Check for updates

Functional Gene Variants in Chronic Obstructive Pulmonary Disease: The Search Continues

Chronic obstructive pulmonary disease (COPD) is the third leading cause of death in the United States (1). Although efforts to curb tobacco usage have the potential to lessen the development of COPD and other tobacco-related diseases, the high degree of variability that exists between tobacco use and severity of airflow limitation supports the concept that genetic factors influence the severity of COPD in smokers. Such observations, including some degree of familial aggregation, indicate that a much more complex genetic pathobiology of COPD exists that merits investigation. Investigators continue to pursue the identification of variants within the human genome in COPD. The quest to identify clinically informative variants remains challenging given the sheer number of genetic variants that lack clinical association and the few that harbor a statistical association with diseases as complex as COPD (2, 3).

To date, several studies have determined that genome-wide association studies (GWASs) may hold the key to uncovering genetic susceptibility to COPD. In one of the early studies, Pillai and colleagues conducted a GWAS for SNPs in COPD cases ($N = 823$) and control ($N = 810$) cohorts from Bergen, Norway, followed by sequential SNP testing in three independent cohorts (4). The investigators ultimately identified associations between COPD and SNPs at the *CHRNA3/5* (cholinergic nicotinic receptor) locus. Understanding the potential links between SNPs and tobacco smoking in the setting of COPD has been addressed in a few studies. Notably, genotyping in over 3,000 patients from four independent cohorts revealed that 24 SNPs with two specific loci, 2q21 and 6p21, were associated with age at initiation of smoking ($N = 3,397$). SNPs on 15q25 (*CHRNA3/5*) and four SNPs (rs7251570, rs4105144, rs1801272, and rs12461383) in the *CYP2A6* locus on chromosome 19q13 were associated with lifetime number of cigarettes smoked per day (5). Despite such encouraging studies, the clinical impact of genetic variants on COPD pathogenesis remains largely theoretical. Thus, many have advocated for a systems biology approach that uses gene sets or pathways and integrates GWASs with other platforms such as protein–protein interactions, termed dense module searching for genome-wide association studies (dmGWAS). Using this approach in established cohorts (COPDGene and GenKOLS), the investigators identified 10 candidate genes, including *KCNK3*, *NEDD4L*, and *RIN3*, by integrating GWAS and protein–protein interactions (3). More recent investigations have focused on elucidating the relationship between COPD gene variants, methylation, and gene expression. For example, Nedeljkovic and colleagues determined in a discovery cohort of 724 and replication cohort of 766 that a gene variant at 19q13.2 (rs7937) was associated with methylation sites within *EGLN2* and *EGLN2* gene expression in blood (6). The discovery of genetic variants that clear the path to functionally relevant targets could then be leveraged as disease-predictive and modifying biomarkers as well as eventual therapeutic targets in a disease that is currently lacking therapies that slow progression.

In this issue of the *Journal*, Benway and colleagues (pp. 92–102) take a natural next step in a GWAS to further identify gene variants within accessible chromatin regions in COPD but focused on clarifying cell-specific functional relevance (Figure 1) (7). Using the assay for transposase-accessible chromatin sequencing, the authors identified open chromatin regions (OCRs) in five human cell types (human bronchial epithelial cell lines [16HBE], primary human bronchial epithelial cells, small airway epithelial cells, alveolar type II pneumocytes [AT-II], and lung fibroblasts). After the identification of total OCRs in each of the cell types, the authors leverage their recent publication to isolate 7,285 variants with a suspected 20 causal SNPs (8). By then conducting an evaluation for both overlap and enrichment within the five cell types, the authors determined that overlap of 250 SNPs, including 22 that are predicted to impact regulatory elements, do in fact exist within OCRs but in a cell-specific manner with, for example, AT-II harboring the strongest enrichment.

