

Exome Sequencing Analysis in Severe, Early-Onset Chronic Obstructive Pulmonary Disease

Dandi Qiao¹, Christoph Lange², Terri H. Beaty³, James D. Crapo⁴, Kathleen C. Barnes⁵, Michael Bamshad⁶, Craig P. Hersh^{1,7}, Jarrett Morrow¹, Victor M. Pinto-Plata⁸, Nathaniel Marchetti⁹, Raphael Bueno¹⁰, Bartolome R. Celli⁷, Gerald J. Criner¹¹, Edwin K. Silverman^{1,7}, and Michael H. Cho^{1,7}; Lung GO, NHLBI Exome Sequencing Project, and COPDGene Investigators

¹Channing Division of Network Medicine, ⁷Division of Pulmonary and Critical Care Medicine, and ¹⁰Division of Thoracic Surgery, Department of Medicine, Brigham and Women's Hospital and Harvard Medical School, Boston, Massachusetts; ²Department of Biostatistics, Harvard School of Public Health, Boston, Massachusetts; ³Johns Hopkins Bloomberg School of Public Health, and ⁵Division of Allergy and Clinical Immunology, Department of Medicine, Johns Hopkins University, Baltimore, Maryland; ⁴National Jewish Health, Denver, Colorado; ⁶Division of Genetic Medicine, Department of Pediatrics, University of Washington and Seattle Children's Hospital, Seattle, Washington; ⁸Department of Critical Care Medicine and Pulmonary Disease, Baystate Medical Center, Springfield, Massachusetts; ⁹Department of Thoracic Medicine and Surgery, and ¹¹Division of Pulmonary and Critical Care Medicine Temple University School of Medicine, Philadelphia, Pennsylvania

Abstract

Rationale: Genomic regions identified by genome-wide association studies explain only a small fraction of heritability for chronic obstructive pulmonary disease (COPD). Alpha-1 antitrypsin deficiency shows that rare coding variants of large effect also influence COPD susceptibility. We hypothesized that exome sequencing in families identified through a proband with severe, early-onset COPD would identify additional rare genetic determinants of large effect.

Objectives: To identify rare genetic determinants of severe COPD.

Methods: We applied filtering approaches to identify potential causal variants for COPD in whole exomes from 347 subjects in 49 extended pedigrees from the Boston Early-Onset COPD Study. We assessed the power of this approach under different levels of genetic heterogeneity using simulations. We tested genes identified in these families using gene-based association tests in exomes of 204

cases with severe COPD and 195 resistant smokers from the COPDGene study. In addition, we examined previously described loci associated with COPD using these datasets.

Measurements and Main Results: We identified 69 genes with predicted deleterious nonsynonymous, stop, or splice variants that segregated with severe COPD in at least two pedigrees. Four genes (*DNAH8*, *ALCAM*, *RARS*, and *GBF1*) also demonstrated an increase in rare nonsynonymous, stop, and/or splice mutations in cases compared with resistant smokers from the COPDGene study; however, these results were not statistically significant. We demonstrate the limitations of the power of this approach under genetic heterogeneity through simulation.

Conclusions: Rare deleterious coding variants may increase risk for COPD, but multiple genes likely contribute to COPD susceptibility.

Keywords: genetic association studies; segregation analysis; chronic obstructive pulmonary disease

(Received in original form June 25, 2015; accepted in final form January 6, 2016)

Supported by NHLBI grants R01 HL084323, P01 HL083069, P01 HL105339, R01 HL075478 and R01 HL089856 (E.K.S.); K08 HL097029 and R01 HL113264 (M.H.C.); K01 HL129039 (D.Q.); and R01 HL089897 (J.D.C.) and by the Alpha-1 Foundation (M.H.C.); funding for GO Exome Sequencing Program was provided by NHLBI grants RC2 HL-103010 (HeartGO), RC2 HL-102923 (LungGO), and RC2 HL-102924 (WHISP); the exome sequencing was performed through NHLBI grants RC2 HL-102925 (BroadGO) and RC2 HL-102926 (SeattleGO). Sequencing for the Boston Early-Onset COPD Study was provided by the University of Washington Center for Mendelian Genomics and was funded by the National Human Genome Research Institute and NHLBI grant 1 U54HG006493 to Drs. Debbie Nickerson, Jay Shendure, and Michael Bamshad. Sequencing for the COPDGene subjects was provided through the NHLBI Exome Sequencing Program (RC2HL102923 [LungGO], RC2HL102926 [SeattleGO]), with additional support through the COPD Foundation. The COPDGene study (NCT00608764) is also supported by the COPD Foundation through contributions made to an Industry Advisory Board comprised of AstraZeneca, Boehringer Ingelheim, Novartis, Pfizer, GlaxoSmithKline, Siemens, and Sunovion.

Author Contributions: Study design: C.L., E.K.S., and M.H.C.; data collection: T.H.B., J.D.C., K.C.B., M.B., C.P.H., J.M., V.M.P.-P., N.M., R.B., B.R.C., G.J.C., E.K.S., and M.H.C.; data quality control and analysis: D.Q. and M.H.C.; statistical support: D.Q., C.L., and M.H.C.; and manuscript writing: D.Q., E.K.S., and M.H.C. All authors revised the manuscript.

Correspondence and requests for reprints should be addressed to Dandi Qiao, Ph.D., Channing Division of Network Medicine, 181 Longwood Avenue, Boston, MA 02115. E-mail: redaq@channing.harvard.edu

This article has an online supplement, which is accessible from this issue's table of contents at www.atsjournals.org

Am J Respir Crit Care Med Vol 193, Iss 12, pp 1353–1363, Jun 15, 2016

Copyright © 2016 by the American Thoracic Society

Originally Published in Press as DOI: 10.1164/rccm.201506-1223OC on January 6, 2016

Internet address: www.atsjournals.org

At a Glance Commentary

Scientific Knowledge on the Subject

Subject: Chronic obstructive pulmonary disease (COPD) is a complex disease, and only a small fraction of heritability can be explained by previously identified genomic loci. Alpha-1 antitrypsin deficiency shows that rare coding variants of large effect can contribute to COPD susceptibility.

What This Study Adds to the Field

Field: We identified candidate genes from exome sequencing in 49 families with severe early-onset COPD and sought supportive evidence of association in 400 unrelated subjects in the COPDGene study. Our results failed to find convincing evidence for a single gene or variant, and, in conjunction with simulations, suggest that although rare variants may contribute to the susceptibility to severe COPD, it is unlikely that the effect can be attributed to only a few genes.

Chronic obstructive pulmonary disease (COPD) is one of the leading causes of death in the United States and worldwide (1). Although the major environmental risk factor, exposure to cigarette smoking, is well known, the response to cigarette smoke is highly variable, and a substantial portion of susceptibility to COPD is due to genetic factors (2). Genome-wide association studies (GWASs) have successfully identified several genomic regions influencing the risk for COPD (3, 4), but, similar to many other common complex diseases, only a small percentage of COPD heritability can be explained by these regions (5). Although extensive work has been performed identifying common variants associated with COPD, few studies have focused on rare variants. However, severe alpha-1 antitrypsin deficiency, a disorder discovered over 50 years ago and found to remarkably increase the risk of COPD, demonstrates that rare coding variants of large effect also influence COPD susceptibility.

The availability of large-scale DNA sequencing at low cost has made genome-wide investigation of the human coding genome (whole-exome sequencing) a

feasible study design. Several studies have been successful in identifying disease-causing variants for complex diseases using association-based tests for exome sequencing data. However, such success relied on large sample sizes (6–8). An alternative approach, used in Mendelian diseases of high penetrance, is to use a direct filtering strategy in a number of unrelated affected individuals or in a few families to identify rare or novel variants in the same gene that exist only in the cases and not in the control subjects. Many studies have identified and validated the candidate genes/variants that were discovered using this approach (9–12). Here, we hypothesized that additional rare functional coding variants influence the development of COPD. Because subjects with severe disease may be enriched for genetic determinants of COPD, our discovery efforts focused on identifying rare, functional variants that segregated with disease in families identified through probands with severe, early-onset disease. We sought additional evidence of association in a set of severely affected cases and resistant smoking control subjects from the COPDGene Study. Some of the results of this study have been previously reported in the form of abstracts (13, 14).

