ORIGINAL RESEARCH

Protective effect of club cell secretory protein (CC-16) on COPD risk and progression: a Mendelian randomisation study

Stephen Milne ,[1,2,3] Xuan Li,[1] Ana I Hernandez Cordero,[1] Chen Xi Yang,[1] Michael H Cho ,[4] Terri H Beaty,[5] Ingo Ruczinski,[6] Nadia N Hansel,[7] Yohan Bossé,[8] Corry-Anke Brandsma ,[9] Don D Sin,[1,2] Maen Obeidat[1]

► Additional material is published online only. To view please visit the journal online (http://dx.doi.org/10.1136/thoraxjnl-2019-214487).

For numbered affiliations see end of article.

Correspondence to
Dr Stephen Milne, Centre for Heart Lung Innovation, St Paul's Hospital, The University of British Columbia, Vancouver, BC V6Z 1Y6, Canada; stephen.milne@hli.ubc.ca

SM and XL contributed equally.

Received 23 December 2019
Revised 25 June 2020
Accepted 10 July 2020
Published Online First 24 August 2020

ABSTRACT

Background The anti-inflammatory pneumoprotein club cell secretory protein-16 (CC-16) is associated with the clinical expression of chronic obstructive pulmonary disease (COPD). We aimed to determine if there is a causal effect of serum CC-16 level on the risk of having COPD and/or its progression using Mendelian randomisation (MR) analysis.

Methods We performed a genome-wide association meta-analysis for serum CC-16 in two COPD cohorts (Lung Health Study (LHS), n=3850 and ECLIPSE, n=1702). We then used the CC-16-associated single-nucleotide polymorphisms (SNPs) as instrumental variables in MR analysis to identify a causal effect of serum CC-16 on 'COPD risk' (ie, case status in the International COPD Genetics Consortium/UK-Biobank dataset; n=35 735 COPD cases, n=222 076 controls) and 'COPD progression' (ie, annual change in forced expiratory volume in 1 s in LHS and ECLIPSE). We also determined the associations between SNPs associated with CC-16 and gene expression using n=1111 lung tissue samples from the Lung Expression Quantitative Trait Locus Study.

Results We identified seven SNPs independently associated ($p<5\times10^{-8}$) with serum CC-16 levels; six of these were novel. MR analysis suggested a protective causal effect of increased serum CC-16 on COPD risk (MR estimate (SE) −0.11 (0.04), p=0.008) and progression (LHS only, MR estimate (SE) 7.40 (3.28), p=0.02). Five of the SNPs were also associated with gene expression in lung tissue (at false discovery rate <0.1) of several genes, including the CC-16-encoding gene *SCGB1A1*.

Conclusion We have identified several novel genetic variants associated with serum CC-16 level in COPD cohorts. These genetic associations suggest a potential causal effect of serum CC-16 on the risk of having COPD and its progression, the biological basis of which warrants further investigation.

Key messages

What is the key question?

► Can genetics help uncover a causal effect of serum club cell secretory protein-16 (CC-16) level on chronic obstructive pulmonary disease (COPD) risk and/or progression?

What is the bottom line?

► There is a protective effect of genetically increased serum CC-16 on both COPD risk and progression (as measured by change in forced expiratory volume in 1 s over time), which may be due to increased expression of the CC-16-encoding gene *SCGB1A1* in the lung.

Why read on?

► This is the first study to demonstrate a possible causal effect of serum CC-16 in people with COPD and highlights the potential for CC-16 as a biomarker or therapeutic target.

INTRODUCTION

Chronic obstructive pulmonary disease (COPD) is expected to be the third leading cause of death worldwide by 2030.[1] Currently, there are no effective treatments to prevent the development of COPD or slow its progression. If disease-modifying therapies are to be developed, a better understanding of the causal mechanisms underlying the risk of COPD, and its progression over time, is critical.

Club cell secretory protein-16 (CC-16) is a 16 kDa pneumoprotein produced predominantly by club cells in the airway epithelium, where it appears to have a protective, anti-inflammatory effect.[2–4] Reduced CC-16 levels are associated with the clinical expression of COPD: serum, sputum and bronchoalveolar lavage fluid (BALF) CC-16 concentrations decrease with increasing disease severity[5–8] and reduced serum CC-16 is associated with faster lung function decline.[9 10] In a mouse model of cigarette smoke exposure, CC-16$^{-/-}$ knockout mice showed greater COPD-like lung changes than wild type mice.[3] However, this finding was not replicated by other investigators.[10] While this discrepancy may be explained by differences in experimental design (including the duration and dose of smoke exposure), it is clear that further efforts to establish a causal role for reduced CC-16 in COPD pathogenesis are required.

Mendelian randomisation (MR) analysis is a promising approach to explore causality in humans by relating genetic variants to disease outcomes via a biological risk factor (eg, protein levels).[11] Due to the random allocation of alleles at meiosis and under the assumption that the genetic variants exert their effects on disease outcomes only via the risk factor of interest, MR analysis is relatively resistant to confounding. Furthermore, since the genomic contribution to the risk factor is constant over a lifetime, MR analysis

To cite: Milne S, Li X, Hernandez Cordero AI, *et al*. *Thorax* 2020;**75**:934–943.

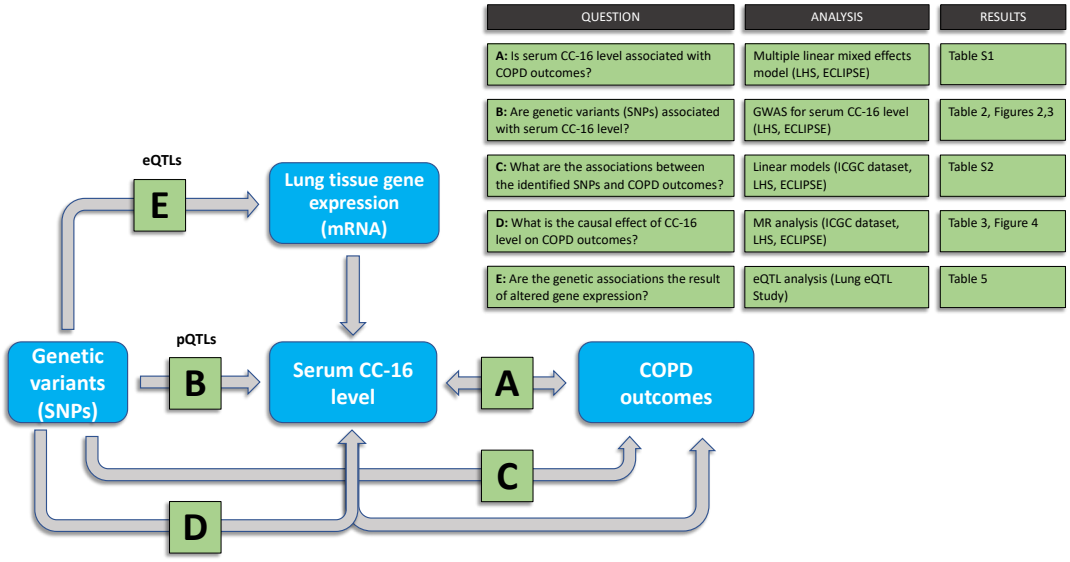

Figure 1 Study workflow. Under the MR framework, genetic variants are associated with the phenotype via the risk factor/exposure. MR analysis therefore attempts to estimate the causal effect of the exposure. The relationship may be mediated by altered gene expression. CC-16, club cell secretory protein-16; COPD, chronic obstructive pulmonary disease; eQTL, Expression Quantitative Trait Locus; ICGC dataset, meta-analysis of International COPD Genetics Consortium and UK Biobank COPD cases and non-COPD controls; LHS, Lung Health Study; MR, Mendelian randomisation; pQTL, protein quantitative trait locus; SNP, single-nucleotide polymorphism.

