



# Surfactant protein D is a causal risk factor for COPD: results of Mendelian randomisation

Ma'en Obeidat <sup>1</sup>, Xuan Li<sup>1</sup>, Stephen Burgess<sup>2,3</sup>, Guohai Zhou<sup>1</sup>, Nick Fishbane<sup>1</sup>, Nadia N. Hansel<sup>4</sup>, Yohan Bossé <sup>5,6</sup>, Philippe Joubert<sup>5</sup>, Ke Hao<sup>7</sup>, David C. Nickle<sup>8</sup>, Maarten van den Berge<sup>9</sup>, Wim Timens<sup>10</sup>, Michael H. Cho<sup>11,12</sup>, Brian D. Hobbs <sup>11,12</sup>, Kim de Jong<sup>13</sup>, Marike Boezen<sup>13</sup>, Rayjean J. Hung<sup>14</sup>, Nicholas Rafaels<sup>15</sup>, Rasika Mathias<sup>16</sup>, Ingo Ruczinski<sup>17</sup>, Terri H. Beaty<sup>18</sup>, Kathleen C. Barnes<sup>15</sup>, Peter D. Paré<sup>1</sup>, Don D. Sin<sup>1</sup> and International COPD Genetics Consortium, Lung eQTL Consortium, Lung Health Study

@ERSpublications  
**Surfactant protein D is a causal risk factor for COPD** <http://ow.ly/n1OG30eUQlf>

**Cite this article as:** Obeidat M, 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 [<https://doi.org/10.1183/13993003.00657-2017>].

**ABSTRACT** Surfactant protein D (SP-D) is produced primarily in the lung and is involved in regulating pulmonary surfactants, lipid homeostasis and innate immunity. Circulating SP-D levels in blood are associated with chronic obstructive pulmonary disease (COPD), although causality remains elusive.

In 4061 subjects with COPD, we identified genetic variants associated with serum SP-D levels. We then determined whether these variants affected lung tissue gene expression in 1037 individuals. A Mendelian randomisation framework was then applied, whereby serum SP-D-associated variants were tested for association with COPD risk in 11 157 cases and 36 699 controls and with 11 years decline of lung function in the 4061 individuals.

Three regions on chromosomes 6 (human leukocyte antigen region), 10 (*SFTPD* gene) and 16 (*ATP2C2* gene) were associated with serum SP-D levels at genome-wide significance. In Mendelian randomisation analyses, variants associated with increased serum SP-D levels decreased the risk of COPD (estimate  $-0.19$ ,  $p=6.46 \times 10^{-3}$ ) and slowed the lung function decline (estimate  $=0.0038$ ,  $p=7.68 \times 10^{-3}$ ).

Leveraging genetic variation effect on protein, lung gene expression and disease phenotypes provided novel insights into SP-D biology and established a causal link between increased SP-D levels and protection against COPD risk and progression. SP-D represents a very promising biomarker and therapeutic target for COPD.

---

This article has supplementary material available from [erj.ersjournals.com](http://erj.ersjournals.com)

Received: March 28 2017 | Accepted after revision: Aug 22 2017

Support statement: M. Obeidat is a Postdoctoral Fellow of the Parker B. Francis Foundation. He is also a recipient of British Columbia Lung Association Research Grant. K. Hao is partially supported by National Natural Science Foundation of China grant no. 21477087. M.H. Cho is recipient of National Institutes of Health grant R01 HL113264, and he also received grant support from GlaxoSmithKline. D. Sin is a tier 1 Canada Research Chair in COPD.

Conflict of interest: Disclosures can be found alongside this article at [erj.ersjournals.com](http://erj.ersjournals.com)

Copyright ©ERS 2017

**Affiliations:** <sup>1</sup>The University of British Columbia Center for Heart Lung Innovation, St Paul's Hospital Vancouver, Vancouver, BC, Canada. <sup>2</sup>Dept of Public Health and Primary Care, University of Cambridge, Cambridge, UK. <sup>3</sup>MRC Biostatistics Unit, University of Cambridge, Cambridge, UK. <sup>4</sup>Pulmonary and Critical Care Medicine, School of Medicine, Johns Hopkins University, Baltimore, MD, USA. <sup>5</sup>Institut Universitaire de Cardiologie et de Pneumologie de Québec, Laval University, Québec, QC, Canada. <sup>6</sup>Dept of Molecular Medicine, Laval University, Québec, QC, Canada. <sup>7</sup>Dept of Genetics and Genomic Sciences, Icahn School of Medicine at Mount Sinai, New York, NY, USA. <sup>8</sup>Merck Research Laboratories, Boston, MA, USA. <sup>9</sup>University of Groningen, University Medical Center Groningen, Dept of Pulmonology, GRIAC Research Institute, Groningen, The Netherlands. <sup>10</sup>University of Groningen, University Medical Center Groningen, Dept of Pathology and Medical Biology, GRIAC Research Institute, Groningen, The Netherlands. <sup>11</sup>Channing Division of Network Medicine, Brigham and Women's Hospital, Boston, MA, USA. <sup>12</sup>Division of Pulmonary and Critical Care Medicine, Brigham and Women's Hospital, Boston, MA, USA. <sup>13</sup>University of Groningen, University Medical Center Groningen, Dept of Epidemiology, GRIAC Research Institute, Groningen, The Netherlands. <sup>14</sup>Lunenfeld-Tanenbaum Research Institute, Sinai Health System, Toronto, ON, Canada. <sup>15</sup>Division of Biomedical Informatics and Personalized Medicine, Dept of Medicine, University of Colorado School of Medicine, Anschutz Medical Campus, Aurora, CO, USA. <sup>16</sup>Division of Genetic Epidemiology, School of Medicine, Johns Hopkins University, Baltimore, MD, USA. <sup>17</sup>Dept of Biostatistics, Bloomberg School of Public Health, Johns Hopkins University, Baltimore, MD, USA. <sup>18</sup>Dept of Epidemiology, Bloomberg School of Public Health, Johns Hopkins University, Baltimore, MD, USA.

**Correspondence:** Don D. Sin, UBC Centre for Heart Lung Innovation, St Paul's Hospital, 1081 Burrard Street, Vancouver, BC V6Z 1Y6, Canada. E-mail: don.sin@hli.ubc.ca

## Introduction

Chronic obstructive pulmonary disease (COPD) is currently the third leading cause of death worldwide [1]. The disease is characterised by airway obstruction that is not fully reversible, and is caused by a complex interaction between genetic and environmental risk factors (e.g. smoking) [2]. Although one-third of the variability in lung function can be attributed to genetic factors, the genes that directly control the risk of COPD remain largely unknown [3].

A critical class of molecules involved in lung growth and homeostasis are surfactants. One surfactant-associated protein that shows correlation with COPD, even in the absence of cigarette smoke, is surfactant protein D (SP-D) [4]. SP-D is a large multimeric collagenous glycoprotein weighing ~43 kDa and is part of the collectin family of proteins [5]. SP-D is produced primarily by type II alveolar cells, as well as by club cells [6, 7]. SP-D plays an important role in regulating pulmonary surfactants, maintaining lipid homeostasis in the lung [8] and promoting innate immunity [9] to protect the lungs from microbial and chemical insults [10]. SP-D deficiency in mice leads to alveolar macrophage activation, increased oxidant stress in the airways and emphysematous changes in lung parenchyma by 3 weeks of age [11]. In humans, single-nucleotide polymorphisms (SNPs) in the surfactant protein D (*SFTPD*) gene on chromosome 10 have been associated with COPD in several different cohorts [12]. The most-studied *SFTPD* variant (rs721917) is a coding SNP, which results in the substitution of methionine for threonine at amino acid 11 (Met11Thr) [13]. This SNP is associated with SP-D serum levels and emphysema [14]. COPD is associated with reduced levels of SP-D in bronchial and bronchoalveolar lavage (BAL) fluids [15], as well as increased serum SP-D levels [16]. Despite these associations it remains uncertain whether SP-D is a biomarker of COPD or is part of the causal pathway leading to COPD.

