Expanded genetic landscape of chronic obstructive pulmonary disease reveals heterogeneous cell type and phenotype associations

Phuwanat Sakornsakolpat[1,2,49], Dmitry Prokopenko[1,3,49], Maxime Lamontagne[4], Nicola F. Reeve[5], Anna L. Guyatt[5], Victoria E. Jackson[5], Nick Shrine[5], Dandi Qiao[1], Traci M. Bartz[6,7,8], Deog Kyeom Kim[9], Mi Kyeong Lee[10], Jeanne C. Latourelle[11], Xingnan Li[12], Jarrett D. Morrow[1], Ma'en Obeidat[13], Annah B. Wyss[10], Per Bakke[14], R Graham Barr[15], Terri H. Beaty[16], Steven A. Belinsky[17], Guy G. Brusselle[18,19,20], James D. Crapo[21], Kim de Jong[22,23], Dawn L. DeMeo[1,24], Tasha E. Fingerlin[25,26], Sina A. Gharib[27], Amund Gulsvik[14], Ian P. Hall[28], John E. Hokanson[29], Woo Jin Kim[30], David A. Lomas[31], Stephanie J. London[10], Deborah A. Meyers[12], George T. O'Connor[32,33], Stephen I. Rennard[34,35], David A. Schwartz[25,36,37], Pawel Sliwinski[38], David Sparrow[39], David P. Strachan[40], Ruth Tal-Singer[41], Yohannes Tesfaigzi[17], Jørgen Vestbo[42], Judith M. Vonk[22,23], Jae-Joon Yim[43], Xiaobo Zhou[1], Yohan Bossé[4,44], Ani Manichaikul[45,46], Lies Lahousse[47,18], Edwin K. Silverman[1,24], H. Marike Boezen[22,23], Louise V. Wain[5], Martin D. Tobin[5,48], Brian D. Hobbs[1,24,50], Michael H. Cho[1,24,50], International COPD Genetics Consortium

1 Channing Division of Network Medicine, Brigham and Women's Hospital, Boston, MA, USA
2 Department of Medicine, Faculty of Medicine Siriraj Hospital, Mahidol University, Bangkok, Thailand
3 Genetics and Aging Research Unit, Department of Neurology, Massachusetts General Hospital, Boston, MA, USA
4 Institut universitaire de cardiologie et de pneumologie de Québec, Québec, Canada
5 Genetic Epidemiology Group, Department of Health Sciences, University of Leicester, Leicester, UK
6 Cardiovascular Health Research Unit, University of Washington, Seattle, WA, USA
7 Department of Medicine, University of Washington, Seattle, WA, USA
8 Department of Biostatistics, University of Washington, Seattle, WA, USA
9 Seoul National University College of Medicine, SMG-SNU Boramae Medical Center, Seoul, South Korea
10 Epidemiology Branch, National Institute of Environmental Health Sciences, National Institutes of Health, Department of Health and Human Services, Research Triangle Park, NC, USA
11 Department of Neurology, Boston University School of Medicine, Boston, MA, USA
12 Department of Medicine, University of Arizona, Tuscson, AZ
13 The University of British Columbia Center for Heart Lung Innovation, St Paul's Hospital, Vancouver, BC, Canada
14 Department of Clinical Science, University of Bergen, Bergen, Norway
15 Department of Medicine, College of Physicians and Surgeons and Department of Epidemiology, Mailman School of Public Health, Columbia University, New York, NY, USA
16 Department of Epidemiology, Johns Hopkins University Bloomberg School of Public Health, Baltimore, MD, USA
17 Lovelace Respiratory Research Institute, Albuquerque, NM, USA
18 Department of Epidemiology, Erasmus Medical Center, Rotterdam, the Netherlands
19 Department of Respiratory Medicine, Ghent University Hospital, Ghent, Belgium
20 Department of Respiratory Medicine, Erasmus Medical Center, Rotterdam, the Netherlands
21 Department of Medicine, Division of Pulmonary and Critical Care Medicine, National Jewish Health, Denver, CO, USA
22 University of Groningen, University Medical Center Groningen, Department of Epidemiology, Groningen, the Netherlands
23 University of Groningen, University Medical Center Groningen, Groningen Research Institute for Asthma and COPD (GRIAC), Groningen, the Netherlands
24 Division of Pulmonary and Critical Care Medicine, Brigham and Women's Hospital, Boston, MA, USA

25 Center for Genes, Environment and Health, National Jewish Health, Denver, CO, USA
26 Department of Biostatistics and Informatics, University of Colorado Denver, Aurora, CO, USA
27 Computational Medicine Core, Center for Lung Biology, UW Medicine Sleep Center, Department of Medicine, University of Washington, Seattle, WA, USA
28 Division of Respiratory Medicine, Queen's Medical Centre, University of Nottingham, Nottingham, UK
29 Department of Epidemiology, University of Colorado Anschutz Medical Campus, Aurora, CO, USA
30 Department of Internal Medicine and Environmental Health Center, School of Medicine, Kangwon National University, Chuncheon, South Korea
31 University College London, London, UK
32 The National Heart, Lung, and Blood Institute's Framingham Heart Study, Framingham, MA, USA
33 Pulmonary Center, Department of Medicine, Boston University School of Medicine, Boston, MA, USA
34 Pulmonary, Critical Care, Sleep and Allergy Division, Department of Internal Medicine, University of Nebraska Medical Center, Omaha, NE, USA
35 Clinical Discovery Unit, AstraZeneca, Cambridge, UK
36 Department of Medicine, School of Medicine, University of Colorado Denver, Aurora, CO, USA
37 Department of Immunology, School of Medicine, University of Colorado Denver, Aurora, CO, USA
38 2nd Department of Respiratory Medicine, Institute of Tuberculosis and Lung Diseases, Warsaw, Poland
39 VA Boston Healthcare System and Department of Medicine, Boston University School of Medicine, Boston, MA, USA
40 Population Health Research Institute, St. George's University of London, London, UK
41 GSK R&D, King of Prussia, PA, USA
42 School of Biological Sciences, University of Manchester, Manchester, UK
43 Division of Pulmonary and Critical Care Medicine, Department of Internal Medicine, Seoul National University College of Medicine, Seoul, South Korea
44 Department of Molecular Medicine, Laval University, Québec, Canada
45 Center for Public Health Genomics, University of Virginia, Charlottesville, VA, USA
46 Department of Public Health Sciences, University of Virginia, Charlottesville, VA, USA
47 Department of Bioanalysis, Ghent University, Ghent, Belgium
48 National Institute for Health Research (NIHR) Leicester Respiratory Biomedical Research Unit, Glenfield Hospital, Leicester, UK
49 These authors contributed equally
50 These authors jointly supervised the work

Correspondence should be addressed to M.H.C. (remhc@channing.harvard.edu).

Summary

Chronic obstructive pulmonary disease (COPD) is the leading cause of respiratory mortality worldwide. Genetic risk loci provide novel insights into disease pathogenesis. To broaden COPD genetic risk loci discovery and identify cell type and phenotype associations, we performed a genome-wide association study in 35,735 cases and 222,076 controls from the UK Biobank and additional studies from the International COPD Genetics Consortium. We identified 82 loci with P value < $5x10^{-8}$; 47 were previously described in association with either COPD or population-based lung function. Of the remaining 35 novel loci, 13 were associated with lung function in 79,055 individuals from the SpiroMeta consortium. Using gene expression and regulation data, we identified enrichment for loci in lung tissue, smooth muscle and several lung cell types. We found 9 shared genomic regions between COPD and asthma and 5 between COPD and pulmonary fibrosis. COPD genetic risk loci clustered into groups of quantitative

imaging features and comorbidity associations**.** Our analyses provide further support to the genetic susceptibility and heterogeneity of COPD.

Background

Chronic obstructive pulmonary disease (COPD) is a disease of enormous and growing global burden[1], ranked third as a global cause of death by the World Health Organization in 2016[2]. Environmental risk factors, predominately cigarette smoking, account for a large fraction of disease risk, but there is considerable variability in COPD susceptibility among individuals with similar smoking exposure. Studies in families and in populations demonstrate that genetic factors account for a substantial fraction of disease susceptibility. Similar to other adult-onset complex diseases, common variants likely account for the majority of population genetic susceptibility[3,4]. Our previous efforts identified 22 genome-wide significant loci[5]. Expanding the number of loci can lead to novel disease pathogenesis insights, not only through discovery of novel biology[6,7], but also through informing more global insights -- such as functional links between loci, and specific cell-types and phenotypes driving COPD genetic risk[5].

We performed a genome-wide association study combining previously described studies from the International COPD Genetics Consortium (ICGC) with additional subjects from the UK Biobank[8], a population-based study of several hundred thousand subjects with lung function and cigarette smoking assessment. We determined, through bioinformatic and computational analysis, the likely set of variants, genes, cell types, and biologic pathways implicated by these associations. Finally, we assessed our genetic findings for relevance to COPD-specific, respiratory, and other phenotypes.

Results

Genome-wide association study of COPD

We included a total of 257,811 individuals from 25 studies in the analysis, including studies from International COPD Genetics Consortium and UK Biobank (**Figure 1**). We defined COPD based on pre-bronchodilator spirometry according to modified Global Initiative for Chronic Obstructive Lung Disease (GOLD) criteria for moderate to very severe airflow limitation[9], as done previously[5]. This definition resulted in 35,735 cases and 222,076 controls (**Supplementary Table 1**). We tested association of COPD and 6,224,355 variants in a meta-analysis of 25 studies using a fixed-effects model. We found no evidence of confounding by population substructure using linkage disequilibrium score regression (LDSC) intercept (1.0377, s.e. 0.0094).

We identified 82 loci (defined using 2-Mb windows) at genome-wide significance ($P < 5 \times 10^{-8}$) (**Figures 1 and 2; and Supplementary Figures 1 and 2**). Forty-seven of 82 loci were previously described as genome-wide significant in COPD[5,10] or lung function[11–19] (**Supplementary Table 2**), leaving 35 novel loci (**Table 1**). We then sought to replicate these loci. Given the strong genetic correlation between population-based lung function and COPD, we tested the lead variant at each locus for association with FEV_1 or FEV_1/FVC in 79,055 individuals from SpiroMeta[20] (**Supplementary Table 3**). We identified 13 loci - *C1orf87, DENND2D, DDX1, SLMAP, BTC, FGF18, CITED2, ITGB8, STN1, ARNTL, SERP2, DTWD1,* and *ADAMTSL3* – that replicated using a Bonferroni correction for a one-sided $P < 0.05/35$; **Table 1**). Although not meeting the strict Bonferroni threshold, additional 14 novel loci were nominally significant in SpiroMeta (consistent direction of effect and one-sided $P < 0.05$): *ASAP2, EML4, VGLL4, ADCY5,*

HSPA4, CCDC69, RREB1, ID4, IER3, RFX6, MFHAS1, COL15A1, TEPP, and *THRA* (**Table 1**), and all 82 loci showed consistent direction of effect with either FEV_1 or FEV_1/FVC ratio in SpiroMeta (**Table 1** and **Supplementary Table 2**). We note that 9 of our 35 novel loci were recently described in a contemporaneous analysis of lung function in UK Biobank[20]. None of the novel loci appeared to be explained by cigarette smoking, and variant effect sizes in ever- and never-smokers and including and excluding self-reported asthmatics were similar (**Supplementary Results**). In addition, we found no significant differences in variant effects by sex (**Supplementary Results**). Including all 82 genome-wide significant variants, we explain up to 7.0% of the phenotypic variance in liability scale, using a 10% prevalence of COPD, and acknowledging that these effects are likely overestimated in the discovery sample. This represents a 48% increase in COPD phenotypic variance explained by genetic loci compared to the 4.7% explained by 22 loci reported in a recent GWAS of COPD[5].

