Candidate genes for COPD: current evidence and research

Woo Jin Kim1
Sang Do Lee2

1Department of Internal Medicine and Environmental Health Center, Kangwon National University, Chuncheon, 2Department of Pulmonary and Critical Care Medicine, Clinical Research Center for Chronic Obstructive Airway Diseases, Asan Medical Center, University of Ulsan College of Medicine, Seoul, South Korea

Abstract: COPD is a common complex disease characterized by progressive airflow limitation; the condition is not fully reversible or curable.1 The global burden of this disease is high, especially in countries with older populations, and this burden is expected to increase as the populace continues aging.2 Although environmental factors, including cigarette smoking and biofuel smoke exposure, are major risk factors, genetic risk factors are also important.3 In addition, an interaction between genetics and environment is believed to drive the development of COPD.

Pathways that play a role in COPD pathogenesis include the response to oxidative stress, the protease–antiprotease imbalance, cell death, and inflammation.4–7 Genetic studies have been performed to identify genetic risk factors and to understand the pathogenesis of COPD. Family-based studies and candidate gene association studies have found associations to many genes and loci. However, α-1 antitrypsin deficiency caused by mutations in SERPINA1 is the only established genetically driven cause of COPD that has a potential intervention so far.8

Although whole-genome and exome sequencing may be the next tools used for the genetic study of COPD, genome-wide association study (GWAS) is currently the most widely used method for the discovery of candidate genes. Several GWASs have discovered novel genes and pathways that are associated with COPD susceptibility. Even more genes have been found to be significantly associated with lung function in the general population. Some of these lung function genes are also associated with COPD susceptibility. The genetic basis of different COPD-related

Introduction

COPD is a common complex disease characterized by progressive airflow limitation; the condition is not fully reversible or curable.1 The global burden of this disease is high, especially in countries with older populations, and this burden is expected to increase as the populace continues aging.2 Although environmental factors, including cigarette smoking and biofuel smoke exposure, are major risk factors, genetic risk factors are also important.3 In addition, an interaction between genetics and environment is believed to drive the development of COPD.

Pathways that play a role in COPD pathogenesis include the response to oxidative stress, the protease–antiprotease imbalance, cell death, and inflammation.4–7 Genetic studies have been performed to identify genetic risk factors and to understand the pathogenesis of COPD. Family-based studies and candidate gene association studies have found associations to many genes and loci. However, α-1 antitrypsin deficiency caused by mutations in SERPINA1 is the only established genetically driven cause of COPD that has a potential intervention so far.8

Although whole-genome and exome sequencing may be the next tools used for the genetic study of COPD, genome-wide association study (GWAS) is currently the most widely used method for the discovery of candidate genes. Several GWASs have discovered novel genes and pathways that are associated with COPD susceptibility. Even more genes have been found to be significantly associated with lung function in the general population. Some of these lung function genes are also associated with COPD susceptibility. The genetic basis of different COPD-related

Keywords: chronic obstructive pulmonary disease, genetics, genome-wide association study
phenotypes, including emphysema and chronic bronchitis, also overlaps with that of COPD susceptibility. After being implicated in disease pathogenesis, these genes can be used as potential drug targets or as biomarkers that can influence diagnosis and personalized treatment.

Currently, the most well-known candidate genes for COPD are \( \text{CHRNA3/5} \) (cholinergic nicotine receptor alpha 3/5), \( \text{IREB2} \) (iron regulatory binding protein 2), \( \text{HHIP} \) (hedgehog-interacting protein), \( \text{FAM13A} \) (family with sequence similarity 13, member A), and \( \text{AGER} \) (advanced glycosylation end product–specific receptor). They have been replicated in multiple populations. None of them are targeted by treatments for COPD yet, and the mechanisms by which they alter COPD risk are still largely unknown. There is some emerging evidence that they may be good targets for treatments or useful as biomarkers. However, more study is required to understand the functional roles of these candidate genes.

Future research is needed to characterize the effect of genetic variants, validate gene function in humans and model systems, and elucidate the genes’ transcriptional and post-transcriptional regulatory mechanisms.\(^9\) In this review, we will cover the genetic and genomic association studies of COPD that identified the candidate genes and review what is currently known about those genes’ functions.