is relatively resistant to bias due to reverse causation (where the disease itself influences the protein level). For these reasons, MR analysis is particularly attractive for establishing a potential causal role for biological factors in complex diseases such as COPD. We have previously demonstrated a potential causal effect of low surfactant protein D in increasing the risk of having COPD and COPD progression using the MR framework.[12] Other molecules associated with COPD outcomes in epidemiological studies, such as C reactive protein,[13] interleukin-6[14] and blood eosinophil count,[15] do not yield evidence of causality when subjected to analysis by MR. Therefore, MR analysis may be useful in identifying the most promising biological pathways for drug development in COPD.

In this study, we aimed to determine if there is any causal effect of serum CC-16 level on the risk of having COPD and the rate of its progression. We performed the largest genome-wide association study (GWAS) for serum CC-16 to date and uncovered one known and six novel genetic loci associated with CC-16 levels. These loci were then used in a MR analysis to show that genetically increased serum CC-16 levels reduce the risk of having COPD and slow the rate of lung function decline in people with established COPD. Additionally, we determined that at least some of these effects may be due to increased lung tissue expression of the CC-16-encoding gene, secretoglobin family 1A member 1 (*SCGB1A1*). Our findings reinforce the importance of CC-16 in COPD pathogenesis.

METHODS

Study overview

The study workflow is outlined in figure 1. We first examined the relationship between serum CC-16 levels and changes in lung function over time. Next, we identified genetic variants (single-nucleotide polymorphisms (SNPs)) independently associated with serum CC-16 levels in a GWAS. We then determined the associations between the SNPs associated with CC-16 levels and two COPD outcomes: (1) the presence of COPD in a case-control dataset ('COPD risk') and (2) the rate of change in lung function ('COPD progression'). Following this, we performed MR analyses for 'COPD risk' and 'COPD progression' using the SNPs associated with CC-16 levels as instrumental variables (IVs) and serum CC-16 level as the risk factor. Finally, we determined which of the CC-16-associated SNPs were also quantitatively associated with gene expression (mRNA levels) in lung tissue. Access to data from each of the studies examined was granted by the respective governing committees. Participant consent and institutional ethics approval was given for each cohort study. For the present analysis, participants were not identified nor contacted, and further institutional ethics approval was not required.

Description of study cohorts

Detailed descriptions of the included studies are presented in the online supplementary methods. Brief summaries are provided below.

Lung Health Study (LHS) cohort:[16] LHS was a longitudinal study examining the effects of smoking cessation intervention and regular inhaled bronchodilator on the rate of lung function decline in 5887 smokers aged 35–60 years with mild-moderate COPD. Lung function was measured annually for the first 5 years and at year 11. At Year 5, the investigators collected blood samples for measurement of blood biomarkers and genotyping (see online supplementary methods).

Evaluation of COPD Longitudinally to Identify Predictive Surrogate Endpoints (ECLIPSE) cohort:[17] ECLIPSE was a longitudinal study of 2746 current or former smokers aged 40–75 years, who were followed for 3 years. COPD cases were defined by spirometry at baseline. Lung function measurements were performed at baseline, 3 months, 6 months and every 6 months thereafter. Blood was collected at baseline for biomarker analysis and genotyping (see online supplementary methods).

2019 International COPD Genetics Consortium (ICGC)/UK Biobank dataset:[18] ICGC is an international collaboration that pooled genome-wide association data of COPD cohort,

case-control and general population studies that all performed spirometry. A recent meta-analysis of 35 735 COPD cases and 222 076 non-COPD controls of European ancestry from the ICGC cohorts and the UK Biobank (a general population cohort of 502 682 volunteers) identified genetic variants with a genome-wide significant association with risk of having COPD. A detailed description of this meta-analysis is provided in the online supplementary methods, but for simplicity this will be referred to as the 'ICGC dataset'. For our analysis, we used the summary statistics resulting from the meta-analysis of COPD case status, hereafter referred to as 'COPD risk'.

Lung Expression Quantitative Trait Locus (eQTL) Study:[19] this study was a meta-analysis of the association between SNPs and gene expression in non-tumour lung tissue from 1111 volunteers (some of which had COPD) undergoing lung resection at three different centres, adjusted for age, sex and smoking status.

Serum CC-16 quantification

Serum CC-16 quantification in the LHS and ECLIPSE has been described previously[8 10] (see online supplementary methods). We transformed serum CC-16 concentrations by their natural logarithm to mitigate the influence of outliers and to approximate normality.

Association between serum CC-16 level and COPD progression

We tested for associations between serum CC-16 concentration and 'COPD progression' (annual change in forced expiratory volume in 1 s (FEV_1)) in the LHS and ECLIPSE using a multiple linear mixed effects (LME) model (Analysis A in figure 1—see online supplementary methods for detailed description of the model). We then combined the results of these two studies in an inverse variance weighted (IVW) meta-analysis. In order to visualise the nature of the association, we also calculated the change in FEV_1 over time using a slope model (see online supplementary methods) and plotted this against serum CC-16 concentration for LHS and ECLIPSE (online supplementary results, figure S1).

Genome-wide association study (GWAS) for serum CC-16 concentration

We determined the effects of genetic variants on serum CC-16 level (Analysis B in figure 1) by GWAS (see online supplementary methods for detailed description). Briefly, we determined the association between each SNP and serum CC-16 concentration (protein quantitative trait loci (pQTLs)) in the LHS and ECLIPSE and combined the results in an IVW fixed-effects meta-analysis. We identified independently associated SNPs using conditional analysis within each 2 Mb gene region.[20] From this, we retained only SNPs having independent, genome-wide significant ($p<5\times10^{-8}$) associations with CC-16 levels.

Associations between CC-16 pQTLs and COPD outcomes

We determined the effects of the CC-16 pQTLs on each COPD outcome. For 'COPD risk', we extracted summary statistics for each CC-16 pQTL's association with the presence of COPD in the ICGC dataset. For 'COPD progression', we tested serum CC-16 pQTLs for possible association with changes in FEV_1 over time in the LHS and ECLIPSE cohorts separately, using a LME model adjusting for covariates including the first five genetic principal components (see online supplementary methods for detailed description) (Analysis C in figure 1).

Mendelian randomisation (MR) analysis

Associations between two measured variables occur frequently in epidemiology, but whether an apparent risk factor is *causally* related to an outcome is often unknown. The MR framework is a form of causal inference testing that aims to determine if a risk factor is on the causal pathway to an outcome. To achieve this, MR incorporates the effects of genetic variants (SNPs) on both risk factor and outcome. By examining only the genetic contributions and assuming all the required assumptions are met, the association tested by MR is in a 'forward' direction (ie, risk factor→outcome) and is not biased by confounders. A concise overview of MR theory and assumptions is provided by Davies *et al*.[21]

Our MR analysis (Analysis D in figure 1) therefore estimates the causal effect of serum CC-16 concentration (ie, the risk factor) on COPD outcomes ("COPD risk" in the ICGC dataset, and "COPD progression" in LHS and ECLIPSE). A detailed description of the analysis is provided in the online supplementary methods. Briefly, we used the MendelianRandomization V.0.2.2 package in R[22 23] to construct an IVW MR model with the CC-16 pQTLs as IVs. The model takes the SNP effects on CC-16 level and on COPD outcomes (calculated above) as inputs and is adjusted for LD between SNPs. We set nominal significance at $p<0.05$ to indicate a causal effect.