Identification of secondary association signals

We used approximate conditional and joint analysis to find secondary signals at each of the 82 genome-wide significant loci. We found 82 secondary signals at 50 loci, resulting in a total of 164 independent associations in 82 loci (**Supplementary Table 4**). Of 50 loci containing secondary associations, 33 were at loci previously described for COPD or lung function, and 6 at Bonferroni-replicated novel loci. Of 82 secondary associations, 20 reached genome-wide significance ($P < 5 \times 10^{-8}$) (**Supplementary Table 4**). Of 61 novel (not previously described in COPD or lung function) independent associations, 21 reached a region-wise Bonferroni-corrected threshold (one-sided $P < 0.05$/novel independent association(s) in each locus) in unconditioned associations from SpiroMeta (**Methods** and **Supplementary Table 4**).

Tissue and specific cell types

In determining the tissue in which COPD genetic variants function to increase COPD risk, lung is the obvious tissue to consider. However, COPD is a systemic disease[21,22] and within the lung, which cell-types collectively contribute to disease pathogenesis is largely unknown. Furthermore, available databases often include cell types (e.g. smooth muscle) from other organs (e.g. the gastrointestinal tract). To identify putative causal tissues and cell types, we assessed the heritability enrichment in integrated genome annotations at the single tissue level[23] and tissue-specific epigenomic marks[24]. Lung tissue showed the most significant enrichment (enrichment=9.25, $P=1.36 \times 10^{-9}$), as previously described, though significant enrichment was also seen in heart (enrichment=6.85, $P=3.83 \times 10^{-8}$) and the gastrointestinal (GI) tract (enrichment=5.53, $P=6.45 \times 10^{-11}$). In an analysis of enriched epigenomic marks, the most significant enrichment was in fetal lung and GI smooth muscle DNase hypersensitivity sites (DHS) ($P= 6.75 \times 10^{-8}$) and H3K4me1 ($P= 7.31 \times 10^{-7}$) (**Supplementary Table 5**). To identify the source of association within lung tissue, we tested for heritability enrichment using single-cell chromatin accessibility[25] (ATAC-Seq) and gene expression (RNA-Seq) from human[26,27] and murine[28] lung (**Supplementary Table 5**). Using LD score regression in murine ATAC-Seq data, we found enrichment of chromatin accessibility in several cell types, including endothelial cells (most significant), type 1, and type 2 alveolar cells (the latter among the highest fold-enrichment [**Supplementary Table 5a**]). Results were varied using LD score regression[24] or SNPsea[29] on single-cell RNA-Seq, with nominal P-values for genes expressed in type 2 alveolar cells, basal-like cells, club cells, fibroblasts and smooth muscle cells (**Supplementary Tables 5b and c**).

Fine-mapping of associated loci

To identify the most likely causal variants at each locus, we performed fine mapping using Bayesian credible sets[30]. Including 160 potential primary and secondary association signals (excluding four variants in the major histocompatibility complex [MHC] region), 61 independent signals had a 99% credible set with fewer than 50 variants; 34 signals had credible sets with fewer than 20 variants (**Supplementary Figure 3**). Eighteen loci had a single variant with a posterior probability of driving association (PPA) greater than 60% including the *NPNT* (4q24) locus, where the association could be fine-mapped to a single intronic variant, rs34712979 (see **Supplementary Results** and **Supplementary Table 6**). Most sets included variants that overlapped genic enhancers of lung-related cell types (e.g., fetal lung fibroblasts, fetal lung, and adult lung fibroblasts) and were predicted to alter transcription binding motifs (**Supplementary Table 6**). Of 61 credible sets with fewer than 50 variants, eight sets contained at least one deleterious variant. These deleterious variants included 1) missense variants affecting *TNS1, RIN3, GPR126, ADAM19, ATP13A2, BTC*, and *CRLF3*; and 2) a splice donor variant affecting a lincRNA - AP003059.2.

Candidate target genes

In most cases, the closest gene to a lead SNP will not be the gene most likely to be the causal or effector gene of disease-associated variants[31–33]. Thus, to identify the potential effector ('target') genes underlying these genetic associations, we integrated additional molecular information including gene expression; gene regulation (open chromatin and methylation data), chromatin interaction, co-regulation of gene expression with gene sets and coding variant data (**Methods** and **Figure 3**).

At 82 loci, 472 genes within +/- 1 Mb of top associated variants were implicated by analysis of least one dataset; 106 genes were implicated by lung gene expression, and an additional 50 genes by >= 2 other datasets (methylation, chromatin interaction, open chromatin regions, similarity in gene sets or deleterious coding variants [**Figure 3**]), for a total of 156 genes meeting more stringent criteria. Excluding loci in the MHC region, the median number of potentially implicated genes per locus was four, with a maximum of 17 genes (7q22.1 and 17q21.1). The median distance of implicated genes to top associated variants was 346 Kb. Among 82 loci, 60 loci (73%) included the nearest gene. We identified 20 genes with supportive evidence from exome sequencing data. Two genes (*ADAM19* and *ADAMTSL3*) were implicated by five datasets (**Figure 3**) and another two (*EML4* and *RIN3*) were implicated by four datasets. A summary of all genes implicated using these approaches is included in **Supplementary Table 7**.

Associated pathways

To gain further functional insight of associated genetic loci, we performed gene-set enrichment analysis using DEPICT[34]. Among 165 enriched gene sets at FDR < 5%, 44% of them were related to the developmental process term, with nominal P for lung development of 1.02×10^{-6}; significant sub-terms included lung alveolus development (P= 0.0003) and lung morphogenesis (P= 0.0005). We also found enrichment of extracellular matrix -related pathways including laminin binding, integrin binding, mesenchyme development, cell-matrix adhesion, and actin filament bundles. Additional pathways of note included histone deacetylase binding, the Wnt receptor signaling pathway, SMAD binding, the MAPK cascade, and the transmembrane receptor protein serine/threonine kinase signaling pathway. Full enrichment analysis results including the top genes for each DEPICT gene set are shown in **Supplementary Table 8**.

Identification of drug targets

GWAS is also useful for identifying drug targets either at the individual gene[17,39,40] or genome-wide level[41,42]. Of 482 candidate target genes, 60 genes could be targeted by at least one approved or in-development drug, totaling 428 drugs with 144 different modes of action (**Supplementary Table 13**). Druggable targets at novel loci for COPD and lung function included *ABHD6, CDKL2, GSTO2, KCNC4*, *PDHB*, *SLK*, and *TRPM7*. We also identified drugs for repositioning in COPD using transcriptome-wide associations and drug-induced gene expression signatures (**Supplementary Results**).

Phenotypic effects of known and novel associations for COPD

To characterize the phenotypic effects of 82 genome-wide significant loci, we performed a phenome-wide association analysis within the deeply phenotyped COPDGene study (**Methods**). We assessed for common patterns of phenotype associations for the 82 loci by using hierarchical clustering across scaled Z scores of phenotype-variant associations. We identified two clusters of variants differentially associated with two sets of phenotypes (**Supplementary Figure 4**). As these two variant-phenotype clusters appeared to be driven by computed tomography (CT) imaging features, we repeated variant clustering limited to quantitative CT imaging features. We again found two clusters of variants, differentiated by association with quantitative emphysema, emphysema distribution, gas trapping, and airway phenotypes (**Figure 4a**). Additionally, we evaluated the association of the 82 genome-wide significant variants in a prior GWAS of emphysema and airway quantitative CT features[35] (**Supplementary Table 9**).

We also examined all genome-wide significant loci in the NHGRI-EBI GWAS Catalog[36] (**Supplementary Figure 5 and Supplementary Table 10**) and looked for trait-associated variants in linkage disequilibrium ($r^2 > 0.2$) with our lead COPD-associated variants. Many variants were associated with anthropometric measures including height and body mass index (BMI), measurements on blood cells (red and white cells), and cancers. COPD is well known to have many common comorbidities, such as coronary artery disease (CAD), type 2 diabetes mellitus (T2D), osteoporosis, and lung cancer. Of these diseases and 13 additional traits, we confirmed previously reported overall genetic correlation (using linkage disequilibrium score regression) of COPD with lung function, asthma, and height, and found evidence of modest correlation between COPD and lung cancer (**Supplementary Results**). However, at individual loci, and using more stringent linkage disequilibrium ($r^2 > 0.6$), we found evidence of shared risk factors for these comorbid diseases and COPD including a genome-wide significant variant near *PABPC4* associated with T2D, four variants with CAD (near *CFDP1, DMWD, STN1*, and *TNS1*), and a variant near *SPPL2C* with bone density (**Figure 4b**).

Identification of loci overlapping with asthma and pulmonary fibrosis

Based on our previous identification of genetic overlap of COPD with asthma, and COPD with pulmonary fibrosis, we examined loci for specific overlap with these two diseases. In asthma, we noted an $r^2 > 0.2$ with one of our variants and previously reported variants at *ID2, ZBTB38, C5orf56, MICA, AGER, HLA-DQB1, ITGB8, CLEC16A,* and *THRA.* In pulmonary fibrosis, in addition to our previously described overlap at *FAM13A*, *DSP*, and 17q21, we noted overlapping associations at *ZKSCAN1* and *STN1* (**Supplementary Table 11**). To more closely examine overlap, we applied a Bayesian method (gwas-pw) of COPD associations from our current GWAS with previous GWASs of asthma (limited to those of European ancestry) and pulmonary fibrosis[37,38]. To mitigate the results of including asthma among our COPD cases, we performed analysis for overlap with asthma removing self-reported asthmatics from UK Biobank for

this analysis (**Methods**). We identified 14 shared genome segments (posterior probability > 70%), 9 with asthma and 5 with pulmonary fibrosis (**Figure 4c** and **Supplementary Table 12**). In addition to the three segments shared with pulmonary fibrosis identified in the previous study[5] (*FAM13A*, *DSP,* and the 17q21 locus, here *CRHR1*), we identified two new segments including loci near *ZKSCAN1* and *STN1* (formerly known as *OBFC1)*. Shared variants between COPD and pulmonary fibrosis all had an opposite effect (i.e., increasing risk for COPD but protective for pulmonary fibrosis). In asthma, we identified five shared segments in the 6p21-22 regions, as well as *ADAM19, ARMC2, ELAVL2,* and *STAT6*. With the exception of *STAT6*, overlapping variants showed the same direction of effect.

Discussion

Genetic factors play an important role in COPD susceptibility. We examined genetic risk of COPD in a genome-wide association study of 35,735 cases and 222,076 controls. We identified 82 genome-wide significant loci for COPD, of which 47 were previously identified in genome-wide association studies of COPD or population-based lung function. Of 35 loci not previously described, 13 replicated in an independent study of population-based lung function. We used several data sources to attempt to assign causal genes at each locus, identifying 156 genes at 82 loci that were supported by either gene expression or a combination of at least 2 other data sources. Our results identify specific genes, cell types, and biologic pathways for targeted study and also suggest a genetic basis for the clinical heterogeneity seen in COPD.