One way of evaluating causality in biomarker-disease associations is to use Mendelian randomisation (MR), which utilises genetic variation to ascertain causal associations [17–19]. MR exploits two unique attributes of genotypes enabling robust causal interrogation: 1) the random allocation of parental alleles to zygotes at meiosis, independent of environmental exposures later in life and 2) the unidirectional flow of biological information from gene to transcript to protein, thus avoiding reverse causation [20]. Within the MR framework, if the relationship between SP-D and COPD is truly causal, the genetic variant(s) influencing serum SP-D levels should also be significantly associated with COPD risk and/or forced expiratory volume in 1 s (FEV<sub>1</sub>) decline over time.

In the present study, we unravelled the genetic architecture of SP-D serum levels in a large cohort of COPD patients to determine SP-D protein quantitative trait loci (pQTLs) and mRNA expression quantitative trait loci (eQTLs) and used MR to demonstrate a causal relationship between serum SP-D and two important phenotypes: COPD risk and FEV<sub>1</sub> decline over time.

## Methods

### Overall study design

The overall study design is depicted in figure 1. First, we identified pQTLs for SP-D using a genome-wide association study (GWAS) of serum SP-D levels in COPD patients, who participated in the Lung Health

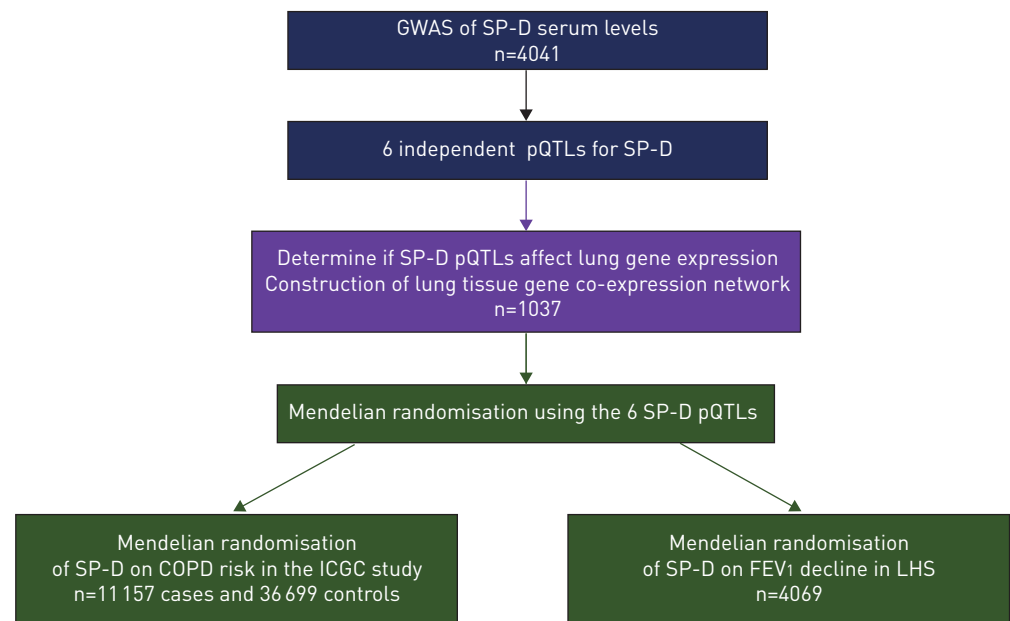


FIGURE 1 Study design. GWAS: genome-wide association study; SP-D: serum surfactant protein D; pQTL: protein quantitative trait loci; COPD: chronic obstructive pulmonary disease; ICGC: International COPD Genetics Consortium; FEV<sub>1</sub>: forced expiratory volume in 1 s; LHS: Lung Health Study.

Study (LHS) [21]. Second, using data from the Lung eQTL Consortium [22], we determined which of these pQTLs were also associated with changes in gene expression (eQTLs) in lung tissue. In addition, we used gene expression data from lung tissue to identify networks of genes strongly related to *SFTPD*. Third, we determined which of the SP-D pQTLs were associated with risk to COPD and COPD progression defined by the rate of decline in FEV<sub>1</sub> over 11 years. Finally, we applied a MR framework to determine whether there was a causal relationship between serum SP-D levels and COPD risk or progression.

### The Lung Health Study

The details of the LHS have been published previously [21, 23]. Briefly, LHS was a multicentre clinical study that evaluated the effects of ipratropium bromide and smoking cessation on lung function decline in current smokers with mild to moderate COPD. The participants were recruited from a total of 10 sites, including nine in the United States and one Canadian site, as previously described [21, 23]. For the first 5 years, the lung function of participants was measured annually, followed by another lung function measurement at year 11. In year 5 of the LHS, venipuncture was performed on 5413 LHS participants. Genotyping was performed using buffy coat samples of 4251 European Americans in the LHS. The details of genotyping and quality control have been described previously [24]. Briefly, samples were genotyped using the Human 660WQuad v.1\_A BeadChip (Illumina, San Diego, CA, USA). Imputation was undertaken using the Michigan Imputation Server [25] using the Haplotype Reference Consortium [26] panel. Variants were excluded if the imputation  $r^2$  was  $<0.7$  and if the minor allele frequency was  $<1\%$ .

### Serum SP-D measurements in the LHS

Using one aliquot of serum samples from blood withdrawn in the year 5 visit, SP-D levels were measured in duplicate in 4754 LHS participants using a commercially available solid-phase sandwich ELISA according to the manufacturer's instructions (BioVendor Laboratory Medicine, Modrice, Czech Republic).

### pQTLs of serum SP-D: GWAS of SP-D serum levels

We performed a GWAS for log serum SP-D levels using SNPTEST [27] assuming an additive genetic model and adjusting for age, sex, body mass index (BMI), smoking status and the first five genetic principal components. The final GWAS dataset included 4041 subjects. Conditional analysis was performed to reveal independent SNPs in the significant regions associated with SP-D levels (online supplementary material). Using this approach, we identified three additional independent SNPs on the chromosome 10 locus.

### Association of SP-D pQTLs with lung tissue expression in the lung eQTL study

We used data from the lung eQTL study to identify potential effect of SP-D pQTLs on lung gene expression. The study details and the subjects' characteristics have been described previously [22]. Briefly, lung eQTLs were derived from a meta-analysis of genotyping and mRNA expression data performed in nontumour lung tissue samples from 1037 patients who underwent lung resection surgery. The eQTL analysis was adjusted for age, sex and smoking status. Genome-wide significant pQTLs for serum SP-D levels were tested for association with lung gene expression (*i.e.* lung eQTLs).

### Association of SP-D pQTLs with COPD risk and progression

The relationship of SNPs identified as pQTLs for serum SP-D with COPD risk was investigated in the International COPD Genetics Consortium (ICGC) study. The ICGC study is the largest COPD GWAS to date, and the lookup was performed in GWAS results for individuals with European ancestry with a sample size of 11 157 cases and 36 699 controls [28].