Testing MR assumptions

MR analysis is only valid if the IVs meet a number of fundamental assumptions.[11 21] We employed a systematic approach to test for violations of these assumptions, as outlined by Burgess *et al*.[24] This included a GWAS catalogue look-up for SNP associations with confounders; testing for weak instrument bias (partial F statistic), reliance on individual IVs (leave-one-out sensitivity analysis) and heterogeneity (Cochran's Q) and performing analyses that are robust to violations of MR assumptions (weighted median MR, MR-Egger and MR-PRESSO). A detailed description of these analyses is provided in the online supplementary methods.

Lung tissue gene expression analysis

To determine if the serum CC-16 pQTLs have any effects on quantitative gene expression in lung tissue, we tested for their associations (at false discovery rate <0.1) with mRNA transcript levels using data from the Lung eQTL Study[19] (Analysis E in figure 1) (see online supplementary methods). For this analysis, we considered only *cis*-eQTLs, within a region 1 Mb either side of the sentinel SNP.

Statistical software

We performed all analyses in R (V.3.4.0; www.r-project.org).[25]

RESULTS

Serum CC-16 level is associated with COPD progression

Table 1 summarises the demographics, lung function and CC-16 concentrations in the two biomarker cohorts (LHS and ECLIPSE). Serum CC-16 concentration was higher in the ECLIPSE subjects, which may reflect reduced renal clearance of CC-16 associated with the older age of this cohort.[26] Serum CC-16 level was significantly associated with changes in FEV_1 in the LHS (2.43 mL/year slower decline in FEV_1 per unit increase in ln(ng/mL) CC-16, p=0.009) but not in ECLIPSE (6.25 mL/year, p=0.11), which had a shorter duration of follow-up (online supplementary results, table S1). When combined by meta-analysis, this translated to a 2.64 mL/year slower decline in

Table 1 Characteristics of the biomarker cohorts for serum CC-16 GWAS

	LHS	ECLIPSE
Total subjects, n	3850	1702
Serum CC-16, ng/mL		
Mean (SD)	3.78 (2.29)	5.64 (3.16)
Median (range)	3.35 (0.33–27.53)	5.01 (0.47–24.25)
Age, years		
Mean (SD)	53.65 (6.73)	63.63 (7.08)
Median (range)	54 (40–67)	64 (40–75)
BMI, kg/m^2		
Mean (SD)	26.92 (4.32)	26.75 (5.59)
Median (range)	26.6 (15.7–48.8)	26.19 (14.53–57.72)
Male sex, n (%)	2489 (64.65)	1145 (67.27)
Current smokers, n (%)*	1984 (51.53)	659 (38.72)
FEV_1, % predicted		
Mean (SD)	75.62 (12.01)	47.57 (15.58)
Median (range)	76.8 (31.6–117.5)	46.30 (13.4–80.6)
FEV_1/FVC, %		
Mean (SD)	62.67 (8.01)	44.59 (11.51)
Median (range)	64.07 (25.27–79.73)	44 (17–75)
GOLD I–II, n (%)	3699 (96.08)	707 (41.54)
GOLD III–IV, n (%)	105 (2.73)	990 (58.17)
Time points for spirometry (years)	0, 1, 2, 3, 4, 5, 11	0, 0.25, 0.5, 1, 1.5, 2, 2.5, 3

*Smoking status determined from study data (see online supplementary methods). BMI, body mass index; CC-16, club cell secretory protein-16; FEV_1, forced expiratory volume in 1 s; FVC, forced vital capacity; GOLD I–II, FEV_1 ≥50% predicted; GOLD III–IV, FEV_1 <50% predicted; GWAS, genome-wide association study; LHS, Lung Health Study.

FEV_1 per unit increase in ln(ng/mL) CC-16 (p=0.004). The relationship between serum CC-16 and FEV_1 decline (slope model) is shown in online supplementary results, figure S1.

Multiple novel loci are associated with serum CC-16 level (pQTLs)

We included 7312348 SNPs in a total of 5552 individuals (n=3850 in LHS, n=1702 in ECLIPSE) in the GWAS. A total of seven pQTLs showed independent, genome-wide significant ($p<5\times10^{-8}$) association with serum CC-16 level in the meta-analysis (figure 2 (Manhattan plot) and figure 3 (gene region plots); table 2; online supplementary results, table S2). The quantile-quantile plot showed deviation from the expected distribution at low p values, with a genomic inflation factor (lambda) of 1.02 consistent with negligible confounding (online supplementary results, figure S4). Of the loci identified in this meta-analysis, only SNP rs3741240 on chromosome 11 (an intronic variant in the 5' untranslated region of *SCGB1A1*, the gene which encodes the CC-16 protein) had been previously identified in a GWAS for serum CC-16 (performed in the ECLIPSE cohort).[27] The remaining six loci are novel pQTLs for serum CC-16: SNP rs4971100 (chromosome 1) is an intronic variant in the tripartite motif containing 46 (*TRIM46*) gene; rs1515498 (chromosome 3) is an intronic variant in the tumour protein p63 (*TP63*) gene; rs37002 is in an intergenic region on the short arm of chromosome 5 (p15.3); rs11032840 is an intronic variant in an uncharacterised region on the short arm of chromosome 11; rs11231085 (chromosome 11) is an intronic variant in the *SCGB1A1* gene and rs7962469 (chromosome 12) is an intronic variant in the activin A receptor type 1B (*ACVR1B*) gene.

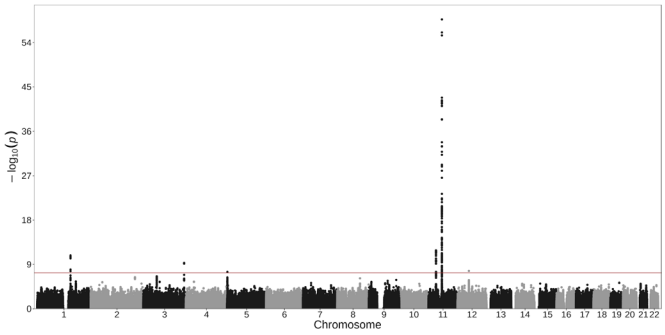

Figure 2 Manhattan plot for serum CC-16 GWAS. Meta-analysis of Lung Health Study and ECLIPSE study GWAS. GWAS p values ($-\log_{10}$ scale) (Y axis) versus single nucleotide polymorphism positions across 22 chromosomes (X axis). Horizontal red line represents the genome-wide significance cut-off of 5×10^{-8}. CC-16, club cell secretory protein-16; GWAS, genome-wide association study.

