Genetic landscape of chronic obstructive pulmonary disease identifies heterogeneous cell-type and phenotype associations

Phuwanat Sakornsakolpat 1,2,51, Dmitry Prokopenko 1,3,51, 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 28,29, John E. Hokanson 30, Woo Jin Kim 31, David A. Lomas 32, Stephanie J. London 10, Deborah A. Meyers 12, George T. O’Connor 33,34, Stephen I. Rennard 35,36, David A. Schwartz 37,38, Pawel Sliwinski 39, David Sparrow 40, David P. Strachan 41, Ruth Tal-Singer 42, Yohannes Tesfaigzi 17, Jørgen Vestbo 43, Judith M. Vonk 22,23, Jae-Joon Yim 44, Xiaobo Zhou 1, Yohan Bossé 4,45, Ani Manichaikul 46,47, Lies Lahousse 18,48, Edwin K. Silverman 1,24, H. Marike Boezen 22,23, Louise V. Wain 5,49, Martin D. Tobin 5,49, Brian D. Hobbs 1,24,52, Michael H. Cho 1,2,4,52, SpiroMeta Consortium 50 and International COPD Genetics Consortium 50

Chronic obstructive pulmonary disease (COPD) is the leading cause of respiratory mortality worldwide. Genetic risk loci provide new insights into disease pathogenesis. 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 associated with \( P < 5 \times 10^{-8} \); 47 of these were previously described in association with either COPD or population-based measures of lung function. Of the remaining 35 new loci, 13 were associated with lung function in 79,055 individuals from the SpiroMeta consortium. Using gene expression and regulation data, we identified functional enrichment of COPD risk loci in lung tissue, smooth muscle, and several lung cell types. We found 14 COPD loci shared with either asthma or pulmonary fibrosis. COPD genetic risk loci clustered into groups based on associations with quantitative imaging features and comorbidities. Our analyses provide further support for the genetic susceptibility and heterogeneity of COPD.

COPD is a disease of enormous and growing global burden3, ranked third as a global cause of death by the World Health Organization in 2016 (ref. 4). 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 have demonstrated that genetic factors account for a substantial fraction of disease susceptibility. Similarly to other adult-onset complex diseases, common variants are likely to account for the majority of population genetic susceptibility5,6. Our previous efforts have identified 22 genome-wide-significant loci6. Expanding the number of loci can lead to new insights into disease pathogenesis, not only through discovery of new biological links at individual loci6,7 but also across loci via identification of functional links and specific cell types and phenotypes6.

We performed a genome-wide association study (GWAS) combining previously described studies from the International COPD Genetics Consortium (ICGC)7 with additional subjects from the UK Biobank7, 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 biological 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 ICGC and UK Biobank (Fig. 1). We defined COPD on the basis of prebronchodilator spirometry according to modified Global Initiative for Chronic Obstructive Lung Disease (GOLD) criteria for moderate to very severe airflow limitation9, as done previously9. 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, by using a fixed-effects model. We found no evidence of confounding by population substructure using the linkage disequilibrium score regression (LDSC) intercept (1.0377, standard error (s.e.) = 0.0094).

A full list of affiliations appears at the end of the paper.

https://doi.org/10.1038/s41588-018-0342-2
We identified 82 loci (defined by using 2-Mb windows) at genome-wide significance ($P < 5 \times 10^{-8}$) (Figs. 1 and 2, and Supplementary Figs. 1 and 2). Forty-seven of the 82 loci were previously described as genome-wide significant in COPD or lung function\textsuperscript{5–10} (Supplementary Table 2), thus leaving 35 new loci (Table 1) at the time of analysis. We then sought to replicate these loci. Given the strong genetic correlation between population-based measures of lung function and COPD, we tested the lead variant at each locus for association with forced expiratory volume in 1 s (FEV$_1$) or FEV$_1$/forced vital capacity (FVC) in 79,055 individuals fromSpiroMeta (Supplementary Table 3). We identified 13 loci—Clorf87, DENND2D, DBX1, SLMAP, BTC, FGF18, CITED2, ITGB8, STN1, ARNTL, SERP2, DTWD1, and ADAMTS13—that replicated by using a Bonferroni correction for one-sided $P$ value ($P < 0.05$/35; Table 1). Although they did not meet the strict Bonferroni threshold, 14 additional new loci were nominally significant in SpiroMeta (consistent direction of effect and one-sided $P < 0.05$): ASAP2, EML4, VGLL4, ADCYS, HSPA4, CCDC69, RREB1, ID4, IER3, RAF6, MFHAS1, COL15A1, TEPP, and THRA (Table 1), and all 82 loci showed a consistent direction of effect in COPD and either FEV$_1$ or FEV$_1$/FVC ratio in SpiroMeta (Table 1 and Supplementary Table 2). We note that 9 of our 35 new loci were recently described in a contemporaneous analysis of lung function in UK Biobank\textsuperscript{11}. None of the new 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 Note). In addition, we found no significant differences in variant effects by sex (Supplementary Note). Including all 82 genome-wide-significant variants, we explained up to 7.0% of the phenotypic variance on the liability scale, using a 10% prevalence for COPD, acknowledging that these effects are likely to be underestimated in the discovery sample. This represents up to a 48% increase in COPD phenotypic variance explained by genetic loci, as compared with the 4.7% explained by the 22 loci reported in a recent GWAS of COPD\textsuperscript{1}.

Identification of secondary association signals. We used approximate conditional and joint analysis\textsuperscript{22} to find secondary signals at each of the 82 genome-wide-significant loci. We found 82 secondary signals at 50 loci, thus resulting in a total of 164 independent associations at 82 loci (Supplementary Table 4). Of the 50 loci containing secondary associations, 33 were at loci previously described for COPD or lung function, and 6 were at Bonferroni-replicated new loci. Of the 82 secondary associations, 20 reached genome-wide significance ($P < 5 \times 10^{-8}$) (Supplementary Table 4). Of the 61 new (not previously described in COPD or lung function) independent associations, 21 reached a region-wise Bonferroni-corrected significance threshold (one-sided $P < 0.05$/new independent association(s) at each locus) in unconditioned associations from SpiroMeta (Methods and Supplementary Table 4).

Tissues and specific cell types. In determining the tissue in which COPD genetic variants function to increase COPD risk, the lung is the obvious tissue to consider. However, COPD is a systemic disease\textsuperscript{12,24}, and within the lung, the cell types collectively contributing to disease pathogenesis are largely unknown. Furthermore, available databases include nonlung cell types, such as smooth muscle in the gastrointestinal (GI) tract, that are relevant to lung biology. To identify putative causal tissues and cell types, we assessed heritability enrichment by using integrated genome annotations at the single-tissue level\textsuperscript{25} and tissue-specific epigenomic marks\textsuperscript{26}. Lung tissue showed the highest enrichment (enrichment $= 9.25$, $P = 1.36 \times 10^{-10}$), as previously described, although significant enrichment was also seen in the heart (enrichment $= 6.85$, $P = 3.83 \times 10^{-8}$) and the 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 (DHSs) ($P = 6.75 \times 10^{-8}$) and H3K4me1 ($P = 7.31 \times 10^{-8}$) (Supplementary Table 5). To identify the source of association within lung tissue, we tested for heritability enrichment by using single-cell chromatin accessibility\textsuperscript{27} (ATAC-seq) and gene expression (RNA-seq) data from human\textsuperscript{5,28} and mouse\textsuperscript{29} lung (Supplementary Table 5). Using LD score regression in mouse ATAC-seq data, we found enrichment of chromatin accessibility in several cell types, including endothelial cells (most significant) as well as type 1 and type 2 alveolar cells (with the latter among the tissues with the highest fold enrichment; Supplementary Table 5b). Results from using LD score regression\textsuperscript{30} or SNPsea\textsuperscript{31} on single-cell RNA-seq data varied, with nominal $P$ values for genes expressed in type 2 alveolar cells, basal-like cells, club cells, fibroblasts, and smooth muscle cells (Supplementary Table 5c,d).