In addition, SP-D pQTLs were tested for association with FEV<sub>1</sub> decline over the 11 years of follow-up in the LHS using a multiple linear mixed effects model. All analyses in LHS were adjusted for age, sex, BMI and smoking status, as noted previously.

### Mendelian randomisation analysis

The independent pQTLs for SP-D identified in this study were used as genetic instruments in MR analyses. We used two MR methods: inverse variance weighting (IVW) and MR-Egger. The IVW MR approach was undertaken by regression of the genetic associations with the outcome (COPD risk from the ICGC study and lung function phenotypes in the LHS) on the genetic associations with observed SP-D serum levels in an inverse variance weighted linear regression with the intercept constrained to zero while accounting for any correlation (linkage disequilibrium) between the tested variants [29, 30]. The MR-Egger approach [31, 32] takes into account the presence of pleiotropy (SNPs associated with multiple unrelated phenotypes) and employs the same IVW approach, but with the intercept unconstrained. The intercept represents the average pleiotropic effect across the tested SNPs on the outcome. If the intercept differs from zero (the MR-Egger test), then there is evidence of directional pleiotropy [32]. In both approaches, we additionally tested for heterogeneity to identify whether or not the effect of SNPs is homogenous on the outcome.

### Weighted gene co-expression analysis

To identify genes whose expression in lung correlated with the SFTPD gene, the weighted gene co-expression analysis (WGCNA) R package [33] was used to cluster modules of co-expressed genes from 1037 lung tissue samples. Details of the WGCNA approach are available in the online supplementary material. Briefly, modules are determined based on the co-expression patterns, and "hub" genes are ones with the highest correlation to their module expression values (referred to as module membership).

TABLE 1 Demographic and clinical characteristics of subjects with genotypes across quintiles of serum surfactant protein D (SP-D) in the Lung Health Study

	Quintile					p-value
	1	2	3	4	5	
<b>Subjects</b>	786	792	812	822	829	
<b>SP-D ng·mL<sup>-1</sup></b>	≤59.7	59.7–81.3	81.3–105.4	105.4–139.5	>139.5	
<b>Age years</b>	52.1±6.8	52.8±6.8	54.0±6.6	53.8±6.6	55.1±6.3	<0.0001
<b>Males %</b>	63.6	65.5	61.6	62.2	63.1	0.524
<b>BMI kg·m<sup>-2</sup></b>	27.9±4.3	27.4±4.4	26.8±4.3	26.4±4.2	26.0±4.2	<0.0001
<b>Smokers<sup>#</sup> %</b>	70.9	78.7	82.9	87.2	92.5	<0.0001
<b>FEV<sub>1</sub> % predicted</b>	78.5±11.6	75.7±12.0	74.9±12.0	74.6±11.9	73.3±12.2	<0.0001

Data are presented as n, range or mean±SD, unless otherwise stated. p-values were generated using ANOVA for continuous variables and a Chi-squared test for categorical variables. BMI: body mass index; FEV<sub>1</sub>: forced expiratory volume in 1 s. #: either continuous or intermittent smokers, defined as smokers (verified objectively using salivary cotinine or exhaled carbon monoxide) at the time of at least one annual follow-up visit. All values are for the year 5 values because this was the visit during which blood samples were collected.

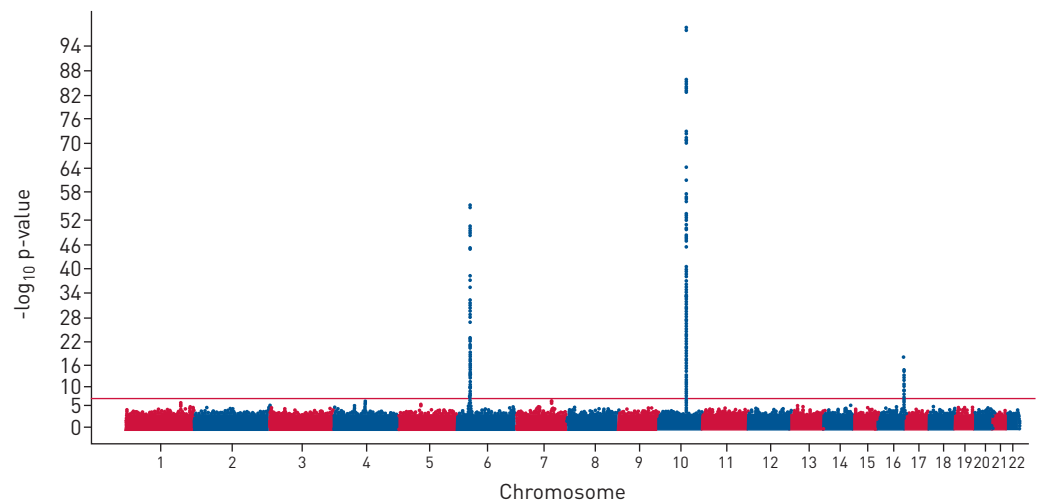


FIGURE 2 Serum surfactant protein D (SP-D) genome-wide association study (GWAS) Manhattan plot showing p-values and single-nucleotide polymorphism positions across 22 autosomal chromosomes. The horizontal red line represents the genome-wide cut-off of  $5 \times 10^{-8}$ .

### Statistical analysis software

All analyses were performed using R (version 3.2.1; [www.r-project.org](http://www.r-project.org)). Detailed and additional methods are available in the online supplementary material.

## Results

### Descriptive demographics of LHS participants

The demographics of LHS participants across the quintiles of serum SP-D levels are shown in table 1. Serum SP-D levels were significantly related to increasing age, decreasing BMI and FEV<sub>1</sub> and smoking status (with continuous smokers having higher average serum SP-D levels than sustained ex-smokers).

### GWAS of serum SP-D levels

Variants with minor allele frequency <1% or imputation quality <0.7 were filtered out, leaving 7 807 998 variants in this GWAS analysis. Quantile–quantile plots are presented in online supplementary figure S1 and showed a deviation from the expected distribution only for low p-values, indicating strong association signals. The genomic inflation factor ( $\lambda$ ) was 1.02, suggesting no systematic deviation in the association statistics due to factors such as population structure.

A total of 1633 SNPs representing three regions on chromosomes 6, 10 and 16 achieved genome-wide significance ( $p < 5 \times 10^{-8}$ ). The Manhattan plot is shown in figure 2 and the region plots are shown in figure 3. The most significantly associated pQTL was an intronic SNP (rs34406153) in the *SFTPD* gene ( $p = 2.02 \times 10^{-99}$ ). The sentinel SNP in the second associated locus was an intergenic SNP (rs12660817) in the human leukocyte antigen (HLA) region on chromosome 6 ( $p = 1.00 \times 10^{-55}$ ). The third locus harboured the intronic SNP (rs9927461) in the ATPase secretory pathway  $\text{Ca}^{2+}$  transporting 2 (*ATP2C2*) gene on chromosome 16 ( $p = 3.36 \times 10^{-18}$ ).

Conditional analyses on chromosome 10 identified an additional three independent signals exceeding genome-wide significance (online supplementary figure S2). The regions on chromosomes 6 and 16 did not harbour any additional independent loci, which were significant at the  $p < 5 \times 10^{-8}$  threshold. Genome-wide significant pQTLs for serum SP-D in these three regions are presented in table 2. SNP rs721917, which was reported recently by the ICGC consortium to be associated with COPD [28] was a pQTL for serum SP-D ( $p = 2.34 \times 10^{-13}$ ). This SNP had a modest linkage disequilibrium with the other sentinel SNPs in the chromosome 10 region, with linkage disequilibrium values ranging from  $r^2 = 0.16$  (SNP rs726014) to  $r^2 = 0.35$  (SNP rs12358676).