Serum CC-16 pQTLs are associated with risk of having COPD and its progression

The effect estimates for individual serum CC-16 pQTLs on COPD outcomes are shown in the online supplementary table S3. One pQTL (rs7962469 on chromosome 12) was significantly associated with 'COPD risk' in the ICGC dataset, translating to a 3% decrease in the odds of having COPD per CC-16-increasing allele (OR=0.97, $p=1.2\times10^{-3}$). Two pQTLs (rs3741240 and rs11231085, both on chromosome 11) were significantly associated with 'COPD progression' in the LHS, translating to 1.63 mL/year (p=0.047) and 2.89 mL/year ($p=5.4\times10^{-4}$) slower decline in FEV_1 per CC-16-increasing allele, respectively. Two pQTLs (rs3741240 and rs37002) were significantly associated with 'COPD progression' in ECLIPSE, translating to 6.96 mL/year slower (p=0.02) and 6.38 mL/year (p=0.04) faster decline in FEV_1 per CC-16-increasing allele, respectively.

Genetically determined serum CC-16 level is causally associated with risk of having COPD and its progression

Our primary analysis for determining a causal effect of serum CC-16 level on COPD outcomes was an IVW multivariable MR analysis using all seven pQTLs for CC-16 as IVs. The IVW MR estimate for 'COPD risk' in the ICGC dataset was significant (beta=–0.11, p=0.008), suggesting a protective effect of genetically increased serum CC-16 level on the risk of having COPD (table 3, figure 4A). The IVW MR estimate for 'COPD progression' in the LHS was statistically significant (beta=7.4, p=0.02), suggesting a protective effect of genetically increased serum CC-16 on FEV_1 decline (table 3, figure 4B). The IVW MR estimate for 'COPD progression' in the ECLIPSE study was in the same direction as in the LHS but it was not statistically significant (table 3; online supplementary results, figure S5).

We undertook a systematic approach to testing for violations of MR assumptions. A GWAS catalogue look-up found no genome-wide significant associations between the serum CC-16 pQTLs and potential confounders (online supplementary results, table S4). The SNP rs7962469 on chromosome 12 has a previously reported association with lung function ($p=8\times10^{-19}$),[28] but given that this is part of the definition of COPD, it was not considered a confounder. The partial F statistic for each SNP was

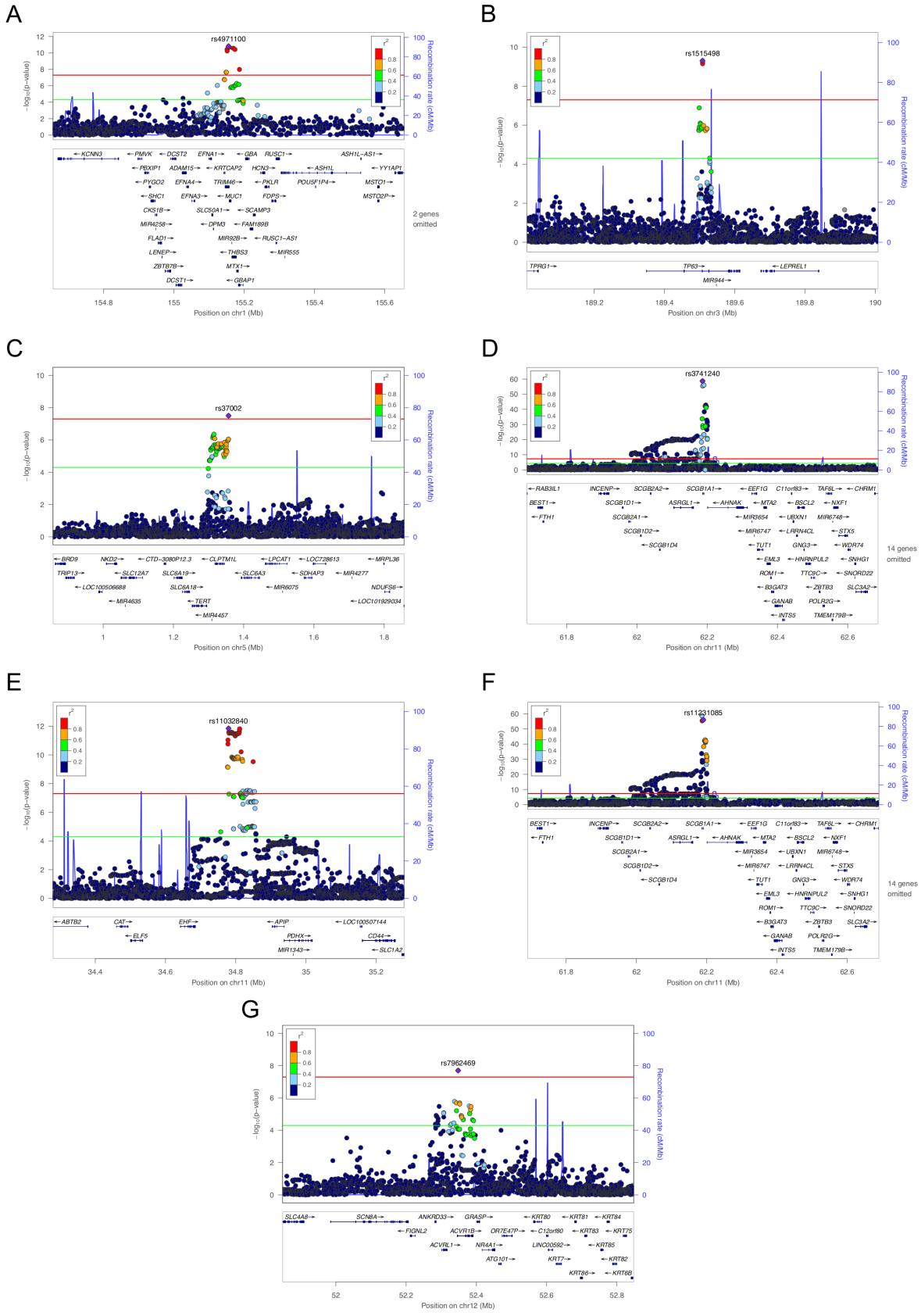

Figure 3 Gene region plots for serum CC-16 GWAS. Meta-analysis of Lung Health Study and ECLIPSE study GWAS. GWAS p values ($-\log_{10}$ scale) (Y axes) versus SNP genomic position (X axes) on (A) chromosome 1, (B) chromosome 3, (C) chromosome 5, (D–F) three independently associated regions on chromosome 11, and (G) chromosome 12. Horizontal red line represents the genome-wide significance cut-off of 5×10^{-8}. Horizontal green line represents p-value cut-off of 5×10^{-5}. Gene names and their coordinates are labelled. The sentinel (most significant) SNP for each region is annotated by rs identifier. r^2: linkage disequilibrium with the sentinel SNP. CC-16, club cell secretory protein-16; GWAS, genome-wide association study; SNP, single-nucleotide polymorphism.