Fine-mapping of associated loci. To identify the most likely causal variants at each locus, we performed fine-mapping using Bayesian credible sets\textsuperscript{31}. 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 Fig. 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 (NC_000004.11:g.106819053C>G>A; Supplementary Note and Supplementary Table 6). Most sets included variants that overlapped enhancer regions in lung-related cell types (for example, fetal lung fibroblasts, fetal lung, and adult lung fibroblasts) and were predicted to alter transcription-factor-binding motifs (Supplementary Table 6). Of the 61 credible sets with fewer than 50 variants, 8 sets contained at least one deleterious variant. These deleterious variants included (i) missense variants affecting TNS1, RIN3, ADGRG6, ADAM19, ATP13A2, BTC, and CRLF3 and (ii) a splice donor variant affecting a long intergenic noncoding RNA (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 for disease-associated variants\textsuperscript{32–34}. Thus, to identify the potential effector (‘target’) genes underlying the genetic associations, we integrated additional molecular information including gene expression, gene regulation (open chromatin and methylation data), chromatin
| rsID       | Common gene | Locus  | Risk allele | P  | P  | P  |
|------------|-------------|--------|-------------|----|----|----|
| rs2603499  |             |        | T/C         | 0.0511 | 1.08-1.11 | 2.0×10−4 |
| rs9229366  |             |        | C/T         | 0.751 | 1.05-1.10 | 1.7×10−8 |
| rs1431861  |             |        | G/A         | 0.34 | 1.06-1.12 | 9.2×10−9 |
| rs469195    |             |        | T/C         | 0.071 | 1.05-1.2 | 6.2×10−9 |
| rs2208926   |             |        | G/A         | 0.74 | 1.05-1.1 | 6.2×10−9 |
| rs2293886   |             |        | T/C         | 0.38 | 1.05-1.2 | 6.2×10−9 |
| rs62259026  |             |        | G/A         | 0.75 | 1.05-1.1 | 6.2×10−9 |
| rs5935808   |             |        | T/C         | 0.38 | 1.05-1.2 | 6.2×10−9 |
| rs646691    |             |        | T/C         | 0.071 | 1.05-1.2 | 6.2×10−9 |
| rs1410722   |             |        | G/A         | 0.74 | 1.05-1.1 | 6.2×10−9 |
| rs60577     |             |        | T/C         | 0.38 | 1.05-1.2 | 6.2×10−9 |
| rs688656    |             |        | G/A         | 0.74 | 1.05-1.1 | 6.2×10−9 |
| rs798565    |             |        | G/A         | 0.74 | 1.05-1.1 | 6.2×10−9 |
| rs34651     |             |        | T/C         | 0.38 | 1.05-1.2 | 6.2×10−9 |
| rs10760580  |             |        | G/A         | 0.74 | 1.05-1.1 | 6.2×10−9 |
| rs967865    |             |        | T/C         | 0.38 | 1.05-1.2 | 6.2×10−9 |
| rs3660361   |             |        | G/A         | 0.74 | 1.05-1.1 | 6.2×10−9 |
| rs66364     |             |        | T/C         | 0.38 | 1.05-1.2 | 6.2×10−9 |
| rs3660361   |             |        | G/A         | 0.74 | 1.05-1.1 | 6.2×10−9 |
| rs66364     |             |        | T/C         | 0.38 | 1.05-1.2 | 6.2×10−9 |
| rs3660361   |             |        | G/A         | 0.74 | 1.05-1.1 | 6.2×10−9 |
| rs66364     |             |        | T/C         | 0.38 | 1.05-1.2 | 6.2×10−9 |
| rs3660361   |             |        | G/A         | 0.74 | 1.05-1.1 | 6.2×10−9 |
| rs66364     |             |        | T/C         | 0.38 | 1.05-1.2 | 6.2×10−9 |
Identification of drug targets. GWAS is also useful for identifying drug targets at either the individual gene level or genome-wide level. Of the 472 candidate target genes, 59 genes were targeted by at least one approved or in-development drug, for a total of 427 drugs with 134 different modes of action (Supplementary Table 9).

Phenotypic effects of COPD-associated variants. To characterize the phenotypic effects of the 82 genome-wide-significant loci, we performed a phenome-wide association analysis within the deeply phenotyped COPDGene study (Methods). We looked 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 differently associated with two sets of phenotypes (Supplementary Fig. 4). Because these two variants–phenotype clusters appeared to be driven by computed tomography (CT) imaging features, we repeated variant clustering but limited analysis 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 (Fig. 4a). Additionally, we evaluated the association of the 82 genome-wide-significant variants in a prior GWAS of emphysema and airway quantitative CT features (Supplementary Table 10).

We also examined all genome-wide-significant loci in the NHGRI-EBI GWAS Catalog (Supplementary Fig. 5 and Supplementary Table 11) and looked for trait-associated variants in LD (r² > 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 blood 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 (by using LD score regression²) of COPD with lung function, asthma, and height and found evidence of a modest interaction, co-regulation of gene expression with gene sets, and coding variant data (Fig. 3 and Methods).

At the 82 associated loci, 472 genes within 1 Mb of the top associated variants were implicated by analysis of at least one dataset; 106 genes were implicated by lung gene expression, and an additional 50 genes were implicated by two or more other datasets (methylation, chromatin interaction, open chromatin regions, similarity in gene sets, or deleterious coding variants; Fig. 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 the implicated genes to the top nearest gene among the implicated genes was 346 kb. Of the 82 loci, 60 (73%) included the associated variants were implicated by analysis of at least one dataset; 106 genes were implicated by lung gene expression, and an additional 50 genes were implicated by two or more other datasets (methylation, chromatin interaction, open chromatin regions, similarity in gene sets, or deleterious coding variants; Fig. 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 the implicated genes to the top nearest gene among the implicated genes was 346 kb. Of the 82 loci, 60 (73%) included the
correlation between COPD and lung cancer (Supplementary Note). However, at individual loci and using a more stringent LD threshold ($r^2 > 0.6$), 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 associated with CAD (near CEDP1, DMWD, STN1, and TNS1), and a variant near SPPL2C associated with bone density (Fig. 4b).

Overlapping loci with asthma and pulmonary fibrosis. On the basis of our previous identification of genetic overlap of COPD with asthma and COPD with pulmonary fibrosis, we examined loci for overlap specifically with these two diseases. In asthma, we noted $r^2 > 0.2$ 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 FM13A, DSP, and 17q21, we noted overlapping associations at ZKSCAN1 and STN1 (Supplementary Table 12). To more closely examine overlap, we applied a Bayesian method (gwas-pw) to COPD associations from our current GWAS and previous GWAS results for asthma (limited to those of European ancestry) and pulmonary fibrosis. To mitigate the effects of including individuals with asthma among our COPD cases, we performed analysis for overlap with asthma after removing self-reported individuals with asthma from UK Biobank for this analysis (Methods). We identified 14 shared genome segments (posterior probability >70%), 9 for asthma and 5 for pulmonary fibrosis (Fig. 4c and Supplementary Table 13). In addition to the three segments shared with pulmonary fibrosis identified in the previous study (FM13A, DSP, and the 17q21 locus—here nearest CRHR1), we identified two new segments including loci near ZKSCAN1 and STN1 (formerly known as OBFC1). Shared variants for COPD and pulmonary fibrosis all had opposite directions of effect in the two diseases (that is, increasing risk for COPD but protective for pulmonary fibrosis). In asthma, we identified five shared segments in the 6p21-22 region, as well as the segments containing ADAM19, ARMC2, ELAVL2, and STAT6. With the exception of the segment with STAT6, overlapping variants showed the same direction of effect.

Discussion

Genetic factors have an important role in COPD susceptibility. We examined genetic risk of COPD in a GWAS of 35,735 cases and 222,076 controls. We identified 82 genome-wide significant loci for COPD, of which 47 were previously identified in GWAS of COPD or population-based measures of lung function. Of the 35 loci not previously described at the time of analysis, 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 the 82 loci that were supported by either gene expression or a combination of at least two other data sources. Our results identify specific genes, cell types, and biological pathways for targeted study and also suggest a genetic basis for the clinical heterogeneity seen in COPD.

Our study supports a role for early life events in the risk of COPD. Gene-set enrichment analysis identified developmental pathways both specific to the lung (for example, lung morphogenesis and lung alveolar development) and related to the lung (for example, the canonical Wnt receptor and MAPK–ERK and nerve growth factor receptor signaling pathways). We also confirmed enrichment of heritability for epimorphic marks in the fetal lung. Our findings are consistent with epidemiological studies demonstrating that a substantial portion of the risk for COPD may develop in early
life: genetic variants may set initial lung function\(^6\) and patterns of growth\(^6,8-10\). Although 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 mouse or ex vivo models—for example, determining whether perturbation of these genes changes proliferation and differentiation of lung epithelial progenitors in induced pluripotent stem cell (iPSC)-derived alveolar type 2 lung cells\(^13\)—could provide experimental evidence for 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\(^41\).

