



Genetic studies as a tool for identifying novel potential targets for treatment of COPD

Ani Manichaikul and Jennifer N. Nguyen

Affiliation: Center for Public Health Genomics, University of Virginia, Charlottesville, VA, USA.

Correspondence: Ani Manichaikul, Center for Public Health Genomics, University of Virginia School of Medicine, West Complex Room 6115, Charlottesville, VA 22903, USA. E-mail: amanicha@virginia.edu

 @ERSpublications

Genetic studies are a useful tool for identification of novel causal biomarkers and therapeutic targets for COPD <http://ow.ly/zZKC30fNecz>

Cite this article as: Manichaikul A, Nguyen JN. Genetic studies as a tool for identifying novel potential targets for treatment of COPD. *Eur Respir J* 2017; 50: 1702042 [<https://doi.org/10.1183/13993003.02042-2017>].

Chronic obstructive pulmonary disease (COPD), characterised by chronic airflow limitation and abnormal response to noxious particles or gases [1], is currently the fourth leading cause of death worldwide [2] and the only major cause of death that has seen continued increases in recent years. While smoking cessation, oxygen therapy and bronchodilators offer some relief from the symptoms of disease, and medications are available to prevent COPD exacerbations [3], there are currently no treatments available to slow or stop progression, reflecting a limited understanding of the molecular underpinnings of the disease. In this issue of the *European Respiratory Journal*, OBEIDAT *et al.* [4] identify surfactant protein-D (SP-D) as a potentially causal risk factor for COPD. Their paper is presented in multiple distinct parts, beginning with a genome-wide association study (GWAS) to identify single-nucleotide polymorphisms (SNPs) from three genomic regions (on chromosomes 6, 10 and 16) associated with circulating levels of SP-D, followed by a Mendelian randomisation study demonstrating that genetically determined lower levels of SP-D are associated with increased risk of COPD based on results from a large-scale consortium based on a GWAS of up to ~11 000 COPD cases and ~37 000 controls recently published by the International COPD Genetics Consortium [5]. The results from the study of OBEIDAT *et al.* [4] are promising and open new potential avenues for early diagnosis and treatment of this disease. How did we get to this novel finding, what steps remain to carry this finding forward towards establishing the role of SP-D in COPD, and how can we stimulate additional discovery of novel therapeutic targets?

Genetic studies of COPD support SP-D as a candidate biomarker

For decades, the *SERPINA1* gene that encodes α_1 -antitrypsin was the only genetic factor known to accelerate decline in pulmonary function and increase risk of COPD. Replacement of α_1 -antitrypsin may slow the decline for a subset of COPD patients with α_1 -antitrypsin deficiency [6]. Over the past decade, GWAS by the Cohorts for Heart and Aging Research in Genomic Epidemiology (CHARGE) [7] and SpiroMeta consortia [8] have identified ~30 gene regions that contain common SNPs significantly associated with forced expiratory volume in 1 s (FEV₁) and its ratio to forced vital capacity (FEV₁/FVC).

Received: Oct 04 2017 | Accepted after revision: Oct 08 2017

Support statement: This work was supported by funding from NIH R01HL131565. Funding information for this article has been deposited with the Crossref Funder Registry.

Conflict of interest: Disclosures can be found alongside this article at erj.ersjournals.com

Copyright ©ERS 2017

GWAS approaches have also been used to identify genetic variants for COPD [5, 9–13]. Major candidate gene regions identified through GWAS of COPD include *FAM13A* [11], *CHRNA3/5/IREB2* [9] and *HHIP* [13], with four novel loci, *EEFSEC*, *DSP*, *MTCL1*, and *SFTPD*, that emerged in the most recently published GWAS of COPD [5].

Before its identification through GWAS, SP-D had been already studied as a biomarker for COPD exacerbations [14] and asthma severity [15, 16]. Interestingly, the lead *SFTPD* SNP identified in association with COPD [5], rs721917, is a protein coding missense mutation providing a rationale for SP-D as the protein implicated by the identified genetic association. The same SNP also has demonstrated association with emphysema [17]. Despite the mounting evidence of SP-D as a biomarker for COPD, thus far there has not been sufficient evidence to establish a causal role for SP-D in the development of COPD.

Mendelian randomisation analysis and its underlying assumptions

Mendelian randomisation provides a useful analytical framework to establish the causality of a candidate biomarker in the pathogenesis of disease. The approach, which can be conducted in the context of an observational study, builds on the random assignment of alleles at meiosis to carry out what some consider a “natural” randomised trial. The basic idea underlying the approach is that we can use data from genetic association studies, including GWAS, to represent the biomarker of interest. In the study by OBEIDAT *et al.* [4], SNPs identified in a GWAS of SP-D levels were used as instruments independent of potential confounders of the relationship between SP-D and COPD. When applied correctly, Mendelian randomisation analysis can provide one piece of evidence to establish the biomarker of interest as a causal factor in the disease of interest [18].