**CHRNA3, CHRNA5, and IREB2**

There are several genes at chromosome 15q25 that have been identified by GWAS for affecting COPD risk, including \( \text{CHRNA3} \), \( \text{CHRNA5} \), and \( \text{IREB2} \).\(^{10-12}\) The COPD cohorts investigated were the Norway case-control cohort (GenKOLS), the family-based ICGN cohort, the NETT (National Emphysema Treatment Trial)/NAS (normative aging study) cohorts, the Boston Early-Onset COPD cohort, and the COPDGene study cohort. The association between \( \text{CHRNA3/5} \) and COPD has been replicated in multiple ethnic populations by direct genotyping.\(^{13-15}\) The \( \text{CHRNA3/5} \) locus is associated with emphysema and smoking intensity in COPD.\(^{16,17}\) The \( \text{CHRNA3/5} \) region is also associated with lung cancer and nicotine addiction. It has been debated whether this common susceptibility region is the result of a common pathogenic pathway for lung cancer and COPD, or if it is simply associated with nicotine addiction, a risk factor for both diseases. In addition, the causal variant within the \( \text{CHRNA3/5} \) locus may be different in lung cancer than in COPD. There is some evidence that this locus has independent roles in the pathogenesis of COPD and smoking behavior.\(^{18}\)

\( \text{CHRNA3/5} \) and \( \text{CHRN} \) are subunits of the nicotine cholinergic receptor, and the cholinergic system is active not only in cholinergic neuronal cells, but also in bronchial epithelial cells and airway inflammatory cells. The proteins are responsive to nicotine and are upregulated during chronic tobacco exposure. A recent study integrating GWAS results with expression quantitative trait loci (eQTL) study results found that single nucleotide polymorphisms (SNPs) in the 15q25 region were associated with the expression of \( \text{IREB2} \) and \( \text{CHRNA3} \) in blood and sputum samples.\(^{19}\) \( \text{CHRNA3/5} \) and \( \text{IREB2} \) may play different roles in the pathogenesis of COPD.

\( \text{IREB2} \) was first identified by characterizing the differential gene expression in lung tissue between COPD patients and controls, and genotyping the SNPs within the candidate regions.\(^{20}\) \( \text{IREB2} \) is a protein that binds iron-responsive elements (IREs), maintains cellular iron metabolism, and is regulated in response to oxygen and iron supply. \( \text{IREB2} \) expression is higher in the lung tissue of COPD cases. The \( \text{Ireb} \)-knockout mouse has abnormal iron metabolism in the brain, which causes cellular dysfunction.\(^{21}\) However, the role of \( \text{IREB2} \) in COPD pathogenesis is still not known. A GWAS of the pulmonary artery measurement obtained by computed tomography (CT) in cohorts from the COPD-Gene study and the Evaluation of COPD Longitudinally to Identify Predictive Surrogate Endpoints (ECLIPSE) study found a genome-wide significant association to \( \text{IREB2} \).\(^{22}\) This suggests a role for \( \text{IREB2} \) in the pathogenesis of pulmonary hypertension in COPD, particularly in the vascular subtype.

**HHIP**

Many recently identified COPD-associated variants are located at chromosome 4q31, upstream of \( \text{HHIP} \). This intergenic region has been associated with COPD susceptibility in several GWASs\(^{10,11,12}\) and has consistently been replicated in multiple ethnicities.\(^{24-27}\) This region is also associated with lung function in the general population.\(^{28-30}\) The \( \text{HHIP} \) locus is associated with fat-free mass and exacerbations in COPD subjects.\(^{16} \) \( \text{HHIP} \) is associated with various CT phenotypes in COPD including distinct patterns of emphysema\(^7\) and the severity of emphysema.\(^{31} \) \( \text{HHIP} \) is more associated with emphysema measurements than with airway phenotypes and has a more significant association in emphysema subgroups.\(^{32}\) This difference may reflect a different pathogenic process driven by \( \text{HHIP} \), or may be driven by correlations between COPD status and imaging measurements. \( \text{HHIP} \) is also associated with adult height in the general population.\(^{33}\)
Considering that the FEV$\text{1}$ (forced expiratory volume in 1 second) prediction is determined by height, there may be genetic factors that control both phenotypes. Using the candidate gene strategy, it was found that the HHIP gene interacted with environmental tobacco smoke in utero, suggesting that this gene is involved in the lung response to smoke exposure in early life.$^{34}$

HHIP encodes a membrane glycoprotein that is an endogenous antagonist for the Hedgehog pathway. Hedgehog signaling is important for the morphogenesis of the lung and other organs.$^{35}$ Although the role of HHIP in COPD is not fully understood, several studies have validated the function of this gene in COPD pathogenesis. The associated SNPs are located upstream of HHIP, suggesting they may affect promoter activity. A lung eQTL study revealed that SNPs associated with COPD affect the expression of HHIP, and the risk allele of rs1828591 decreases its expression.$^{36}$ Zhou et al$^{37}$ reported that HHIP expression is reduced in COPD lung tissues and the genomic region upstream of HHIP interacts with the HHIP promoter. The risk allele of a variant in the HHIP enhancer region reduces promoter activity via a differential binding affinity to transcription factors.