Our study supports the role of early life events in the risk of COPD. Gene set enrichment analysis identified developmental pathways both specific to the lung (e.g., lung morphogenesis and lung alveolar development) and related to the lung (e.g., the canonical Wnt receptor[43,44], the MAPK/ERK, and the nerve growth factor receptor signaling pathways). We also confirmed enrichment of heritability in epigenomic marks of fetal lung. Our findings are consistent with epidemiologic studies demonstrating that a substantial portion of the risk of COPD may develop in early life: genetic variants may set initial lung function[45] and patterns of growth[45–47]. While further work will be needed to confirm the causal variants and genes affected by our variants, testing the role of these genes in lung development-relevant murine or ex-vivo models – for example, determining whether the perturbation of these genes changes proliferation and differentiation of lung epithelial progenitors in induced pluripotent cell-derived lung alveolar type 2 cells[44] – could provide experimental evidence of the role of these genes in early life susceptibility. Ultimately, the goal of this work would be to identify targets for or subsets of high risk individuals early in the disease course, or molecular candidates that may affect lung repair and regeneration[48].

Apart from genes related to lung development, our analyses highlighted several genes and pathways already of interest in COPD therapy (e.g. *CHRM3* / acetylcholine receptor inhibitors, the *MAPK* pathway) – supporting the role of genetic analyses in finding therapeutic targets[17,49] – and newer genes that could inform future functional studies. We identified interleukin 17 receptor D (*IL17RD*), as a potential effector gene at the at 3p14 locus. Numerous studies have examined the role of IL-17A in COPD[50], and IL17RD can differentially regulate pathways employed by IL-17A[51] . Chitinase acidic (*CHIA*) at 1p13.3, which encodes a protein that degrades chitin[52], exhibits lung-specific expression[53,54]. *CHIA* variants have been associated with FEV_1[55], asthma[56–59], and acid mammalian chitinase activity[58,60]. We identified several

potential effector genes related to extracellular matrix, cell adhesion, cell-cell interactions, and elastin-associated microfibrils[61–63], some of which have been previously identified in studies of lung function[14]. These include integrin family members that mediate cell-matrix communication (e.g., *ITGA1, ITGA2*, *ITGA8*[64–66]), an integrin ligand encoding gene (*NPNT*[67]), and genes encoding matrix proteins (e.g., *MFAP2* and *ADAMTSL3*). *ADAMTSL3* plays a role in cell-matrix interactions related to the assembly of fibrillin and microfibril biogenesis[68–70] and of our candidate effector genes was supported by the greatest number of bioinformatic analyses. Recombinant forms of other *ADAMTS*-like proteins demonstrate experimental evidence of promoting and enhancing fibrillin and microfibril deposition and assembly[71,72]. *ADAMTSL3* may play a role in preventing emphysematous destruction of lung tissue by ADAMTS in COPD.

In addition to identifying the effector gene, knowing the effector cell type is critical for functional studies. We identified an overall enrichment of epigenomic marks in lung tissue and smooth muscle (also identified in studies of lung function[15]). This latter association was found in gastrointestinal tissue cell types; respiratory smooth muscle is absent in the analyzed datasets. We also performed analyses of single-cell data in an attempt to identify the specific lung cell types in which our top variants are potentially functioning. We found evidence for enrichment of several cell types, including but not limited to endothelial cells, alveolar type 2 cells, and basal-like cells. Each of these cell types has been postulated to have a role in the development of COPD[73–75], and our data is consistent with the likely heterogeneity of lung cell types contributing to COPD susceptibility. The lung is comprised at least 40 different resident cell types[76], most of which are not distinctly represented in these datasets. Thus, while our findings support the investigation of specific cell types for further functional studies, they also highlight the need for profiling of lung-relevant cell types and loci-specific analyses.

Characterization of functional variant effects could lead to better disease subtyping and more targeted therapy for COPD. Cluster analysis on hundreds of COPD-associated features in the more extensively phenotyped COPDGene cohort revealed heterogeneous effects of genetic variants on COPD-related phenotypes, including computed tomography (CT) measurements of airway abnormalities and emphysema – well-described sources of heterogeneity in COPD[77–79]. Analyzing hundreds of diseases/traits in GWAS Catalog, we identified overlapping associations with various diseases/traits in multiple organ systems, comorbidities such as coronary artery disease, bone mineral density, and type 2 diabetes mellitus (T2D). The COPD-associated *PABPC4* locus was associated with T2D[80] and C-reactive protein (CRP) level[81]. Although a causal gene in this locus and its contribution to COPD is unknown, its association with T2D may suggest a shared disease pathway and drug targets. Together, the identification of variable COPD risk loci associations with sub-phenotypes and other diseases[83,84] may have potential for more nuanced approaches to therapy for COPD. Overall, our phenotype, gene, and pathway analyses illustrate the utility of both searching for enrichment of genetic signals overall, and performing a more detailed identification of the effects of individual variants or groups of variants.

We performed additional specific analysis in two diseases that overlap with COPD, asthma and pulmonary fibrosis. While a genome-wide genetic correlation of COPD and asthma has been previously described[5], our analysis is the first to identify specific shared genetic segments between asthma and COPD. While the effects at most of these shared segments were concordant in direction, one of the segments of particular interest was near *STAT6*, which had opposite directions of effect in the two diseases. *STAT6* plays a role in T helper (Th) type 2-dependent inflammation, and is activated by interleukin-4 and interleukin-13 (IL-4 and IL-13)[85]. IL-13, in turn, has been found to be increased in

asthmatic airways[86] but decreased in severe emphysema[87]. In pulmonary fibrosis, variants at all overlapping loci have an opposite direction of effect compared to COPD[5]. These effects raise the possibility that specific therapies for one disease could increase the risk of the other disease, which may be worth evaluating in treatment trials. The reasons why genetic effects are divergent between COPD and fibrosis are unclear, but these identified opposite effects could point to molecular switches that influence why some smokers develop emphysema while others develop pulmonary fibrosis. While pulmonary fibrosis is an uncommon disease and specifically excluded in several of our COPD case-control cohorts, interstitial lung abnormalities are increasingly being recognized as a potential precursor to fibrosis and an inverse relationship between these abnormalities and emphysema has been previously identified[88]. Mechanistically, some have hypothesized that the divergent derangement of Wnt and Notch signaling pathways[89] and mesenchymal cell fate[90] may be responsible for the distinct development of these two diseases. We also describe an overlapping region at the *STN1* (previously known as *OBFC1*) locus. *STN1* plays a role in telomere maintenance[91]; shortened telomeres have been observed in both COPD and IPF[92,93], and rare genetic variants in the telomerase pathway have been implicated in both pulmonary fibrosis and emphysema – albeit with concordant effects on either disease[94].

While our study is the largest genome-wide association study of COPD, individuals meeting our criteria for COPD in the UK Biobank may be different from other studies, especially for smoking history. We used the same definition of COPD as in our prior analysis[5], which included non-smokers. Our use of pre-bronchodilator spirometry to define COPD (allowing us to maximize sample size) as well as population-based lung function for replication could bias our findings against variants that are only associated with more severe forms of COPD. We did not exclude other causes of airway obstruction such as asthma, noting that asthma frequently overlaps with and is misdiagnosed in COPD[95]. We performed several additional analyses to determine whether our results were driven by, or markedly different, by smoking status, asthma, or use of pre- instead of post-bronchodilator spirometry to define COPD. The results of these additional analyses did not indicate a substantial impact of these factors on our overall findings, and together with prior analyses[5,15], suggest that bias due to these factors is likely small. However, our study was not designed to identify differences between subgroups, and we cannot rule out a role for studying more severe disease or disease subtypes. We note that the alpha-1 antitrypsin locus (*SERPINA1*) was identified as genome-wide significant in smaller studies of emphysema and in smokers with severe COPD[96]. In the current study, the association of the PiZ allele had $P = 2.2 \times 10^{-5}$ using moderate-to-severe cases (FEV_1 < 80% predicted), and a smaller P-value (1.4×10^{-6}) in severe cases (FEV_1 < 50% predicted) despite a smaller sample size, a phenomenom we have previously described[10]. Thus, despite the strong overlap of COPD with quantitative spirometry, new loci may be identified through studies of sufficiently large subsets of COPD patients and with more specific and homogeneous COPD phenotypes. Given suggestive evidence for replication using a related (but not identical) phenotype for additional novel loci beyond the 13 meeting a Bonferroni-corrected threshold for significance, we chose to include all loci significant in discovery in subsequent analyses, recognizing that we likely included some false positive associations. Our study focused on relatively common variants, predominantly in individuals of European ancestry; more detailed studies of rare variants, the human leukocyte antigen (HLA) regions, and other ethnicities are warranted, but broader multi-ethnic analyses are limited by the number of cases in currently available cohorts. Although COPD sex differences have been reported[97], we did not identify significant sex-specific differences in effect sizes of the 82 top variants. Future studies including more subjects and methodological advances may be needed to elucidate this effect.

The global burden of COPD is increasing. Our work finds a substantial number of new loci for COPD and uses multiple lines of supportive evidence to identify potential genes and pathways for both existing and novel loci. Further investigation of the genetic overlap of COPD with other respiratory diseases and the phenotypic effects of top loci finds new shared loci for asthma and idiopathic pulmonary fibrosis and suggests heterogeneity across COPD-associated loci. Together, these insights provide multiple new avenues for investigation of the underlying biology and the potential therapeutics in this deadly disease.

Methods

Study populations

The UK Biobank is a population-based cohort consisting of 502,682 individuals[8]. To determine lung function, we used measures of forced expiratory volume in 1 second (FEV_1) and forced vital capacity (FVC) derived from the spirometry blow volume-time series data, subjected to additional quality control based on ATS/ERS criteria[98] (**Supplementary Methods**). As in our previous study[5], we defined COPD using pre-bronchodilator spirometry according to modified Global Initiative for Chronic Obstructive Lung Disease (GOLD) criteria for moderate to very severe airflow limitation[9]: FEV_1 less than 80% of predicted value (using reference equations from Hankinson et al.[99]), and the ratio between FEV_1 and FVC less than 0.7. Consistent with our previous analyses and enrollment criteria for COPD case-control datasets[100], we did not exclude individuals based on self-reported asthma. Genotyping was performed using Axiom UK BiLEVE array and Axiom Biobank array (Affymetrix, Santa Clara, California, USA) and imputed to the Haplotype Reference Consortium (HRC) version 1.1 panel[101].

We invited participants in the prior International COPD Genetics Consortium (ICGC) COPD genome-wide association study to provide case-control association results (with the exception of the 1958 British Birth Cohort, to avoid overlapping samples with the replication sample). ICGC cohorts performed case-control association analysis based on pre-bronchodilator measurements of FEV_1 and FEV_1/FVC, and cases were identified using modified GOLD criteria, as above. Studies were imputed to 1000 Genomes reference panels. Detailed cohort descriptions and cohort-specific methods have been previously published[5] (**Supplementary Methods**).