Apart from genes related to lung development, our analyses highlighted several genes and pathways already of interest in COPD therapy (including the COPD target/therapy relationships exemplified by CHRM3/acyethylcholine receptor inhibitors and the MAPK pathway/p38 inhibitors)—supporting the role of genetic analyses in finding therapeutic targets\(^28,32\)—and newer genes that may inform future functional studies. We identified IL17RD (encoding IL-17 receptor D) as a potential effector gene at the 3p14 locus. Numerous studies have examined the role of IL-17A in COPD\(^4\), and IL17RD can differentially regulate pathways employed by IL-17A\(^4\). CHIA (chitinase acidic) at 1p13.3, which encodes a protein that degrades chitin\(^65\), exhibits lung-specific expression\(^66,67\). CHIA variants have been associated with FEV\(_1\), asthma\(^68-70\), and acid mammalian chitinase activity\(^71,72\). We identified several potential effector genes related to extracellular matrix, cell adhesion, cell–cell interactions, and elastin-associated microfibrils\(^73,74\), some of which have previously been identified in studies of lung function\(^7\). These include integrin family members that mediate cell–matrix communication (for example, ITGA1, ITGA2, and ITGA8\(^75-77\)), an integrin-ligand encoding gene (NPNT\(^7\)), and genes encoding matrix proteins (for example, MFAP2 and ADAMTSL3). ADAMTSL3 has a role in cell–matrix interactions related to assembly of fibrillin and microfibrillogenesis\(^74,78\), and of our candidate effector genes, it was supported by the greatest number of bioinformatic analyses. Recombinant forms of other ADAMTS-like proteins have shown evidence in experiments of promoting and enhancing fibrillin and microfibril deposition and assembly\(^74,78\). ADAMTSL3 may have a role in preventing emphysematous destruction of lung tissue by ADAMTs in COPD.

In addition to identifying the effector gene for a variant, knowing the effector cell type is critical for functional studies. We identified an overall enrichment of epimark genotypes in lung tissue and GI smooth muscle (also identified in studies of lung function\(^8\)); respiratory smooth muscle was absent from 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\(^88-90\), and our data are consistent with the likely heterogeneity of lung cell types contributing to COPD susceptibility. The lung comprises at least 40 different resident cell types\(^4\), most of which were 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 locus-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 showed heterogeneous effects of genetic variants on COPD-related phenotypes, including CT measurements of airway abnormalities and emphysema—well-defined sources of heterogeneity in COPD\(^90-92\). Analyzing...
hundreds of diseases and traits in the GWAS Catalog, we identified overlapping associations with various diseases and traits in multiple organ systems, including comorbidities such as CAD, bone mineral density, and T2D. The COPD-associated PABPC4 locus was associated with T2D and C-reactive protein (CRP) levels. Although a causal gene in this locus and its contribution to COPD are unknown, the association of this locus with T2D may suggest a shared disease pathway and drug targets. In sum, the identification of variable associations for COPD risk loci with subphenotypes and other diseases 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 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. Genome-wide genetic correlation of COPD with asthma has previously been described, but our analysis is, to our knowledge, the first to identify specific genetic segments shared by asthma and COPD. While the effects at most of these shared segments were concordant in direction, one of the segments of particular interest near STAT6 had opposite directions of effect in the two diseases. STAT6 has a role in helper T cell type 2–dependent inflammation and is activated by IL-4 and IL-13. IL-13 in turn has been found at increased levels in asthmatic airways but at decreased levels in severe emphysema. In pulmonary fibrosis, variants at all overlapping loci have an opposite direction of effect in comparison to COPD. This observation raises the possibility that specific therapies for one disease could increase the risk for the other disease, which may be worth evaluating in treatment trials. The reasons why genetic effects are divergent in COPD and fibrosis are unclear, but these identified opposite effects may be worth evaluating in specific genetic segments shared by asthma and COPD. While the direction of effect in comparison to COPD is consistent across studies, our current understanding of the biology underlying these pathways is limited.

5. This observation raises the possibility that specific therapies for one disease could increase the risk for the other disease, which may be worth evaluating in treatment trials. The reasons why genetic effects are divergent in COPD and fibrosis are unclear, but these identified opposite effects may be worth evaluating in specific genetic segments shared by asthma and COPD. While the direction of effect in comparison to COPD is consistent across studies, our current understanding of the biology underlying these pathways is limited.

Although our study is a large GWAS of COPD, individuals meeting our criteria for COPD in UK Biobank may be different from those in other studies, especially with respect to smoking history. We used the same definition of COPD as in our previous analysis, which included nonsmokers. Our use of prebronchodilator spirometry to define COPD (allowing us to maximize sample size) as well as population-based measures of 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. We performed several additional analyses to determine whether our findings were driven by or markedly different as a result of smoking status, asthma, or use of pre- instead of postbronchodilator spirometry to define COPD. The results of these additional analyses did not indicate a substantial effect of these factors on our overall findings and, together with prior analyses, suggest that bias due to these factors is likely to be 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) has been identified as genome-wide significant in smaller studies of emphysema and in smokers with severe COPD. In the current study, the PiZ allele (NC_000014.8:g.94844947C>T, rs28929470) was associated with $P=2.2 \times 10^{-5}$ by using moderate to severe cases (FEV1 <80% predicted) and with a smaller P value (1.4 \times 10^{-5}) in severe cases (FEV1 <50% predicted) despite a smaller sample size, a phenomenon that we have previously described. Thus, despite the strong overlap of COPD with quantitative spirometry measures, new loci may be identified through studies of sufficiently large subsets of patients with COPD and with more specific and homogenous COPD phenotypes. Given the suggestive evidence for replication using a related (but not identical) phenotype for additional new 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 probably 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 ancestry groups are warranted, but broader multiancestry analyses are limited by the number of cases in currently available cohorts. Although COPD sex differences have been reported, we did not identify significant sex-specific differences in the effect sizes of the 82 top variants. Future studies including more subjects and methodological advances may be needed to elucidate this.

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

Online content
Any methods, additional references, Nature Research reporting summaries, source data, statements of data availability and associated accession codes are available at https://doi.org/10.1038/s41588-018-0342-2.

Received: 25 June 2018; Accepted: 20 December 2018; Published online: 25 February 2019

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. World Health Organization. Global Health Estimates 2016: Disease burden by Cause, Age, Sex, by Country and by Region, 2000–2016. https://www.who.int/healthinfo/global_burden_disease/estimates/en/index1.html (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, 941–947 (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 β-catenin. Am. J. Respir. Crit. Care Med. 194, 185–197 (2016).

7. Lao, T. et al. Hhip haplinsufficiency sensitizes mice to age-related emphysema. Proc. Natl. Acad. Sci. USA 113, E6811–E6816 (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. Bulik-Sullivan, B. K. et al. LD Score regression distinguishes confounding from polygenicity in genome-wide association studies. Nat. Genet. 49, 205–210 (2017).

11. 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).

12. Wilk, J. B. 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).

13. Soler Artigas, M. et al. Genome-wide association analysis of lung function and chronic obstructive pulmonary disease identifies new loci and potential druggable targets. Nat. Genet. 49, 416–425 (2017).

14. Wain, L. et al. Novel genetic loci for lung function and interstitial lung disease. PLoS Genet. 10, e1004290 (2014).

15. Repapi, E. et al. Genomic analysis of breast cancer identifies multiple loci associated with lung function. Nat. Genet. 42, 36–44 (2010).

16. Wain, L. et al. Genome-wide association studies identify five loci associated with lung function. Nat. Genet. 42, 45–52 (2010).

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

18. 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).

19. Shrim, N. et al. New genetic signals for lung function highlight pathways and pleiotropy, and chronic obstructive pulmonary disease associations across multiple ancestries. Preprint at https://www.biorxiv.org/content/early/2018/05/02/343293 (2018).