### The effect of SP-D pQTLs on lung mRNA expression

Given that SP-D is predominantly synthesised in the lung, the three loci harbouring pQTLs for serum SP-D were evaluated in lung tissue to determine whether these regions also had an effect on lung gene expression (*i.e.* lung eQTLs) [22]. The top eQTL results (based on the lowest eQTL p-value per gene) for each of the loci are shown in table 3. The pQTL SNPs in the *SFTPD* locus on chromosome 10 were also lung eQTLs for *SFTPD* and the SNPs' effects on serum protein levels were significantly related to their

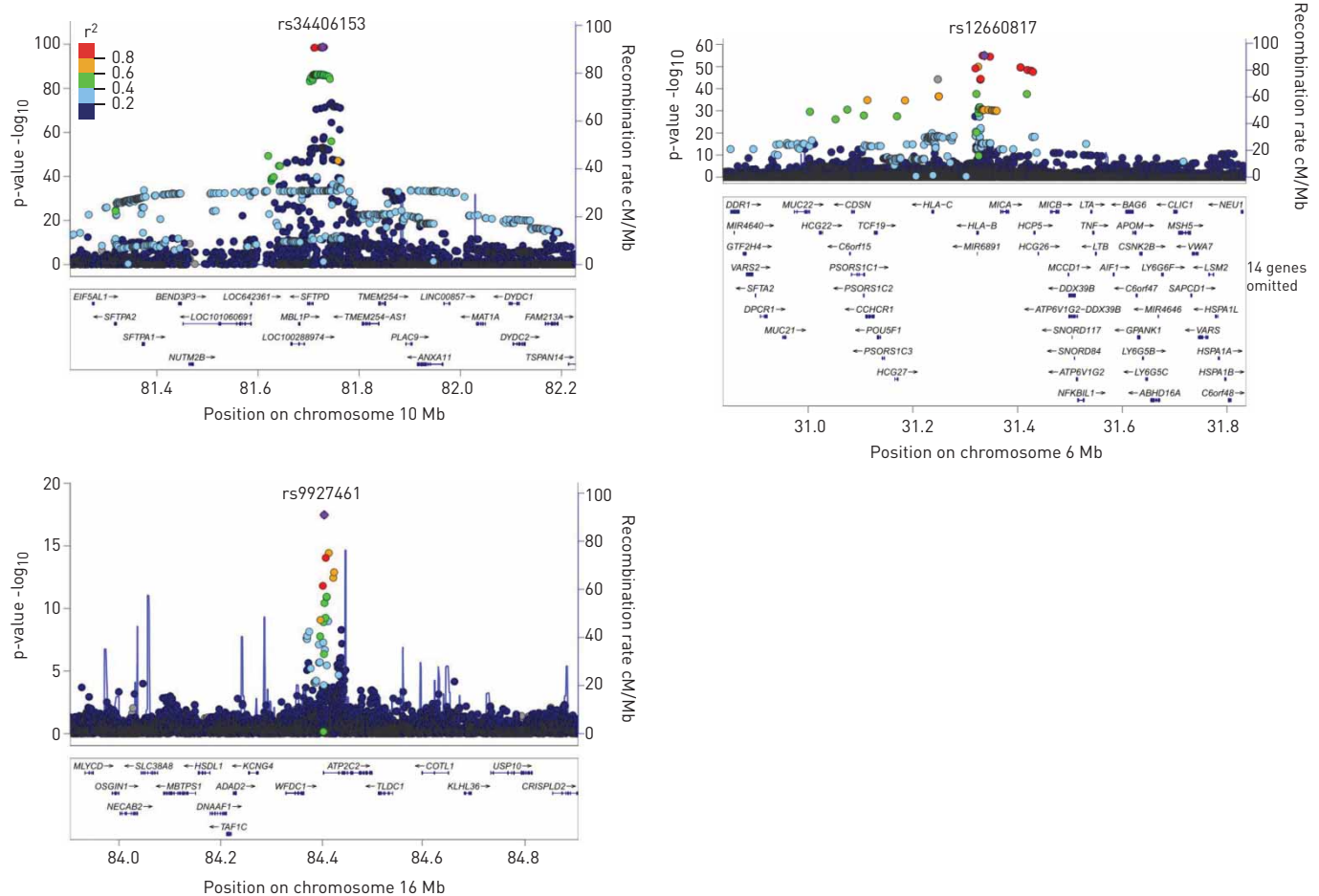


FIGURE 3 Region plots of the serum surfactant protein D (SP-D) protein quantitative trait loci (pQTLs) with gene names and their corresponding coordinates. The sentinel single-nucleotide polymorphism (SNP) is shown as a purple diamond and the colour coding of SNPs reflects the degree of linkage disequilibrium with the sentinel SNP using 1000G reference.

effect on mRNA expression in lung tissue (figure 4). We additionally report in table 2 the effect of SP-D pQTLs identified in our GWAS on lung tissue expression for the SP-D encoding gene *SFTPD* (i.e. lung eQTLs for *SFTPD*). Three of the four pQTLs in table 2 were significant eQTLs for *SFTPD* gene expression in lung tissue ( $p < 0.01$ ) with the same direction of effect to the protein levels in serum. Together, these

TABLE 2 The protein expression trait loci (pQTL) for serum surfactant protein D (SP-D)

	Chromosome	Position (hg19)	ALT/REF	MAF	Imputation <sup>#</sup> $r^2$	$\beta$	SE	SP-D pQTL p-value	Variance <sup>‡</sup> %	eQTL estimate	<i>SFTPD</i> eQTL p-value
<b>rs12660817</b>	6	31336757	G/A	0.055	1	0.366	0.023	$1.01 \times 10^{-55}$	6.010	0.135	0.152
<b>rs34406153</b>	10	81728069	A/G	0.177	0.996	-0.296	0.014	$2.02 \times 10^{-99}$	9.000	-0.150	$6.37 \times 10^{-3}$
<b>rs726014</b>	10	81708744	C/T	0.104	0.998	-0.343	0.017	$9.55 \times 10^{-87}$	8.201	-0.232	$4.97 \times 10^{-4}$
<b>rs144420336</b>	10	81716293	A/G	0.069	0.973	0.275	0.021	$4.94 \times 10^{-39}$	3.882	-0.034	0.749
<b>rs12358676</b>	10	81755644	G/T	0.464	0.988	0.093	0.011	$1.19 \times 10^{-18}$	1.802	0.123	$2.76 \times 10^{-3}$
<b>rs721917</b>	10	81706324	G/A	0.422	1	-0.078	0.011	$2.34 \times 10^{-13}$	1.071	0.042	0.304
<b>rs9927461</b>	16	84403561	A/C	0.148	0.975	-0.131	0.015	$3.36 \times 10^{-18}$	1.788	-0.040	0.512

$\beta$ , SE and p-values for chromosome 10 pQTLs are from the SP-D pQTL genome-wide association study. The expression quantitative trait loci (eQTL) estimates are based on the same allele coding in column ALT/REF [alternative allele, used to derive the statistics/reference allele]. SNP: single-nucleotide polymorphism; hg19: human genome version 19; MAF: minor allele frequency. <sup>#</sup>: imputation quality control metrics; SNP rs721917 is directly genotyped, while the rest were imputed; <sup>‡</sup>: the variance explained by that particular SNP obtained from univariate linear regression of log serum SP-D levels on SNP dosage.