Based on the strong genetic overlap of lung function and COPD[5], we performed lookups of select significant variants for FEV_1 and FEV_1/FVC in the SpiroMeta consortium meta-analysis[20]. Briefly, SpiroMeta comprised of a total of 79,055 individuals from 22 studies imputed to either the 1000 Genomes Project Phase 1 reference panel (13 studies) or the HRC (9 studies). Each study performed linear regression adjusting for age, age^2, sex, and height, using rank-based inverse normal transforms, adjusting for population substructure using principal components or linear mixed models, and performing separate analyses for ever- and never- smokers or using a covariate for smoking (for studies of related subjects). Genomic control was applied to individual studies, and results were combined using a fixed-effects meta-analysis[20].

Genome-wide association analysis

In UK Biobank, we performed logistic regression of COPD, adjusting for age, sex, genotyping array, smoking pack-years, ever smoking status, and principal components of genetic ancestry. Association analysis was done using PLINK 2.0 alpha[102] (downloaded on December 11, 2017) with Firth-fallback settings, using Firth regression when quasi-complete separation or regular-logistic-regression convergence failure occurred. We performed a fixed-effects meta-analysis of all ICGC cohorts and UK

Biobank using METAL (version 2010-08-01)[103]. We assessed population substructure and cryptic relatedness by linkage disequilibrium (LD) score regression intercept[104]. We defined a genetic locus using a 2-Mb window (+/-1 Mb) around a lead variant, with conditional analyses as described below.

To maximize our power to identify existing and discover new loci, we examined all loci at the genome-wide significance value of $P < 5 \times 10^{-8}$. We first characterized loci as being previously described (evidence of prior association with lung function[11–19,105,106] or COPD[5,10,107]) or novel. We defined previously reported signals if they were in the same LD block in Europeans[108] and in at least moderate LD ($r^2 >= 0.2$). For novel loci we attempted replication through association of each lead variant with either FEV_1 or FEV_1/FVC ratio in SpiroMeta, using one-sided p values with Bonferroni correction for the number of novel loci examined. Novel loci failing to meet a Bonferroni-corrected P value were assessed for nominal significance (one-sided $P < 0.05$) or directional consistence with FEV_1 and FEV1/FVC ratio in SpiroMeta.

Cigarette smoking is the major environmental risk factor for COPD and genetic loci associated with cigarette smoking have been reported[5,109]. While we adjusted for cigarette smoking in our analysis, we further examined these effects by additionally testing for association of each locus with cigarette smoking and by looking at two separate analyses of ever- and never- smokers in UK Biobank. We tested for sex-specific genetic effects of genome-wide significant variants via a stratified analysis and interaction testing, using a 5% Bonferroni-corrected threshold to determine significance (**Supplementary Methods**).

Identification of independent associations at genome-wide significant loci

We identified specific independent associations at genome-wide significant loci using GCTA-COJO[110]. This method utilizes an approximate conditional and joint analysis approach requiring summary statistics and representative LD information. As the UK Biobank provided the predominant sample, we used 10,000 randomly drawn unrelated individuals from this discovery dataset as a LD reference sample. We scaled genome-wide significance to a 2-Mb region, resulting in a locus-wide significant threshold of 8×10^{-5}, or 2×10^{-6} for variants in the major histocompatibility complex (MHC) region (chr6:28477797-33448354 in hg19[111]). We created regional association plots via LocusZoom using 1000 Genomes EUR reference data (Nov2014 release)[112].

Identification and prioritization of tissues and cell types, candidate variants, genes, and pathways

Identification of enriched tissues and specific cell types

We used LD Score Regression (LDSC) to estimate the enrichment of functional annotations[113] and specifically expressed gene regions[24] on disease heritability. We utilized LDSC baseline models (e.g., conserved region, promoter flanking region), tissue-specific annotations from the Roadmap Epigenomics Program[24], integrated tissue annotations from GenoSkyline[23], and cell type-specific chromatin accessibility data[25] (ATAC-Seq). We used four single-cell gene expression (RNA-Seq) datasets to identify specific cell types (**Supplementary Methods**), including 1) lung epithelial cells from normal and pulmonary fibrosis human lung[26] (Gene Expression Omnibus [GEO] accession GSE86618), 2) human induced pluripotent stem cells (iPSCs)-derived putative alveolar type 2 cells[27] (GSE96642), 3) mouse lungs at embryonic day 18.5 (E18.5) and 4) postnatal day 1 (P1) by Whitsett et al. (available at LungMAP[28]). We also used SNPsea[29] to identify enriched cell types in genome-wide significant loci (**Supplementary Methods**). We reported only estimates of coefficients and P values for the Roadmap

annotations and gene expression datasets, as theses analyses used –h2-cts, which does not report fold enrichment.

Fine-mapping of independent association signals at genome-wide significant loci

We used Bayesian fine-mapping at each locus to identify the credible set: the set of variants with a 99% probability of containing a causal variant. Briefly, for each genome-wide significant loci we calculated approximate Bayes factors[30] of association. We then selected variants in each locus, so that their cumulative posterior probability was equal or greater than 0.99 using an unscaled variance. At loci with multiple independent associations, we used statistics from approximate conditional analysis with GCTA software on each index variant adjusting for other independent variants in the loci. Otherwise, we used unconditioned statistics from our meta-analysis. We characterized variant effects in credible sets using variant annotations from Ensembl Variant Effect Predictor[114].

Identification of target genes

We used several computational approaches with corresponding available datasets to identify target genes in genome-wide significant loci. We used two methods that utilized gene expression data: 1) S-PrediXcan and 2) DEPICT. We used S-PrediXcan[115] to identify genes with genetically regulated expression associated with COPD. We used data from the Lung-eQTL consortium[116,117] (1,038 lung tissue samples) as an eQTL and gene expression reference database. S-PrediXcan is the extension of PrediXcan[118] that test for association between a trait and imputed gene expression using summary statistics. Here, we performed S-PrediXcan using models for protein-coding genes +/- 1 Mb from top-associated variants at genome-wide significant loci. We used DEPICT (Data-driven Expression Prioritized Integration for Complex Traits)[34] to prioritize genes from 'reconstituted' gene sets.

We also used additional information on gene regulation, including epigenetic data: 1) regulatory fine mapping, 2) mQTL, and 3) chromosome conformation capture. We used regulatory fine mapping (regfm[119]) to overlap 99% credible interval (CI) variants at each GWAS locus with open chromatin regions based on DNAse hypersensitivity sites (DHS). DHS cluster accessibility state was then associated with gene expression levels (for 13,771 genes) from 22 tissues in the Roadmap Epigenomics Project[120]. Using both the 99% CI and DHS overlap, as well as the DHS state and transcript level association, regfm calculates a posterior probability of association of each gene +/- 1 Mb of the lead SNP at each GWAS locus. We also searched for overlapping methylation quantitative trait loci (mQTL) data from lung tissue, as recently described[121]. To determine whether these signals co-localized (rather than being related due to linkage disequilibrium), we performed colocalization analysis between our GWAS and mQTL in genome-wide significant loci using eCAVIAR[122] (eQTL and GWAS CAusal Variants Identification in Associated Regions, **Supplementary Methods**). We also sought information from publicly available chromosome conformation capture data[123]. We queried association statistics of chromatin contact (i.e., long range chromatin interactions) between top associated variants and gene promoters nearby in a lung (fetal lung fibroblast cell line (IMR90) and human lung tissue[123]) using HUGIn[124] (Hi-C Unifying Genomic Interrogator). We retained only the strongest associations (i.e., smallest P value) for each cell line/primary cell in the analysis.

Finally, we searched for signals from deleterious variants by querying consequences of variants within 99% credible sets containing fewer than 50 variants (**Supplementary Methods**). We also searched for rare coding variants, based on exome sequencing results in the COPDGene, Boston Early-Onset COPD, and International COPD Genetics Network studies, as previously described[125]. In brief, we performed

exome sequencing on 485 severe COPD cases and 504 smoking resistant controls from the COPDGene study and 1,554 subjects ascertained through 631 probands with severe COPD from the Boston Early-Onset COPD study (BEOCOPD) and the ICGN study. We performed single-variant analyses using Firth and efficient resampling methods (SKAT R package[126]) for the COPDGene data (case-control) and generalized linear mixed models (GMMAT) for the BEOCOPD-ICGN data (using lung function). Gene-based analyses were conducted using burden, SKAT, and SKAT-O tests with asymptotic and efficient resampling methods (SKAT package) combined with Fisher's method for the COPDGene data, and using SKAT-O tests (MONSTER) for the BEOCOPD-ICGN data. Two variant-filtering criteria were considered: deleterious variants (predicted by FATHMM) with minor allele frequency (MAF) < 0.01, and functional variants (moderate effect predicted by SNPEff) with MAF < 0.05. We also applied a gene-based segregation test (GESE) to the ultra-rare (MAF < 0.1%) and loss-of-function variants in the BEOCOPD-ICGN data on the severe COPD affection status. In gene-based analyses, we combined results from all methods above and retained only most significant P values for each gene.

For each dataset described above, we used Bonferroni-corrected P values, or a fixed posterior probability threshold to determine target genes at each locus. We reported protein-coding genes +/-1 Mb from a top associated variant. We restricted our search to genes from the GRCh37 server in biomaRt[114] with updated HUGO Gene Nomenclature Committee (HGNC) names (downloaded from https://genenames.org on June 7, 2018). For each locus, we used a 5% Bonferroni-corrected threshold (i.e., P < 0.05 divided by number of genes at that locus) to determine significance for 4 data types: gene expression data, chromatin conformation capture data, co-regulation of gene expression, and exome sequencing results. For two remaining datasets, we used a fixed posterior probability (of gene association with a GWAS locus) threshold of 0.1 for regfm and eCAVIAR. We considered genes that were implicated by gene expression or >= 2 combination of other datasets (e.g., methylation and chromatin conformation capture data) as target genes.

Identification of pathways

To identify enriched pathways in COPD-associated loci, we performed gene-set enrichment analysis using the "reconstituted" genes sets from DEPICT, as described above[34]. We defined significant gene sets using false discovery rate (FDR) < 5%.

Effects on COPD-related and other phenotypes

COPD is a complex and heterogeneous disorder, comprised of different biologic processes and specific phenotypic effects. In addition, many loci discovered by GWAS have pleiotropic effects. To identify these effects, we performed analyses of a) identification of overlapping genetic loci between related disorders (asthma and pulmonary fibrosis) b) genetic association studies of our genome-wide significant findings using COPD-related phenotypes, including a cluster analysis to identify groups of variants that may be acting via similar mechanisms; c) look up of top variants in prior COPD-related quantitative computed tomography (CT) imaging feature GWAS, d) look up of associations with other diseases/traits using GWAS Catalog, and e) estimate the genetic correlation between COPD and other diseases/traits.

To identify overlapping loci between COPD and other respiratory disorders, we used gwas-pw[127] to perform pairwise analysis of GWAS. This method searches for shared genomic segments[108] using adaptive significance threshold, allowing detection of sub genome-wide significant loci. We identified shared segments or variants using posterior probability of colocalizing greater than 0.7[127]. We obtained GWAS summary statistics from two previous studies of pulmonary fibrosis[38] and asthma in Europeans[37].

