Attempted Replication of Reported Chronic Obstructive Pulmonary Disease Candidate Gene Associations

Craig P. Hersh, Dawn L. DeMeo, Christoph Lange, Augusto A. Litonjua, John J. Reilly, David Kwiatkowski, Nan Laird, Jody S. Sylvia, David Sparrow, Frank E. Speizer, Scott T. Weiss, and Edwin K. Silverman

Channing Laboratory, Pulmonary and Critical Care Division, and Hematology/Oncology Division, Department of Medicine, Brigham and Women's Hospital; Harvard Medical School; Department of Biostatistics, Harvard School of Public Health; and Veterans Affairs Medical Center, Boston, Massachusetts

Case-control studies have successfully identified many significant genetic associations for complex diseases, but lack of replication has been a criticism of case-control genetic association studies in general. We selected 12 candidate genes with reported associations to chronic obstructive pulmonary disease (COPD) and genotyped 29 polymorphisms in a family-based study and in a case-control study. In the Boston Early-Onset COPD Study families, significant associations with quantitative and/or qualitative COPD-related phenotypes were found for the tumor necrosis factor (TNF)-α $-308G>A$ promoter polymorphism ($P < 0.02$), a coding variant in surfactant protein B (SFTPB Thr131Ile) ($P = 0.03$), and the $(GT)_{31}$ allele of the heme oxygenase (HMOX1) promoter short tandem repeat ($P = 0.02$). In the case-control study, the SFTPB Thr131Ile polymorphism was associated with COPD, but only in the presence of a gene-by-environment interaction term ($P = 0.01$ for both main effect and interaction). The 30-repeat, but not the 31-repeat, allele of HMOX1 was associated ($P = 0.04$). The TNF $-308G>A$ polymorphism was not significant. In addition, the microsomal epoxide hydrolase "fast" allele (EPHX1 His139Arg) was significantly associated in the case-control study ($P = 0.03$). Although some evidence for replication was found for SFTPB and HMOX1, none of the previously published COPD genetic associations was convincingly replicated across both study designs.

Keywords: association studies; case-control studies; emphysema; genetics; single nucleotide polymorphism

Case-control association analysis is a commonly used study design in the field of complex trait genetics. Susceptibility genes successfully identified using this approach include the associations between Factor V Leiden and venous thromboembolism (1), Apolipoprotein E4 and Alzheimer's disease (2), and peroxisome proliferator–activated receptor-γ (PPARγ) and type 2 diabetes mellitus (3). However, candidate gene case-control association studies have been criticized because of a lack of replication (4, 5). Often the first published study reports a significant association between a candidate gene polymorphism and a disease of interest, but subsequent studies are unable to confirm the association (6).

Candidate genes can be selected based on previous genetic linkage analysis results (positional candidate genes), but they are more commonly chosen based on known or presumed mechanisms in disease pathophysiology or based on the results of previous association studies. In the example of PPARγ above, Altshuler and colleagues identified 16 published genetic associations for type 2 diabetes mellitus (3). They genotyped the previously associated polymorphisms in a parent-child trios study, to protect against spurious results due to population stratification. Significant associations ($P < 0.05$) were then tested for replication in two case-control studies and 1 sibling-pair study. By the use of this method, they confirmed only one of the previously reported associations. The common allele of the PPARγ Pro12-Ala polymorphism conferred a modest (1.25-fold), but significant increase in risk for type 2 diabetes mellitus.

In this study of chronic obstructive pulmonary disease (COPD), we used a study design similar to that of Altshuler and colleagues in their study of diabetes. We sought to determine whether any of the published COPD candidate gene associations would withstand the test of replication and whether a convincing COPD genetic association could be found using this approach. Twelve candidate genes with reported significant associations to COPD were identified in the published literature (Table 1 and Table E1 in the online supplement). A total of twenty-nine polymorphisms (24 single nucleotide polymorphisms [SNPs], 1 insertion/deletion [indel], 3 short tandem repeats [STRs], and 1 null deletion) were genotyped in both a family-based COPD study and a case-control COPD study. Results from this study have been previously reported as an abstract (7).

MATERIALS AND METHODS

Study Populations

Details of subject recruitment and phenotyping in the Boston Early-Onset COPD Study have been reported previously (8). Severe, early-onset COPD probands had an $FEV_1 < 40\%$ predicted, age < 53 yr, and did not have severe α_1-antitrypsin deficiency. First-degree relatives, older second-degree relatives, spouses, and other family members with COPD were invited to participate. This analysis included 949 individuals from 127 extended pedigrees. Ninety-eight percent of the Boston Early-Onset COPD Study participants were white.

The case-control study identified cases from the National Emphysema Treatment Trial (NETT) (9, 10). Subjects participating in NETT had an $FEV_1 \leqslant 45\%$ predicted (11), evidence of hyperinflation on pulmonary function testing, and bilateral emphysema on high-resolution chest CT scan. For this analysis, cases included 304 non-Hispanic white participants in the NETT Genetics Ancillary Study; nearly all were current or former smokers. Control subjects were participants in the Normative Aging Study (NAS), a longitudinal study of healthy men conducted by the Veterans Administration (VA) in Greater Boston, starting in the 1960s (12). Control subjects included 441 white male

TABLE 1. COPD CANDIDATE GENES AND POLYMORPHISMS STUDIED, ALONG WITH DETAILS OF STUDIES THAT SUPPORT AND DO NOT SUPPORT AN ASSOCIATION*