TABLE 3 Lung tissue expression quantitative trait loci (eQTL) results for genetic variants that were serum surfactant protein D (SP-D) protein quantitative trait loci (pQTL)

Chromosome	SNP	Position (b37)	Alleles <sup>#</sup>	Z. Laval <sup>¶</sup>	Z. Groningen <sup>+</sup>	Z. UBC <sup>§</sup>	Lung eQTL p-value <sup>f</sup>	Gene symbol <sup>##</sup>	SP-D GWAS p-value
6	rs9380221	31150873	T C	-15.8	-17.8	-13.7	$3.87 \times 10^{-163}$	<i>PSORS1C3</i>	$8.61 \times 10^{-9}$
10	rs12267802	81749610	A G	2.9	1.7	3.4	$3.75 \times 10^{-06}$	<i>SFTPD</i>	$9.22 \times 10^{-30}$
16	rs12445532	84404091	G A	-9.3	-4.2	-3.7	$4.53 \times 10^{-27}$	<i>ATP2C2</i>	$3.72 \times 10^{-11}$

SNP: single-nucleotide polymorphism; GWAS: genome-wide association study; *PSORS1C3*: psoriasis susceptibility 1 candidate 3; *SFTPD*: surfactant protein D; *ATP2C2*: ATPase, Ca<sup>2+</sup> transporting, type 2C, member 2. <sup>#</sup>: the SNP alleles; the first letter is the coded/effective allele; <sup>¶</sup>: eQTL z-score in the Laval cohort; <sup>+</sup>: z-score in the Groningen cohort; <sup>§</sup>: z-score in the University of British Columbia (UBC) cohort; <sup>f</sup>: eQTL meta-analysis p-value from the UBC, Groningen and Laval studies; <sup>##</sup>: the gene that is regulated by the eQTL.

data suggest SNPs in the *SFTPD* gene region affect serum SP-D by altering the levels of gene expression in lung tissue. In contrast, the pQTL SNPs in chromosomes 6 and 16 were not related to *SFTPD* mRNA levels in the lung, indicating that they affect serum levels through other mechanisms (online supplementary figure S3).

#### Mendelian randomisation analyses

The six pQTL SNPs for serum SP-D levels were tested for association with COPD risk and progression (defined by FEV<sub>1</sub> decline over 11 years). The associations with COPD risk were extracted from the ICGC study meta-analysis. The association with COPD progression was performed using FEV<sub>1</sub> decline over 11 years in participants in the LHS. The combined SNP allelic effects on serum SP-D levels and risk of COPD and FEV<sub>1</sub> decline were then evaluated using two different MR frameworks: IVW and MR-Egger. In the ICGC cohort, IVW MR analysis revealed that SNPs associated with increased serum SP-D levels decreased the risk of COPD (estimate -0.19 change in COPD log odds ratio for every unit increase in log SP-D,  $p=6.46 \times 10^{-03}$ ; figure 5a). In the LHS, IVW MR analysis revealed that those SNPs that increased serum SP-D levels slowed the FEV<sub>1</sub> decline over 11 years (estimate 0.0038 less decline in FEV<sub>1</sub> slope per unit increase in log SP-D,  $p=7.68 \times 10^{-3}$ ; figure 5b). Similarly, the MR-Egger method demonstrated a significant protective relationship between SP-D and FEV<sub>1</sub> decline (estimate 0.0082 less decline in FEV<sub>1</sub> slope per unit increase in log SP-D,  $p=1.97 \times 10^{-02}$ ) as well as COPD risk (estimate -0.42 change in COPD

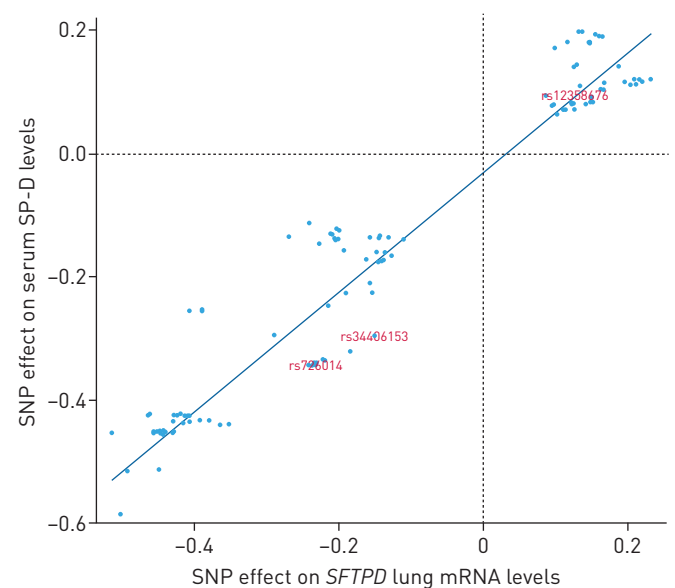


FIGURE 4 Serum surfactant protein D (SP-D) serum protein quantitative trait loci (pQTLs) on chromosome 10 as lung expression quantitative trait loci (eQTLs). Single-nucleotide polymorphisms (SNPs) plotted are those that have a significant effect on mRNA levels using a nominal p-value <0.05. The three SNPs shown are SNPs that were genome-wide significant pQTLs for SP-D and in addition were significant as lung eQTLs. The graph shows that SNPs associated with increased serum levels in the genome-wide association study (GWAS) are also associated with increased lung mRNA expression and *vice versa*. *SFTPD*: surfactant protein D gene.

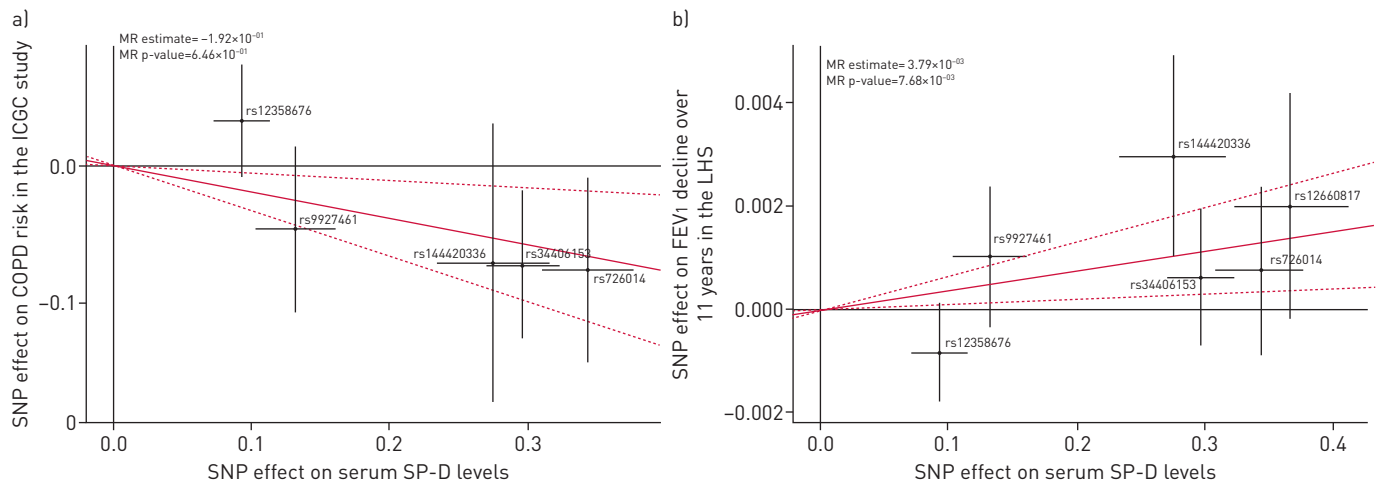


FIGURE 5 Inverse variance weighting Mendelian randomisation (MR) of serum surfactant protein D (SP-D) on chronic obstructive pulmonary disease (COPD) risk and progression. a) Estimates of SP-D on COPD risk; b) estimates of effect of SP-D on forced expiratory volume in 1 s (FEV<sub>1</sub>) decline. Estimates are derived from a serum genome-wide association study (GWAS) for SP-D and from the International COPD Genetics Consortium (ICGC) and Lung Health Study (LHS) GWAS for COPD risk and FEV<sub>1</sub> decline, respectively. Error bars represent 95% confidence intervals. The single-nucleotide polymorphism (SNP) reference (rs) numbers are shown. The slope of the line is the instrumental variable regression estimate of the effect of SP-D on COPD risk and FEV<sub>1</sub> decline.