20. Vongpatanasin, W. et al. Conditional and joint multiple-SNP analysis of GWAS summary statistics identifies additional variants influencing complex traits. Nat. Genet. 44, 369–375 (2012).

21. Agusti, A. & Soriano, J. B. COPD as a systemic disease. Lancet 373, 133–138 (2009).

22. 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. 10, e1004933 (2014).

23. Finucane, H. K. et al. Partitioning heritability by functional annotation using genome-wide association summary statistics. Nat. Genet. 47, 1228–1235 (2015).

24. Cusanovich, D. A. 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).

25. Finucane, H. K. et al. Partitioning heritability by functional annotation using genome-wide association summary statistics. Nat. Genet. 47, 1228–1235 (2015).

26. 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).

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

28. Slowikowski, K., Hu, X. & Raychaudhuri, S. SNPsea: an algorithm to identify cell types, tissues and pathways affected by risk loci. Bioinformatics 30, 2496–2497 (2014).

29. Wakefield, J. A Bayesian measure of the probability of false discovery in genetic epidemiology studies. Am. J. Hum. Genet. 81, 208–227 (2007).

30. Schlessinger, J. & Pestka, J. V. Obesity, adipocyte browning and human disease. Nat. Rev. Endocrinol. 10, 112–119 (2014).

31. Derksen, J. C. et al. Longitudinal modeling of lung function trajectories in smokers with and without COPD. Am. J. Respir. Crit. Care Med. 198, 1033–1042 (2018).

32. Boucherat, O., Morissette, M. C., Provenccher, 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, 362–375 (2016).
88. Giordano, R. J. et al. Targeted induction of lung endothelial cell apoptosis and is required for neutrophilia. *Nat. Commun.* 3, 1119 (2012).

65. O’Leary, N. A. et al. Reference Sequence (RefSeq) database at NCBI: current status, taxonomic expansion, and functional annotation. *Nucleic Acids Res.* 34, D733–D745 (2016).

66. 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).

75. Gibson, M. A., Hughes, J. L., Fanning, J. C. & Cleary, E. G. The major histocompatibility class I subunit α1 binds to fibrillin-1 and promotes fibrillin-1 fibril formation. *J. Biol. Chem.* 286, 11626–11637 (2011).

76. Massaro, G. D. et al. Retinoic acid receptor-β activation of the transforming growth factor-β-activating integrin subunit β1, 4870–4882 (2010).

77. Markovics, J. A. et al. Interleukin-1β: an endogenous inhibitor of β-1 integrin in atopic asthma and serum total IgE levels. *J. Allergy Clin. Immunol.* 122, 202–208 (2008).

78. Gibson, M. A., Hughes, J. L., Fanning, J. C. & Cleary, E. G. The major histocompatibility class I subunit α1 binds to fibrillin-1 and promotes fibrillin-1 fibril formation. *J. Biol. Chem.* 286, 11626–11637 (2011).

79. Araya, J. et al. Squamous metaplasia amplifies pathologic epithelial–mesenchymal interactions in COPD patients. *J. Clin. Invest.* 117, 3551–3562 (2007).

80. Zelts, D. & Gullberg, D. The integrin–collagen connection: a glue for tissue repair. *J. Cell. Sci.* 129, 1284 (2016).

81. Hall, N. G., Klenotic, P., Anand-Apte, B. & Apte, S. S. ADAMTSL-3/β3punctin-2, a novel glycoprotein in extracellular matrix related to the ADAMTS family of metalloproteases. *Matrix Biol.* 22, 501–510 (2003).

82. Apte, S. S. A disintegrin-like and metalloprotease (reprolysin-type) with thrombospondin type 1 motif (ADAMTS) superfamily: functions and mechanisms. *J. Biol. Chem.* 284, 31493–31497 (2009).

83. 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, 17156–17167 (2011).

84. Gabriel, L. A. R. et al. ADAMTS4, a secreted glycoprotein widely distributed in the eye, binds fibrillin-1 microfibrils and accelerates microfibril biogenesis. *Invest. Ophthalmol. Vis. Sci.* 53, 461–469 (2012).

85. Tsutsumi, K. et al. ADAMTS-6 is a novel extracellular matrix protein that binds to fibrillin-1 and promotes fibrillin-1 fibril formation. *J. Biol. Chem.* 285, 4870–4882 (2010).

86. 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).

87. Crystal, R. G. Airway basal cells. The ‘smoking gun’ of chronic obstructive pulmonary disease. *Am. J. Respir. Crit. Care Med.* 190, 1355–1362 (2014).

88. Giordano, R. J. et al. Targeted induction of lung endothelial cell apoptosis causes emphysema-like changes in the mouse. *J. Biol. Chem.* 283, 29447–29460 (2008).
competed to study concept and design. All authors, including those whose initials are not listed above, contributed to critical review and editing of the manuscript and approved the final version of the manuscript.

Competing interests
M.H.C., E.K.S., L.V.W., M.D.T., D.A.L., and L.P.H. have received grant funding from GlaxoSmithKline (GSK). E.K.S. has received honoraria from Novartis for continuing medical education seminars and travel support from GSK. L.P.H. has received grant support from BL. R.T.-S. is an employee and shareholder of GSK. J.V. has received personal fees from GSK, Chiesi Pharmaceuticals, BI, Novartis, and AstraZeneca. D.L.D. has received grants from the National Institutes of Health for research on COPD and personal fees from Novartis. D.A.L. has received honoraria from GSK and chaired the Respiratory Therapy Area Board from 2012 to 2015. Outside the submitted work, L.L. reports expert consultation for Boehringer Ingelheim and Novartis and unrestricted grants from AstraZeneca and Chiesi.

Additional information
Supplementary information is available for this paper at https://doi.org/10.1038/s41588-018-0342-2.
Reprints and permissions information is available at www.nature.com/reprints.
Correspondence and requests for materials should be addressed to M.H.C.

Publisher's note: Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

© The Author(s), under exclusive licence to Springer Nature America, Inc. 2019
SpiroMeta Consortium

Nick Shrine5, Anna L. Guyatt5, Chiara Batini5, Jing Hua Zhao53, Matthias Wielchers54, Understanding Society Scientific Group55, Stefan Weiss56, Katherine A. Kentistou57,58, James P. Cook59, Jennie Hui60,61,62,63, Stefan Karrasch64,65,66, Medea Imboden57,68, Sarah E. Harris69,70, Jonathan Marten71, Stefan Enroth72, Shona M. Kerr71, Ida Surakkak73,74, Veronique Vitart71, Terho Lehtimäki75, Ralf Ewert76, Christian Gieger77, Georg Homuth56, Peter K. Joshi57, Claudia Langenberg78, Lars Lind79, Jian’an Luan78, Anubha Mahajan80, Alison Murray81, David J. Porteous69,70, Rajesh Rawal77,82, Blair H. Smith83, Paul R. H. J. Timmers57, Olli T. Raitakari84,85, Ulf Gyllensten72, Igor Rudan57, Ian J. Deary69,88, Nicole M. Probst-Hensch67,68, Holger Schulz64,66, Alan L. James60,89,90, James F. Wilson57,71, Beate Stubbe76, Eleftheria Zeggini91,92, Marjo-Riitta Jarvelin54,93,94,95,96, Nick Wareham78, Caroline Hayward77, Andrew P. Morris59,80, David P. Strachan41, Ian P. Hall88,29, Martin D. Tobin5,49 and Louise V. Wain5,49