Genes/Pathways	Study Author and Year	Association Results	Country (Ethnicity)	Case Definition
Protease-antiprotease				
α_1-Antichymotrypsin (SERPINA3)	Poller 1992, 1993	Initial positive	Sweden	Airway resistance > 0.4kPa/liter/s
	Ishii 2000	Replicated	Japan	Clinical diagnosis, including PFTs
	Sandford 1998	Negative	Canada	FEV_1 < 80% pred, FEV_1/FVC < 0.7
	Benetazzo 1999	Negative	Italy	FEV_1 < 70% pred, BD response < 12%
Matrix metalloproteinase-1 (MMP1)	Joos 2002	Initial positive	United States & Canada (whites)	Rapid decline vs. no decline in FEV_1 over 5 yr in patients with COPD (Lung Health Study)
Matrix metalloproteinase-9 (MMP9)	Minematsu 2001	Initial positive	Japan	Emphysema on chest CT scan, ≥ 10 pack-years
	Joos 2002	Negative	United States & Canada (whites)	Rapid decline vs. no decline in FEV_1 over 5 yr in patients with COPD (Lung Health Study)
Tissue inhibitor of metalloproteases-2 (TIMP2)	Hirano 2001	Initial positive	Japan	FEV_1 < 70% pred, FEV_1/FVC < 0.7
Antioxidant				
Heme oxygenase (HMOX1)	Yamada 1999	Initial positive	Japan	Reduced FEV_1/FVC, $D_{L_{CO}}$; chest CT scan
	He 2002	Negative	United States & Canada (whites)	Rapid decline vs. no decline in FEV_1 over 5 yr in patients with COPD (Lung Health Study)
Xenobiotic metabolism				
Microsomal epoxide hydrolase (EPHX1)	Smith 1997	Initial positive	Scotland	COPD: FEV_1 < 1.5 liters. Emphysema: lung pathology at lung cancer resection
	Sandford 2001	Replicated	United States & Canada (whites)	Rapid decline vs. no decline in FEV_1 over 5 yr in patients with COPD (Lung Health Study)
	Yoshikawa 2000	Negative	Japan	FEV_1 < 70% pred
	Takeyabu 2000	Negative	Japan	FEV_1/FVC < 0.7; chest CT scan
	Yim 2000	Negative	Korea	FEV_1 < 80% pred, FEV_1/FVC < 0.7, BD response < 12%
Glutathione S-Transferase P1 (GSTP1)	Ishii 1999	Initial positive	Japan	FEV_1/FVC < 0.7
	He 2002	Replicated	United States & Canada (whites)	Rapid decline vs. no decline in FEV_1 over 5 yr in patients with COPD (Lung Health Study)
	Yim 2002	Negative	Korea	FEV_1 < 80% pred, FEV_1/FVC < 0.7, BD response < 12%
Glutathione S-transferase M1 (GSTM1)	Harrison 1997	Initial positive	Scotland	Emphysema on lung pathology at lung cancer resection
	Baranova 1997	Replicated	France	Chronic bronchitis with "moderate" or "severe" airflow obstruction
	Yim 2002	Negative	Korea	FEV_1 < 80% pred, FEV_1/FVC < 0.7, BD response < 12%
	He 2002	Negative	United States & Canada (whites)	Rapid decline vs. no decline in FEV_1 over 5 yr in patients with COPD (Lung Health Study)
Inflammation				
Tumor Necrosis Factor-α (TNF)	Huang 1997	Initial positive	Taiwan	FEV_1 < 80% pred, FEV_1/FVC < 0.7, chronic bronchitis
	Sakao 2001	Replicated	Japan	FEV_1 < 80% pred, FEV_1/FVC < 0.7; chronic bronchitis excluded
	Sandford 2001	Negative	United States & Canada (whites)	Rapid decline vs. no decline in FEV_1 over 5 yr in patients with COPD (Lung Health Study)
	Higham 2000	Negative	England	FEV_1/FVC ≤ 80% pred, BD response ≤ 15%, ≥ 10 pack-years
	Patuzzo 2000	Negative	Italy	FEV_1 < 70% pred, BD response < 12%
	Ferrarotti 2003	Negative	Italy	FEV_1 < 50% pred, FEV_1/FVC < 70% pred, BD response < 12%, $D_{L_{CO}}$ < 50% pred,
	Ishii 2000	Negative	Japan	Clinical diagnosis, including PFTs
Immune response				
Surfactant protein B (SFTPB)	Guo 2001	Initial positive	Mexico	FEV_1 < 70% pred, FEV_1/FVC < 0.7, > 10 pack-years
	Seifart 2002	Replicated	Germany	FEV_1 < 70% pred, FEV_1/FVC < 0.8, chronic bronchitis
Surfactant protein D (SFTPD)	Guo 2001	Initial positive	Mexico	FEV_1 < 70% pred, FEV_1/FVC < 0.7, > 10 pack-years
Vitamin D binding protein (group-specific complement, GC)	Horne 1990	Initial positive	Canada	Clinical diagnosis
	Schellenberg 1998	Replicated	Canada	FEV_1 < 80% pred, FEV_1/FVC < 0.7
	Ishii 2001	Replicated	Japan	FEV_1/FVC < 70%, ≥ 10 pack-years
	Ito 2004	Replicated	Japan	Clinical diagnosis, > 20 pack-years
	Kauffmann 1983	Negative	France	Never-smokers, low lung function, ≤ 40 yr old
	Kasuga 2003	Negative	United States & Canada (whites)	Low lung function among patients with COPD (Lung Health Study)
	Sandford 2001	Negative	United States & Canada (whites)	Rapid decline vs. no decline in FEV_1 over 5 yr in patients with COPD (Lung Health Study)

Definition of abbreviations: BD, bronchodilator; COPD, chronic obstructive pulmonary disease; $D_{L_{CO}}$, diffusing capacity for carbon monoxide; PFT, pulmonary function test.
* Complete citations for this table are available in the online supplement (Table E1).

NAS participants with a history of at least 10 pack-years of cigarette smoking and without airflow obstruction at their most recent visit (FEV_1 > 80% predicted [13] and FEV_1/FVC > 90% predicted [11]).