A key to translating causal variants to clinical relevance lies in determining their functional impact. In this case, the authors have used machine learning (deltaSVM) and fine mapping as tools for predicting functional sequelae of variants in regulatory elements in the COPD GWAS loci credible sets and cell types. Through this approach, they were able to predict previously described unique variants of 4q31 upstream of *HHIP* in the various cell types, including rs13140176 in AT-II and 16HBE, several variants in 16HBE, and SNPs specific to primary human bronchial epithelial cells and AT-II (9). Similarly, a functional variant in the previously described 4q22.1 locus in *FAM13A* was predicted in 16HBE and lung fibroblast cells, whereas several additional SNPs were predicted to affect chromatin accessibility across all cell types (10). The authors go on to suggest that the deltaSVM model may predict additional causal variants in cases in which fine mapping is less accurate. Lastly, the authors focus in on AT-II and identify additional predicted functional variants within several known loci, including 22q12.3 (*SYN3*), 5q15 (in between *SPATA9* and *RHOBTB3*), 1p36.13 (*MFAP2*), and 5q35.1 (near *FGF18*).

The impetus for the current study lies in the paucity of data regarding functional variants within COPD and lack of lung cell–specific characterization. Here, the authors have leveraged previous data coupled with an integrated approach that incorporates the assay for transposase-accessible chromatin sequencing, fine mapping, and machine learning to predict causal variants within OCRs in several lung cell types. The result is a thought-provoking study that highlights AT-II cells as harboring multiple potential causal variants and opens the door for additional allied validation

This article is open access and distributed under the terms of the Creative Commons Attribution Non-Commercial No Derivatives License 4.0 (https://creativecommons.org/licenses/by-nc-nd/4.0/). For commercial usage and reprints, please contact Diane Gern (dgern@thoracic.org).

Originally Published in Press as DOI: 10.1165/rcmb.2021-0169ED on May 4, 2021

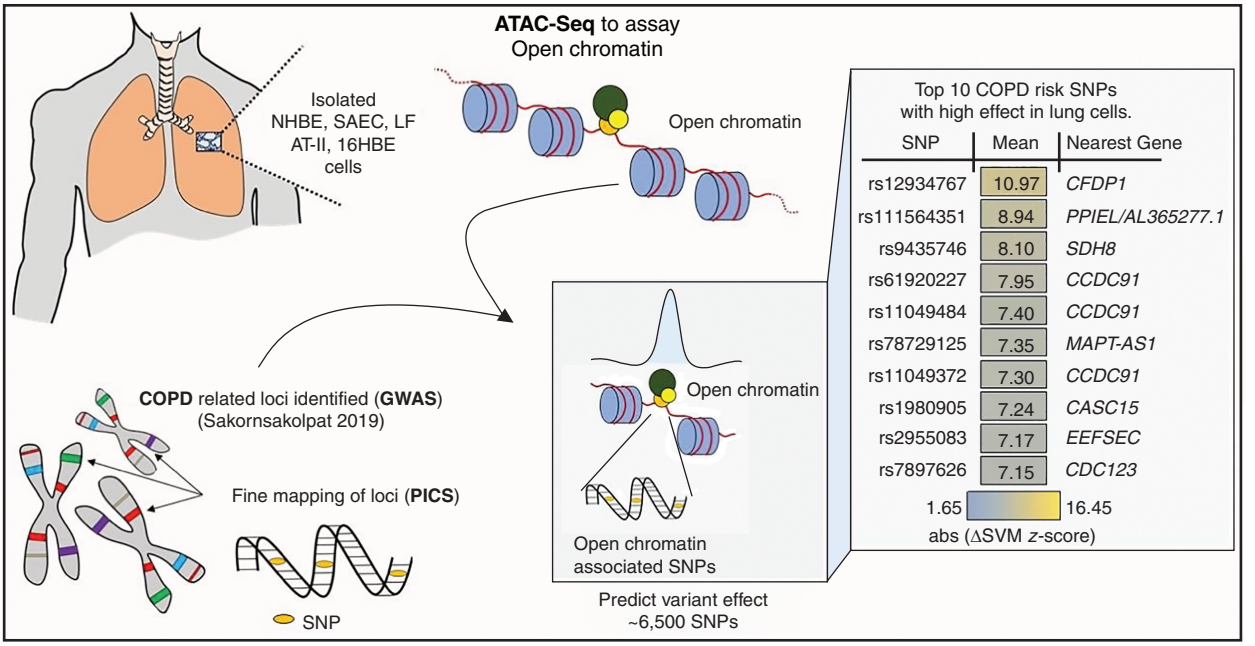

Figure 1. Overview of the work by Benway and colleagues (7). 16HBE = human bronchial epithelial cell lines; ATAC-Seq = assay for transposase-accessible chromatin sequencing; AT-II = alveolar type II pneumocytes; COPD = chronic obstructive pulmonary disease; GWAS = genome-wide association study; LF = lung fibroblasts; NHBE = primary human bronchial epithelial cells; PICS = Probabilistic Identification of Causal SNPs; SAEC = small airway epithelial cells.