log odds ratio for every unit increase in log SP-D,  $p=3.51 \times 10^{-03}$ ). The MR-Egger intercept p-value was not significant ( $p=0.15$ ), indicating the absence of pleiotropy in the MR IVW SNPs. The heterogeneity p-value for all MR tests was not significant ( $p>0.05$ ), indicating that the SNPs had a relatively similar effect on COPD risk and progression. Results from all MR analyses are shown in table 4 and the MR-Egger plots are presented in online supplementary figure S4.

#### WGCNA and SFTPD co-expression in lung tissue

To further understand the biology of SP-D, we searched for genes that were strongly correlated with *SFTPD* gene expression in lung tissue. Using the consensus network approach, we identified 38 distinct modules containing strongly co-expressed genes in the lung expression data. The module containing the *SFTPD* gene included 364 different genes, which were enriched for pathways related to lipid, cholesterol and fatty acids metabolic and biosynthesis processes (online supplementary table S1). This network construction additionally allowed us to identify “hub” genes, which, by definition, had the highest module membership value for a given module. The hub gene in the module containing *SFTPD* was surfactant-associated 3 (*SFTA3*). The top 50 genes with the strongest correlation with *SFTPD* are shown in online supplementary figure S4. Genes strongly co-expressed with *SFTPD* included *CRTAC1*, *CACNA2D2* and *S100*.

#### Discussion

Despite its high prevalence and impact, there are currently no biomarkers or disease-modifying therapies for COPD. Establishing causality for selected proteins and pathways is one promising step toward their development as both biomarkers and therapeutic targets.

TABLE 4 Summary of results from Mendelian randomisation (MR) analyses

	Estimate±SE	p-value	MR heterogeneity p-value	Egger intercept	
				Estimate±SE	p-value
MR of SP-D on FEV1 decline					
IVW MR	0.0038±0.0014	7.68×10 <sup>−03</sup>	0.0508	NA	NA
MR-Egger	0.0082±0.0035	1.97×10 <sup>−02</sup>	0.0872	−0.0011±0.0008	0.1532
MR of SP-D on COPD risk in the ICGC study					
IVW MR	−0.1923±0.0706	6.46×10 <sup>−03</sup>	0.1698	NA	NA
MR-Egger	−0.4238±0.1452	3.51×10 <sup>−03</sup>	0.3834	0.0533±0.0291	0.067693

SP-D: surfactant protein D; FEV<sub>1</sub>: forced expiratory volume in 1 s; IVW: inverse variance weighting; COPD: chronic obstructive pulmonary disease; ICGC: International COPD Genetics Consortium; NA: not applicable.



In this study, we determined the effect of genetic variation on SP-D protein levels in serum, mRNA levels in lung tissue and disease phenotypes to identify a potential causal role for SP-D in the aetiology of COPD. We identified six independent loci that were significantly associated with serum SP-D levels, including four independent SNPs at the *SFTPD* locus on chromosome 10, plus two distal regions of the genome (in the HLA locus on chromosome 6 and the *ATP2C2* gene region on chromosome 16). Importantly, integrating with lung tissue-specific eQTL data, we showed that SNPs in the *SFTPD* gene probably impact serum SP-D by modulating *SFTPD* mRNA levels in the lung. In contrast, the genetic signal on chromosome 16 appears to be related to expression of the *ATP2C2* gene, and thus its effect on serum SP-D is likely to be indirect. Most importantly, using two MR approaches, we showed that SNPs that were associated with higher serum SP-D levels imparted lower risk for COPD and FEV<sub>1</sub> decline over time with no evidence for heterogeneity across SNPs on these phenotypes. Together, these data suggest a causal role of SP-D in the pathogenesis of COPD.

A study by KIM *et al.* [34] identified the three loci that were pQTLs for SP-D on chromosomes 6, 10 and 16. A more recent study by SUN *et al.* [35] additionally identified SNP rs2146192 in the *SFTPD* gene region as associated with SP-D blood levels in COPD subjects. This SNP is in modest linkage disequilibrium with the top SP-D pQTL in our study ( $r^2=0.53$ ). Using conditional analyses, SUN *et al.* identified two other independent pQTL SNPs: one on chromosome 10 and the other in the HLA region on chromosome 6. In the current study, we confirm and extend the findings of KIM *et al.* and SUN *et al.* We replicated the same three loci from the study by KIM *et al.* and identified several additional SNPs in the *SFTPD*, HLA and *ATP2C2* regions that act as pQTLs for SP-D levels. In total, the three loci explained 30.7% of the variation in serum SP-D levels. The strongest pQTLs for SP-D in this study were in the *SFTPD* gene encoding the SP-D protein, suggesting that serum levels are driven in part by SP-D production in lung tissue. Consistent with this notion, we observed that the SNPs' effect on *SFTPD* mRNA levels in lung tissue is directionally consistent with their effect on serum SP-D levels. In addition, there were no distal genetic determinants of *SFTPD* expression in lung tissue suggesting the SP-D serum pQTLs on chromosome 10 captured most of the regulatory effects on SP-D production in the lung.

The identification of distal pQTLs may provide important insights into SP-D biology. The chromosome 16 locus contains the *ATP2C2* gene, which is expressed in many tissues (including the lungs) and functions as a transporter of  $\text{Ca}^{2+}$  ions [36]. However, SNPs in the *ATP2C2* region had no effect on *SFTPD* mRNA levels in lungs; mRNA levels of *ATP2C2* were positively correlated with *SFTPD* mRNA expression ( $r=0.2$ ,  $p=6.9\times 10^{-11}$ ). Together, these data suggest *ATP2C2* may regulate serum SP-D through an indirect mechanism such as post-translational modification, transport and/or clearance, among others, or through a shared functional pathway.

In lung tissue, *SFTPD* was mapped to a module of 364 strongly co-expressed genes which were enriched for pathways related to lipid and cholesterol biosynthesis. Surfactant-associated 3 (*SFTA3*) was the "hub" gene and its expression was positively related to *SFTPD*. *SFTA3* is an immunoregulatory protein in the lung, and may play an important role in inflammation and immune defence [37]. Other genes that were strongly co-expressed with *SFTPD* included cartilage acidic protein 1 (*CRTAC1*), the calcium channel, voltage-dependent,  $\alpha 2/\delta$  subunit 2 (*CACNA2D2*) and S100 calcium-binding protein A14 (*S100A14*) genes. *CRTAC1* is upregulated (>10-fold) in type 2 epithelial cells as they differentiate in culture [38]. Previous studies have shown that binding of SP-D to phosphatidylinositol [39], toll like receptor (TLR)2 and TLR4 [40] and carbohydrate structures on the surface of bacteria, viruses and fungi is calcium dependent [41–46]. Calcium may also cause SP-D oligomers to self-aggregate, as has been shown previously with surfactant protein A [47]. Further work is needed to unravel the mechanisms underlying calcium-related gene associations with SP-D gene expression and serum levels.