For the overlap analysis of COPD with asthma, we examined the influence of the inclusion of individuals with self-reported asthma on both the overlap of discrete GWAS loci (using gwas-pw) and genome-wide genetic correlation (using LD score regression) by performing these analyses in the meta-analysis of ICGC studies and the UK Biobank (with individuals with asthma removed from cases in the latter). To assess heterogeneous effects of COPD susceptibility loci on COPD-related features (phenotypes), we evaluated associations of our genome-wide significant SNPs with 121 detailed phenotypes (e.g., lung function, computed tomography-derived metrics, biomarkers, and comorbidities) available in 6,760 COPDGene non-Hispanic whites. We calculated Z-scores for each SNP-phenotype combination relative to the COPD risk allele to create a SNP by phenotype Z-score matrix. We tested each COPD-related phenotype with at least one nominally significant association with one of our genome-wide significant COPD SNPs, leaving us with 107 phenotypes. We then oriented all Z-scores to be positive (based on sign of median Z score) in association with each phenotype to avoid clustering based on direction of association. To avoid clustering phenotypes only by strength of association with SNPs, we scaled Z-scores within each phenotype by subtracting mean Z-scores and dividing by the standard deviation of Z-scores within each phenotype. We then scaled Z-scores across SNPs to circumvent clustering of SNPs according only to relative strength of association with phenotypes. We then performed hierarchical clustering of the scaled Z-scores of associations between SNPs and phenotypes to identify clusters of SNPs and phenotypes for all 107 phenotypes as well as in the subset of 26 quantitative imaging phenotypes. We identified optimal number of clusters using the Calinski index[128,129]. To identify features that independently predict cluster membership, we fitted a logistic regression model via penalized maximum likelihood using the glmnet package[130]. We determined optimal regularization parameters using 10-fold cross validation. We further examined top variant associations with COPD-related traits through a look-up of top variants in a prior GWAS of 12,031 subjects with quantitative emphysema and airway CT features[35]. To examine overlap of our COPD results with other traits, we downloaded genome-wide significant associations from the GWAS Catalog[36,131] ($P < 5 \times 10^{-8}$). Between a pair of COPD- and trait- associated variants within the same LD block in Europeans[108], we computed the LD using the European ancestry panel[132] and considered the overlap if variants were in at least in moderate LD ($r^2 >= 0.2$). We estimated genetic correlation between COPD and other diseases/traits using a web engine for LDSC, LD Hub[133]. We assessed the results using a 5% Bonferroni-corrected significance level.

Identification of drug targets

We queried our target genes using the Drug Repurposing Hub[134], available at https://clue.io. This resource contains comprehensive annotations of launched drugs, drugs in phases 1-3 of clinical development, previously approved and preclinical or tool compounds, curated using publicly available sources (e.g., ChEMBL and Drugbank) and proprietary sources. We performed drug-gene expression similarity analysis[135] (the Query, https://clue.io) using a ranked gene set from a gene-based association test[115] (**Supplementary Methods**).

Figures

Figure 1 Study design

COPD, chronic obstructive pulmonary disease; FEV_1, force expiratory volume in one second; FVC, forced vital capacity.

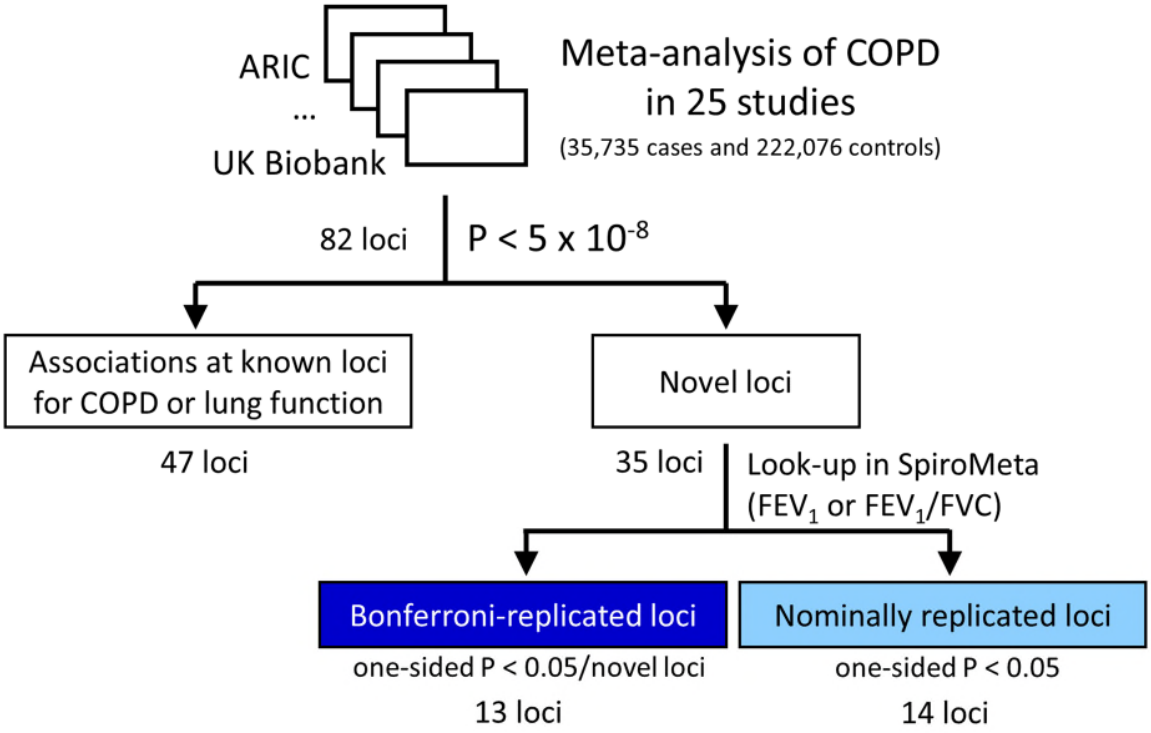


Figure 2 Manhattan plot

Loci are labeled with the closest gene to the lead variant. Colors indicates variants at novel loci which replicated using a Bonferroni-corrected threshold in SpiroMeta (dark blue, one sided P < 0.05/35) and a nominally significant threshold (light blue, one-sided P < 0.05).

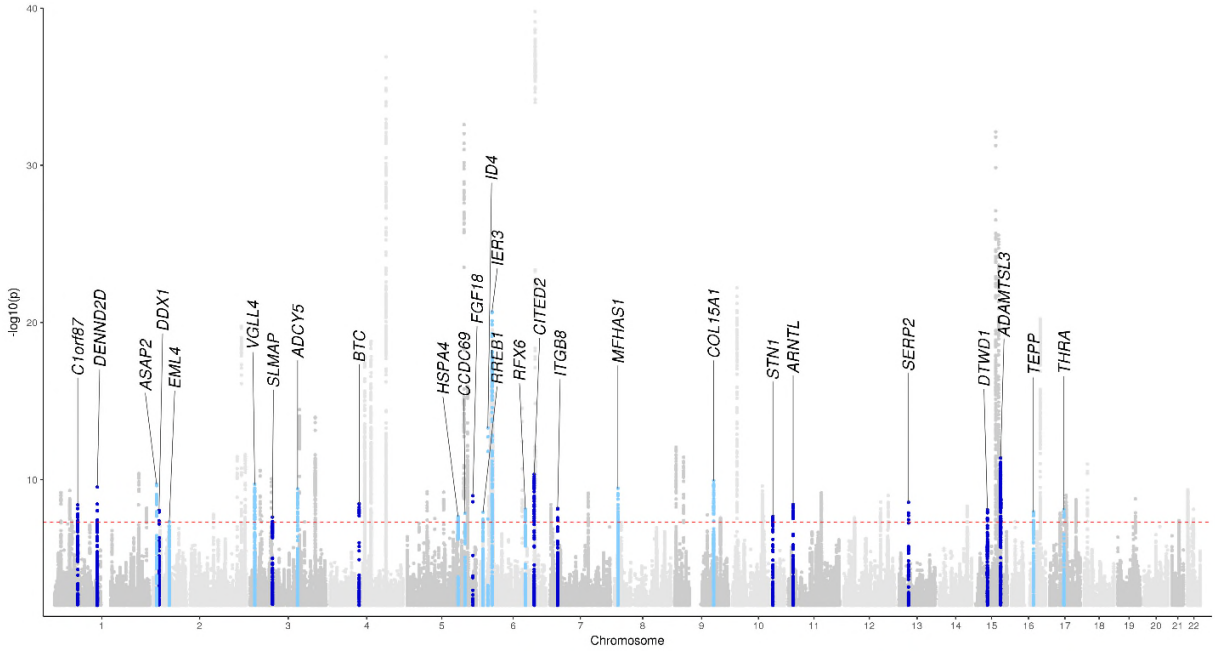
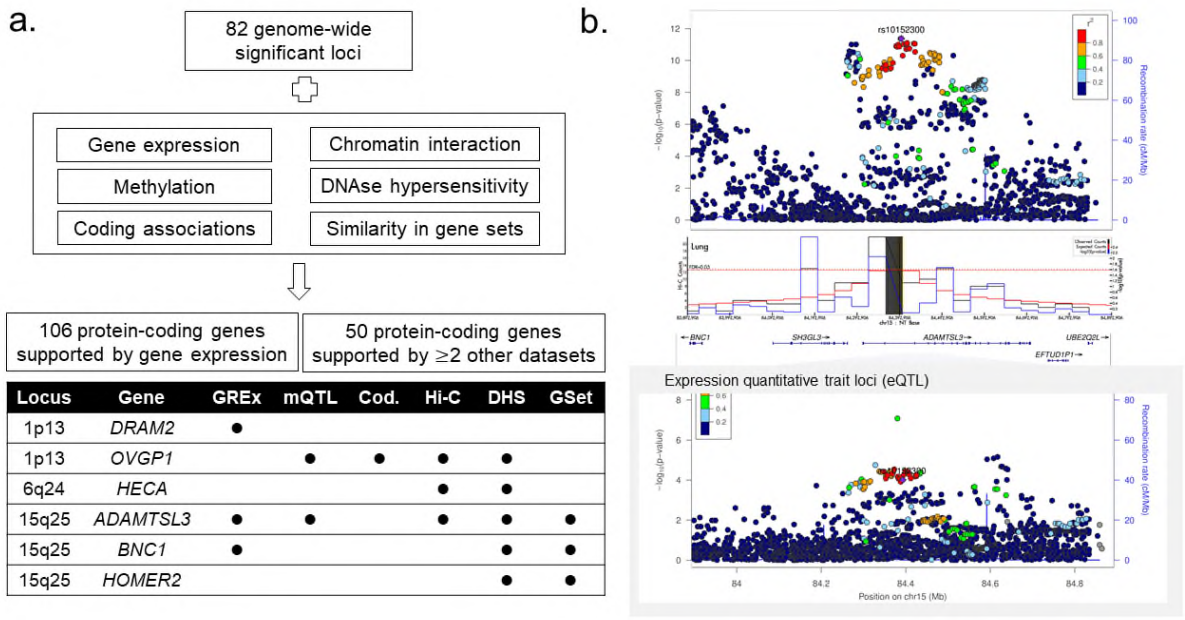