studies. Importantly, the reported variants within accessible regulatory regions are predictive, and thus functional validation to determine the true biological effect is still required. Second, the authors by their own admission recognize that additional cell types within the lung as well as primary cells derived from patients with COPD should undergo similar evaluation for variants to determine clinical relevance and provide additional mapping. Benway and colleagues present a novel and timely study in their efforts to further define COPD causal variants within OCRs in a cell-specific manner. This represents one of several important steps toward further understanding the pathobiology of COPD with an eye toward developing novel therapies. ■

Author disclosures are available with the text of this article at www.atsjournals.org.

Mario Acunzo, Ph.D.
Patrick Nana-Sinkam, M.D.
Division of Pulmonary and Critical Care Medicine
Virginia Commonwealth University
Richmond, Virginia

References

1. Petersen H, Vazquez Guillamet R, Meek P, Sood A, Tesfaigzi Y. Early endotyping: a chance for intervention in chronic obstructive pulmonary disease. *Am J Respir Cell Mol Biol* 2018;59:13–17.
2. Hobbs BD, de Jong K, Lamontagne M, Bossé Y, Shrine N, Artigas MS, *et al.*; COPDGene Investigators; ECLIPSE Investigators; LifeLines Investigators; SPIROMICS Research Group; International COPD Genetics Network Investigators; UK BiLEVE Investigators; International COPD Genetics Consortium. Genetic loci associated with chronic obstructive pulmonary disease overlap with loci for lung function and pulmonary fibrosis. *Nat Genet* 2017;49:426–432.
3. McDonald ML, Mattheisen M, Cho MH, Liu YY, Harshfield B, Hersh CP, *et al.*; GenKOLS, COPDGene and ECLIPSE study investigators. Beyond GWAS in COPD: probing the landscape between gene-set associations, genome-wide associations and protein-protein interaction networks. *Hum Hered* 2014;78:131–139.
4. Pillai SG, Ge D, Zhu G, Kong X, Shianna KV, Need AC, *et al.*; ICGN Investigators. A genome-wide association study in chronic obstructive pulmonary disease (COPD): identification of two major susceptibility loci. *PLoS Genet* 2009;5:e1000421.
5. Siedlinski M, Cho MH, Bakke P, Gulsvik A, Lomas DA, Anderson W, *et al.*; COPDGene Investigators; ECLIPSE Investigators. Genome-wide association study of smoking behaviours in patients with COPD. *Thorax* 2011;66:894–902.
6. Nedeljkovic I, Lahousse L, Carnero-Montoro E, Faiz A, Vonk JM, de Jong K, *et al.* COPD GWAS variant at 19q13.2 in relation with DNA methylation and gene expression. *Hum Mol Genet* 2018;27:396–405.
7. Benway CJ, Liu J, Guo F, Du F, Randell SH, Cho MH, *et al.*; International COPD Genetics Consortium. Chromatin landscapes of human lung cells predict potentially functional chronic obstructive pulmonary disease genome-wide association study variants. *Am J Respir Cell Mol Biol* 2021;65:92–102.
8. Sakornsakolpat P, Prokopenko D, Lamontagne M, Reeve NF, Guyatt AL, Jackson VE, *et al.*; SpiroMeta Consortium; International COPD Genetics Consortium. Genetic landscape of chronic obstructive pulmonary disease identifies heterogeneous cell-type and phenotype associations. *Nat Genet* 2019;51:494–505.
9. Prokopenko D, Sakornsakolpat P, Fier HL, Qiao D, Parker MM, McDonald MN, *et al.* Whole-genome sequencing in severe chronic obstructive pulmonary disease. *Am J Respir Cell Mol Biol* 2018;59:614–622.
10. Sharma A, Kitsak M, Cho MH, Ameli A, Zhou X, Jiang Z, *et al.* Integration of molecular interactome and targeted interaction analysis to identify a COPD disease network module. *Sci Rep* 2018;8:14439.