REVIEW ARTICLE

Complex genetics of pulmonary diseases: lessons from genome-wide association studies and next-generation sequencing

NIMA POULADI, CHRISTIAN BIME, JOE G. N. GARCIA, and YVES A. LUSSIER

TUCSON, ARIZ AND CHICAGO, ILL

The advent of high-throughput technologies has provided exceptional assistance for lung scientists to discover novel genetic variants underlying the development and progression of complex lung diseases. However, the discovered variants thus far do not explain much of the estimated heritability of complex lung diseases. Here, we review the literature of successfully used genome-wide association studies (GWASs) and identified the polymorphisms that reproducibly underpin the susceptibility to various noncancerous complex lung diseases or affecting therapeutic responses. We also discuss the inherent limitations of GWAS approaches and how the use of next-generation sequencing technologies has furthered our understanding about the genetic determinants of these diseases. Next, we describe the contribution of the metagenomics in understanding the interactions of the airways microbiome with lung diseases. We then highlight the urgent need for new integrative genomics-phenomics methods to more effectively interrogate and understand multiple downstream "omics" (eg, chromatin modification patterns). Finally, we address the scarcity of genetic studies addressing under-represented populations such as African Americans and Hispanics. (Translational Research 2015;■:1–18)

Abbreviations: DNAseq = DNA sequencing; GWAS = genome-wide association study; NGS = next-generation sequencing

From the Department of Medicine, University of Arizona, Tucson, Ariz; Center for Biomedical Informatics and Biostatistics, University of Arizona, Tucson, Ariz; BIO5 Institute, University of Arizona, Tucson, Ariz; University of Arizona Health Sciences Center, University of Arizona, Tucson, Ariz; Arizona Respiratory Center, University of Arizona, Tucson, Ariz; Institute for Genomics and Systems Biology, Argonne National Laboratory and University of Chicago, Chicago, Ill.

Submitted for publication February 16, 2015; revision submitted April 27, 2015; accepted for publication April 29, 2015.

Reprint requests: Yves A. Lussier, BIO5 Institute, University of Arizona, 1657 East Helen Street, P.O. Box 210240, Tucson, AZ 85721; e-mail: yves@email.arizona.edu.

1931-5244/$ - see front matter

http://dx.doi.org/10.1016/j.trsl.2015.04.016

INTRODUCTION

High-throughput technologies have revolutionized our understanding of the molecular mechanisms underlying the development and progression of complex lung diseases. Importantly, genome-wide association studies (GWASs) have played a critical role as an advanced high-density genotyping approach that allows for characterizing the contribution of single-nucleotide polymorphisms (SNPs) scattered across the genome to the genetic susceptibility of individuals. Specifically, GWAS benefits from the variety of array platform technologies, by them the genotyping of numerous common variants is possible. Until the advent of GWAS in 2007, the genetic research of complex inheritance was labor-intensive and biased,

focusing on identifying genetic variants based on prior knowledge. GWAS offered a substantial higher rate of discovery and was not limited by the need for an a priori hypothesis or prior knowledge. However, the limited number of probes in GWAS platforms (eg, 1 million probes) compared with more than 79,000,000 variant sites thus far identified by phase 3 release of 1000 Genomes Project (June 24, 2014) indicates a remaining bias. Importantly, the rapidly reducing cost of next-generation DNA sequencing (DNAseq) paves the way for a paradigm shift in discovery: personal polymorphisms inaccessible by biased methods.

Previous reviews in lung diseases focused primarily on GWAS publications including descriptions of disease-SNP associations observed in only 1 dataset. The present study focused on replicated GWAS discoveries and new discoveries in whole genome sequencing of patients with a variety of lung diseases. We summarize the contribution of GWASs in identifying polymorphisms that confer risk to the development of specific pulmonary disorders and the GWASs that may influence patient responses to therapy. In addition, we have included a summary of next-generation sequencing (NGS) DNAseq findings benefiting from various NGS technologies. The advent of NGS has made unbiased sequencing of DNA faster and cheaper than that of Sanger sequencing method. We did not include the lung cancer polymorphisms, as the number of discoveries in that specialty would require an entire different review. Finally, we reflect on the potential considerations and avenues for future studies that would aim to catalog the genetic determinants of complex lung diseases.

METHODS

We used NHGRI GWAS catalog[1] to find the GWAS publications that are related to various primary noncancerous lung diseases; we found 81 curated publications. Among them, we focused on publications with results of discovery cohorts that have been separately and significantly replicated in their replication cohorts. The tables comprise an exhaustive listing of reproducible variants reported in the studied article. The authors might have mapped a certain polymorphism onto more than 1 gene. Therefore, the number of assigned genes might exceed the number of polymorphisms in several studies. We also used query terms of the various pulmonary diseases plus terms related to next-generation/whole exome/whole genome sequencing in PubMed to find the publications where NGS have been used. We used the original articles and the GeneCards (http://www.genecards.org/) to annotate the functions of the genes.

RESULTS

GWAS and complex pulmonary diseases. We divided the studies reviewed into GWAS and NGS categories with GWASs subdivided into 3 sections: obstructive pulmonary diseases (asthma and chronic obstructive pulmonary diseases [COPD]), restrictive pulmonary diseases, and miscellaneous lung diseases. We also provide a systems biology interpretation of discovered variants. The NHGRI GWAS catalog reports 207 lung disease–associated SNPs from 32 publications. Of these, 134 SNP-disease associations were reproducible in independent cohorts. However, we identified that only 61% of these SNPs were reported by the NHGRI with the correct identifier as originally published in the references. We also identified 13 reproducible lung disease SNPs in these NHGRI-reported publications that were missing in the NHGRI catalog. Here, we review the entirety of these and provide the rectified results in tables.

Obstructive pulmonary diseases. ***Asthma.*** Asthma is a common chronic inflammatory disorder of the airways, characterized by episodic and reversible airflow obstruction, airway hyper-responsiveness, and underlying inflammation.[2] In the United States, children, women, racial and ethnic minorities, residents of inner cities, and economically disadvantaged populations have a disproportionately higher burden of asthma morbidity and mortality when compared with the general population.[3-5] Among children younger than 18 years, asthma is the third major cause of the hospital admissions.[6] Of note, the pattern of its prevalence is significantly different with respect to various ages, countries with different economic infrastructure, and degree of severity.[6] Duffy et al estimated that asthma heritability is about 60%,[7] highlighting the underlying role of genetic determinants in asthma development.

Asthma susceptibility. Asthma genetic variants have been the most studied among pulmonary diseases. Moffatt et al[8] conducted the first asthma GWAS and uncovered the association of variants of ORMDL3, which downregulates the sphingolipid synthesis, with childhood-onset asthma (Table I). Himes et al subsequently performed a GWAS across asthma patients of various ethnicities[24] and discovered novel PDE4D polymorphism in combined populations of Hispanics and European descent but not in African American individuals (Table I). The authors also corroborated the association of the ORMDL3 variants with asthma. DeWan et al in a GWAS found evidence in favor of the association of variants of PDE11A gene with childhood allergic asthma.[25] Interestingly, both PDE11A and PDE4D genes belong to the phosphodiesterase superfamily of genes, implying the potential underlying role of this superfamily in the development of asthma.[25]