Methods

Details of the EOCOPD (Boston Early-Onset COPD) study have been described previously (15). Briefly, probands were aged 53 years or younger with prebronchodilator

FEV₁ of 40% predicted or lower, physician-diagnosed COPD, and without severe alpha-1 antitrypsin deficiency (16). All first-degree relatives, older second-degree relatives, and additional affected family members were eligible for enrollment. We selected 49 of the largest pedigrees with at least two individuals affected with COPD for whole-exome sequencing analysis. Details of the COPDGene Study (17), a multicenter epidemiologic and genetic study of 10,192 smokers, have also been described previously. Severe (GOLD [Global Initiative for Chronic Obstructive Lung Disease] spirometry grade 3 or 4 with post-bronchodilator FEV₁ < 50% predicted and FEV₁/FVC < 0.7) cases were under 63 years old and had substantial emphysema by quantitative computed tomography scan (>15% at –950 HU). Resistant smoking control subjects were frequency matched on pack-years of cigarette smoking, with normal lung function, age greater than 65 years, and no significant emphysema (<5% at –950 HU). Sequencing for both cohorts was performed at the University of Washington (Seattle, WA), using NimbleGen V2 capture (Roche NimbleGen, Inc., Madison, WI) and the Illumina platform (Illumina, Inc., San Diego, CA). A subset of the COPDGene subjects was sequenced via the NHLBI Exome Sequencing Program, and the EOCOPD subjects were sequenced as part of the Center for Mendelian Genomics.

We used the extended pedigrees to identify rare, predicted deleterious variants segregating with severe (GOLD grade 3 or 4)

Table 1. Baseline Characteristics of 347 Individuals Chosen from the Boston Early-Onset Chronic Obstructive Pulmonary Disease Study for Whole-Exome Sequencing

Characteristics	GOLD 2–4*	GOLD 1†	Normal Spirometry‡	PRISm
N	155	16	148	28
Age, yr	56 (14)	52 (14)	42 (14)	53 (16)
Pack-years	43 (0–133)	32 (0–90)	14 (0–163)	30 (0–136)
FEV ₁ % predicted§	41 (11–78)	86 (80–105)	97 (80–127)	73 (58–79)
FEV ₁ /FVC ratio§	44 (21–69)	64 (60–69)	79 (70–92)	77 (71–87)

Definition of abbreviations: GOLD = Global Initiative for Chronic Obstructive Lung Disease; PRISm = preserved ratio impaired spirometry.

PRISm (55): FEV₁ < 80% predicted and FEV₁/FVC ≥ 0.7. Mean (SE) are presented for age, and mean (range) are presented for pack-years, FEV₁% predicted, and FEV₁/FVC ratio for each phenotype group.

*GOLD 2–4: FEV₁ < 80% predicted and FEV₁/FVC < 0.7.

†GOLD 1: FEV₁ ≥ 80% predicted and FEV₁/FVC < 0.7.

‡Normal spirometry: FEV₁ ≥ 80% predicted and FEV₁/FVC ≥ 0.7.

§Maximum of pre- and post-bronchodilator values.

Table 2. Genes with Rare Deleterious Variants Segregating in at Least Two Pedigrees

Gene	No. Var.	No. Seg. (Novel)	Variant Type
<i>STRN4</i>	5	2 (0)	2 NS
<i>PDE3B</i>	6	2 (1)	1 stop gained, 1 NS
<i>HSPA5</i>	7	2 (0)	2 NS
<i>ZUFSP</i>	7	2 (1)	1 stop gained, 1 NS
<i>TKT</i>	9	2 (1)	1 splice site donor, 1 NS
<i>POLA1</i>	9	2 (2)	2 NS
<i>ADAMTS1</i>	11	2 (0)	2 NS
<i>GBF1</i>	14	2 (0)	2 NS
<i>RPAP1</i>	17	2 (0)	2 NS
<i>PLCB1</i>	17	2 (0)	2 NS
<i>TPR</i>	19	2 (1)	2 NS
<i>KIAA1274</i>	21	2 (0)	2 NS
<i>ARHGEF17</i>	22	2 (1)	2 NS
<i>PLXNA1</i>	23	2 (0)	2 NS
<i>ODZ3</i>	25	2 (0)	2 NS
<i>TULP4</i>	25	2 (0)	2 NS
<i>ZFYVE26</i>	27	2 (1)	2 NS
<i>CUL9</i>	30	2 (0)	2 NS
<i>PLXNA4</i>	35	3 (0)	1 stop gained, 2 NS
<i>HERC1</i>	38	3 (1)	3 NS
<i>EVPL</i>	42	2 (0)	2 NS
<i>DNAH8</i>	55	2 (1)	2 NS
<i>LRP1B</i>	55	2 (0)	2 NS
<i>RYR3</i>	56	2 (1)	2 NS
<i>PDCD2L</i>	3	2 (1)	1 stop gained, 1 NS
<i>DIABLO</i>	5	2 (0)	2 NS
<i>CAT</i>	5	2 (0)	2 NS
<i>KDEL2C</i>	6	2 (1)	2 NS
<i>PALM2</i>	6	2 (0)	2 NS
<i>P2RX5</i>	6	2 (1)	1 stop gained, 1 NS
<i>MAP7D2</i>	6	2 (2)	2 NS
<i>SEZ6L2</i>	8	2 (1)	1 splice site acceptor, 1 NS
<i>ARSK</i>	8	2 (1)	2 NS
<i>BMP5</i>	8	2 (0)	2 NS
<i>OR9G4</i>	10	2 (0)	2 NS
<i>ALCAM</i>	10	2 (1)	1 start lost, 1 NS
<i>RHBDF2</i>	10	2 (0)	2 NS
<i>ILF3</i>	10	2 (2)	2 NS
<i>AC093158.1</i>	10	2 (0)	2 NS
<i>RXFP1</i>	10	2 (1)	2 NS
<i>BCAR3</i>	10	2 (0)	2 NS
<i>RABGGTA</i>	11	2 (1)	2 NS
<i>CCDC19</i>	12	2 (0)	2 NS
<i>CACNA1F</i>	13	2 (2)	2 NS
<i>RARS</i>	14	2 (1)	2 NS
<i>CYP11B1</i>	14	2 (0)	2 NS
<i>BAI1</i>	15	2 (1)	2 NS
<i>LRGUK</i>	15	2 (0)	2 NS
<i>KIAA1239</i>	21	2 (1)	2 NS
<i>EP300</i>	23	2 (0)	2 NS
<i>NEDD4</i>	25	2 (0)	2 NS
<i>SCN7A</i>	25	2 (0)	2 NS
<i>POLR1A</i>	25	2 (1)	2 NS
<i>LAMB3</i>	28	2 (0)	2 NS
<i>HEATR8</i>	29	2 (0)	2 NS
<i>NPHP4</i>	32	2 (1)	2 NS
<i>CHTF18</i>	33	2 (0)	2 NS
<i>FLNB</i>	35	2 (0)	2 NS
<i>ODZ4</i>	37	2 (1)	2 NS
<i>PCDH15</i>	39	2 (0)	2 NS
<i>PKHD1</i>	55	2 (1)	2 NS
<i>TTN</i>	373	2 (0)	2 NS
<i>MLYCD</i>	6	2 (0)	2 NS
<i>FMNL1</i>	9	2 (0)	2 NS

(Continued)

COPD, and absent in smoking control subjects (normal spirometry, age >40 yr, with at least 5 pack-years of cigarette smoking). Segregating variants were stop, splice, or nonsynonymous variants characterized as deleterious by the Combined Annotation-Dependent Depletion score (18) and Condel (19) with minor allele frequency (MAF) of less than 0.1% in public datasets. We used simulations to assess the power of this approach under different levels of genetic heterogeneity. In addition to our segregation analysis, we performed a secondary analysis using a recently described gene discovery method, pVAAST (version 2.1.6) (20), that attempts to combine rare variant association with linkage to discover disease genes.

To attempt to validate associations identified by our approach, we performed gene-based association tests in severe cases versus resistant smoking control subjects from COPDGene. Our primary analysis was a burden test on affection status for variants with MAF less than 5% using permutation-based *P* values in ScoreSeq (21), adjusting for sex and ancestry. We considered evidence of replication if either: (1) there was an enrichment of segregating genes with consistent direction of effect and *P* value less than 0.05 using the burden test (equivalent to a one-sided *P* <0.025); or (2) if any individual gene met significance after Bonferroni correction (corrected *P* <0.05). To determine if our top genes demonstrated enrichment in any functional categories, we performed analyses using TopGO and STRING (7, 22, 23). We additionally assessed lung tissue expression of segregating genes in patients with severe COPD and with normal spirometry. Finally, we examined genes near loci previously reported in association with COPD or lung function in GWAS or in Mendelian diseases related to COPD for evidence of either segregation in our family-based study or association in our case-control study. Additional analyses and methods are available in the online supplement.