All studies were approved by the appropriate institutional review boards. Participants in the Boston Early-Onset COPD Study and the NETT Genetics Ancillary Study gave written informed consent. Anonymized data were used for the Normative Aging Study participants, as approved by the IRBs of Partners Healthcare System and of the Boston VA.

Candidate Genes and Genotyping

Twelve candidate genes with reported associations to COPD were identified from the published literature (Tables 1 and E1). Nineteen polymorphisms were initially genotyped in both the family-based and case-control studies (Table 2). Four SNPs (EPHX1 His139Arg, EPHX1 Tyr113His, GSTP1 Ile105Val, and tumor necrosis factor [TNF] −308G>A) were genotyped using the 5′ to 3′ exonuclease assay in TaqMan (Applied Bio-

TABLE 2. MARKERS GENOTYPED IN EARLY-ONSET COPD FAMILIES AND IN THE CASE-CONTROL STUDY

Gene	Polymorphism	Role	Presumed Function
SERPINA3	rs4934	Ala-15Thr	Signal peptide
	rs17473	Pro229Ala (Bonn-1)	Reduced serum α_1-antichymostrypsin levels
	rs1800463	Leu55Pro (Bochum-1)	
MMP1	rs1799750	G-1607GG	Increased gene expression
MMP9	D20S838	(CA) repeat	Reduced gene expression with increased length of repeat
TIMP2	rs2277698	Ser101Ser	
HMOX1	—	(GT) repeat	May affect gene expression
EPHX1	rs868966	Promoter	
	rs1877724	Intron	
	rs1051740	Tyr113His ("slow")	Reduced enzyme activity
	rs2292566	Exon, synonymous	
	rs2260863	Intron	
	rs2234922	His139Arg ("fast")	Increased enzyme activity
	rs1051741	Exon, synonymous	
	rs360063	3' UTR	
GSTP1	rs947894	Ile105Val	Altered enzyme activity
	rs1799811	Ala114Val	
GSTM1	—	Null	Absent enzyme activity
TNF	rs1800629	Promoter (-308G$>$A)	Higher baseline and induced expression
	rs361525	Promoter (-238)	
	rs1800610	Intron ($+489$)	
	rs3091257*	3' UTR	
	rs769178	3' UTR	
SFTPB	rs1130866	Thr131Ile	Alters glycosylation site
	D2S388	(CA) Repeat	
SFTPD	rs2243639	Thr180Ala	
	rs721917	Met31Thr	
GC	rs4588	Thr436Lys (Gc-2)	Decreased conversion into macrophage activating factor
	rs7041	Asp432Glu (Gc-1S)	

See Table E1 in the online supplement for references.

* Due to a high rate of Mendelian errors in the early-onset COPD families and deviation from HWE in the NAS control subjects, this SNP was not analyzed for association in either study population.

systems, Foster City, CA) (14), using ABI TaqMan Assays-on-Demand. The 1–base pair indel in MMP1 (rs1799750) was also genotyped using TaqMan. Details of this assay and the remaining SNP and STR assays are listed in Table E2. The remaining SNPs, including the ten additional TNF and EPHX1 SNPs below, were genotyped with unlabeled mini-sequencing reactions and mass spectrometry in Sequenom (San Diego, CA) (15). Sequenom genotyping protocols are available on our website (http://www.innateimmunity.net). For three STR markers, PCR was performed using fluorescent-labeled and unlabeled primers, and product sizes were assessed by capillary electrophoresis on an ABI 3100 machine. For D2S388 and D20S838, the primers listed in the public database were used (http://www.ncbi.nlm.nih.gov/). The HMOX1 STR assay was designed per Yamada and coworkers (16). Product sizes were compared with the human genome sequence (http://genome.ucsc.edu/) to determine the number of repeats. The glutathione S-transferase M1 (GSTM1) null deletion was assessed using a quantitative real-time PCR (TaqMan) assay (17). Primers and probes for GSTM1 and a two-copy gene control (BRCA1) were from the previous report (17).

Four additional SNPs in TNF (Table 2) were selected using a linkage disequilibrium tagging algorithm (http://www.innateimmunity.net), based on genotype data available from SeattleSNPs (http://pga.mbt.washington.edu/). Tag SNPs were selected using an LD threshold defined by $r^2 \geqslant 0.9$ and a minor allele frequency $\geqslant 0.05$. Six additional SNPs in EPHX1 (Table 2) were selected from public databases (http://snpper.chip.org/, http://snp500cancer.nci.nih.gov/).

Statistical Analysis

In the family-based study, deviations from Mendelian inheritance were tested with Pedcheck (18). The data were analyzed using the extended pedigree family-based association test, implemented in PBAT (http://www.biostat.harvard.edu/~clange/default.htm) (19). Quantitative (FEV$_1$ and FEV$_1$/FVC) and qualitative (mild-to-severe and moderate-to-severe airflow obstruction) traits were analyzed under additive genetic models, adjusting for age, sex, height, smoking status, and pack-years of cigarette smoking, including quadratic terms where appropriate. Dominant genetic models and gene-by-environment interactions were tested in secondary analyses.

The case-control data were analyzed in SAS/Genetics (SAS Institute, Cary, NC). Hardy-Weinberg equilibrium was assessed in control subjects using an exact test. Odds ratios and chi-square statistics were calculated from 2×2 tables of allele frequencies. Logistic regression was used to control for age and pack-years in both additive and dominant genetic models. Gene-by-smoking and gene-by-gene interactions were tested by including appropriate cross-product terms in the regression models. Haplotype analysis was performed using the expectation-maximization algorithm and score tests, implemented in haplo.stats (20).