These studies suggest that the genetic variation of the HHIP region affects the risk of COPD by affecting HHIP expression in lung tissues. HHIP silencing in an airway epithelial cell line leads to a change in gene expression, and these differentially expressed genes are enriched in pathways related to the extracellular matrix and cell growth, which are processes relevant to COPD pathogenesis.$^{38}$ Recently, Lao et al$^{39}$ found that Hhip-haploinsufficient mice have increased airspace size after cigarette smoke exposure, increased lung compliance, and increased numbers of lymphoid aggregates. The functions of the genes with altered expression in Hhip$^{-/-}$ mice exposed to cigarette smoke were enriched in the pathway of lymphocyte activation. They used haploinsufficient mice because Hhip$^{-/-}$ mice die shortly after birth due to lung branching morphogenesis failure.

HHIP was also found to be associated with lung cancer by a candidate gene study.$^{40}$ The Hedgehog pathway is a critical mediator of cigarette-smoke-induced lung cancer, and it may act as a common pathway for the development of COPD and lung cancer.$^{41}$

**FAM13A**

A GWAS using three COPD cohorts, GenKOLS, NETT/NAS, and the ECLIPSE study, identified variants at chromosome 4q22 in the gene *FAM13A*.$^{42}$ These are some of the most highly associated SNPs in COPD and are located in an intron. These associations have been replicated in a subset of the patients in the COPDGene study and the cohort of the International COPD Genetics Network. They have also been replicated in Asian populations, assayed using the candidate gene strategy.$^{43,44}$ *FAM13A* was first found to associate with lung function in a GWAS using the general population,$^{28}$ and it is associated with lung function in asthmatic subjects.$^{45}$ GWASs using COPD cases with chronic bronchitis in the COPDGene study, GenKOLS, and ECLIPSE cohorts identified a significant association with *FAM13A*,$^{46}$ whereas several GWASs for emphysema did not identify a genome-wide association. The odds ratios of *FAM13A* SNPs for COPD with chronic bronchitis were significantly higher than those for nonchronic bronchitis COPD, suggesting that *FAM13A* is more related to the pathogenesis of the chronic bronchitis subtype. Of note, *FAM13A* is also associated with idiopathic pulmonary disease,$^{47}$ but the expression of *FAM13A* in lung tissues does not differ by case/control status or by genotype.

*FAM13A* was originally identified in cattle near a quantitative trait locus affecting milk production, and is expressed in the kidney, pancreas, lung, and thymus.$^{48}$ Although the function of FAM13A has not been extensively studied, its RhoGAP domain may be related to COPD. Rho GTPases are key regulators of cytoskeletal dynamics, are involved in the pulmonary endothelial barrier, and are dysregulated in several lung diseases.$^{49}$ A lung eQTL study suggested that the expression of *FAM13A* may be associated with particular SNPs.$^{50}$ In the case of COPD, the FAM13A risk allele is associated with increased *FAM13A* expression in the lung, although expression does not differ in lung tissues between COPD cases and controls.$^{51}$ A recent study by Jin et al$^{52}$ found that FAM13A activates Wnt signaling by increasing the stability of β-catenin. Although depletion of FAM13A in a lung cancer cell line reduces Wnt signaling activity, *FAM13A*-knockout mice are viable and *FAM13A*-mutant lungs are morphologically indistinguishable from wild-type lungs, and Wnt signaling remains normal in Fam13a-knockout lungs. They also found that Akt regulates the phosphorylation of FAM13A, which can lead to cytoplasmic sequestration of FAM13A. Considering that Akt has a role in the pathogenesis of COPD,$^{51}$ FAM13A may contribute to lung disease through aberrant Akt signaling. Further work is needed to validate the functional role of FAM13A in the pathogenesis of COPD.

**AGER**

GWASs of lung function in the general population have found that chromosome 6p21 is associated with FEV$\text{1}$/FVC.
(forced vital capacity) and FEV$_1$, which are important physiologic parameters of COPD. This association was investigated in COPD patients identified from the population cohort using spirometry criteria, and the study found a suggestive association between COPD risk and $AGER$, although it was not statistically significant. A candidate gene study in NETT/NAS, GenKOLS, ECLIPSE, and a subset of the COPDGene study cohort found that it was associated with COPD susceptibility, although a subsequent GWAS did not find a significant association. On the other hand, an association has been found in multiple ethnic populations.