Results

Variant Filtering Analysis in the Boston EOCOPD Study

Of 351 total subjects, 4 were excluded based on pedigree, racial, or sex mismatches. Baseline characteristics of the analyzed

Table 2. (Continued)

Gene	No. Var.	No. Seg. (Novel)	Variant Type
COX10	11	2 (0)	2 NS
PLD1	13	2 (0)	2 NS
NID1	25	2 (2)	2 NS
PLEKHG1	25	2 (0)	2 NS
CFTR	27	2 (0)	2 NS

Definition of abbreviations: No. Seg. = number of pedigrees harboring a segregating variant; NS = nonsynonymous; No. Var. = total number of variants.

No. Var. is the total number of variants in the gene meeting our primary filtering criteria (minor allele frequency <0.1%, Combined Annotation-Dependent Depletion and Condel scores). No. Seg. (Novel) is the number of pedigrees harboring a variant with a segregating variant, with the number of segregating novel variants in the parenthesis. Variant type indicates the functional impact of the segregating variants, defined by the genetic variant annotation and effect prediction toolbox SnpEff (56); shown are the number of variants with the specified impact and that segregate in at least one pedigree.

347 Boston EOCOPD Study individuals are shown in Table 1. After quality control, 124,288 substitution variants remained for analysis. Out of the 347 sequenced subjects, we identified 107 affected subjects with severe and very severe COPD; we also identified 34 unaffected current or former smokers with normal lung function.

A total of 10,144 rare (MAF < 0.1%) predicted functional (nonsynonymous, stop, or splice) variants were assessed for segregation in EOCOPD pedigrees. Using a dominant model, 1,164 variants in 1,081 genes segregated with COPD affection status in at least one family, of which 261 variants in 256 genes were private (not previously reported, and seen in only one family). Among these 1,081 genes, 2 segregated in 3 different families, and 67 segregated in 2 different families. Among the 69 genes that segregated in at least 2 families, 5 of the genes contained only segregating novel variants (not present in any public database used in estimating the minor allele frequencies and also not present in the ExAC database [24]; see Table 2). A total of 5 of the 69 genes included 1 stop variant that segregated in a family; no genes had more than 1 segregating stop variant. A flow chart of our analysis is shown in Figure 1, and a list of the segregating genes can be found in Table 2.

No genes had variants that segregated in at least two families with an autosomal recessive model using our filtering criteria. Additional analyses using a less-stringent MAF threshold identified nine genes with variants of allele frequency less than 5% segregating in a recessive model (see the online supplement). With a compound heterozygous model, three genes with

variants of allele frequency less than 1% segregated in at least two families using our filtering criteria (see the online supplement). Gene set enrichment tests using TopGO (22) and enrichment tests on the 69 genes within the protein–protein interaction network using STRING (7) did not identify significant enrichment (see the online supplement). In addition to our segregation analysis, we explored alternative methods of identifying novel disease genes in our pedigrees. An analysis with pVAAST (version 2.1.6) (20) identified three genes that were statistically significant after Bonferroni correction (*SAA2*, *SAA4*, and *LIMS1*); however, none of these had supportive evidence of association in COPDGene. Of the top genes ($P < 1 \times 10^{-4}$) identified in this analysis, only one overlapped with genes from our segregation analysis (*RXFP1*; see the online supplement).

Power of Segregation Analyses under Genetic Heterogeneity

Using simulations, we assessed the power of our approach to identify highly penetrant variants under different levels of genetic heterogeneity. Details of the simulations can be found in the online supplement. Figure 2 shows the probability of observing at least two (Figure 2A) and four (Figure 2B) segregating families for different numbers of causal genes. As expected, the probability goes down dramatically as the number of causal genes increases. In a one-parent/one-offspring (P-1O) scenario, the probability of detecting any one of the causal genes drops to 68% for 25 causal genes; the probability is even smaller for families with one parent and three offspring.

However, the probability of observing a null gene segregating in at least two families (i.e., false positive) for P-1O in this scenario is 7.9%. With a larger number of variants and higher MAFs, this probability goes up, which means there would be more false positives for larger genes with greater variation (more variants with larger MAFs) using the segregation criteria.

We also plot the probability of observing (at least) n segregating P-1O families for different numbers of causal genes (Figure 2c) where n is 1–10. We can see that with 10 (or fewer) causal genes, we would have relatively high probability to observe one causal gene segregating in at least four families (75%). With 25 or more causal genes, this probability drops dramatically, and we are unlikely to observe at least four families harboring segregating variants in the same gene. These simulations suggest that our results are consistent with at least a moderate level of genetic heterogeneity of severe COPD.

Association Analysis in COPDGene

In the COPDGene dataset, a total of 204 severe cases and 195 resistant smoking

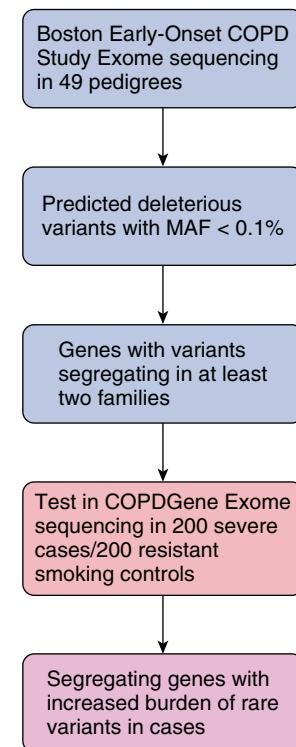


Figure 1. A flow chart of the study design. COPD = chronic obstructive pulmonary disease; MAF = minor allele frequency.

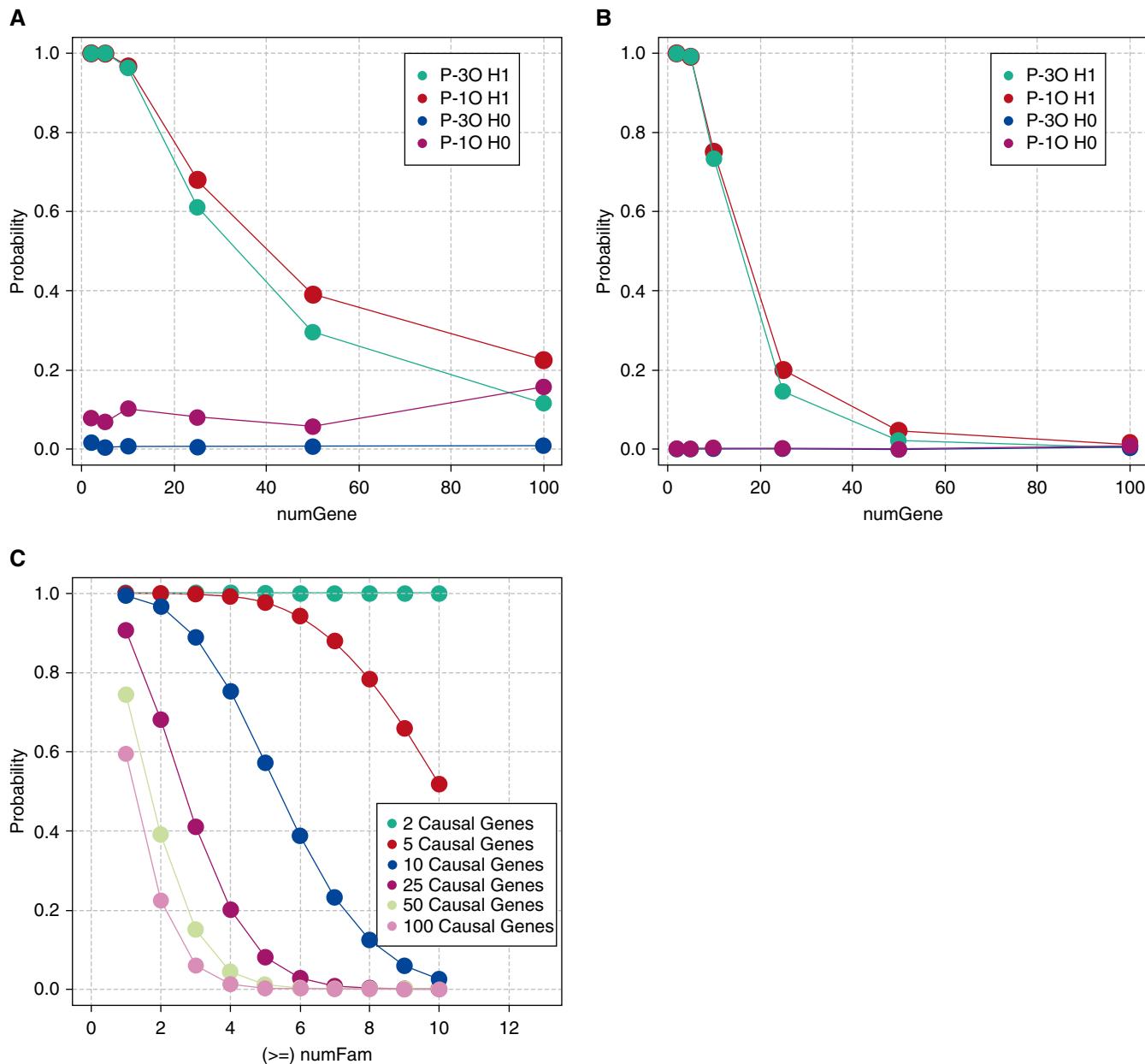


Figure 2. Effect of genetic heterogeneity on probability of identifying segregating variants. (A and B) The probability of identifying at least two (A) or four (B) segregating families is shown on the y-axis for different numbers of causal genes (x-axis) for 49 three-offspring (P-3O) or 49 one-parent with one-offspring (P-1O) scenarios under the alternative (H1) and null (H0) hypotheses. (C) Plot of the probability of observing at least n segregating P-1O families (x-axis) for different numbers of causal genes. numFam = number of segregating families; numGene = number of causal genes.

control subjects with 206,324 variants remained after quality control (see the online supplement). Baseline characteristics of the subjects are shown in Table 3.