Table 2 GWAS for serum CC-16 level

SNP rsID	Chr	Position†	Effect allele	Alt. allele	Nearby gene	LHS				ECLIPSE				Meta-analysis	
						Effect allele freq.	Effect on CC-16 (SE)‡	P value	Variance explained %	Effect allele freq.	Effect on CC-16 (SE)‡	P value	Variance explained %	Effect on CC-16 (SE)‡	P value
rs4971100	1	155 155 731	A	G	*TRIM46*	0.42	0.079 (0.014)	5.69×10^{-9}*	0.58%	0.41	0.062 (0.018)	5.66×10^{-4}	0.60%	0.073 (0.011)	1.55×10^{-11}*
rs1515498	3	189 508 302	A	G	*TP63*	0.64	0.074 (0.014)	1.38×10^{-7}	0.58%	0.66	0.061 (0.018)	8.62×10^{-4}	0.43%	0.069 (0.011)	4.94×10^{-10}*
rs37002	5	1 356 944	C	T	–	0.52	0.048 (0.014)	6.95×10^{-4}	0.28%	0.5	0.084 (0.018)	3.92×10^{-6}	1.14%	0.062 (0.011)	3.15×10^{-8}*
rs11032840	11	34 779 464	G	T	–	0.6	0.063 (0.014)	3.72×10^{-6}	0.38%	0.6	0.098 (0.018)	2.75×10^{-8}*	1.29%	0.076 (0.011)	1.44×10^{-12}*
rs3741240	11	62 186 542	G	A	*SCGB1A1*	0.65	0.178 (0.014)	1.1×10^{-37}*	3.95%	0.64	0.171 (0.017)	4.20×10^{-22}*	4.99%	0.176 (0.011)	2.08×10^{-59}*
rs11231085	11	62 190 448	G	C	*SCGB1A1*	0.35	0.182 (0.014)	1.94×10^{-39}*	3.66%	0.35	0.162 (0.019)	6.03×10^{-18}*	4.28%	0.175 (0.011)	9.13×10^{-57}*
rs7962469	12	52 348 259	A	G	*ACVR1B*	0.34	0.060 (0.015)	3.87×10^{-5}	0.38%	0.33	0.072 (0.019)	1.25×10^{-4}	0.47%	0.064 (0.011)	1.99×10^{-8}*

Separate GWAS for each cohort, combined by inverse variance weighted meta-analysis.
*P value less than genome-wide significant cut-off of 5×10^{-8}.
†Build hg19 of human reference genome.
‡ln(ng/mL) change in CC-16 per allele.
ACVR1B, activin receptor 1B; Chr, chromosome; ECLIPSE, Evaluation of COPD Longitudinally to Identify Predictive Surrogate Endpoints; GWAS, genome-wide association studies ; LHS, Lung Health Study; MAF, minor allele frequency; rsID, reference SNP cluster identifier; *SCGB1A1*, secretoglobin family 1A member 1; SNP, single nucleotide polymorphism; *TP63*, tumour protein p63; *TRIM46*, tripartite motif containing 46.

Table 3 Inverse variance weighted MR analysis for effect of genetically determined serum CC-16 level on COPD outcomes

Cohort	Outcome variable	Inverse-variance weighted MR	
		MR estimate β (SE)	Estimate p value
COPD risk			
ICGC dataset	COPD (yes/no)	–0.11 (0.04)	0.008*
COPD progression			
LHS	Change in FEV_1 (mL/year)	7.40 (3.28)	0.02*
ECLIPSE	Change in FEV_1 (mL/year)	19.14 (12.59)	0.13

*P<0.05.
COPD, chronic obstructive pulmonary disease; FEV_1, forced expiratory volume in 1 s; ICGC, International COPD Genetics Consortium; LHS, Lung Health Study; MR, Mendelian randomisation.

Figure 4 MR plots. Inverse variance weighted regression model, adjusted for linkage disequilibrium between SNPs, intercept constrained to zero. The model relates the per-allele effects of the SNPs on serum CC-16 level in the GWAS meta-analysis (X axes) to their per-allele effects on (A) risk of having COPD in the ICGC dataset, and (B) change in FEV_1 over time in the LHS (Y axes). Red line represents the estimated effect. Error bars represent 95% CIs. Individual SNPs are annotated by their rs identifier. CC-16, club cell secretory protein-16; COPD, chronic obstructive pulmonary disease; FEV_1, forced expiratory volume in 1 s; ICGC, 2019 International COPD Genetics Consortium; LHS, Lung Health Study; MR, Mendelian randomisation; SNP, single-nucleotide polymorphisms.

Table 4 Tests of MR assumptions

		Heterogeneity test	Weighted median MR		MR-Egger			MR-PRESSO
Cohort	Outcome variable	Cochran's Q p value	Estimate β (SE)	Estimate p value	Estimate β (SE)	Estimate p	Intercept p	Global test p
COPD risk								
ICGC dataset	COPD (yes/no)	0.16	−0.09 (0.04)	0.033*	−0.03 (0.11)	0.75	0.45	0.29
COPD progression								
LHS	Change in FEV_1 (mL/year)	0.05	11.12 (3.82)	0.004*	22.18 (8.50)	0.009*	0.05	0.08
ECLIPSE	Change in FEV_1 (mL/year)	0.21	33.48 (13.29)	0.012*	68.57 (29.54)	0.02*	0.06	0.27

See main text for methods and interpretation of each test.
*P<0.05.
COPD, chronic obstructive pulmonary disease; FEV_1, forced expiratory volume in 1 s; ICGC, International COPD Genetics Consortium; LHS, Lung Health Study; MR, Mendelian randomisation.

>10 in both LHS and ECLIPSE (online supplementary results, table S5), which suggested that these genetic variants did not contribute to significant weak instrument bias.

Cochran's Q test was not significant for any of the analyses, suggesting that there was no significant heterogeneity when all seven SNPs were included as IVs (table 4). Leave-one-out sensitivity analyses (online supplementary results, figure S6) suggested that the average causal effect of serum CC-16 is driven largely by the strongest pQTLs (rs3741240 and rs11231085, both nearby the *SCGB1A1* gene on chromosome 11). Interestingly, when the SNP rs37002 on chromosome 5 was left out, the IVW MR estimate for 'COPD progression' in ECLIPSE became significant. This SNP has a statistically significant and relatively large effect on FEV_1 decline in this cohort, but in the opposite direction to the other SNPs (online supplementary results, table S3).

Finally, we performed additional MR variations that are robust to violations of MR assumptions. MR-median tests were significant for all three analyses (table 4), suggesting that the majority of IVs used were valid instruments. The MR-Egger estimate was significant for 'COPD progression' in both LHS (beta=22.18, p=0.009) and ECLIPSE (beta=68.57, p=0.02); for all three outcome analyses, the intercept terms were not significant (MR-Egger plots are shown in online supplementary results, figures S7–S9). The MR-PRESSO 'global' tests were not significant for any of the outcome analyses (table 4). The MR-Egger and MR-PRESSO results together suggest no major directional pleiotropy in the analyses.

Multiple CC-16 pQTLs affect lung tissue gene expression

Four of the pQTLs for serum CC-16 level were also lung eQTLs (table 5). SNPs rs3741240 and rs11231085 on chromosome 11 were associated with levels of *SCGB1A1* mRNA ($p=4.13\times10^{-12}$ and $p=3.82\times10^{-7}$, respectively), with allele effects in the same direction as their effects on serum CC-16 level. SNP rs4971100 (chromosome 1) was associated with mRNA levels for the glucosylceramidase beta (*GBA*) gene ($p=4.58\times10^{-8}$) in lung tissue. SNP rs7962469 (chromosome 12) was associated with *ACVR1B* gene expression ($p=3.31\times10^{-116}$) with the allele effect in the same direction as its effect on serum CC-16 level.