In the case-control study, we had previously genotyped a panel of 44 unlinked SNPs to test for population stratification using the method of Pritchard and Rosenberg (21). There was no compelling evidence for overt population stratification ($\chi^2_{44\,d.f.} = 58.5$, $P = 0.07$) (22).

RESULTS

Study Subjects

Characteristics of participants in the Boston Early-Onset COPD Study are shown in Table 3A. The probands were predominantly female, as has been previously reported (8, 23). Probands had severe airflow obstruction (mean FEV$_1$ = 19.2% predicted). Details of the participants in NETT and NAS are found in Table 3B. The majority of NETT subjects were men (63.8%), while the NAS control subjects were all men. Ages of cases

TABLE 3A. CHARACTERISTICS OF PARTICIPANTS IN THE BOSTON EARLY-ONSET COPD STUDY

	Probands	First-Degree Relatives	Second-Degree Relatives	Spouses
Subjects	127	503	273	46
Female (%)	95 (74.8%)	287 (57.1%)	157 (57.5%)	14 (30.4%)
Age (± SD)	48.1 (± 4.7)	42.0 (± 17.9)	53.2 (± 19.5)	51.0 (± 5.8)
Pack-years (± SD)	38.9 (± 21.9)	17.3 (± 23.9)	19.5 (± 24.5)	34.0 (± 34.1)
FEV_1, % predicted (± SD)*	19.2 (± 7.4)	83.3 (± 21.9)	84.7 (± 19.0)	86.3 (± 20.4)
FEV_1/FVC, % predicted (± SD)*	39.1 (± 13.3)	86.2 (± 15.1)	89.7 (± 12.5)	90.2 (± 11.3)
Mild-to-severe airflow obstruction (%)*†	127 (100%)	133 (26.4%)	71 (26.0%)	10 (21.7%)
Moderate-to-severe airflow obstruction (%)*‡	127 (100%)	67 (13.3%)	27 (9.9%)	3 (6.5%)

* Spirometry data was not available for eight participants.

† Mild-to-severe airflow obstruction: $FEV_1 < 80\%$ predicted, with $FEV_1/FVC < 90\%$ predicted.

‡ Moderate-to-severe airflow obstruction: $FEV_1 < 60\%$ predicted, with $FEV_1/FVC < 90\%$ predicted.

TABLE 3B. CHARACTERISTICS OF SUBJECTS IN THE CASE-CONTROL STUDY

	Cases: NETT	Control Subjects: NAS	P value*
Total subjects	304	441	
Male (%)	194 (63.8%)	441 (100%)	< 0.0001
Age (± SD)	67.3 (± 6.0)	67.5 (± 8.5)	n.s.
Pack-years (± SD)	67.4 (± 31.6)	38.5 (± 26.5)	< 0.0001
FEV_1, % predicted (± SD)	24.8 (± 6.5)	92.2 (± 11.4)	< 0.0001
FEV_1/FVC, % predicted (± SD)	41.5 (± 8.2)	102.6 (± 6.8)	< 0.0001

Definition of abbreviations: NAS, Normative Aging Study; NETT, National Emphysema Treatment Trial; n.s., not significant.

* *t* Test or χ^2 test, as appropriate.

and control subjects were similar, but the NETT cases had a significantly greater smoking history (67.4 versus 38.5 pack-years, $P < 0.0001$). The 304 NETT cases had severe COPD (mean $FEV_1 = 24.8\%$ predicted).

Family-Based Study

In the Boston Early-Onset COPD Study families, all markers were tested for association with the postbronchodilator phenotypes, including FEV_1, FEV_1/FVC, mild-to-severe airflow obstruction ($FEV_1 < 80\%$ predicted, with $FEV_1/FVC < 90\%$ predicted), and moderate-to-severe airflow obstruction ($FEV_1 < 60\%$ predicted, with $FEV_1/FVC < 90\%$ predicted). Significant results from the extended pedigree family-based association test are shown in Table 4. The strongest associations were found with the TNF promoter $-308G>A$ SNP for both quantitative and qualitative COPD-related phenotypes (postbronchodilator), in additive genetic models, adjusting for age, sex, height, smoking status (ever versus never), and pack-years. Similar results were obtained using prebronchodilator spirometry phenotypes (data not shown).

A coding variant in surfactant protein B (SFTPB), Thr131Ile, was associated with the qualitative phenotype moderate-to-severe airflow obstruction, but a STR near SFTPB (D2S388) was not associated.

In addition, the 31-repeat (137-bp) allele (allele frequency in the extended pedigrees = 0.07) of the Heme oxygenase (HMOX1) promoter STR also was found to be associated with postbronchodilator FEV_1 and FEV_1/FVC. None of the other variants— two additional STRs, the MMP1 indel, and the other 12 SNPs— were found to be associated with qualitative or quantitative COPD-related traits in the early-onset COPD families. The GSTM1 deletion could not be analyzed in the family-based study design, as heterozygotes for the deletion could not be distinguished consistently from the wild-type. One marker, the α_1-antichymotrypsin Pro229Ala SNP (Bonn-1), was found to be monomorphic in the extended pedigrees and was not genotyped in the case-control study.

To assess for gene-by-environment interactions, the TNF $-308G>A$ and SFTPB Thr131Ile variants were examined in an

TABLE 4. EXTENDED PEDIGREE FAMILY-BASED ASSOCIATION TESTING IN BOSTON EARLY-ONSET COPD STUDY FAMILIES, USING ADDITIVE GENETIC MODELS

Phenotype (postbronchodilator)	Candidate Gene Polymorphisms		
	TNF $-308G>A$	SFTPB Thr131Ile	HMOX1 137-bp Allele
	(60 informative families)	(78 informative families)	(13 informative families)
FEV_1	0.004	n.s.	0.02
FEV_1/FVC	0.004	n.s.	0.02
Mild-to-severe airflow obstruction*	0.02	n.s.	0.08
Moderate-to-severe airflow obstruction†	0.04	0.03	0.05

P values < 0.1 are shown. None of the other markers listed in Table 1 showed significant association

* Mild-to-severe airflow obstruction: $FEV_1 < 80\%$ predicted, with $FEV_1/FVC < 90\%$ predicted.