Among the 69 genes with variants segregating in at least two families, we identified four genes that showed a nominally significantly ($P < 0.05$, using the T5 test in ScoreSeq [21]) increased burden of deleterious variants in cases (P value for enrichment of genes with nominally

significant result in COPDGene among the 69 genes = 0.079). Details of the variants segregating in these four genes are shown in Table E2 in the online supplement, and individual pedigree diagrams are shown in Figures 3–6 (the sexes of the individuals have been modified to maintain anonymity). The counts of deleterious rare variants for all four genes in the COPDGene dataset are presented in Table 4. Of note, in the COPDGene case-control dataset, all of the

deleterious rare variants with MAF less than 1% in activated leukocyte cell adhesion molecule (ALCAM; also known as CD166) are present in the COPD cases.

Gene Expression in Lung Tissue

We sought complementary evidence of a role in COPD pathogenesis for genes identified through segregation by assessing differential expression in COPD lung tissue. Among the 69 genes segregating in at least

Table 3. Baseline Characteristics of Individuals Chosen from the COPDGene Study for Whole-Exome Sequencing

Group	Male Sex (%)	Age (yr)	Pack-Years	FEV ₁ % Predicted	Emphysema (% < -950 HU)
Cases of COPD (n = 204)	51	57.4 ± 4.0	50.3 ± 19.5	30.0 ± 10.1	29.5 ± 9.9
Resistant smoking control subjects (n = 195)	44	70.3 ± 3.9	49.8 ± 20.0	99.7 ± 9.7	1.8 ± 1.4

Definition of abbreviations: COPD = chronic obstructive pulmonary disease; HU = Hounsfield units.

Means (SE) are presented for age, pack-years, FEV₁ % predicted, and computed tomography emphysema for each phenotype group.

2 families, 63 of them had evidence for expression in the lung by RNA-Seq (fragments per kilobase of exon per million fragments mapped > 0.5 in more than 50% of samples) in the Lung Genomics Research Consortium (<http://www.lung-genomics.org/>; [25]; $P = 0.035$ using hypergeometric test). In lung tissue from 111 subjects with severe COPD and 40 control subjects with normal spirometry (26), 13 of the 69 genes segregating in at least two EOCOPD families showed nominally ($P < 0.05$) significant difference in expression in COPD (enrichment $P = 0.144$); two of them (*PDCD2L* and *HSPA5*) showed significant (false discovery rate < 0.05) differences in expression in COPD ($P = 0.03$ for enrichment in false discovery rate-corrected differential expressed genes). Among the four genes that are also nominally significant in the COPDGene, *RARS* (arginyl-transfer RNA [tRNA] synthetase) was differentially

expressed in the lung at a nominal level of significance (unadjusted $P = 0.038$).

Evidence for Segregation and Association of Candidate Genes

We compiled a list of genes (Table E8) based on proximity to published genome-wide associations for COPD and lung function (3, 27–31), and also genes from Mendelian syndromes with manifestations of emphysema (32). Of these genes, 96 had variants present in the family-based data, and eight of the genes had evidence of a rare, predicted deleterious variant that segregated in at least one family (Table 5). Although none of these genes had variants that segregated in at least two families, we noted that a second variant in *FAM13A* (rs114577372) just above the MAF cutoff (0.2%) also appeared to segregate with affection in one pedigree. Five of these genes reached a nominal significance level in the case-control dataset. Results for the

variant segregation analysis and the case-control association analyses are shown in Table 5. For *CCDC38*, we also examined the specific single-nucleotide polymorphism, rs10859974, previously described in association with protection from development of COPD; 61 cases and 83 control subjects harbored this single-nucleotide polymorphism, directionally consistent with a prior report (33).

Discussion

The reasons for varied susceptibility to COPD remain largely unknown. Although most efforts have focused on identifying common variants through candidate gene studies and GWAS, a subset of patients with COPD have monogenic diseases, including severe α_1 -antitrypsin deficiency and cutis laxa. We sought to identify novel genetic candidates for severe, early-onset disease by

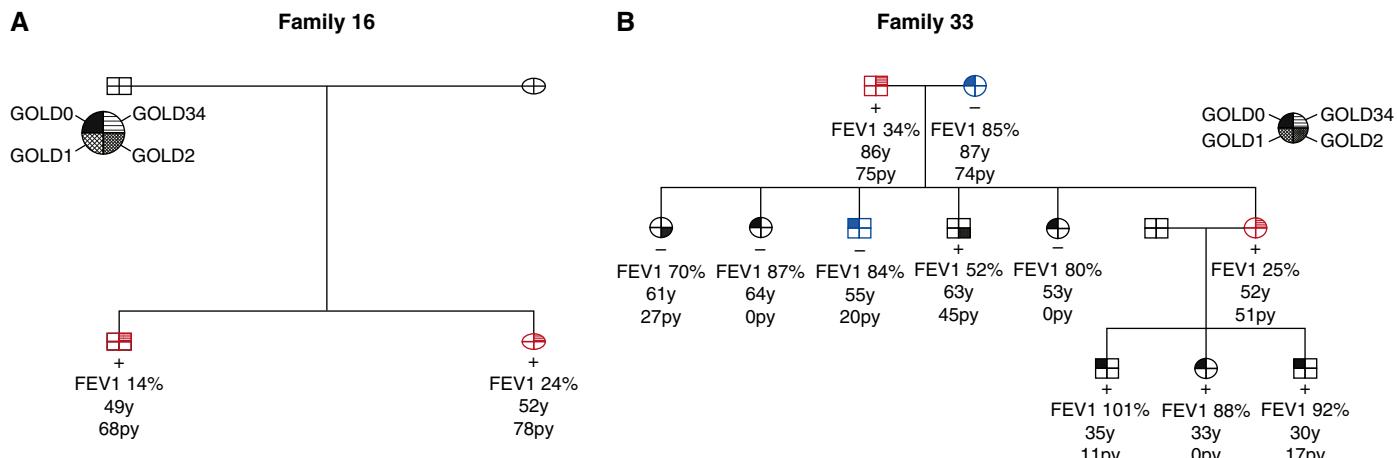


Figure 3. Pedigree plots of the families with segregating variants in activated leukocyte cell adhesion molecule (*ALCAM*). Individuals colored in red are the cases of severe and very severe chronic obstructive pulmonary disease (COPD); individuals colored in blue are control subjects with normal spirometry, number of pack-years above 5, and age above 40 years. COPD severity (or GOLD [Global Initiative for Chronic Obstructive Lung Disease] 0, for those with normal spirometry) is indicated by solid symbols. An open circle/square indicates neither case nor control; “+” symbol under the circle indicates that the individual carries a segregating variant in the gene of interest; “–” symbol indicates the individual does not carry the variant; otherwise the variant status is unknown for the subject. Below the carrier status, phenotype FEV₁% predicted, age, and pack-years (py) are shown for the sequenced subjects.