DISCUSSION

By performing the largest GWAS-pQTL analysis for serum-CC16 levels to date (n=5552), we have identified one known and six novel loci associated with serum CC-16 level. Furthermore, using the MR framework, we have demonstrated a potential protective effect of genetically increased serum CC-16 level on both the risk of having COPD and on COPD progression (measured as change in FEV_1 over time). Our findings expand on the existing body of literature on the biology of CC-16 and support a possible causal role for CC-16 in COPD pathogenesis.

We identified genetic variants independently associated with serum CC-16 level (pQTLs) in six different loci across five chromosomes. The most significantly associated SNP was rs3741240 ($p=2.08\times10^{-59}$ in the GWAS meta-analysis), which is in a non-coding region of the CC-16 gene *SCGB1A1* on chromosome 11. This SNP was also the top association in the only previously published GWAS for serum CC-16 level.[27] In the same gene region, we have identified a second, independently associated SNP rs11231085 by conditional analysis ($p=9.13\times10^{-57}$ in the GWAS meta-analysis). These results reflect the increased statistical power of GWAS meta-analysis to detect associated genetic variants.

To our knowledge, this is the first study to demonstrate that variants in the *SCGB1A1* gene are associated with lung function decline. The SNP rs3741240 has been previously associated with both airway hyperresponsiveness[29] and asthma,[30] but was

Table 5 Lung tissue eQTL analysis

SNP rsID	Chr	Position*	Effect allele	Alternate allele	Probe set	Gene	Allele effect on gene expression†	Allele effect p
rs4971100	1	155 155 731	A	G	100300160_TGI_at	*GBA*	−0.23	4.58×10^{-8}
rs37002	5	1 356 944	C	T	100307694_TGI_at	Unannotated	0.93	7.47×10^{-161}
rs3741240	11	62 186 542	G	A	100143118_TGI_at	*SCGB1A1*	0.31	4.13×10^{-12}
rs11231085	11	62 190 448	G	C	100143118_TGI_at	*SCGB1A1*	0.22	3.82×10^{-7}
rs7962469	12	52 348 259	A	G	100127400_TGI_at	*ACVR1B*	0.82	3.31×10^{-116}

Expression data from the Lung eQTL Study.
*Build hg19 of human reference genome.
†Change in gene expression level per effect allele.
ACVR1B, activin receptor 1B; Chr, chromosome; eQTL, expression quantitative trait loci; *GBA*, glucosylceramidase beta; rsID, reference SNP cluster identifier; *SCGB1A1*, secretoglobin family 1A member 1; SNP, single nucleotide polymorphism.

not associated with COPD outcomes in the previously published ECLIPSE GWAS.[27] The SNP rs11231085 has not previously been associated with lung function or lung disease. Our lung tissue eQTL analysis suggests that genetically increased serum CC-16 levels may be due to increased lung tissue expression of the *SCGB1A1* gene. The precise mechanisms by which *SCGB1A1* gene expression and CC-16 protein production support lung function are unclear.

The remaining genetic variants we found to be associated with serum CC-16 levels were in loci distal to the *SCGB1A1* gene, and each of these associations with CC-16 was novel. SNP rs7962469 on chromosome 12 has been previously associated with risk of lung disease. This SNP is an intronic variant in the *ACVR1B* gene, which codes for a member of the transforming growth factor-beta (TGF-β) receptor superfamily also known as activin-like kinase-4 (ALK4). We found this variant was also associated with lung tissue expression of *ACVR1B* (ie, it is an eQTL), although how expression of the ALK4 receptor is related to serum levels of CC-16 remains unclear. Members of the TGF-β superfamily, including ALK4, are potent regulators of gene transcription, and the downstream *Smad* signalling pathway has been shown to be dysregulated in COPD.[31] The association between *ACVR1B* variants and serum CC-16 level may therefore be due to modulation of *SCGB1A1* transcription by ALK4 receptor activity. Additionally, ALK4 and *Smad* signalling are involved in the differentiation and proliferation of lung cells including airway epithelial cells.[31] It is therefore possible that *ACVR1B* variants relate to CC-16 levels due to altered numbers of differentiated club cells secreting the CC-16 protein. Interestingly, variants in or near the *ACVR1B* gene region have been associated with lung function[28] and the risk of having COPD.[32 33] The rs7962469 variant, through its effects on *ACVR1B* expression in lung tissue, has also been causally associated with emphysema distribution on chest imaging in the COPDGene cohort.[32] The link between this gene and CC-16 biology is intriguing and warrants further investigation.

The links between the other novel SNPs and serum CC-16 level are less clear. SNP rs1515498 on chromosome 3 is an intronic variant in the *TP63* gene, which itself is related to the risk of lung cancer[34 35] but has also recently been associated with the risk of having COPD in a Taiwanese population.[36] SNP rs4971100 on chromosome 1 is an eQTL for the *GBA* gene in lung tissue, which encodes a cell membrane protein involved in lipid processing. Abnormalities in the *GBA* gene are associated with the lysosomal storage disorder Gaucher's disease[37] as well as the neurodegenerative disorder Parkinson disease.[38] These distal gene associations highlight the complexity of CC-16 biology.

The association between serum CC-16 levels and lung function is well documented in observational and epidemiological studies. CC-16 may be critical to lung development.[39] Lower serum CC-16 levels have also been associated with accelerated FEV_1 decline in early adulthood[40] and increased risk of incident COPD.[39] Despite the consistency of these observations, confounding through reverse causality and environmental exposures remained a possibility. Since MR analysis quantifies only the genomic contribution to serum CC-16 level, it is more resistant to confounding by these factors and argues for a unidirectional association. This is particularly important in light of the disparate effects of cigarette smoke: cumulative chronic exposure is strongly associated with reduced serum CC-16 concentration;[41] yet acute smoke exposure causes a transient increase in serum and commensurate decrease in BALF CC-16 concentration, possibly by disrupting epithelial cell integrity and/or increased vascular permeability.[42] CC-16 appears to have a protective effect on the airway epithelium,[2–4] most likely through regulation of inflammation via the nuclear factor κB pathway.[3] Genetic and environmental factors that reduce the availability of CC-16 in the lung may therefore increase susceptibility to inflammatory insults, with resultant airway and/or parenchymal changes leading to accelerated lung function decline.

Through MR analysis, we found a significant effect of serum CC-16 level on COPD progression in the LHS. However, we did not find a significant effect in the ECLIPSE study. Leave-one-out sensitivity analysis suggested this was due to the influence of the SNP rs37002 on chromosome 5, which had a relatively large effect on FEV_1 change in ECLIPSE but in the opposite direction to the other CC-16-increasing alleles, differences in the study populations (LHS, younger cohort with mild-moderate COPD, larger sample size and longer follow-up; ECLIPSE, older cohort with more severe COPD, smaller sample size and shorter follow-up) may also contribute. Additionally, we did not find a significant relationship between serum CC-16 level and change in FEV_1 in the ECLIPSE study alone, which is in contrast to previous analyses.[9] This may be due to differences in the models used to determine change in FEV_1. Nevertheless, the statistically non-significant trend was in the same direction as previously reported effects of serum CC-16 level on change in FEV_1.[9 10] Further studies in larger, independent cohorts will be needed to confirm the causal effect of serum CC-16 level on change in FEV_1.