† Moderate-to-severe airflow obstruction: $FEV_1 < 60\%$ predicted, with $FEV_1/FVC < 90\%$ predicted.

TABLE 5A. RESULTS OF CASE-CONTROL ASSOCIATION ANALYSIS: BIALLELIC GENOTYPES (SNPS, INSERTION/DELETION, NULL ALLELE)

Gene	Polymorphism	Allele Frequency: Cases	Allele Frequency: Control Subjects	*P* Value Allele Test	*P* Value Genotype Trend Test (adjusted)*
SERPINA3	Ala-15Thr	0.47	0.49	n.s.	n.s.
	Leu55Pro (Bochum-1)	0.01	0.005	n.s.	n.s.
MMP1	G-1607GG	0.49	0.48	n.s.	n.s.
TIMP2	Ser101Ser	0.12	0.10	n.s.	0.08
EPHX1	His139Arg	0.17	0.22	0.02	0.03
	Tyr113His	0.29	0.28	n.s.	n.s.
GSTP1	Ile105Val	0.35	0.30	0.05	0.1
	Ala114Val	0.08	0.09	n.s.	n.s.
GSTM1	null	0.58[†]	0.54[†]	n.s.	n.s.
TNF	−308G>A	0.15	0.17	n.s.	n.s.
SFTPB	Thr131Ile	0.46	0.44	n.s.	n.s.
SFTPD	Thr180Ala	0.40	0.43	n.s.	n.s.
	Met31Thr	0.43	0.41	n.s.	n.s.
GC	Thr436Lys (GC-2)	0.29	0.27	n.s.	n.s.
	Asp432Glu (GC-1S)	0.43	0.42	n.s.	n.s.

Definition of abbreviation: n.s., not significant.

P values ⩽ 0.1 are shown

* Additive genetic model, adjusted for age, pack-years.

[†] Frequency of GSTM1 null (−/−) genotype.

TABLE 5B. RESULTS OF CASE-CONTROL ASSOCIATION ANALYSIS: SHORT TANDEM REPEATS

Gene	Microsatellite	Allele	Allele Frequency: Cases	Allele Frequency: Control Subjects	*P* Value Trend Test (adjusted)*
MMP9	D20S838	106	0.56	0.56	n.s.
		120	0.19	0.17	n.s.
		122	0.12	0.14	n.s.
		124	0.05	0.06	n.s.
HMOX1	(GT) repeat	121	0.20	0.19	n.s.
		123	0.08	0.08	n.s.
		135	0.43	0.48	0.04
		137	0.08	0.07	n.s.
		149[†]	0.04	0.03	n.s.
SFTPB	D2S388	259	0.23	0.19	0.09
		263	0.36	0.38	n.s.
		265	0.11	0.14	n.s.
		269	0.24	0.26	n.s.

Definition of abbreviation: n.s., not significant.

Alleles with frequency > 0.05 are listed.

* Additive genetic model, comparing each allele to all others, adjusted for age, pack-years.

[†] For HMOX1 (GT) repeat, allele 149 (frequency = 0.033) included to represent "L" class in Yamada and colleagues (16).

analysis stratified by smoking status as well as in a model including an interaction term for pack-years. These analyses did not show evidence of gene-by-environment interaction effects.

In the analysis of the additional TNF SNPs in the early-onset COPD families, one SNP in the 3′ untranslated region (UTR), rs3091257, had a high rate of Mendelian errors. In addition, this SNP deviated from Hardy-Weinberg proportions in the NAS control subjects and was not analyzed further in either cohort. None of the three additional TNF SNPs were associated with COPD-related phenotypes in the family-based study.

In the early-onset COPD families, none of the eight SNPs in EPHX1 were found to be associated in the primary analysis. In a secondary analysis using a dominant genetic model, one intronic SNP (rs1877724) was marginally associated with postbronchodilator values for both FEV_1 ($P = 0.02$) and moderate-to-severe airflow obstruction ($P = 0.02$).

Case-Control Study

As noted above, the TNF rs3091257 SNP was out of Hardy-Weinberg Equilibrium (HWE) in the NAS control subjects and was removed from the analysis. All other SNPs and STRs tested in the study were found to be in HWE in the NAS control subjects. The results of the case-control association analysis of the biallelic markers (SNPs, insertion/deletion, and GSTM1 null allele) are shown in Table 5A. The only significant association was with a coding variant in microsomal epoxide hydrolase (EPHX1), His139Arg, referred to as the "fast" allele ($P = 0.03$ for an additive genetic model, adjusting for age and pack-years in a logistic regression); the fast allele appeared to be protective (odds ratio 0.73; 95% confidence interval 0.56, 0.96). The positive family-based association with TNF −308G>A was not replicated in the case-control study. The SFTPB Thr131Ile polymorphism was not significant in the primary case-control analysis. However, in a logistic regression model adjusting for age and pack-years that included and an interaction between SFTPB Thr131Ile and pack-years, both the main SNP effect (additive model) and the interaction term became significant ($P = 0.01$ for both main effect and interaction). In analyses of both additive and dominant genetic models, none of the other biallelic markers were pre-

dictive of COPD in the case-control study, in models that did and did not include gene-by-environment interactions.