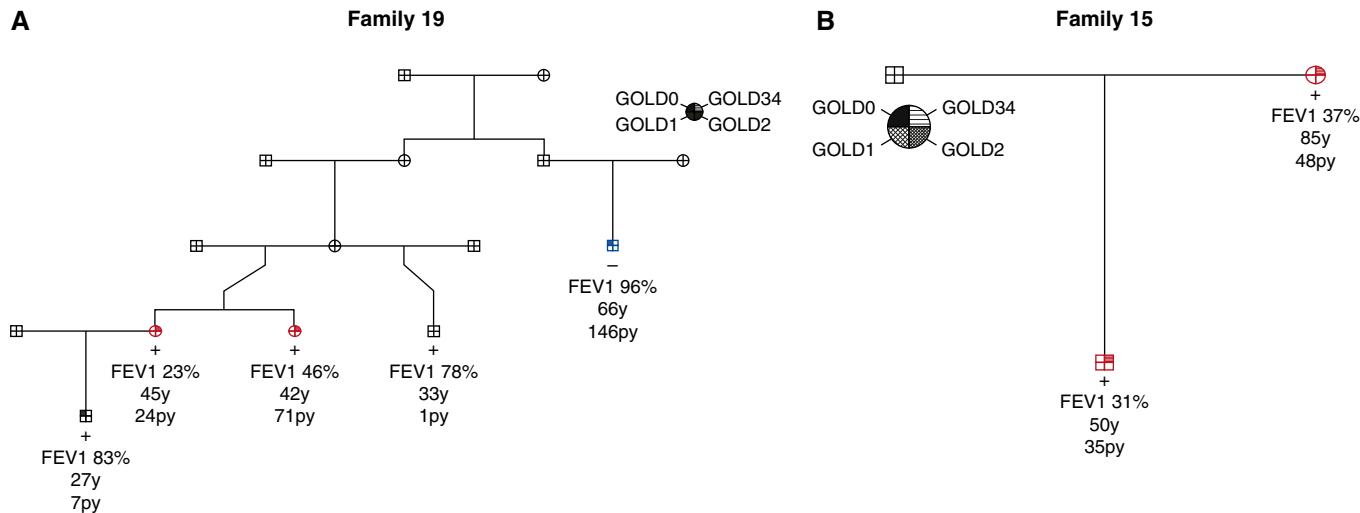


Figure 4. Pedigree plots of the families with segregating variants in dynein, axonemal, heavy chain 8 (*DNAH8*). Individuals colored in *red* are the cases of severe and very severe chronic obstructive pulmonary disease (COPD); individuals colored in *blue* are control subjects with normal spirometry, number of pack-years above 5, and age above 40 years. COPD severity (or GOLD [Global Initiative for Chronic Obstructive Lung Disease] 0, for those with normal spirometry) is indicated by *solid* symbols. An *open circle/square* indicates neither case nor control; “+” symbol under the *circle* indicates that the individual carries a segregating variant in the gene of interest; “-” symbol indicates the individual does not carry the variant; otherwise the variant status is unknown for the subject. Below the carrier status, phenotype FEV₁% predicted, age, and pack-years (py) are shown for the sequenced subjects.

performing a filtering-based segregation analysis in exome sequencing data from the family-based Boston EOCOPD Study. We considered MAF, predicted functional impact, and segregation with severe COPD in families.

To strengthen evidence for segregation, we further required genes to have segregating variants in at least two different pedigrees, and identified 69 genes with rare, predicted

deleterious variants that segregated in at least two families. The absence of genes with segregating functional variants in more than three families suggests that coding genetic determinants of severe EOCOPD may be highly heterogeneous. In an independent dataset of cases with severe COPD and resistant smoking control subjects, four of these genes had an increased burden of rare

deleterious variants in cases. *ALCAM* is a member of a subfamily of Ig receptors (34) and is expressed in many cells, including lung endothelial cells, where it appears to control flux into the alveolar airspace (35) and inhibits the transendothelial migration of monocytes in rat lung (36). In a transcriptomic analysis of human lung development, *ALCAM* expression was

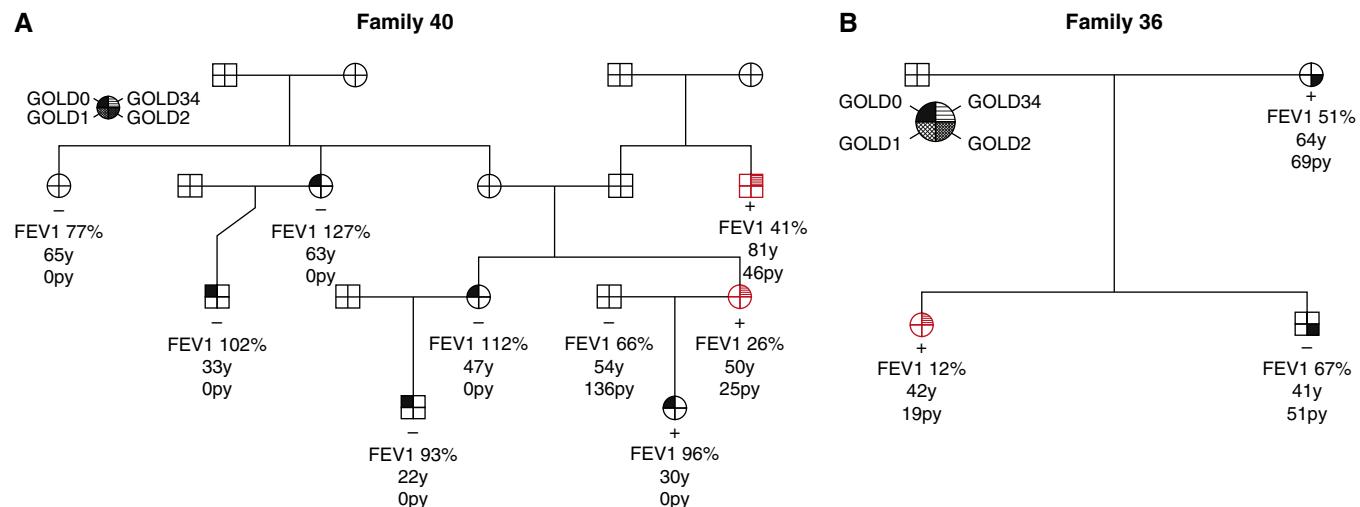


Figure 5. Pedigree plots of the families with segregating variants in arginyl-TRNA synthetase (*RARS*). Individuals colored in *red* are the cases of severe and very severe chronic obstructive pulmonary disease (COPD); individuals colored in *blue* are control subjects with normal spirometry, number of pack-years above 5, and age above 40 years. COPD severity (or GOLD [Global Initiative for Chronic Obstructive Lung Disease] 0, for those with normal spirometry) is indicated by *solid* symbols. An *open circle/square* indicates neither case nor control; “+” symbol under the *circle* indicates that the individual carries a segregating variant in the gene of interest; “-” symbol indicates the individual does not carry the variant; otherwise the variant status is unknown for the subject. Below the carrier status, phenotype FEV₁% predicted, age, and pack-years (py) are shown for the sequenced subjects.

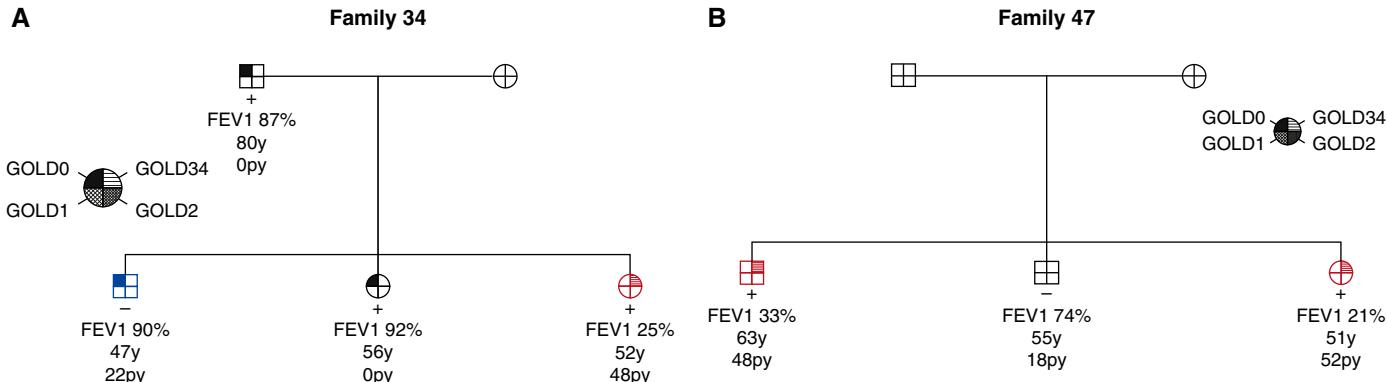


Figure 6. Pedigree plots of the families with segregating variants in golgi brefeldin A resistant guanine nucleotide exchange factor 1 (*GBF1*). Individuals colored in red are the cases of severe and very severe chronic obstructive pulmonary disease (COPD); individuals colored in blue are control subjects with normal spirometry, number of pack-years above 5, and age above 40 years. COPD severity (or GOLD [Global Initiative for Chronic Obstructive Lung Disease] 0, for those with normal spirometry) is indicated by solid symbols. An open circle/square indicates neither case nor control; “+” symbol under the circle indicates that the individual carries a segregating variant in the gene of interest; “—” symbol indicates the individual does not carry the variant; otherwise the variant status is unknown for the subject. Below the carrier status, phenotype FEV₁% predicted, age, and pack-years (py) are shown for the sequenced subjects.