Figure 3 Identification of target genes

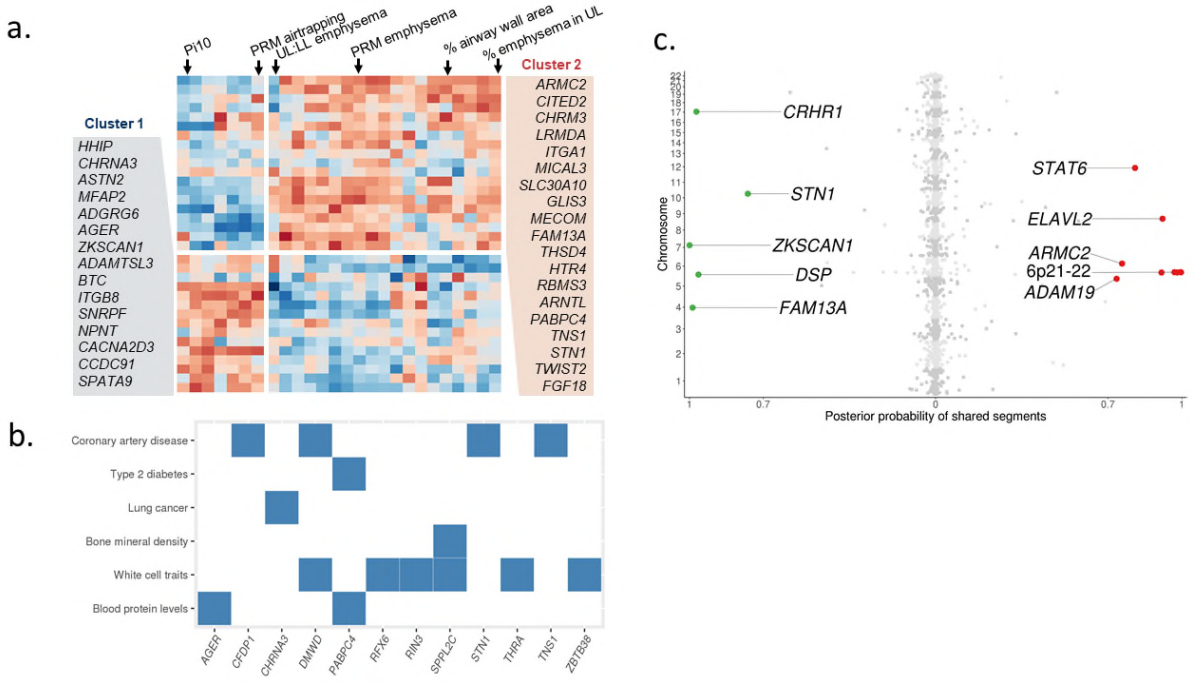
(a) Overview of datasets used to identify target genes at genome-wide significant loci (b) Regional association plots at *ADAMTSL3* locus showing GWAS (top), chromatin interaction in lung tissue (middle) and expression quantitative trait loci (bottom). GREx, Genetically regulated gene expresssion (significant associations identified by S-PrediXcan); mQTL, methylation quantitative trait loci (colocalized signals between GWAS and mQTL at posterior probability > 0.1); Cod., Coding associations (significant single variant or gene-based association tests for deleterious coding variants); Hi-C, significant chromatin interaction identified in human lung or the IMR90 cell line; DHS, DNase hypersensititvity sites (using regulatory fine-mapping or regfm, Methods); GSet, target genes identified by DEPICT using reconstituted gene sets.



Locus	Gene	GREx	mQTL	Cod.	Hi-C	DHS	GSet
1p13	*DRAM2*	●					
1p13	*OVGP1*		●	●	●	●	
6q24	*HECA*				●	●	
15q25	*ADAMTSL3*	●	●		●	●	●
15q25	*BNC1*	●				●	●
15q25	*HOMER2*					●	●

Figure 4 Effects on COPD-related and other phenotypes

(a) Heatmap of scaled computed tomography (CT) quantitative imaging associations with the 34 genome-wide significant variants (known and replicated novel associations) with at least nominal (P<0.05) association with any CT imaging feature in COPDGene non-Hispanic white participants. Red coloring suggests more positive association with a phenotype and blue suggests less positive association with a phenotype. Cluster 1 variants are more associated with airway imaging features and Cluster 2 variants are more associated with emphysema imaging features. Variants are referred to by the closest gene. (b) Overlapping of genome-wide significant loci of COPD and select traits from GWAS Catalog (c) Genome-wide overlapping results between COPD with pulmonary fibrosis (left) and asthma (right). PRM emphysema, emphysema quantified by parametric response mapping; UL, upper lobe of the lung; LL, lower lobe of the lung; Pi10, an airway feature calculated from regressing the square root of the airway wall area with the airway internal perimeter.



Tables

Table 1 Meta-analysis results showing 35 loci novel for COPD and lung function

(see the Excel file)

Acknowledgements

Please refer to the **Supplementary Note** for full acknowledgements.

Author contributions

P.S. contributed to the study concept and design, data analysis, and manuscript writing. D.P., B.D.H., M.H.C. contributed to the study concept and design, data analysis, statistical support, and manuscript writing. A.B.W., K.d.J., S.J.L., D.P.S. contributed to the study concept and design and data analysis. P.B., R.G.B., J.D.C., A.G., D.A.M., G.T.O., S.I.R., D.A.S., R.T.-S., Y.T., E.K.S. contributed to the study concept and design and data collection. T.H.B., J.E.H. contributed to the study concept and design and to statistical support. I.P.H., H.M.B., L.V.W., M.D.T. contributed to the study concept and design. 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.

Competing financial interests

M.H.C., E.K.S., L.V.W., M.D.T., and I.P.H. have received grant funding from GSK. E.K.S. has received honoraria from Novartis for Continuing Medical Education Seminars and travel support from GlaxoSmithKline. I.P.H. has received grant support from BI. R.T.-S. is an employee of GSK. D.A.S. has financial support from Eleven P15. J.V. has received personal fees from GSK, Chiesi Pharmaceuticals, BI, Novartis, and AstraZeneca.

Supplementary Figures

Supplementary Figure 1 Forest plots for 82 genome-wide significant associations

Supplementary Figure 2 Regional association plots for 82 genome-wide significant associations

Supplementary Figure 3 Distribution of number of variants in 99% credible sets

Supplementary Figure 4 Heatmap of associations of 60 index variants and phenotypes in COPDGene

Supplementary Figure 5 Associations of index variants and traits in NHGRI-EBI GWAS Catalog

Supplementary Figure 6 Power analysis for sex-difference analysis

Supplementary Figure 7 Scatter plot of COPD odd ratio of nominally significant SNPs in meta-analysis of a subset of COPD case-control cohorts* using a pre- and post-bronchodilator definition of COPD

Supplementary Figure 8 Comparison of odds ratios (OR) including and excluding individuals with asthma of 82 genome-wide significant variants

Supplementary Tables

See the Excel file.

Supplementary Table 1 Cohort baseline characteristics in COPD cases and controls

Supplementary Table 2 Meta-analysis results showing 47 previously reported loci for COPD or lung function

Supplementary Table 3 Cohort baseline characteristics in replication studies from SpiroMeta consortium

Supplementary Table 4 Multiple independent associations within the same 2-Mb window identified using approximate conditional and joint analysis

Supplementary Table 5 Heritability enrichment in cell-type specific epigenomic mark from Roadmap Epigenomic Project

Supplementary Table 6 Functional annotation of variants with posterior probability of association greater than 0.6

Supplementary Table 7 Candidate target genes

Supplementary Table 8 Gene sets significantly enriched at FDR < 0.05 using DEPICT

Supplementary Table 9 Lookup associations for quantitative computed tomography (QCT) features

Supplementary Table 10 Association results from GWAS catalog

Supplementary Table 11 GWAS catalog for idiopathic pulmonary fibrosis and asthma

Supplementary Table 12 Overlapping loci between COPD with asthma and pulmonary fibrosis

Supplementary Table 13 Drug for candidate target genes

Supplementary Table 14 Meta-analysis results showing 35 loci novel for COPD and lung function and its stratified analysis in ever-smokers and never-smokers

Supplementary Table 15 Meta-analysis results showing 47 loci previously reported for COPD or lung function and its stratified analysis in ever-smokers and never-smokers

Supplementary Table 16 Sex-specific association statistics for 82 genome-wide significant variants

Supplementary Table 17 Numbers of individuals diagnosed with COPD based on pre- and post-bronchodilator spirometry with and without asthma cases

Supplementary Table 18 Association statistics for all 82 primary and secondary associations based on pre- and post-bronchodilator spirometry in COPDGene

Supplementary Table 19 Meta-analysis results showing 35 loci novel for COPD and lung function and additional analysis excluding asthma

Supplementary Table 20 Meta-analysis results showing 47 loci previously reported for COPD or lung function and additional analysis excluding asthma

Supplementary Table 21 Results of the analysis including and excluding individuals with asthma in UK Biobank

Supplementary Table 22 Genetic correlation between COPD and other traits/diseases

References

1. GBD 2015 Chronic Respiratory Disease Collaborators. Global, regional, and national deaths, prevalence, disability-adjusted life years, and years lived with disability for chronic obstructive pulmonary disease and asthma, 1990-2015: a systematic analysis for the Global Burden of Disease Study 2015. *Lancet. Respir. Med.* **5,** 691–706 (2017).
2. *Global Health Estimates 2016: Deaths by Cause, Age, Sex, by Country and by Region, 2000-2016*. (2018).
3. Fuchsberger, C. *et al.* The genetic architecture of type 2 diabetes. *Nature* **536,** 41–47 (2016).
4. Zhou, J. J. *et al.* Heritability of chronic obstructive pulmonary disease and related phenotypes in smokers. *Am. J. Respir. Crit. Care Med.* **188,** 941–7 (2013).
5. 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).
6. Jiang, Z. *et al.* A Chronic Obstructive Pulmonary Disease Susceptibility Gene, FAM13A, Regulates Protein Stability of beta-Catenin. *Am J Respir Crit Care Med* **194,** 185–197 (2016).
7. Lao, T. *et al.* Hhip haploinsufficiency sensitizes mice to age-related emphysema. *Proc. Natl. Acad. Sci. U. S. A.* **113,** E4681-7 (2016).
8. Sudlow, C. *et al.* UK biobank: an open access resource for identifying the causes of a wide range of complex diseases of middle and old age. *PLoS Med.* **12,** e1001779 (2015).

9. Vogelmeier, C. F. *et al.* Global Strategy for the Diagnosis, Management, and Prevention of Chronic Obstructive Lung Disease 2017 Report. GOLD Executive Summary. *Am. J. Respir. Crit. Care Med.* **195,** 557–582 (2017).

10. Cho, M. H. *et al.* Risk loci for chronic obstructive pulmonary disease: a genome-wide association study and meta-analysis. *Lancet Respir Med* **2,** 214–225 (2014).

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. Repapi, E. *et al.* Genome-wide association study identifies five loci associated with lung function. *Nat Genet* **42,** 36–44 (2010).

13. 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).

14. Soler Artigas, M. *et al.* Genome-wide association and large-scale follow up identifies 16 new loci influencing lung function. *Nat Genet* **43,** 1082–1090 (2011).

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,** 769–781 (2015).

16. Soler Artigas, M. *et al.* Sixteen new lung function signals identified through 1000 Genomes Project reference panel imputation. *Nat Commun* **6,** 8658 (2015).