The matrix metalloproteinase-9 (MMP9) STR (D20S838) was initially analyzed by grouping alleles into "Small" and "Large" repeat numbers, according to the method of Joos and colleagues (24); alleles $\leq$ 110 bp (16 repeats in the previous report) were classified as small, and those $\geq$ 112 bp (17 repeats) were considered large. The HMOX1 promoter polymorphism was analyzed as per Yamada and coworkers (16). Alleles were classified as Small ($<$ 129 bp, $<$ 27 repeats), Medium (129–139 bp, 27–32 repeats) and Large ($\geq$ 141 bp, $\geq$ 33 repeats). Neither STR marker was significantly associated in these analyses.

The MMP9, HMOX1, and SFTPB STRs were then analyzed by comparing each allele with a frequency of at least 0.05 to all other alleles, using both additive and dominant genetic models (Table 5B). The 30-repeat allele (135 bp, allele frequency in NETT cases = 0.43) of the HMOX1 STR was significantly associated in the case-control analysis (adjusted $P = 0.04$, additive genetic model), whereas the 31-repeat allele (137 bp, allele frequency in Boston Early-Onset COPD Study families = 0.07) had been significant in the family-based study. The 135-bp allele was the most common repeat size and was underrepresented in the cases (allele test, odds ratio = 0.84; 95% CI, 0.68–1.04).

None of the three additional SNPs in TNF were significant in the case-control study; no evidence of gene-by-smoking interaction was found in this analysis. In the case-control study, none of the additional SNPs in EPHX1 was found to be associated with COPD. However, in a model that included a gene-by-smoking (pack-years) interaction, both the main effect ($P = 0.02$) and the interaction ($P = 0.03$) were significant for a silent coding variant in exon 3 (rs2292566). This SNP was not in linkage disequilibrium with the fast allele in exon 4 that was found to be associated with COPD ($r^2 < 0.1$ in NAS control subjects).

Gene-by-Gene Interactions

Previous authors have examined gene-by-gene interactions in COPD, specifically among genes involved in xenobiotic metabolism (25, 26). The GSTM1 deletion was tested in two-way interactions with GSTP1 Ile105Val, the EPHX1 slow allele, and the EPHX1 fast allele in the case-control study. In additive models for the SNPs, none of the interactions was significant, in models including or excluding the main SNP effects. Similarly, three-way interactions between the GSTM1 deletion, GSTP1 Ile105Val, and either the EPHX1 fast or slow alleles were not significant.

We also tested for interactions between the EPHX1 fast polymorphism, TNF −308 G>A, and SFTPB Thr131Ile, the SNPs that had been significant in either the family-based or case-control study. None of the two-way interactions nor the three-way interaction were significant.

Haplotype Analysis

In the case-control population, two-SNP haplotype analyses in SFTPD and GC were not significant, though there was a trend toward association for GSTP1 (global score test, $P = 0.06$). Haplotype analyses in TNF, using all four SNPs as well as using two SNP sliding windows, were not significant. The eight-SNP haplotype analysis in EPHX1 was not significant, and none of the two- or four-SNP sliding window haplotypes of EPHX1 was associated with COPD.

DISCUSSION

In this study, we selected 29 polymorphisms in 12 genes that had been reported to be associated with COPD in the published literature and genotyped these variants in a family-based study of early-onset COPD and a case-control COPD study. The most

significant association in the family-based study (TNF −308G>A) was not replicated in the case-control study, and the strongest association in the case-control study (EPHX1 fast allele) was not found in the family-based analysis. Two variants showed modest evidence for replication across both study designs. A coding SNP in surfactant protein B (Thr131Ile) that was marginally associated ($P = 0.03$) with one qualitative trait (moderate-to-severe airflow obstruction) in the Boston Early-Onset COPD families was not associated in the primary case-control analysis, but did show association ($P = 0.01$) when an SNP-by-smoking interaction was included. An STR in HMOX1 was significant in both studies, though different alleles were associated in each cohort.

The associations with SFTPB and HMOX1 merit further investigation. The different effects of gene-by-smoking interaction in the analyses of SFTPB Thr131Ile in the family-based and case-control studies and the different alleles of the HMOX1 repeat driving the associations in the two study designs suggest that these polymorphisms are not the functional variants affecting COPD susceptibility. The effects that we detected may be due to linkage disequilibrium with nearby functional variants. Analysis of additional SNPs in these genes will be required to confirm these genetic associations. Despite our positive results, we cannot exclude that these may be spurious associations due to the multiple comparisons performed.

Several explanations have been proposed to explain the lack of replication that is commonly seen in case-control association studies in complex trait genetics (5, 6). Small sample sizes may lead to inadequate power to detect an association in the initial study or to replicate true associations in subsequent studies. In fact, the majority of the COPD candidate gene association studies listed in Table 1 enrolled fewer than 100 cases and 100 control subjects. Insufficient power is not likely to explain the lack of replication seen in our study. Using the example of TNF −308G>A with a 17% minor allele frequency in the NAS control subjects, our case-control study had 90% power ($\alpha = 0.05$) to detect an odds ratio of 1.8, in an additive model; this odds ratio is less than reported in either of the two published COPD association studies with significant results (27, 28). However, if the true odds ratio were lower than 1.8 in other populations, then the power to detect significant associations would be reduced.

Spurious associations may result from multiple testing in studies that assess many genes, markers, and phenotypes (29). No consensus exists on the optimal method to adjust for multiple testing in case-control genetic association studies, though replication in an independent study may provide the strongest evidence for true association. Multiple testing was a potential problem in our family-based study, given the multiple genes and phenotypes tested, though the independent case-control sample provided an opportunity to confirm the findings from the family-based study.

Genotyping errors usually bias toward no association, though systematic errors may lead to false positive results. Deviation from Hardy-Weinberg equilibrium (HWE) in the control group may be a sign of genotyping error (29). We found that only one of the markers tested deviated from HWE, and that SNP was excluded from the association analyses. Departure from Mendelian transmission of alleles is another indication of genotyping error that is only applicable to family-based studies; besides the excluded SNP above, only a small number of Mendelian inconsistencies were found in our study.