Although our primary MR analysis suggested a causal effect of serum CC-16 level on COPD outcomes, we took steps to ensure that the causal inferences do not violate the fundamental assumptions of MR.[43] These steps included: (1) ensuring that pQTLs for serum CC-16 were independent of each other by accounting for linkage disequilibrium structure; (2) investigating the strength of each instrument to minimise the potential for weak instrument bias; (3) minimising the influence of confounders by adjusting for relevant covariates and searching a GWAS catalogue for known SNP-trait associations; (4) testing for the influence of outliers using weighted median MR and Cochran's Q test for heterogeneity (the presence of which would suggest that variability in the SNP associations is greater than would be expected by chance alone and thus may indicate bias) and (5) testing for evidence of directional pleiotropy that is, one or more of the IVs is associated with the outcome via a risk factor other than serum CC-16, through MR-Egger and MR-PRESSO analyses. On balance, the outcomes of this strategy suggested no major violation of the MR assumptions.

Our results complement previous in vitro and animal studies investigating a causal role for CC-16 in COPD pathogenesis. Among the most convincing evidence is the study by Laucho-Contreras and colleagues[3] in which CC-16-/- knockout mice showed signs of accelerated emphysema-like lung changes, airway remodelling, alveolar apoptosis and increased lung inflammation following cigarette smoke exposure. These effects were attenuated by transfection with a CC-16-producing adenoviral vector. However, these changes were not observed in a similar study by Park and colleagues, despite the use of an identical strain of mouse and a similar experimental design.[10] The reasons for this discrepancy are unclear. Our results therefore contribute important evidence from human studies that CC-16 is implicated in the development and progression of COPD.

CONCLUSION

Using GWAS and the MR framework, we have provided evidence for a potential causal effect of serum CC-16 level in COPD pathogenesis and lung function decline. These results complement previous attempts to establish a causal role for CC-16 in

animal models. Further investigation of serum CC-16 and associated pathways as potential biomarkers or therapeutic targets in COPD is warranted.

Author affiliations
[1]Centre for Heart Lung Innovation, St Paul's Hospital, The University of British Columbia, Vancouver, British Columbia, Canada
[2]Division of Respiratory Medicine, The University of British Columbia, Vancouver, British Columbia, Canada
[3]Faculty of Medicine and Health, The University of Sydney, Sydney, New South Wales, Australia
[4]Channing Division of Network Medicine and Division of Pulmonary and Critical Care Medicine, Brigham and Women's Hospital, Boston, Massachusetts, USA
[5]Department of Epidemiology, Bloomberg School of Public Health, Johns Hopkins University, Baltimore, Maryland, USA
[6]Department of Biostatistics, Bloomberg School of Public Health, Johns Hopkins University, Baltimore, Maryland, USA
[7]Pulmonary and Critical Care Medicine, School of Medicine, Johns Hopkins University, Baltimore, Maryland, USA
[8]Institut universitaire de cardiologie et de pneumologie de Québec, Department of Molecular Medicine, Laval University, Québec City, Québec, Canada
[9]University of Groningen Department of Pathology and Medical Biology, University Medical Centre Groningen, Groningen, The Netherlands

Acknowledgements This research has been conducted using the UK Biobank Resource. We thank the International COPD Genetics Consortium investigators for providing genetic association summary statistics.

Contributors SM and XL devised the study, analysed the data, wrote and revised the draft manuscripts and contributed equally to this work. AIHC, CXY and MC contributed to the data analysis pipeline and interpretation of results and revised the draft manuscripts. TB, IR, NNH, YB, C-AB and DDS contributed to data collection and revised the draft manuscripts. MO contributed to study design, the data analysis pipeline and the interpretation of results and revised the draft manuscript. All authors approved the final manuscript. MO serves as senior author with full access to the data and responsibility for data integrity.

Funding There were no direct financial sponsors for the submitted work. SM and AIHC are supported by the MITACS Accelerate program. MHC is supported by R01 R01 HL135142, R01 HL137927, R01 HL089856, R01 HL147148 and R01HL133135.

Disclaimer The content is solely the responsibility of the authors and does not necessarily represent the official views of the NIH. The funding body has no role in the design of the study and collection, analysis, and interpretation of data and in writing the manuscript.

Competing interests SM reports personal fees from Novartis, Boehringer Ingelheim and Menarini, and non-financial support from Draeger Australia, outside the submitted work. MHC has received grant support from GSK and Bayer and consulting fees from Genentech. DDS reports grants from Merck, personal fees from Sanofi-Aventis, Regeneron and Novartis, and grants and personal fees from Boehringer Ingelheim and AstraZeneca, outside the submitted work. MO is currently employed by Novartis Pharmaceuticals. YB holds a Canada Research Chair in Genomics of Heart and Lung Diseases. DDS holds the De Lazzari Family Chair at HLI and a Tier 1 Canada Research Chair in COPD. MO is a Fellow of the Parker B Francis Foundation and a Scholar of the Michael Smith Foundation for Health Research (MSFHR).

Patient consent for publication Not required.

Provenance and peer review Not commissioned; externally peer reviewed.

Data availability statement Data are available on reasonable request. Summary statistics from ICGC/UK Biobank GWAS meta-analysis are available through the Database of Genotypes and Phenotypes (dbGaP) repository (accession no. phs000179.v5.p2) and from the UK Biobank. Summary GWAS data from Lung Health Study (accession no. phs000335.v3.p2) and ECLIPSE (accession no. phs001252.v1.p1) are available through dbGaP. Lung eQTL Study data are available through dbGap (accession no. phs001745.v1.p1) and by direct application to the Lung eQTL Consortium.

ORCID iDs
Stephen Milne http://orcid.org/0000-0001-5274-1298
Michael H Cho http://orcid.org/0000-0002-4907-1657
Corry-Anke Brandsma http://orcid.org/0000-0001-8911-3658