In the LHS, there was a modest negative correlation between serum SP-D levels and FEV<sub>1</sub> ( $r=-0.063$ ,  $p<1\times 10^{-4}$ ), consistent with several other studies [48, 49]. In addition, serum SP-D levels were correlated with accelerated FEV<sub>1</sub> decline using a linear mixed effect regression (estimates  $-0.0068$ ,  $p=6.60\times 10^{-23}$ ). Superficially, these relationships appear to be paradoxical to the main findings of the present study (opposite direction of association). However, cross-sectional relationships between lung function and serum SP-D have to be interpreted with caution because of the possibility of reverse causation (*i.e.* pathological processes that contribute to COPD causing increased leakage of SP-D from lungs into the systemic circulation) and confounding (*e.g.* smoking). For instance, we and others have previously shown that serum SP-D is affected not only by lung synthesis of SP-D, but also by permeability of the alveolar–capillary interface, which is perturbed in COPD and increased by active smoking and lung inflammation [48, 50]. Furthermore, it is notable that in the LHS none of the study participants were taking inhaled or systemic steroids at the time of blood assessment. Thus, the current results were not impacted by steroids. Genetic analyses avoid most of these pitfalls since the causal pathway between genetic variants and protein and disease is unidirectional. In the present study, MR results supported a significant protective role for SP-D on both risk of COPD and FEV<sub>1</sub> decline.

There is a huge unmet clinical need to identify novel biomarkers and therapeutic targets for COPD. SP-D represents an excellent candidate since 1) it is a lung-specific protein that can be reliably assayed in blood, BAL fluid and lung tissue; 2) it tracks well with disease severity and exacerbations [49]; 3) it is responsive to steroid treatment [49]; and 4) it has inherent functional and biological attributes that suggest a role in COPD pathogenesis [10]. Moreover, this study is the first to our knowledge to report a novel causal link between genetically elevated SP-D serum levels and decreased risk for COPD and decelerated FEV<sub>1</sub> decline.

The current study has a number of limitations. First, LHS participants were all smokers with mild or moderate COPD at the time of recruitment, and this may have affected the serum measures of SP-D as shown by the discordance between genetic and observational analyses. Second, the serum and lung mRNA expression were measured in different individuals, which prevents direct comparisons between the two sites. However, the use of genetic variation provided a robust method to estimate and extrapolate the findings for mRNA and protein between lung tissue and blood. Third, SP-D was detected in serum using ELISA, which may not differentiate between intact (dodecameric) and monomeric or trimeric forms of SP-D. Further quantitative studies of different multimeric SP-D protein structures are warranted.

In conclusion, the current study investigated the effect of genetic variation on serum SP-D levels, lung tissue gene expression and COPD risk and progression. The findings provided novel insights into SP-D biology and support a novel causal link between increased SP-D levels and protection against COPD risk and accelerated FEV<sub>1</sub> decline. SP-D represents a very promising biomarker and therapeutic target for COPD which warrants further study.

### Acknowledgements

Authors' contributions are as follows. Conception and design of the study: M. Obeidat, P.D. Paré and D.D. Sin; lung expression quantitative trait loci data collection and analysis: K. Hao, D.C. Nickle, M. van den Berge, W. Timens, Y. Bossé and P. Joubert; International COPD Genetics Consortium genome-wide association study data collection and analysis: M.H. Cho, B.D. Hobbs, K. de Jong and M. Boezen; Lung Health Study data analyses: M. Obeidat, X. Li, G. Zhou, N. Fishbane, N.N. Hansel, T.H. Beaty, N. Rafaels, R. Mathias, I. Ruczinski and K.C. Barnes; statistical support and advice on study conduct: R. Hung and S. Burgess; writing the manuscript: M. Obeidat, P.D. Paré and D.D. Sin; all co-authors discussed results and implications and commented on the manuscript at all stages.

### References

- 1 World Health Organization (WHO). The Top 10 Causes of Death. Geneva, WHO, 2014.
- 2 Vestbo J, Hurd SS, Agustí AG, *et al.* Global strategy for the diagnosis, management, and prevention of chronic obstructive pulmonary disease: GOLD executive summary. *Am J Respir Crit Care Med* 2013; 187: 347–365.
- 3 Hukkinen M, Kaprio J, Broms U, *et al.* Heritability of lung function: a twin study among never-smoking elderly women. *Twin Res Hum Genet* 2011; 14: 401–407.
- 4 Wert SE, Yoshida M, LeVine AM, *et al.* Increased metalloproteinase activity, oxidant production, and emphysema in surfactant protein D gene-inactivated mice. *Proc Natl Acad Sci USA* 2000; 97: 5972–5977.
- 5 Atochina EN, Beers MF, Hawgood S, *et al.* Surfactant protein-D, a mediator of innate lung immunity, alters the products of nitric oxide metabolism. *Am J Respir Cell Mol Biol* 2004; 30: 271–279.
- 6 Voorhout WF, Veenendaal T, Kuroki Y, *et al.* Immunocytochemical localization of surfactant protein D (SP-D) in type II cells, Clara cells, and alveolar macrophages of rat lung. *J Histochem Cytochem* 1992; 40: 1589–1597.
- 7 Madsen J, Kliem A, Tornøe I, *et al.* Localization of lung surfactant protein D on mucosal surfaces in human tissues. *J Immunol* 2000; 164: 5866–5870.
- 8 Fisher JH, Sheftelyevich V, Ho YS, *et al.* Pulmonary-specific expression of SP-D corrects pulmonary lipid accumulation in SP-D gene-targeted mice. *Am J Physiol Lung Cell Mol Physiol* 2000; 278: L365–L373.
- 9 Wright JR. Immunoregulatory functions of surfactant proteins. *Nat Rev Immunol* 2005; 5: 58–68.
- 10 Sin DD, Pahlavan PS, Man SF. Surfactant protein D: a lung specific biomarker in COPD? *Ther Adv Respir Dis* 2008; 2: 65–74.
- 11 Wert SE, Yoshida M, LeVine AM, *et al.* Increased metalloproteinase activity, oxidant production, and emphysema in surfactant protein D gene-inactivated mice. *Proc Natl Acad Sci USA* 2000; 97: 5972–5977.
- 12 Foreman MG, Kong X, DeMeo DL, *et al.* Polymorphisms in surfactant protein-D are associated with chronic obstructive pulmonary disease. *Am J Respir Cell Mol Biol* 2011; 44: 316–322.
- 13 Bowler RP. Surfactant protein D as a biomarker for chronic obstructive pulmonary disease. *COPD* 2012; 9: 651–653.
- 14 Ishii T, Hagiwara K, Kamio K, *et al.* Involvement of surfactant protein D in emphysema revealed by genetic association study. *Eur J Hum Genet* 2012; 20: 230–235.
- 15 Honda Y, Takahashi H, Kuroki Y, *et al.* Decreased contents of surfactant proteins A and D in BAL fluids of healthy smokers. *Chest* 1996; 109: 1006–1009.
- 16 Mutti A, Corradi M, Goldoni M, *et al.* Exhaled metallic elements and serum pneumoproteins in asymptomatic smokers and patients with COPD or asthma. *Chest* 2006; 129: 1288–1297.
- 17 White J, Sofat R, Hemani G, *et al.* Plasma urate concentration and risk of coronary heart disease: a Mendelian randomisation analysis. *Lancet Diabetes Endocrinol*; 4: 327–336.
- 18 Davey Smith G, Ebrahim S. 'Mendelian randomization': can genetic epidemiology contribute to understanding environmental determinants of disease? *Int J Epidemiol* 2003; 32: 1–22.
- 19 Voight BF, Peloso GM, Orho-Melander M, *et al.* Plasma HDL cholesterol and risk of myocardial infarction: a mendelian randomisation study. *Lancet* 380: 572–580.
- 20 Davey Smith G, Hemani G. Mendelian randomization: genetic anchors for causal inference in epidemiological studies. *Hum Mol Genet* 2014; 23: R89–R98.