Importantly, Mendelian randomisation analysis relies on multiple assumptions in order to make a valid inference regarding the causality of the underlying biomarker of interest [19, 20]. Here, we present the three underlying assumptions and discuss the extent to which these assumptions hold true in the context of the current study.

1) Assumption 1: the SNPs used to construct the instrument should be robustly associated with SP-D. OBEIDAT *et al.* [4] used a stringent threshold for genome-wide significance to identify the SNPs taken forward for Mendelian randomisation analyses, such that we have relatively strong evidence that the assumption is met. However, as with any GWAS study, it is possible that some of the identified SNPs represent false positives. Although some of the protein quantitative trait loci reported in this study have been reported in previous studies [21, 22], some of the SP-D associated SNPs used in Mendelian randomisation analyses by OBEIDAT *et al.* [4] were newly reported in this study and have yet to be confirmed in analysis of independent datasets.

2) Assumption 2: the SNPs used to construct the instrument should be unrelated to any external confounders of the relationship between SP-D and COPD. SP-D has been associated with COPD even after individuals have ceased smoking [23], suggesting that the relationship between SP-D and COPD may be at least partially independent of this major risk factor for COPD. However, it is difficult to provide concrete “proof” that this assumption is met, as there are potentially additional confounders of the relationship between SP-D and COPD that have not yet been identified.

3) Assumption 3: the SNPs used to construct the instrument for SP-D are related to COPD only through their modulation of SP-D levels. While SP-D region SNPs have previously been reported in genetic association studies of COPD [5, 24] and emphysema [17], it is noteworthy that the top associated SNP in these studies was the protein coding missense mutation rs721917, providing a strong rationale to pinpoint SP-D as the underlying mediator. However, numerous other genes lie within ~200 kb of the *SFTPD* gene that encodes SP-D, these include *TMEM25A* encoding a transmembrane protein, the antisense RNA gene *NUTM2B-AS1*, and the mannose-binding lectin pseudogene, *MBL1P*. As the functionality of some GWAS-identified regions has been shown to act over regulatory regions that span megabase distances [25], additional assessment of genes within the region of the SP-D locus may be needed to establish whether or not *SFTPD* is the lone gene underlying this genomic signal. That said, we agree with the authors that there is reasonable evidence to surmise SP-D might be the primary and possibly sole gene underlying the identified genomic signal at this locus.

Mendelian randomisation analysis: strengths, limitations and extensions

The core strength of the Mendelian randomisation approach is that, applied correctly, it has the ability to inform on the causal nature of the relationship of the identified biomarker, in our case SP-D, and the disease outcome of interest, COPD in this case. However, as noted earlier, there are multiple key assumptions necessary to make valid inference regarding causality based on the results of a Mendelian randomisation study, and it remains somewhat unclear whether or not these assumptions held true for the

current study. In addition, while the direction of effect provided by the reported Mendelian randomisation analysis is informative in conceptualising the general relationship between SP-D levels and risk of COPD, the corresponding effect estimates obtained by Mendelian randomisation analysis probably do not generalise to the increased risk of COPD that may correspond to decreased levels of measured SP-D at the population level. Therefore, a successful Mendelian randomisation study is merely the beginning, and in the future we would like to see: 1) population-level longitudinal studies used to confirm the prognostic value of SP-D as a biomarker for COPD; and 2) randomised controlled clinical trials to assess the value of increasing SP-D for prevention of COPD.

Given the promising results obtained by the current Mendelian randomisation study of SP-D for COPD, this study can also be viewed as a proof of concept for characterisation of other candidate biomarkers for COPD. The growing availability of biomarker studies for COPD present a good resource for future Mendelian randomisation studies that could be used to assess the causal nature of those biomarkers in COPD. In addition, existing and forthcoming 'omics studies provide the opportunity to apply the Mendelian randomisation approach in a high-throughput manner. The methodologies for high-throughput extension of Mendelian randomisation-type approaches are quickly gaining traction in the genetic community [26, 27], providing systematic frameworks for integration of genomic and transcriptomic datasets. In the coming years, we expect it will become feasible and fruitful to leverage these new statistical methodologies, combined with growing genomic datasets for COPD and related traits, as well as lung-specific 'omics studies to identify novel causal biomarkers for COPD, leading to novel therapeutic targets for treatment of COPD.