17. 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).

18. Wyss, A. B. *et al.* Multiethnic meta-analysis identifies ancestry-specific and cross-ancestry loci for pulmonary function. *Nat. Commun.* **9,** 2976 (2018).

19. Jackson, V. E. *et al.* Meta-analysis of exome array data identifies six novel genetic loci for lung function. *Wellcome open Res.* **3,** 4 (2018).

20. Shrine, N. *et al.* New genetic signals for lung function highlight pathways and pleiotropy, and chronic obstructive pulmonary disease associations across multiple ancestries. *bioRxiv* (2018).

21. Agusti, A. & Soriano, J. B. COPD as a systemic disease. *COPD* **5,** 133–8 (2008).

22. Barnes, P. J. & Celli, B. R. Systemic manifestations and comorbidities of COPD. *Eur. Respir. J.* **33,** 1165–85 (2009).

23. Lu, Q. *et al.* Systematic tissue-specific functional annotation of the human genome highlights immune-related DNA elements for late-onset Alzheimer's disease. *PLoS Genet.* **13,** e1006933 (2017).

24. Finucane, H. K. *et al.* Heritability enrichment of specifically expressed genes identifies disease-relevant tissues and cell types. *Nat. Genet.* **50,** 621–629 (2018).

25. Cusanovich, D. A. *et al.* A Single-Cell Atlas of In Vivo Mammalian Chromatin Accessibility. *Cell* **174,** 1309–1324.e18 (2018).

26. Xu, Y. *et al.* Single-cell RNA sequencing identifies diverse roles of epithelial cells in idiopathic

pulmonary fibrosis. *JCI insight* **1,** e90558 (2016).

27. Jacob, A. *et al.* Differentiation of Human Pluripotent Stem Cells into Functional Lung Alveolar Epithelial Cells. *Cell Stem Cell* **21,** 472–488.e10 (2017).

28. Ardini-Poleske, M. E. *et al.* LungMAP: The Molecular Atlas of Lung Development Program. *Am. J. Physiol. Lung Cell. Mol. Physiol.* **313,** L733–L740 (2017).

29. Slowikowski, K., Hu, X. & Raychaudhuri, S. SNPsea: an algorithm to identify cell types, tissues and pathways affected by risk loci. *Bioinformatics* **30,** 2496–7 (2014).

30. Wakefield, J. A Bayesian measure of the probability of false discovery in genetic epidemiology studies. *Am. J. Hum. Genet.* **81,** 208–27 (2007).

31. Visscher, P. M. *et al.* 10 Years of GWAS Discovery: Biology, Function, and Translation. *Am. J. Hum. Genet.* **101,** 5–22 (2017).

32. Zhou, X. *et al.* Identification of a chronic obstructive pulmonary disease genetic determinant that regulates HHIP. *Hum. Mol. Genet.* **21,** 1325–35 (2012).

33. Claussnitzer, M., Hui, C.-C. & Kellis, M. FTO Obesity Variant and Adipocyte Browning in Humans. *N. Engl. J. Med.* **374,** 192–3 (2016).

34. Pers, T. H. *et al.* Biological interpretation of genome-wide association studies using predicted gene functions. *Nat. Commun.* **6,** 5890 (2015).

35. Cho, M. H. *et al.* A Genome-Wide Association Study of Emphysema and Airway Quantitative Imaging Phenotypes. *Am. J. Respir. Crit. Care Med.* **192,** 559–69 (2015).

36. MacArthur, J. *et al.* The new NHGRI-EBI Catalog of published genome-wide association studies (GWAS Catalog). *Nucleic Acids Res.* **45,** D896–D901 (2017).

37. Demenais, F. *et al.* Multiancestry association study identifies new asthma risk loci that colocalize with immune-cell enhancer marks. *Nat. Genet.* **50,** 42–53 (2018).

38. 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).

39. Sanseau, P. *et al.* Use of genome-wide association studies for drug repositioning. *Nat. Biotechnol.* **30,** 317–20 (2012).

40. Lencz, T. & Malhotra, A. K. Targeting the schizophrenia genome: a fast track strategy from GWAS to clinic. *Mol. Psychiatry* **20,** 820–6 (2015).

41. Lamb, J. *et al.* The Connectivity Map: using gene-expression signatures to connect small molecules, genes, and disease. *Science* **313,** 1929–35 (2006).

42. Sirota, M. *et al.* Discovery and preclinical validation of drug indications using compendia of public gene expression data. *Sci. Transl. Med.* **3,** 96ra77 (2011).

43. Skronska-Wasek, W. *et al.* Reduced Frizzled Receptor 4 Expression Prevents WNT/β-Catenin-driven Alveolar Lung Repair in Chronic Obstructive Pulmonary Disease. *Am. J. Respir. Crit. Care Med.* **196,** 172–185 (2017).

44. Sakornsakolpat, P. *et al.* Integrative genomics identifies new genes associated with severe COPD and emphysema. *Respir. Res.* **19,** 46 (2018).

45. Bui, D. S. *et al.* Childhood predictors of lung function trajectories and future COPD risk: a prospective cohort study from the first to the sixth decade of life. *Lancet. Respir. Med.* (2018). doi:10.1016/S2213-2600(18)30100-0

46. McGeachie, M. J. *et al.* Patterns of Growth and Decline in Lung Function in Persistent Childhood Asthma. *N. Engl. J. Med.* **374,** 1842–1852 (2016).

47. Ross, J. C. *et al.* Longitudinal Modeling of Lung Function Trajectories in Smokers with and without COPD. *Am. J. Respir. Crit. Care Med.* (2018). doi:10.1164/rccm.201707-1405OC

48. Boucherat, O., Morissette, M. C., Provencher, S., Bonnet, S. & Maltais, F. Bridging Lung Development with Chronic Obstructive Pulmonary Disease. Relevance of Developmental Pathways in Chronic Obstructive Pulmonary Disease Pathogenesis. *Am. J. Respir. Crit. Care Med.* **193,** 362–75 (2016).

49. Nelson, M. R. *et al.* The support of human genetic evidence for approved drug indications. *Nat. Genet.* **47,** 856–60 (2015).

50. Miossec, P. & Kolls, J. K. Targeting IL-17 and TH17 cells in chronic inflammation. *Nat. Rev. Drug Discov.* **11,** 763–76 (2012).

51. Mellett, M. *et al.* Orphan receptor IL-17RD tunes IL-17A signalling and is required for neutrophilia. *Nat. Commun.* **3,** 1119 (2012).

52. O'Leary, N. A. *et al.* Reference sequence (RefSeq) database at NCBI: current status, taxonomic expansion, and functional annotation. *Nucleic Acids Res.* **44,** D733-45 (2016).

53. Saito, A., Ozaki, K., Fujiwara, T., Nakamura, Y. & Tanigami, A. Isolation and mapping of a human lung-specific gene, TSA1902, encoding a novel chitinase family member. *Gene* **239,** 325–31 (1999).

54. Fagerberg, L. *et al.* Analysis of the human tissue-specific expression by genome-wide integration of transcriptomics and antibody-based proteomics. *Mol. Cell. Proteomics* **13,** 397–406 (2014).

55. Aminuddin, F. *et al.* Genetic association between human chitinases and lung function in COPD. *Hum. Genet.* **131,** 1105–14 (2012).

56. Birben, E. *et al.* The effects of an insertion in the 5'UTR of the AMCase on gene expression and pulmonary functions. *Respir. Med.* **105,** 1160–9 (2011).

57. Chatterjee, R., Batra, J., Das, S., Sharma, S. K. & Ghosh, B. Genetic association of acidic mammalian chitinase with atopic asthma and serum total IgE levels. *J. Allergy Clin. Immunol.* **122,** 202–8, 208.e1–7 (2008).

58. Ober, C. & Chupp, G. L. The chitinase and chitinase-like proteins: a review of genetic and functional studies in asthma and immune-mediated diseases. *Curr. Opin. Allergy Clin. Immunol.* **9,** 401–8 (2009).

59. Heinzmann, A. *et al.* Joint influences of Acidic-Mammalian-Chitinase with Interleukin-4 and Toll-like receptor-10 with Interleukin-13 in the genetics of asthma. *Pediatr. Allergy Immunol.* **21,** e679-86 (2010).

60. Okawa, K. *et al.* Loss and Gain of Human Acidic Mammalian Chitinase Activity by Nonsynonymous SNPs. *Mol. Biol. Evol.* **33,** 3183–3193 (2016).

61. Yang, J. *et al.* Rootletin, a novel coiled-coil protein, is a structural component of the ciliary rootlet. *J. Cell Biol.* **159,** 431–40 (2002).

62. Gibson, M. A., Hughes, J. L., Fanning, J. C. & Cleary, E. G. The major antigen of elastin-associated microfibrils is a 31-kDa glycoprotein. *J. Biol. Chem.* **261,** 11429–36 (1986).

63. Massaro, G. D. *et al.* Retinoic acid receptor-beta: an endogenous inhibitor of the perinatal formation of pulmonary alveoli. *Physiol. Genomics* **4,** 51–7 (2000).

64. Markovics, J. A. *et al.* Interleukin-1beta induces increased transcriptional activation of the transforming growth factor-beta-activating integrin subunit beta8 through altering chromatin architecture. *J. Biol. Chem.* **286,** 36864–74 (2011).

65. Kitamura, H. *et al.* Mouse and human lung fibroblasts regulate dendritic cell trafficking, airway inflammation, and fibrosis through integrin αvβ8-mediated activation of TGF-β. *J. Clin. Invest.* **121,** 2863–75 (2011).

66. Araya, J. *et al.* Squamous metaplasia amplifies pathologic epithelial-mesenchymal interactions in COPD patients. *J. Clin. Invest.* **117,** 3551–62 (2007).

67. Zeltz, C. & Gullberg, D. The integrin-collagen connection - a glue for tissue repair? *J. Cell Sci.* **129,** 1284 (2016).

68. Hall, N. G., Klenotic, P., Anand-Apte, B. & Apte, S. S. ADAMTSL-3/punctin-2, a novel glycoprotein in extracellular matrix related to the ADAMTS family of metalloproteases. *Matrix Biol.* **22,** 501–10 (2003).

69. Apte, S. S. A disintegrin-like and metalloprotease (reprolysin-type) with thrombospondin type 1 motif (ADAMTS) superfamily: functions and mechanisms. *J. Biol. Chem.* **284,** 31493–7 (2009).

70. Kutz, W. E. *et al.* ADAMTS10 protein interacts with fibrillin-1 and promotes its deposition in extracellular matrix of cultured fibroblasts. *J. Biol. Chem.* **286,** 17156–67 (2011).

71. Gabriel, L. A. R. *et al.* ADAMTSL4, a secreted glycoprotein widely distributed in the eye, binds fibrillin-1 microfibrils and accelerates microfibril biogenesis. *Invest. Ophthalmol. Vis. Sci.* **53,** 461–9 (2012).

72. Tsutsui, K. *et al.* ADAMTSL-6 is a novel extracellular matrix protein that binds to fibrillin-1 and promotes fibrillin-1 fibril formation. *J. Biol. Chem.* **285,** 4870–82 (2010).