- 21 Anthonisen NR, Connett JE, Kiley JP, *et al.* Effects of smoking intervention and the use of an inhaled anticholinergic bronchodilator on the rate of decline of FEV<sub>1</sub>. The Lung Health Study. *JAMA* 1994; 272: 1497–1505.
- 22 Hao K, Bossé Y, Nickle DC, *et al.* Lung eQTLs to help reveal the molecular underpinnings of asthma. *PLoS Genet* 2012; 8: e1003029.
- 23 Kanner RE, Connett JE, Williams DE, *et al.* Effects of randomized assignment to a smoking cessation intervention and changes in smoking habits on respiratory symptoms in smokers with early chronic obstructive pulmonary disease: the Lung Health Study. *Am J Med* 1999; 106: 410–416.
- 24 Hansel NN, Ruczinski I, Rafaels N, *et al.* Genome-wide study identifies two loci associated with lung function decline in mild to moderate COPD. *Hum Genet* 2013; 132: 79–90.
- 25 Das S, Forer L, Schönherr S, *et al.* Next-generation genotype imputation service and methods. *Nat Genet* 2016; 48: 1284–1287.
- 26 McCarthy S, Das S, Kretzschmar W, *et al.* A reference panel of 64,976 haplotypes for genotype imputation. *Nat Genet* 2016; 48: 1279–1283.
- 27 Marchini J, Howie B, Myers S, *et al.* A new multipoint method for genome-wide association studies by imputation of genotypes. *Nat Genet* 2007; 39: 906–913.
- 28 Hobbs BD, De Jong K, Lamontagne M, *et al.* Genetic loci associated with chronic obstructive pulmonary disease overlap with loci for lung function and pulmonary fibrosis. *Nat Genet* 2017; 49: 426–432.
- 29 Burgess S, Dudbridge F, Thompson SG. Combining information on multiple instrumental variables in Mendelian randomization: comparison of allele score and summarized data methods. *Stat Med* 2016; 35: 1880–1906.
- 30 Johnson T. Efficient Calculation for Multi-SNP Genetic Risk Scores. American Society of Human Genetics Annual Meeting, San Francisco, CA, USA, 2012.
- 31 Holmes MV, Ala-Korpela M, Davey Smith G. Mendelian randomization in cardiometabolic disease: challenges in evaluating causality. *Nat Rev Cardiol* 2017; 14: 577–590.
- 32 Bowden J, Davey Smith G, Burgess S. Mendelian randomization with invalid instruments: effect estimation and bias detection through Egger regression. *Int J Epidemiol* 2015; 44: 512–525.
- 33 Langfelder P, Horvath S. WGCNA: an R package for weighted correlation network analysis. *BMC Bioinformatics* 2008; 9: 559.
- 34 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: 1238–1247.
- 35 Sun W, Kechris K, Jacobson S, *et al.* Common genetic polymorphisms influence blood biomarker measurements in COPD. *PLoS Genet* 2016; 12: e1006011.
- 36 Vanoevelen J, Dode L, Van Baelen K, *et al.* The secretory pathway Ca<sup>2+</sup>/Mn<sup>2+</sup>-ATPase 2 is a Golgi-localized pump with high affinity for Ca<sup>2+</sup> ions. *J Biol Chem* 2005; 280: 22800–22808.
- 37 Schicht M, Rausch F, Finotto S, *et al.* SFTA3, a novel protein of the lung: three-dimensional structure, characterisation and immune activation. *Eur Respir J* 2014; 44: 447–456.
- 38 Ballard PL, Lee JW, Fang X, *et al.* Regulated gene expression in cultured type II cells of adult human lung. *Am J Physiol Lung Cell Mol Physiol* 2010; 299: L36–L50.
- 39 Ogasawara Y, Kuroki Y, Akino T. Pulmonary surfactant protein D specifically binds to phosphatidylinositol. *J Biol Chem* 1992; 267: 21244–21249.
- 40 Ohya M, Nishitani C, Sano H, *et al.* Human pulmonary surfactant protein D binds the extracellular domains of Toll-like receptors 2 and 4 through the carbohydrate recognition domain by a mechanism different from its binding to phosphatidylinositol and lipopolysaccharide. *Biochemistry* 2006; 45: 8657–8664.
- 41 Meschi J, Crouch EC, Skolnik P, *et al.* Surfactant protein D binds to human immunodeficiency virus (HIV) envelope protein gp120 and inhibits HIV replication. *J Gen Virol* 2005; 86: 3097–3107.
- 42 Ferguson JS, Voelker DR, McCormack FX, *et al.* Surfactant protein D binds to *Mycobacterium tuberculosis* bacilli and lipoarabinomannan via carbohydrate-lectin interactions resulting in reduced phagocytosis of the bacteria by macrophages. *J Immunol* 1999; 163: 312–321.
- 43 Murray E, Khamri W, Walker MM, *et al.* Expression of surfactant protein D in the human gastric mucosa and during *Helicobacter pylori* infection. *Infect Immun* 2002; 70: 1481–1487.
- 44 Nadesalingam J, Reid KB, Palaniyar N. Collectin surfactant protein D binds antibodies and interlinks innate and adaptive immune systems. *FEBS Lett* 2005; 579: 4449–4453.
- 45 Hickling TP, Bright H, Wing K, *et al.* A recombinant trimeric surfactant protein D carbohydrate recognition domain inhibits respiratory syncytial virus infection *in vitro* and *in vivo*. *Eur J Immunol* 1999; 29: 3478–3484.
- 46 Madan T, Eggleton P, Kishore U, *et al.* Binding of pulmonary surfactant proteins A and D to *Aspergillus fumigatus* conidia enhances phagocytosis and killing by human neutrophils and alveolar macrophages. *Infect Immun* 1997; 65: 3171–3179.
- 47 van Iwaarden JF, Teding van Berkhout F, Whitsett JA, *et al.* A novel procedure for the rapid isolation of surfactant protein A with retention of its alveolar-macrophage-stimulating properties. *Biochem J* 1995; 309: 551–555.
- 48 Tkacova R, McWilliams A, Lam S, *et al.* Integrating lung and plasma expression of pneumo-proteins in developing biomarkers in COPD: a case study of surfactant protein D. *Med Sci Monit* 2010; 16: CR540–CR544.
- 49 Lomas DA, Silverman EK, Edwards LD, *et al.* Serum surfactant protein D is steroid sensitive and associated with exacerbations of COPD. *Eur Respir J* 2009; 34: 95–102.
- 50 Winkler C, Atochina-Vasserman EN, Holz O, *et al.* Comprehensive characterisation of pulmonary and serum surfactant protein D in COPD. *Respir Res* 2011; 12: 29.