53Cardiovascular Epidemiology Unit, Department of Public Health and Primary Care, University of Cambridge, Cambridge, UK. 54Department of Epidemiology and Biostatistics, MRC–PHE Centre for Environment & Health, School of Public Health, Imperial College London, London, UK. 55Interfaculty Institute for Genetics and Functional Genomics, Department of Functional Genomics, University Medicine Greifswald, Greifswald, Germany. 56Centre for Global Health Research, Usher Institute for Population Health Sciences and Informatics, University of Edinburgh, Edinburgh, UK. 57Centre for Cardiovascular Sciences, Queen’s Medical Research Institute, University of Edinburgh, Edinburgh, UK. 58Department of Biostatistics, University of Liverpool, Liverpool, UK. 59Busselton Population Medical Research Institute, Sir Charles Gairdner Hospital, Nedlands, Western Australia, Australia. 60School of Population Health, University of Western Australia, Crawley, Western Australia, Australia. 61PathWest Laboratory Medicine of WA, Sir Charles Gairdner Hospital, Crawley, Western Australia, Australia. 62School of Pathology and Laboratory Medicine, University of Western Australia, Crawley, Western Australia, Australia. 63Institute of Epidemiology, Helmholtz Zentrum München—German Research Center for Environmental Health, Neuherberg, Germany. 64Institute and Outpatient Clinic for Occupational, Social and Environmental Medicine, Ludwig-Maximilians-Universitat, Munich, Germany. 65Institute of Molecular Medicine Finland (FIMM), University of Helsinki, Helsinki, Finland. 66Institute of Epidemiology and Social Medicine, University of Bergen, Bergen, Norway. 67Department of Biostatistics, University of Oxford, Oxford, UK. 68Department of Public Health Institute, Basel, Switzerland. 69Institute of Public Health and Primary Care, University of Cambridge, Cambridge, UK. 70Department of Clinical Chemistry, Fimlab Laboratories, and Finnish Cardiovascular Research Center–Tampere, Faculty of Medicine and Life Sciences, University of Tampere, Tampere, Finland. 71Department of Population Health and Genomics, Ninewells Hospital and Medical School, University of Dundee, Dundee, UK. 72Department of Clinical Pathology and Nuclear Medicine, Turku University Hospital, Turku, Finland. 73Research Centre of Applied and Preventive Cardiovascular Medicine, University of Turku, Turku, Finland. 74Department of Clinical Chemistry, Fimlab Laboratories, and Finnish Cardiovascular Research Center–Tampere, Faculty of Medicine and Life Sciences, University of Tampere, Tampere, Finland. 75Department of Pulmonary Physiology and Sleep Medicine, Sir Charles Gairdner Hospital, Nedlands, Western Australia, Australia. 76School of Medicine and Pharmacology, University of Western Australia, Crawley, Western Australia, Australia. 77Institute for Molecular Medicine Finland (FIMM), University of Helsinki, Helsinki, Finland. 78National Institute for Health and Welfare (THL), Helsinki, Finland. 79Department of Clinical Chemistry, Fimlab Laboratories, and Finnish Cardiovascular Research Center–Tampere, Faculty of Medicine and Life Sciences, University of Tampere, Tampere, Finland. 80Wellcome Trust Centre for Human Genetics, University of Oxford, Oxford, UK. 81Department of Medical Sciences, Aberdeen Biomedical Imaging Centre, University of Aberdeen, Aberdeen, UK. 82Institute of Epidemiology and Social Medicine, University of Münster, Münster, Germany. 83Division of Population Health and Genomics, Ninewells Hospital and Medical School, University of Dundee, Dundee, UK. 84Department of Clinical Pathology and Nuclear Medicine, Turku University Hospital, Turku, Finland. 85Research Centre of Applied and Preventive Cardiovascular Medicine, University of Turku, Turku, Finland. 86Department of Clinical Chemistry, Fimlab Laboratories, and Finnish Cardiovascular Research Center–Tampere, Faculty of Medicine and Life Sciences, University of Tampere, Tampere, Finland. 87Division of Population Health and Genomics, Ninewells Hospital and Medical School, University of Dundee, Dundee, UK. 88Department of Clinical Pathology and Nuclear Medicine, Turku University Hospital, Turku, Finland. 89Department of Primary Health Care, Oulu University Hospital, OYS, Oulu, Finland. 90School of Medicine and Pharmacology, University of Oulu, Oulu, Finland. 91Wellcome Trust Centre for Human Genetics, University of Oxford, Oxford, UK. 92Institute of Medical Sciences, Aberdeen Biomedical Imaging Centre, University of Aberdeen, Aberdeen, UK. 93Center for Translational Genomics, Helmholtz Zentrum München—German Research Center for Environmental Health, Neuherberg, Germany. 94School of Medicine and Pharmacology, University of Western Australia, Crawley, Western Australia, Australia. 95School of Medical Sciences, University of Aberdeen, Aberdeen, UK. 96Research Centre of Applied and Preventive Cardiovascular Medicine, University of Turku, Turku, Finland. 97Department of Clinical Pathology and Nuclear Medicine, Turku University Hospital, Turku, Finland. 98School of Medicine and Pharmacology, University of Western Australia, Crawley, Western Australia, Australia. 99Department of Pulmonary Physiology and Sleep Medicine, Sir Charles Gairdner Hospital, Nedlands, Western Australia, Australia. 100Institute of Molecular Medicine Finland (FIMM), University of Helsinki, Helsinki, Finland. 101Department of Clinical Chemistry, Fimlab Laboratories, and Finnish Cardiovascular Research Center–Tampere, Faculty of Medicine and Life Sciences, University of Tampere, Tampere, Finland. 102Department of Pulmonary Physiology and Sleep Medicine, Sir Charles Gairdner Hospital, Nedlands, Western Australia, Australia. 103On behalf of the International COPD Genetics Consortium.

International COPD Genetics Consortium

Alvar Agusti97, Wayne Anderson98, Nawar Bakerly99,100, Per Bakke14, Robert Bals101, Kathleen C. Barnes37, R. Graham Barr15, Terri H. Beaty16, Eugene R. Bleecker12, H. Marike Boezen22,23, Yohan Bossé4,45, Russell Bowler21, Christopher Brightling49,102, Marleen de Bruijne103,104, Peter J. Castaldi1, Bartolome Celli24, Michael H. Cho1,24, Harvey O. Coxson105, James D. Crapo21, Ron Crystal106, Pim de Jong107, Asger Dirksen108, Jennifer Dy109, Marilyn Foreman110, Judith Garcia-Aymerich111,112,113, 114, Jason Gill114, Rachel Goble114, Ronald L. Graham114, John J. D. Hackett115, James R. Harst116, A. Michael Halushka117,8,118, Jonathan E. Harding119, Janne Hartikainen119, Steven M. Hinders120, Richard A. Hogg121, Brian S. Hackett122, James C. S. Hare123, Ralf Ewert76, J. An K. Ando124, fores are limited.

International COPD Genetics Consortium

Alvar Agusti97, Wayne Anderson98, Nawar Bakerly99,100, Per Bakke14, Robert Bals101, Kathleen C. Barnes37, R. Graham Barr15, Terri H. Beaty16, Eugene R. Bleecker12, H. Marike Boezen22,23, Yohan Bossé4,45, Russell Bowler21, Christopher Brightling49,102, Marleen de Bruijne103,104, Peter J. Castaldi1, Bartolome Celli24, Michael H. Cho1,24, Harvey O. Coxson105, James D. Crapo21, Ron Crystal106, Pim de Jong107, Asger Dirksen108, Jennifer Dy109, Marilyn Foreman110, Judith Garcia-Aymerich111,112,113, 114, Jason Gill114, Rachel Goble114, Ronald L. Graham114, John J. D. Hackett115, James R. Harst116, A. Michael Halushka117,8,118, Jonathan E. Harding119, Janne Hartikainen119, Steven M. Hinders120, Richard A. Hogg121, Brian S. Hackett122, James C. S. Hare123, Ralf Ewert76, J. An K. Ando124, fores are limited.
Pierre Ghevenois¹¹⁴, Soumitra Ghosh⁴², Hester Gietema¹¹⁵, Amund Gulsvik¹⁴, Ian P. Hall²⁸,²⁹, Nadia Hansel¹¹⁶, Craig P. Hersh¹,²⁴, Brian D. Hobbs¹,²⁴, Eric Hoffman¹⁷, Noor Kalsheker¹¹⁸, Hans-Ulrich Kauczor¹⁷⁹, Woo Jin Kim³¹, Deog Kyeom Kim⁹, Tarja Laitinen¹²⁰, Diether Lambrechts¹²¹,¹²², Sang-Do Lee¹²³, Augusto A. Litonjua²⁴, David A. Lomas³², Stephanie J. London¹⁰, Daan W. Loth¹⁸, Sharon M. Lutz¹²⁵, David Lynch¹²⁶, William MacNee¹²⁷, Merry-Lynn McDonald¹²⁸, Deborah A. Meyers¹², John D. Newell¹¹⁷, Borge G. Nordestgaard¹²⁹,¹³⁰, George T. O’Connor³³,³⁴, Ma’en Obeidat¹³⁰, Yeon-Mok Oh¹²³, Peter D. Paré¹³,¹³¹, Massimo Pistolesi¹³², Dirkje S. Postma²³, Milo Puhan¹³³, Elizabeth Regan¹³⁴, Stephen S. Rich⁴⁶, Joon Beom Seo¹³⁵, Andrea Short¹³⁶, Edwin K. Silverman¹²⁴, David Sparrow⁴⁰, Berend Stoel¹³⁷, David P. Strachan⁴¹, Nicola Sverzellati¹³⁸, Ruth Tal-Singer¹⁴², Gerbenter Riet¹³⁹, Yohannes Tesfaigzi¹⁷, Martin D. Tobin⁵,⁴⁹, Edwin J. R. Van Beek¹⁴⁰, Bram van Ginneken¹⁴¹, Jørgen Vestbo⁴³, Claus F. Vogelmeier¹⁴², Louise V. Wain⁵,⁴⁹, Adam Wanner¹⁴³, George Washko¹⁴⁴, Els Wauters¹⁴⁵, Emiel F. M. Wouters¹⁴⁶, Robert P. Young¹⁴⁷ and Loems Zeigler-Heitbrock¹⁴⁸