The 6p21 region that showed significant association with COPD includes many genes, including $TNXB$, $PPT2$, $AGER$, and $NOTCH4$. However, $AGER$ has a potential functional variant, rs2070600, and it has been studied the most in the pathogenesis of COPD. A GWAS of percent emphysema determined by CT using the Multi-Ethnic Study of Atherosclerosis cohort identified a significant association with the $AGER$/PPT region. This region did associate with emphysema severity and gas trapping in a GWAS using cohorts from the COPDGene, ECLIPSE, NETT, and GenKOLS studies.

The protein product of $AGER$, the receptor for advanced glycan end products (RAGE), is a multiligand receptor of the immunoglobulin superfamily and interacts with molecules implicated in homeostatic function, inflammation, and development. RAGE levels are increased in the airway and alveolar walls of COPD lungs. RAGE expression in mice increases after cigarette smoke exposure, and cigarette-smoking-induced inflammatory responses by alveolar macrophages are diminished in RAGE-knockout mice. Transgenic mice with upregulated RAGE have impaired alveolar morphogenesis during lung development, distal airspace enlargement, and increased alveolar cell apoptosis. Another study using RAGE transgenic mice found incremental dilatation of alveolar spaces, as well as pronounced inflammation in the peripheral lung and alveolar destabilization. A promoter variant of $AGER$ in cystic fibrosis patients is associated with poor lung function, and it increases expression in airway epithelial cell lines, suggesting it is a modifier of lung disease severity.

The soluble isoform of RAGE (sRAGE) contains the RAGE extracellular domain and can bind to circulating proinflammatory ligands, preventing RAGE activation. Mice that are exposed to chronic hypoxia have downregulated pulmonary RAGE protein and increased levels of sRAGE, which might be adaptive to and protective against chronic hypoxia. Circulatory levels of sRAGE are reduced in COPD patients. Reduced sRAGE levels are associated with increased emphysema in two COPD cohorts. Decreased plasma sRAGE levels are also associated with the progression of airflow limitation over time. In patients of the Treatment of Emphysema with a Selective Retinoid Agonist (TESRA) and ECLIPSE studies, sRAGE is associated with diffusing capacity, emphysema, and COPD disease status, and the variant rs2070600 is associated with circulating sRAGE levels. The significant association between rs2070600 and plasma sRAGE levels was also found in Dutch diabetes mellitus and control subjects. RAGE has been studied in metabolic diseases, and decreased levels of sRAGE are linked to vascular complications. RAGE contributes to the pathogenesis of COPD in the lung probably via the regulation of inflammation and apoptosis, and further study of the functions of this gene may lead to it being identified as a potential therapeutic target.

Other candidate genes

There have been several more regions identified in GWASs of COPD. A GWAS using subjects from the ECLIPSE, NETT/NAS, GenKOLS, and COPDGene studies identified chromosome 19q13 as being associated with COPD, along with the previously identified HHIP, FAM13A, and 15q25 regions. Chromosome 19q13 contains $CYP2A6$, $RAB4B$, $MIA$, and $EGLN$, which could potentially be involved in COPD pathogenesis, and EGLN2 was found to be dysregulated in the airway epithelium of smokers. A GWAS using the full COPDGene cohort identified additional associations with $TGFB2$, $MMP12$, and $RIN3$. TGFB2 and MMP12 have been previously studied in COPD or related phenotypes, whereas RIN3 has not been studied in COPD and needs to be investigated further. SERPINE2 was identified using a linkage analysis of gene expression changes in lung tissue. A recent GWAS of airway thickness identified rs734556 on chromosome 2q, which is associated with SERPINE2 expression.

An association study of SNP in $GSTDS$ on chromosome 4q24, which was originally shown to be associated with lung function in general population, was performed in COPD case and control subjects. Variant in GSTDS was significantly associated with COPD susceptibility. SNPs in this region were also tested in GenKOLS, NETT, ECLIPSE, and part of COPDGene, and association with COPD susceptibility was significant. $GSTCD$ and $INTS12$ genes in 4q24 are expressed in human lung tissues and airway cells and expression of these genes was altered by TGFβ1 in human airway smooth muscle cells. Top SNPs associated with airflow obstruction defined by FEV$_1$ and FEV$_1$/FVC less than lower limits of normal were tested in the COPD meta-analysis. In this analysis, $ADAM19$ on
chromosome 5q33 was nominally associated with COPD.\textsuperscript{12} This gene was also associated with FEV\textsubscript{i}/FVC in the CHARGE consortium.\textsuperscript{28} This region was nominally associated with COPD susceptibility.\textsuperscript{54} A GWAS of emphysema in the three COPD cohorts including the Norway cohort, ECLIPSE, and NETT revealed that \textit{BICD1} was the most associated with the presence or absence of emphysema.\textsuperscript{76} The emphysema-predominant mild COPD of COPDGene study was nominally associated with \textit{BICD1}.