REFERENCES

1 Mathers CD, Loncar D. Projections of global mortality and burden of disease from 2002 to 2030. *PLoS Med* 2006;3:e442.
2 Pang M, Liu H-Y, Li T, *et al*. Recombinant Club cell protein 16 (CC16) ameliorates cigarette smoke-induced lung inflammation in a murine disease model of COPD. *Mol Med Rep* 2018;18:2198–206.
3 Laucho-Contreras ME, Polverino F, Gupta K, *et al*. Protective role for Club cell secretory protein-16 (CC16) in the development of COPD. *Eur Respir J* 2015;45:1544–56.
4 Hong KU, Reynolds SD, Giangreco A, *et al*. Clara cell secretory protein-expressing cells of the airway neuroepithelial body microenvironment include a label-retaining subset and are critical for epithelial renewal after progenitor cell depletion. *Am J Respir Cell Mol Biol* 2001;24:671–81.
5 Pilette C, Godding V, Kiss R, *et al*. Reduced epithelial expression of secretory component in small airways correlates with airflow obstruction in chronic obstructive pulmonary disease. *Am J Respir Crit Care Med* 2001;163:185–94.
6 Bernard A, Marchandise FX, Depelchin S, *et al*. Clara cell protein in serum and bronchoalveolar lavage. *Eur Respir J* 1992;5:1231–8.
7 Braido F, Riccio AM, Guerra L, *et al*. Clara cell 16 protein in COPD sputum: a marker of small airways damage? *Respir Med* 2007;101:2119–24.
8 Lomas DA, Silverman EK, Edwards LD, *et al*. Evaluation of serum CC-16 as a biomarker for COPD in the ECLIPSE cohort. *Thorax* 2008;63:1058–63.
9 Vestbo J, Edwards LD, Scanlon PD, *et al*. Changes in forced expiratory volume in 1 second over time in COPD. *N Engl J Med* 2011;365:1184–92.
10 Park HY, Churg A, Wright JL, *et al*. Club cell protein 16 and disease progression in chronic obstructive pulmonary disease. *Am J Respir Crit Care Med* 2013;188:1413–9.
11 Smith GD, Ebrahim S. 'Mendelian randomization': can genetic epidemiology contribute to understanding environmental determinants of disease? *Int J Epidemiol* 2003;32:1–22.
12 Obeidat Ma'en, Li X, Burgess S, *et al*. Surfactant protein D is a causal risk factor for COPD: results of Mendelian randomisation. *Eur Respir J* 2017;50:1700657.
13 Dahl M, Vestbo J, Zacho J, *et al*. C reactive protein and chronic obstructive pulmonary disease: a Mendelian randomisation approach. *Thorax* 2011;66:197–204.
14 van Durme YMTA, Lahousse L, Verhamme KMC, *et al*. Mendelian randomization study of interleukin-6 in chronic obstructive pulmonary disease. *Respiration* 2011;82:530–8.
15 Amini M, Vonk JM, Abbasi A, *et al*. Blood eosinophil count and metabolic, cardiac and pulmonary outcomes: a Mendelian randomization study. *Twin Res Hum Genet* 2018;21:89–100.
16 Connett JE, Kusek JW, Bailey WC, *et al*. Design of the lung health study: a randomized clinical trial of early intervention for chronic obstructive pulmonary disease. *Control Clin Trials* 1993;14:3–19.
17 Vestbo J, Anderson W, Coxson HO, *et al*. Evaluation of COPD longitudinally to identify predictive surrogate end-points (ECLIPSE). *Eur Respir J* 2008;31:869–73.
18 Sakornsakolpat P, Prokopenko D, Lamontagne M, *et al*. Genetic landscape of chronic obstructive pulmonary disease identifies heterogeneous cell-type and phenotype associations. *Nat Genet* 2019;51:494–505.
19 Hao K, Bossé Y, Nickle DC, *et al*. Lung eQTLs to help reveal the molecular underpinnings of asthma. *PLoS Genet* 2012;8:e1003029.
20 Yang J, Ferreira T, Morris AP, *et al*. Conditional and joint multiple-SNP analysis of GWAS summary statistics identifies additional variants influencing complex traits. *Nat Genet* 2012;44:369–75. S1–3.
21 Davies NM, Holmes MV, Davey Smith G. Reading Mendelian randomisation studies: a guide, glossary, and checklist for clinicians. *BMJ* 2018;362:k601.
22 Yavorska OO, Burgess S. MendelianRandomization: an R package for performing Mendelian randomization analyses using summarized data. *Int J Epidemiol* 2017;46:1734–9.
23 Yavorska OO, Burgess S. MendelianRandomization: Mendelian Randomization Package version 0.2.2 [online software package]. Available: https://cran.r-project.org/web/packages/MendelianRandomization/ [Accessed 31 Oct 2019].
24 Burgess S, Davey Smith G, Davies N, *et al*. Guidelines for performing Mendelian randomization investigations [version 2; peer review: 1 approved, 1 approved with reservations]. *Wellcome Open Research* 2020;4.
25 R Core Team. *R: A language and environment for statistical computing [program]*. Vienna, Austria: R Foundation for Statistical Computing, 2017.
26 Hermans C, Dong P, Robin M, *et al*. Determinants of serum levels of surfactant proteins A and B and Clara cell protein CC16. *Biomarkers* 2003;8:461–71.
27 Kim DK, Cho MH, Hersh CP, *et al*. Genome-wide association analysis of blood biomarkers in chronic obstructive pulmonary disease. *Am J Respir Crit Care Med* 2012;186:1238–47.
28 Kichaev G, Bhatia G, Loh P-R, *et al*. Leveraging polygenic functional enrichment to improve GWAS power. *Am J Hum Genet* 2019;104:65–75.
29 Taniguchi N, Konno S, Hattori T, *et al*. The CC16 A38G polymorphism is associated with asymptomatic airway hyper-responsiveness and development of late-onset asthma. *Ann Allergy Asthma Immunol* 2013;111:376–81.
30 Laing IA, Goldblatt J, Eber E, *et al*. A polymorphism of the CC16 gene is associated with an increased risk of asthma. *J Med Genet* 1998;35:463–7.
31 Verhamme FM, Bracke KR, Joos GF, *et al*. Transforming growth factor-β superfamily in obstructive lung diseases. more suspects than TGF-β alone. *Am J Respir Cell Mol Biol* 2015;52:653–62.

32 Boueiz A, Pham B, Chase R, *et al*. Integrative genomics analysis identifies ACVR1B as a candidate causal gene of emphysema distribution. *Am J Respir Cell Mol Biol* 2019;60:388–98.
33 Morrow JD, Cho MH, Platig J, *et al*. Ensemble genomic analysis in human lung tissue identifies novel genes for chronic obstructive pulmonary disease. *Hum Genomics* 2018;12:1.
34 Hosgood HD, Wang W-C, Hong Y-C, *et al*. Genetic variant in TP63 on locus 3q28 is associated with risk of lung adenocarcinoma among never-smoking females in Asia. *Hum Genet* 2012;131:1197–203.
35 Miki D, Kubo M, Takahashi A, *et al*. Variation in TP63 is associated with lung adenocarcinoma susceptibility in Japanese and Korean populations. *Nat Genet* 2010;42:893–6.
36 Huang H-C, Lin FC-F, Wu M-F, *et al*. Association between chronic obstructive pulmonary disease and PM2.5 in Taiwanese nonsmokers. *Int J Hyg Environ Health* 2019;222:884–8.
37 Hruska KS, LaMarca ME, Scott CR, *et al*. Gaucher disease: mutation and polymorphism spectrum in the glucocerebrosidase gene (GBA). *Hum Mutat* 2008;29:567–83.
38 Sidransky E, Lopez G. The link between the GBA gene and parkinsonism. *Lancet Neurol* 2012;11:986–98.
39 Guerra S, Halonen M, Vasquez MM, *et al*. Relation between circulating CC16 concentrations, lung function, and development of chronic obstructive pulmonary disease across the lifespan: a prospective study. *Lancet Respir Med* 2015;3:613–20.
40 Zhai J, Insel M, Addison KJ, *et al*. Club cell secretory protein deficiency leads to altered lung function. *Am J Respir Crit Care Med* 2019;199:302–12.
41 Robin M, Dong P, Hermans C, *et al*. Serum levels of CC16, SP-A and SP-B reflect tobacco-smoke exposure in asymptomatic subjects. *Eur Respir J* 2002;20:1152–61.
42 Van Miert E, Dumont X, Bernard A. CC16 as a marker of lung epithelial hyperpermeability in an acute model of rats exposed to mainstream cigarette smoke. *Toxicol Lett* 2005;159:115–23.
43 Burgess S, Bowden J, Fall T, *et al*. Sensitivity analyses for robust causal inference from Mendelian randomization analyses with multiple genetic variants. *Epidemiology* 2017;28:30–42.