References

- 1 Gerald LB, Bailey WC. Global Initiative for Chronic Obstructive Lung Disease. *J Cardiopulm Rehabil* 2002; 22: 234–244.
- 2 Lozano R, Naghavi M, Lim S, *et al.* Global and regional mortality from 235 causes of death for 20 age groups in 1990 and 2010: a systematic analysis for the Global Burden of Disease Study 2010. *Lancet* 2012; 380: 2095–2128.
- 3 Wedzicha JA, Calverley PMA, Albert RK, *et al.* Prevention of COPD exacerbations: a European Respiratory Society/American Thoracic Society guideline. *Eur Respir J* 2017; 50: 1602265.
- 4 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.
- 5 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.
- 6 Teschler H. Long-term experience in the treatment of α 1-antitrypsin deficiency: 25 years of augmentation therapy. *Eur Respir Rev* 2015; 24: 46–51.
- 7 Psaty BM, O'Donnell CJ, Gudnason V, *et al.* Cohorts for Heart and Aging Research in Genomic Epidemiology (CHARGE) consortium: design of prospective meta-analyses of genome-wide association studies from 5 cohorts. *Circ Cardiovasc Genet* 2009; 2: 73–80.
- 8 Artigas MS, Loth DW, Wain LV, *et al.* Genome-wide association and large-scale follow up identifies 16 new loci influencing lung function. *Nat Genet* 2011; 43: 1082–1090.
- 9 Wilk JB, Shrine NRG, Loehr LR, *et al.* Genome-wide association studies identify CHRNA5/3 and HTR4 in the development of airflow obstruction. *Am J Respir Crit Care Med* 2012; 186: 622–632.
- 10 Pillai SG, Ge D, Zhu G, *et al.* A genome-wide association study in chronic obstructive pulmonary disease (COPD): identification of two major susceptibility loci. *PLoS Genet* 2009; 5: e1000421.
- 11 Cho MH, Boutaoui N, Klanderman BJ, *et al.* Variants in FAM13A are associated with chronic obstructive pulmonary disease. *Nat Genet* 2010; 42: 200–202.
- 12 Cho MH, Castaldi PJ, Wan ES, *et al.* A genome-wide association study of COPD identifies a susceptibility locus on chromosome 19q13. *Hum Mol Genet* 2012; 21: 947–957.
- 13 Cho MH, McDonald M-LN, Zhou X, *et al.* Risk loci for chronic obstructive pulmonary disease: a genome-wide association study and meta-analysis. *Lancet Respir Med* 2014; 2: 214–225.
- 14 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.
- 15 Benfante A, Battaglia S, Paterno A, *et al.* Serum surfactant protein D is a potential biomarker of severity of asthma. *Am J Respir Crit Care Med* 2015; 191: A4311.
- 16 Benfante A, Battaglia S, Principe S, *et al.* Asthmatics with high levels of serum surfactant protein D have more severe disease. *Eur Respir J* 2016; 47: 1864–1867.
- 17 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.
- 18 Davey Smith G, Hemani G. Mendelian randomization: genetic anchors for causal inference in epidemiological studies. *Hum Mol Genet* 2014; 23: R89–R98.
- 19 Solovieff N, Cotsapas C, Lee PH, *et al.* Pleiotropy in complex traits: challenges and strategies. *Nat Rev Genet* 2013; 14: 483–495.
- 20 VanderWeele TJ, Tchetgen Tchetgen EJ, Cornelis M, *et al.* Methodological challenges in mendelian randomization. *Epidemiology* 2014; 25: 427–435.
- 21 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.
- 22 Sun W, Kechris K, Jacobson S, *et al.* Common genetic polymorphisms influence blood biomarker measurements in COPD. *PLoS Genet* 2016; 12: e1006011.
- 23 Sims MW, Tal-Singer RM, Kierstein S, *et al.* Chronic obstructive pulmonary disease and inhaled steroids alter surfactant protein D (SP-D) levels: a cross-sectional study. *Respir Res* 2008; 9: 13.

- 24 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.
- 25 Smemo S, Tena JJ, Kim K-H, *et al.* Obesity-associated variants within FTO form long-range functional connections with IRX3. *Nature* 2014; 507: 371–375.
- 26 Gusev A, Ko A, Shi H, *et al.* Integrative approaches for large-scale transcriptome-wide association studies. *Nat Genet* 2016; 48: 245–252.
- 27 Gamazon ER, Wheeler HE, Shah KP, *et al.* A gene-based association method for mapping traits using reference transcriptome data. *Nat Genet* 2015; 47: 1091–1098.