significantly associated with the second principal component, which included the surfactant-associated claudin *CLDN18*, *CXCL5*, and the major histocompatibility complex class II molecules *HLA-DRA* and *HLA-DRB1* (37). *ALCAM* also may have a role in regulation of proteolysis; in a melanoma model, *ALCAM* truncation impaired matrix metalloproteinase-2 activation (38). *DNAH8* encodes for an outer arm dynein heavy chain (39) and is a candidate for involvement in primary ciliary dyskinesia (6). Recently, variants in *DNAH8* were identified in association with heavy smoking (40). *RARS* cytoplasmic is a member for the aminoacyl-tRNA synthetase family, an enzyme that catalyzes the ligation of amino acids to their specific tRNAs (39). Although a lung-specific genetic defect involving *RARS* has not, to our knowledge, been described, murine knockouts of p38, an aminoacyl-tRNA synthetase cofactor,

develop respiratory distress, and p38 is required for lung cell differentiation (41). Antibodies against several synthetases (but not arginyl-tRNA) lead to the antisynthetase syndrome, which includes interstitial lung disease as one of its manifestations (42). Finally, *GBF1* (Golgi-specific brefeldin A-resistance guanine nucleotide exchange factor 1) is a guanine nucleotide exchange factor that regulates the recruitment of proteins to membranes (43). *GBF1* appears to be important for neutrophil chemotaxis and superoxide production involving phosphatidyl inositol 3-kinase (44). Phosphatidyl inositol 3-kinase inhibitors are an area of therapeutic development in COPD (45).

We also note the appearance of variants in cystic fibrosis transmembrane conductance regulator (*CFTR*) that segregated with affection status in two families. The first, rs121908752, has been previously described

as pathogenic (46); the second, rs149279509, is of unclear effect (47). Neither of these families appeared to harbor other possibly pathogenic *CFTR* variants or the ΔF508 mutation (rs121909001). Several studies have demonstrated a potential role for *CFTR* in COPD (48), although a clear effect of these variants has not been demonstrated (49). Although our main findings focused on a dominant model for segregation, we also examined a recessive model. Although none of these genes demonstrated evidence for segregation in more than one family, one of the genes from this analysis was *ABCA13*, which segregated in a family with three siblings with severe COPD. A second model, attempting to model potential compound heterozygotes (as true phasing was not available), identified a few genes containing pairs of variants that segregated with COPD status in at least two pedigrees (see the online supplement).

In addition to identifying novel disease candidates, we also attempted to identify evidence for rare variants in previously published candidate genes for COPD or lung function. A total of 12 genes were shown to be segregating in families or have a marginally significant *P* value in the case-control dataset. We identified segregating variants in *FAM13A*, common variants in and near which have been strongly associated with COPD and lung function (3). A recent sequencing study of 100 smokers without COPD identified a nonsynonymous variant in *CCDC38* associated with protection from development of COPD; this locus has also been associated with FEV₁/FVC in cohort

Table 4. Segregating Genes with Marginally Significant *P* Values in the COPDGene Case-Control Dataset

Gene	T5 <i>P</i> Value	MAC for Most Deleterious Variants	
		Cases	Control Subjects
<i>ALCAM</i>	0.041	3	0
<i>RARS</i>	0.039	2	1
<i>DNAH8</i>	0.026	29	20
<i>GBF1</i>	0.048	6	2

Definition of abbreviation: MAC = minor allele count.

T5 *P* values are two-sided (with increased burden of rare variants in cases) obtained for variants with allele frequency less than 5% using a burden test. MAC values are shown for variants with allele frequency less than 0.1% most likely to be deleterious (see main text).

Table 5. Results for Published Candidate Genes in the Variant Segregation Analysis and Case–Control Association Analysis

Gene	Source	EOCOPD (No. of Pedigrees)	SNP Location	COPDGene P Value	Effect
<i>CCDC38</i>	Lung function	1	12-96272113	0.0457	Protective
<i>DNER</i>	Lung function	1	2-230231749	0.6453	Deleterious
<i>EGLN2</i>	COPD	1	19-41307024	0.9071	Protective
<i>FAM13A</i>	COPD	1	4-89950680	0.6424	Deleterious
<i>GPR126</i>	Lung function	1	6-142723802	0.9141	Deleterious
<i>MECOM</i>	Lung function	1	3-169099162	0.0588	Protective
<i>NMBR</i>	Lung function	1	6-142409506	0.7902	Deleterious
<i>KCNJ2</i>	Lung function	1	17-68171232	—	—
<i>CYP2A6</i>	COPD	0	—	0.0021	Protective
<i>RHOBTB3</i>	Lung function	0	—	0.0382	Protective
<i>THSD4</i>	Lung function	0	—	0.0330	Protective
<i>ZNF323</i>	Lung function	0	—	0.0347	Protective

Definition of abbreviations: COPD = chronic obstructive pulmonary disease; EOCOPD = Early-Onset Chronic Obstructive Pulmonary Disease Study; SNP = single-nucleotide polymorphism.

Genes that are segregating in at least one EOCOPD pedigree or have a *P* value less than 0.05 using the T5 test in the COPDGene dataset are listed here. No. of Pedigrees indicates the number of pedigrees with a segregating variant. COPDGene *P* value refers to the result from the T5 test.

studies. In our dataset, we found a marginally significant *P* value in the protective direction in the burden test on the case–control dataset, and a directionally consistent allele count in the control subjects versus cases for rs10859974, supporting the protective effect of this variant. Of note, this study also described *ABCA13* and *GBF1* as genes harboring three or more rare variants in resistant smoking control subjects; these data would be inconsistent with our findings, illustrating the challenges of these analyses and potential benefits of examining a severely affected group (37). We also identify several other genes at GWAS loci that may harbor rare variants. However, none of our findings meets the thresholds imposed for multiple testing corrections, and further study is required to confirm these findings.

Identification of rare variants in complex disease has been challenging. Several recent successes have demonstrated statistically significant results, most relying on very large sample sizes (6, 50). In contrast, since the initial reports of using exome sequencing to identify the etiology of a Mendelian disorder, many studies have used filtering/segregation approaches with varied success (18). A limited number of studies have also identified high penetrance variants in complex disease, including pulmonary disease (51–53), although none has identified such variants in COPD.

Our analysis has several limitations. First, our threshold chosen for MAF and functional scores are rather arbitrary. A well-defined consensus for these thresholds is not clear, particularly in the setting of different genetic risk factors (locus

heterogeneity) and the presence of affected subjects who may have varying amounts of genetic risk. More detailed phenotypic data could help identify similar Mendelian subtypes (21). Our decision to use a 0.1% allele frequency reflects in part the strong relationship between cigarette smoking and COPD, as a pathogenic variant could be present at some frequency potentially much higher than 0.1% in nonsmokers. Although we acknowledge the limitations of bioinformatics predictive tools, leaving out Combined Annotation-Dependent Depletion score/Condel predictions for filtering appeared to strongly bias our selection for larger genes, which are more likely to harbor variation (e.g., *TTN*). Another essential limitation of segregation analysis is that it is not yet possible, to our knowledge, to provide a valid *P* value on these segregation events, due to the complicated family structure and missing parental information. A test incorporating these ideas to correct for gene size or family structure has been developed (54); however, it can be applied to pairs of cases with the same relationships only, and no solution currently exists to extend this approach to more complex pedigrees. COPD is strongly related to cigarette smoking. Unaffected subjects in the family dataset were younger and had fewer pack-years than control subjects in the case–control dataset. Thus, it is possible that the segregating variants we identify are related to nicotine addiction. This and other differences between the studies could be a source of bias. Our association analysis in COPDGene is also likely underpowered, as power greatly depends on

the MAF of the causal variants and the effect size (see the online supplement); thus, causal genes may not reach nominal significance, and our identified genes with nominal significance could be false positives.