These associations require more replications and further fine-mapping studies are needed to find the causal variants of COPD, as well as studies to functionally validate the identified genes.

**Future directions**

Subtyping or classification, drug development, and risk prediction are potential clinical applications of the associated genes. Phenotype heterogeneity can be investigated using association studies on various COPD-related phenotypes. Although the associations merely reflect the correlations between COPD status and phenotypes, there may be subtypes with different genetic alterations. Considering that treatment with PDE4 inhibitor is effective only for the chronic bronchitis subtype, there may be a mechanism that is unique to this subtype. Candidate genes can be used to determine personalized treatment because they may help identify a subtype-unique pathogenesis, variation in a drug-action site, or variable drug metabolism.\textsuperscript{76}

More regions have been identified in GWASs on FEV\textsubscript{i} and FEV\textsubscript{i}/FVC in the general population, probably because of the larger sample sizes. GWASs using a greater sample size of COPD subjects may find more candidate genes.\textsuperscript{77} Another approach for finding more candidate genes is to identify rare variation using exome sequencing or arrays.

Functional studies for COPD candidate genes are summarized in Table 1. Although there are a few potential drug targets that are implicated in the novel COPD pathogenesis, it will take many years to apply them in the treatment.

**Conclusion**

Recently, several candidate genes associated with COPD risk have been identified using GWAS. Replication and functional validation studies may lead to clinical applications for these genes such as novel therapeutics, subtyping, and risk prediction for COPD.

**Disclosure**

The authors report no conflicts of interest in this work.

**References**


**Table 1 Current candidate genes for COPD and their potential functional mechanisms in the COPD pathogenesis**

<table>
<thead>
<tr>
<th>Genes</th>
<th>Chromosomal loci</th>
<th>Functional studies</th>
</tr>
</thead>
<tbody>
<tr>
<td>CHRNA3/5</td>
<td>1q525</td>
<td>Responsive to nicotine; upregulated during tobacco exposure</td>
</tr>
<tr>
<td>IREB2</td>
<td>1q525</td>
<td>Higher expression in COPD lung tissue; knock out causes cellular dysfunction</td>
</tr>
<tr>
<td>HHIP</td>
<td>4q31</td>
<td>Reduced expression in COPD lung tissue; regulate pathway related to the extracellular matrix</td>
</tr>
<tr>
<td>FAM13A</td>
<td>4q22</td>
<td>Activate Wnt signaling; Akt regulates cytoplasmic sequestration of FAM13A</td>
</tr>
<tr>
<td>AGER</td>
<td>6p21</td>
<td>Impair alveolar morphogenesis; increase alveolar cell apoptosis; Alveolar destabilization</td>
</tr>
</tbody>
</table>

**Abbreviations:** CHRNA3/5, cholinergic nicotine receptor alpha 3/5; IREB2, iron regulatory binding protein 2; HHIP, hedgehog-interacting protein; FAM13A, family with sequence similarity 13, member A; AGER, advanced glycosylation end product-specific receptor.


Eur Respir J. 2010;36:89–95.


Respir Res. 2015;16:49.


Am J Respir Crit Care Med. 2015;192:559–569.


35. Chuang PT, Kawcak T, McMahon AP. Feedback control of mammalian Hedgehog signaling by the Hedgehog-binding protein, Hip1, modulates Fgf signaling during branch morphogenesis of the lung. 


Genome Med. 2015;7:12.


47. Fingerlin TE, Murphy E, Zhang W, et al. Genome-wide association study identifies multiple susceptibility loci for pulmonary fibrosis. 


50. Corvol H, Hodges CA, Drumm ML, et al. Moving beyond genetics: is the type 1 diabetes-associated

51. Jin Z, Chung JW, Mei W, et al. Regulation of nuclear-cytoplasmic shuttling and function of Family with sequence similarity 13, member A (Fam13a), by B56-containing PP2As and Akt. 