¹²⁹Respiratory Institute, Hospital Clinic, University of Barcelona, IDIABIPS, CIBERES, Barcelona, Spain. ¹³⁰Department of Medicine, University of North Carolina at Chapel Hill, Chapel Hill, NC, USA. ¹³¹Salford Royal NHS Foundation Trust, Salford, UK. ¹³²Manchester Academic Health Sciences Centre, University of Manchester, Manchester, UK. ¹³³Department of Internal Medicine V-Pulmonology, Allergology and Critical Care Medicine, Saarland University, Homburg, Germany. ¹³⁴Department of Respiratory Medicine, Allergy and Thoracic Surgery, Glenfield Hospital, University of Leicester, Leicester, UK. ¹³⁵Biomedical Imaging Group Rotterdam, Departments of Radiology and Medical Informatics, Erasmus Medical Center, Rotterdam, the Netherlands. ¹³⁶Department of Computer Science, University of Copenhagen, Copenhagen, Denmark. ¹³⁷Department of Radiology, University of British Columbia, Vancouver, British Columbia, Canada. ¹³⁸Department of Genetic Medicine and Division of Pulmonary and Critical Care Medicine, Department of Medicine, Weill Cornell Medical College, New York, NY, USA. ¹³⁹Department of Respiratory Medicine, University Medical Center Utrecht, Utrecht, the Netherlands. ¹⁴⁰Department of Respiratory Medicine, Herlev and Gentofte Hospital, Copenhagen, Denmark. ¹⁴¹Department of Electrical and Computer Engineering, Northeastern University, Boston, MA, USA. ¹⁴²Morehouse School of Medicine, Atlanta, GA, USA. ¹⁴³Global, Barcelona, Spain. ¹⁴⁴Universitat Pompeu Fabra (UPF), Barcelona, Spain. ¹⁴⁵CIBER Epidemiología y Salud Publica (CIBERESP), Barcelona, Spain. ¹⁴⁶Department of Radiology, Hôpital Erasme, Université Libre de Bruxelles, Brussels, Belgium. ¹⁴⁷Department of Radiology, Maastricht University Medical Center, Maastricht, the Netherlands. ¹⁴⁸Department of Medicine, Johns Hopkins University, Baltimore, MD, USA. ¹⁴⁹Department of Radiology, University of Iowa, Iowa City, IA, USA. ¹⁵⁰School of Life Sciences, University of Nottingham, Nottingham, UK. ¹⁵¹Department of Diagnostic and Interventional Radiology, University of Heidelberg, Translational Lung Research Center Heidelberg, Heidelberg, Germany. ¹⁵²Department of Pulmonary Diseases and Clinical Allergology, Turku University Hospital, University of Turku, Turku, Finland. ¹⁵³Vesalius Research Center (VRC), VIB, Leuven, Belgium. ¹⁵⁴Laboratory for Translational Genetics, Department of Oncology, KU Leuven, Leuven, Belgium. ¹⁵⁵Department 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, Republic of Korea. ¹⁵⁶Division of Pediatric Pulmonary Medicine, Department of Pediatrics, Golisano Children’s Hospital at Strong, University of Rochester Medical Center, Rochester, NY, USA. ¹⁵⁷Department of Population Medicine, Harvard Pilgrim Health Care Institute and Harvard Medical School, Boston, MA, USA. ¹⁵⁸Department of Radiology, National Jewish Health, Denver, CO, USA. ¹⁵⁹Department of Respiratory Medicine, University of Edinburgh, Edinburgh, UK. ¹⁶⁰Division of Pulmonary, Allergy and Critical Care Medicine, University of Alabama at Birmingham, Birmingham, AL, USA. ¹⁶¹Department of Clinical Biochemistry and Copenhagen General Population Study, Herlev and Gentofte Hospital, Copenhagen University Hospital, Herlev, Denmark. ¹⁶²Faculty of Health and Medical Sciences, University of Copenhagen, Copenhagen, Denmark. ¹⁶³Department of Medicine, University of British Columbia, Institute for Heart and Lung Health, St. Paul’s Hospital, Vancouver, British Columbia, Canada. ¹⁶⁴Department of Clinical and Experimental Medicine, University Hospital Careggi, Florence, Italy. ¹⁶⁵Epidemiology, Biostatistics and Prevention Institute (EBPI), University of Zurich, Zurich, Switzerland. ¹⁶⁶Division of Rheumatology, Department of Medicine, National Jewish Health, Denver, CO, USA. ¹⁶⁷Department of Radiology and Research Institute of Radiology, Asan Medical Center, University of Ulsan College of Medicine, Songpa-gu, Seoul, Republic of Korea. ¹⁶⁸School of Social Sciences, University of Manchester, Manchester, UK. ¹⁶⁹Leiden University Medical Center, Leiden, the Netherlands. ¹⁷⁰Radiology, Department of Medicine and Surgery, University of Parma, Parma, Italy. ¹⁷¹Department of General Practice, Academic Medical Center, University of Amsterdam, Amsterdam, the Netherlands. ¹⁷²Edinburgh Imaging, University of Edinburgh, Edinburgh, UK. ¹⁷³Department of Radiology and Nuclear Medicine, Radboud University Medical Center, Nijmegen, the Netherlands. ¹⁷⁴Pulmonary Diseases, University of Marburg, Marburg, Germany. ¹⁷⁵Division of Pulmonary, Allergy, Critical Care and Sleep Medicine, Department of Medicine, University of Miami, Miller School of Medicine, Miami, FL, USA. ¹⁷⁶Department of Radiology, Brigham and Women’s Hospital, Boston, MA, USA. ¹⁷⁷Laboratory of Pneumology, Department of Chronic Diseases, Metabolism and Ageing, KU Leuven, Leuven, Belgium. ¹⁷⁸Respiratory Medicine, Maastricht University Medical Center, Maastricht, the Netherlands. ¹⁷⁹School of Biological Sciences and Faculty of Medical and Health Sciences, University of Auckland, Auckland, New Zealand. ¹⁸⁰EvA Study Center, Helmholtz Zentrum München, Gauting, Germany.
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 to estimate enrichment of functional annotations and specifically expressed gene regions in disease-associated loci we identified. We used a logistic regression baseline models (for example, conserved region or promoter-flanking region), tissue-specific annotations from the Roadmap Epigenomis Project, integrated tissue annotations from GenoSkyline, and cell-type-specific chromatin accessibility data (ATAC-seq). We used four single-cell gene expression (RNA-seq) datasets to identify specific cell types (Supplementary Note), including (i) lung epithelial cells from normal and pulmonary fibrosis human lung (Gene Expression Omnibus (GEO) accession GSE86618), (ii) human iPSC-derived putative alveolar type 2 cells (GSE96612), and (iii) mouse lungs at embryonic day (E) 18.5 and (iv) mouse lungs at postnatal day (P) 1 from Whittet et al. (unpublished; available at LungMAP ). We also used SNPsea to identify enriched cell types in genome-wide-significant loci. We performed lookups of select significant variants for FEV1 and FEV1/FVC in the SpiroMeta consortium meta-analysis. Briefly, SpiroMeta comprised 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, by using rank-based inverse normal transforms, adjusting for population substructure by 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 relatedness). We defined a genetic locus as being previously described (evidence of prior association) if its posterior probability was greater than or equal to 0.99, by using unscaled variance. To determine whether these signals colocalized (rather than being related owing to LD), we performed colocalization analysis between our GWAS loci and mQTLs in genome-wide-significant loci by using eCAVIAR. We also sought information from publicly available chromosome-conformation capture data. We queried association statistics of chromatin contact (that is, long-range chromatin interactions) between top associated variants and gene promoters near the lung (fetal lung fibroblast cell line (IMR90) and human lung tissue) using HUGI. We retained only the strongest associations (that is, those with the smallest P values) for each cell line or primary cell type in the analysis. We then searched for signals from deleterious variants by querying the consequences of variants within 99%-credible sets containing fewer than 50 variants (Supplementary Note). We also searched for rare coding variants, on the basis of exome-sequencing results in the COPDGene, Boston Early-Onset COPD (BEOCOPD), and International COPD Genetics Network (ICGN) studies, as previously described. We performed de novo 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 BECOCPD and ICGN studies. Details on statistical tests for single-variant and gene-based analyses are summarized in the Supplementary Note.