Our study is, to our knowledge, the first whole-exome analysis of severe COPD. Our analysis focused on a unique cohort of pedigrees identified through a proband with severe, early-onset disease to enrich for rare variants of high penetrance. We sought supportive evidence in a second, case–control cohort also selected for severe disease and resistant smoking control subjects. Our analysis prioritizes a set of promising genes for further study, but also suggests that rare coding variants that markedly increase susceptibility to severe COPD are likely to be found in many genes. Future studies could include replication in additional subjects with severe COPD and functional analysis for the effects of the given variants. Our work illustrates many of the challenges of rare variant analysis in a complex disorder with known Mendelian subtypes, and suggests that genetic heterogeneity is a likely limitation for this approach. Our experience suggests that additional methods to identify rare variants of high penetrance in families may be needed. ■

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

Acknowledgment: The authors thank Nan Laird, Shamil Sunyaev, Goo Jun, and the members of the University of Washington Center for Mendelian Genomics for their insightful comments, and all study participants. The authors acknowledge the support of the NHLBI

and the contributions of the research institutions, study investigators, field staff, and study participants in creating this resource for biomedical

research. They also thank the Exome Aggregation Consortium and the groups that provided exome variant data. The content of this article is solely

responsibility of the authors and does not necessarily represent the official views of the NHLBI or the National Institutes of Health.

References

- Mannino DM, Homa DM, Akinbami LJ, Ford ES, Redd SC. Chronic obstructive pulmonary disease surveillance—United States, 1971–2000. *Respir Care* 2002;47:1184–1199.
- Hersh CP, Demeo DL, Silverman EK. Chronic obstructive pulmonary disease. In: Silverman EK, Shapiro SD, Lomas DA, Weiss ST, editors. *Respiratory genetics*. London: Hodder Arnold; 2005. pp. 253–296.
- Cho MH, Bautaoui N, Klanderman BJ, Sylvia JS, Zinetti JP, Hersh CP, DeMeo DL, Hunninghake GM, Litonjua AA, Sparrow D, et al. Variants in FAM13A are associated with chronic obstructive pulmonary disease. *Nat Genet* 2010;42:200–202.
- Pillai SG, Ge D, Zhu G, Kong X, Shianna KV, Need AC, Feng S, Hersh CP, Bakker P, Gulsvik A, 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.
- Zhou JJ, Cho MH, Castaldi PJ, Hersh CP, Silverman EK, Laird NM. Heritability of chronic obstructive pulmonary disease and related phenotypes in smokers. *Am J Respir Crit Care Med* 2013;188:941–947.
- Crosby J, Peloso GM, Auer PL, Crosslin DR, Stitzel NO, Lange LA, Lu Y, Tang ZZ, Zhang H, Hindy G, et al.; TG and HDL Working Group of the Exome Sequencing Project, National Heart, Lung, and Blood Institute. Loss-of-function mutations in APOC3, triglycerides, and coronary disease. *N Engl J Med* 2014;371:22–31.
- Jensen LJ, Kuhn M, Stark M, Chaffron S, Creevey C, Muller J, Doerks T, Julian P, Roth A, Simonovic M, et al. STRING 8—a global view on proteins and their functional interactions in 630 organisms. *Nucleic Acids Res* 2009;37:D412–D416.
- Gilissen C, Arts HH, Hoischen A, Spruijt L, Mans DA, Arts P, van Lier B, Stehouwer M, van Reeuwijk J, Kant SG, et al. Exome sequencing identifies WDR35 variants involved in Sensenbrenner syndrome. *Am J Hum Genet* 2010;87:418–423.
- Flannick J, Thorleifsson G, Beer NL, Jacobs SB, Grarup N, Burtt NP, Mahajan A, Fuchsberger C, Atzmon G, Benediktsson R, et al.; Go-T2D Consortium; T2D-GENES Consortium. Loss-of-function mutations in SLC30A8 protect against type 2 diabetes. *Nat Genet* 2014;46:357–363.
- Ng SB, Buckingham KJ, Lee C, Bigham AW, Tabor HK, Dent KM, Huff CD, Shannon PT, Jabs EW, Nickerson DA, et al. Exome sequencing identifies the cause of a mendelian disorder. *Nat Genet* 2010;42:30–35.
- Knowles MR, Leigh MW, Ostrowski LE, Huang L, Carson JL, Hazucha MJ, Yin W, Berg JS, Davis SD, Dell SD, et al.; Genetic Disorders of Mucociliary Clearance Consortium. Exome sequencing identifies mutations in CCDC114 as a cause of primary ciliary dyskinesia. *Am J Hum Genet* 2013;92:99–106.
- Bamshad MJ, Ng SB, Bigham AW, Tabor HK, Emond MJ, Nickerson DA, Shendure J. Exome sequencing as a tool for Mendelian disease gene discovery. *Nat Rev Genet* 2011;12:745–755.
- Qiao D, Lange C, Beaty TH, Crapo JD, Silverman EK, Cho MH. Whole exome sequencing analysis of severe, early-onset COPD in extended pedigrees [abstract]. Presented at the American Society of Human Genetics Conference. October 18–22, 2014, San Diego, CA.
- Qiao D, Lange C, Silverman EK, Cho MH. Whole exome sequencing analyses of severe, early-onset COPD in extended pedigrees [abstract]. Presented at the American Thoracic Society International Conference. May 16–21, 2014, San Diego, CA.
- Silverman EK, Chapman HA, Drazen JM, Weiss ST, Rosner B, Campbell EJ, O'Donnell WJ, Reilly JJ, Ginns L, Mentzer S, et al. Genetic epidemiology of severe, early-onset chronic obstructive pulmonary disease. Risk to relatives for airflow obstruction and chronic bronchitis. *Am J Respir Crit Care Med* 1998;157:1770–1778.
- Lieberman J. Alpha-1-antitrypsin deficiency and related disorders. In: *Principles and practice of medical genetics*, 5th ed. Vol. 2. New York: Churchill Livingstone; 1983. pp. 911–935.
- Regan EA, Hokanson JE, Murphy JR, Make B, Lynch DA, Beaty TH, Curran-Everett D, Silverman EK, Crapo JD. Genetic epidemiology of COPD (COPDGene) study design. *COPD* 2010;7:32–43.
- Kircher M, Witten DM, Jain P, O'Roak BJ, Cooper GM, Shendure J. A general framework for estimating the relative pathogenicity of human genetic variants. *Nat Genet* 2014;46:310–315.
- González-Pérez A, López-Bigas N. Improving the assessment of the outcome of nonsynonymous SNVs with a consensus deleteriousness score, Condel. *Am J Hum Genet* 2011;88:440–449.
- Hu H, Roach JC, Coon H, Guthery SL, Voelkerding KV, Margraf RL, Durttschi JD, Tavtigian SV, Shankaracharya, Wu W, et al. A unified test of linkage analysis and rare-variant association for analysis of pedigree sequence data. *Nat Biotechnol* 2014;32:636–669.
- Lin DY, Tang ZZ. A general framework for detecting disease associations with rare variants in sequencing studies. *Am J Hum Genet* 2011;89:354–367.
- Alexa A, Rahnenführer J, Lengauer T. Improved scoring of functional groups from gene expression data by decorrelating GO graph structure. *Bioinformatics* 2006;22:1600–1607.
- Alexa A, Rahnenführer J. TopGO: enrichment analysis for gene ontology, R package version 2.8.0; 2010.
- Exome Aggregation Consortium. Cambridge, MA: ExAC [accessed 2014 Dec]. Available from: <http://exac.broadinstitute.org>
- Uhlén M, Fagerberg L, Hallström BM, Lindskog C, Oksvold P, Mardinoglu A, Sivertsson Å, Kampf C, Sjöstedt E, Asplund A, et al. Proteomics: tissue-based map of the human proteome. *Science* 2015;347:1260419.
- Morrow J, Qiu W, DeMeo DL, Houston I, Pinto Plata VM, Celli BR, Marchetti N, Criner GJ, Bueno R, Washko GR, et al. Network analysis of gene expression in severe COPD lung tissue samples [abstract]. *Am J Respir Crit Care Med* 2015;191:A1253.
- Cho MH, Castaldi PJ, Wan ES, Siedlinski M, Hersh CP, Demeo DL, Himes BE, Sylvia JS, Klanderman BJ, Zinetti JP, et al.; ICGN Investigators; ECLIPSE Investigators; COPDGene Investigators. A genome-wide association study of COPD identifies a susceptibility locus on chromosome 19q13. *Hum Mol Genet* 2012;21:947–957.
- Soler Artigas M, Loth DW, Wain LV, Gharib SA, Obeidat M, Tang W, Zhai G, Zhao JH, Smith AV, Huffman JE, et al.; International Lung Cancer Consortium; GIANT consortium. Genome-wide association and large-scale follow up identifies 16 new loci influencing lung function. *Nat Genet* 2011;43:1082–1090.
- Wilk JB, Chen TH, Gottlieb DJ, Walter RE, Nagle MW, Bandler BJ, Myers RH, Borecki IB, Silverman EK, Weiss ST, et al. A genome-wide association study of pulmonary function measures in the Framingham Heart Study. *PLoS Genet* 2009;5:e1000429.
- Cho MH, McDonald ML, Zhou X, Mattheisen M, Castaldi PJ, Hersh CP, Demeo DL, Sylvia JS, Zinetti J, Laird NM, et al.; NETT Genetics, ICGN, ECLIPSE and COPDGene Investigators. Risk loci for chronic obstructive pulmonary disease: a genome-wide association study and meta-analysis. *Lancet Respir Med* 2014;2:214–225.
- Repapi E, Sayers I, Wain LV, Burton PR, Johnson T, Obeidat M, Zhao JH, Ramasamy A, Zhai G, Vitart V, et al.; Wellcome Trust Case Control Consortium; NSHD Respiratory Study Team. Genome-wide association study identifies five loci associated with lung function. *Nat Genet* 2010;42:36–44.
- Hancock DB, Eijgelsheim M, Wilk JB, Gharib SA, Loehr LR, Marciante KD, Franceschini N, van Durme YM, Chen TH, Barr RG, et al. Meta-analyses of genome-wide association studies identify multiple loci associated with pulmonary function. *Nat Genet* 2010;42:45–52.