Failure to demonstrate HWE may also be a sign of population stratification, which refers to differences in allele frequency between cases and controls due to ethnic differences and not due to disease status (30). Population stratification can lead to spurious association in case-control studies (21). Careful matching of cases and control subjects on ethnicity provides some protection

against population stratification. Several statistical methods, based on genotype data from additional unlinked markers elsewhere in the genome, are available to test for stratification and control for its effects if present (21, 31, 32). None of the published COPD genetic association studies have employed these formal tests. We tested a modest sized panel of SNPs in our case-control study and found little evidence for stratification.

The issues described may lead to false positive (multiple testing, population stratification) or false negative (small sample size, genotyping error) results. However, true differences may lead to inconsistent results. COPD is a heterogeneous disease and published association studies have used different phenotype definitions. For example, studies of TNF have defined cases on the basis of airflow obstruction (28), emphysema (33), decline in lung function (34), or chronic bronchitis (27). It is possible that a given genetic variant may confer susceptibility to a specific COPD-related phenotype. In our case-control study, the NETT cases all had emphysema confirmed by chest CT scan. Radiographic evidence of emphysema was not a requirement for entry into the Boston Early-Onset COPD Study, though many probands did have chest CT scans showing emphysema (8). In our family-based study we analyzed quantitative and qualitative traits, based on spirometry, but the case-control study used COPD diagnosis as a binary outcome. However, we used strict spirometric criteria to define cases and controls, so the overall conclusions should not be affected. The power may be greater using quantitative versus qualitative traits, however.

For the majority of the genes studied, we genotyped only one or two markers per gene, as has been done in most of the previously reported studies. This method relies on the assumption that the variants tested have functional effects on COPD susceptibility. If another variant in or near the gene were the causal variant, then the true association could be easily missed. Different linkage disequilibrium patterns with the functional variant may lead to variable results in different populations. In two genes, TNF and EPHX1, we tested additional SNPs and used haplotype analysis to study these genes more thoroughly. However, this did not strengthen our findings.

Genetic heterogeneity may also explain the varying results among case-control association studies, especially those done in different ethnic groups. Many of the COPD association studies in Table 1 have shown inconsistent results in white and Asian populations. True differences may be the result of different genetic determinants of disease in diverse populations, variation in gene–environment interaction due to specific environmental exposures, or different patterns of linkage disequilibrium between the tested marker and the causal variant (35). Though most of our study subjects were whites from the United States, it is possible that severe early-onset COPD represents a unique disease subtype with different genetic determinants than the usually seen, later-onset COPD. However, many of the family members in the Boston Early-Onset COPD Study had less severe airflow obstruction, consistent with more usual forms of COPD. Nevertheless, variants in several genes studied, including TNF-α and surfactant protein B, may primarily increase susceptibility to severe early-onset COPD. These results should be interpreted with caution due to the multiple tests performed; replication in an independent cohort is still required.

This study highlights the major difficulty with using a candidate gene approach to uncover susceptibility genes for COPD, namely the lack of replication commonly seen in candidate gene studies. Future candidate gene association studies need to employ rigorous genetic epidemiology methods, including adequate sample sizes, control for multiple testing, and testing for population stratification. A more systematic approach to COPD genetics, starting with genome-wide linkage analysis followed by positional candidate gene association testing and/or SNP-based fine mapping, may lead to more consistent results in the search for genetic determinants of COPD.

References

1. Ridker PM, Hennekens CH, Lindpaintner K, Stampfer MJ, Eisenberg PR, Miletich JP. Mutation in the gene coding for coagulation factor V and the risk of myocardial infarction, stroke, and venous thrombosis in apparently healthy men. *N Engl J Med* 1995;332:912–917.
2. Saunders AM, Strittmatter WJ, Schmechel D, George-Hyslop PH, Pericak-Vance MA, Joo SH, Rosi BL, Gusella JF, Crapper-MacLachlan DR, Alberts MJ. Association of apolipoprotein E allele epsilon 4 with late-onset familial and sporadic Alzheimer's disease. *Neurology* 1993;43:1467–1472.
3. Altshuler D, Hirschhorn JN, Klannemark M, Lindgren CM, Vohl MC, Nemesh J, Lane CR, Schaffner SF, Bolk S, Brewer C, *et al.* The common PPARgamma Pro12Ala polymorphism is associated with decreased risk of type 2 diabetes. *Nat Genet* 2000;26:76–80.
4. Freely associating. *Nat Genet* 1999;22:1–2.
5. Hirschhorn JN, Altshuler D. Once and again-issues surrounding replication in genetic association studies. *J Clin Endocrinol Metab* 2002;87: 4438–4441.
6. Ioannidis JP, Ntzani EE, Trikalinos TA, Contopoulos-Ioannidis DG. Replication validity of genetic association studies. *Nat Genet* 2001;29: 306–309.
7. Hersh CP, DeMeo DL, Lange C, Chapman HA, Reilly JJ, Sylvia JS, Speizer FE, Weiss ST, Silverman EK. Attempted replication of reported COPD associations using a family-based study design [abstract]. *Am J Respir Crit Care Med* 2004;169:A506.
8. 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.
9. The National Emphysema Treatment Trial Research Group. Rationale and design of The National Emphysema Treatment Trial: a prospective randomized trial of lung volume reduction surgery. *Chest* 1999;116: 1750–1761.
10. Fishman A, Martinez F, Naunheim K, Piantadosi S, Wise R, Ries A, Weinmann G, Wood DE. A randomized trial comparing lung-volume-reduction surgery with medical therapy for severe emphysema. *N Engl J Med* 2003;348:2059–2073.
11. Crapo RO, Morris AH, Gardner RM. Reference spirometric values using techniques and equipment that meet ATS recommendations. *Am Rev Respir Dis* 1981;123:659–664.
12. Bell B, Rose CL, Damon H. The Normative Aging Study: an interdisciplinary and longitudinal study of health and aging. *Aging Hum Dev* 1972;3:5–17.
13. O'Connor GT, Sparrow D, Weiss ST. A prospective longitudinal study of methacholine airway responsiveness as a predictor of pulmonary-function decline: the Normative Aging Study. *Am J Respir Crit Care Med* 1995;152:87–92.
14. Holland PM, Abramson RD, Watson R, Gelfand DH. Detection of