73. Ghosh, M. *et al.* Exhaustion of Airway Basal Progenitor Cells in Early and Established Chronic Obstructive Pulmonary Disease. *Am. J. Respir. Crit. Care Med.* **197,** 885–896 (2018).

74. Crystal, R. G. Airway basal cells. The 'smoking gun' of chronic obstructive pulmonary disease. *Am. J. Respir. Crit. Care Med.* **190,** 1355–62 (2014).

75. Giordano, R. J. *et al.* Targeted induction of lung endothelial cell apoptosis causes emphysema-like changes in the mouse. *J. Biol. Chem.* **283,** 29447–60 (2008).

76. Franks, T. J. *et al.* Resident cellular components of the human lung: current knowledge and goals for research on cell phenotyping and function. *Proc. Am. Thorac. Soc.* **5,** 763–6 (2008).

77. Boschetto, P. *et al.* Predominant emphysema phenotype in chronic obstructive pulmonary. *Eur. Respir. J.* **21,** 450–4 (2003).

78. Castaldi, P. J. *et al.* Cluster analysis in the COPDGene study identifies subtypes of smokers with distinct patterns of airway disease and emphysema. *Thorax* **69,** 415–22 (2014).

79. Cerveri, I. *et al.* The rapid FEV(1) decline in chronic obstructive pulmonary disease is associated with predominant emphysema: a longitudinal study. *COPD* **10,** 55–61 (2013).

80. Bonàs-Guarch, S. *et al.* Re-analysis of public genetic data reveals a rare X-chromosomal variant associated with type 2 diabetes. *Nat. Commun.* **9,** 321 (2018).

81. Dehghan, A. *et al.* Meta-analysis of genome-wide association studies in >80 000 subjects identifies multiple loci for C-reactive protein levels. *Circulation* **123,** 731–8 (2011).

82. Snoeck-Stroband, J. B. *et al.* Chronic bronchitis sub-phenotype within COPD: inflammation in sputum and biopsies. *Eur. Respir. J.* **31,** 70–7 (2008).

83. Hersh, C. P. *et al.* Non-emphysematous chronic obstructive pulmonary disease is associated with diabetes mellitus. *BMC Pulm. Med.* **14,** 164 (2014).

84. Higami, Y. *et al.* Increased Epicardial Adipose Tissue Is Associated with the Airway Dominant Phenotype of Chronic Obstructive Pulmonary Disease. *PLoS One* **11,** e0148794 (2016).

85. Chung, K. F. & Barnes, P. J. Cytokines in asthma. *Thorax* **54,** 825–57 (1999).

86. Kroegel, C., Julius, P., Matthys, H., Virchow, J. C. & Luttmann, W. Endobronchial secretion of interleukin-13 following local allergen challenge in atopic asthma: relationship to interleukin-4 and eosinophil counts. *Eur. Respir. J.* **9,** 899–904 (1996).

87. Boutten, A. *et al.* Decreased expression of interleukin 13 in human lung emphysema. *Thorax* **59,** 850–4 (2004).

88. Washko, G. R. *et al.* Lung volumes and emphysema in smokers with interstitial lung abnormalities. *N. Engl. J. Med.* **364,** 897–906 (2011).

89. Chilosi, M., Poletti, V. & Rossi, A. The pathogenesis of COPD and IPF: distinct horns of the same devil? *Respir. Res.* **13,** 3 (2012).

90. Kulkarni, T., O'Reilly, P., Antony, V. B., Gaggar, A. & Thannickal, V. J. Matrix Remodeling in Pulmonary Fibrosis and Emphysema. *Am. J. Respir. Cell Mol. Biol.* **54,** 751–60 (2016).

91. Wan, M., Qin, J., Songyang, Z. & Liu, D. OB fold-containing protein 1 (OBFC1), a human homolog of yeast Stn1, associates with TPP1 and is implicated in telomere length regulation. *J. Biol. Chem.* **284,** 26725–31 (2009).

92. Albrecht, E. *et al.* Telomere length in circulating leukocytes is associated with lung function and disease. *Eur. Respir. J.* **43,** 983–92 (2014).

93. Armanios, M. Telomerase and idiopathic pulmonary fibrosis. *Mutat. Res.* **730,** 52–8 (2012).

94. Stanley, S. E. *et al.* Telomerase mutations in smokers with severe emphysema. *J. Clin. Invest.* **125,** 563–70 (2015).

95. Tinkelman, D. G., Price, D. B., Nordyke, R. J. & Halbert, R. J. Misdiagnosis of COPD and asthma in

primary care patients 40 years of age and over. *J. Asthma* **43,** 75–80

96. Foreman, M. G. *et al.* Alpha-1 Antitrypsin PiMZ Genotype Is Associated with Chronic Obstructive Pulmonary Disease in Two Racial Groups. *Ann. Am. Thorac. Soc.* **14,** 1280–1287 (2017).

97. Han, M. K. *et al.* Gender and chronic obstructive pulmonary disease: why it matters. *Am. J. Respir. Crit. Care Med.* **176,** 1179–84 (2007).

98. Miller, M. R. *et al.* Standardisation of spirometry. *Eur. Respir. J.* **26,** 319–38 (2005).

99. 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).

100. Regan, E. A. *et al.* Genetic epidemiology of COPD (COPDGene) study design. *COPD* **7,** 32–43 (2010).

101. McCarthy, S. *et al.* A reference panel of 64,976 haplotypes for genotype imputation. *Nat. Genet.* **48,** 1279–83 (2016).

102. Chang, C. C. *et al.* Second-generation PLINK: rising to the challenge of larger and richer datasets. *Gigascience* **4,** 7 (2015).

103. Willer, C. J., Li, Y. & Abecasis, G. R. METAL: fast and efficient meta-analysis of genomewide association scans. *Bioinformatics* **26,** 2190–2191 (2010).

104. Bulik-Sullivan, B. K. *et al.* LD Score regression distinguishes confounding from polygenicity in genome-wide association studies. *Nat. Genet.* **47,** 291–5 (2015).

105. Lutz, S. M. *et al.* A genome-wide association study identifies risk loci for spirometric measures among smokers of European and African ancestry. *BMC Genet.* **16,** 138 (2015).

106. Loth, D. W. *et al.* Genome-wide association analysis identifies six new loci associated with forced vital capacity. *Nat Genet* **46,** 669–677 (2014).

107. 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).

108. Berisa, T. & Pickrell, J. K. Approximately independent linkage disequilibrium blocks in human populations. *Bioinformatics* **32,** 283–5 (2016).

109. Consortium, T. and G. & Tobacco and Genetics Consortium. Genome-wide meta-analyses identify multiple loci associated with smoking behavior. *Nat. Genet.* **42,** 441–7 (2010).

110. 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).

111. Genome Refernce Consortium. Human Genome Region MHC. Available at: https://www.ncbi.nlm.nih.gov/grc/human/regions/MHC?asm=GRCh37.p13.

112. Pruim, R. J. *et al.* LocusZoom: regional visualization of genome-wide association scan results. *Bioinformatics* **26,** 2336–7 (2010).

113. Finucane, H. K. *et al.* Partitioning heritability by functional annotation using genome-wide association summary statistics. *Nat. Genet.* **47,** 1228–35 (2015).

114. Durinck, S., Spellman, P. T., Birney, E. & Huber, W. Mapping identifiers for the integration of genomic datasets with the R/Bioconductor package biomaRt. *Nat. Protoc.* **4,** 1184–91 (2009).

115. Barbeira, A. N. *et al.* Integrating Predicted Transcriptome From Multiple Tissues Improves Association Detection. *bioRxiv* (2018).

116. Hao, K. *et al.* Lung eQTLs to help reveal the molecular underpinnings of asthma. *PLoS Genet.* **8,** e1003029 (2012).

117. Lamontagne, M. *et al.* Leveraging lung tissue transcriptome to uncover candidate causal genes in COPD genetic associations. *Hum. Mol. Genet.* **27,** 1819–1829 (2018).

118. Gamazon, E. R. *et al.* A gene-based association method for mapping traits using reference transcriptome data. *Nat Genet* **47,** 1091–8 (2015).

119. Shooshtari, P., Huang, H. & Cotsapas, C. Integrative Genetic and Epigenetic Analysis Uncovers Regulatory Mechanisms of Autoimmune Disease. *Am. J. Hum. Genet.* **101,** 75–86 (2017).

120. Bernstein, B. E. *et al.* The NIH Roadmap Epigenomics Mapping Consortium. *Nat. Biotechnol.* **28,** 1045–8 (2010).

121. Morrow, J. D. *et al.* Human Lung DNA Methylation Quantitative Trait Loci Colocalize with COPD Genome-wide Association Loci. *Am. J. Respir. Crit. Care Med.* (2018). doi:10.1164/rccm.201707-1434OC

122. Hormozdiari, F. *et al.* Colocalization of GWAS and eQTL Signals Detects Target Genes. *Am. J. Hum. Genet.* **99,** 1245–1260 (2016).

123. Schmitt, A. D. *et al.* A Compendium of Chromatin Contact Maps Reveals Spatially Active Regions in the Human Genome. *Cell Rep.* **17,** 2042–2059 (2016).

124. Martin, J. S. *et al.* HUGIn: Hi-C Unifying Genomic Interrogator. *Bioinformatics* **33,** 3793–3795 (2017).

125. Qiao, D. *et al.* Whole exome sequencing analysis in severe chronic obstructive pulmonary disease. *Hum. Mol. Genet.* (2018). doi:10.1093/hmg/ddy269

126. Lee, S., Fuchsberger, C., Kim, S. & Scott, L. An efficient resampling method for calibrating single and gene-based rare variant association analysis in case-control studies. *Biostatistics* **17,** 1–15 (2016).

127. Pickrell, J. K. *et al.* Detection and interpretation of shared genetic influences on 42 human traits. *Nat. Genet.* **48,** 709–17 (2016).

128. Caliński, T. & Harabasz, J. A dendrite method for cluster analysis. *Commun. Stat.* **3,** 1–27 (1974).

129. Dimas, A. S. *et al.* Impact of type 2 diabetes susceptibility variants on quantitative glycemic traits reveals mechanistic heterogeneity. *Diabetes* **63,** 2158–71 (2014).

130. Friedman, J., Hastie, T. & Tibshirani, R. Regularization Paths for Generalized Linear Models via Coordinate Descent. *J. Stat. Softw.* **33,** 1–22 (2010).

131. EMBL-EBI. GWAS Catalog. Available at: https://www.ebi.ac.uk/gwas/. (Accessed: 10th April 2018)

132. Machiela, M. J. & Chanock, S. J. LDlink: a web-based application for exploring population-specific

haplotype structure and linking correlated alleles of possible functional variants. *Bioinformatics* **31,** 3555–7 (2015).

133. Zheng, J. *et al.* LD Hub: a centralized database and web interface to perform LD score regression that maximizes the potential of summary level GWAS data for SNP heritability and genetic correlation analysis. *Bioinformatics* **33,** 272–279 (2017).

134. Corsello, S. M. *et al.* The Drug Repurposing Hub: a next-generation drug library and information resource. *Nat. Med.* **23,** 405–408 (2017).

135. Subramanian, A. *et al.* A Next Generation Connectivity Map: L1000 Platform and the First 1,000,000 Profiles. *Cell* **171,** 1437–1452.e17 (2017).