33. Wain LV, Sayers I, Soler Artigas M, Portelli MA, Zeggini E, Obeidat M, Sin DD, Bossé Y, Nickle D, Brandsma CA, et al. Whole exome re-sequencing implicates CCDC38 and cilia structure and function in resistance to smoking related airflow obstruction. *PLoS Genet* 2014; 10:e1004314.

34. Bowen MA, Patel DD, Li X, Modrell B, Malacko AR, Wang WC, Marquardt H, Neubauer M, Pesando JM, Francke U, et al. Cloning, mapping, and characterization of activated leukocyte-cell adhesion molecule (ALCAM), a CD6 ligand. *J Exp Med* 1995;181: 2213–2220.

35. Ofori-Acquah SF, King J, Voelkel N, Schaphorst KL, Stevens T. Heterogeneity of barrier function in the lung reflects diversity in endothelial cell junctions. *Microvasc Res* 2008;75:391–402.

36. Masedunskas A, King JA, Tan F, Cochran R, Stevens T, Sviridov D, Ofori-Acquah SF. Activated leukocyte cell adhesion molecule is a component of the endothelial junction involved in transendothelial monocyte migration. *FEBS Lett* 2006;580:2637–2645.

37. Kho AT, Bhattacharya S, Tantisira KG, Carey VJ, Gaedigk R, Leeder JS, Kohane IS, Weiss ST, Mariani TJ. Transcriptomic analysis of human lung development. *Am J Respir Crit Care Med* 2010;181:54–63.

38. Lunter PC, van Kilsdonk JW, van Beek H, Cornelissen IM, Bergers M, Willems PH, van Muijen GN, Swart GW. Activated leukocyte cell adhesion molecule (ALCAM/CD166/MEMD), a novel actor in invasive growth, controls matrix metalloproteinase activity. *Cancer Res* 2005; 65:8801–8808.

39. Chapelin C, Duriez B, Magnino F, Goossens M, Escudier E, Amselem S. Isolation of several human axonemal dynein heavy chain genes: genomic structure of the catalytic site, phylogenetic analysis and chromosomal assignment. *FEBS Lett* 1997;412:325–330.

40. Wain LV, Shrine N, Miller S, Jackson VE, Ntalla I, Artigas MS, Billington CK, Kheirallah AK, Allen R, Cook JP, et al. Novel insights into the genetics of smoking behaviour, lung function, and chronic obstructive pulmonary disease (UK BiLEVE): a genetic association study in UK Biobank. *Lancet Respir Med* 2015;3:769–781.

41. Kim MJ, Park BJ, Kang YS, Kim HJ, Park JH, Kang JW, Lee SW, Han JM, Lee HW, Kim S. Downregulation of FUSE-binding protein and c-myc by tRNA synthetase cofactor p38 is required for lung cell differentiation. *Nat Genet* 2003;34:330–336.

42. Park SG, Schimmel P, Kim S. Aminoacyl tRNA synthetases and their connections to disease. *Proc Natl Acad Sci USA* 2008;105: 11043–11049.

43. Claude A, Zhao BP, Kuziemsky CE, Dahan S, Berger SJ, Yan JP, Arnold AD, Sullivan EM, Melançon P. GBF1: a novel Golgi-associated BFA-resistant guanine nucleotide exchange factor that displays specificity for ADP-ribosylation factor 5. *J Cell Biol* 1999;146:71–84.

44. Mazaki Y, Nishimura Y, Sabe H. GBF1 bears a novel phosphatidylinositol-phosphate binding module, BP3K, to link PI3K γ activity with Arf1 activation involved in GPCR-mediated neutrophil chemotaxis and superoxide production. *Mol Biol Cell* 2012;23:2457–2467.

45. Barnes PJ. New anti-inflammatory targets for chronic obstructive pulmonary disease. *Nat Rev Drug Discov* 2013;12:543–559.

46. Sosnay PR, Siklosi KR, Van Goor F, Kaniecki K, Yu H, Sharma N, Ramalho AS, Amaral MD, Dorfman R, Zielenski J, et al. Defining the disease liability of variants in the cystic fibrosis transmembrane conductance regulator gene. *Nat Genet* 2013;45: 1160–1167.

47. Cystic fibrosis mutation database. Toronto: Cystic Fibrosis Centre at the Hospital for Sick Children; 2011.

48. Raju SV, Jackson PL, Courville CA, McNicholas CM, Sloane PA, Sabbatini G, Tidwell S, Tang LP, Liu B, Fortenberry JA, et al. Cigarette smoke induces systemic defects in cystic fibrosis transmembrane conductance regulator function. *Am J Respir Crit Care Med* 2013;188:1321–1330.

49. Raju SV, Tate JH, Peacock SK, Fang P, Oster RA, Dransfield MT, Rowe SM. Impact of heterozygote CFTR mutations in COPD patients with chronic bronchitis. *Respir Res* 2014;15:18.

50. Jørgensen AB, Frikke-Schmidt R, Nordestgaard BG, Tybjærg-Hansen A. Loss-of-function mutations in APOC3 and risk of ischemic vascular disease. *N Engl J Med* 2014;371:32–41.

51. de Jesus Perez VA, Yuan K, Lyuksyutova MA, Dewey F, Orcholski ME, Shuffle EM, Mathur M, Yancy L Jr, Rojas V, Li CG, et al. Whole-exome sequencing reveals TopBP1 as a novel gene in idiopathic pulmonary arterial hypertension. *Am J Respir Crit Care Med* 2014; 189:1260–1272.

52. Emond MJ, Louie T, Emerson J, Zhao W, Mathias RA, Knowles MR, Wright FA, Rieder MJ, Tabor HK, Nickerson DA, et al.; National Heart, Lung, and Blood Institute (NHLBI) GO Exome Sequencing Project; Lung GO. Exome sequencing of extreme phenotypes identifies DCTN4 as a modifier of chronic *Pseudomonas aeruginosa* infection in cystic fibrosis. *Nat Genet* 2012;44:886–889.

53. Ma L, Roman-Campos D, Austin ED, Eyries M, Sampson KS, Soubrier F, Germain M, Tréguoët DA, Borczuk A, Rosenzweig EB, et al. A novel channelopathy in pulmonary arterial hypertension. *N Engl J Med* 2013;369:351–361.

54. Ionita-Laza I, Makarov V, Yoon S, Raby B, Buxbaum J, Nicolae DL, Lin X. Finding disease variants in Mendelian disorders by using sequence data: methods and applications. *Am J Hum Genet* 2011; 89:701–712.

55. Wan ES, Castaldi PJ, Cho MH, Hokanson JE, Regan EA, Make BJ, Beaty TH, Han MK, Curtis JL, Curran-Everett D, et al. COPDGene Investigators. Epidemiology, genetics, and subtyping of preserved ratio impaired spirometry (PRISM) in COPDGene. *Respir Res* 2014; 15:89.

56. Cingolani P, Platts A, Wang LL, Coon M, Nguyen T, Wang L, Land SJ, Lu X, Ruden DM. A program for annotating and predicting the effects of single nucleotide polymorphisms, SnpEff: SNPs in the genome of *Drosophila melanogaster* strain *w¹¹¹⁸*, *iso-2*; *iso-3*. *Fly (Austin)* 2012; 6:80–92.