009).
31. Foreman, M.G. *et al.* Polymorphisms in surfactant protein-D are associated with chronic obstructive pulmonary disease. *Am J Respir Cell Mol Biol* **44**, 316-22 (2011).
32. Fingerlin, T.E. *et al.* Genome-wide association study identifies multiple susceptibility loci for pulmonary fibrosis. *Nat Genet* **45**, 613-20 (2013).
33. Mathai, S.K. *et al.* Desmoplakin (DSP) Variants are Associated with Idiopathic Pulmonary Fibrosis. *American Journal of Respiratory and Critical Care Medicine* (2015).
34. Pickrell, J.K. *et al.* Detection and interpretation of shared genetic influences on 42 human traits. *Nat Genet* **48**, 709-17 (2016).
35. Washko, G.R. *et al.* Lung volumes and emphysema in smokers with interstitial lung abnormalities. *N Engl J Med* **364**, 897-906 (2011).
36. Chilosi, M., Poletti, V. & Rossi, A. The pathogenesis of COPD and IPF: distinct horns of the same devil? *Respir Res* **13**, 3 (2012).
37. Stanley, S.E. *et al.* Telomerase mutations in smokers with severe emphysema. *J Clin Invest* **125**, 563-70 (2015).

38. Soriano, J.B. *et al.* The proportional Venn diagram of obstructive lung disease: two approximations from the United States and the United Kingdom. *Chest* **124**, 474-81 (2003).
39. Welter, D. *et al.* The NHGRI GWAS Catalog, a curated resource of SNP-trait associations. *Nucleic Acids Res* **42**, D1001-6 (2014).
40. Moffatt, M.F. *et al.* A large-scale, consortium-based genomewide association study of asthma. *N Engl J Med* **363**, 1211-21 (2010).
41. Gonzalez, J.R. *et al.* A common 16p11.2 inversion underlies the joint susceptibility to asthma and obesity. *Am J Hum Genet* **94**, 361-72 (2014).
42. Thorgeirsson, T.E. *et al.* Sequence variants at CHRNB3-CHRNA6 and CYP2A6 affect smoking behavior. *Nat Genet* **42**, 448-53 (2010).
43. Consortium, T.a.G. Genome-wide meta-analyses identify multiple loci associated with smoking behavior. *Nat Genet* **42**, 441-7 (2010).
44. Finucane, H.K. *et al.* Partitioning heritability by functional annotation using genome-wide association summary statistics. *Nat Genet* **47**, 1228-35 (2015).
45. Bulik-Sullivan, B. *et al.* An atlas of genetic correlations across human diseases and traits. *Nat Genet* **47**, 1236-41 (2015).
46. Stanley, S.E. *et al.* Loss-of-function mutations in the RNA biogenesis factor NAF1 predispose to pulmonary fibrosis-emphysema. *Sci Transl Med* **8**, 351ra107 (2016).
47. Coram, M.A. *et al.* Leveraging Multi-ethnic Evidence for Mapping Complex Traits in Minority Populations: An Empirical Bayes Approach. *Am J Hum Genet* **96**, 740-52 (2015).
48. Castaldi, P.J. *et al.* The association of genome-wide significant spirometric loci with chronic obstructive pulmonary disease susceptibility. *Am J Respir Cell Mol Biol* **45**, 1147-53 (2011).
49. Winkler, T.W. *et al.* Quality control and conduct of genome-wide association meta-analyses. *Nat Protoc* **9**, 1192-212 (2014).
50. Abecasis, G., Li, Y. & Willer, C. METAL MetaAnalysis Helper. Version release 2011-03-25. URL: http://genome.sph.umich.edu/wiki/METAL_Program. (2011).
51. Willer, C.J., Li, Y. & Abecasis, G.R. METAL: fast and efficient meta-analysis of genomewide association scans. *Bioinformatics* **26**, 2190-1 (2010).
52. Yang, J., Lee, S.H., Goddard, M.E. & Visscher, P.M. GCTA: a tool for genome-wide complex trait analysis. *Am J Hum Genet* **88**, 76-82 (2011).
53. Yang, J. *et al.* Conditional and joint multiple-SNP analysis of GWAS summary statistics identifies additional variants influencing complex traits. *Nat Genet* **44**, 369-75, S1-3 (2012).
54. Lamontagne, M. *et al.* Refining susceptibility loci of chronic obstructive pulmonary disease with lung eqtls. *PLoS One* **8**, e70220 (2013).
55. Bosse, Y. *et al.* Molecular signature of smoking in human lung tissues. *Cancer Res* **72**, 3753-63 (2012).
56. Lamontagne, M. *et al.* Genetic regulation of gene expression in the lung identifies CST3 and CD22 as potential causal genes for airflow obstruction. *Thorax* **69**, 997-1004 (2014).
57. Langmead, B., Trapnell, C., Pop, M. & Salzberg, S.L. Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome Biol* **10**, R25 (2009).

58. Purcell, S. *et al.* PLINK: a tool set for whole-genome association and population-based linkage analyses. *Am J Hum Genet* **81**, 559-75 (2007).
59. Chang, C.C. *et al.* Second-generation PLINK: rising to the challenge of larger and richer datasets. *Gigascience* **4**, 7 (2015).
60. R Core Team (2016). R: A language and environment for statistical computing. (R Foundation for Statistical Computing. URL: http://www.R-project.org/, Vienna, Austria).
61. Hardin, M. *et al.* The clinical and genetic features of COPD-asthma overlap syndrome. *Eur Respir J* **44**, 341-50 (2014).
62. Fingerlin, T.E. *et al.* Genome-wide imputation study identifies novel HLA locus for pulmonary fibrosis and potential role for auto-immunity in fibrotic idiopathic interstitial pneumonia. *BMC Genet* **17**, 74 (2016).
63. Consortium, C.A.D. *et al.* Large-scale association analysis identifies new risk loci for coronary artery disease. *Nat Genet* **45**, 25-33 (2013).
64. 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).
65. Locke, A.E. *et al.* Genetic studies of body mass index yield new insights for obesity biology. *Nature* **518**, 197-206 (2015).
66. Zheng, H.F. *et al.* Whole-genome sequencing identifies EN1 as a determinant of bone density and fracture. *Nature* **526**, 112-7 (2015).