Identification of independent associations at genome-wide-significant loci. We identified specific independent associations at genome-wide-significant loci by using GCTA-COJO. This method uses an approximate conditional and joint analysis approach requiring summary statistics and representative LD information. Because UK Biobank provided the predominant sample, we used 10,000 randomly drawn unrelated individuals from this discovery database as an LD reference sample. We scaled genome-wide significance to a 2-Mb region, providing a locus-wide-significance threshold of 8 × 10^-5. We first characterized loci as being previously described (evidence of prior association with lung function19-22,23 or COPD24,25) or new. We defined previously reported loci if they achieved the same LD score as in the same LD reference sample or moderate LD (r^2 ≥ 0.2) with previously described loci. For new loci, we attempted replication through association of each lead variant with either FEV1, or FEV1/FVC ratio in SpiroMeta, using one-sided P values with Bonferroni correction for the number of new loci examined. New loci failing to meet a Bonferroni-corrected P-value threshold were assessed for nominal significance (one-sided P < 0.05) or directed consistency with FEV1, or 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. 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 additional 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 Note). Identification of independent associations at genome-wide-significant loci. We identified specific independent associations at genome-wide-significant loci by using GCTA-COJO. This method uses an approximate conditional and joint analysis approach requiring summary statistics and representative LD information. Because UK Biobank provided the predominant sample, we used 10,000 randomly drawn unrelated individuals from this discovery database as an LD reference sample. We scaled genome-wide significance to a 2-Mb region, thus resulting in a locus-wide-significance threshold of 8 × 10^-5, or 2 × 10^-4 for variants in the MHC region (chr.6: 28,477,797–33,448,354 in hg19). We created regional association plots via LocusZoom, using 1,000 Genomes European (EUR) reference data10 (Nov2014 release).
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 regfim and eCAVIAR. We considered genes that were implicated by gene expression or a combination of two or more other datasets (for example, 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 by using the reconstituted gene sets from DEPICT, as described above. We defined significant gene sets using FDR <5%.

Identification of drug targets. We queried our target genes by using the Drug Repurposing Hub. This resource contains comprehensive annotations of launched drugs, drugs in phases 1–3 of clinical development, previously approved and preclinical or tool compounds, curated by using publicly available sources (for example, ChEMBL and DrugBank) and proprietary sources. We performed drug-gene expression similarity analysis of the Query, using a ranked gene set from a gene-based association test (Supplementary Note).

Effects on COPD-related and other phenotypes. COPD is a complex and heterogeneous disorder comprising different biological processes and specific phenotypic effects. In addition, many loci discovered by GWAS have pleiotropic effects. To identify these effects, we performed (i) identification of overlapping genetic loci with related disorders (asthma and pulmonary fibrosis); (ii) genetic association studies of our genome-wide-significant findings by using COPD-related phenotypes, including a cluster analysis to identify groups of variants that might be acting via similar mechanisms; (iii) lookup of top variants in prior COPD-related quantitative CT imaging feature GWAS; (iv) lookup of associations with other diseases and traits by using the GWAS Catalog; and (v) estimation of the genetic correlation between COPD and other diseases and traits.

To identify overlapping loci between COPD and other respiratory disorders, we used gwas-pw to perform pairwise analysis of GWAS. This method searches for shared genomic segments by using an adaptive significance threshold, allowing detection of loci below genome-wide significance. We identified shared segments or variants by using a posterior probability of colocalization greater than 0.7 (ref. 110). We obtained GWAS summary statistics from previous studies of pulmonary fibrosis and asthma in European populations. For the overlap analysis of COPD with asthma, we examined 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 UK Biobank (with individuals with asthma removed from cases in the latter instance). 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 (for example, lung function, CT imaging-derived metrics, biomarkers, and comorbidities) available for 6,760 COPDGene non-Hispanic white individuals. 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, thus leaving us with 107 phenotypes. We then oriented all z scores to be positive (according to the sign of the median z score) in association with each phenotype to avoid clustering based on the direction of association. To avoid clustering phenotypes only by strength of association with SNPs, we scaled the z scores within each phenotype by subtracting the mean z scores and dividing by the standard deviation of the z scores within each phenotype. We then scaled z scores across SNPs to circumvent clustering of SNPs according to only the 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 performed clustering of variants both in the set of all genome-wide significant variants in discovery and in the subset of known variants plus new variants meeting a strict Bonferroni threshold in SpiroMeta replication (Supplementary Note). We further examined top variant associations with COPD-related traits through lookup of top variants in a prior GWAS of 12,031 subjects with quantitative emphysema and airway CT features. To examine the overlap of our COPD results with other traits, we downloaded genome-wide significant associations from the GWAS Catalog \( P < 5 \times 10^{-8} \), downloaded on 10 April 2018. We computed the LD for a pair of COPD- and trait-associated variants within the same LD block in Europeans by using the European-ancestry panel and considered overlap to be present if variants were in at least moderate LD \( r^2 \geq 0.2 \). We estimated genetic correlation between COPD and other diseases and traits by using a web engine for LDSC, LD Hub. We assessed the results by using a 5% Bonferroni-corrected significance level.

Reporting Summary. Further information on research design is available in the Nature Research Reporting Summary linked to this article.

Data availability
Genome-wide association summary statistics are available at the database of Genotypes and Phenotypes (dbGaP) under accession phs000179.v5.p2 and via the UK Biobank. Derived phenotypic data for COPD case-control status are also available from UK Biobank.

References
110. Miller, M. R. et al. Standardisation of spirometry. Eur. Respir. J. 26, 319–338 (2005).
111. 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, 179–187 (1999).
112. Regan, E. A. et al. Genetic epidemiology of COPD (COPDGene) study design. COPD 7, 32–43 (2010).
113. McCarthy, S. et al. A reference panel of 64,976 haplotypes for genotype imputation. Nat. Genet. 48, 1279–1283 (2016).
114. Chang, C. C. et al. Second-generation PLINK: rising to the challenge of larger and richer datasets. Gigascience 4, 7 (2015).
115. Willer, C. J., Li, Y & Abecasis, G. R. METAL: fast and efficient meta-analysis of genomewide association scans. Bioinformatics 26, 2190–2191 (2010).
116. 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).
117. Loth, D. W. et al. Genome-wide association analysis identifies six new loci associated with forced vital capacity. Nat. Genet. 46, 669–677 (2014).
118. 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).
119. Berisa, T. & Pickrell, J. K. Approximately independent linkage disequilibrium blocks in human populations. Bioinformatics 32, 283–285 (2016).
120. Tobacco and Genetics Consortium. Genome-wide meta-analyses identify multiple loci associated with smoking behavior. Nat. Genet. 42, 441–447 (2010).
121. Pruim, R. J. et al. LocusZoom: regional visualization of genome-wide association scan results. Bioinformatics 26, 2356–2357 (2010).
122. Hormozdiari, F. et al. Colocalization of GWAS and eQTL signals detects target genes. Am. J. Hum. Genet. 99, 1245–1260 (2016).
123. Martin, J. S. et al. HUGIn: Hi-C unifying genomic interrogator. Gigascience 3, 3793–3795 (2017).
124. 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, 1184–1191 (2009).
125. 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, 3555–3557 (2015).
Reporting Summary

Nature Research wishes to improve the reproducibility of the work that we publish. This form provides structure for consistency and transparency in reporting. For further information on Nature Research policies, see Authors & Referees and the Editorial Policy Checklist.