specific polymerase chain reaction product by utilizing the 5′—3′ exonuclease activity of Thermus aquaticus DNA polymerase. *Proc Natl Acad Sci USA* 1991;88:7276–7280.

15. Sun X, Ding H, Hung K, Guo B. A new MALDI-TOF based minisequencing assay for genotyping of SNPS. *Nucleic Acids Res* 2000;28: E68.

16. Yamada N, Yamaya M, Okinaga S, Nakayama K, Sekizawa K, Shibahara S, Sasaki H. Microsatellite polymorphism in the heme oxygenase-1 gene promoter is associated with susceptibility to emphysema. *Am J Hum Genet* 2000;66:187–195.

17. Welch R, Yadavalli S, Puti V, Rothman N, Chanock SJ, Yeager M. A high-throughput, real-time PCR (TaqMan) method for detecting the heterozygous state for GSTM1 and GSTT1 [abstract]. *Am J Hum Genet* 2003;73Suppl:247.

18. O'Connell JR, Weeks DE. PedCheck: a program for identification of genotype incompatibilities in linkage analysis. *Am J Hum Genet* 1998;63:259–266.

19. Lange C, DeMeo D, Silverman EK, Weiss ST, Laird NM. PBAT: tools for family-based association studies. *Am J Hum Genet* 2004;74:367–369.

20. Schaid DJ, Rowland CM, Tines DE, Jacobson RM, Poland GA. Score tests for association between traits and haplotypes when linkage phase is ambiguous. *Am J Hum Genet* 2002;70:425–434.

21. Pritchard JK, Rosenberg NA. Use of unlinked genetic markers to detect population stratification in association studies. *Am J Hum Genet* 1999; 65:220–228.

22. Celedon JC, Lange C, Raby BA, Litonjua AA, Palmer LJ, DeMeo DL, Reilly JJ, Kwiatkowski DJ, Chapman HA, Laird N, *et al.* The transforming growth factor-{beta}1 (TGFB1) gene is associated with chronic obstructive pulmonary disease (COPD). *Hum Mol Genet* 2004;13: 1649–1656.

23. Silverman EK, Weiss ST, Drazen JM, Chapman HA, Carey V, Campbell EJ, Denish P, Silverman RA, Celedon JC, Reilly JJ, *et al.* Gender-related differences in severe, early-onset chronic obstructive pulmonary disease. *Am J Respir Crit Care Med* 2000;162:2152–2158.

24. Joos L, He JQ, Shepherdson MB, Connett JE, Anthonisen NR, Pare PD, Sandford AJ. The role of matrix metalloproteinase polymorphisms in the rate of decline in lung function. *Hum Mol Genet* 2002;11:569–576.

25. Yim JJ, Park GY, Lee CT, Kim YW, Han SK, Shim YS, Yoo CG. Genetic susceptibility to chronic obstructive pulmonary disease in Koreans: combined analysis of polymorphic genotypes for microsomal epoxide hydrolase and glutathione S-transferase M1 and T1. *Thorax* 2000;55: 121–125.

26. He JQ, Ruan J, Connett JE, Anthonisen NR, Pare PD, Sandford AJ. Antioxidant gene polymorphisms and susceptibility to a rapid decline in lung function in smokers. *Am J Respir Crit Care Med* 2002;166:323–328.

27. Huang SL, Su CH, Chang SC. Tumor necrosis factor-alpha gene polymorphism in chronic bronchitis. *Am J Respir Crit Care Med* 1997;156:1436–1439.

28. Sakao S, Tatsumi K, Igari H, Shino Y, Shirasawa H, Kuriyama T. Association of tumor necrosis factor alpha gene promoter polymorphism with the presence of chronic obstructive pulmonary disease. *Am J Respir Crit Care Med* 2001;163:420–422.

29. Silverman EK, Palmer LJ. Case-control association studies for the genetics of complex respiratory diseases. *Am J Respir Cell Mol Biol* 2000;22: 645–648.

30. Cardon LR, Palmer LJ. Population stratification and spurious allelic association. *Lancet* 2003;361:598–604.

31. Pritchard JK, Stephens M, Rosenberg NA, Donnelly P. Association mapping in structured populations. *Am J Hum Genet* 2000;67:170–181.

32. Devlin B, Roeder K. Genomic control for association studies. *Biometrics* 1999;55:997–1004.

33. Ferrarotti I, Zorzetto M, Beccaria M, Gile LS, Porta R, Ambrosino N, Pignatti PF, Cerveri I, Pozzi E, Luisetti M. Tumour necrosis factor family genes in a phenotype of COPD associated with emphysema. *Eur Respir J* 2003;21:444–449.

34. Sandford AJ, Chagani T, Weir TD, Connett JE, Anthonisen NR, Pare PD. Susceptibility genes for rapid decline of lung function in the Lung Health Study. *Am J Respir Crit Care Med* 2001;163:469–473.

35. Hirschhorn JN, Lohmueller K, Byrne E, Hirschhorn K. A comprehensive review of genetic association studies. *Genet Med* 2002;4:45–61.