Statistical parameters

When statistical analyses are reported, confirm that the following items are present in the relevant location (e.g. figure legend, table legend, main text, or Methods section).

- The exact sample size (n) for each experimental group/condition, given as a discrete number and unit of measurement
- An indication of whether measurements were taken from distinct samples or whether the same sample was measured repeatedly
- The statistical test(s) used AND whether they are one- or two-sided
- A description of all covariates tested
- A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons
- A full description of the statistics including central tendency (e.g. means) or other basic estimates (e.g. regression coefficient) AND variation (e.g. standard deviation) or associated estimates of uncertainty (e.g. confidence intervals)
- For null hypothesis testing, the test statistic (e.g. F, t, r) with confidence intervals, effect sizes, degrees of freedom and P value noted
- Give P values as exact values whenever suitable.
- For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings
- For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes
- Estimates of effect sizes (e.g. Cohen's d, Pearson's r), indicating how they were calculated
- Clearly defined error bars
- State explicitly what error bars represent (e.g. SD, SE, CI)

Our web collection on statistics for biologists may be useful.

Software and code

Policy information about availability of computer code

Data collection

We derived various metrics on spirometry quality control using R. Software for genotype imputation: IMPUTE2 (ARIC, CHS AA, KARE, MESA, SPIROMICS), minimac (CHS EUR, COPDGene, ECLIPSE, FHS, LifeLines, NETT-NAS, GenKOLS), MaCH (COPDGene, ECLIPSE, FHS, NETT-NAS, GenKOLS), and Michigan Imputation Server (eQTL).
We generated genome-wide summary statistics for UK Biobank using PLINK v.2.0 alpha, https://www.cog-genomics.org/plink/2.0/). We performed a fixed-effects meta-analysis using METAL (v.2010-08-01, http://csg.sph.umich.edu/abecasis/Metal/download/). Regional association plots were generated using LocusZoom (v.1.4, https://github.com/statgen/locuszoom-standalone). We used LDSC (v.1.0, https://github.com/bulik/lsc/) and SNPseqa (v.1.0.3, https://github.com/slowkow/snpseqa) to perform enrichment analysis of cell types. Fine-mapping was performed using Wakefield’s method (v.2007-04-17, http://faculty.washington.edu/jonno/BCDP.R). We queried variant annotation using BioMart (v.2.36.1, http://www.biomart.org/) and HaploReg (v.4.1, http://archive.broadinstitute.org/mammals/haploreg/haploreg.php). For target gene identification, we tested association of genetically regulated gene expression and COPD using S-PrediXcan (v.0.6.1, https://github.com/hakulimlab/MetaXcan); colocalization analysis using cAVIAR (v.2.1, https://github.com/hormozd/caviar); coding associations using SKAT (v.1.3.2.1, https://cran.r-project.org/package=SKAT), GESE (v.2.0.1, https://scholar.harvard.edu/dqjao/geese), GMAT (v.0.7, https://github.com/lin-lab/GMAT), and MONSTER (v.1.3, http://www.stat.uchicago.edu/~mcpeek/software/MONSTER/index.html); DNase hypersensitivity sites using regfm (v.1.0.0, https://github.com/cotsgapslab/regfm); and reconstituted gene sets using DEPICT (v.1.40611, https://data.broadinstitute.org/mpg/depict/). We analyzed drug–gene expression signatures using the Query (https://clue.io). Other analyses were done using custom codes in R (v.3.5.0, https://www.r-project.org/).

Data

Policy information about availability of data

All manuscripts must include a data availability statement. This statement should provide the following information, where applicable:

- Accession codes, unique identifiers, or web links for publicly available datasets
- A list of figures that have associated raw data
- A description of any restrictions on data availability

The genome-wide association summary statistics is available at the database of Genotypes and Phenotypes (dbGaP) under accession phs000179.v5.p2 and via the UK Biobank. Derived phenotypic data for COPD case control status is also available in the UK Biobank.

Field-specific reporting

Please select the best fit for your research. If you are not sure, read the appropriate sections before making your selection.

- [x] Life sciences  [ ] Behavioural & social sciences  [ ] Ecological, evolutionary & environmental sciences

For a reference copy of the document with all sections, see nature.com/authors/policies/ReportingSummary-flat.pdf

Life sciences study design

All studies must disclose on these points even when the disclosure is negative.

Sample size

UK Biobank is a population-based cohort consisting of 502,682 individuals which was designed to detect genetic effect in multiple traits. The sample size calculation for each trait (e.g., COPD) was not pre-defined. For studies in International COPD Genetics Consortium (ICGC), most studies pre-determined the sufficient sample size according to their designs (e.g., case-control, population-based or family-based studies). Based on the post-hoc sample size calculation, the power to detect genetic association of COPD in UK Biobank was 1 (relative risk=1.1, minor allele frequency=0.3, COPD prevalence=0.1, Ncases=21081, Ncontrols=179713, significance=0.05).

Data exclusions

The exclusion criteria were pre-established. For UK Biobank, we excluded samples based on 1) spirometry quality control (QC) 2) genotype QC and 3) Genetic ancestry. Detailed descriptions of exclusion criteria were described in Supplementary Methods. Briefly, of 502,682 individuals, 445,754 individuals had at least two measures of FEV1 and FVC and spirometry metrics, age, sex, height, and smoking. We included 324,299 individuals who had reproducible and acceptable measures of spirometry. For genotyping QC, we included 486,367 individuals after excluding outlying heterozygosity or missingness, sex mismatch, >10 3rd degree relatives, and putative sex chromosome aneuploidy. We performed further analyses on individuals with European ancestry identified by K-means clustering. For cohorts part of the International COPD Genetics Consortium, we removed subjects based on phenotype and genotype quality control as previously described. As in UK Biobank, we followed ATS/ERS guidelines for spirometry QC in ICGC studies. As in UK Biobank, ICGC studies followed ATS/ERS guidelines in performing the spirometry QC. We performed genotyping QC using the same rationale as in UK Biobank.

Replication

After we identified novel genome-wide significant associations from the meta-analysis of COPD, we looked up their associations with FEV1 or FEV1/FVC in SpiroMeta, given the high genetic correlation of COPD and lung function. We determined significance in replication using a Bonferroni correction at 5% for a number of novel associations for COPD or lung function, requiring a consistent direction of effect (one-sided Bonferroni-corrected P value < 0.05). We also considered the nominal associations (P < 0.05) and directionally consistent associations. Of 35 novel, 13 loci reached Bonferroni-corrected significance threshold (P < 0.05) and 14 loci were nominally significant (P < 0.05) in the replication studies in the SpiroMeta. All 35 novel loci were directionally consistent. Loci failing to replicate may be due to the limited sample size in the SpiroMeta consortium.

Randomization

This was observational study. Study individuals were not allocated to experimental groups.

Blinding

Study individuals were not allocated to groups; blinding was not required.
### Reporting for specific materials, systems and methods

| Materials & experimental systems | Methods |
|----------------------------------|---------|
| n/a Involved in the study        | n/a Involved in the study |
| ☒ Unique biological materials    | ☒ ChIP-seq |
| ☒ Antibodies                    | ☒ Flow cytometry |
| ☒ Eukaryotic cell lines          | ☒ MRI-based neuroimaging |
| ☒ Palaeontology                  |         |
| ☒ Animals and other organisms    |         |
| ☒ Human research participants    |         |

### Human research participants

**Policy information about studies involving human research participants**

**Population characteristics**

Characteristics of participants across 25 studies are described fully in Supplementary Table 1. We included adult men and women, smokers and non-smokers. Within the UK Biobank and the majority of the studies in the ICGC, we included individuals of European ancestry. We tested for association between genetic variants and COPD adjusting for covariates including smoking status, and stratification by smoking status.

**Recruitment**

Participants were recruited based on study design of each individual study (e.g., case-control, population-based, and family-based studies). Informed consent was obtain for all participants and participants for each individual study were recruited in accordance with the standards of the committee or governing body with jurisdiction over human subjects research.